THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XXXV JULY 1956 NUMBER 4

The Effect of Surface Treatments on Point-Contact Transistor Characteristics

By J. H. FORSTER and L. E. MILLER

(Manuscript received January 23, 1956)

A description is given of the electrical properties of formed point contacts on germanium. A useful technique for observation of the equipotentials surrounding such contacts is described. The contrasting properties of donor-free and donor-doped contacts, used as diodes or transistor collectors are emphasized.

It is shown that unformed point contacts (which have electrical properties largely determined by a surface barrier layer), may exhibit analogous differences. Such changes are produced by chemical treatments calculated to influence properties of a soluble germanium oxide film on the surface.

The above information is applied to a study of transistor forming as it is done in present point-contact transistor processing. It is shown that high yields from the forming process can be expected on oxidized surfaces, and that chemical washes which remove soluble germanium oxide drastically lower forming yields. These and other effects are evaluated as sources of variability in forming yield.

TABLE OF CONTENTS

1. INTRODUCTION

The point-contact transistor, on the basis of several years use in the field in Bell System applications, has proved itself to be rugged and dependable. For certain military applications, a lasting demand exists for high-speed point-contact transistors. The adaptation of cartridge type units to a hermetically sealed structure has been completed, with further benefits to reliability. To date, the point-contact transistor is one of the few transistors to successfully pass all military specifications for shock, vibration, and high acceleration. Thus, although there are at present limitations to the electrical characteristics that can be built into a point-contact transistor which make it unsuitable for use in some switching circuits, there are many applications in which this type of transistor can give consistent and reliable performance. In fact, applications exist wherein the specific requirements are uniquely satisfied by the point-contact transistor.

However, the basic operational principles of this kind of device are not as well understood as would be desirable for facilitating developmental studies for manufacture. Although considerable effort has been

expended towards the analysis and understanding of the physical mechanisms of the point-contact transistor since its announcement in 1948, a complete design theory for these transistors is not available. This lack probably results partially from a more general interest in the readily designable junction transistor types, and partially from the relative complexity of the device itself. Actually the physical mechanisms which account for the operation of this device have their counterparts in at least three basically unique devices: the point diode, the junction transistor, and the filamentary transistor.

Thus, although the empirical knowledge of point-contact transistor design and operation is large enough to allow a reasonable degree of designability, and manufacture of these transistors in large quantities is possible, there are, from time to time, manufacturing problems which are often difficult to solve without sound theoretical understanding of the physical mechanisms which make the device work.

This article is concerned with describing the results of a general study of the physical properties of a few specific kinds of point contacts. The kinds of contact studied have been those of specific interest to those concerned with manufacture and processing of point-contact transistors. This investigation was conducted in parallel with the final development for manufacture of the hermetically sealed point-contact transistor. The study of these properties has led to practical solutions of several problems encountered during manufacture of point-contact transistors, and has provided experimental data which is of interest in consideration of the basic physical mechanisms involved in the operation of the point-contact transistor.

The work to be described, primarily experimental in nature, follows in Sections 2, 3 and 4. In section 2, the properties of formed, or electrically pulsed point contacts, and their relation to the source of output characteristic anomalies often responsible for lowering forming yields in point-contact transistor production is discussed. The properties of point contacts which have received no electrical forming in the conventional sense are considered in section 3. The electrical properties of these contacts, used as diodes or transistor collectors, are shown to be dependent on chemical history of the etched germanium surface. Thus "chemical forming" of point contacts is possible. Section 4 deals with application of these results to forming problems which arise during manufacture of point-contact transistors. The important relation between the chemical history of the surface and the forming on that surface is considered.

2. PROPERTIES OF FORMED POINT CONTACTS

2.1 *Effects of Electrical Forming on Point Contacts*

The simplest form of point-contact transistor collector is a metal to semiconductor contact which has not been subjected to excessive power dissipation either in short high energy pulses, or in the form of more prolonged aging at lower power levels. Such contacts will be referred to as unformed contacts, and their properties will be discussed in detail in Section 3. Unformed point-contact transistors sometimes exhibit power gain, but in general they are not suitable for use as active devices because the gain, although it may be highly variable from unit to unit, is usually low. The electrical characteristics of such contacts depend on a metal-semiconductor contact at the semiconductor surface, and control of these properties requires exacting control of surface preparation, surrounding ambient, and mechanical stability of the point.

In early experiments, Brattain[1] used electrical forming to improve both the power gain and stability of the transistor. For present purposes, the process of electrical forming will be defined as the passage of a short pulse of reverse current through a point contact which produces permanent changes in the electrical properties of the contact. This is usually accomplished by charging a condenser to several hundred volts, and subsequently discharging it through a resistor in series with the transistor collector. Bardeen and Pfann,[2] investigating electrical forming of phosphor bronze points on etched germanium surfaces, indicate, as a possible explanation of their data, that the forming pulse changes the height of the potential barrier at the germanium surface. This would, in absence of large surface conductivity, increase the reverse current through the point and increase the efficiency of hole collection by the point.[3] Thus, the formed point may, according to theory,[4] act as a collector with a current multiplication (α) greater than unity. Thermal and potential probing of an n-germanium surface under a formed phosphor bronze point indicates, according to Valdes,[5] that an appreciable volume of germanium is converted to p-type conduction. Thus, the reverse current through a formed point probably depends on the characteristics of a p-n junction a small distance from the point, rather than on a potential barrier at the germanium surface.

A characteristic of the point-contact transistor is that the current gain can be substantially greater than unity. The current gain, α, is usually defined as the current multiplication at constant voltage, that is:

$$\alpha = -\frac{\partial I_c}{\partial I_e}\bigg|_{v_c} \tag{1}$$

where I_c and I_e are the collector and emitter currents. The α can be considered as the product of three terms, that is:

$$\alpha = \alpha_i \beta \gamma \tag{2}$$

where γ and β represent the injection efficiency and transport factor respectively for minority carriers. The term α_i is the "intrinsic" current multiplication of the collector itself. As mentioned above, there are theoretical reasons to account for an α_i as large as $(1 + b)$, where b is the ratio μ_n/μ_p of the mobilities of electrons and holes, and thus the term α_i may be roughly as large as 3.1. The average current gain, $\bar{\alpha}$, taken over a large interval of emitter current, is seldom found to be greater than this value, and is usually about 2.5. However, the small signal α at low emitter current usually is found to be considerably larger than 3.1.

Several mechanisms have been proposed to account for this excess current gain at low emitter bias in formed transistors. The most generally known of these are the p-n hook hypothesis and the trapping model.[6, 7]

The experiments to be described in this section will be concerned primarily with the characteristics of formed points as transistor collectors, and thus with the transport factor β. The subject of the origin of the intrinsic α_i will be discussed further in a later section.

The experiment of Valdes indicates that the properties of a formed point contact depend on the physical properties of a small region of germanium near the point, produced by impurity diffusion from the point or imperfections introduced during the forming pulse. A highly idealized representation of the physical situation is shown in Fig. 1. This is a radial model of a formed point contact on a semi-infinite block of n-germanium (respectively ρ), with a hemispherical p-layer (radius $\simeq r_0$). The electron and hole concentrations in the formed layer near the junction are designated as n_p and p. If a reverse bias V_c is applied to the point, a potential difference $V(r_1) - V(r_2) = V_J$ results from the resistance of the junction at r_0. For $r \geqq r_2$, at distances well outside r_0, the potential $V(r)$ and the magnitude of the field $E(r)$ are given by

$$V = \frac{\rho I}{2\pi r} \tag{3}$$

$$E = \frac{\rho I}{2\pi r^2} \tag{4}$$

where I is the total current through the point. For

$$| V_c - V(r_1) | \ll | V_J |, \; V(r_2) \simeq V_c - V_J,$$

and the junction resistance limits the magnitude of the drift field that can be set up near the point. For example, if the lifetime τ_n of electrons in the p-layer[8] is substantially lower than τ_p, that of holes in the germanium bulk, the reverse current density across the junction can be increased by an increase in n_p, and junction resistance lowered.

Pfann[9] reports a substantial increase in the reverse current of formed point contacts with donor concentration of the point wire. The increase in n_p will depend on the distribution of donors in the p-layer after the forming pulse. A high donor concentration near the collector point may

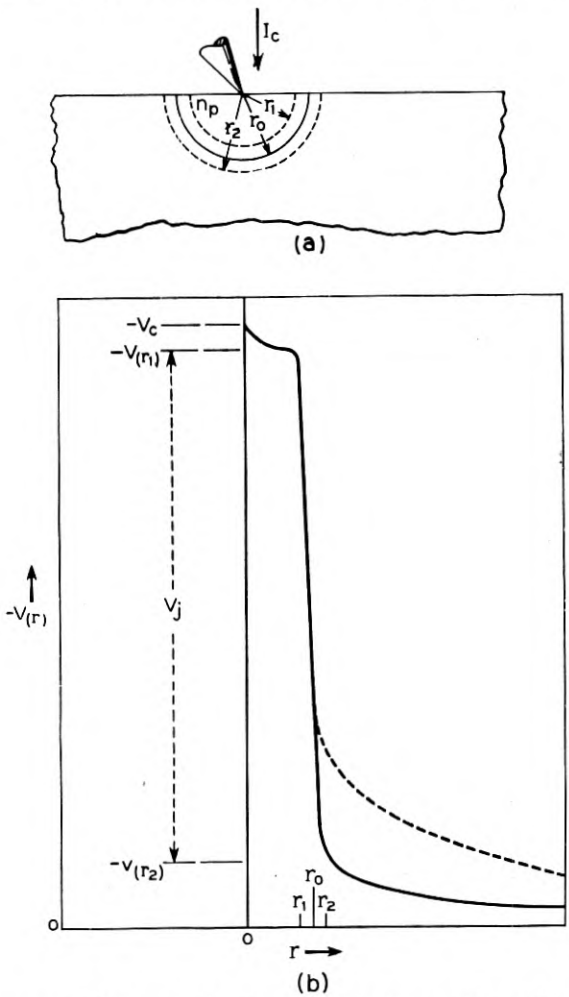


Fig. 1 — Formed point contact under reverse bias — schematic representation.

form an n-type inversion layer under the point (p-n hook) which, when the point is under reverse bias, acts as an electron emitter. Such a situation might arise as a result of diffusion of impurities from the collector point at the high temperature reached during the forming pulse. An acceptor element, such as copper, with a high diffusion coefficient[10] might penetrate substantially farther into the germanium than donor elements such as phosphorous or antimony[11] with lower diffusion constants. Thus, the donor concentration near the point might be substantially higher than the acceptor concentration if the solubility of the acceptor element is low.

On the other hand, an appreciable number of donor atoms may penetrate the germanium as far as do the acceptors. Thus, the equilibrium value of n_p may be increased simply by decreasing the effective concentration of acceptors in the p-layer. Such a case might arise when a collector point such as copper is doped with a suitable amount of a donor element with a large diffusion coefficient and limited solubility.

The observation of regions of melted germanium[12] under heavily formed points gives evidence for a somewhat different interpretation of the forming process. It has been suggested[13] that forming is essentially a remelt process. For example, forming of a phosphor-bronze point may produce a copper-germanium eutectic, allowing the introduction of a sizeable phosphorus concentration in the remelt region which is maintained after freezing. Thus the depth of penetration of the donor element depends upon the size of the remelt region, and the penetration of the acceptor element depends upon its solid state diffusion coefficient. This mechanism can lead either to the formation of a p-n hook, or at least to a lyer of p-germanium with a high equilibrium electron concentration.

Whatever the reason for the decrease in resistance of the collector barrier, if it is sufficient, the magnitude of $E(r)$ for $r > r_2$ can be increased by forming to sufficient value to ensure efficient collection of holes and a transport factor β close to unity.

It would then be expected that for a formed donor-free point, such as the beryllium-copper alloy points often used as unformed emitters, the formed p-region would have a high acceptor concentration, n_p would be small, and under reverse bias, the magnitude of V_J would be large, with $|I_{co}|$, $|V(r_2)|$, and average α small, [solid curve, Fig. 1(b). On the other hand, a formed phosphor bronze point of the kind conventionally used to make transistor collectors, should exhibit under reverse bias, a lesser magnitude of V_J, with $|I_{co}|$, $|V(r_2)|$, and α as much as an order of magnitude larger, (dashed line in Fig. 1(b)].

2.2 *Donor-Free and Donor-Doped Contacts*

The qualitative picture of the conventional formed contact given above has been substantially supported by the work of Valdes, who observed a large increase in floating potential near the reverse biased collector after the forming pulse and a substantial p-region in the bulk of the germanium after forming.

Experiments have been directed to a comparison of the properties as diodes and collectors, between two kinds of points. Phosphor bronze points of the type used as transistor collectors, and beryllium copper points, normally used as emitters, were investigated. Thus a direct comparison can be made between donor-doped and donor-free points which have been given similar forming pulses. The forming pulses were of the capacitor discharge type, with voltage and RC values similar to those used in conventional transistor forming. The points used were of the cantilever variety, and the n-germanium was zone-leveled material in the 3 to 4 ohm-cm range. Two points were supported in a double-ended micro-manipulator which allowed freedom of movement in 3 dimensions for each point.

2.2.1 *Potential Probes*

Conventionally, point-contact transistors are made on a superoxol-etched wafer. This etch leaves a rough surface which is unsuitable for accurate potential probing. Some measurements were made of the floating potentials on this kind of surface, but accurate results were difficult to obtain. In a later section it is shown that the kind of etch used in surface preparation can have profound effects on the degree of forming obtained. However, it is shown that forming characteristics of an "aged" CP_4-etched surface are quite similar to the superoxol surface. Thus this kind of surface was used, since its topographical uniformity allows very reproducible results in the measurement of floating potentials.

Fig. 2 is a comparison of the floating potentials for the two kinds of transistor points examined. The log-log plot shows the magnitude of the floating potential, V_p, near the reverse biased collector as a function of r, the distance of the probe from the collector measured between centers of the two points. The bars represent the uncertainty in measurement of the linear distance. Three curves are shown. The lowest Curve I represents the potential near a Be-Cu point formed with a conventional forming pulse. Curve II is a plot of the potential near a similarly formed phosphor bronze point, while Curve III represents data obtained using such a point more heavily formed. In all cases the magnitude of the floating potential decreases inversely as the distance from the point, and is given

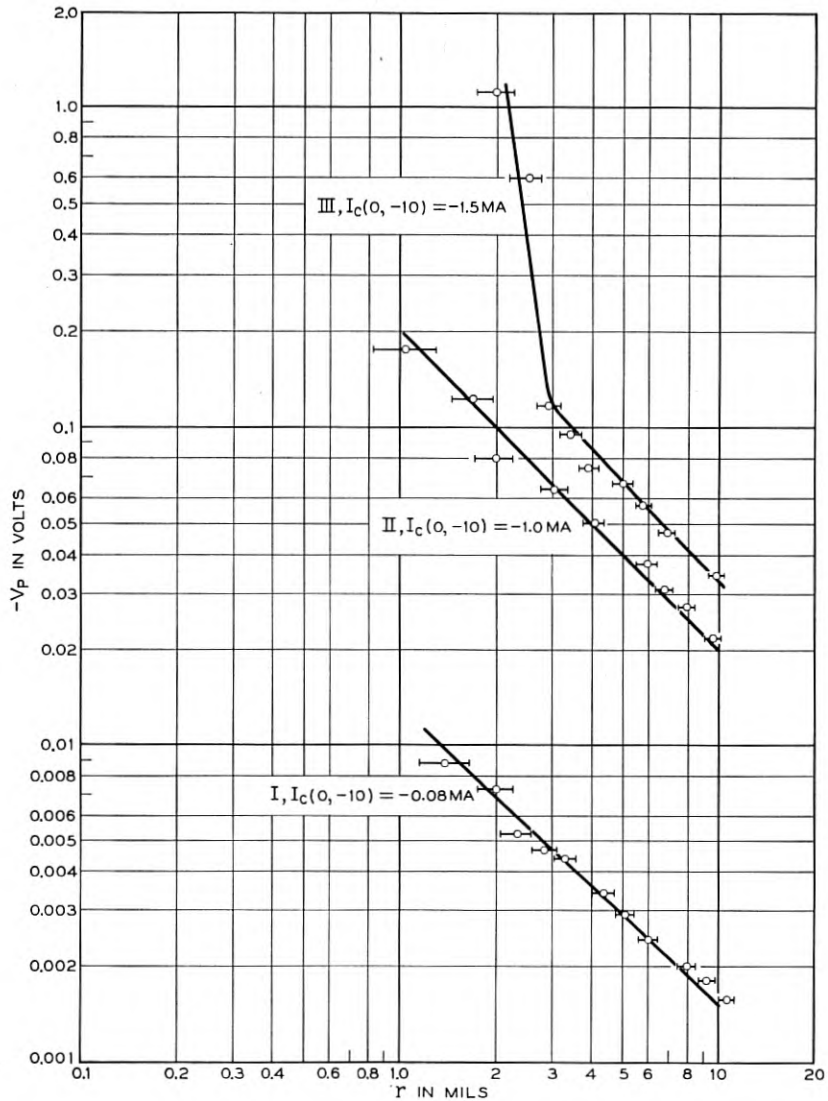


Fig. 2 — Comparison of floating potentials near formed points.

by $\rho I/2\pi r$ where ρ is well within the range of the measured resistivity (3–4 ohm-cm).

Thus the effect of adding the donor to the point wire is to increase the reverse current and increase the floating potential near the point by an order of magnitude. One would therefore expect an accompanying

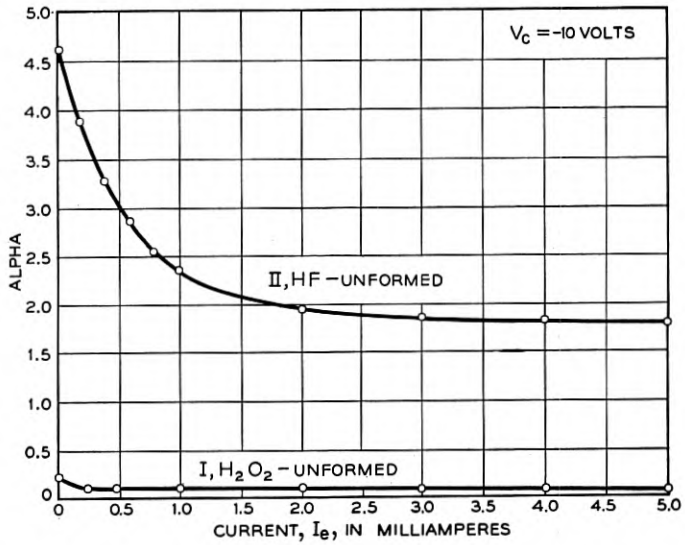


Fig. 3 — Comparison of alpha-emitter-current characteristics of formed points.

increase in the drift field near the point and a corresponding increase in α. Fig. 3 indicates that such is the case. The small signal α is plotted as a function of emitter current in Curves I and II. The point spacing in this case is 2.5 mils. It is interesting to note that the peak at low emitter currents is present in both cases, in spite of the fact that presence of a p-n hook is not likely when the Be-Cu point is formed.

It is thus apparent that the forming the Be-Cu point produces a structure which more closely resembles a p-n junction. The effect of adding the donor is to reduce the resistance of the junction. Further contrast between these two kinds of contacts is demonstrated by comparing forward currents through the contacts and their capacities. In Table I, a summary of all the contrasting properties is given. All values quoted are representative values.

2.2.2 *Use of the Copper Plating Technique*

During the investigation of these contact properties, an interesting way of illustrating their physical properties was developed. This technique, borrowed from junction transistor technology, can be used to identify visually the boundary between the formed region and the bulk germanium in a metallographic section of a point-contact transistor. It further appears that modifications of the technique will enable determination of

TABLE I

Contact	Formed Be Cu	Formed Phosphor Bronze
$I_c(0, -10)$ ma	−0.01 ma	− 1.0 ma
$I_c(6, -5)$ ma	−1.0 ma	−14.0 ma
$I_c(0, +0.5)$ ma	2.8 ma	0.8
Peak value of α	0.25	4.5
α (5.0, −10)	0.1	1.7
Capacity ($Vc = -5V$)	3.0 $\mu\mu$f	< 0.1 $\mu\mu$f

the equipotentials surrounding a collector or emitter point under bias, and visualization of current flow patterns in point contact transistors under bias operating conditions.

Use of this technique in identification of formed transistor properties is quite simple. A transistor container (including only the completed header, wafer, and point-contact structure) is filled with araldite plastic, which is allowed to harden. The collector point is then electrically formed. The plastic is necessary to ensure that the collector point does not subsequently move from the formed area. The can itself is then embedded in a plastic block, which is lapped down to expose a cross section of the unit, Fig. 4(a) and (b). Both the collector point and the base electrode are well masked, Fig. 5. A droplet of $CuSO_4$ solution of fairly low concentration is placed on the germanium, so that it is in physical contact only with with the germanium and the masking plastic. In order to identify the formed region, a reverse bias of 20 volts or so is applied between the collector point and the base contact for a time usually of 0.1 second or less. Actually, best results have been obtained by applying the reverse

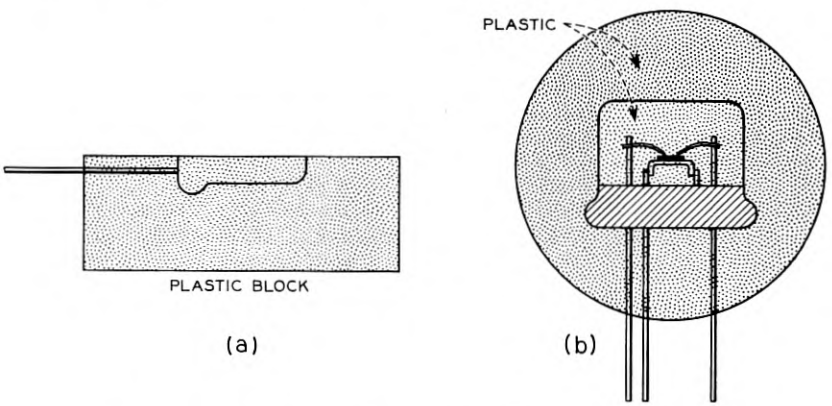


Fig. 4 — Preparation of a transistor for copper plating.

bias in the form of a condenser discharge pulse. Care must be taken to avoid changes in contact characteristics resulting from the plating pulse. The deposit of copper does not appear instantly after pulse application, but may require several seconds before becoming visible. At the instant the deposit becomes visible, the plating solution is washed off.

Fig. 6(a) and 6(b) show the results of the plating operation on a formed collector point and a formed emitter point. Both pulses were similar to, though somewhat "heavier" than those usually used to form transistors. These units were plated under the conditions illustrated in Fig. 5(a). The floating potential in the vicinity of the reversed bias point can be measured as a function of the distance, r, from its center, using an auxiliary tungsten point. Qualitatively this potential is shown as a function of the distance, r, in Fig. 5(b). In this case most of the drop in magnitude of the potential appears within a radius, r, less than 0.002 inches, provided surface conductivity is small. The conductivity of the plating solution is kept small to ensure that the potential distribution in the germanium is not altered by presence of the solution. Under these conditions, it is assumed that, although copper ions in solution are attracted towards the highly negative regions of the germanium, the main current flow is through the germanium, except for regions of high potential gradients. In these regions some of the current will be carried by ions in the solution, by-passing the region. If the formed region boundary is a sharp p-n junction, one would expect a plating pattern as observed in Fig. 6(b) and 6(d), as is observed with the donor-free emitter point. For the more complicated structure produced by forming the

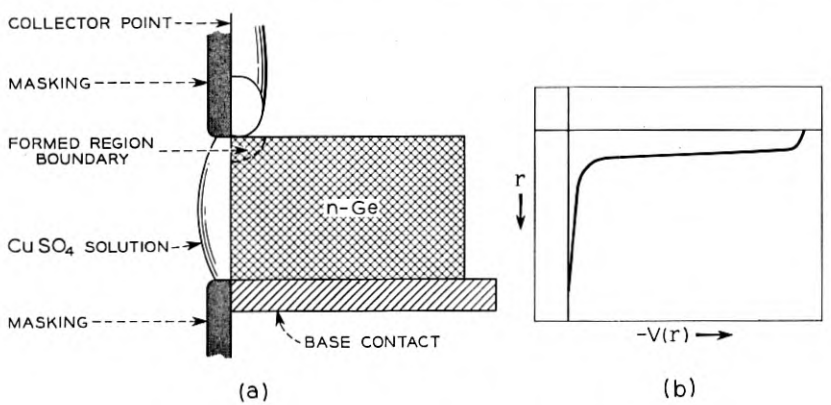


Fig. 5 — Experimental conditions for copper plating.

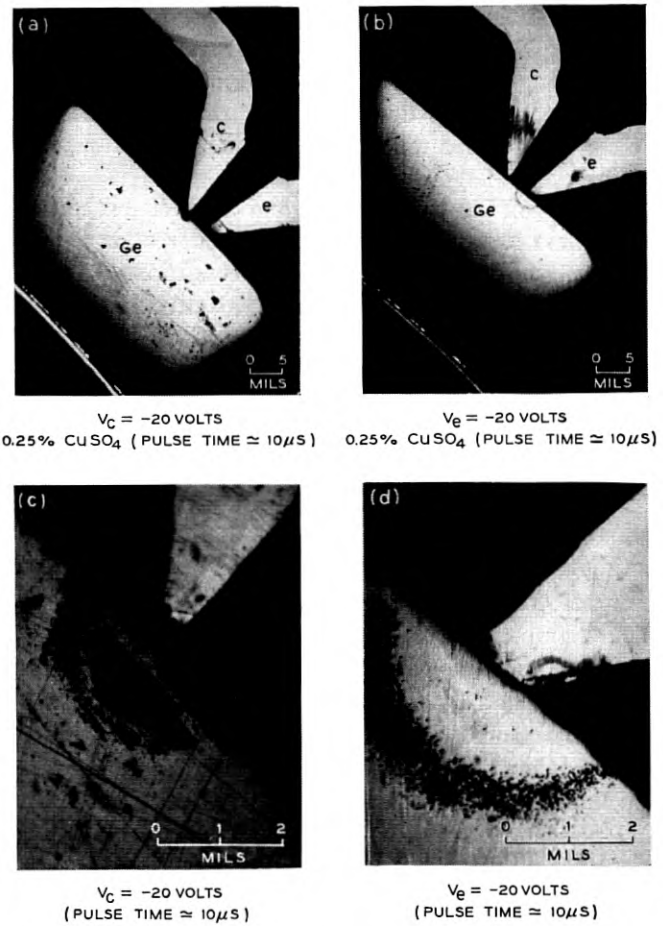


Fig. 6 — Copper plated formed layers in point-contact transistors.

collector, the pattern obtained is more difficult to interpret, Fig. 6(a) and (c). However, in both cases the disturbed areas are roughly comparable in shape and size.

Differences in the forward characteristics of the collector and emitter points may also be graphically observed by means of the plating technique. In Figs. 7(a) and 7(b) are sketches of patterns obtained by applying forward bias to contacts for plating. In this case a more concentrated solution is used, and the plating time is longer. In Fig. 7(a) is shown the pattern obtained when an unformed collector point is biased for-

ward during the plating pulse. The copper deposits to within the order of a diffusion length from the emitter point. Fig. 7(b) shows the pattern obtained by plating the region near a forward biased formed collector. Here again the copper has deposited over practically all of the base wafer surface, except for a much smaller hemispherical region near the collector point.

By adjustment of the plating time and solution concentration, the almost radial field in the bulk germanium under a reverse-biased collector point can be detected. Under similar conditions, an emitter point biased to the same voltage shows a plating pattern similar to that of Fig. 6(b), with little evidence of the radial field. This would be expected from the potential plots shown in Fig. 2.

These techniques serve merely to illustrate graphically the differences in the two types of contact. Although both points when formed give rise to a formed region in the bulk germanium of similar size and shape, the diode characteristics of the junction under the donor-doped point are degraded.

The plating technique may also be adjusted to allow sensitivity to the current flow pattern in a transistor with both points biased to operating values. The example shown in Fig. 8 demonstrates visually the bulk nature of the current flow in the point contact transistor. Here the copper plates out on the negative regions of the crystal and is noticeably absent from the regions of high hole density under the emitter point. In the region to the left of the collector indicated by the arrow, the plating is partially obscured by masking. The size of the copper-free region under the emitter point may be reduced to substantially zero for the same I_e by increasing the bias applied to the collector.

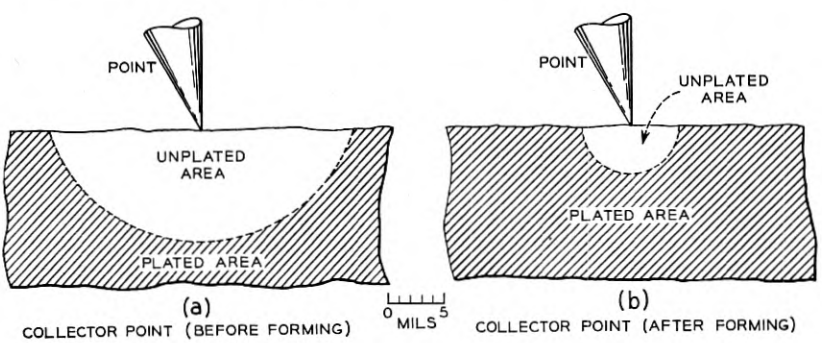
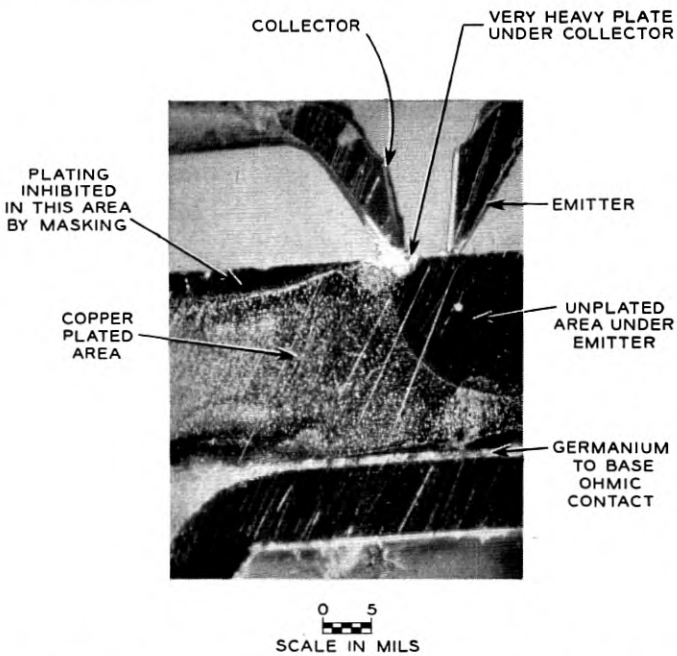


Fig. 7 — The effect of forming and current flow in point-contact collectors.

2.3 *Under-Formed and Over-Formed Contacts*

One of the problems encountered in the large-scale manufacture of point-contact transistors is the variation in the forming yield. Thus, forming to a specified criterion of transistor performance does not always result in a uniform product. Although considerable care may be taken to ensure uniformity of all bulk properties and forming technique, a large variation may be encountered in the output characteristics of the transistors. In Section 4, a prime factor in determining the efficiency of forming is shown to be the chemical history of the germanium surface. Uncontrollable variations in surface conditions may therefore often account for much of the variations in results of a specific forming technique.

Such variations often manifest themselves merely as differences in degree, but may show up as differences in kind, taking the form of anomalous output characteristics. These have been classified by L. E. Miller[14] into three qualitatively different phenomena. The first of these, referred



UNIT OPERATED AT LOW I_e; PLATED 0.25% $CuSO_4$, 20 SECONDS
$I_e = 0.5$ MA, $V_C = -20$ V, $\alpha = 0.1$

Fig. 8 — Flow geometry for a low alpha point-contact transistor.

to as the type (1) anomaly, is of interest here since it represents a collector contact whose physical properties are between the extremes listed in Section 2.2. Miller has shown that the source of this kind of output characteristic can be identified as the formed area under the collector point.

Essentially this anomaly consists of an abrupt rise in the current gain as the collector voltage V_c is increased at constant emitter current. Beyond the critical value of V_c, the characteristic of the unit resembles that of a well formed transistor. One is led to consider that such a contact is under-formed, in the sense that at low V_c, collection of holes is inadequate. Further support is lent to such a definition by the data of Miller, which shows a definite increase in the occurrence of anomalous units with a decrease in the I_{co} of the contact. Such an increase occurs regardless of whether the I_{co} decrease is obtained by decreasing the donor concentration of the point wire, or by increasing the time constant of the forming pulse. In Table II are compared collector capacity and I_{co} measurements made in units with and without output characteristic anomalies. The capacity of these anomalous collectors also appears to range between the two extremes listed in Table I. Thus there is evidence that these collectors are intermediate between the extremes cited in Table I in the sense that at low reverse biases the drift field is low, and the properties of the formed barrier resemble those of a formed donor-free point.

The results of detailed investigation of the properties of such anomalous characteristics now being conducted will be published at a later date. The present experimental results indicate that the instability occurs when the extra current to the collector, ΔI_c, reaches a critical value. In this respect, increasing the transport factor β, by increasing V_c, or increasing the emitter current are equivalent. At a roughly critical ΔI_c, the transition between a low α and a higher value of α occurs. After the transition, the unit behaves like a conventional point contact transistor, with a current multiplication on the order of $(1 + b)$ at higher values of I_e. Thus the origin of this kind of anomaly may lie in the lowering of the formed barrier by the space charge of the holes, a mechanism suggested by Bardeen.[15]

TABLE II

	$I_c(I_e = 0, V_c = -10 \text{ volts})$	$C_c(I_e = 0, V_c = -10 \text{ volts})$
Typical Transistor	1.0 ma	0.1 $\mu\mu$f
Typical Anomalous Transistor	0.2 ma	0.5 $\mu\mu$f

The other anomalous collector characteristics considered by Miller have their origin in the relation between the transport factor and the properties of the emitter at various operating conditions. In view of the relations existing between the occurrence of these anomalies and the I_{co} of the collector contact, there is some justification for classification of these contacts as "over-formed."

3. PROPERTIES OF UNFORMED POINT CONTACTS

3.1 *Physical Properties of Metal-Semiconductor Contacts*

The classical ideas on the nature of the rectifying metal-semiconductor contact have undergone substantial revision since the consideration by Bardeen[16] of the importance of surface states and the work on the point contact transistory by Bardeen and Brattain. According to Bardeen's model, the nature of the space charge layer at such a contact is to be considered largely independent of the metal used for contact, and is primarily dependent on the charge residing in localized states at the germanium surface. Thus the rectifying properties of the metal semiconductor contact in air are expected to be largely independent of the work function of the contact metal.

The question of the exact nature of the surface charges is not yet readily answerable. Charges may arise which consist of electrons and holes residing in surface states of the type proposed by Tamm.[17] On the other hand, other surface charges may arise as a result of adsorbed impurity ions, or from adsorbed atoms or molecules having electrical dipole moments. Brattain and Bardeen[18] have shown that the space charge layer is dependent on the surrounding ambient and have indicated that charge may reside on the outer surface of a film (presumably an oxide layer) at the germanium surface as well as in surface states of the type mentioned above, which are presumably those responsible for surface recombination processes.

Thus, it is the surface charge on the semiconductor, rather than the nature of the metal, which primarily determines the nature of the potential barrier which exists at a metal semiconductor junction.

A schematic electron energy diagram for the contact between a metal and an n-type semiconductor is shown in Fig. 9. The potential barrier φ_0, and the nature of the space charge layer in the semiconductor are determined by the surface charge system and the bulk properties of the semiconductor. In turn, the surface charge system is dependent upon such factors as the ambient at the germanium surface and the chemical history of the surface.

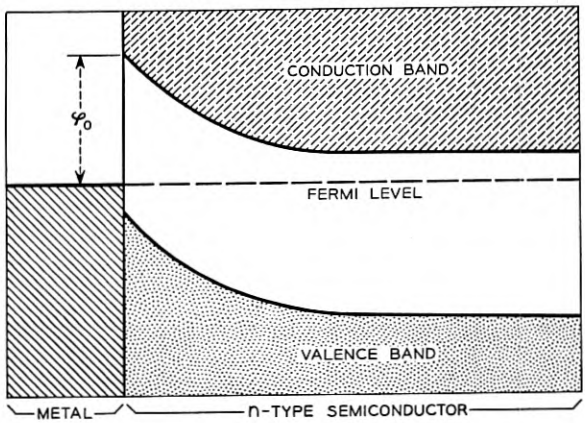


Fig. 9 — Electron energy diagram for a metal-semiconductor contact.

The experiments of Brown[19] indicate that the presence of charge on the surface of p-germanium can alter the space charge in the crystal near its surface and, in some cases, produces an inversion layer of n-germanium at the surface. Garrett and Brattain[20] have shown that a change of ambient from sparked oxygen to dry oxygen to wet oxygen can increase I_{co} and floating potential on n-p-n junction transistors, and the process is reversible. Their interpretation is that sparked oxygen builds up a film, presumably germanium oxide. Oxygen atoms on the surface, negatively charged, can give rise to a p-type inversion layer on n-germanium. Moisture apparently counteracts this negative charge, and humid oxygen can cause an n-type inversion layer on p-germanium, which can be removed with a dry oxygen ambient.

Thus, the electrical resistance of an unformed metal-germanium contact on an etched germanium surface can be expected to be extremely sensitive to any chemical treatment which tends to affect the constitution of the oxide layer present on the surface, regardless of the metal used for contact in air. Bardeen and Brattain,[1] in early transistor experiments, have shown that such is the case. They have used transistor collector points on germanium surfaces which, after etching, were subjected to an oxidation treatment (heating in air).

In this section are described experiments which seem to indicate that the reverse resistance of unformed diodes on etched n-germanium surfaces can be decreased by chemical surface treatment, and the magnitude of the floating potential near such contacts is increased to sufficient extent that the point can serve as a multiplying collector. Average α for

these points approaches values found in electrically formed collectors. Subsequent parts of this section will be concerned with description of the experiments involved and comparison of the electrical characteristics of these points with those of conventionally formed points.

The effections of electrical forming on donor-doped and donor-free point contacts have been described in earlier sections. It has been stressed that the addition of the donor element to the point results in a contact with degraded diode characteristics, but which serves as an excellent collector.

The possibility of an analogous situation in an *unformed* point collector exists, with the electrical forming of the donor-doped point being replaced by a suitable chemical treatment of the surface. The experiments described below indicate that such is the case.

3.2 *Experimental Procedures*

The germanium used in these experiments was zone-leveled material. The n-germanium was in the 3 to 4 Ω-cm range. Originally, experiments were run using slices, about 0.025″ in thickness, soldered on flat brass blocks, with the brass well masked with polystyrene. Germanium dice, already mounted on standard base-header assemblies used in a hermetic-seal transistor process pilot line, were also used.

The ground surface of a slice was given a three-minute chemical etch (CP_4 or superoxol), washed in pure water (conductivity <0.1 micromho), and blown dry in a nitrogen stream. This surface could then be exposed for several minutes to 24 per cent HF, hot zinc chloride-ammonium chloride solder flux, or other chemical treatments as the experiment might require. These solutions were applied to the slice or die in the form of large droplets, so the solution did not come in contact even with the masking. Later, in order to make doubly sure that contamination from the base or base contact was not involved, all experiments were repeated using a two-inch length of a zone leveled bar with a base contact soldered on one end, and the other end, freshly ground between treatments, used as the surface under examination. The etching was done by lowering one end of the bar about one-half inch into the etch, leaving the contact end a good distance from the etch. The etched surface could subsequently be exposed to any desired chemical treatment. After the chemical treatment, the sample surface was again washed in low conductivity water for several minutes and blown dry with nitrogen.

The sample, after chemical treatment, was placed on a double ended manipulator base, used to control the position and pressure of two canti-

lever points on the treated surface. The electrical characteristics of a beryllium copper point, operating as transistor collector on the treated surface, could then be investigated. An auxiliary etched tungsten point doubled as a potential probe and as an emitter. A switching arrangement allowed oscilloscope presentation of the I_c-V_c collector family and the alpha-emitter current sweep, measurement of the emitter floating potential on a high impedance VTVM, and determination of other transistor parameters for any desired position of the emitter point.

Phosphor bronze collector points were not used since it was found that, on certain chemically etched surfaces the mere application of a negative bias of 15–40 volts for a few seconds sometimes is sufficient to cause electrical forming of the point in the sense that I_{co} and average α are increased by an appreciable amount.

The beryllium copper points were carefully cleaned to prevent contamination by donor elements. Their cleanliness was then tested by other methods described in Section 3.3.6.

With this arrangement, most of the electrical properties of a given manipulator unit could be inspected during the time the unit "survived." These electrical measurements were made in room air (R. H. between 20 and 30 per cent), although provision was made for directing a continuous stream of dry nitrogen at the points and surrounding surface.

3.3 *Experimental Results*

3.3.1 *Unformed Transistors on Superoxol Etched* Surfaces*

A striking difference was observed in the electrical characteristics of unformed collector points on the various n-germanium surfaces examined. In particular, surprisingly large values of $I_c(0, -10)$ and $I_c(6, -5)$, (the latter taken as a measure of average α), were encountered on the superoxol etched surface subsequently "soaked" for about 10 minutes with 24 per cent HF. At these locations the unformed transistor action was quite similar to that observed with a conventional phosphor bronze point formed on a freshly etched surface.

These large values were found only in specific locations on the treated surface, there being a random fluctuation of $I_c(0\ -10)$ and $I_c(6, -5)$ with location of the points on the surface. However, no such large values of these parameters were found (together) on surfaces freshly etched in superoxol. The α as a function of emitter current for the unformed points (2.5 mil spacing) on a superoxol etched surface, before (Curve I) and after (Curve II) HF treatment is shown in Fig. 10. Comparison with

* One part 30 per cent H_2O_2, one part 48 per cent HF and four parts water.

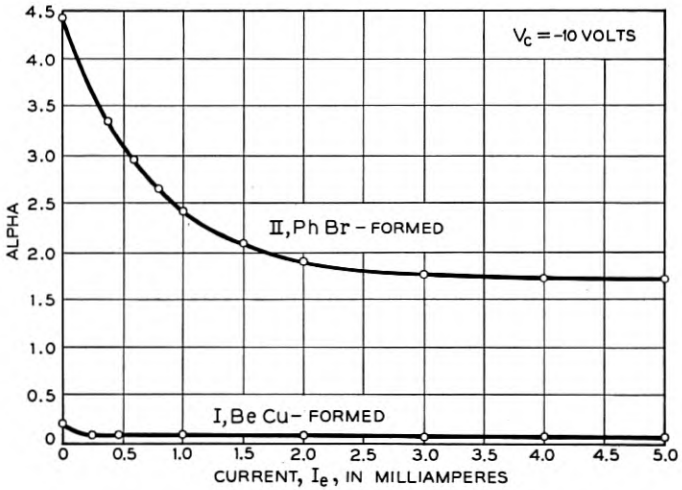


Fig. 10 — Comparison of alpha-emitter-current characteristics for unformed collectors.

Curve II, Fig. 3, indicates that the $\alpha(I\varepsilon)$, obtained after the HF treatment, is comparable to that of a phosphor bronze collector formed conventionally on the same etched surface before treatment. (It turns out that conventional electrical forming on the etched surface after the HF treatment is more difficult, and in cases as referred to above, where the α is not initially high, requires an excessive number of pulses to bring the α to a normal value.)

In Table III are listed the maximum and minimum values of some transistor parameters found on the same superoxol-etched surface before and after the HF treatment (point spacing about 2 mils).

It is seen that the effect of the subsequent HF treatment after the superoxol etch is at least in some locations on the treated surface to increase the $I_c(0, -10)$ and the average α, in some cases to values approaching those encountered in conventionally formed point-contact transistors. There is also a lowering of the forward current of the unformed collector point after the HF treatment. It is not to be implied from this table that the I_{co} is always found to be low on fresh superoxol-etched surfaces. Actually high values of $I_c(0, -10)$ have been occasionally found on surfaces freshly etched in superoxol. However, these collectors seldom have high values of average α, and it is suspected that here the higher reverse current is associated with excessive surface conductivity. Treatment of such a surface with HF always serves to increase the average α, and decrease the forward emitter current, with

TABLE III

Parameter	After 3 Min. Superoxol Etch		After Subsequent 10 Min. "Soak" in 24% HF	
	Max. Value Observed	Min. Value Observed	Max. Value Observed	Min. Value Observed
$I_c(0, -10)$ ma.....	−0.16	−0.06	− 0.98	−0.20
$I_c(6, -5)$ ma.....	−7.0	−0.95	−13.5	−7.0
$I_c(0, +0.5)$ ma.....	2.8	2.0	2.3	1.3
Peak value of α....	3.0	0.15	6.0	2.0
α (5.0, −10).....	0.50	0.09	2.0	0.5

no significant changes in the extreme values of $I_c(0, -10)$ encountered initially. Some of the unformed units have collector families quite similar to those of an electrically formed point-contact transistor. However, the resemblance ends when stability of operation is considered. When the unformed units are operated in room ambient, hysteresis loops are occasionally observed, either in the I_c-V_c output characteristic sweep, or the α-emitter current sweep. This hysteresis can be eliminated by directing a stream of dry nitrogen across the germanium surface in the vicinity of the points. It is not known whether the hysteresis is thermal or electrolytic in nature. The operation of these unformed units, even in the absence of hysteresis, is extremely erratic and unstable. Operating a unit at a high power level will cause loss of α and I_{co}, and mechanical shock delivered to the collector point while the unit operates under bias may cause loss or gain of α and I_{co}. In cases where I_{co} (and α) are low when the collector point is initially set down on the treated surface, an increase in I_{co} and α may be brought about by mechanical motion of the point, (such as "tapping" the manipulator base, or dragging the point across the surface). In other cases the high α and I_{co} are found immediately after the point is set down on the freshly treated surface, without any such procedure. None of these effects is observed to an appreciable degree on a freshly etched surface without further treatment.

The effect of zinc chloride-ammonium chloride solder flux on fresh superoxol-etched surfaces was also investigated. In this case, after the etch, the surface was immersed in almost boiling solder flux for about ten minutes. The effect of this surface treatment on the performance of the unformed transistors was entirely similar to the results quoted in connection with the HF treatment. The treatment increased the reverse collector current and average α, and decreased the forward collector current, on the average. Magnitudes of $I_c(6, -5)$ as high as 14 ma were observed on surfaces treated in this way.

3.3.2 *Unformed Transistors on CP_4-Etched Surfaces*

With reference to unformed point contact properties, the CP_4-etched surface is not at all similar to the superoxol etched surface. If two beryllium-copper points are put down on a ground surface freshly etched in CP_4, and operated as a transistor, high values of $I_c(0, -10)$ and $I_c(6, -5)$ are often encountered. However, after an hour or so in room air, both these parameters decrease and after an overnight exposure to room air, the properties of the surface with regard to the transistor action resemble those of a surface freshly etched in superoxol. At this point, a treatment in 24 per cent HF will return $I_c(0, -10)$ and $I_c(6, -5)$ to their originally high values. These effects are summarized in Table IV.

3.3.3 *Diode Characteristics on Electro-Etched Surfaces*

It has been found that the rectification properties of unformed point diodes may also be changed conveniently by changing the conditions during an electrolytic etch in KOH solution. These results are summarized in Table V which represents typical variation in reverse current, I_r, with surface variation attainable by adjusting the current density and etching time. In each case the measurements represent data taken on germanium cut from adjacent sections of the same ingot and given the surface treatment noted in the table. In general the electro-etched and chemically etched results agree; that is, any treatment which appears most likely to leave an oxide film (such as the use of a high current density during electro-etching) will yield a diode with improved rectification characteristics.

3.3.4 *Output Characteristic Anomalies*

In the process of examining these chemically treated surfaces, some of the superoxol-etched n-germanium surfaces were given additional

TABLE IV

Parameter	Value after 3 Min. CP_4 Etch		Value after 16 Hrs. in Room Air		Value after 10 Min. in 24% HF	
	Max. Value Observed	Min. Value Observed	Max. Value Observed	Min. Value Observed	Max.	Min.
$I_c(0,-10)$ ma..	−1.7	−0.30	−0.10	−0.04	−1.0	−0.06
$I_c(6, -5)$ ma..	−13.3	−11.0	−7.0	−2.0	−17.5	−8.0
Peak value of α.....	4.5	2.5	2.0	0.75	9.0	3.0
$\alpha(5.0, -10)$....	1.8	1.0	1.0	0.25	2.0	0.75

TABLE V

Etch Treatment in 0.1% KOH	I_r (−10 volts)
10 ma for 30 sec	−0.16 ma
5 ma for 30 sec	−0.37
2.5 ma for 30 sec	−0.55
5 ma for 1 min	−0.04
2.5 ma for 1 min	−0.18
1.75 ma for 1 min	−0.74

treatments in H_2O_2 (superoxol strength). In general, no great differences were observed in the unformed alpha and $I_c(0, -10)$ after the treatment. However, in isolated cases, unformed units made on etched p-germanium treated in this way exhibit output characteristic anomalies of the type characterized by Miller as type (1). It was later found that the same surface treatment can produce a similar result on etched n-germanium surfaces, again only in isolated locations on the surface. An output characteristic of this form is shown in Fig. 11. This unformed unit was made on a superoxol-etched n-germanium surface with a subsequent three-minute soak in H_2O_2. This characteristic was extremely sensitive to variation in point pressure.

Miller has also referred to output anomalies of types (2) and (3), which are usually associated with close point spacing in conventional point-contact transistors. Such types of anomaly have been observed in unformed units (with high average alpha) made on HF treated surfaces.

3.3.5 *Floating Potential Measurements*

In all cases where the $I_c(0, -10)$ and average alpha on etched surfaces are increased by the HF or solder flux treatment, these increases are accompanied by an increase in the magnitude of the floating potential near the reverse-biased collector. In Fig. 12 the magnitude of the floating potential V_p of a sharp tungsten probe near the reverse-biased collector is shown as a function of r, the distance of the probe from the collector (r is approximately the distance between the center of the two point contacts). The surface used in this experiment was prepared by chemical polish for three minutes in CP_4 and subsequent storing in room air for sixteen hours. This provided a smooth surface which resembled, at least with regard to electrical characteristics, a freshly etched superoxol surface.

Curve I represents the potential-distance plot for an unformed BeCu point on the aged superoxol-etched surface. Curve II represents a similar plot for an unformed BeCu point taken after the surface was given a ten-minute soak in 24 per cent HF.

The measured resistivity of the germanium used in this experiment was 3.3 to 3.6 Ω-cm. It can be seen from Curves I and II that increase in the magnitude of the floating potential near the unformed point on the etched surface after the HF treatment is, to a rough extent, proportional to the increase in $I_c(0, -10)$ produced by the treatment. Values of $2\pi V_p r/I$ taken from lines of slope (-1) drawn for best fit through points on the individual curves give reasonable agreement with the measured resistivity. For curve I, $2\pi V_p r/I = 3.3$ ohm-cm, and for Curve II, $2\pi V_p r/I = 3.5$ ohm-cm.

By comparing Curves I and II of Fig. 2 with Curves I and II of Fig. 12, it can be seen that the effect of treating the surface under the unformed point with HF is analogous to adding donor to the formed point

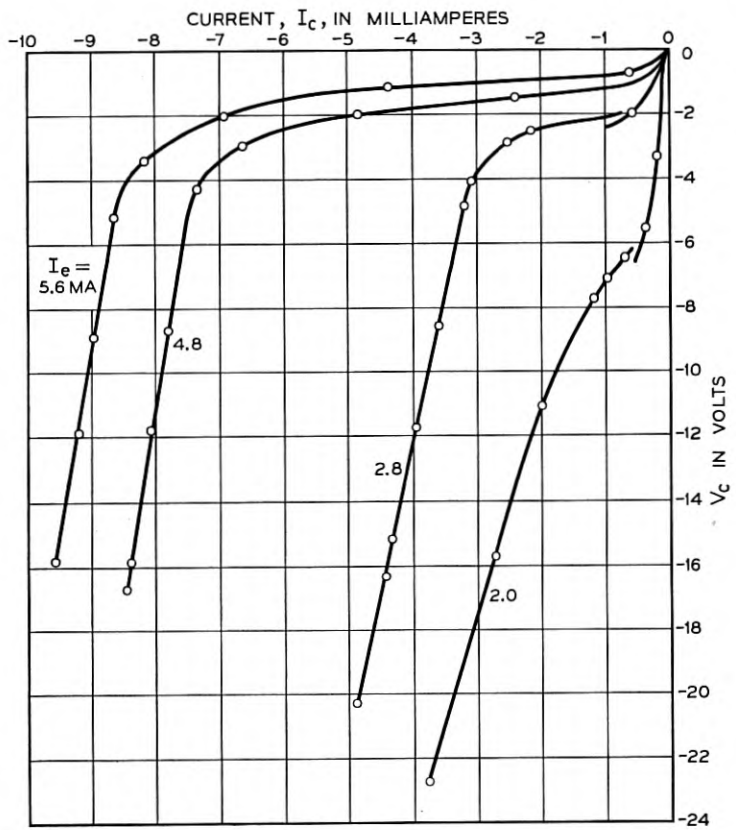


Fig. 11 — Type (1) collector anomaly observed in unformed unit (n-type germanium).

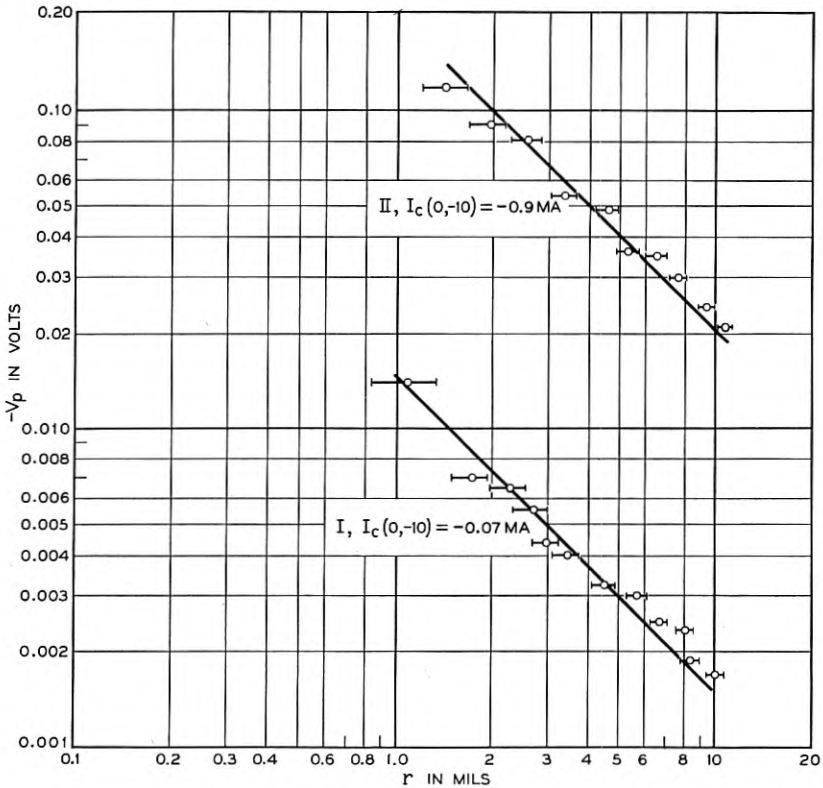


Fig. 12 — Comparison of floating potentials for unformed point-contact collectors.

on the etched surface. It seems reasonable to ascribe the increased negative floating potential after the HF treatment to an increase in current density through the surface under the point, rather than to any increase in surface conductivity. It is worth noting that on superoxol-etched surfaces, the negative floating potential near an unformed collector point can often be increased by an order of magnitude by blowing a stream of dry nitrogen near the point. This effect may possibly be a result of excess surface conductivity, but in these cases is *not* accompanied by any appreciable changes in $I_c(0, -10)$ or average alpha.

3.3.6 *Contamination of Collector Points and Surfaces*

Past experience with use of point-contacts as transistor collectors indicates that experiments may often be confused or confounded by unsus-

pected contamination of the points used. For this reason, particular attention was given to chemical processing of the beryllium copper points used in the preceding experiments. These points were chemically cleaned to remove oxides and unwanted contaminants, and carefully washed before use. Several lots were processed at different times, and all experiments repeated on the different lots, with no contradictory results.

It is particularly important that the point be free from donor elements, since it has been observed that that phosphor bronze points or "poisoned" beryllium copper points washed with a lithium chloride solution often exhibit on superoxol etched surfaces a kind of "forming" after the application of reverse bias. The symptoms of this are a sudden increase in I_{co} which take place as the reverse bias is increased above 15–20 volts. The alpha emitter current sweep shows evidence of excessive noise in such a case, and it is not until the collector is given a conventional forming pulse that this excessive noise is eliminated, and the unit becomes stable in operation.

A donorless point can be reasonably identified by the fact that electrical pulsing, heavy or light, will not increase the initially low average alpha on a superoxol-etched surface to values much above 1.0, although I_{co} may be increased or decreased depending on the type of condenser discharge used. The beryllium copper points used were tested on superoxol-etched surface to make sure they showed no tendency to form electrically.

If high values of alpha can be found when these points are used as unformed collectors on the surfaces treated in HF or solder flux, the question arises whether such values may be attributable to presence of a donor element left on the surface in some mysterious way by the chemical treatment. If such is the case, the donor might, at high enough reverse bias, be responsible for an increased alpha in a manner similar to that observed in connection with the forming in under bias of phosphor bronze collectors on etched surfaces. Two precautions were taken in this connection. No reverse bias greater than 10 volts was ever applied intentionally to these collectors during experiments (with exception of the unit in Figure 11), and secondly, forming characteristics of both phosphor bronze points and the beryllium copper points on this type of surface were investigated.

It was found that on a superoxol surface treated with HF or the solder flux, a phosphor bronze point would form to a high average α, but this invariably required more forming pulses than on a superoxol etched surface. "One-shot" forming is common for a superoxol etched surface, whereas after the HF or solder-flux treatment, forming to high average

TABLE VI

	Point			
	Beryllium Copper Collector		Phosphor Bronze Collector	
	Occasion			
	Before Forming	After Four Forming Pulses	Before Forming	After Four Forming Pulses
$I_c(0, -10)$	−0.75	−0.50	−0.80	−2.5
$\alpha(5.0, -10)$	1.5	0	1.5	2.0
Noisy	Yes	Yes	Yes	No

alpha invariably requires at least three and sometimes many more "shots", although it can be done. This type of formed unit does not exhibit excessive noise in the α-I_e sweeping gear. However, pulsing of the beryllium copper points on the latter kind of surface, in similar fashion, invariably results in loss of alpha and never eliminates the excessive noise. Initially, the pulsing decreases the I_{co} magnitude, but continued pulsing will eventually cause large increases in this case. These results provide circumstantial evidence, at least, that the treated surface and the point are operationally free of any donor element and that the transistor collector barrier involved is at the germanium surface. For example, in Table VI are given some typical data obtained during pulsing of points on a superoxol-etched surface after treatment with near boiling zinc chloride-ammonium chloride solder-flux. A tungsten emitter was used.*

3.4 *Discussion of Experimental Results*

3.4.1 *Effects of the Chemical Treatment on the Superoxol-Etched Surfaces*

It might be presumed that an inversion layer and a relatively high surface conductivity is responsible for the increase in negative floating potential and reverse current observed on the superoxol-etched n-germanium surface after the HF treatment. On the other hand, if it be assumed that at the etched surface, in room air, an inversion layer exists which does not introduce excessive surface conductivity, one can say that the effect of the HF treatment is merely to raise the surface potential, (i.e., to reduce the barrier height for electrons). This might

* Alpha values are usually lower in any given situation when the conventional chisel-type beryllium copper emitter point is replaced by an etched tungsten point.

account for the increase in reverse current density* and a proportional increase in the magnitude of the floating potential near the point. In this case the geometry of current flow across the contact should remain relatively unchanged as indicated by the floating potential measurements. In this way the effect of the HF treatment is somewhat analogous to the addition of a small donor concentration near the surface to counteract the inversion layer. Since soluble oxide layers[21] have been identified on etched germanium surfaces, it is not unlikely that HF (known to dissolve germanium oxide)[22] might act to reduce the effective thickness of an oxide layer. Such a hypothesis is in agreement with the results of other experimenters,[23] who have attributed a surface inversion layer under the point of an n-germanium rectifier to the presence of germanium oxide. They have presumed the oxide is essential to the formation of a good point contact rectifier. The fact that, for a given ambient, the surface potential is determined by the oxide layer thickness has been postulated by Kingston.[24]

3.4.2 *CP_4-Etched Surfaces*

Sullivan,[25] in connection with an experimental investigation of humidity stability of electrolytically-etched and chemically-etched p-n grown junction diodes, shows that CP_4 chemically-etched surfaces become more stable with respect to humidity variation after humidity exposure and cycling at room temperature. Referring to the fact that electron diffraction studies fail to reveal a crystalline oxide film on CP_4 chemically-polished surfaces and to the results of Law,[26] which indicate that oxide films may be formed slowly at room temperature on exposure to water vapors, he attributes the changes of stability on the CP_4 polished surface to the building up of an oxide film. If such a change can take place on the CP_4 chemically-polished surface on exposure to humid room air, then the results of Section 3.3 can be understood under the assumption that the action of the HF treatment is to remove the oxide film.

After the chemical polish, values of $I_c(0, -10)$ and average alpha for the unformed units are high, as might be expected if the polishing operation leaves the germanium surface with no appreciable oxide film. As the oxide film builds up on continued exposure to room air, both of these parameters are reduced. The subsequent application of HF tends to restore these parameters to their original values by removal of some of this oxide film. Thus, the results of this section are in accord with the

* Evidence for an increase in surface recombination velocity on HF treated surfaces is given in Section 4.2.3.

hypothesis discussed in the previous section to account for the effect of HF on the unformed transistors.

Such evidence, however, is at best only indirect evidence for the build-up of an oxide layer on prolonged exposure to room air. In experiments with grown p-n junction diodes, the authors have found great variations in the length of time required for the electrical properties of the diodes to recover after short wash periods in low conductivity water. Thus the slow changes mentioned above may at this point result from simply a longer time required for the surface to "dry out" after the washing treatment. However, a substantial difference in the physical properties of the oxide layer left by the two etches concerned is still implied. In this connection it is also worth noting that hysteresis effects appear primarily in unformed units made on HF treated surfaces.

The results of these experiments have important implications in the technology of point contact transistors. The results of an application of these results to transistor forming procedures are given in the following section.

4. RELATION OF GERMANIUM SURFACE PROPERTIES TO TRANSISTOR FORMING

4.1 *Pilot Production Problems*

The pilot production and early manufacturing stages of cartridge-type point-contact transistors has generally been characterized by periods during which the forming yields have been very high and similar periods of very low yield. Often these alternate periods occurred during the use of germanium taken from the same rod-grown or zone-leveled crystal. Considerable effort has been expended in attempting to correlate these variations in yield to variations, from crystal to crystal, or in different portions of the same crystal, or such bulk properties as resistivity or minority carrier lifetime. Although these properties of germanium do have some effect on device parameters such as average alpha, reverse emitter current, and I_{co}, there has not been any positive indication that variations in yield are attributable to the amount of variation of bulk properties normally found in the germanium which meets the specifications of the particular device concerned.

This problem was compounded during the early stages of the development of the process for hermetically sealing the point-contact transistor. It was found that although reasonable yields were obtained in the cartridge process, equivalent transistors in the hermetically sealed structure were made only with greatly reduced yield. Further, although micro-

manipulator units could be made with no difficulty, the same material fabricated into a completed structure showed completely different characteristics. In the course of investigation of this problem, it was ofund that the nature of the germanium surface treatment and specifically treatments calculated to produce or react with germanium oxide can profoundly affect the "formability" of the germanium surface as well as a number of other transistor parameters in the fabricated units.

It is the purpose of this section to emphasize the importance of considering the surface properties of germanium in attempting to solve such specific problems of development encountered in devices of this type. In particular, the striking variability of transistor forming on etched germanium surfaces subjected to varying chemical treatments and ambients will be described, as well as the effects of such pre-forming treatments on the parameters of the finished units. The experiments discussed in the previous section indicate how changes in the double layer at the germanium surface can influence the characteristics of an unformed point diode. In turn, the experiments below indicate how the characteristics of the unformed diode are related to the device properties of the transistor collector produced by forming the diode.

4.2 *Experimental Results*

4.2.1 *Pilot Process Forming Yields*

The forming yield of a point-contact transistor is determined by the values of the acceptance criteria and the allowable limits for each of these. Often, different criteria as well as different forming techniques are used for different transistors, so that direct comparison of results is quite complex. There are, however, certain common requirements placed on all point-contact transistors:

(a) The unit is formed so that the average alpha is roughly two or more. The collector current at a relatively high emitter current and low collector voltage is usually an approximate measure of this value, $I_c(6, -5)$ for example.

(b) The collector current with no emitter current flowing should be as low as is commensurate with the first objective.

The other transistor parameters are either directly or indirectly related to these. The number of pulses required to achieve the minimum forming objective, therefore, is one direct measure of the formability of a particular transistor; the average alpha obtained after pulsing is another. However, one must consider both average alpha and I_{co}, since while forming to a given average alpha, the I_{co} may increase prohibi-

TABLE VIII

Treatment	No. of Pulses to Form	Average $I_c(6, -5)$	Average $I_c(0, -20)$	Figure of Merit $\frac{I_c(6, -5)}{I_c(0, -20)}$
3 min. in normal superoxol etch	2	−15.5 ma	−0.69 ma	22.4
1 min. in 48% HF	4	−10.2	−3.2	3.2
1 min. in 30% H_2O_2	1	−17.7	−1.9	9.3

oxides which are present so that a good solder joint may be made. Since the oxide on chemically-etched germanium is likely of the soluble form, one might assume that the results of Table VII imply that the action of the flux and heat tends to dissolve or remove this layer. Also implied by the data is that the presence of such an oxide layer is essential to efficient forming.

The experiments summarized in Table VIII further substantiate this hypothesis. These data represent manipulator transistors made on the *same* germanium wafers which had been treated in succession to a normal superoxol etch, a treatment in 48 per cent hydrofluoric acid, and a treatment in hydrogen peroxide, superoxol strength. Since the soluble form of germanium dioxide is known to react with hydrofluoric acid,[22] it is presumed that the action of the HF is to partially or wholly remove any oxide left by the etch. The H_2O_2 tends to restore the original surface conditions left by the etch. Each figure represents the average of five transistors formed to the 2N21 acceptance criterion, ($V_c(3, -5.5) \leqq 2.0$ volts).

In this case the hydrogen peroxide treated units have an extremely high average alpha, but the I_{co} is also higher than for normally etched units. In terms of the device properties, a unit with a more or less typical average alpha with a low I_{co} is more desirable than the one with an extremely high average alpha but accompanying high I_{co}. It has not been determined whether the I_{co} would be lower for the superoxol treated units if it had been possible to form to the same average alpha as the normally etched units. This is an important piece of device design information which is currently under investigation.

It is clear from these experiments that the nature of the germanium surface, and most probably the nature of the germanium oxide layer on it, to a large extent, determines the properties of the transistor formed on this surface. Direct application of this knowledge to the fabrication process of the hermetically sealed point contact transistor has been carried out by N. P. Burcham.

4.2.2 *Relation of Unformed Diode Characteristics to Transistor "Formability"*

From the results of the previous sections, it appears that superoxol-etched germanium surfaces treated with reagents in which germanium dioxide is soluble provide point contact diode characteristics unsuited to electrical pulse forming. Part of this difficulty, manifested in the inability to reach a specified value of average α without a prohibitive increase in I_{co}, probably results from a lower injection efficiency, γ, for the emitter on such a surface. This seems reasonable in view of the lower forward and higher reverse currents indicated in Table III produced by an HF soak. In Section 4.2.3 evidence will be shown that surface recombination is greater on n-type germanium surfaces treated with HF. This effect can also lead to difficulty in forming to high α without increase in I_{co}, since, for the same drift field, one would expect more minority carriers to die at the surface during their transit to the collector.

On the other hand, there is evidence for believing that the nature of the forming process itself may be quite different on an HF treated surface. Fig. 14(a) shows the time dependence of the collector voltage during a typical condenser discharge forming pulse.

The envelope of the voltage pulse follows roughly an exponential decay of a condenser-resistor series combination. However, inspection shows that during the discharge time, the resistance of the combination

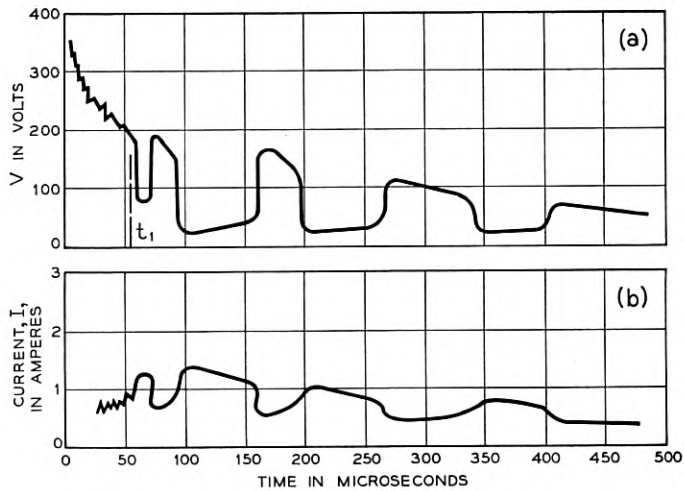


Fig. 14 — Collector current and voltage versus time for a condenser discharge forming pulse.

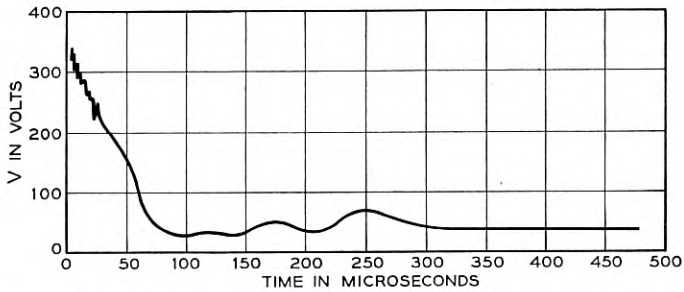


Fig. 15 — Forming voltage pulse for HF treated surface.

undergoes a succession of breakdown and recovery intervals. In Fig. 14(b) is the accompanying plot of current against time. Comparison of these two plots shows that following the application of the voltage, the resistance of the contact decreases until a rather sudden more rapid decrease in resistance occurs, taking place at time t_1. In view of this time scale, the first decrease can be attributed to a heating of the contact, a form of thermal breakdown at the metal-semiconductor surface.[27] Any reason invoked to account for the second more rapid decrease in resistance must account for the short time (a few μs) in which this change occurs. In any event, shortly after the second "breakdown," a quenching results, with the collector resistance returning to a value nearer to its original value. This sequence of events is roughly repeated until the condenser is discharged.

The properties of the contact at nominal reverse voltage and currents are usually changed as soon as one such condenser discharge pulse has occurred, and often one such pulse is sufficient to reach the forming objective. A typical forming pulse obtained under similar conditions to those for Fig. 14 is shown in Fig. 15, with the exception that the surface has been treated in HF for a few minutes. On this case it is apparent that the second, rapid breakdown is entirely absent. The well-defined forming pulse of Fig. 14 is usually obtained on surfaces with good pre-forming diode characteristics, and results in production of a usable transistor.

From results of the previous sections it is well established that etched surfaces treated with reagents in which germanium dioxide is soluble provide point contact diode characteristics unsuited to electrical pulse forming.

It is often assumed, on the basis of the results of Valdes,[5] that forming effects result from the diffusion of impurities from the point into the semiconductor during the forming pulse. Since the high temperature required for such diffusion results from the power dissipated at the metal

to semiconductor contact, more efficient forming probably results on surfaces which display very low initial saturation currents. On surfaces which produce a poor initial rectifying diode, the local energy of the forming pulse may be dissipated too far out into the bulk of the semiconductor. This situation would result in inefficient forming.

Since the low-voltage diode characteristics and the forming are probably related, one should be able to predict the "formability" of any particular surface. Fig. 16 shows that this can be done qualitatively. In the graph each point represents the average of at least five units formed on electro-etched surfaces to the forming objective, $V_c(3, -5.5) \leqq 2.0$ volts. Fig. 16(a) represents the reverse emitter current before forming plotted on a log scale versus the percentage of units taking more than five pulses to form. The reverse emitter current rather than the reverse collector current is a desirable preforming parameter to use since this pre-

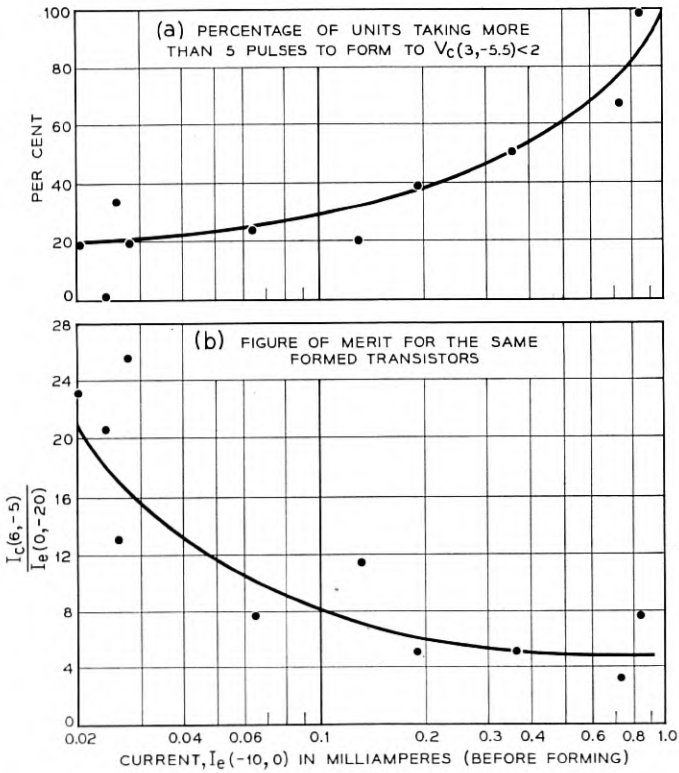


Fig. 16 — Relation of forming to pre-forming characteristics: electro-etched surfaces.

cludes any premature forming which could occur. This curve shows that a low reverse emitter current (high back impedance) is associated with easy forming and that a high reverse emitter current is associated with hard forming. Fig. 16(b) represents data on the same group of units with $I_c(6, -5)/I_c(0, -20)$ plotted versus the reverse emitter current on a log abscissa. It is significant to note that the figure of merit is consistently high for units with low reverse emitter current and low for units with high reverse emitter currents. It was possible to achieve this wide range in reverse currents on the same material by adjusting the current density in the manner summarized by Table V. In each case a high current density results in the low reverse currents.

Some other oxidizing agents may be used interchangeably with the materials just discussed. A dilute nitric acid solution produces a surface on which excellent diode properties are observed and good forming results on these surfaces. It has also been found that a treatment in potassium cyanide results in a surface which appears to be well oxidized. There are, however, some indications that certain chemical treatments tend, more than others, to passivate the germanium surface to any subsequent treatment.

Although it has been shown that variations in the surface oxide layer markedly affect the transistor made on that particular surface, variations in forming yield such as illustrated by the manipulator line in Fig. 13 are still unaccounted for. The etching procedure in the fabrication of the point contact transistor has always been one of the most carefully controlled steps. It therefore becomes necessary to examine the process for some subtle interaction between the germanium surface and the ambient to which the surface is subjected during processing.

4.2.3 *Controlled Ambient Experiments*

The experiment summarized by Fig. 17 represents a "dry box" experiment designed to investigate the effect of ambient on the forming yield. Ten germanium wafers were mounted on hermetic seal headers, they were electro-etched, and then five treated for one minute in HF. The wafers were rinsed in deionized water, dried for three minutes in a stream of nitrogen, and placed in a nitrogen dry box where the relative humidity was maintained at less than 1 per cent. One micromanipulator transistor was formed on each wafer immediately and then at subsequent intervals of one day, always in widely different locations on the wafer. These manipulations were carried out inside the dry box using rubber gloves so that at no time was the RH greater than 1 per cent. After two days the box was opened to room air and the experiment continued.

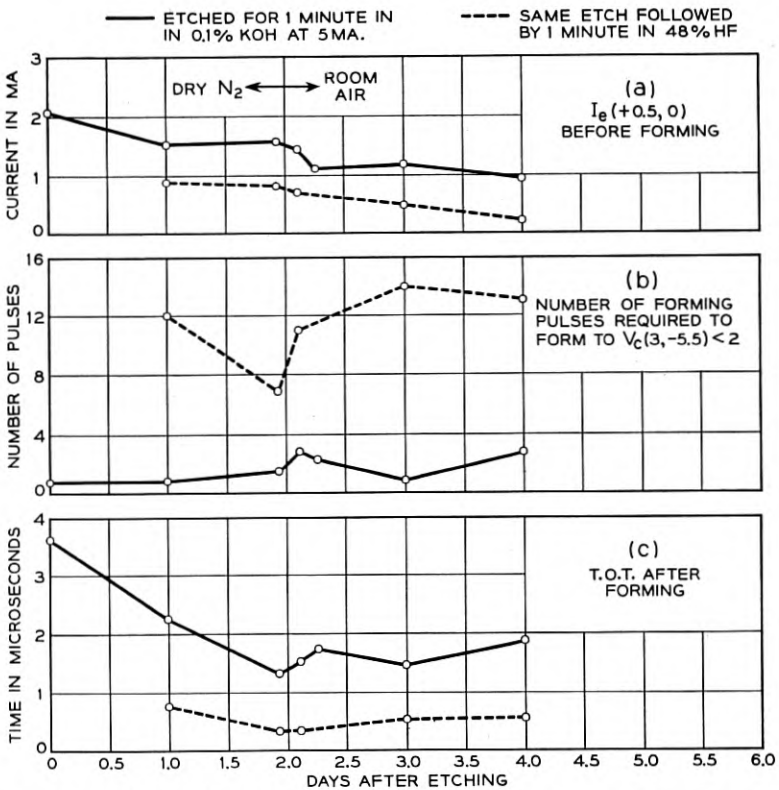


Fig. 17 — Effect of storage ambient on transistor characteristics — electro-etched surfaces.

Each point on Fig. 17 represents the average of five units on five different wafers.

The difference in the electrical properties of the two surfaces in air already noted in previous sections is observed. In addition an increase in surface recombination is indicated on the HF treated surface by a decrease in the turn-off-time measurement (TOT).* Finally, any influence of ambient on the electrical properties of the two surfaces used is apparently small.

4.2.4 *A Statistical Survey Experiment on Transistor Forming*

The experiment described here was designed to check some of the effects noted in earlier sections as well as to investigate possible interactions between the germanium surface and various ambients experienced during the processing of point contact units. The experimental design

* TOT is a nonparametric measurement indicative of the switching speed when used in a specific circuit.

TABLE IX — EXPERIMENTAL DESIGN OF RANDOMIZED BLOCK EXPERIMENT

Ambient Shelf Conditions	Surface Treatments					
	A Normal Superoxol Etch	B Electro Etch 5 ma. for 1 min. in 0.1% KOH	C 1 min. H_2O_2 after Normal Etch	D 1 min. H_2O_2 after Normal Etch and Shelf in Ambient	E 1 min. HF after Normal Etch	F 1 min. HF after Normal Shelf in Ambient
Formed immediately after treatment	x	x	x	x	x	x
Formed after shelf in room ambient	x	x	x	x	x	x
Formed after shelf over drierite	x	x	x	x	x	x
Formed after shelf in dry N_2	x	x	x	x	x	x
Formed after shelf at 76.5% RH	x	x	x	x	x	x

Note: Shelf represents storage for 24 hours.

used is a 5 × 6 randomized block experiment with multiple subgroups.[28] Table IX shows the general plan of the experiment. The six columns represent different etch treatments, and the five rows represent some possible variations in storage conditions. Each subgroup represents five transistors, and the experiment represents a total of 150 transistors made on germanium from the same zone-leveled slice, given 30 different treatments. Although nine measurements were made for each transistor, the figure of merit appeared to be most significantly dependent on the treatments.

As expected from the results already quoted, the major variability was found in units formed on surfaces freshly treated with HF, with considerable improvement in formability during storage. However, the looked for influence of storage ambients does not appear when the column F has been removed from consideration. One concludes that the variation between treatments is small, and the effect of ambient is even less than the effect of the treatments. Thus when surface treatment does not vary to extremes, the effect of storage ambient is relatively minor. Thus variations found in such experiments as exemplified on the manipulator line in Fig. 13 must be attributed to a still unknown factor.

4.2.5 *Effect of Contamination Before Etching*

Since etching removes the damaged surface and is usually done with highly corrosive materials, it seems unlikely that any contamination

before etching could affect the efficiency of etch. There have, however, been some indications that this does occur. Certain chemical treatments appear to passivate the surface to any subsequent treatment, for example, the results in Sections 4.2.3 and 4.2.4. The electro-etched surface followed by an HF treatment does not change rapidly with time in room air, while the superoxol-etched surface followed by an HF treatment changes quite rapidly. Surfaces which have been etched in CP_4 and subsequently treated in HF appear to be as stable as electro-etched surfaces. Subsequent treatments in superoxol do not appear to result in significant changes in the surface characteristics. Experiments on unetched germanium wafers indicate that none of the components of CP_4 alone will prevent normal etching, but if an unetched wafer is treated with a combination of 50 per cent nitric acid plus 48 per cent HF for a few moments, the surface will be stabilized as to retard the formation of the normal pyramidal etch pattern when the surface is etched in superoxol etch. Taken together these observations may imply that certain types of oxide surfaces are more stable than others and perhaps may even be passivated to subsequent environmental conditions.

With this background of information it becomes more believable that chemical treatments before etching could affect the surface of the germanium resulting from the subsequent etching. It is not unreasonable to believe that any variation in surface potential resulting from pre-etch treatment might influence the reaction between the etchant and the germanium. An experiment was performed using gold-bonded bases to isolate the contribution of the solder flux normally used in the base-wafer attachment. Twenty wafers from the same slice were divided into four subgroups of five. The groups were treated in such a way that any effects of HF or solder flux soaking before superoxol etching could be detected.

The results of this experiment do indicate that presence of flux before etching significantly affects the collector currents and turn-off time of transistors made on such surfaces. Although there was no apparent difference in forming yield between sub-groups, it is felt that this variation would show up as a difference in forming yield in a process where the forming efficiency is decreased somewhat by the impregnant.

4.3 *Conclusions*

Treatment of an etched surface with germanium dioxide solvents such as HF or KOH degrades the surface to such an extent that transistor forming efficiency is decreased. A similar effect is produced by corrosive flux and heat. Thus, pre-forming measurements may be used to predict

the formability of a particular germanium surface. It is shown that poor diode characteristics are usually associated with poor forming yields. One convenient way of controlling the diode characteristics to ensure successful forming is to etch electrolytically. High current density results in the most desirable surface characteristics. Electro-etched germanium which has been subsequently treated in hydrofluoric acid shows little tendency to oxidize either in room air or dry nitrogen ambient, while superoxol-etched germanium, given the same HF treatment, changes quite rapidly in room air presumably due to oxidation of surface. Sullivan[25] has also observed differences in the stability of electro-etched and chemically-treated surfaces.

Different surfaces can be prepared chemically which show more than the amount of variation normally found in pilot and manufacturing process lines. However, extreme variations in storage ambients have relatively little significant effects on any of these surfaces. It is therefore concluded that although certain chemical treatments may affect forming, the variations in process yields are not attributable to interaction between the germanium surface and storage ambients.

The results of Sections 4.2.2 and 4.2.3 suggest the possibility of passivation of the germanium surface. An electro-etched surface followed by an HF treatment exhibits a higher degree of stability to ambient than does a superoxol-etched surface treated in the same way. Treatment of a lapped germanium surface with two components of CP_4 ($HF + HNO_3$) will inhibit subsequent etching in superoxol.

The possibility that contamination *before* etching may affect the characteristics of the germanium surface after etching is considered. Experiments show that contamination of the germanium with corrosive zinc chloride-ammonium chloride flux before etching significantly affects the rectification properties of the germanium surface obtained after etching. The surface recombination velocity (in so far as it is determinative of the turn-off time of the transistor) is also significantly affected. However, on the basis of the results quoted here, it is not possible to conclude that such contamination can account for an appreciable amount of the unassignable variability in forming yields experienced in pilot and manufacturing process lines involving soldered base-wafer connections.

5. GENERAL CONCLUDING REMARKS

The experiments which have been described have implications which are important in both design and processing of point-contact transistors. These are summarized below:

5.1 *Point-Contact Transistors with High Current Gain*

In most switching applications the combination of high current gain and low reverse current is desirable. The measurements of current gain, taken together with the potential probe measurements in Section 2.2.1, indicate that, for the structures used here, the reverse collector current at operating voltage must be large enough to set up a substantial drift field before efficient collection of holes can occur. If this condition is not met, either the unit has low gain at all values of emitter current (unformed), or develops a bistability of the kind described in Section 2.3 (partially formed). For a given structure, the drift field can be increased by increasing resistivity of the germanium at the expense of increased base resistance. Here thermal stability of the contact also provides a limit. A more likely expedient, in the case of germanium, is to decrease the area of the formed collector junction by using sharper points and modified forming technique. The limits here are produced by reliability requirements for mechanical stability of the point structure.

5.2 *Current Multiplication in Unformed Transistors*

Many experiments have reported on junction transistors with high current gains which are attributable to the p-n hook mechanism. The high values of current gain observed with conventionally formed point contact transistors have been attributed to various mechanisms, among which is the hypothesis of a p-n hook structure,[6] primarily in the bulk of the germanium, introduced by the pulsing of the donor-doped point. In particular, at small emitter currents small signal α-values in conventionally formed collectors may reach values as high as ten, and values of α as large as 100 are encountered in formed collectors exhibiting anomalous output characteristics. However, the average α over a 6-ma emitter current range is usually near the value of 3.1 which would be expected from the mobility ratio of holes and electrons with the Type-A transistor geometry. The increase in reverse current of a formed collector by addition of donor to the point wire may result from the production of a hook structure. However, information is needed concerning the importance of the hook structure in accounting for the high values of α encountered at low emitter currents, or in connection with collector characteristic anomalies in conventionally formed point-contact transistors.

The unformed transistors discussed in this article differ from electrically formed units in that the collector barrier is the one at the metal-semiconductor surface. It has been found that certain chemical treatments can produce a collector barrier which allows an increased reverse

current flow and a substantial drift field near the emitter. Some of these units show an α value at all emitter currents quite comparable in magnitude to that of conventionally formed collectors, and surface treatment alone can also introduce in these unformed collector characteristics anomalies similar to those found in some formed units. It is difficult to visualize a p-n hook structure arising at the germanium surface as a result of the chemical treatments discussed. If such a possibility is precluded, the p-n hook mechanism does not seem necessary to the attaining of high α values at low emitter currents, or an α emitter current dependence of the kind normally observed in anomaly-free units. To account for values of α obtained with unformed collectors at low emitter currents, other mechanisms, such as the suggestion of Shockley, involving hole trapping in the germanium under the collector point[6, 7] or the suggestion of Van Roosbroeck,[29] involving conductivity modulation, might in this case be more suitable.

Further, unformed transistors made by appropriate chemical treatments can duplicate qualitatively the electrical characteristics of conventionally formed units, including alpha-emitter current dependence and output characteristic anomalies of types (1), (2) and (3). These phenomena can thus occur under circumstances where a well-defined hook structure is improbable.

5.3 *Surface Properties and Transistor Forming*

It has been found that a major factor in determining the forming yield of point-contact transistors is the chemical history of the surface. Thus in processing of point-contact transistors, major attention should be paid to ensuring chemical control of the base wafer surface if the forming yield is to be kept high. On the other hand, considerable variation may apparently be tolerated in storage ambients. Of course it has not been shown that such variations in storage conditions do not have an effect on subsequent reliability of the product. Processes which permit exposure of surfaces to solder fumes either before or after etching are to be regarded with suspicion. Monitoring of the reverse emitter diode characteristics should prove useful as a means of securing proper control of the pre-forming surface.

ACKNOWLEDGEMENT

The authors wish to acknowledge the help of M. S. Jones, who carried out many of the experiments mentioned here, and N. Carthage who did the electroetching work. The continued support and encouragement of N. J. Herbert has been greatly appreciated.

REFERENCES

1. J. Bardeen and W. H. Brattain, Physical Principles Involved in Transistor action, Phys. Rev. **75,** p. 1213, April 15, 1949.
2. J. Bardeen and W. G. Pfann, Effects of Electrical Forming on the Rectifying Barriers of n- and p-Germanium Transistors, Phys. Rev. **77,** p. 401–402, Feb. 1, 1950.
3. W. Shockley, Electrons and Holes in Semiconductors, D. VanNostrand Company, New York, N. Y., p. 110.
4. Reference 3, p. 111.
5. L. B. Valdes, Transistor Forming Effects in n-Type Germanium, Proc. I.R.E. **40,** p. 446, April, 1952.
6. W. Shockley, Theories of High Values of Alpha for Collector Contacts on Germanium, Phys. Rev. **78,** p. 294–295, May 1, 1950.
7. W. R. Sittner, Current Multiplication in the Type A Transistor, Proc. I.R.E., **40,** pp. 448–454, April, 1952.
8. Valdes (Reference 5) reports large concentrations of copper present in the p-germanium under heavily formed phosphor-bronze points.
9. W. G. Pfann, Significance of Composition of Contact Point in Rectifying Junctions on Germanium, Phys. Rev. **81,** p. 882, March 1, 1951.
10. C. S. Fuller and J. D. Struthers, Copper as an Acceptor Element in Germanium, Phys. Rev. **87,** p. 526, Aug. 1, 1952.
11. C. S. Fuller, Diffusion of Acceptor and Donor Elements into Germanium, Phys. Rev. **86,** p. 136, April 1, 1952.
12. Reference 5, p. 448.
13. Personal communication, H. E. Corey, Jr.
14. L. E. Miller, Negative Resistance Regions in the Collector Characteristics of Point Contact Transistors, Proc. I.R.E., **40,** p. 65–72, Jan. 1, 1956.
15. Reference 1, p. 1225.
16. John Bardeen, Surface States and Rectification at a Metal Semiconductor Contact, Phys. Rev., **71,** p. 717–727, May, 15, 1947.
17. I. Tamm, über eine Mögliche Art der Elektronenbindung an Kristalloberflächen, Physik, Zeits, Sowjetunion, **1,** 1932, p. 733.
18. W. H. Brattain and J. Bardeen, Surface Properties of Germanium, B. S. T. J., **32,** pp. 1–41, Jan., 1953.
19. W. L. Brown, n-Type Surface Conductivity on p-Type Germanium, Phys. Rev. **91,** pp. 518–527, Aug. 1, 1953.
20. W. H. Brattain and C. G. B. Garrett, private communication.
21. R. D. Heidenreich, private communication.
22. O. H. Johnson, Germanium and its Inorganic Compounds, Chem. Rev. **51,** pp. 431–469, 1952.
23. M. Kikurchi and T. Onishi, A Thermo-Electrical Study of the Electrical Forming of Germanium Rectifiers, J. App. Phys., **24,** pp. 162–166, Feb., 1953.
24. R. H. Kingston, Water-Vapor Induced n-Type Surface Conductivity on p-Type Germanium, Phys. Rev., **98,** 1766–1775, June 15, 1955.
25. M. V. Sullivan, personal communication.
26. J. T. Law, A Mechanism for Water Induced Excess Reverse Current on Grown Germanium n-p Junctions, Proc. I. R. E., 42, pp. 1367–1370, Sept., 1954.
27. E. Billig, Effect of Minority Carriers on the Breakdown of Point Contact Rectifiers, Phys. Rev. **87,** p. 1060, Sept. 15, 1952.
28. G. W. Snedcor, Statistical Methods, The Iowa State College Press, Ames, Iowa, 1946.
29. W. VanRoosbroeck, Design of Transistors with Large Current Amplification, J. App. Phys., **23,** p. 1411, Dec., 1952.

The Design of Tetrode Transistor Amplifiers

By J. G. LINVILL and L. G. SCHIMPF

(Manuscript received March 7, 1956)

The design of tetrode transistor amplifiers encounters problems of the type that occurs with other transistor uses. Desired frequency characteristics, limitations of parasitic elements, and other practical considerations impose constraints on the range of terminations that can be employed. With many transistors, one can terminate a transistor so that it will oscillate without external feedback; this oscillation or other exceedingly sensitive terminations must be avoided.

The two-port parameters of the transistor in any orientation in which it is to be used constitute the fixed or given information which is the starting point of the amplifier design. Using this starting point, methods are developed by which one can select, on simple bases, the kinds of terminations that will be suitable. To facilitate the design of amplifiers, a set of charts has been developed from which one can read power gain and input impedance as functions of the load termination.

Illustrative tetrode amplifiers are described. These include a common base 20-mc video amplifier, a common-emitter 10-mc video amplifier, an IF amplifier centered at 30 mc, and an IF amplifier centered at 70 mc. Predicted and measured gains are compared.

INTRODUCTION

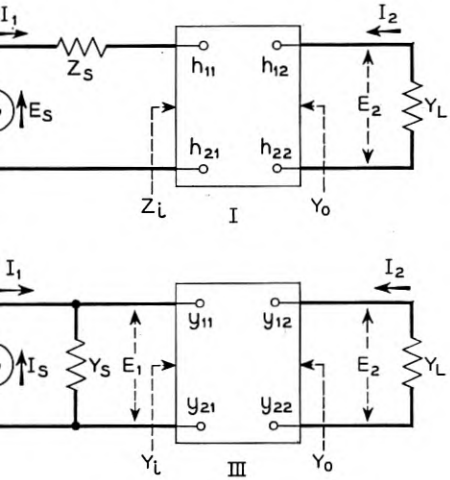
Junction tetrode transistors[1] of the type currently produced for research purposes at Bell Telephone Laboratories are suitable for high-frequency applications. They are being studied for use in video amplifiers, as IF amplifiers where the center frequency is below 100 mc, for oscillators up to 1,000 mc and for very fast pulse circuits.

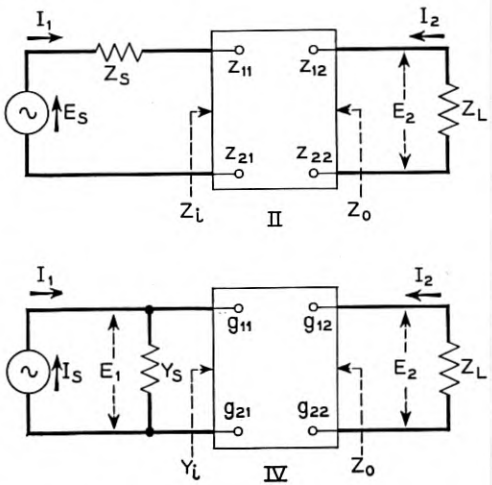
Their application in amplifiers brings up design considerations similar to those encountered for other transistors but with differences resulting

[1] R. L. Wallace, L. G. Schimpf and E. Dickten, A Junction Transistor Tetrode for High-Frequency Use, Proc. I.R.E., **40**, pp. 1,395–1,400, Nov. 1952.

from different parameter values and variation. The analysis presented in this paper regarding amplifier design was motivated by the study of tetrodes, but the results are equally applicable for other types.

The design of an amplifier begins with a characterization of the transistor which is suitable for the study of its performance as an amplifier. From this characterization, or functional representation, one





CORRESPONDING QUANTITIES			
I	II	III	IV
h_{11}	z_{11}	y_{11}	g_{11}
h_{12}	z_{12}	y_{12}	g_{12}
h_{21}	z_{21}	y_{21}	g_{21}
h_{22}	z_{22}	y_{22}	g_{22}
I_1	I_1	E_1	E_1
E_1	E_1	I_1	I_1
E_2	I_2	E_2	I_2
I_2	E_2	I_2	E_2
E_S	E_S	I_S	I_S
Z_S	Z_S	Y_S	Y_S
Y_L	Z_L	Y_L	Z_L
Z_i	Z_i	Y_i	Y_i
Y_o	Z_o	Y_o	Z_o

RELATIONSHIPS BELOW ARE BETWEEN QUANTITIES IN COLUMN I. CORRESPONDING RELATIONSHIPS ARE WRITTEN DIRECTLY FOR CORRESPONDING QUANTITIE IN ANY OTHER COLUMN.

(1) $E_1 = I_1 h_{11} + E_2 h_{12}$

(2) $I_2 = I_1 h_{21} + E_2 h_{22}$

(3) $Z_i = h_{11} - \dfrac{h_{12} h_{21}}{Y_L + h_{22}}$

(4) $Y_o = h_{22} - \dfrac{h_{12} h_{21}}{Z_S + h_{11}}$

(5) $I_2 = \dfrac{h_{21} E_S Y_L}{(h_{11} + Z_S)(h_{22} + Y_L) - h_{12} h_{21}}$

Fig. 1 — Two-port parameters with summary of relationships.

determines the potentialities of amplifiers employing the transistor and designs a suitable amplifier circuit. This step involves answering two questions: What performance, maximum power gain for instance, is it possible to obtain? What source and load impedances should the transistor be associated with?

Two-Port Parameters of Transistors

For circuit applications, the two-port parameters are the most convenient for characterization of the transistor. These parameters implicitly but completely characterize the device from the performance standpoint.

Four sets of two-port parameters are illustrated in Fig. 1. Any set can be calculated from any other set, and the choice of the set to employ is determined only by convenience in the use of available measuring equipment and the preference of the designer. The relationships between parameters, input and output impedances, voltage and current ratios are summarized on Fig. 1. The same expressions given there for h's can be used for any parameter set so long as one uses the corresponding quantities applicable to the desired parameter set.

Though the transistor can be operated as an amplifier with the base, emitter or collector common between the input and output terminal pairs, the two-port parameters for any of the connections can be used to calculate the parameters for any other connection.

For determination of the two-port parameters of tetrode transistors, R. L. Wallace suggested the use of two-terminal impedance measurements with subsequent calculation of the two-port parameters of interest from these. The impedances indicated in Fig. 2 have proved simple to measure at typical operating points with conventional high-frequency bridges. These impedances have been measured at a set of frequencies extending to 30 mc. Because of the number of transistors measured it has been economical to program a digital computer to calculate two-port parameters and other quantities of interest from the measured two-terminal impedances.

THE RELATIONSHIPS OF TRANSISTOR PARAMETERS TO AMPLIFIER PERFORMANCE

Any of the sets of two-port parameters implicitly characterize all of the linear properties of the transistor for the range of frequencies for which the parameters have been measured. As mentioned before, it is necessary to translate the parameters into answers to the following questions. How much amplification can the transistor give at a particular

frequency? What impedance should it be supplied from? What impedance should it feed? What gain will be obtained using a pair of impedances different from the optimum ones? The answering of these and related questions amounts to establishing a convenient means of translating the parameter values into the quantities of interest applying to the amplifier. Such a convenient translating means for solving these problems is described in this section.

Earlier explicit solutions to special cases of the problem are well known. Wallace and Pietenpol[2] have given simple expressions in terms of the transistor parameters for matching input and output impedances and the maximum available gain when the transistor has purely real parameters. An implicit solution for optimum source and load impedances for maximum gain in the complex case has been known for a long time. It is simply that the transistor be terminated at the input and output by conjugate matching impedances. The implicit nature of this solution arises from the fact that the input impedance is a function of the load impedance, and the output impedance is a function of the source impedance for transistors with internal feedback. The solution for optimum source and load impedance from this approach amounts to the solution of simultaneous quadratic equations with complex unknowns and becomes involved.

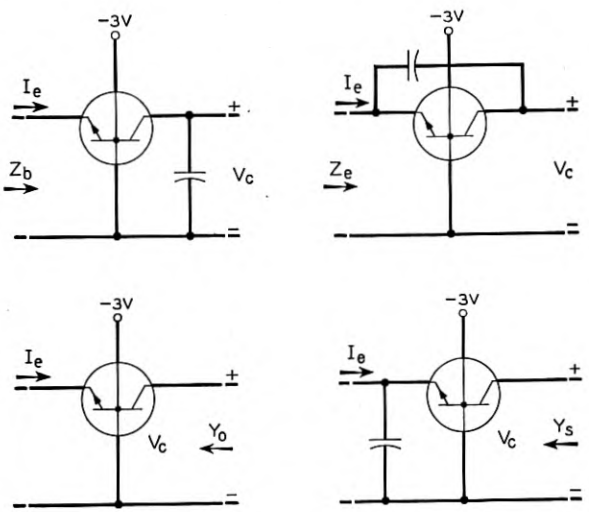


Fig. 2 — Two terminal impedance measurements for determination of two-port parameters.

[2] R. L. Wallace and W. J. Pietenpol, Some Circuit Properties of n-p-n Transistors, Proc. I.R.E., **39**, pp. 753–67, July, 1951.

From the approach to the problem taken in this paper, one solves first for the maximum power gain and subsequently determines the optimum terminations. It turns out that the solutions leads to explicit relationships for optimum performance and terminations and also leads to charts from which power gains and input impedance can be read for any terminations.

In all expressions to be developed, the h parameters are used. Precisely the same expressions can be obtained for z's, y's, or g's provided that one uses the corresponding quantities in the table of Fig. 1.

The maximum power gain is a quantity of primary interest in transistors since the transistor ordinarily has a resistive component in its driving-point impedance. Thus voltage or current amplification is constrained by the limited power gain attainable. In some cases, however, because of the inherent feedback internal to the device, instability can result simply from proper passive terminations without application of any additional feedback. Such cases are distinct because of this property. Transistors exhibiting this possibility are said to be potentially unstable at the frequency in question.

A quantity of interest presented here and derived later is a particular power gain defined for h-parameters as

$$\frac{\text{power out}}{\text{power in}} = \frac{P_{00}}{P_{i0}} = \frac{|h_{21}|^2}{4h_{11r}h_{22r} - 2Re(h_{12}h_{21})} \tag{1}$$

where h_{11r} and h_{22r} mean the real part of h_{11} and of h_{22}. $Re(h_{12}h_{21})$ means the real part of the product of h_{12} and h_{21}. Unless the amplifier is potentially unstable, the quantity P_{00}/P_{i0} is within 3 db of the maximum available gain for the transistor.

The matter of potential instability of the transistor is of great interest. Certainly the transistor is potentially unstable if P_{00}/P_{i0} is negative. Otherwise potential instability is indicated by greater than unity values of the criticalness factor

$$C = 2\frac{P_{00}}{P_{i0}}\left|\frac{h_{12}}{h_{21}}\right| \tag{2}$$

If the transistor is not potentially unstable the maximum available gain is $K_G(P_{00}/P_{i0})$ where

$$K_G = \frac{2(1 - \sqrt{1 - C^2})}{C^2} \tag{3}$$

For $0 \leqq C \leqq 1$, $1 \leqq K_G \leqq 2$. A plot of K_G as a function of C is shown in Fig. 3. The function is seen to be exceedingly flat near $K_G = 1$ for

C between zero and 0.6. Thus the value P_{00}/P_{i0} in the majority of cases where the transistor is not potentially unstable is a close approximation to the maximum available gain.

The optimum source and load impedances can be expressed in terms of the transistor parameters and other quantities given in terms of them by the following relationships where the transistor is not potentially unstable.

$$G = 1 \lfloor Arg\ (\overline{-h_{12}h_{21}}) = e^{j\theta} \tag{4}$$[3]

$$Z_s \text{ opt} = \overline{Z}_{in} = h_{11} - \frac{h_{12}h_{21}}{2h_{22r}}\left(1 - \frac{CK_GG}{2}\right) \tag{5}$$

$$Y_L \text{ opt} = -h_{22} + \frac{2h_{22r}}{1 - \frac{CK_GG}{2}} \tag{6}$$

Though explicit relationships for ideal terminations and for the maximum power gain which one can achieve with a transistor are of interest, such terminations limit the band width of the amplifiers. Therefore, it is important to have convenient means for evaluating power gain and input impedance for other than ideal terminations in order to realize a desired bandwidth. A chart which facilitates computation of these quantities is now developed from an analysis which leads to the other results quoted above.

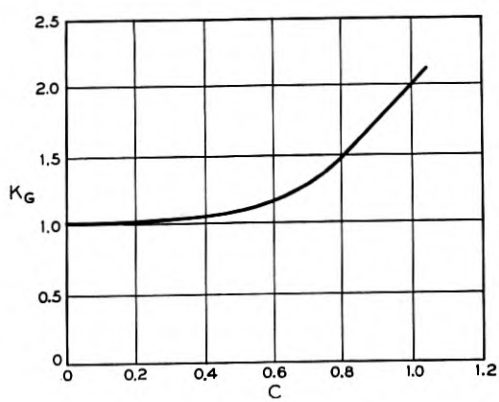


FIG. 3 — K_G plotted as a function of C.

[3] If $\overline{-h_{12}h_{21}}$ is $c + jd$, then $\theta = \tan^{-1}(d/c)$; $G = e^{j\theta}$ and $\overline{-h_{12}h_{21}}$ is the conjugate of $-h_{12}h_{21}$.

Power Flow in a Two-Port Device

A convenient point of departure in the analysis of power amplification in a transistor or other linear two-port device is the arrangement shown in Fig. 4. The two-port is supplied by a unit current at the frequency of interest and at reference phase at the input terminal pair. The output of the two-port is connected to a voltage source of the same frequency. The input-current and output-voltage time functions are

$$i_1 = Re\sqrt{2}\varepsilon^{j\omega t} = Re\sqrt{2}I_1\varepsilon^{j\omega t} \tag{7}$$

and

$$e_2 = Re\sqrt{2}(a + jb)\varepsilon^{j\omega t} = Re\sqrt{2}(L + jM)\left(\frac{-h_{21}}{2h_{22}r}\right)\varepsilon^{j\omega t} = Re\sqrt{2}E_2e^{j\omega t} \tag{8}$$

In (8), L and M are introduced for simplicity in some later relationships.

The whole analysis is essentially a study of power flow in the circuit shown in Fig. 4 as L and M of (8) are varied. All possible terminations and excitations can be simulated simply by varying L and M. Under some conditions the voltage source will absorb power; under others it will supply power to the two-port. Ordinarily the current source supplies power to the two-port, but for appropriate ranges of L and M if the two-port is potentially unstable, the transistor may supply power both to the current source and the voltage source. The problem of evaluating maximum power gain is simply finding the values of L and M corresponding to the greatest ratio of power out to power in. The load impedance to which this situation corresponds is $E_2/-I_2$. The input impedance for this condition is simply E_1/I_1, and the optimum source impedance is the complex conjugate of the latter quantity.

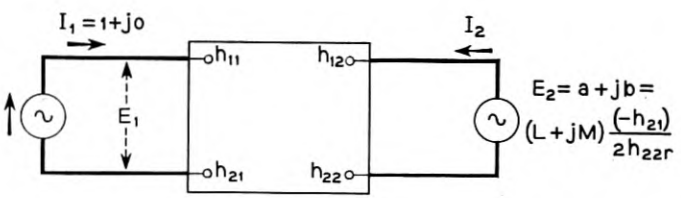


Fig. 4 — A two-port device supplied by a current source and feeding into a voltage source.

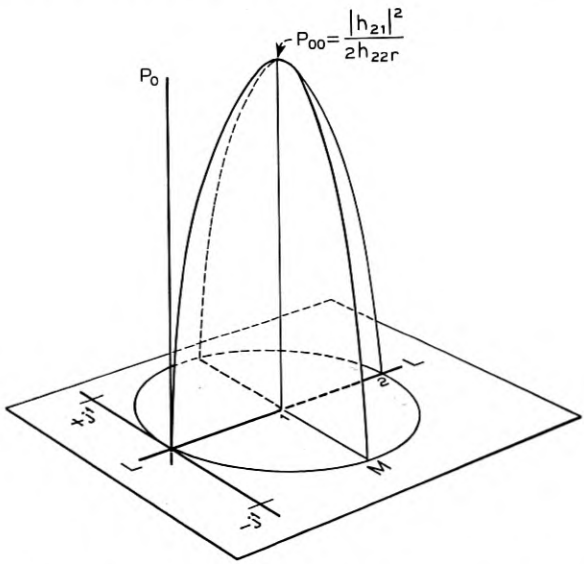


Fig. 5 — Sketch of power output as a function of L and M.

PROJECTION IN L-M PLANE OF GRADIENT LINE IS G OR $\text{Arg}-h_{12}h_{21}$

SLOPE OF PLANE ALONG G IS $\left|\frac{h_{21}h_{12}}{2h_{22r}}\right|$

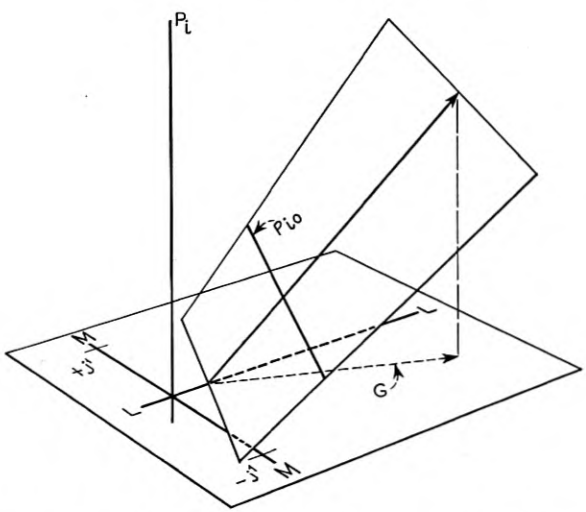


Fig. 6 — Sketch of power input as a function of L and M.

The output power can be readily evaluated in terms of L and M.

$$I_2 = I_1 h_{21} + E_2 h_{22} \tag{9}$$

$$I_2 = (1 + j0)h_{21} + (L + jM)h_{22} \frac{(-h_{21})}{2h_{22r}} \tag{10}$$

$$\text{Power out} = P_0 = Re(-\bar{E}_2 I_2) \tag{11}$$

$$P_0 = Re\left[\frac{(L - jM)\bar{h}_{21}}{2h_{22r}} h_{21} - (L^2 + M^2) \frac{h_{22} \mid h_{21} \mid^2}{4h_{22r}}\right] \tag{12}$$

$$= L \frac{\mid h_{21} \mid^2}{2h_{22r}} - (L^2 + M^2) \frac{\mid h_{21} \mid^2}{2h_{22r}} \tag{13}$$

On the basis of (13) the power output plotted as a function of L and M is a paraboloid as shown in Fig. 5, having the pertinent dimensions indicated there. Only within the circle centered at $L = 1$, $M = 0$ and passing through the origin does one obtain positive power output. The apex of the paraboloid corresponds to

$$P_0 = P_{00} = \frac{\mid h_{21} \mid^2}{4h_{22r}} \tag{14}$$

The input power can similarly be evaluated in terms of L and M.

$$E_1 = I_1 h_{11} + E_2 h_{12} \tag{15}$$

$$= (1 + j0)h_{11} + (L + jM) \frac{(-h_{21})}{2h_{22r}} h_{12} \tag{16}$$

$$\text{Power in} = P_i = Re[E_1 I_1] \tag{17}$$

$$P_i = Re\left[h_{11} + (L + jM) \frac{(-h_{21})h_{12}}{2h_{22r}}\right] \tag{18}$$

$$= h_{11r} - LRe \frac{(h_{12}h_{21})}{2h_{22r}} + MIm \frac{(h_{12}h_{21})}{2h_{22r}} \tag{19}$$

where $Im[(h_{12}h_{21})/2h_{22r}]$ means the imaginary part of the expression in parenthesis.

On the basis of Eq. 19 the input power plotted as a function of L and M is simply an inclined plane having the properties indicated on Figure 6.

Since Figures 5 and 6 turn out to be such simple geometrical figures the problem of finding the point of maximum ratio of P_0 to P_i is very simple and other interpretations are easy to make. First, a negative value of $P_{i0}(P_i$ at 1, 0) certainly indicates potential instability for both input and output terminations receive power from the two-port. Even if the plane of P_i intersects the L-M plane within the unit circle centered at

1, 0, then the two-port is potentially unstable since on one side of the intersection both input and output terminations receive power from the two-port. The change in P_i from the minimum value found on the unit circle centered at 1, 0 to P_{i0} divided by P_{i0} is the criticalness factor, C. Values of C greater than unity indicate potential instability.

The power input at 1, 0 is

$$P_{i0} = \frac{2h_{11r}h_{22r} - Re(h_{12}h_{21})}{2h_{22r}} \tag{20}$$

Using (14) and (20), one obtains

$$\frac{P_{00}}{P_{i0}} = \frac{|h_{21}|^2}{4h_{11r}h_{22r} - 2Re(h_{12}h_{21})} \tag{21}$$

$$C = \frac{\frac{|h_{12}h_{21}|}{2h_{22r}}}{\frac{2h_{11r}h_{22r} - Re(h_{12}h_{21})}{2h_{22r}}} = 2\frac{P_{00}}{P_{i0}}\left|\frac{h_{12}}{h_{21}}\right| \tag{22}$$

Now if the plane of power input, Fig. 6, is parallel to the L-M plane and above it, certainly the point of maximum power gain is the apex of the paraboloid, 1, 0 in Fig. 5. If the plane is inclined but always above the unit circle centered at 1, 0 certainly the point of maximum power gain is downward along the gradient line which lies above the point 1, 0. This must be so since for any contour of equal power out (a circle of fixed elevation around the paraboloid) the minimum power input (or greatest gain) lies along the line of steepest descent from 1, 0 in Fig. 6. Thus the problem of evaluation of the maximum available gain reduces to the simple problem of finding the abscissa of Fig. 7 where the ratio of ordinates of the parabola and straight line is a maximum. The parabola

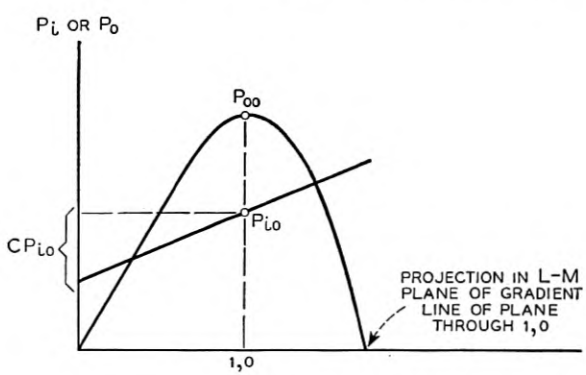


Fig. 7 — Section of paraboloid and inclined plane of Figs. 5 and 6.

and straight line are sections of the paraboloid and plane through the gradient line of the plane over 1, 0.

A straightforward analysis indicates that the point in the L-M plane where the maximum of P_0/P_i occurs is at

$$L + jM = 1 - \frac{CK_G G}{2} \tag{23}$$

where these quantities are defined as in (2), (3), and (4). The power gain at this optimum point is K_G times that obtained at 1, 0. One finds that the maximum gain is only two times P_{00}/P_{i0} even if C approaches unity which corresponds to the marginal case of potential instability.

The analysis just described leads to the maximum values of power gain and to the best terminating impedances. For many design problems these answers are a guide but one may prefer to use other than optimum values for other compelling reasons. For such a case charts from which one can get the pertinent quantities are very helpful.

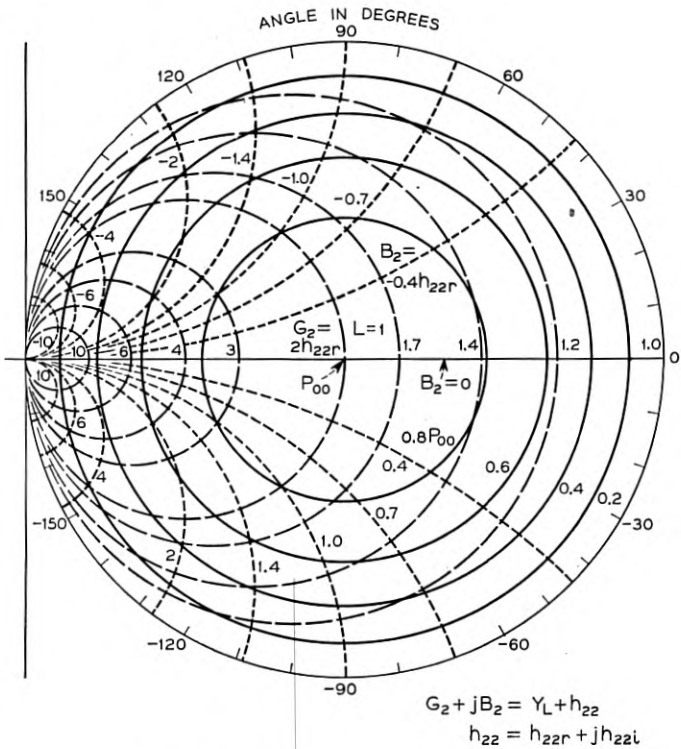


Fig. 8 — Gain and impedance chart.

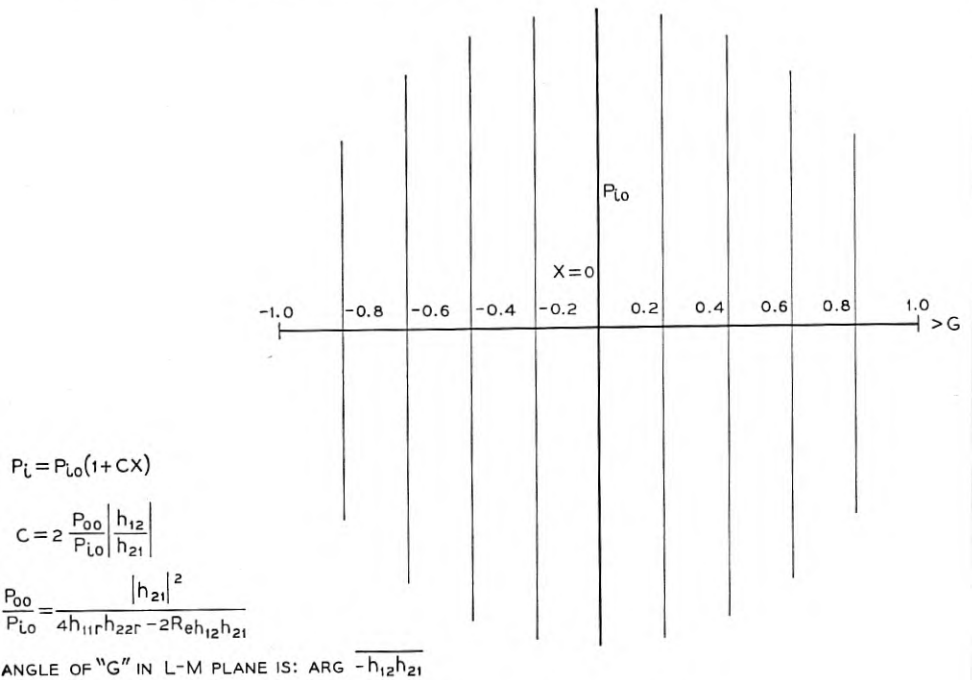


Fig. 9(a) — Input power as a function of X.

Development of Transmission and Impedance Charts

The same point of departure employed in the evaluation of optimum cases leads to a convenient set of charts. Equation 12 shows that a set of concentric circles centered at 1, 0 are loci in the L-M plane of constant power output for a unit current source at the input. It is convenient to plot these as is done on Figure 8, showing P_0 as a fraction of P_{00}.

$$\frac{P_0}{\frac{|h_{21}|^2}{4h_{22r}}} = \frac{P_0}{P_{00}} = 1 - (L-1)^2 - M^2 \tag{24}$$

Since Y_L, the load admittance, is $-I_2/E_2$, using (10) one obtains

$$\frac{-I_2}{E_2} = Y_L = -h_{22} + \frac{2h_{22r}}{L + jM} \tag{25}$$

Now it is clear that if one defines G_2 and B_2 by

$$Y_2 = G_2 + jB_2 = Y_L + h_{22} = \frac{2h_{22r}}{L + jM} \tag{26}$$

loci of constant real and imaginary parts of Y_2 become the mutually orthogonal circles shown in Fig. 8. Thus the value of $L + jM$ is determined by the load admittance and two-port parameters.

Contours representing constant input power, with equal increments of power between successive contours, are always parallel equally-spaced lines in the L-M plane. However, as may be seen from (19) and Fig. 6 different cases have different directions for the line normal to the contours, (the gradient line) and also different power increments for a given spacing of equal-power-input contours. It is convenient to define a new variable X which is the component along the gradient line of the vector starting at $L = 1$, $M = 0$ and going to L, M. Thus

$$P_i = P_{i0}(1 + CX) \tag{27}$$

Equation 27 suggests Fig. 9(a) which shows loci of constant power input plotted as a function of X. If Fig. 9(a) is shown on a transparent ma-

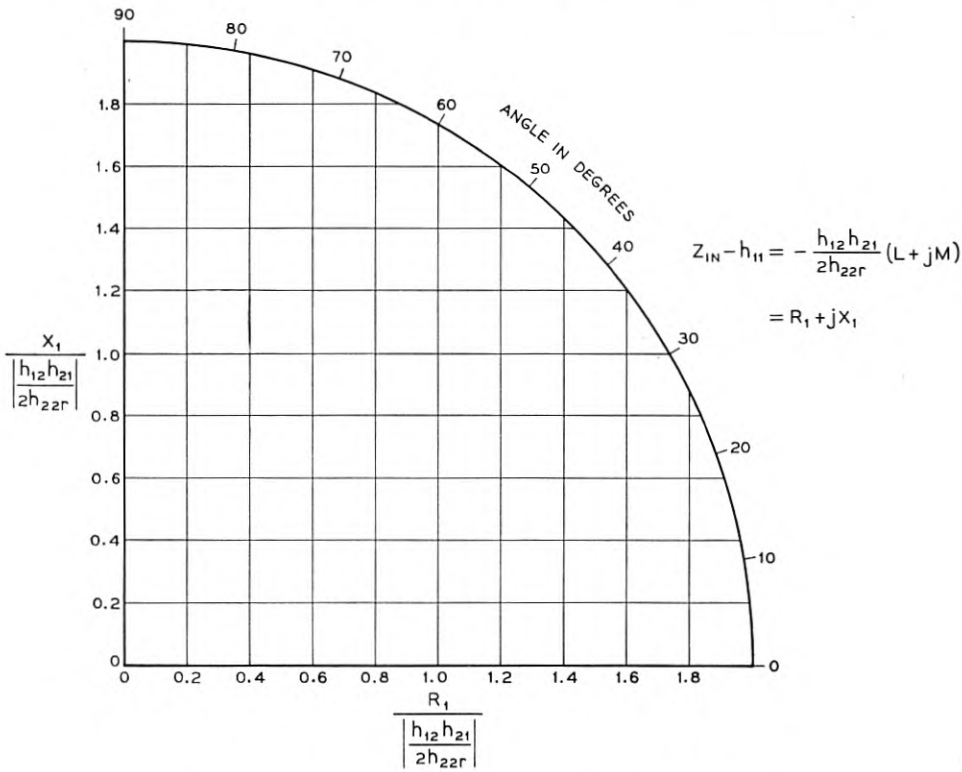


Fig. 9(b) — Input impedance as a function of L and M.

terial, its center (at $X = 0$ along the gradient line) can be superposed with the point $L = 1$, $M = 0$ of Fig. 8. With the gradient line of Fig. 9(a) oriented at the argument of $\overline{-h_{12}h_{21}}$, or θ in the L-M plane, one can easily determine graphically the power gain at any point in the L-M plane compared to the power gain at 1, 0. With Fig. 9(a) superposed on Fig. 8 as just described the viewer gets a bird's-eye-impression of the paraboloid of power output and the inclined plane of power input simultaneously. With such a bird's-eye view, it is easy to assess possibilities for power gain with all possible angles of load termination.

The evaluation in input impedance is done through use of (16) from which is obtained

$$\frac{E_1}{I_1} = Z_{in} = h_{11} + (L + jM)\frac{(-h_{12}h_{21})}{2h_{22r}}, \qquad \text{or} \tag{28}$$

$$Z_{in} = h_{11} + (L + jM)(e^{-j\theta})\left|\frac{h_{12}h_{21}}{2h_{22r}}\right| \tag{29}$$

For evaluating the second component of (29), it is convenient to have a second transparent overlay, Fig. 9(b), consisting of a rectangular grid to the same scale as the L-M plane, Fig. 8, with coordinates marked as

$$Re\frac{(Z_{in} - h_{11})}{\left|\frac{h_{12}h_{21}}{2h_{22r}}\right|} = \frac{R_1}{\left|\frac{h_{12}h_{21}}{2h_{22r}}\right|}$$

and

$$Im\frac{(Z_{in} - h_{11})}{\left|\frac{h_{12}h_{21}}{2h_{22r}}\right|} = \frac{X_1}{\left|\frac{h_{12}h_{21}}{2h_{22r}}\right|}$$

This overlay is placed over the L-M plane with the

$$\frac{R_1}{\left|\frac{h_{12}h_{21}}{2h_{22r}}\right|}$$

axis making the angle θ with respect to the L axis. Thus on the rectangular overlay for any point in the L-M plane, one reads

$$\frac{Z_{in} - h_{11}}{\left|\frac{h_{12}h_{21}}{2h_{22r}}\right|}$$

PARTICULAR DESIGNS OF TETRODE TRANSISTOR AMPLIFIERS

The charts and optimum relationships developed in the preceding section are convenient starting points in the design of amplifiers. They

do not ordinarily constitute a finished solution, however, since practical constraints frequently modify the design used. Moreover, all of the relationships are expressed on a single frequency basis, and many times the amplifier must operate over a range of frequencies broad enough that parameters change significantly over the range.

Four amplifier designs are described in this section: a single stage, common-base, 20-mc video amplifier; a common-emitter, 10-mc video amplifier; an IF amplifier at 30 mc and a 60 to 80-mc IF amplifier. Parameter measurements made with bridges support the first three designs.

Parameter values and associated constants of a typical tetrode transistor are given in Table I. The quantities shown there reveal some interesting facts about the typical tetrode transistor represented. First, in the common-base connection the tetrode is potentially unstable at 30 mc but not at the lower frequencies. The common-emitter amplifier is potentially unstable at 1 and 3 mc. Second, the power gains of common-emitter and common-base stages are about the same at 30 mc, the common-emitter connection giving more gain at low frequencies.

The matter of potential instability requires further consideration from a practical point of view. Potential instability at a frequency neither implies that a stable amplifier cannot be built at that particular frequency, nor does it imply that one can obtain an unlimited amount of stable amplification at that frequency. It does mean that by simultaneously tuning output and input one can adjust for oscillation. The region of potential instability corresponds to a region in which the input resistance may be negative for appropriate loads. Instability is avoided in the physical amplifier if one supplies the amplifier from a sufficiently high impedance that the input loop impedance always has a positive real part. To operate the amplifier with such a load that it presents a negative resistance to the source is attended by the difficulty that the amplification is more sensitive to changes in the source impedance than it is when the input resistance is positive. Hence the possible higher gain with internal positive feedback goes along with a greater sensitivity to changing termination impedance.

A Common-Base 20-Mc Video Amplifier

The data presented in Table 1 gives a quite comprehensive picture of possibilities for amplifier designs. To it must be added a practical fact. It is difficult to connect the load impedance without adding about 2 $\mu\mu$f of capacitance. This means that any termination considered must include about this amount of capacitance. By a theorem regarding

TABLE I — PARAMETERS AND ASSOCIATED CONSTANTS OF TETRODE No. 668

Freq. Mc	0.6	1.0	3.0	10.0	30.0
Common Base					
h_{11}	$49 + j0.0$	$49 + j2.0$	$49 + j2.0$	$50 + j2.0$	$51 + j4.0$
h_{12}	$(1.1 + j0.24)\cdot 10^{-3}$	$(1.1 + j0.40)\cdot 10^{-3}$	$(1.2 + j1.3)\cdot 10^{-3}$	$(1.3 + j4.6)\cdot 10^{-3}$	$(2.1 + j11)\cdot 10^{-3}$
h_{21}	$-0.93 + j0.015$	$-0.93 + j0.023$	$-0.92 + j0.055$	$-0.88 + j0.13$	$-0.82 + j0.24$
h_{22}	$(3.3 + j5.3)\cdot 10^{-6}$	$(3.7 + j8.9)\cdot 10^{-6}$	$(6.8 + j25)\cdot 10^{-6}$	$(28 + j57)\cdot 10^{-6}$	$(35 + j140)\cdot 10^{-6}$
P_{00}/P_{i0}	310	310	230	87	40
C	0.78	0.79	0.89	0.94	1.3
$-\theta$	11°	20°	45°	65°	64°
$\left\lvert\frac{h_{12}h_{21}}{2h_{22r}}\right\rvert$	160	150	120	77	150
Common Emitter					
h_{11}	$720 - j160$	$660 - j210$	$440 - j270$	$220 - j210$	$130 - j140$
h_{12}	$(2.0 + j2.8)\cdot 10^{-3}$	$(3.0 + j4.3)\cdot 10^{-3}$	$(8.0 + j6.9)\cdot 10^{-3}$	$(1.5 + j1.0)\cdot 10^{-3}$	$(18 - j2.2)\cdot 10^{-3}$
h_{21}	$13 - j3.3$	$12 - j4.3$	$7.2 + j5.6$	$2.8 - j4.1$	$0.98 - j2.7$
h_{22}	$(6.4 + j6.4)\cdot 10^{-5}$	$(8.6 + j9.8)\cdot 10^{-5}$	$(1.9 + j17)\cdot 10^{-5}$	$(34 + j18)\cdot 10^{-5}$	$(40 + j17)\cdot 10^{-5}$
P_{00}/P_{i0}	1570	1330	567	121	44.4
C	0.80	1.1	1.3	0.75	0.57
$-\theta$	−139°	−145°	−177°	128°	103°
$\left\lvert\frac{h_{12}h_{21}}{2h_{22r}}\right\rvert$	359	381	250	111	65

passive impedances[4] this puts an upper limit on the level of impedance presented by the load over a band of frequencies. The greatest possible constant level of load impedance over 20 mc is

$$|Z| = \frac{2}{C\omega} = \frac{2}{2.10^{-12} \cdot 2.10^{7} \cdot 2\pi} = 7{,}960\Omega \qquad (30)$$

Thus number, though not strictly applicable to this case, nonetheless gives a measure of the sort of value which one can expect. Hence one observes that for a broad-band video amplifier the load impedance is certainly going to be considerably less than $1/|h_{22}|$ which up to 10 mc is not less than 15,000 ohms. Moreover, the gain, if it is to be uniform, will certainly be limited by the gain obtainable at 20 mc.

Recognition that the load admittance will be a number of times h_{22r},

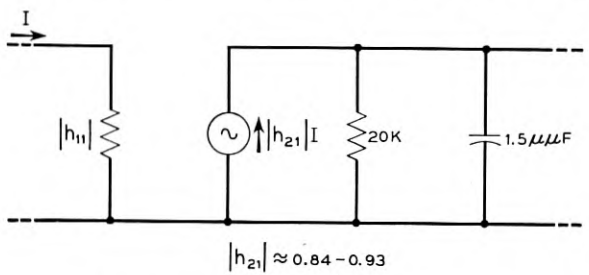


Fig. 10 — A rough approximant for the common base video amplifier. The variation of $|h_{21}|$ is a function of frequency and not variation between units.

five to ten, means that in Fig. 8 one will be operating near the origin where $(L + jM)$ is much less than one. Thus Z_{in} in (28) will be approximately h_{11}. Moreover, by superposing Fig. 9(a) on Fig. 8 at the correct angle for a frequency of 30 mc ($\theta = -64°$) one observes that negative power input occurs only in the small section of circle cut-off by a chord running from the 80° to the 155° points on the periphery. This region is quite a way from the likely point of operation. Thus, this points out that the low impedance termination precludes instability due to internal feedback.

If the amplifier is supplied by a 75-ohm source, its output admittance at 30 mc (Equation 4, Figure 1) is $(7.0 + j20) \cdot 10^{-5}$ mho. At 10 mc the output admittance is $(4.2 + j8.8) \cdot 10^{-5}$ mhos.

These computations reveal that the amplifier in the common-base connection appears quite like the model shown in Fig. 10. Clearly, the

[4] H. W. Bode, Network Analysis and Feedback Amplifier Design, D. Van Nostrand Co., New York, 1945.

about 1 db at frequencies up to 10 mc. Comparison of the measured and computed values is shown on Fig. 15 for a load of 500 ohms with no high-frequency compensation. The low-frequency gain is higher than for the common base connection but the response is down 3 db at 7 mc. By using the combination of R_1 in parallel with 800 $\mu\mu f$ in the emitter circuit, negative feedback is introduced at low frequencies which results in the reduction of low frequency gain tending to make the response more uniform. In addition the L-C network has been added in the output to compensate for the drop of $|h_{21}|$ with increasing frequency and the increasing effect of the output capacitance.

This results in the response shown as the dotted curve on Fig. 15. The low-frequency gain has been reduced to 17.5 db, but the response is now flat to within ± 0.3 db up to 13 mc and is 3 db down at 18 mc.

Although the data given on video amplifiers shows the results obtained using one transistor, similar response curves were obtained from some 6 or 8 units.

An I-F Amplifier Centered at 30 Mc.

The design of an IF amplifier at 30 mc is distinct from the preceding two cases in that one can use matching techniques over the narrow band.

Reference to Table 1 reveals that the common-base connection provides more potential gain at 30 mc than the common emitter connection; in fact, the common-base connection can be made to oscillate with certain terminations. The common-base connection is chosen for the 30-mc amplifier.

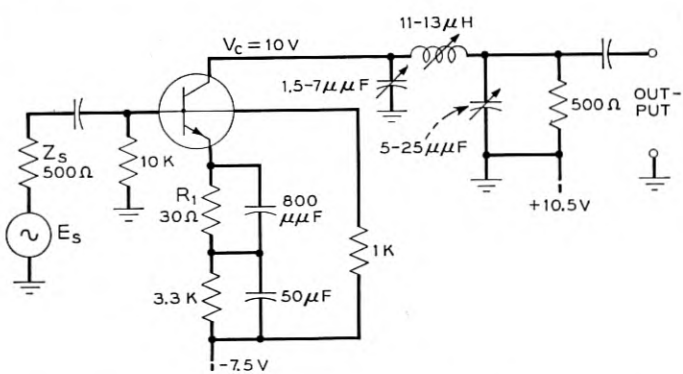


Fig. 14 — Circuit of a common emitter video amplifier. R_1 in parallel with 800 $\mu\mu f$ and the LC network in the output circuit peak the response at 10 to 12 mc.

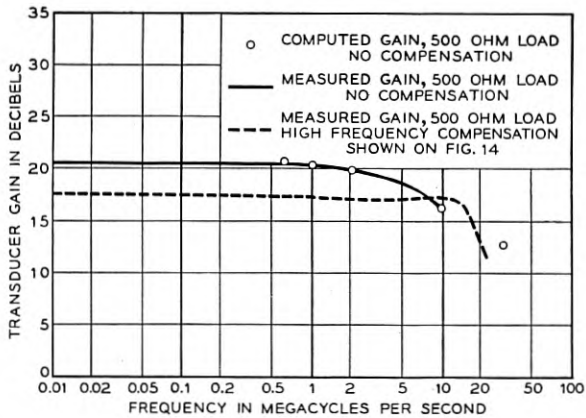


Fig. 15 — Computed and measured response of a common emitter amplifier.

In the design of the IF amplifier one is interested in a moderate range of frequencies. It will generally be true that the most frequency dependent parameters are the output and load admittances, since the load is to be tuned. One can take as a suitable load a parallel combination of a fixed conductance with a frequency dependent susceptance, the sort of termination typical of tuned circuits. Thus on Fig. 8, the locus of $G_2 + jB_2$ is one of the $G_2 =$ Const. circles.

Superposition of Fig. 9(b) on Fig. 8 with the

$$\frac{R_1}{\left|\dfrac{h_{12}h_{21}}{2h_{22r}}\right|}$$

axis making an angle of $-64°$ with the L axis reveals that $Z_{in} - h_{11}$ has a negative real part on the upper left edges of all of the contours of constant G_2. On the $G_2 = 2h_{22r}$ contour, $Re(Z_{in})$ reaches a minimum of 22.5 ohms. We select a load with $G_L = 2h_{22r}(G_2 = 3h_{22r})$ to avoid low values of input resistance resulting from the internal feedback.

Superposition of Fig. 9(a) on Fig. 8 with the gradient line making an angle of $-64°$ through the point L, $M = 1, 0$ reveals that the maximum value of P_0/P_i on the $G_2 = 3h_{22r}$ circle is 1.87 P_{00}/P_{i0} and it occurs for $B_2 = -2h_{22r}$. The input impedance at this point is $36 + j87$ ohms.

For an amplifier one is primarily interested in

$$\frac{P_0}{\text{Power Available from Source}}$$

(which is called transducer gain) rather than P_0/P_i, the quantities just

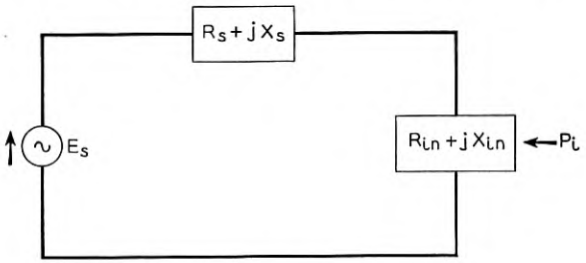


Fig. 16 — Typical input circuit.

read from the charts. From the source-load arrangement shown in Fig. 16, one readily computes

$$\frac{P_i}{\text{Power Available from Source}} = \frac{\dfrac{|E_s|^2 R_{in}}{(R_s + R_{in})^2 + (X_s + X_{in})^2}}{\dfrac{|E_s|^2}{4R_s}} = \frac{4R_s R_{in}}{(R_s + R_{in})^2 + (X_s + X_{in})^2} \tag{32}$$

The source impedance selected for the amplifier is $75 - j87$ ohms at 30 mc. The 75 ohms is selected to reduce the effect of variations in input impedance when it is reduced further by the internal feedback. The 87 ohms of capacitive reactance is selected to tune the input reactance at the peak of response. Using Fig. 8 with the overlay of Fig. 9(a) along with (32) under the assumption that X_s varies insignificantly over the frequencies involved one obtains Table II. This table shows the variation of transducer gain as the value of B_2 is changed as well as indicating the value of B_2 required for the maximum gain. Thus, if the total output capacitance is known, the load admittance required to give the maximum gain at the desired frequency can be computed. As will be shown

TABLE II — EVALUATION OF TRANSDUCER GAIN OF I-F AMPLIFIER

B_2	$-5h_{22r}$	$-4h_{22r}$	$-3h_{22r}$	$-2h_{22r}$	$-h_{22r}$	Oh_{22r}	h_{22r}
P_0/P_i	50	64	73	75	59	50	38
Z_{in}	$22 + j48$	$23 + j58$	$28 + j73$	$36 + j87$	$67 + j99$	$96 + j94$	$120 + j67$
P_i/Power available from source	0.61	0.66	0.78	0.87	0.98	0.98	0.94
Transducer gain	30	42	57	66	58	49	36
Gain, db	14.8	16.2	17.5	18.2	17.6	16.9	15.5

below, the bandwidth at which the response is down a given number of db can also be computed.

From Table II one observes that this design provides a gain of about 18 db with half power frequencies where the susceptance B_2 has changed by $\pm 3h_{22r}$ mhos from its value of $-2h_{22r}$ at the center of the pass band. The value of h_{22i} corresponds to approximately 1 $\mu\mu$f of capacitance and if the stray capacitance amounts to 3.5 $\mu\mu$f, then the bandwidth is $\Delta B2/2C$ since the slope of the susceptance of a tuned circuit is

$$2C \frac{\text{mhos}}{\text{rad/sec}}$$

Thus the bandwidth is approximately

$$\frac{6 \cdot 3.5 \cdot 10^{-5}}{2 \cdot 2.5 \cdot 10^{-12} \cdot 6.28}$$

or 3.7 mc. This is the actual value of load capacitance measured on an experimental amplifier with a vacuum tube voltmeter connected to the output. The measured response of this amplifier with a load of $Y_L = (68 - j215) \cdot 10^{-6}$ at 30 mc ($G_L \doteq 2h_{22r}$) shows a peak gain of 18.3 db and half power points separated by 3.8 mc. For a given value of G_L, the bandwidth of the amplifier will vary inversely with the total capacitance in the output circuit. The same gain as obtained in the sample given above, can be obtained over a narrower band by increasing the load capacitance. Since the minimum capacitance is fixed, if one wishes to increase the width of the pass band, a higher value of G_2 must be used. In the same manner as is used to arrive at the data shown on Table II, Table III is computed for a value of $G_2 = 6h_{22r}$ ($G_L = 5h_{22r}$).

In this case, the maximum value of P_0/P_i occurs when $B_2 = -3h_{22r}$. The source impedance is selected to be $75 - j45$ ohms at 30 mc and the remainder of the table is computed. The maximum computed gain is approximately 16 db with half power frequencies where the susceptance

TABLE III — EVALUATION OF TRANSDUCER GAIN OF I.F. AMPLIFIER

B_2	$-8h_{22r}$	$-5h_{22r}$	$-3h_{22r}$	$-2h_{22r}$	$-h_{22r}$	$+h_{22r}$	$+4h_{22r}$
P_0/P_i	25	33	43	40	39	31	21
Z_{in}	$35 + j22$	$35 + j35$	$50 + j45$	$56 + j46$	$65 + j48$	$77 + j39$	$83 + j20$
P_{in}/Power available from source	0.83	0.86	0.96	0.97	0.99	0.99	0.97
Transducer gain	21	28	41	39	39	31	20
Gain db	13.2	14.5	16.1	15.9	15.9	14.9	13.0

B_2 has changed by $\pm 6h_{22r}$ mhos from its value at the center of the band. Using the same value of circuit capacitance as above, the indicated bandwidth is about 7.4 mc. The measured response on an amplifier with this value of load impedance indicates a gain of 16.1 db at 30 mc, with the frequencies at the half power point separated by 7.6 mc.

Often it is desirable to build tuned amplifiers to work between like impedances in which case at least the output network must perform both the function of selectivity and impedance transformation. An example of a simple network to perform these functions is shown on Fig. 17. The impedance transforming properties of such a circuit are well known. With a given value of load resistance, the load admittance presented to the transistor can be made to have a given value at a certain frequency. However, since the circuit performs both the function of impedance transformation and selectivity the bandwidth is determined by the output impedance selected. This circuit does not present a fixed value of conductance as a function of frequency but for frequencies near the maximum gain it is a fair approximation to assume it constant. The output circuit of Fig. 17 was designed to present a load admittance such that $G_2 + jB_2 = 3h_{22r} - j2h_{22r}$ at 30 mc. This is the same condition as computed in Table II so one would expect the same value of maximum gain. However, in order to present the proper value of load impedance, a total load capacitance of about 4 $\mu\mu$f must be used. This indicates a bandwidth of 3.3 mc between the half power points. The measured response of this amplifier is shown on Fig. 18 as the solid line. The points indicate the computed maximum gain and the frequencies at which the gain is down 3 db.

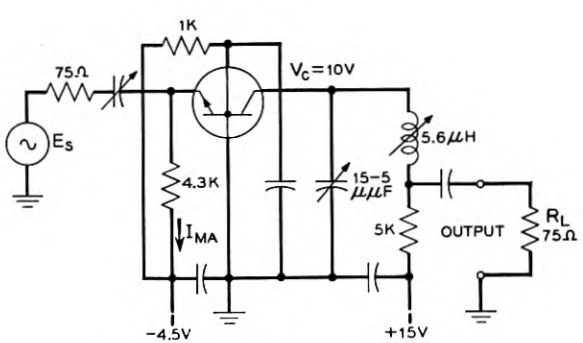


Fig. 17 — Simple tuned amplifier. The output circuit performs both the functions of impedance transformation and selectivity.

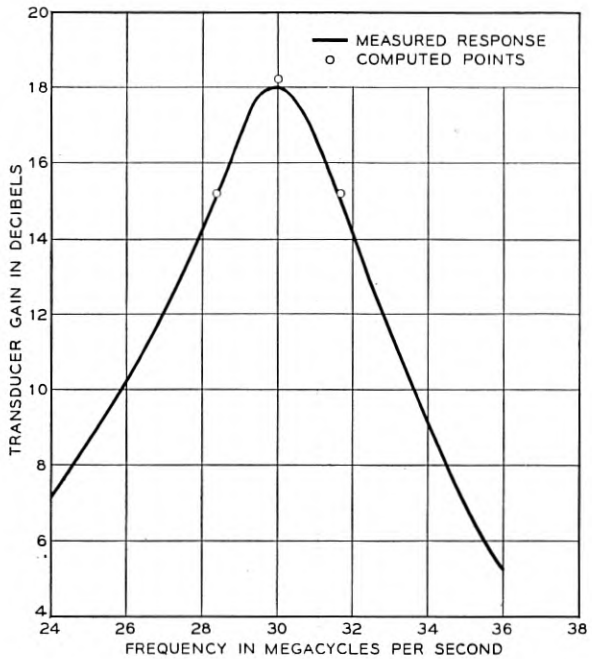


Fig. 18 — Measured and computed response of the stage shown on Fig. 17.

An IF Amplifier Centered at 70 Mc.

Although we do not have complete data on the parameter values of tetrode transistors in this frequency range, amplifiers with a center frequency of 70 mc have been built and their performance measured. The amplifier was designed to provide a flat gain characteristic over the frequency range from 60 to 80 mc. The stage was designed with the equivalent of a double tuned transformer, interstage circuit with the transformer being replaced by the equivalent tee section. The selective circuit is terminated at its output into the load resistance in the case of the last stage or by the input impedance of the following transistor when it is used as an interstage network. The impedance transformation of the network is approximately 75 ohms to 1,500 ohms so it is essentially unterminated at the collector. By using a sweeping oscillator, such a stage can be adjusted to result in a fairly flat frequency response. A typical stage is shown on Fig. 19. The output terminals are connected to either the load or the next emitter. The response obtained from a 3-stage amplifier is shown on Fig. 20. In order to determine the variation of gain

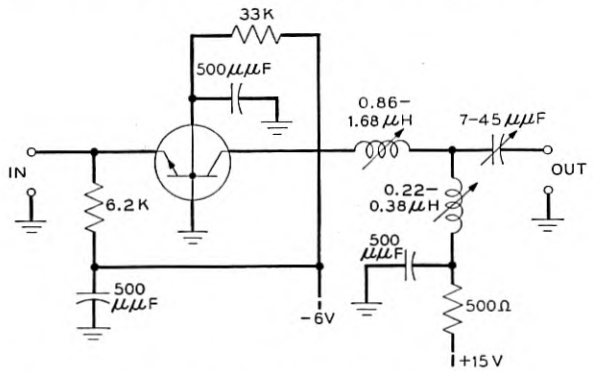


Fig. 19 — Circuit of a 60 to 80-mc band pass amplifier stage.

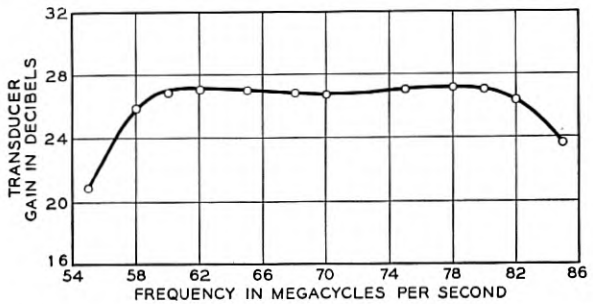


Fig. 20 — Gain of a 3-stage band pass amplifier working between 75-ohm impedances. Each stage uses the circuit shown on Fig. 19.

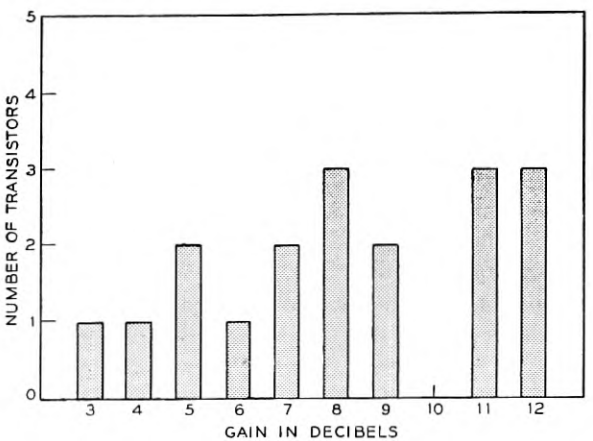


Fig. 21 — Variation of gain for a group of transistors used in the circuit of Fig. 19.

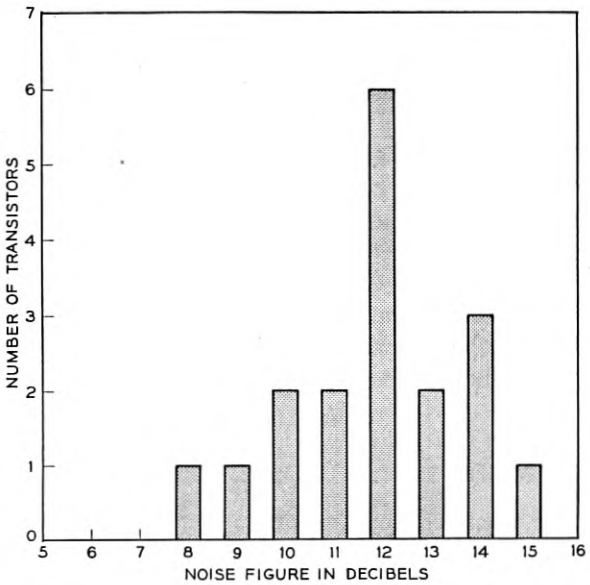


FIG. 22 — Noise figure for a group of transistors used in the circuit of Fig. 19.

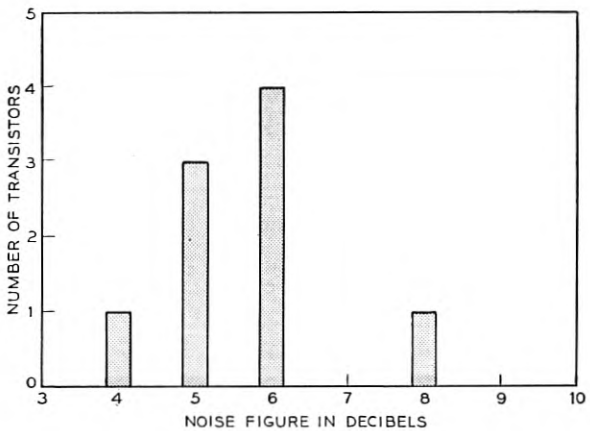


Fig. 23 — Noise figure for a group of transistors used in a 10-mc bandpass amplifier.

between various transistors, 18 tetrodes were measured in the first stage of the amplifier. If the measured gain of each transistor is rounded off to the nearest db and the number of transistors having this gain plotted as the abscissa, the results shown on Fig. 21 are obtained. Of the 18 transistors measured, 11 have a gain of 8 db or greater. Similar data has been obtained on the noise figure of the same 18 transistors, the results being shown on Fig. 22. In general, the transistors having the highest gain also have the lowest noise figure. The noise figure depends to some extent on the source impedance but a 75-ohm source results in a noise figure which is within a few tenths of a db of the minimum. The value of the noise figure does not vary a great deal as the collector voltage and emitter current are changed except that if the collector voltage is lowered below 6 or 8 volts the gain decreases and in general the noise figure increases.

Noise Figure at 10 Mc.

Although not described here, bandpass amplifiers centered at 10 mc with a 200-kc pass band have been constructed using tetrode transistors. A gain of slightly over 20 db per stage can be realized at this frequency. The noise figure of transistors tried in this circuit is shown on Fig. 23, the data being shown in the same manner as described above. At 10 mc the noise figures are lower than at 70 mc. The remarks made above concerning variation of noise figure with operating conditions also apply to this case.

ACKNOWLEDGMENTS

We are happy to acknowledge the advice and encouragement given us by R. L. Wallace, Jr., and others in the Laboratories. We also wish to express our thanks to E. Dickten who fabricated the transistors used to obtain the experimental data presented. W. F. Wolfertz made the transistor parameter measurements used in the computations. R. H. Bosworth and C. E. Scheideler were responsible for construction of the circuits and some of the gain measurements. We also wish to thank W. R. Bennett for his aid in preparing the manuscript.

The Nature of Power Saturation in Traveling Wave Tubes

By C. C. CUTLER

(Manuscript received February 2, 1956)

The non-linear operating characteristics of a traveling wave tube have been studied using a tube scaled to low frequency and large size. Measurements of electron beam velocity and current as a function of RF phase and amplitude show the mechanism of power saturation.

The most important conclusions are:

I. There is an optimum set of parameters ($QC = 0.2$ and $\gamma r_0 = 0.5$) giving the greatest efficiency.

II. There is a best value of the gain parameter "C" which leads to a best efficiency of about 38 per cent.

III. A picture of the actual spent beam modulation is now available which shows the factors contributing to traveling wave tube power saturation.

INTRODUCTION

The highest possible efficiency of the traveling wave tube has been estimated from many different points of view. In his first paper on the subject[1] J. R. Pierce showed that according to small signal theory, when the dc beam current reaches 100 per cent modulation an efficiency of

$$\eta = \frac{C}{2} \tag{1}$$

is indicated,* and thus the actual efficiency might be limited to something like this value. Upon later consideration[2] he concluded that the ac convection current could be twice the dc current and that one might expect an efficiency of

$$\eta = 2C \tag{2}$$

He also considered the effects of space charge, and concluded on the

* Symbols are consistent with Reference 2 and are listed at the end of this paper.

same basis that under high space charge and elevated voltage conditions, efficiencies might be as high as

$$\eta = 8C \tag{3}$$

J. C. Slater[3] on the other hand considered the motion of electrons in a traveling wave and concluded that the maximum possible reduction in beam velocity would also indicate a limiting efficiency of $2C$. Taking a more realistic account of the electron velocity, Pierce[2] showed that these considerations lead to a value of

$$\eta = -4y_1C \tag{4}$$

which, since y_1 ranges between $-\frac{1}{2}$ and -2, leads to the same range of values as the other predictions.

None of these papers purport to give a physical picture of the overloading phenomenon, but only specify clear limitations to the linear theory. L. Brillouin[4] on the other hand found a stable solution for the flow of electrons bunched in the troughs of a traveling wave. This he supposed to represent the limiting high level condition of traveling wave tube operation. His results give an efficiency of

$$\eta = 2bC \tag{5}$$

In the first numerical computations of the actual electron motion in a traveling wave tube in the nonlinear region of operation, Nordsieck[5] predicted efficiencies ranging between 2.5 and 7 times C and showed that there would be a considerable reduction in efficiency for large diameter beams, due to the non-uniformity of circuit field across the beam diameter. He also gave some indication of the electron dynamics involved. Improving on this line of attack, Poulter[6] calculated some cases including the effect of space charge and large values of C.

Tien, Walker and Wolontis[7] carried computations still further for small values of C by including the effect of small beam radii upon the space charge terms, and showed that space charge and finite (small) beam radii result in much smaller efficiencies than were previously predicted. J. E. Rowe[8] got similar results and gave more information on the effects of finite values of C. Computations for large values of C by Tien[9] showed that a serious departure from the small C conditions takes place above values of $C = 0.1$ if space charge is small (i.e., below $QC = 0.1$) and above $C = 0.05$ for larger values of space charge. They indicated that a maximum value of efficiency as high as 40 per cent should be possible using $C = 0.15$, $QC = 0.1$ and elevated beam voltages.

These five papers give some insight into the electron dynamics of power

saturation, but still involve questionable approximations which make it desirable to compare predictions with the actual situation.

Theoretical considerations of the effects of attenuation upon efficiency have not led to conclusions coming even close to the observed results. Measured characteristics[10, 11] show that the effect of attenuation is very large, but that attenuation may be appropriately distributed to attain stability and isolation between input and output of the tube without degrading the output power.

There are also several papers in the French and German periodicals which deal with the question of traveling wave tube efficiency. Some of these are listed in References 12 through 20.

This paper describes measurements of efficiency and of beam modulation made on a traveling wave tube scaled to large size,* and low frequencies. The construction of the tube, shown in Fig. 1, and the measurement of its parameters were much more accurate than is usual in the design of such tubes. The results are believed to be generally applicable to tubes having similar values of the normalized parameters.

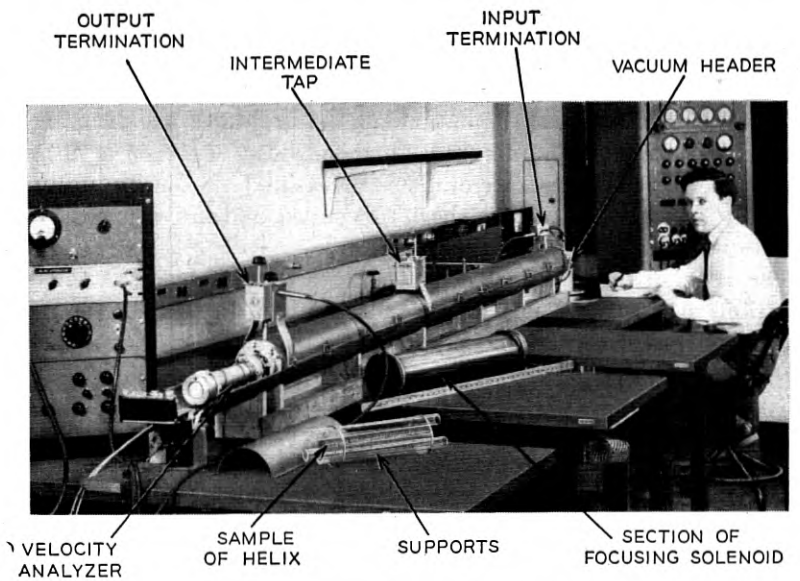


Fig. 1 — The scale model traveling wave tube. The tube is 10 feet long with a copper helix supported by notched glass tubing from an aluminum cylinder overwound with a focusing solenoid. It is continuously pumped and readily demountable.

* See Appendix.

Two kinds of measurements are described. First, the efficiency and power output are determined for various conditions of operation, and second the spent beam ac velocity and current are measured. The principal results are shown in Figs. 2 to 4 which give the obtainable efficiencies, and in Figs. 7 to 10 which show some of the factors which contribute to power saturation. These figures are discussed in detail later. The most significant phenomenon is the early formation of an out-of-phase bunch of electrons which have been violently thrown back from the initial bunch, absorbing energy from the circuit wave, and inhibiting its growth. The final velocity of most of the electrons is near to that of the circuit wave which would lead to a value of

$$\text{limiting efficiency } \eta = -2y_1C \tag{6}$$

if the wave velocity maintained its small signal value. Actually the wave slows down, under the most favorable conditions giving rise to a somewhat higher efficiency. For other conditions, space charge, excess electron velocity, or nonuniformity of the circuit field enter in various ways to prevent the desired grouping of electrons and result in lower efficiencies.

The observed efficiencies are a rather complicated function of QC, γr_0 and C. To compare with efficiencies obtained from practical tubes one must account for circuit attenuation and be sure that some uncontrolled factor such as helix non-uniformity and secondary emission is not seriously affecting the tubes' performance. Measured efficiencies of several carefully designed tubes have been assembled and are compared with the results of this paper in Table I.

The results of these measurements compare favorably with the computations of Tien, Walker and Wolontis[7], and of Tien[9]. There are, however some important differences which are discussed in a later section.

TRAVELING WAVE TUBE EFFICIENCY MEASUREMENTS

Reasoning from low level theory, efficiency should be a function of the gain parameter, "C," the space charge parameter "QC," the circuit attenuation, and (for large beam sizes), the relative beam radius "γr_0." It was soon found that efficiency is a much more complicated function of γr_0 than expected. The initial objective was to determine the effect of QC, C, and γr_0 separately on efficiency, but it was necessary to give a much more general coverage of these parameters, not assuming any of them to be small.

Most of the measurements have been made with small values of loss

TABLE I

Laboratory	Freq. mc.	QC	γr_0	C	η measured	η (from Fig. 3)	η (From Fig. 3 with allowance for circuit attenuation[10]
McDowell*	4,000	0.27	0.62	0.078	**19.5**	26	**21.6**
	6,000	0.29	0.8	0.058	**13.2**	16.2	**12.5**
Brangaccio and Cutler†	4,000	0.61	0.87	0.041	**11**	6	**6**
Danielson and Watson*	11,000	0.35	1.2	0.05	**6.6**	7	**4.8**
R. R. Warnecke[16, 17, 18]	870	0.32	0.3	.125	**27**	33	**33**
W. Kleen and W. Friz[15]	4,000	0.5	0.43	0.05	**7.8**	11.5	**5.7**
W. Kleen‡	4,000	0.2	0.94	0.1	**20**	26	**22**
L. Brück§	3,500	0.19	0.6	0.065	**15**	23	**18.5**
Hughes Aircraft Co.	3,240	0.19	0.94	0.12	**39**	31	**29**
	9,000	0.15	1.3	0.11	**25**	15.5	**12.7**

* At Bell Telephone Laboratories.
† Reference 10 (a slight beam misalignment could account for most of this difference).
‡ Siemens & Halske, Munich, Germany.
§ Telefunken, Ulm, Germany.

and of the gain parameter, where efficiency is proportional to C, as expected from small-signal small-C predictions. This reduces the problem to a determination of η/C versus QC and γr_0.

Many measurements of this kind have been made, and the data are summarized in Figs. 2 and 3, with efficiency shown as a function of QC and γr_0. In Fig. 2 we have the efficiency when the beam voltage is that which gives maximum low-level gain. Fig. 3 shows the efficiency obtained when the beam potential is raised to optimize the power output, and contours of constant efficiency have been sketched in. There is significantly higher efficiency than before in the region of maximum efficiency, but not much more elsewhere.

Fig. 4 shows how efficiency varies with C for a small value of QC, a representative value of γr_0, and with beam voltage increased to maximize the output. This indicates a maximum of about 38 per cent at $C = 0.14$.

Some of the computed results of Tien, Walker and Wolontis,[7] and of Tien[9] are also indicated in the figures. Their results generally indicate somewhat greater efficiencies than were observed, but in the most significant region the comparison is not too bad as will be seen in a later section.

The measurements are for conditions having negligible circuit loss near the tube output. There are no new data on the effect of loss, but earlier results[10] have been verified by measurements at Stanford University[11] and are still believed to be a satisfactory guide in tube design.

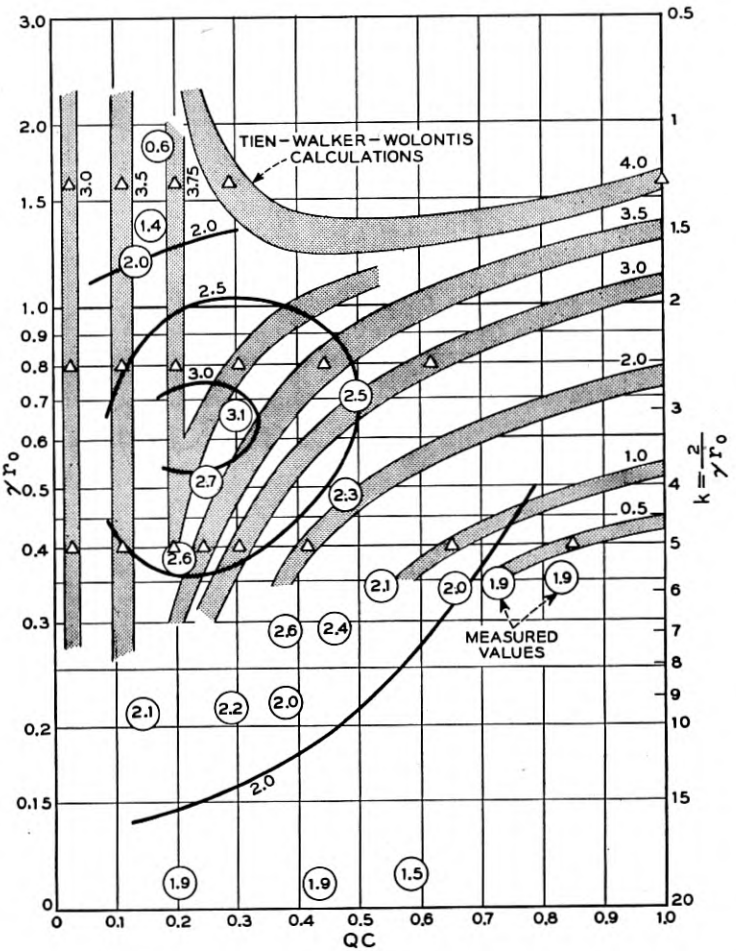


Fig. 2 — Values of efficiency/C as a function of QC and γr_0 at the voltage giving maximum gain per unit length. The shaded contours and triangular points are from the computations[7] of Tien, Walker and Wolontis. The circled points are from the measurements and the line contours are estimated lines of constant efficiency. The most significant difference is for large beam radii, where the RF field varies over the beam radius in a way not accounted for in the computations.

SPENT BEAM CHARACTERISTICS

The scale model traveling wave tube was followed by a velocity analyzer as sketched in Fig. 5 and described in the Appendix. A sample of the beam at the output end of the helix is passed through a sweep circuit to separate electrons according to phase, and crossed electric and

magnetic fields to sort them according to velocity. The resulting beam draws a pattern on a fluorescent screen as shown in Fig. 6 from which charge density and velocity can be measured as a function of signal phase. The velocity coordinate is determined by photographing the ellipse with several different beam potentials, as in Fig. 6(a), and the phase coordinate is measured along the ellipse. From pictures like this a complete determination of electron behavior is obtained from the linear region up to and above the saturation level.

The results of such a run are plotted in Fig. 7. The upper lefthand

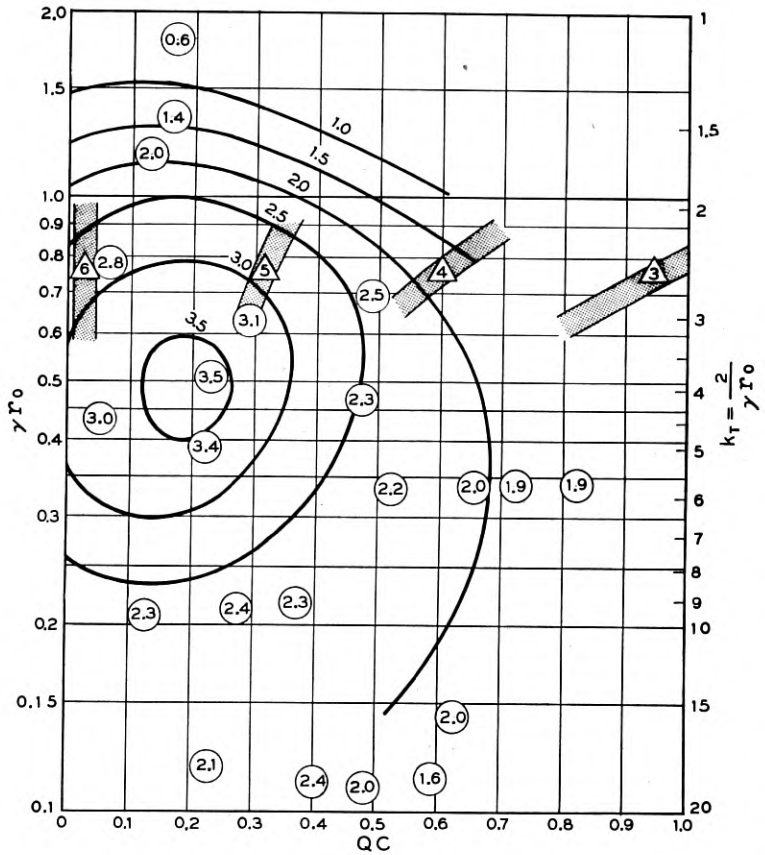


Fig. 3 — Values of efficiency/C as a function of QC and γr_0 at elevated beam voltage. Raising the beam voltage has little effect at large QC and small γr_0, and less than expected anywhere. Again the triangular points are from Tien, Walker and Wolontis,[7] and the line contours are estimated from the measured data.

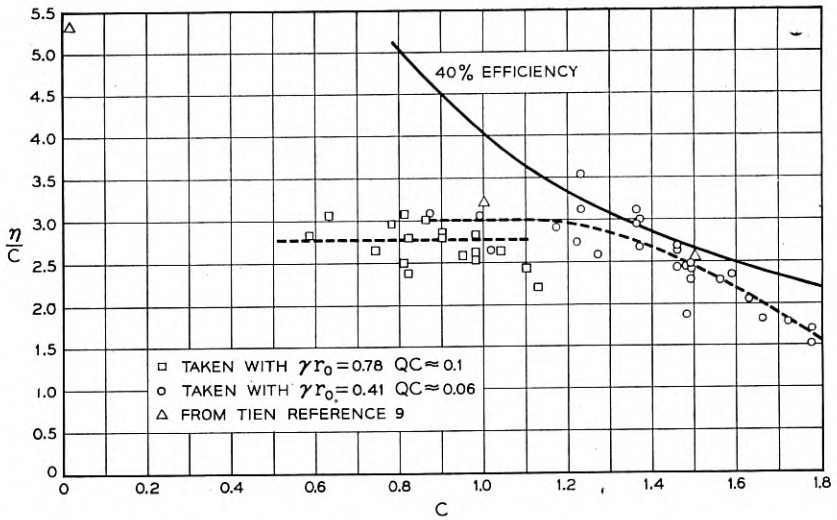


Fig. 4 — Efficiency/C for large values of C and with elevated beam voltage. Efficiency seriously departs from proportionality to C at $C = 0.14$, where a maximum efficiency of about 38 per cent is measured.

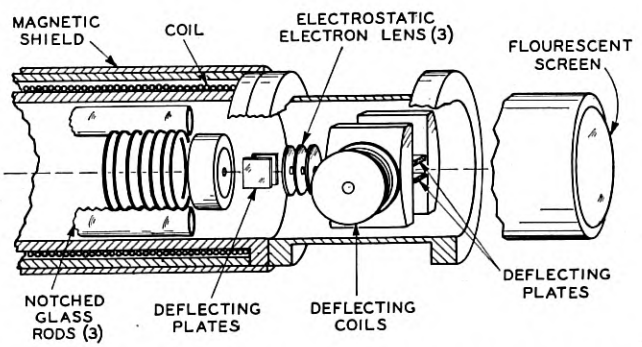


Fig. 5 — The velocity analyzer. A sample of the spent electron beam is accelerated to a high potential, swept transversely with a synchronous voltage, sorted with crossed electric and magnetic fields, and focused onto a fluorescent screen.

pattern, Fig. 7(a), is representative of the low level (linear) conditions (22 db below the drive for saturation output). The dashed curve represents the voltage on the circuit, inverted so that electrons can be visualized as rolling down hill on the curve. The phase of this voltage relative to the electron ac velocity is computed from small signal theory, but

everything else in Fig. 7, including subsequent variations of phase, are measured. The solid line patterns represent the ac velocity, and the shaded area, the charge density corresponding to that velocity. Thus in each pattern we have a complete story of (fundamental) circuit voltage, electron velocity and current density as a function of phase, for a particular signal input level. The velocity and current modulations at small signal levels check calculated values well, and it is not difficult to visualize the dynamics giving this pattern.

Consider first the situation in the tube at small signal amplitudes. At the input an unmodulated electron beam enters the field of an electromagnetic wave moving with approximately the same velocity as the electrons. The electrons are accelerated or decelerated depending upon their phase relative to the wave, and soon are modulated in velocity. The velocity modulation causes a bunching of the electrons near the potential maxima (i.e., the valleys in the inverted potential wave shown) and these bunches in turn induce a new electromagnetic wave component onto the circuit roughly in quadrature following the initial wave. The addition of this component gives a net field somewhat retarded from the initial wave and larger in amplitude. Continuation of this process

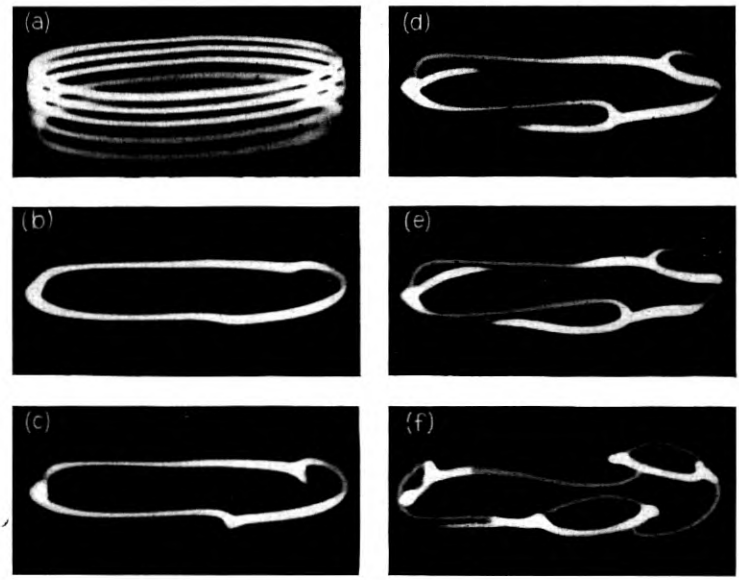


Fig. 6 — Velocity analyzer patterns. The beam sample is made to traverse an ellipse at ⅓ the signal frequency. Current density modulation appears as intensity variation, and velocity variation as vertical deflection from the ellipse.

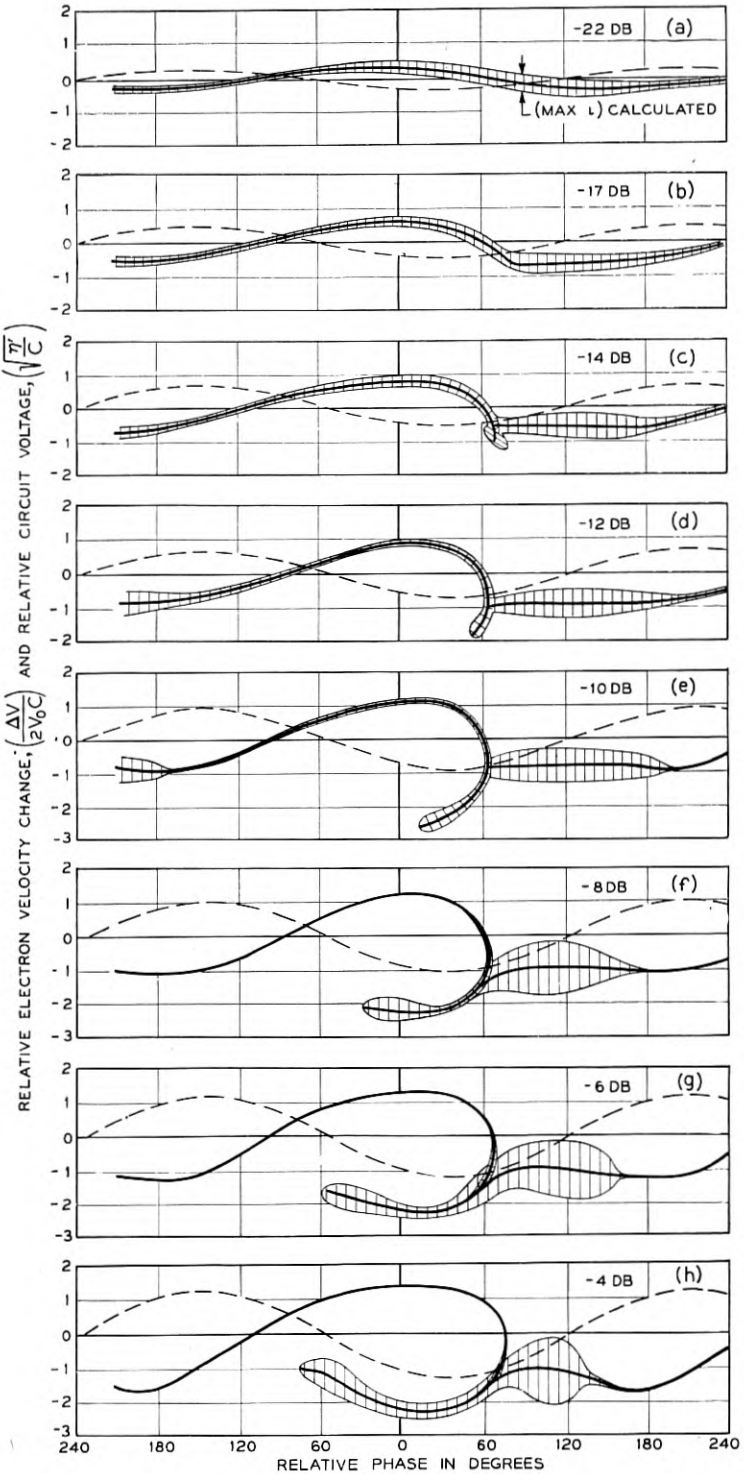
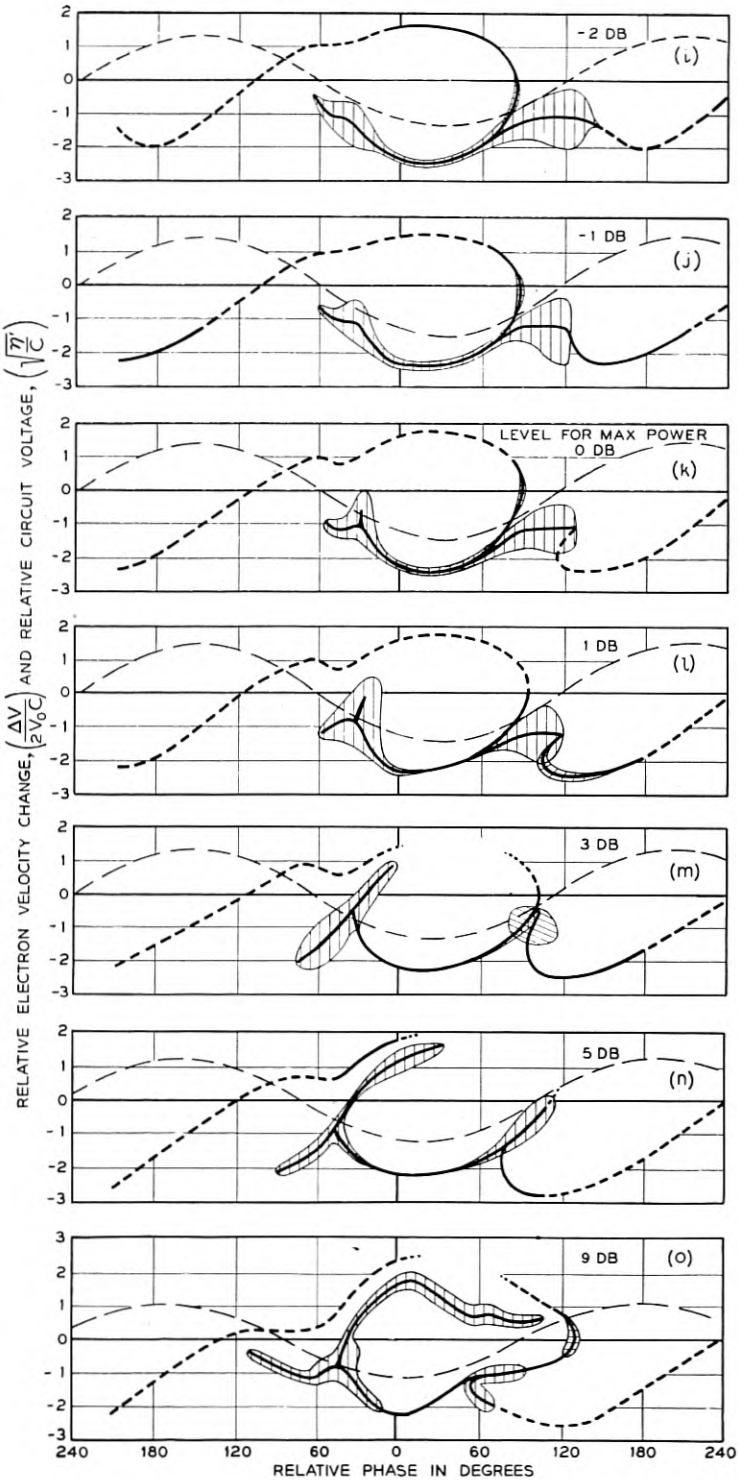


Fig. 7 — Curves of current and velocity as a function of phase for various input levels. The velocity becomes multivalued at a very low level, a tail forming a nucleus for a second electron bunch which eventually caused saturation in the output. For this run $C = 0.1\ Q, C = 0.06, \gamma r_0 = 0.4$ and $b = 0.26$.


-2 DB
(i)
-1 DB
(j)
LEVEL FOR MAX POWER
0 DB
(k)
1 DB
(l)
3 DB
(m)
5 DB
(n)
9 DB
(o)
RELATIVE ELECTRON VELOCITY CHANGE, $\left(\frac{\Delta V}{2V_0C}\right)$ AND RELATIVE CIRCUIT VOLTAGE, $\left(\sqrt{\frac{M}{C}}\right)$
RELATIVE PHASE IN DEGREES
240 180 120 60 0 60 120 180 240

may be seen to give a resultant increasing wave traveling somewhat slower than the initial wave, and thus slower than the electron velocity. Returning to Fig. 7 we see that electrons in the decelerating field [from +30 to +210° in Fig. 7(a)] have been slowed down, and because of their initial velocity being faster than the wave velocity, have moved forward in the wave giving a region of minimum velocity somewhat in advance of the point of maximum retarding field (greatest negative slope in the wave potential). Also, bunching due to acceleration and deceleration of electrons has produced a maximum of electron current density which, because of the initial excess electron velocity, is somewhat to the right of the potential maximum (downward).

As the level is increased the modulation increases and at 17 db below saturation drive, Fig. 7(b), some nonlinearity is evident. The velocity and current are no longer sinusoidal, but show the beginnings of a cusp in the velocity curve and a definite non-sinusoidal bunching of the electrons in the retarding field region (between +30 and 210°).

In the next pattern, Fig. 7(c), at 14 db below saturation a definite cusp has formed with a very sharp concentration of electrons extending significantly below the velocities of the other electrons. We already have a wide range of velocities in the vicinity of the cusp, and at this level the single valued velocity picture of the traveling wave tube breaks down. Although it cannot be distinctly resolved, the study of many such pictures leaves little doubt that the cusp and its later development is really a folding of the velocity line.

The next pattern at 12 db below saturation drive, Fig. 7(d), shows a greater development of the spur and a somewhat greater consolidation of current in the main bunch between +60° and +180°. It is interesting that the velocity in this region has not changed significantly. In order for this to be true the space charge field must just compensate for the circuit field. In the vicinity of the 60° point the space charge field obviously must reverse, accounting for the very sharp deceleration evident in the very rapid development of the low velocity spur. The decelerating field must be far from that of the wave, inasmuch as the electrons just behind the cusp are much more sharply decelerated than those preceding the cusp. We conclude that there are very sharply defined space charge fields much stronger than the helix field. At this relatively low drive, the velocity spread has already achieved its maximum peak value.

The succeeding three patterns show a continuing growth of the spur, a continued bleeding of electrons from the higher velocity regions, and a consolidation of the main bunch just in advance of the spur. Presumably the increased concentration of space charge in the bunch has kept

pace with the increasing helix field, so that the net decelerating field still balances to nearly zero. At 4 db below the saturation drive, Fig. 7(h), the spur has moved well into the accelerating region, and has been speeded up. The main bunch of electrons is still to the right of the spur, and has been consolidated into a 60° interval. The few electrons in advance of this region evidently no longer find the space charge field sufficient to balance the circuit field, and are being decelerated into a second low velocity loop.

The next three patterns show a continued growth of this second low velocity loop, further consolidation of the 'main bunch', and the rapid formation of a second bunch in the accelerating field at the end of the spur. It is interesting that at saturation drive, Fig. 7(k) the two bunches are very nearly equal, and in equal and opposite circuit fields, nearly 180° apart. The reason for the saturation is that while the main bunch is still giving up energy to the wave, the new one is absorbing energy at an equal rate. The fundamental component of electron current is evidently small, and is in quadrature with the circuit field. The current density in the dashed regions is less than 1 per cent of that in the bunches, and probably more than 95 per cent of the electrons are in the two bunches. Two new effects are observable at this level. The second electron bunch has begun to come apart, presumably because of strong localized space charge forces. These forces are also evident in the kink in the velocity pattern drawn by the fast electrons at the same phase as the second bunch.

Since the majority of the current is in the two bunches at a reduced velocity of

$$\frac{\Delta V}{2V_0C} = -1.1$$

one would expect an output efficiency of

$$\frac{\Delta V}{V_0} = 2.2C$$

The actual measured efficiency

$$\frac{\text{RF power output}}{\text{DC power input}}$$

was 2.0 C. Under the conditions described, (6) would give 1.4 C.

At still higher drive levels the pattern continues to develop, electrons from the first bunch falling back into the second, which in turn continues to divide, one part accelerating ahead into a new spur, and the other

slowing down and falling further back in phase. At 9 db above saturation, Fig. 7(o), the pattern is quite complex, and at still higher levels it is utterly indescribable.

It is interesting that the velocity gives a line pattern, even though a multivalued one. It is reasonable to suppose that the development of the spur is really a folding of the velocity line so that the spur is really a double line. Thus, at the 9 db level, and at 0° phase, for instance, there must be electrons originating from five different parts of the initial distribution. In an attempt to verify this the resolution of the velocity analyzer was adjusted so that a difference in velocity of 2 per cent of the overall spread could be observed, but there was no positive indication of more than one velocity associated with any line shown.

There has been a long-standing debate as to whether or not electrons are trapped in the circuit field, or continue to override the wave at large amplitudes. The observations indicate that with low values of space charge and near synchronous voltage the electrons are effectively trapped in the wave until well above saturation amplitude. In other circumstances this is not the case, as we shall see.

SPACE CHARGE EFFECTS

The data of Fig. 7 were taken with a very small value of the space charge parameter QC, so small in fact as to be almost negligible as far as low level operation is concerned. Yet the space charge forces evidently played a very strong role in the development of the velocity and current patterns. It is doubtful that space charge would ever be negligible in this respect, because if the space charge parameter were smaller, the bunching would be more complete, the electron density in the bunch would be greater limited only by the balance of space charge field and circuit field in the bunch. The effect of decreasing QC further therefore is a greater localization of the space charge forces, rather than a reduction of their magnitude, at least until the bunch becomes short compared to the beam radius.

Increasing the value of the space charge parameter has quite the opposite effect. In Fig. 8 are shown three velocity-current distributions at the saturation level, for different values of QC. It can be seen that a result of increased space charge is a greater spread of velocities, and a wider phase distribution of current.

With the introduction of space charge, the velocity difference between the electrons and the circuit wave at low levels is increased. Consequently electrons spend a longer time in the decelerating field before being thrown back in the low velocity spur, and thus lose more energy. The

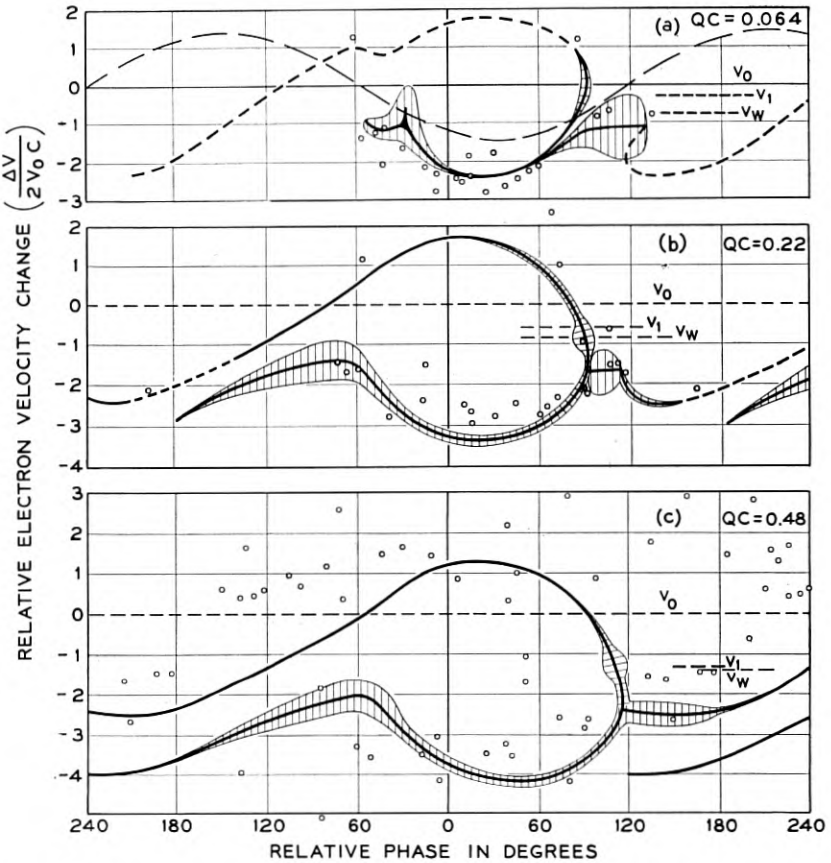


Fig. 8 — A comparison showing the effect of the space charge parameter QC on the velocity and current at overload. The points represent the disc electrons of the computations[7] of Tien, Walker and Wolontis. For this run $\gamma r_0 = 0.4$ and b is chosen for maximum x_1 .

greater reduction of velocity results in a faster and farther retarding of the current in the spur before the retarded electrons recover velocity in the accelerating region. Also the larger space charge forces prevent as tight bunching of the electrons anywhere, so that at overload they are spread over a much wider phase interval (about 360° for $QC = 0.5$). Space charge also prevents electrons from the forward part of the bunch from being trapped so that more electrons escape ahead of the decelerating field and more current is found in the upper half of the velocity curve. This very likely is the reason that efficiency decreases when QC is increased above about 0.3.

EFFECT OF BEAM SIZE

In small signal operation, decreasing the beam radius below that which assures a constant circuit field throughout the beam has no effect except that accounted for by its effect on QC. Fig. 9 shows that for large signals, however, it has a pronounced effect. When the beam is made smaller (with QC maintained by changing frequency and beam current), the slowed up tail is formed at a much lower signal level (not shown), by a very few electrons which begin to collect in the accelerating region before the beam is strongly modulated. As the level is increased, the current is redistributed, more going into the tail without much alteration in the shape of the velocity pattern, and with no strong bunching at any part of the curve. This result is exaggerated in Fig. 9(c) by measuring with a

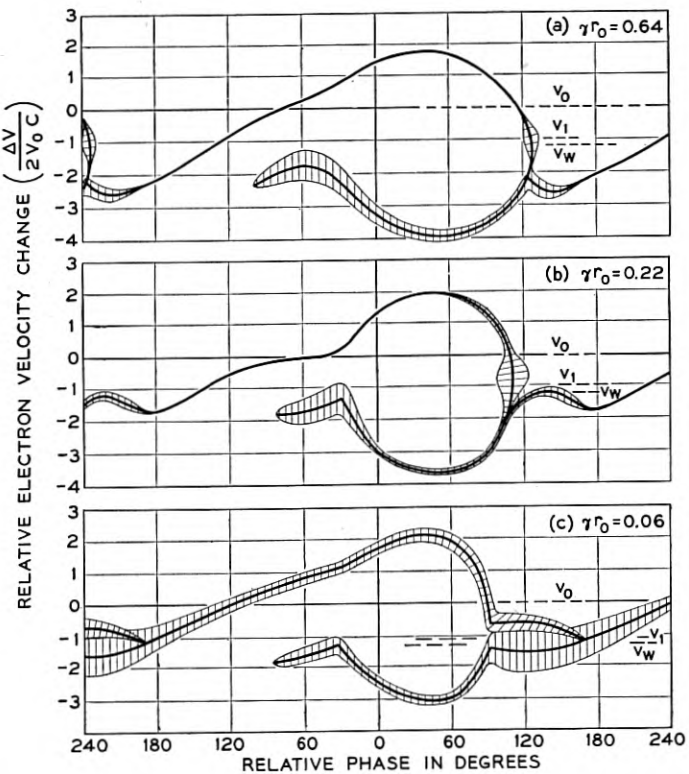


Fig. 9 — Curves of current and velocity as influenced by γr_0. Space charge becomes a very potent factor near overload, especially when the beam is small. For this run $QC = 0.34$ and $b = 1.0$.

ridiculously small beam. By comparison with curves taken for larger beams, the tail is diminutive, electrons are much more uniformly distributed over all velocities and phases, and a peculiar splitting of velocities in the main bunch is found. The latter indicates that electrons entering from the higher velocity region move forward in the bunch, and the rest gradually retard. The smaller reduction in velocities, and the spread of electrons into the higher velocity regions is consistent with the lower efficiency measured (Fig. 2).

To explain the observed difference in high level performance of tubes with different size beams we must consider the character of the ac longitudinal space charge field. The coulomb field from an elemental length of an electron beam is inversely proportional to the square of the distance from the element

$$E = \text{Const}\,\frac{q\Delta z}{(z - z_1)^2} \tag{7}$$

provided $(z - z_1) \gg r_0$ and $(z - z_1) \ll a$.

For $(z - z_1)$ not small compared to a, (i.e., circuit radius not awfully large) the field would drop even faster with $(z - z_1)$ due to the shielding effect of the circuit. On the other hand, very near to the beam element $(z - z_1 \ll r_0)$, the field is approximately that of a disc, which is nearly independent of z, i.e.,

$$E = \text{Const}\cdot\frac{q\Delta z}{\pi r_0} \tag{8}$$

independent of z for $z \ll r_0$.

Thus to a fair approximation the space charge field may be considered to be uniform for an axial distance of the order of a half a beam radius, and to drop rapidly at greater distances. For a given current element, a small diameter beam has an intense field extending only a short distance, while an equal charge element in a larger beam has a weaker longitudinal field extending to a greater distance.

At low amplitudes the extent of the forces makes no difference in operation, for a sinusoidal current gives a sinusoidal space charge field in either case. However, at large amplitudes, a sharp change in current density has a very high short range space charge field if the beam is small, or a much smaller smoothed out long range field if the beam is large. For $\gamma r_0 = 0.5$ which appears to be an optimum compromise between the effects of space charge and field non-uniformity, the space charge field could scarcely be confined closer than about $\pm 30°$ in phase. On the other hand, a sharp bunching of electrons in a beam having

$\gamma r_0 = .05$ would have 100 times the space charge field, extending however only one tenth as far from the current discontinuity.

Returning to Fig. 9 we can see how these considerations enter into the development of the beam modulation. In the case of the small beam, Fig. 9(c), at the very beginning of the formation of a cusp, the strong highly concentrated space charge force causes a rapid deceleration of nearby electrons, resulting in the relatively early formation of a diminutive tail. The very high localized space charge force also prevents as tight bunching of electrons, forcing some to move forward and continuously repopulate the accelerating part of the wave. The relatively early falling apart of the initial bunch and the greater acceleration of the overriding electrons evidently give the latter enough velocity to penetrate the main bunch of electrons and form the second class of electrons in the main bunch, 90°–150° in Fig. 9(c). Thus the net result of reducing the beam size is a severe aggravation of space charge debunching effects, with a consequent reduction in efficiency. To get high efficiency, we conclude, the beam should not be small. It should not be larger than $\gamma r_0 = 0.7$ however, for then the circuit field is not uniform enough over the beam cross-section to excite it properly, resulting in a loss in efficiency as is evident in Figs. 2 and 3.

EFFECT OF INCREASED BEAM VOLTAGE

It is common practice in the operation of traveling wave tubes to elevate the beam voltage, taking a sacrifice in gain in order to obtain increased power output. The effects on the beam modulation are shown in Fig. 10. In Fig. 10(a), the voltage is somewhat below that giving maximum gain. The curve is characteristic of what we have already seen but the bunching is less pronounced and the velocities are less reduced. In Fig. 10(b) the voltage is somewhat above that giving maximum gain and the curve is much like that of Fig. 8 except that the decelerated electrons are slowed by a greater amount, consistent with the increased separation of electron and wave velocity, and also with the measured increase in power output.

Increasing the beam voltage still further gives only a slight increase in efficiency. Fig. 10(c) shows that even though electrons are slowed to still lower velocities, and the velocity spread is increased, many more electrons override the circuit wave and are accelerated, thereby offsetting the greater contribution of the slower electrons. This is much like what was seen with increasing space charge (QC) and indeed the effects are almost equivalent. As one would expect therefore, little is gained by elevating the beam voltage if the space charge is large, the

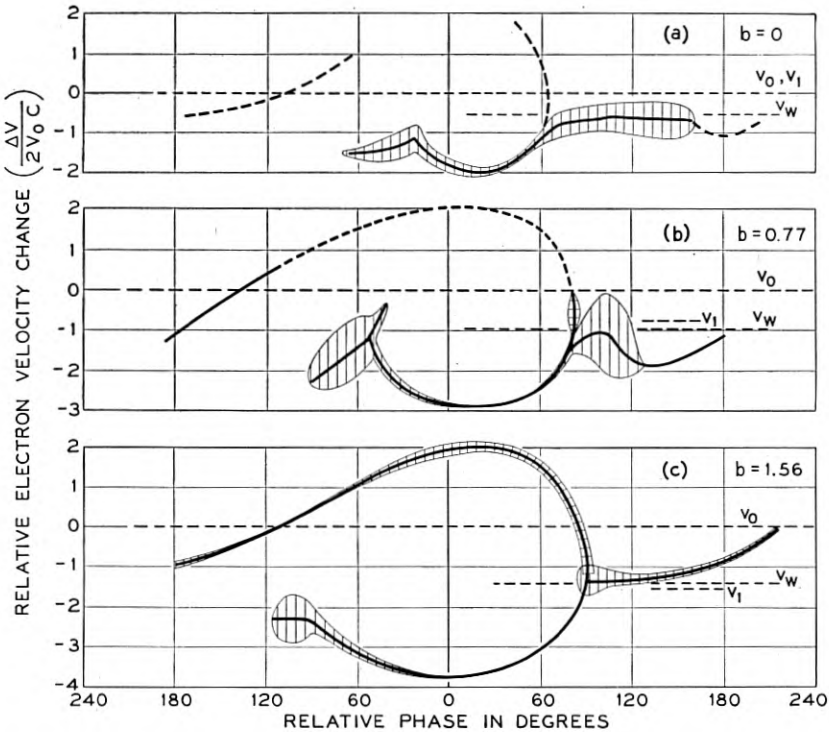


Fig. 10 — The influence of beam velocity on ac velocity and current. When the velocity is raised too high, the electrons are not effectively trapped by the wave, and override into the accelerating field. With large QC and/or small γr_0 the electrons override in any case, and little is gained by increasing b. For this case $QC = 0.13$ and $\gamma r_0 = 0.21$.

main effect being to push more electrons forward into the accelerating region.

ELECTRIC FIELD IN THE BEAM

Besides telling a clear story of the non-linear dynamics of the traveling wave tube, the foregoing curves contain a lot of information about average current and velocity distributions. From the current or velocity curves we can in turn deduce the distribution of longitudinal electric field in the beam. Figs. 11(a) and (b) show the instantaneous current as a function of phase, taken from the curves of Figs. 8(a) and (b). The infinite differential in the velocity curve necessarily gives a pole in the charge density (at about 88°). The total charge in the vicinity of the

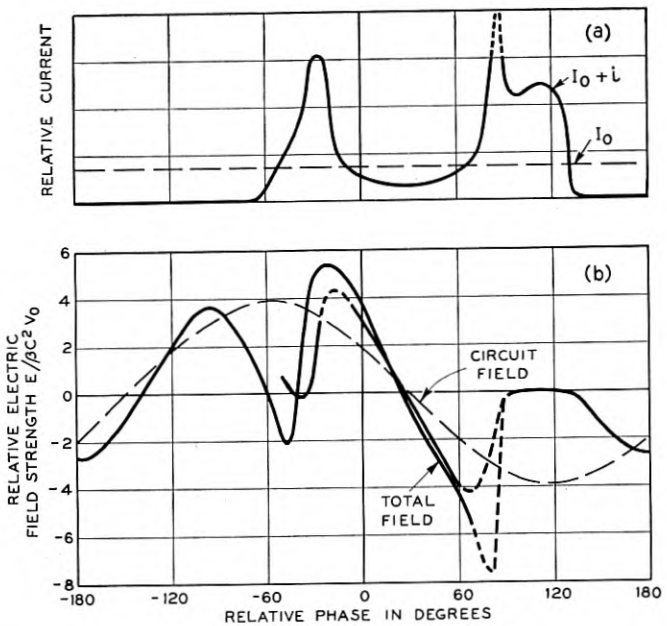


FIG. 11 — AC current and electric field in the beam. The upper curve comes directly from Fig. 8(a). The lower curve is deduced by an approximate method from the velocity curve of Fig. 8(a). The double value below 90° is partly due to inconcistency between the two parts of the velocity curve, and partly due to the nature of the approximation.

pole, and the range of the space charge force (dependent upon QC and γr_0) determines its effect upon the electron dynamics.

Most of the current is incorporated in the two bunches nearly 180° apart, as we have seen, each bunch having a current density many times the average.

We might obtain the space charge fields from the current density, but this would require a rather definite knowledge of the characteristic space charge field versus distance as influenced by beam diameter. It would also be pushing the accuracy of charge density measurement, which is crude at best. A better way is to compute the electron acceleration from the velocity curves. This may be done by taking two velocity patterns at slightly different signal levels, and tracing electrons from one to the next, using the measured velocity to determine the relative phase shift of any electron.

In the appendix it is shown that a close approximation to this is

$$E_\Phi = 2\beta C^2 V_0 \left[\frac{(V_0 - V_w) + \Delta V}{2V_0 C}\right] \frac{d}{d\Phi}\left(\frac{\Delta V}{2V_0 C}\right) \qquad (10)$$

where the parameters are all obtained from a single velocity curve, and

E_Φ is field strength in volts meter at phase Φ

$\frac{\Delta V}{2V_0C}$ is the value of the ordinate of the velocity characteristic of of interest (Figures 7 to 10) and

$\left(\frac{V_0 - V_w}{2V_0C}\right)$ is the value of the ordinate corresponding to the wave velocity. (To be precise, the wave velocity at the associated output level, but to a reasonable approximation, that of the wave velocity at low levels. (This value is indicated by V_w in the velocity curves.)

The total electric field has been computed for the case of Figs. 8(a) and (b) and is given in Figs. 11(b) and 12(b) together with the circuit field calculated for the associated power level and plotted with an arbitrarily chosen phase. In each case it is seen that the space charge field is comparable in magnitude to the circuit field, is far from sinusoidal, and

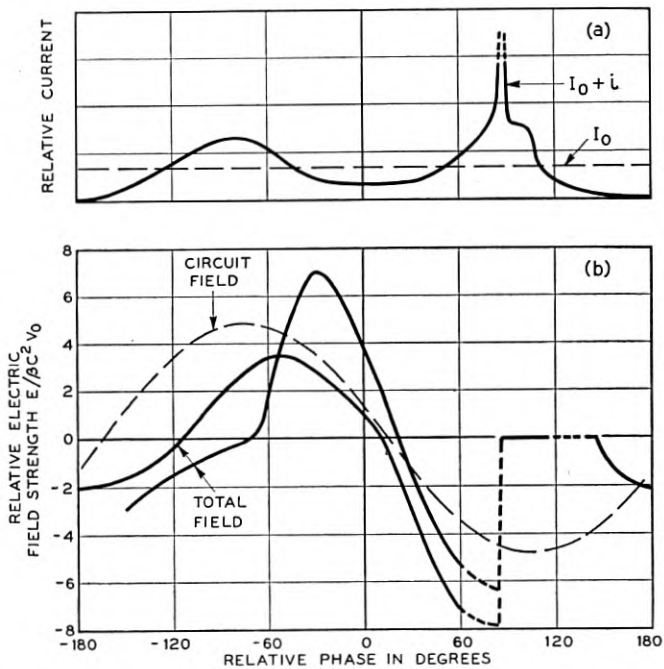


Fig. 12 — AC current and electric field in the beam deduced from Fig. 8(b). The greater space charge results in a less defined bunch, and smoother space charge field than in Fig. 11.

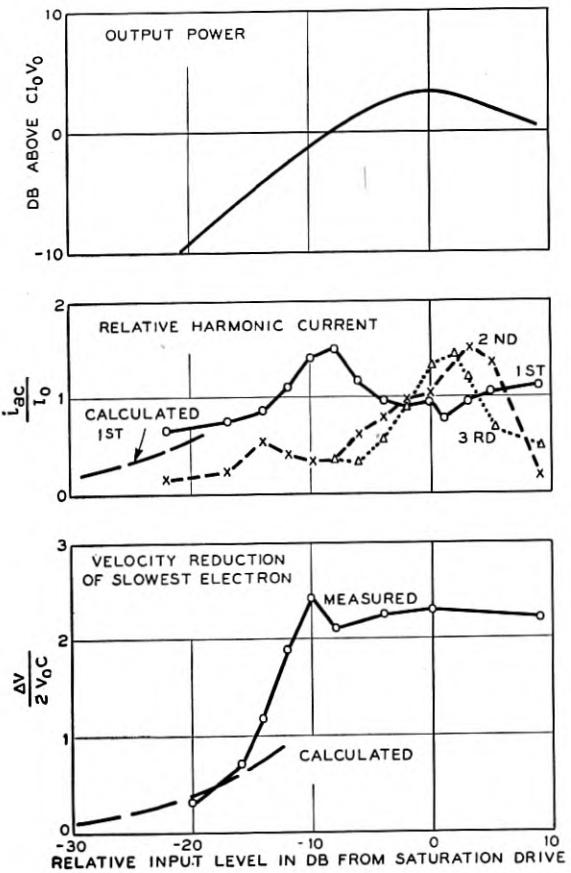


Fig. 13 — Curves of output level, fourier component amplitudes of beam current, and peak velocity as a function of input level for low space charge. These curves were deduced from Fig. 8(a).

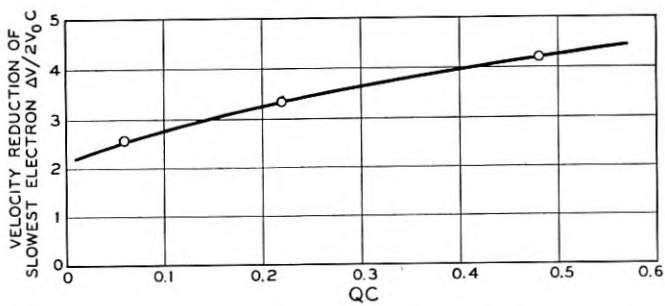


Fig. 14 — Maximum velocity reduction as a function of space charge (from Fig. 8). The velocity reduction is about 3.5 y_1 .

agrees qualitatively with what would be expected from the associated curve of beam current.

To determine the curves of Figs. 11 and 12 is rather stretching the accuracy of the measurements as can be seen by the large discrepancy in the field calculated from the two parts of the velocity curve which of course should be identical. The figures do give an interesting qualitative picture of traveling wave tube behavior however, and are included here for that reason.

OVERALL VELOCITY SPREAD

Of more practical importance is the overall velocity spread in the spent beam. It is often desirable to reduce the power dissipation in a traveling wave tube by operating the collector at a potential below that of the electron beam, and it is interesting to see how far one might go. Fig. 13 shows how the velocity reduction of the slowest electron, together with the output level and fourier current components of beam current vary with input level. For small amplitudes, the low level theory accurately predicts the velocity, but near overload, as we have seen, the minimum velocity drops sharply to a value several times lower than that projected from small signal theory.

The maximum velocity spread dependence upon the space charge parameter QC is shown in Fig. 14. Similar data for values of the other parameters may be obtained from the velocity diagrams.

From the foregoing data, one can deduce the amount of reduction of collector potential that should be theoretically possible without turning back any electrons. An idealized unipotential anode could collect all the current at a potential ΔV (in the foregoing figures) above the cathode, decreasing the dissipated power by a factor of $\Delta V/V_0$ below the dc beam power.

STOPPING POTENTIAL MEASUREMENTS

Information on spent beam velocity has also been obtained by a stopping-potential measurement at the collector of a more conventional 4,000-mc traveling wave tube.* Two fine mesh grids were closely spaced to a flat collecting plate, and collector current was measured as a function of the potential of second grid. The first grid was very dense, to prevent reflected electrons from returning into the helix. One curve taken with this arrangement is shown in Fig. 15 and for comparison we have

* Similar measurements have been reported by Atsumi Kondo, Improvement of the Efficiency of the Traveling Wave Tube, at the I.R.E. Annual Conference on Electron Tube Research, Stanford University, June 18, 1953.

plotted the distribution predicted from Fig. 9(b). The *RF* losses in the 4,000-mc tube were not negligible, and probably account for slightly smaller power output and greater proportion of higher velocity electrons.

COMPARISON WITH COMPUTED CURVES

Non-linear calculations of traveling wave tube behavior have been made by Tien, Walker and Wolontis[7] and by Tien[9] covering the same region of parameter values as is reported here. In Figs. 2, 3, 4 and 9 are shown some of their data on our coordinates. The similarity of the results over much of the range is rather reassuring. It is interesting that in order to make the computations it was necessary to assume two space charge factors, just as was found experimentally. There are, however, some significant differences:

1. In general, the computed values give a higher value of efficiency than is measured, by about 25 per cent. Thus, the computations indicate

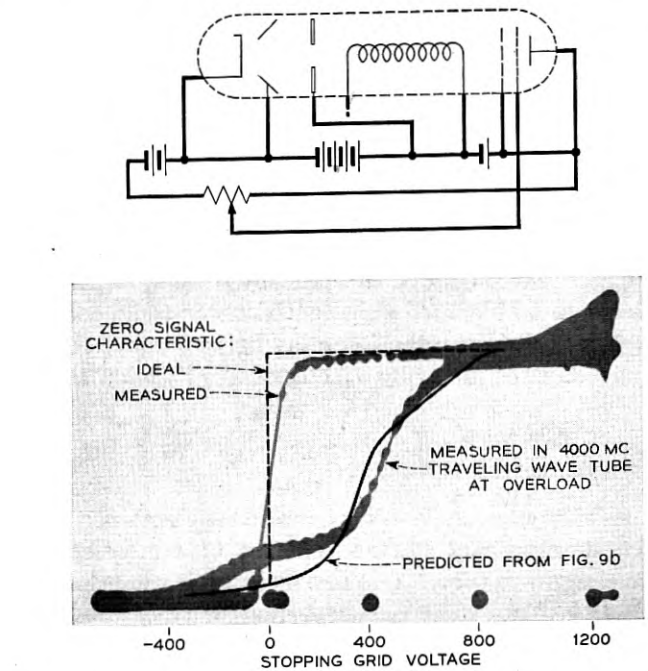


Fig. 15 — Collected current versus stopping potential. The oscilloscope curve is for a 4,000-mc tube, and the other that predicted from the scale model measurements. By integrating current as a function of velocity for Figs. 7–10 stopping potential distributions can be deduced for other conditions.

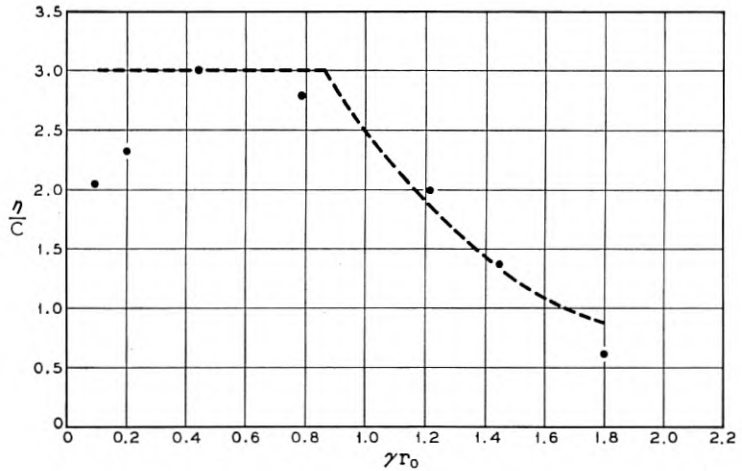


Fig. 16 — Efficiency versus γr_0 for small QC. The dashed curve is proportional to the amount of beam current in the circuit field strength having at least 85 per cent of the intensity at the edge of the beam. This illustrates the fact that for large beams only the edge of the beam is effective.

that with the reasonable values of $QC = .25$ and $\gamma r_0 = 0.8$ ($k_T = 2.5$), the efficiency would be about $3.8C$, whereas the measured value is $3.1C$.

2. The largest discrepancy in the measured and computed value of η/C is for large values of γr_0 (small k_T), where the computations show a steady increase in efficiency instead of a sharp decrease. This arises because the computational model assumed the electric field to be uniform across the beam, whereas in the actual tube it varies as $I_0(\gamma r)$, and for large values of γr_0 the field is weak near the beam axis. This effect is shown in Fig. 16 where η/C is plotted versus γr_0 for small values of QC, on the same scale with a curve proportional to the square of the fraction of the beam within a cylindrical shell such that

$$1 - \frac{I_0(\gamma r_1)}{I_0(\gamma r_0)} = 0.85 \tag{11}$$

where r_1 is the inside radius, and r_0 the outside beam radius (i.e., the fraction of the beam in a field greater than 85 per cent of that at the beam edge).

No serious studies of velocity were made for large beams, but on cursory examination it was evident that the beam modulation varied considerably over the cross section when the beam was very large, and scarcely at all when it was smaller than around $\gamma r_0 < 0.8$.

3. The observed effect of small beam radius upon efficiency is not as pronounced as was found in the computations. The reason is not known but may be due to modulation of the beam diameter at large signal levels. This effect would be negligible with the larger γr_0's, due to the focusing fields being relatively much larger.

4. The computations, and also those of Nordsieck,[5] Poulter[6] and Rowe[8] indicate a much higher efficiency than has been observed at elevated beam voltages and small C and QC. The reason for this may be that the limited number of "electrons" used in the computational models fail to adequately account for the very sharp space charge cusp that forms under low QC conditions, or that interpolation between their points should not be linear, as assumed in making the comparison. On the other hand it would be difficult to be sure that nonuniformities in electron emission were not influencing the measurements in the case of the large beams by giving a larger QC than calculated.

5. The increase in efficiency to be had by elevation of beam voltage is much smaller than is indicated by the computations. This may be a real difference, or it may be that at elevated voltages, the measurements are beginning to feel the influence of overloading in the attenuator. The margin of safety on attenuator overloading is not as great as one would like at the higher frequencies.

6. The velocity curves, Fig. 8, compare the computed and measured data on three runs. For small QC, Fig. 8(a), the agreement is remarkably good considering the fact that in the computation only 24 "electrons" were used to describe a rather complicated function. The effect of the lumping of space charge in the artificial 'disc' electrons causes a scatter-of points which is different from that in an actual tube as is especially apparent in Figure 8c. In spite of this the computational results indicate a velocity spread and current distribution not greatly different from that observed.

CONCLUSIONS

The large scale model traveling wave tube is a means for the determination of non-linear behavior, and has been valuable in determining relationships and limitations important to efficient operation of such tubes. It has shown that there is a broad optimum in tube parameters around $C = 0.14$ $QC = 0.2$ and $\gamma r_0 = 0.5$ for which values it is possible to obtain efficiencies well above 30 per cent. The measured ac beam velocity and current near overload show that it is unlikely that significant increase in efficiency can be obtained by any simple expedients such

as operations on the helix pitch alone, or the use of an auxiliary output circuit.

The results being in normalized form, are believed to be generally applicable to conventional traveling wave tube design. With determination of an equivalence in beams, they should even be a useful guide in the design of tubes using hollow beams or other configurations.

The work described could not have been done without the able assistance of G. J. Stiles and L. J. Heilos and the helpful council of many of my colleagues at Bell Telephone Laboratories.

Appendix

Scale model tube design

There were a larger number of factors to be accounted for in the design of this tube. Its proportions should be such as to make it representative of the usual design of traveling wave tube. Its size should be such as to make it easy to define the electron beam boundary, and to dissect the beam. The size should also be such that the electron beam velocity analysis could be done before the beam character would be changed either by space charge, or its velocity spread. The voltage should be low so that further acceleration in the velocity analyzer would not lead to an inconveniently high voltage. Finally, the availability of suitable measuring gear over a 3–1 frequency range, and the size of the laboratory must be considered. All of these factors led to low frequency operation, limited principally by the laboratory size and the mechanics of construction.

A moderate perveance of around 0.2×10^{-6} was taken, with a γa of 1.2 and γr_0 of 0.8 in a representative helix with small impedance reduction due to dielectric and space harmonic loading. This is representative of practical tube design in the microwave range and is centered on the parameter values of most general interest. At a frequency of 100 mc and a beam potential of 400 volts this resulted in a helix 10 feet long and $1\frac{1}{2}$ inches in diameter, with an electron beam 1 inch in diameter. The choice of frequency was finally determined by the availability of measuring equipment, and the voltage was selected to give a convenient size for dissection of the electron beam.

By changing frequency, beam current, and beam diameter it was possible to cover a reasonable range of γr_0, and QC, and to make some observations into the region of large C operation.

In all of the measurements described, a very strong uniform magnetic field was used to confine the beam, and therefore scaling of the magnetic

focusing field need not be considered. The electron beam was produced in a gridded gun and is thus near to the ideal confined flow, which is the only focusing arrangement which is known to determine a reasonably uniform boundary to the beam. The beam size and straightness was checked using a fluorescent screen at the collector end.

NORMALIZING FACTORS

The measurements described are expressed relative to the linear theory, in Pierce's[2] notation, which are generally used in the design of traveling wave tubes. Thus, instead of being presented in the terms of measurement or simply normalized to efficiency, perveance, impedance, etc., they are expressed in terms of C, QC, γr_0, etc., with normalized fields, currents and velocities. In this way the results become adjuncts to the linear theory and are more easily applied to tube design. Electron velocity is plotted on the same scale as the relative velocity parameters b and y_1 used in low level theory, (i.e., normalized to $\Delta V/2V_0C$). Efficiency is normalized as η/C, which for C less than 0.1 is relatively independent of C. Field strength in the linear region is proportional to

$$\sqrt{\frac{\eta'}{C}}$$

(η' being efficiency measured at the appropriate signal level). Solving the equation for C^3,

$$C^3 = \frac{E^2}{2\beta^2 P}\frac{I_0}{2V_0} \tag{12}$$

gives us

$$\sqrt{\frac{\eta'}{C}} \approx \frac{E}{\beta C^2 V_0} \tag{13}$$

which we use as the normalizing parameter for electric field. Circuit potential is the integral of circuit field over a quarter period, giving a normalized parameter V/V_0C^2. For convenience in the use of common coordinates, circuit potential was plotted as $-V/2V_0C^2$ in Figure 7.

The other curves are plotted as values relative to dc quantities or to saturation level.

Strictly speaking, the results hold only for tubes having the same proportions as the model. Practically, however, as long as the helix impedance and radius (ka or γa) are not different by orders of magnitude from the values used, and as long as the perveance is low (below 2×10^{-6} for

instance), the results are believed to be significant for tubes having the indicated values of γr_0 and QC.

HELIX IMPEDANCE

It is important to the measurements to have an accurate evaluation of the helix impedance. Several methods of measurement have been discussed in the literature.[21, 22] That described by R. Kompfner was selected, wherein the circuit impedance is correlated with the beam current and voltage which gives a null in the output signal. When the beam voltage and current are adjusted to give zero transmission for a lossless section of helix (neglecting space charge) $CN = 0.314$ and $\delta V/V_0 \cong 1/N$. Using the measured length of the helix, and measuring the voltage and current giving the null in signal transmission, we can compute C, and thus the impedance and velocity (synchronous voltage) of the helix.

The impedance was calculated by P. K. Tien,[23] and the results are compared in Fig. 17. The measured impedance at the high frequency end was much too low until space charge in the beam was accounted for in interpreting the measurements. Fortunately, in the absence of attenu-

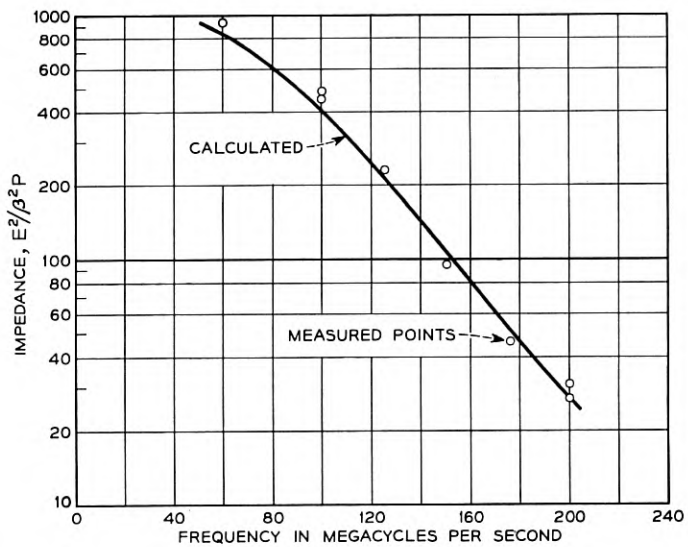


Fig. 17 — Helix impedance as a function of frequency. The impedance was calculated taking into account dielectric loading and wire size. It was measured using the Kompfner dip method, taking account of space charge.

ation, the conditions for start of oscillation in a backward wave oscillator are the same as for the output null in a traveling wave tube. Space charge was first accounted for using the results of H. Heffner[24, 25] giving an excellent check between predicted and computed helix impedance. Later C. F. Quate[26] showed that the same measurement could be used to determine the space charge parameter QC as well as the helix impedance. Since thermal velocity effects and the uncertainty of some of the assumptions used in evaluating the small signal effects of space charge cast some doubt on the proper evaluation of this term, further measurements were made on this factor, and a satisfactory correlation between the observed value of QC and that computed from the Fletcher[27] curves was obtained.

TOTAL ACCELERATING FIELDS

From the velocity characteristics shown in Figs. 7 through 10, we can deduce the electron accelerations, and thus the electric fields at any point. While the curves are actually diagrams of velocity as a function of phase, they closely correspond to the velocity-time or distance distribution of the electrons in the traveling wave tube. Knowing these characteristics we can deduce the motion of any element of charge, and thus the force under which it moves. It is observed that over most of the curve the shape of the velocity pattern does not change nearly so rapidly as the redistribution of electrons within the pattern. Thus, we can approximate the situation at any amplitude by assuming the velocity pattern to be constant, and that electrons move within the pattern according to simple particle dynamics. This is a good approximation except where the acceleration is high (i.e., vertical crossings of the wave velocity line).

Consider then an element of the velocity pattern at phase Φ_1 and velocity $(u_0 + \Delta u)$. In an interval dt this element will move a distance

$$(u_0 + \Delta u)\, dt \tag{14}$$

and will change velocity by

$$du = E \frac{e}{m} dt \tag{15}$$

At the same time the wave will have moved a distance $v\, dt$, resulting in a relative change in phase between wave and current element of

$$d\Phi = \beta(u_0 - v + \Delta u)\, dt \tag{16}$$

In terms of equivalent differences the term in brackets can be written

$$(u_0 - v + \Delta u) = \sqrt{2 \frac{e}{m} V_0}\, C \left(\frac{V_0 - V_w + \Delta V}{2 V_0 C} \right) \tag{17}$$

from (16) and (17) we can write:

$$\frac{du}{dt} = \frac{du}{d\Phi}\frac{d\Phi}{dt}$$
$$= \frac{d}{d\phi}\left(\frac{\Delta V}{2V_0C}\sqrt{2\frac{e}{m}V_0}\,C\right)\left[\beta\sqrt{\frac{e}{m}V_0}\,C\left(\frac{V_0 - V_w + \Delta V}{2V_0C}\right)\right] \tag{18}$$

giving (from 15)

$$\frac{E}{\beta V_0 C^2} = 2\left[\left(\frac{V_0 - V_w}{2V_0C}\right) + \left(\frac{\Delta V}{2V_0C}\right)\right]\frac{d}{d\Phi}\left(\frac{\Delta V}{2V_0C}\right) \tag{19}$$

β, V_0 and C are constants of the tube, the first inner parenthesis may be calculated from the tube constants and is shown in the curves. $\Delta V/V_0C$ and its differential are the value and the slope of the velocity curve in question.

The important approximations here are that the velocity-phase curves are representative of velocity-distance characteristics, which is true for small values of C, and that the electrons move roughly tangent to the given velocity pattern. By comparing several patterns at different signal levels it is observed that this is true to a fair accuracy over most of the curve. Also it is assumed that the wave velocity at large amplitudes is the same as that for small signals, which is not quite true. The resulting curves give at least a qualitative picture of the field distribution within a traveling wave tube, and serve to emphasize the importance of space charge fields in determining the non-linear characteristics.

ELECTRIC FIELD OF THE HELIX WAVE

In order to see what part of the field is due to space charge we must evaluate the corresponding helix fields. A value for this can be derived from the basic traveling wave tube equations assuming the helix fields to be sinusoidal and not seriously affected in impedance by the beam (small C again). By definition

$$\frac{E^2}{2\beta^2 P}\frac{I_0}{4V_0} = C^3 \tag{18}$$

and

$$\frac{\eta'}{C} = \frac{P}{I_0V_0C} \tag{19}$$

where η' is normalized power level, i.e., efficiency corresponding to the signal level E of interest. From this we deduce for the normalized circuit

field

$$\frac{E}{\beta V_0 C^2} = 2\sqrt{2}\sqrt{\frac{\eta'}{C}} \tag{20}$$

which integrates to give a normalized ac circuit voltage

$$\frac{V}{2\sqrt{2}V_0 C^2} = \sqrt{\frac{\eta'}{C}} \tag{21}$$

RELATIVE PHASE BETWEEN WAVE, VELOCITY AND CURRENT

The velocity analyzer provides no convenient measure of relative phase between the helix wave and the beam modulation. Therefore we compute the relation of helix field and beam modulation for a small signal, and for large amplitudes measure the phase of each relative to that at small amplitudes.

Pierce gives the relationship[2] $$v = \frac{-\eta\Gamma V}{u_0(j\beta_e - \Gamma)} \tag{22}$$

which using (9) and the fact that $$\beta_e C\delta = j\beta_e - \Gamma \tag{23}$$

gives for the small signal beam modulation

$$\frac{\Delta V}{2V_0 C} = -j\frac{\sqrt{2}}{\delta}\sqrt{\frac{\eta}{C}} = \frac{\sqrt{2}}{|\delta|}\sqrt{\frac{\eta'}{C}}\underline{\left|\left(\tan^{-1}\frac{y_1}{x_1}\right) - \frac{\pi}{2}\right.} \tag{24}$$

Similarly we have for the small signal current modulation

$$i = I_0\sqrt{2\frac{\eta}{C}}\cdot\delta^2 = I_0\sqrt{2\frac{\eta}{C}}\,|\,\delta^2\,|\,\underline{\left|2\tan^{-1}\frac{x_1}{y_1}\right.} \tag{25}$$

The value of $\delta(= x_1 + jy_1)$ is given in Fig. 18, drawn from data supplied by P. K. Tien, from Pierce[2] and from Birdsall and Brewer.[28] This figure was also used as a basis for determining the values of y_1 and b used in several of the curves.

MEASUREMENT OF POWER

The output power, and relative output phase was measured using a micro-oscilloscope.[29] The subharmonic of the signal was used for a sweep voltage, and phase was measured from the shape of the observed lissajou figures. The oscilloscope deflection was compared with the dc deflection from a battery standard, and checked on occasion with a bolometer power meter at the operating frequency.

THE VELOCITY ANALYZER

There are many ways in which one may separate velocities in an electron stream. Crossed electric and magnetic fields were used in this ex-

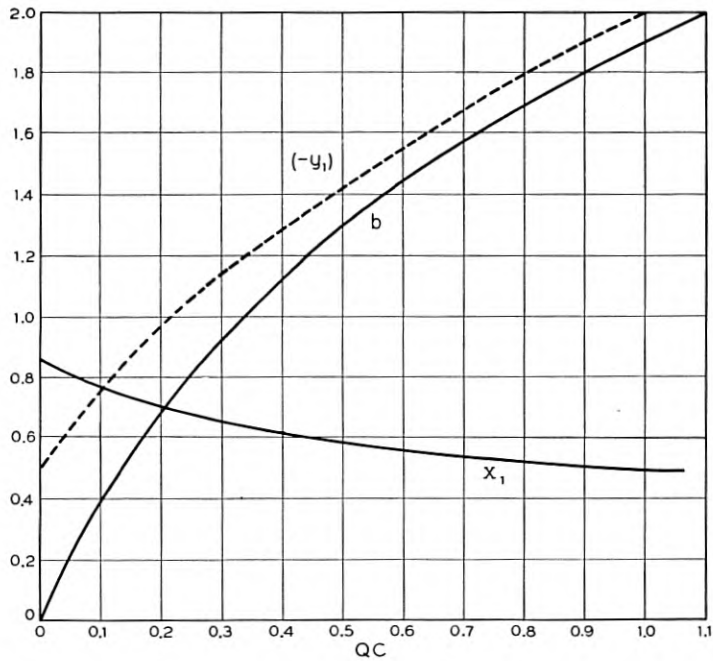


Fig. 18 — Increasing wave propagation factors used in interpreting the measurements. These are the maximum value of x_1 and the corresponding value of b and y_1 for given values of QC.

periment because a simple control of sensitivity was important in order to study velocity differences ranging from 1 per cent up to as much as 100 per cent of the dc beam velocity.

The velocity analyzer is sketched in Fig. 5. It consists of an aperture which transmits only a few microamperes of the electron stream; a magnetic pole piece (not shown) terminating the focusing field; a pair of horizontal deflection plates; an electrostatic lens system; pole pieces and deflection plate to provide a region with crossed electric and magnetic fields; and finally a drift tube, a post deflection acceleration electrode and fluorescent screen. The whole assembly is raised 1,000 volts above the helix potential and the 0.001″ aperture is very close to the end of the helix, so that the electrons are very quickly accelerated to a high voltage. By this means, the region of debunching outside of the helix field is kept below 1.4 radians transit angle and the velocity spread within the analyzer is reduced by a factor of four. Space charge within the analyzer is entirely negligible because of the small current transmitted.

In order to discriminate in phase before the electrons are scrambled

due to their spread in velocity, the horizontal sweeping plates are mounted just as close to the aperture as is deemed practical. The observed velocity spreads in the beam were such as to give less than 0.2 radians error in phase under the worst conditions.

The horizontal deflecting plates were driven synchronously with a sub-harmonic of the RF input to the helix, and the resulting deflection served to separate electrons according to phase in the final display.

Placing the focusing lens after the deflection plates results in a considerable reduction in deflection sensitivity. However, undesirable magnification of the pinhole aperture dictated that the lens could not be close to it, and it was important to initiate the deflection as early as possible. The lens consists of three discs, the center one being biased to about 800 volts above the mean voltage of the rest of the system.

Immediately after the lens there are two iron pole pieces and two insulated electric deflection plates which extend parallel to the beam for 1¼ inches. The pole pieces provide a dc magnetic field up to about 20 gausses induced by small coils outside of the envelope, and the electric deflection plates are biased with up to a corresponding 50 volts dc polarized to oppose the magnetic deflection of the beam. The electric and magnetic fields are adjusted so that the normal unmodulated electron beam traverses the region with no deflection and strikes the center of the fluorescent screen. In the crossed field region

$$\frac{E}{B} = \sqrt{2\frac{e}{m}V_0}. \tag{26}$$

Electrons having greater or lesser velocity are deflected parallel to the electric field, and give a corresponding deflection from the center of the fluorescent screen.

To get a display in which the various elements are not hopelessly entangled, it was necessary to sweep the trace in an initial ellipse at a subharmonic rate. The sweep voltage was applied to the horizontal deflection plates, with just a little applied to the vertical plates through a phase shifter. The relative phase of any part of the trace was measured from the ellipse, and the velocity sensitivity was calibrated by observing the ellipse deflection as a function of the dc beam potential, as shown in Fig. 6(a). There is a small error due to the sensitivity of deflection to velocity, and due to distortion of the ellipse by fringing fields.

In order to measure velocity and current density in the displayed pattern, the fluorescent screen was photographed, and the negative projected in a microcomparator. It was assumed that with the small currents used, the light intensity was proportional to current, and the film linearity was calibrated by making exposures of several different durations. The trace density was measured with a densitometer, sweeping

over the trace width to account for variations in focus for different parts of the pattern. Admittedly, the process is not very accurate, but it does give a rough measure of current density and helps considerably in interpreting the observed velocity patterns.

NOMENCLATURE

a	Circuit radius
b	Parameter relating electron velocity to that of the cold circuit wave $u_0 - v_1/u_0C = \Delta V/2V_0C$
B	Magnetic field
β	the axial phase constant ω/v_1
C	The gain parameter $= (E^2/2\beta^2P)\ (I_0/4V_0)$
γ	Radial phase constant $\cong \beta = \omega/v_1$
δ_1	Complex propagation constant for the increasing wave
E	Electric field
E_Φ	Electric field at phase Φ
e/m	Charge to mass ratio of the electron
I_0	Beam current in amperes
$I_0(\ )$	Modified Bessel function
k_T	Tien's constant k, $= 2/\gamma r_0$
ka	Circuit circumference measured in (air) wavelengths
N	Number of wavelengths
η	Maximum efficiency
η'	Efficiency at an intermediate power level
P	RF power obtainable from the circuit
QC	Space charge parameter
q	Charge per unit length in the electron beam
r	Radial distance from the axis
r_0	Beam radius
t	Time variable
u	Electron velocity
u_0	DC beam velocity
v	AC velocity of the electron beam
v_1	Wave velocity
V_0	DC beam voltage
T_w	Voltage corresponding to the wave velocity
ΔV	Voltage difference corresponding to the difference in velocity of an electron and the dc beam velocity
δV	Difference between synchronous voltage and that giving the Kompfner dip
Φ	Relative phase
z	Distance measured along the beam

REFERENCES

1. Pierce, J. R., Theory of the Beam Type Traveling Wave Tube, Proc. I.R.E., **35,** pp. 111–123, Feb., 1947.
2. Pierce, J. R., Traveling Wave Tubes, D. VanNostrand Co., Chapter XII.
3. Slater, J. C., Microwave Electronics, D. VanNostrand Co., 1950, pp. 298.
4. Brillouin, L., The Traveling Wave Tube (Discussion of Waves of Large Amplitudes), J. Appl. Phys., **20,** p. 1197, Dec., 1949.
5. Nordsieck, A., Theory of the Large Signal Behavior of Traveling Wave Amplifier, Proc. I.R.E., **41,** pp. 630–647, May, 1953.
6. Poulter, H. C., Large Signal Theory of the Traveling Wave Tube, Tech. Report No. 73 Electronics Research Laboratory, Stanford University, Stanford, California, Jan., 1954.
7. Tien, P. K., Walker, L. R., and Wolontis, V. M., A Large Signal Theory of Traveling Wave Amplifiers, Proc. I.R.E., **43,** pp. 260–277, Mar. 1955.
8. Rowe, J. E., A Large Signal Analysis of the Traveling Wave Amplifier, Technical Report No. 19, Electron Tube Laboratory, University of Michigan.
9. Tien, P. K., A Large Signal Theory of Traveling Wave Amplifiers Including the Effects of Space Charge and Finite C, B.S.T.J., **34,** Mar., 1956.
10. Brangaccio, D. J., and Cutler, C. C., Factors Affecting Traveling Wave Tube Power Capacity, Trans. I.R.E. Professional Group of Electron Devices, **PGED 3,** June, 1953.
11. Crumly, C. B., Quarterly Status Progress Report No. 26, Electronics Research Laboratory, Stanford University, Stanford, California, pp. 10–12.
12. Doehler, O., et Kleen, W., Phenomènes non Linéaires dans les Tubes a Propagation D'onde" Annales de Radioélectricité (Paris), **3,** pp. 124–143, 1948.
13. Doehler, O., et Kleen, W., Sur le Rendement du Tube a Propagation D'onde," Annales de Radioélectricité, Tome IV No. 17 Juillet, 1949 pp. 216–221.
14. Berterotiére, R., et Convert, G., Sur Certains Effets de la Charge D'espace dans les Tubes a Propagation D'onde, Annales de Radioélectricité, Tome V, No. 21, Juillet, 1950.
15. Klein, W., und Friz, W., Beitrag zum Verhalten von Wanderfeldrähren bei Hohen Engangspegeln, F.T.Z., pp. 349–357, July, 1954.
16. Warnecke, R. R., L'évolution des Principes des Tubes Électroniques Modernes pour Micro-ondes, Convegno di Elellronica e Televisione, Milano, p. 12–17, Aprile, 1954.
17. Warnecke, R. R., Sur Quelques Résultats Récemment Obtenus dans le Domaine des Tubes Electroniques pour Hyperfréquences, Annales de Radioélectricité, Tome IX, No. 36, Avril, 1954.
18. Warnecke, R., Guenard, P., and Doehler, O., Phénomènes fondamentaux dans les Tubes à onde Progressive, Onde Electrique, France, **34,** No. 325, p. 323–338, 1954.
19. Brück, L., und Lauer, R., Die Telefunken Wanderfeldröhre TL6, Die Telefunken-Röhre Heft 32, pp. 1–21, Februar, 1955.
20. Brück, L., Vergleich der Verschiedenen Formeln für den Wirkungsgrad einer Wanderfeldröhre, Die Telefunken-Röhre Heft 32, pp. 23–37, Februar, 1955.
21. Cutler, C. C., Experimental Determination of Helical Wave Properties, Proc. I.R.E., **36,** pp. 230–233, Feb., 1948.
22. Kompfner, R., On the Operation of the Traveling Wave Tube at Low Level, Journal British I.R.E., **10,** p. 283, Aug.-Sept., 1950.
23. Tien, P. K., Traveling-Wave Tube Helix Impedance, Proc. I.R.E., **41,** pp. 1617–1623, Nov., 1953.
24. Heffner, H., Analysis of the Backward-Wave Traveling-Wave Tube, Proc. I.R.E., **42,** pp. 930–937, June, 1954.
25. Johnson, H. R., Kompfner Dip Conditions, Proc. I.R.E., **43,** p. 874, July, 1955.
26. Quate, C. F., Power Series Solution and Measurement of Effective QC in Traveling-Wave Tubes, Oral presentation at Conference on Electron Tube Research, University of Maine, June, 1954.
27. Fletcher, R. C., Helix Parameters in Traveling Wave Tube Theory, Proc. I.R.E., **38,** pp. 413–417, Apr., 1950.
28. Birdsall, C. K., and Brewer, G. R., Traveling Wave Tube Characteristics for Finite Values of C, Trans. I.R.E., **PGED-1,** pp. 1–11, Aug., 1954.
29. Pierce, J. R., Traveling Wave Oscilloscope, Electronics, **22,** Nov., 1949.

The Field Displacement Isolator

By S. WEISBAUM and H. SEIDEL

(Manuscript received February 7, 1956)

A nonreciprocal ferrite device (field displacement isolator) has been constructed with reverse to forward loss ratios of about 150 in the region from 5,925 to 6,425 mc/sec. The forward loss is of the order of 0.2 db while the reverse loss is 30 db. These results are obtained by using a single ferrite element, spaced from the sidewall of the guide. The low forward loss suggests the existence of an electric field null at the location of a resistance strip on one face of the ferrite. We discuss the various conditions, derived theoretically, under which the electric field null may be obtained and utilized. Furthermore, a method of scaling is demonstrated which permits ready design to other frequencies.

I. INTRODUCTION

The need for passive nonreciprocal structures has long been recognized.[1] In the microwave field, Hogan's gyrator[2] paved the way for an increasingly important class of such devices. The isolator, in particular, has emerged as one of the more useful ferrite components. It performs the function, as its name implies, of isolating the generator from spurious mismatch effects of the load. Unlike lossy pads, which consume generator power, the isolater provides a unidirectionally low loss transmission path.

A. G. Fox, S. E. Miller and M. T. Weiss[3] have pointed out that nonreciprocal ferrite devices may exploit any of the following waveguide effects:

1. Faraday rotation
2. Gyromagnetic resonance
3. Field displacement
4. Nonreciprocal phase shift

In the present paper we shall discuss an isolator, based upon the field displacement effect, which was developed to meet the following requirements for a proposed microwave relay system (5,925–6,425 mc/sec):

1. Forward loss 0.2 db

2. Reverse loss 20 db
3. Return loss 30 db

The field displacement isolator employs an ordinary rectangular waveguide and requires no specialized adaptation to the rest of the guide system. It is relatively compact and does not require excessive magnetic fields. In contrast to the field displacement structure of Reference 3, in which a symmetrically disposed pair of ferrite slabs is used, the present unit (see Fig. 1) contains only a single slab. Other differences of a more substantial nature may be noted — in the present case the slab is displaced from the guide wall, it occupies a partial height of the waveguide, and it employs a novel disposition of the absorption material on one face. These features result in a broadband device.

In the analysis presented in this paper the isolator field characteristics for a full height slab are determined by exact solution of Maxwell's equations, as opposed to the "point-field" perturbation approximation used in Reference 3. An exact solution of the partial height geometry of the experimental device would be exceedingly difficult to obtain. However, such a solution did not appear to be essential for this investigation since good correspondence has been obtained between the experimental results and the idealized full height slab calculations.

The following performance of the isolator was obtained from 5,925–6,425 mc/sec:

1. Forward loss $\sim$ 0.2 db

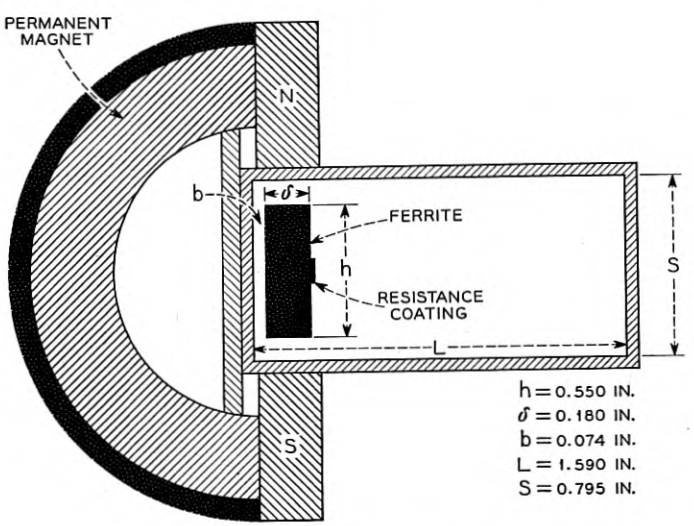


Fig. 1 — Field displacement isolator.

2. Reverse loss $\sim$ 30 db
3. Return loss $\sim$ 30 db

The extremely low forward loss strongly suggested the existence of an electric field null in the plane of the resistance material. Consequently, a theoretical investigation of the null condition was made and a set of criteria established for the existence and utilization of the null. (E. H. Turner[4] independently developed the same null conditions.) An extension of the analysis leads directly to a set of scaling laws which permits the ready design of isolators of comparable performance at other frequency bands.

II. DESCRIPTION OF OPERATION

In Section IIA we will show how the "point-field" approach[3] is used to predict the qualitative behavior of the structure and in Section IIB we will apply a more rigorous analysis to the determination of the optimum design parameters.

A. Qualitative

Prior to introducing the actual isolator configuration, we shall review some elementary properties both of the ferrite medium and of an unloaded rectangular waveguide. It is in terms of these properties that we can understand, in a qualitative sense, the interaction of an *rf* wave with a ferrite in such a waveguide. Since the behavior of a ferrite medium in the presence of a static magnetic field and a small *rf* field has been discussed in the literature[5] the following resumé is not intended to be detailed. It is presented, however, to maintain continuity.

If a static magnetic field is applied to a ferrite medium the unpaired electron spins, on the average, will line up with the field. If now an *rf* magnetic field, transverse to the dc field, excites the spin system these electrons will precess, in a preferential sense, about the static field. The precession gives rise to components of transverse permeability at right angles to the *rf* magnetic field, leading to a tensor characterization of the medium. This tensor has been given by Polder[5] and may be diagonalized in terms of circularly polarized wave components. Corresponding to the appropriate sense of polarization we use the designation + and −. When the polarization is in the same sense as the natural precessional motion of the spin system, gyromagnetic resonance occurs for an appropriate value of the static magnetic field. The scalar permeabilities μ_- and μ_+ are shown in Fig. 2 as functions of the internal static magnetic field as would be observed at an arbitrary frequency.

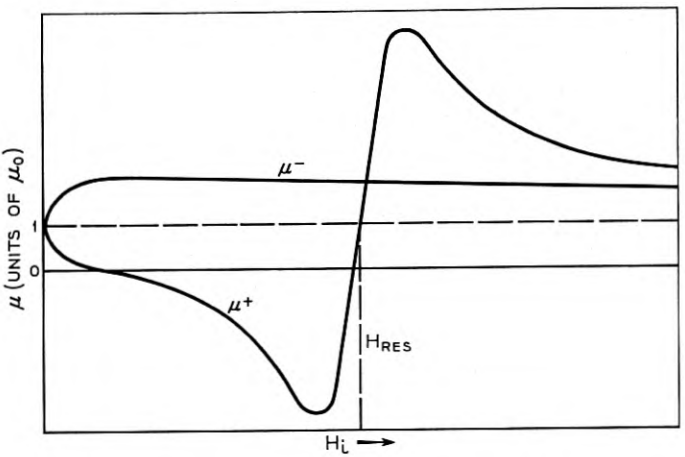
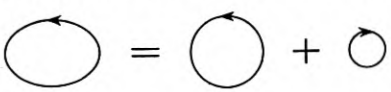
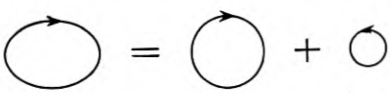


Fig. 2 — Permeability versus magnetic field.

Clearly, in employing a ferrite medium, we intend to use the basic difference between the scalar permeabilities μ_+ and μ_-. To this end we may exploit the fact that the magnetic field configuration at any given point in a rectangular waveguide is, in general, elliptically polarized. Traveling loops of magnetic intensity appear in Fig. 3 for the fundamental (TE_{10}) mode. At point P an observer sees a counterclockwise elliptically polarized magnetic intensity if the wave is traveling in the (+y) direction.* The propagating wave may be decomposed into two oppositely rotating circularly polarized waves of different amplitudes:



For propagation in the (−y) direction the *rf* polarization is reversed:



Let us now consider the actual experimental configuration shown in Fig. 4 (the partial height geometry was chosen on an experimental basis, in that it gave VSWR considerably less than that for a full height ferrite slab). The precession of the spin magnetic moments is counterclockwise

* It is evident that a point converse to P exists symmetrically to the right of center. This is utilized in a double slab isolator which has been investigated by S. Weisbaum and H. Boyet, I.R.E., **44**, p. 554, April, 1956.

looking along the direction of the dc magnetic field shown in Fig. 4. Since the major component of circular polarization for $(+y)$ propagation is also counterclockwise the permeability will be less than unity for this direction of propagation. This occurs provided we are using small static fields, as is readily verified from Fig. 2. The permeability will be greater than unity for $(-y)$ propagation. Physically, this is equivalent to energy being crowded out of the ferrite for $(+y)$ propagation and to energy being crowded in, in the reverse direction. The electric field will thus be distorted as shown in a qualitative way in Fig. 5. The vertical dimension in this figure serves both to identify the guide configuration and to provide an ordinate for the electric field intensity.

The fields as shown in Fig. 5 merely represent a qualitative picture of the distributions in the guide and are not intended to be exact. There is no question, however, that the electric fields at the ferrite face are dif-

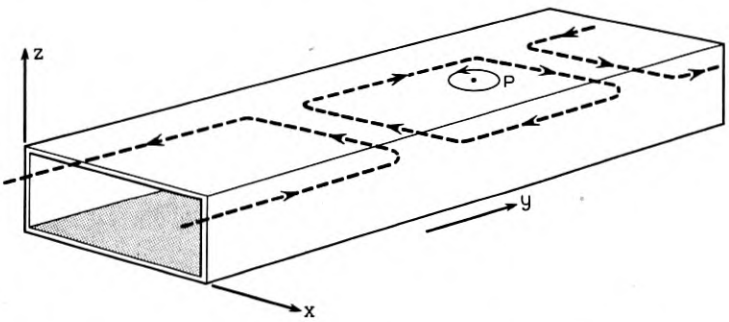


Fig. 3 — Magnetic field configuration — Dominant TE_{10} mode.

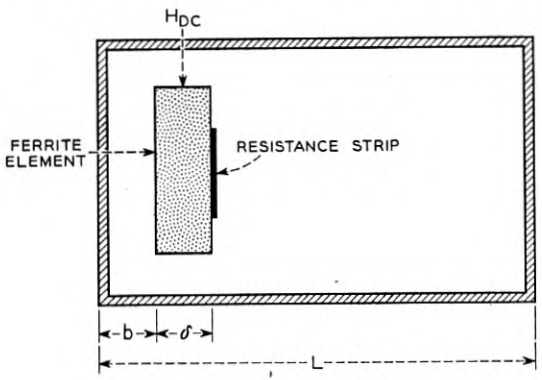


Fig. 4 — Experimental configuration.

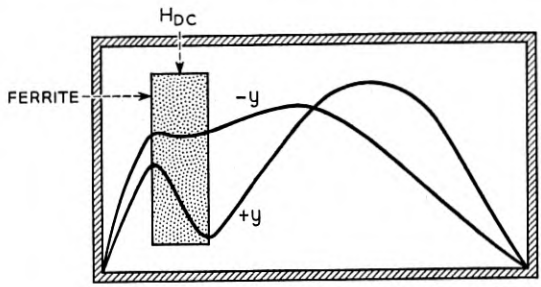


Fig. 5 — Electric field distortion.

ferent in magnitude corresponding to the two directions of propagation. Hence, if resistance material is placed at the interior face of the ferrite (see Fig. 1) we may expect to absorb more energy in one direction of propagation.

B. Analysis of Electric Field Null:-Full Height Ferrite

The description we have given in Section IIA is based on a perturbation approach and does not take into account the higher order interaction effects of the ferrite and the propagating wave. In this section we consider an analysis of the idealized case, namely that of a full height ferrite slab, and impose the condition of an electric field null at the face of the ferrite for the forward direction of propagation. While this too does not represent the true experimental situation, we believe it to be a better approximation than the "point-field" perturbation viewpoint.

The fields of the various regions shown in Fig. 6 are described as follows:

$$\begin{aligned} E_z^{(1)} &= \sin \alpha_1 x \\ E_z^{(2)} &= Ae^{-i\alpha_2 x'} + Be^{i\alpha_2 x'} \quad \text{where} \quad x' = x - a \\ E_z^{(3)} &= V \sin \alpha_1 x'' \quad \text{where} \quad x'' = x - L \end{aligned} \qquad \text{(II — 1)}$$

where

α_j = transverse wave number in the j^{th} region
a = transverse dimension from narrow wall to ferrite face
L = broad waveguide dimension
x = variable dimension along broad face
z = height variable
A, B, V = constants

Setting up the wave equation, there results

$$\alpha_2^2 = K^2\left[\frac{\varepsilon_r}{\mu_r}(\mu_r^2 - k_r^2) - 1\right] + \alpha_1^2 \qquad \text{(II—2)}$$

where μ_r and k_r are the relative diagonal and off-diagonal terms of the Polder tensor, respectively, K is the free space wave number and ε_r is the relative dielectric constant.

$$\mu_r = 1 + \frac{4\pi M_S \gamma \omega_0}{\omega_0^2 - \omega^2}$$

$$k_r = \pm \frac{4\pi M_S \gamma \omega}{\omega_0^2 - \omega^2}$$

$\gamma = 2.8 \times 10^6$ cycles/sec/oersted
$4\pi M_S$ = saturation magnetization in gauss
H_0 = static magnetization in oersteds
$\omega_0 = \gamma H_0$
$K = \dfrac{2\pi}{\lambda}$

The following transcendental equation results from satisfying the boundary conditions on E and H:[6]

$$\frac{\tan \alpha_1 a[\mu_r \alpha_2 + k_r \beta \tan \alpha_2 \delta] + (\mu_r^2 - k_r^2)\alpha_1 \tan \alpha_2 \delta}{(\beta^2 - K^2 \mu_r \varepsilon_r) \tan \alpha_1 a \tan \alpha_2 \delta + \alpha_1(\mu_r \alpha_2 - k_r \beta \tan \alpha_2 \delta)} + \frac{\tan \alpha_1 b}{\alpha_1} = 0 \qquad \text{(II—3)}$$

where β is the propagation constant.

The minimum nontrivial value of α_1 causing a null to appear at the ferrite face is $\alpha_1 = \pi/a$. Placing this value in (II — 3) produces the following transcendental equation for the null:

$$\frac{\frac{\pi}{a}(\mu_r^2 - k_r^2) \tan \alpha_2 \delta}{\mu_r \alpha_2 - k_r \beta \tan \alpha_2 \delta} + \tan \alpha_1 b = 0 \qquad \text{(II—4)}$$

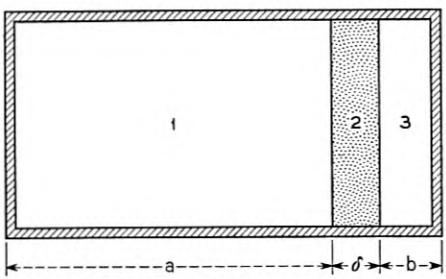


Fig. 6 — Full height geometry.

where $b = L\text{-}a\text{-}\delta$. Equation (II — 4) demonstrates that the null condition is nonreciprocal since, in general, the solutions differ for k_r positive and k_r negative. The quantity k_r has the same sign as the direction of the dc magnetic field; reversing the sign of k_r is equivalent to reversing the direction of propagation.

A numerical analysis of equation (II — 4) has led to the conclusion that the null condition is most broadband when $| \mu_r | < | k_r |$.* We use the criterion $| \mu_r | = | k_r |$ to determine a critical magnetic field:

$$H_c = \frac{\omega}{\gamma} - 4\pi M_s \qquad \text{(II — 5)}$$

Clearly we require $\omega/\gamma > 4\pi M_s$ for physically realizable solutions. The saturation magnetization ($4\pi M_s$) is subject to the following:

1. A choice of too large a $4\pi M_s$ might create a mode problem and in addition will not satisfy the limit on $4\pi M_s$ implied in (II — 5).

2. $4\pi M_s$ must be sufficiently large so that the field needed to make $| \mu_r | < | k_r |$ not be excessive.

3. $\gamma\sqrt{H(H + 4\pi M)}$ (this being the slab resonance frequency for small slab thickness[7]) must be sufficiently far from the operating frequency to avoid loss due to resonance absorption. In addition, this condition improves the frequency insensitivity of the null.

Further analytic considerations are presented in Section IV.

III. EXPERIMENTAL DESIGN CONSIDERATIONS

Aside from the partial height nature of the slab, there are two other basic factors in the experimental situation which are not present in the analysis of Section IIB (see also Section IV). First, the ferrite has both finite dielectric and magnetic loss. Second, higher order modes may be present. These deviations from the simplified analysis are by no means trivial and it would not be surprising if one found a considerable modification of the analytic results. As it turns out, there are broad areas of general agreement between the theoretical and experimental results and in no case examined here does one find a basic inconsistency. In considering the various parameters which must be adjusted to optimize the broadband performance of the isolator we will point out, where possible, how the theoretical results are modified by the factors mentioned above. The parameters of interest are:

* This is partially evident from equation II — 4. The quantity $| \mu_r |$ must be less than $| k_r |$ if the angle $(\alpha_1 b)$ is to be small and in the first quadrant. Second quadrant solutions cause the guide cross section to be excessively large, with attendant higher mode complication.

A. The saturation magnetization ($4\pi M_s$) and the applied magnetic field (H_{DC}).

B. The ferrite height.

C. The thickness (δ) of the ferrite and its distance (b) from the nearest sidewall.

D. The placement of the resistance material and its resistivity (ρ).

E. The length of the ferrite (ℓ).

A. $4\pi M_s$ and H_{DC}

Theoretically, minimum forward loss occurs with a true null at the face of the loss film and has been given in the condition $|\mu_r| < |k_r|$. Although this inequality is required in the full height slab analysis, experiment (Fig. 7) indicates the low loss region to be so broad as to extend well into the low field, or $|\mu_r| > |k_r|$ region.

There is inherent loss in the ferrite so that a more accurate statement of the bandwidth of operation is that in which the losses in the film are of equal order to the ferrite losses at the band edges. Even discounting ferrite losses, it will be shown in Section IV that we have a good analytic basis for the observed broadness of the low loss region. In general, therefore, we need not be as restrictive as the null analysis of Section IIB would imply. It is not surprising then that optimum operation actually occurs in the region $|\mu_r| > |k_r|$. There are several reasons why this may be so:

1. Shift of operation occurs due to the partial height nature of the ferrite slab.

2. Reverse loss has a peak in the low field region, requiring a compromise of low forward loss and high reverse loss for best isolation ratios (see Fig. 8).

3. Optimum compromise between low ferrite loss and low film loss must be made.

The internal magnetic field, determining $|\mu_r|$ and $|k_r|$, differs from the applied field by the demagnetization of the ferrite slab. Although not ellipsoidal, it may nonetheless be considered to have an average demagnetization which has been computed, for this case, to be 460 oersteds. A further complication in knowledge of the internal field is the proximity effect of the pole pieces. This latter correction was obtained experimentally and, all in all, it was determined that the internal field for optimum operation was of the order of 300 oersteds. For the given ferrite and the range of frequency of operation, this internal field corresponds to the condition that $|\mu_r| > |k_r|$, as stated above.

Taking all effects into account, it was found that optimum permanent magnet design occurred for an air gap field of 660 oersteds.

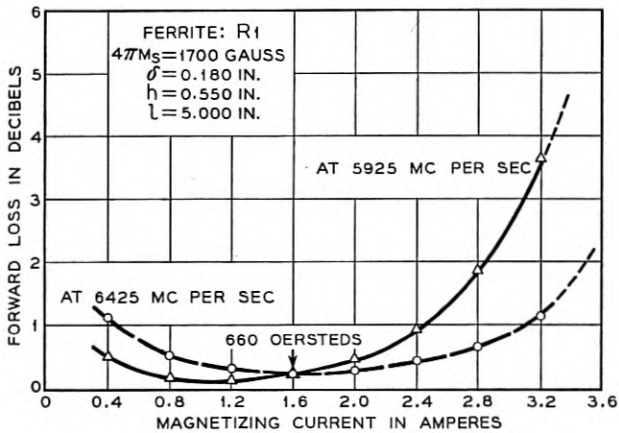


Fig. 7 — Forward loss versus magnetizing current.

Using the experimental values $4\pi M_S$ = 1,700 gauss and internal magnetic field = 300 oersteds, the frequency at which ferromagnetic resonance occurs was estimated to be about 2200 mc/sec. This value is sufficiently far from our operating range (5,925–6,425 mc/sec) that we would expect a negligible loss contribution due to resonance absorption. This is confirmed by the low forward loss actually observed.

B. Ferrite Height

We have already pointed out that when the ferrite height is reduced from full height a more reasonable VSWR is obtained. This is due to the fact that we have relieved the stringent boundary requirements at the

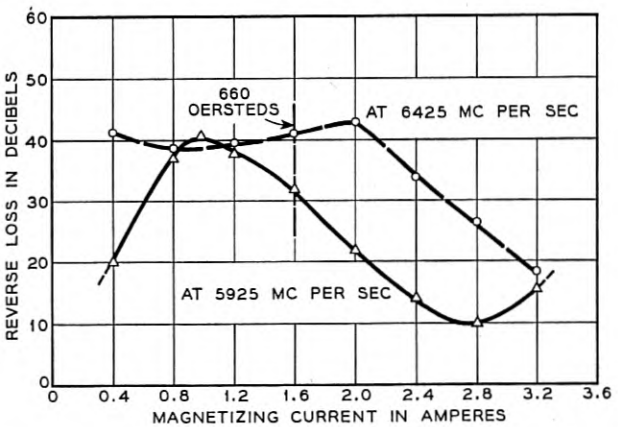


Fig. 8 — Reverse loss versus magnetizing current.

top and bottom faces of the ferrite and approach, in a sense, a less critical rod type geometry. A ferrite height of 0.550″ gave a VSWR $\sim$ 1.05 over the band. With full height slabs (0.795″), VSWR values as high as 10:1 have been observed for typical geometries.

C. δ and b

Experimentally, we have examined various ferrite thicknesses at different distances from the sidewall until optimum broadband performance was obtained. Table I shows the ferrite distance from the wall which gave the best experimental results (highest broadband ratios, low forward loss, high reverse loss) for each thickness δ of one of the BTL materials. It is interesting to note that the empirical quantity $\delta + b/2 - 2t$, where t is the thickness of the resistive coating, is very nearly constant (within a few mils) for the stated range of δ and for this type of design.*

TABLE I

δ (mils)	b (mils)	$\delta + \frac{b}{2}$ (mils)	t (mils)	$\delta + \frac{b}{2} - 2t$ (mils)
201	11	206.5	3	200.5
189	35	206.5	3	200.5
186	42	207.0	3	201.0
176	65	208.5	3	202.5
189	42	210.0	6	198.0

In Section IV a theoretical calculation using the null condition at 6175 mc/sec for a full height ferrite gives

$$\delta = 180 \text{ mils}$$

$$b = 38.7 \text{ mils}$$

so that $\delta + b/2 = 199.3$ mils. In the theoretical case t is assumed to be very small. It will be noted that the theoretical result for $\delta + b/2$ (with small t) agrees quite well with the experimental $\delta + b/2 - 2t$. The question of the possible physical significance of this quantity is being investigated.

D. Placement of Resistance Material and Choice of Resistivity

The propagating mode with a full height ferrite slab is of a TE_0 variety, the zero subscript indicating that no variation occurs with re-

* In one design of the isolator we used a General Ceramics magnesium manganese ferrite with $\delta = 0.180''$, $b = 0.074''$ and $t = 0.009''$ so that $\delta + b/2 - 2t = 199$ mils, in good agreement with Table I.

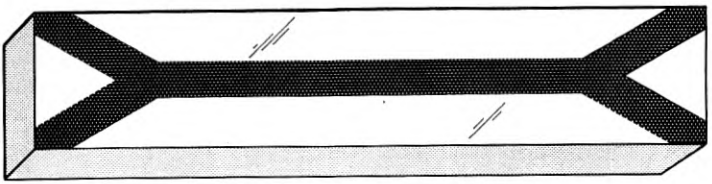


Fig. 9 — Distribution of small tangential electric fields at interior ferrite face.

Fig. 10 — Resistance configuration.

spect to height. A field null in this construction therefore extends across the entire face of the full height ferrite and all of this face is then "active" in the construction of an isolator. This field situation no longer accurately applies to the partial height slab. The departure of the ferrite from the top wall creates large fringing fields extending from the ferrite edges, and large electric fields may exist tangential to the ferrite face close to these edges. We would therefore expect the null condition to persist only in a small region about the vertical center of the ferrite face. We may, however, also expect longitudinally fringing modes (TM-like) to be scattered at the input edge of the ferrite slab so that a longitudinal field maximum will exist at the central region of the ferrite. However, this is a higher mode, so that this maximum decays rapidly past the leading edge.

Considering all the effects, the distribution of small tangential electric fields at the ferrite face may be expected to appear as shown in Fig. 9. Experimentally, we have utilized this low loss region and have avoided the decay region of the higher TM-like modes by using the resistance configuration shown in Fig. 10. The resistivity is uniform and about 75

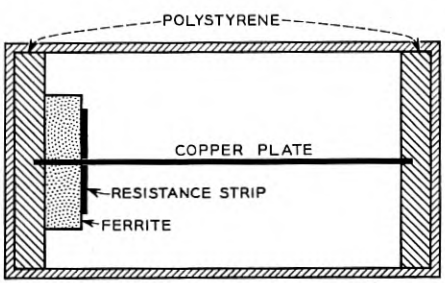


Fig. 11 — Elimination of longitudinal components.

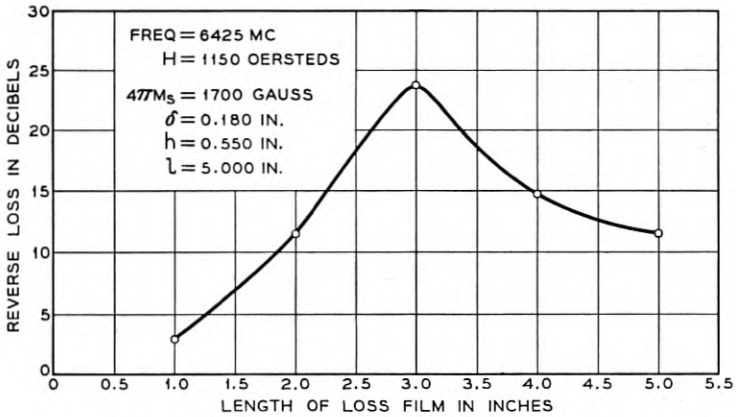


Fig. 12 — Attenuation versus length of resistance strip.

ohms/square. Variations of about ±30 ohms/square about this value result in little deterioration in performance.

Some further discussion of the perturbed dominant mode is of interest. We may think of the height reduction as primarily a dielectric discontinuity where we have effectively added a negative electric dipole density to a full height slab. Since this addition is smaller for the forward case (where there was initially a small electric field) than for the reverse case, we may expect the longitudinal components to be smaller for the forward propagating mode. The other type of longitudinal electric field,

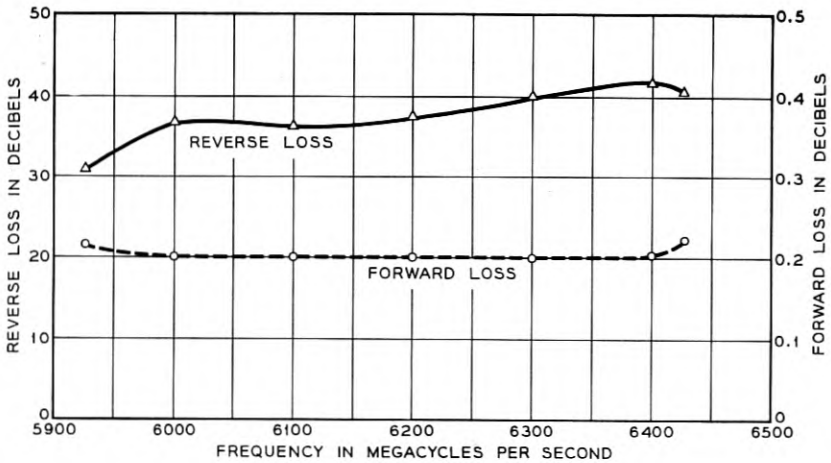


Fig. 13 — Loss versus frequency.

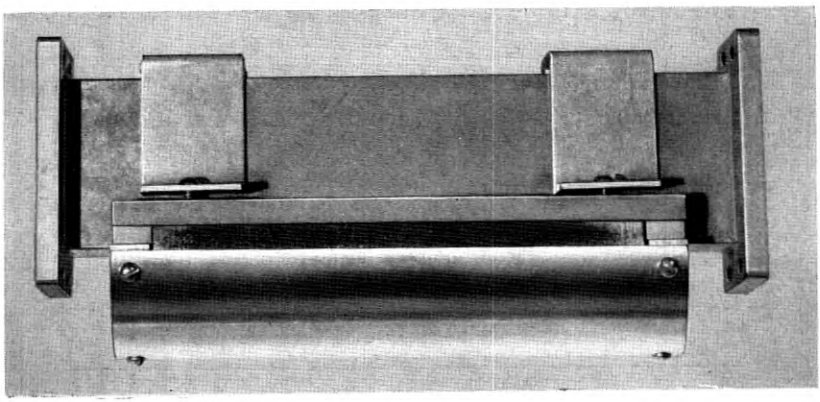


Fig. 14 — Isolator model.

which occurs due to the scattering of the TM-like longitudinal modes. decays rapidly and is not of consequence in an experiment now to be described. This experiment was designed to demonstrate the nonreciprocal nature of the longitudinal electric fields associated with the distorted dominant mode. It also shows that the existence of these components is significant as a loss mechanism for the reverse direction of propagation in the isolator. The geometry employed is shown in Fig. 11.

The copper plate was inserted to minimize longitudinal electric field components, and we may therefore expect to obtain less reverse loss than in the condition of its absence. The result of this experiment was that the reverse loss decreased from about 25 db* without the plate to 18 db with the plate. The forward loss was unaffected.

E. Determination of Length

Given a dominant mode distribution in a waveguide, attenuation will be a linear function of length, once this mode has been established. Consequently, one would expect that doubling the loss film length would double the isolator reverse loss. The isolator does not exhibit this behavior, however, as is illustrated in Fig. 12.

This occurrence might be explained by the appearance of still another longitudinal mode, peculiar in form to gyromagnetic media alone, which propagates simultaneously with the transverse electric mode, and is essentially uncoupled to the loss material. The maximum reverse loss

* This experiment was conducted with a different ferrite than that employed in the eventual design.

thus obtainable is limited by the scattering into this mode. The character of these singular modes will be discussed in a subsequent paper.

Results

The performance of the isolator as a function of frequency is shown in Fig. 13. Fig. 14 shows a completed model of the isolator.

IV. FURTHER ANALYSIS

While an exact characteristic equation is obtainable for the overall geometry of the full height isolator, including the lossy film, the expressions which result are sufficiently complex to be all but impossible to handle. However, if the resistance film is chosen to have small conductivity we may utilize a simple perturbation approach in which the field at the ferrite face is assumed to be unaffected by the presence of the loss film. A quantity η may then be defined* so that

$$\eta = \frac{|E_R|^2}{P} \qquad \text{(IV — 1)}$$

For small conductance values η is proportional to attenuation to first order in either direction of propagation. E_R, in equation (IV — 1), is the electric field adjacent to the film and P is the power flowing across the guide cross section. The loss in the ferrite material is not taken into account in this approximation, but it would naturally have a deteriorating effect on the isolator characteristics.

The ratio of the values of η corresponding to backward and forward direction of propagation defines the isolation ratio, given in db/db, for the limit of very small conductivity.

Fig. 15 shows a calculated curve of the forward value of η and Fig. 16 shows the backward case. The isolation ratio shown in Fig. 17 demonstrates surprisingly large bandwidth for values of the order of 200 db/db. Fig. 18 portrays propagation characteritics for both forward and backward power flows and provides the interesting observation, in conjunction with Fig. 16, that peak reverse loss occurs in the neighborhood of $\lambda = \lambda_g$.

Fig. 19 is a plot of α_1, the transverse wave number, over the frequency range. The flatness of the forward wave number means that the position of null moves very little with frequency across the band. Hence the lossless transmission in the forward direction is broadband. Since the forward and backward wave numbers have such radically different

* See Appendix

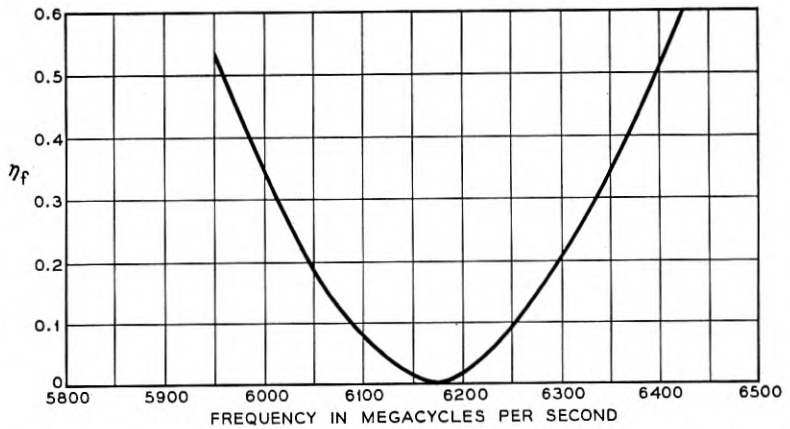


Fig. 15 — Relative attenuation — forward direction

rates of variation, a simple adjustment of parameters may be made to cause the forward null and maximum reverse attentuation to appear at the same frequency, resulting in an optimum performance.

The occurrence of the reverse maximum loss in the region of $\lambda = \lambda_g$ may roughly be explained as follows. As the transverse air wave number decreases, the admittance of the guide, defined on a power flow basis, increases. The electric field magnitude distribution must therefore generally decrease in such a fashion as to cause the overall power flow to re-

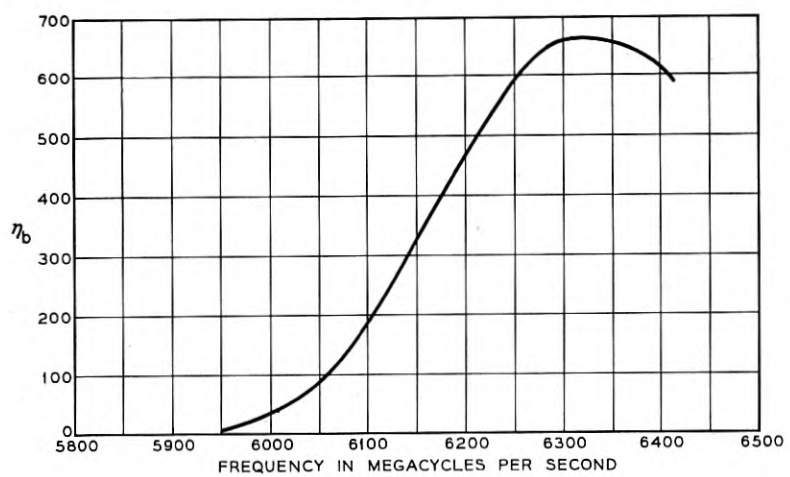


Fig. 16 — Relative attenuation — backward direction

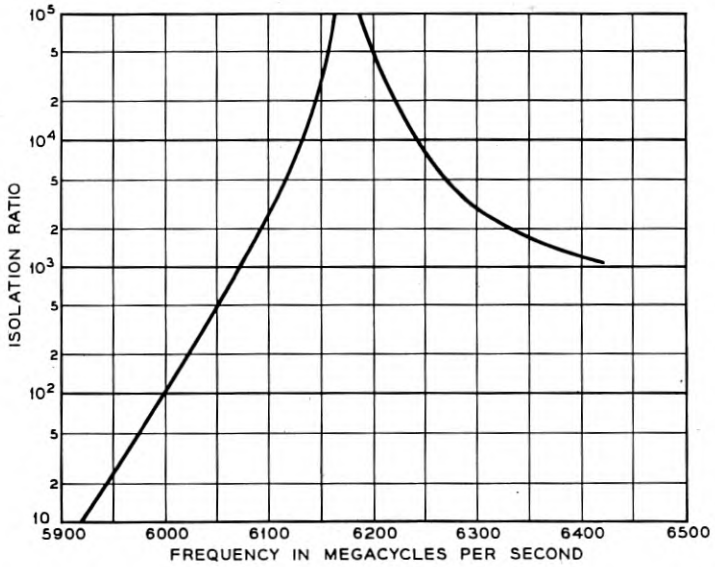


Fig. 17 — Ideal isolation characteristics.

main constant. On the other hand as the transverse air wave number decreases through real values, the electric field adjacent to the ferrite becomes relatively large. At $\lambda = \lambda_g$ the distribution is linear with relatively large dissipation at the ferrite face. As the transverse air wave number increases through imaginary values the distribution becomes exponential such that the field adjacent to the ferrite is always the maximum for the air region and the growth of the field at the face of the ferrite would not seem to be so great as formerly. One would therefore expect a maximum reverse loss somewhere in the region $\lambda = \lambda_g$.

The above considerations plus the transcendental equation for the null show consistency with the experimental design values which were:

$$\delta = 0.180''$$

$$L = 1.59$$

$$4\pi M_S = 1{,}700 \text{ gauss}$$

Using $H_{DC} = 600$ oersteds in the calculation we obtain the spacing from the guide wall $b = 0.0387''$. The fact that we used 600 oersteds for the full height slab calculation as opposed to the internal field of 300 oersteds found experimentally for the partial height slab should not be a source of confusion. It has been indicated earlier that the peak reverse loss shifts

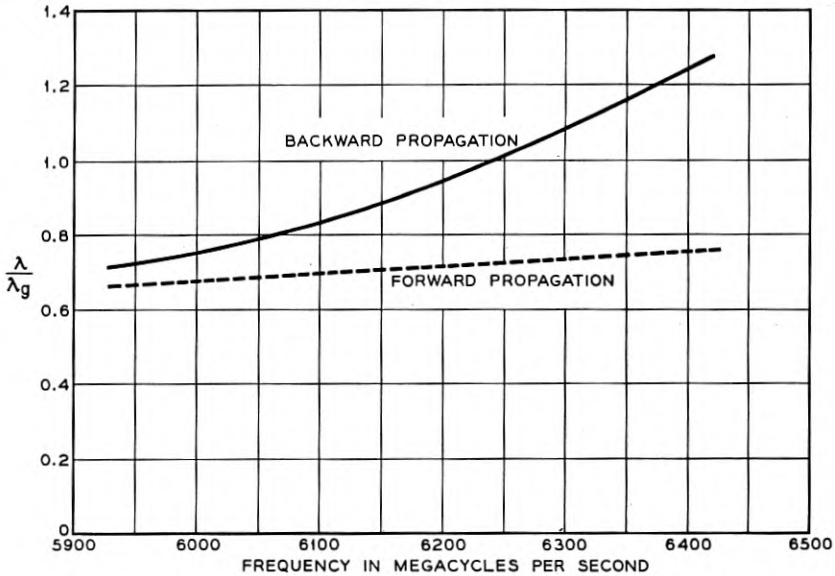


Fig. 18 — Ferrite isolator characteristics.

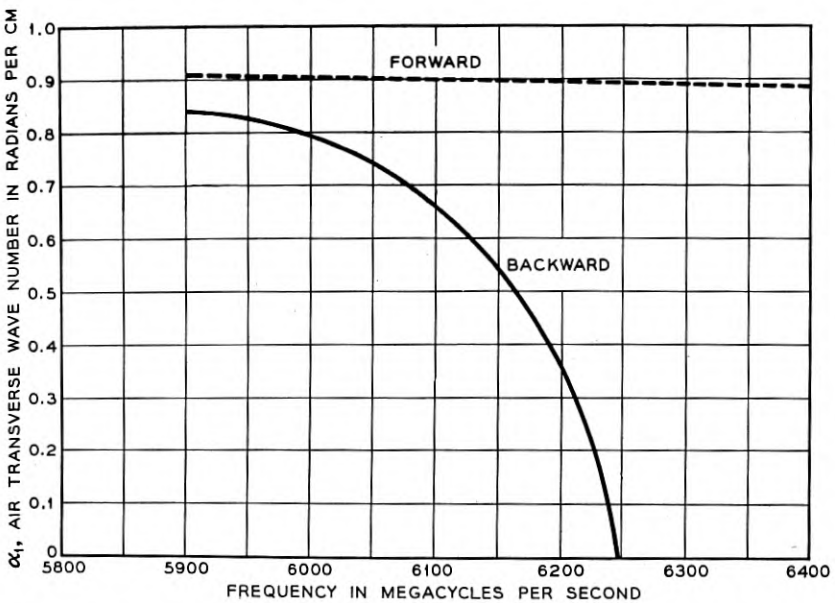


Fig. 19 — Transverse characteristics of a ferrite isolator

with ferrite height reduction. It is not inconsistent therefore to choose 600 oersteds for the full height analysis in contrast to the value determined from the experiment.

V. SCALING

Once the optimum set of parameters has been decided upon for a given frequency range (e.g., 5,925–6,425 mc/sec, $\delta = 0.180''$, $b = 0.074''$, $\ell = 5''$, $h = 0.550''$, $4\pi M_S = 1{,}700$ gauss, $H_{DC} = 660$ oersteds) it is a simple matter to scale these parameters to other frequency ranges. From Maxwell's equations:

$$\text{Curl } H = i\omega\varepsilon E + gE$$

$$\text{Curl } E = -i\omega T \cdot H$$

where T is the permeability tensor, and g is the conductivity in mhos/meter. The first of Maxwell's equations suggest that frequency scaling may be accomplished by permitting both the curl and the conductance to grow linearly with respect to frequency. The curl, which is a spatial derivative operator, may be made to increase appropriately by shrinking all dimensions by a $1/\omega$ factor, which will keep the field configuration the same in the new scale.

Having imposed this condition on the first equation we must satisfy the second of Maxwell's equations by causing T to remain unchanged with frequency. T is a tensor given as follows for a cartesian coordinate system:

$$T = \begin{pmatrix} \mu_r & ik_r & 0 \\ -ik_r & \mu_r & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{(V — 1)}$$

for a magnetizing field in the z direction. The components may be expanded in the following fashion:

$$\mu_r = \frac{1 + 4\pi\left(\frac{\gamma M_S}{\omega}\right)\left(\frac{\gamma H}{\omega}\right)}{\left(\frac{\gamma H}{\omega}\right)^2 - 1}$$

$$k_r = \frac{4\pi\left(\frac{\gamma M_S}{\omega}\right)}{\left(\frac{\gamma H}{\omega}\right)^2 - 1} \qquad \text{(V — 2)}$$

where $4\pi M_s$ is the saturation magnetization in gauss and γ is the magnetomechanical ratio. The Polder tensor evidently remains unchanged if M_s and H are both scaled directly with frequency.

Since the field distributions are assumed unchanged relative to the scale shift, normal and tangential E and H field components continue to satisfy the appropriate boundary equalities at interfaces. Then, invoking the uniqueness theorem, the guide characteristics are only as presumed and the model has been properly scaled as a function of frequency.

The scaling equations are:

$$\begin{aligned} d_1 &= \frac{\omega_2}{\omega_1} d_2 \\ g_1 &= \frac{\omega_1}{\omega_2} g_2 \\ M_{s_1} &= \frac{\omega_1}{\omega_2} M_{s_2} \\ (H_0)_1 &= \frac{\omega_1}{\omega_2} (H_0)_2 \end{aligned} \qquad \text{(V—3)}$$

where d is any linear dimension.

CONCLUSION

An isolator with low forward loss and high reverse loss can be constructed by a proper choice of parameters. Once a suitable design has been reached the scaling technique can be used to reach a suitable design for other frequencies.

As yet, a theoretical analysis of this problem has been carried out only for a full height ferrite.

ACKNOWLEDGMENT

We would like to thank F. J. Sansalone for his assistance in developing the field displacement isolator. We would also like to thank Miss M. J. Brannen for her competent handling of the numerical computations.

APPENDIX

It is desirable to establish an isolator figure of merit. A simple quantity characterizing the isolator action is the normalized rate of power

loss in the resistive strip, for an idealized ferrite, in the low conductive limit of such a strip. Let

$$\eta = \frac{|E_r|^2}{P}$$

where η is the appropriate quantity, E_r is the field at the resistance, and P is the total power flow across the guide cross-section. This figure of merit is related to the rate of change of the attenuation constant (A) with respect to strip conductance in the following manner:

$$\frac{dA}{dg} = 0.04343\eta h \text{ (db)(ohms)/cm}$$

where h is the fractional height of the loss strip, and g is the reciprocal of the surface resistivity in ohms/square.

The total power flow may be divided into integrations of the Poynting vector over the three regions of the guide cross-section. The following results are obtained normalized to $E_r = \sin \alpha_1 a$:

Region 1: $0 \leqq x \leqq a$

$$P_y^{(1)} = \frac{\beta}{2\omega\mu_0}\left(a - \frac{\sin 2\alpha_1 a}{2\alpha_1}\right)$$

Region 2: $a \leqq x \leqq a + \delta$

$$P_y^{(2)} = \frac{\beta}{2\omega\mu_0}\left(\frac{\mu_r\delta}{\mu_r^2 - k_r^2}(d_1^2 + d_2^2) + \frac{\sin 2\alpha_2\delta}{2\alpha_2(\mu_r^2 - k_r^2)} \cdot\left[\mu_r(d_1^2 - d_2^2) - \frac{k_r}{\beta}\alpha_2(2d_1d_2)\right] + \frac{1 - \cos 2\alpha_2\delta}{2\alpha_2(\mu_r^2 - k_r^2)} \cdot\left[\mu_r(2d_1d_2) + \frac{k_r\alpha_2}{\beta}(d_1^2 - d_2^2)\right]\right)$$

Region 3: $a + \delta \leqq x \leqq L$

$$P_y^{(3)} = \frac{\beta}{2\omega\mu_0}\left(b - \frac{\sin 2\,\alpha_1 b}{2\alpha_1}\right)\left(\frac{d_1 \cos \alpha_2\delta + d_2 \sin \alpha_2\delta}{\sin \alpha_1 b}\right)^2$$

where $$d_1 = \sin \alpha_1 a$$

and

$$d_2 = \frac{1}{\mu_r\alpha_2}[(\mu_r^2 - k_r^2)\alpha_1 \cos \alpha_1 a + k_r\beta \sin \alpha_1 a]$$

BIBLIOGRAPHY

1. Tellegen, B. D. H., Philips Res. Rep., **3,** 1948.
2. Hogan, C. L., B.S.T.J. **31,** 1952.
3. Fox, A. G., Miller, S. E., and Weiss, M. T., B.S.T.J. **34,** p. 5., Jan. 1955.
4. Turner, E. H., URSI Michigan Symposium on Electromagnetic Theory, June, 1955.
5. Polder, D., Phil. Mag., **40,** 1949.
6. Lax, B., Button, K. J., Roth, L. M., Tech Memo No. 49, M.I.T. Lincoln Laboratory, Nov. 2, 1953.
7. Kittel, C., Phys. Rev., **73,** 1948.

Transmission Loss Due to Resonance of Loosely-Coupled Modes in a Multi-Mode System

By A. P. KING and E. A. MARCATILI

(Manuscript received January 17, 1956)

In a multi-mode transmission system the presence of spurious modes which resonate in a closed environment can produce an appreciable loss to the principal mode. The theory for the evaluation and control of this effect under certain conditions has been derived and checked experimentally in the particularly interesting case of a TE_{01} *transmission system, where mode conversion to* TE_{02}, TE_{03} $\cdots$ *is produced by tapered junctions between two sizes of waveguide.*

INTRODUCTION

In a transmission system, the presence of a region which supports one or more spurious modes can introduce a large change in the transmission loss of the principal mode when the region becomes resonant for one of the spurious modes. This phenomenon can occur even when the mode conversion is low and the waveguide increases in cross section smoothly to a region which supports more than one mode. In general, the conditions required to resonate the various spurious modes are not fulfilled simultaneously and, in consequence, interaction takes place between the principal mode and only one of the spurious modes for each resonating frequency. Under these conditions the resonating environment can be visualized as made of only two coupled transmission lines, one carrying the desirable mode and the other the spurious one. This simplification makes it possible to calculate the transmission loss as a function of (1) the coefficient of conversion between the two modes and (2) the attenuation of the modes in the resonating environment. The theory has shown good agreement with the measurement of transmission loss of the TE_{01} mode in a pipe wherein a portion was tapered to a larger diameter which can support the TE_{02} mode.

TRANSMISSION LOSS OF A WAVEGUIDE WITH A SPURIOUS MODE RESONATING REGION

Let us consider a single-mode waveguide connected to another of different cross-section that admits two modes. Since these two modes are orthogonal, the junctions may be considered as made of three single-mode lines connected together, provided we define the elements of the scattering matrix properly. The three modes, or lines in which they travel, are indicated by the subscripts 0, 1, and 2, as shown in Fig. 1. If a_0, a_1, a_2 and b_0, b_1, b_2 are the complex amplitudes of the electric field of the incident and reflected waves respectively, then

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = [S] \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

where

$$[S] = \begin{bmatrix} \Gamma_{00} & \Gamma_{01} & \Gamma_{02} \\ \Gamma_{01} & \Gamma_{11} & \Gamma_{12} \\ \Gamma_{02} & \Gamma_{12} & \Gamma_{22} \end{bmatrix} \tag{1}$$

is the scattering matrix.[1]

This specific type of change of cross section may be treated as a three-port junction.

Now, if a length ℓ of a two mode waveguide is terminated symmetrically at both ends with a single mode waveguide (Fig. 2), each joint is described by the same matrix (1), and the connecting two mode wave guide has the following scattering matrix:

$$\begin{bmatrix} 0 & e^{-j\theta_1} & 0 & 0 \\ e^{-j\theta_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{-j\theta_2} \\ 0 & 0 & e^{-j\theta_2} & 0 \end{bmatrix} \tag{1'}$$

in which

$$j\theta_1 = \gamma_1 \ell = (\alpha_1 + j\beta_1)\ell$$

$$j\theta_2 = \gamma_2 \ell = (\alpha_2 + j\beta_2)\ell,$$

γ_1 and γ_2 are the propagation constants of modes 1 and 2.

[1] N. Marcuvitz, Waveguide Handbook, **10**, M.I.T., Rad. Lab. Series, McGraw-Hill, New York, 1951, pp. 107–8.

Matrices 1 and 1′ describe the system completely and from them, the transmission coefficient results,

$$\Gamma = \frac{b_0'}{a_0}$$

$$= \Gamma_{01}^2 e^{-j\theta_1} \frac{A - \Gamma_{22}^2 e^{-j2\theta_2} A^* \left(1 + \left|\frac{\Gamma_{12}}{\Gamma_{22}}\right|^2 + \frac{\Gamma_{00}}{\Gamma_{01}} \frac{\Gamma_{02}^*}{|\Gamma_{22}|^2} \Gamma_{12}\right)^2}{[1 - (\Gamma_{12}^2 - \Gamma_{11}\Gamma_{22})e^{-j(\theta_1+\theta_2)}]^2 - (\Gamma_{11}e^{-j\theta_1} + \Gamma_{22}e^{-j\theta_2})^2} \tag{2}$$

where

$$A = 1 + \left(\frac{\Gamma_{02}}{\Gamma_{01}}\right)^2 e^{-j(\theta_2-\theta_1)}$$

A^* is the complex conjugate of A

Γ_{01}^* is the complex conjugate of Γ_{01}

Γ_{02}^* is the complex conjugate of Γ_{02}

Furthermore, let us make the following simplifying assumptions

$$\Gamma_{00} = 0 \tag{3}$$

$$|\Gamma_{02}| \ll 1 \tag{4}$$

$$\sum_{\beta=0}^{\beta=2} \Gamma_{\beta n}\Gamma_{\beta m}^* = \begin{cases} 1, & \text{if} \quad m = n = 0, 1, 2 \\ 0, & \text{if} \quad m \neq n \end{cases} \tag{5}$$

Equation (3) indicates that if in Fig. 1, lines 1 and 2 were matched, line 0 would also be matched looking toward the junction. Equation (4) states that almost all the transmission is made from 0 to 1, or that there is small mode conversion to the spurious mode 2. Equation (5) assumes that the transition is nondissipative. The first two conditions are fulfilled when the transition is made smoothly. The last is probably the most stringent one, especially if the transition is a long tapered waveguide section, but it is always possible to imagine the transition as lossless and attribute its dissipation to the waveguides.

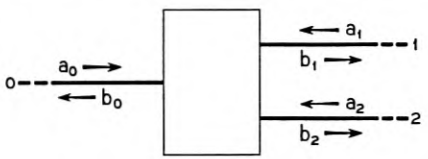


Fig. 1 — Schematic of a three-port junction.

From (2), (3), (4) and (5)

$$\Gamma = \frac{\Gamma_{01}^2 e^{-j\theta_1}}{|\Gamma_{22}|^2} \frac{1}{1 + \left|\frac{\Gamma_{11}}{\Gamma_{22}}\right| e^{-j\varphi}} \left\{1 - \frac{2|\Gamma_{12}|^2(1 - \cos\varphi)}{1 - [\Gamma_{22}e^{-j\theta_2} + \Gamma_{11}e^{-j\theta_1}]^2}\right\} \quad (6)$$

where

$$\Gamma_{11} = |\Gamma_{11}| e^{j\varphi_{11}}$$

$$\Gamma_{22} = |\Gamma_{22}| e^{j\varphi_{22}}$$

$$\varphi = \theta_1 - \theta_2 - \varphi_{11} + \varphi_{22}$$

In order to understand this expression physically, let us suppose first that there is no attenuation. The transmission coefficient Γ becomes 0 when the following equations are fulfilled simultaneously

$$\beta_2 \ell - \varphi_{22} = p\pi \qquad p = 0, 1, 2, 3 \cdots \quad (7)$$

and

$$\varphi = (2q + 1)\pi \qquad q = 0, 1, 2, 3 \cdots \quad (8)$$

The first of these equations states that the line carrying the feebly coupled mode must be at resonance, since this condition is satisfied when the electric length of this line is modified by a multiple of π radians. The second condition, (8), implies that both paths, in lines 1 and 2, must differ in such a way that electromagnetic waves coming through them must arrive in opposite phase at the end of the two-mode waveguide. This is quite clear if we think that, in order to get complete reflection, signals coming through lines 1 and 2 must recombine again with the same intensity and opposite phase. In order to get both modes with the same intensity, the converted mode must be built up through resonance; the opposite phase is obtained by an appropriate electric length adjustment. When attenuation is present, Γ will not be 0, and conditions (7) and (8) for minimum transmission are modified only slightly if the

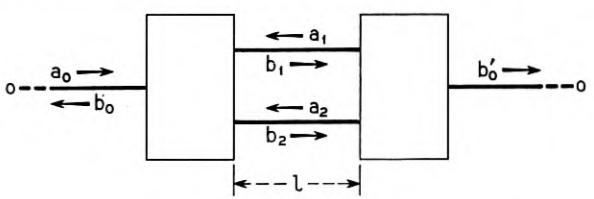


Fig. 2 — Schematic of a two-mode waveguide terminated symmetrically on each side with a single-mode waveguide.

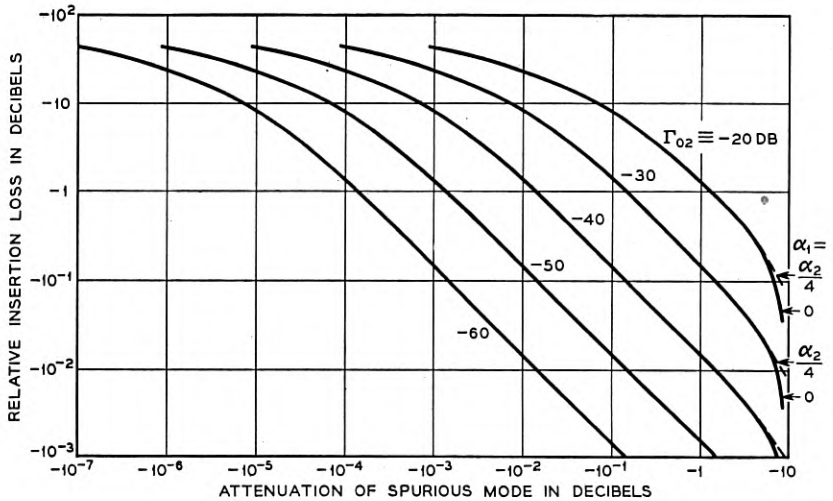


Fig. 3 — Relative insertion loss as a function of the spurious mode attenuation and mode conversion level.

attenuation is low, but the general interpretation of the phenomenon is still the one given above.

From (6) we can calculate the extreme values of $|\Gamma|$ differentiating with respect to ℓ, and we define the relative insertion loss I in db, as the ratio between the minimum and maximum transmitted power expressed in db.

$$I = 20 \log_{10} \left| \frac{\Gamma_{\min}}{\Gamma_{\max}} \right| = 20 \log_{10} \frac{B}{C} \frac{1 - \dfrac{2 |\Gamma_{12}|^2 (1 + \cosh \alpha \ell)}{1 - C^2 |\Gamma_{22}|^2 e^{-2\alpha_2 \ell}}}{1 - \dfrac{2 |\Gamma_{12}|^2 (1 - \cosh \alpha \ell)}{1 + B^2 |\Gamma_{22}|^2 e^{-2\alpha_2 \ell}}} \tag{9}$$

where

$$B = 1 + \left| \frac{\Gamma_{12}}{\Gamma_{22}} \right|^2 e^{-\alpha \ell} \qquad C = 1 - \left| \frac{\Gamma_{12}}{\Gamma_{22}} \right|^2 e^{-\alpha \ell} \qquad \alpha = \alpha_1 - \alpha_2$$

For the most important practical case, that is, when the maximum value attainable by $\cosh \alpha \ell$ is of the order of 1, and knowing from (3), (4) and (5) that

$$|\Gamma_{12}|^2 = |\Gamma_{02}|^2 (1 - |\Gamma_{02}|^2)$$

$$|\Gamma_{22}|^2 \cong 1 - 2 |\Gamma_{02}|^2$$

$$I \cong 20 \log_{10} (1 + 2|\Gamma_{02}|^2 e^{-\alpha \ell}) \left\{1 - \frac{2|\Gamma_{02}|^2(1 + \cosh \alpha\ell)}{1 - e^{-2\alpha_2 \ell} + 2|\Gamma_{02}|^2(1 + e^{-\alpha\ell})e^{-2\alpha_2 \ell}}\right\} \tag{10}$$

From this expression we deduce

(a), I is strongly reduced when $\alpha_2 \ell \gg \Gamma_{02}$.

(b), Attenuation in line 1 is not an important factor until $\alpha_1 \ell$ and $|\alpha_1 - \alpha_2| \ell$ are of the order of 1. In other words, for low attenuation in both lines, $\alpha_2 \ell$ assumes a major importance in the determination of I because it influences the conditions of resonance. That the effect of $\alpha_1 \ell$ is small is shown in Fig. 3 (dotted line for the particular case $\alpha_1 = \alpha_2/4$).

In order to handle the general problem, (10) has been plotted in Fig. 3. We can enter with any two and obtain the third following quantities: I_1, relative insertion loss in db; $10 \log_{10} e^{-2\alpha_2 \ell}$, attenuation in db of the spurious mode in the resonating environment; and $20 \log_{10} \Gamma_{02}$ conversion level at the junction, in db, of power in the spurious mode relative to that in the first line.

APPLICATION OF THESE RESULTS TO A TE_{01} TRANSMITTING SYSTEM

The results of the preceding section have been checked experimentally by measuring the relative insertion loss of different lengths of 7/8" diameter round waveguide tapered at both ends to round waveguides of 7/16" diameter. This waveguide is shown in Fig. 4 with a schematic diagram of the measuring set. In the round transmission line A-B, section A will propagate only TE_{01}. Section B, which has been expanded by means of the conical taper T_1, can support TE_{02} and TE_{03} in addition to the principal TE_{01} mode. This section is a closed region to the spurious modes (TE_{02}, TE_{03}) whose length can be adjusted to resonate each one of these modes. A sliding piston provides a means for varying the length, ℓ, of section B.

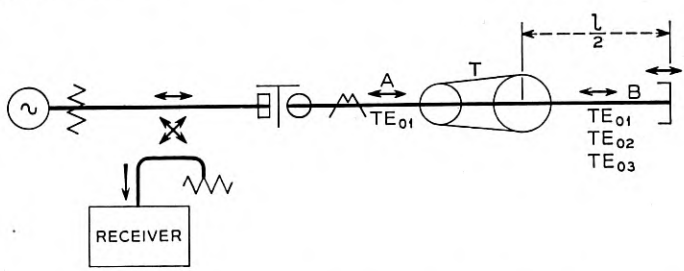


Fig. 4 — Circuit used to measure TE_{01} insertion loss due to resonance of the TE_{02} and TE_{03} modes.

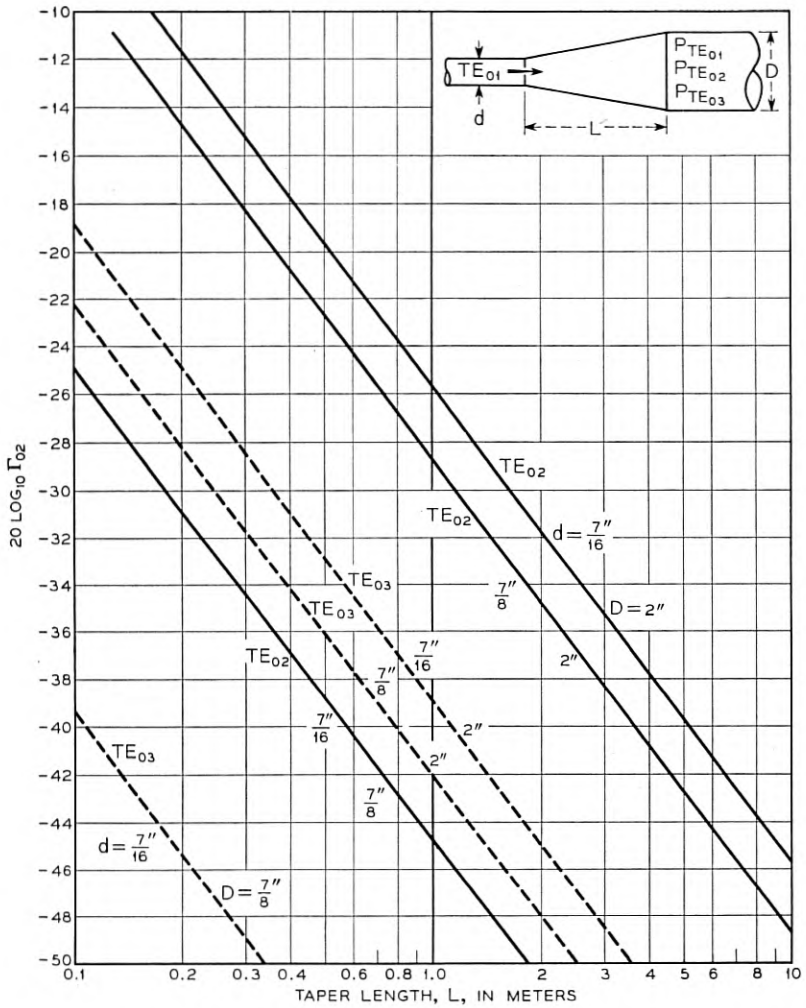


Fig. 5 — Mode conversion of TE_{02} and TE_{03} relative to TE_{01} generated by a conical taper.

The relative levels of TE_{02} and TE_{03} conversions, which have been calculated from unpublished work of S. P. Morgan are shown plotted in Fig. 5 for the waveguide sizes employed in the millimeter wavelength band. The conversions, 20 $\log_{10} \Gamma_{02}$, are plotted in terms of the TE_{02} and TE_{03} powers relative to the TE_{01} mode power and are expressed in db as a function of the taper length L, in meters.

Fig. 6 shows the theoretical and experimental values obtained for TE_{01} relative insertion loss. Since the minimum length of pipe tested is

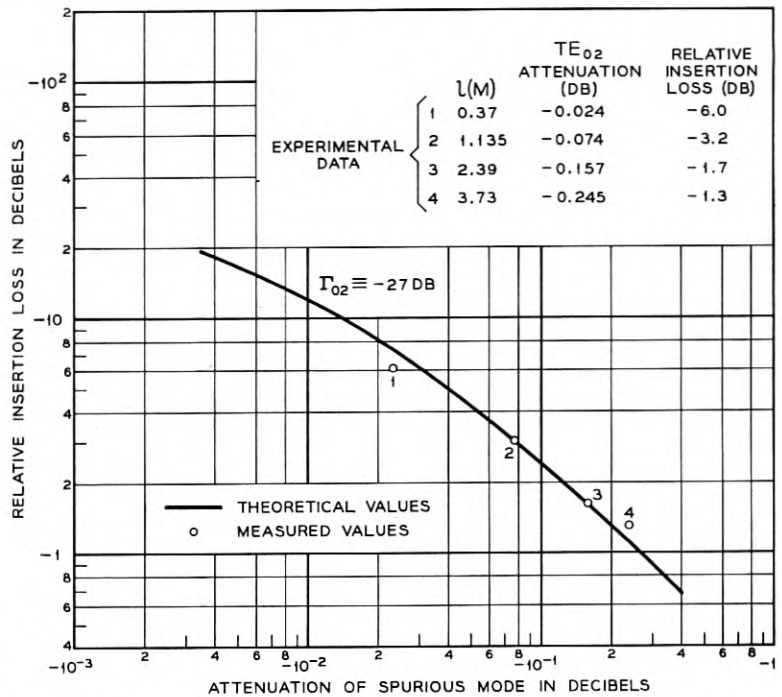


Fig. 6 Theoretical and measured relative insertion loss in the TE_{01} transmission system of Fig. 4.

several times the length of the tapers, the losses in the transitions are fairly small compared to the losses in the multimode guide and this justifies assumption (5). The resonance due to the other modes is too small to be appreciable. This is understandable since, according to (10), the value of the mode conversion for the TE_{03} (Fig. 5) and the attenuation for the shortest length of pipe tested, the calculated relative insertion loss is less than -0.1 db.

CONCLUSIONS

The resonance of spurious modes in a closed environment can produce a large insertion loss of the transmitting mode. In a fairly narrow band device it is possible to avoid this problem by selecting a proper waveguide size for the closed environment. In a broad-band system the losses can be minimized by providing a high attenuation and a low mode conversion for the spurious mode. For example, it may be noted, by referring to Fig. 3, that mode conversion as high as -20 db with a spurious mode loss of -8 db results in only an -0.1 db insertion loss for the transmitting mode.

Measurement of Atmospheric Attenuation at Millimeter Wavelengths

By A. B. CRAWFORD and D. C. HOGG

(Manuscript received September 20, 1955)

A frequency-modulation radar technique especially suited to measurement of atmospheric attenuation at millimeter wavelengths is described. This two-way transmission method employs a single klystron, a single antenna and a set of spaced corner reflectors whose relative reflecting properties are known. Since the method does not depend on measurements of absolute antenna gains and power levels, absorption data can be obtained more readily and with greater accuracy than by the usual one-way transmission methods.

Application of the method is demonstrated by measurements in the 5-mm to 6-mm wave band. The results have made it possible to assign an accurate value for the line-breadth constant of oxygen at atmospheric pressure; the constant appropriate to the measurements lies between 600 and 800 MCS per atmosphere.

INTRODUCTION

It is well known that certain bands in the microwave region are attenuated considerably due to absorption by water vapour and oxygen in the atmosphere. A theory of absorption for both gases was given by Van Vleck.[1] Numerous measurements have been made on the gases when confined to waveguides or cavities[2] and several when unconfined in the free atmosphere.[3] Nevertheless, there is some uncertainty regarding the line-breadth constants which should be used in calculating water vapour and oxygen absorption. In particular, at atmospheric pressure there is doubt as to the amount of absorption on the skirts of the bands where the absorption is small. The present work was undertaken to test a new method of measurement and to improve the accuracy of experimental data measured in the free atmosphere.

The method of measurement is one of comparison of reflections from

spaced corner reflectors whose relative reflecting properties are known. The free-space attenuation is readily calculated and any measured attenuation in excess of this represents absorption by the atmospheric gases.

A description of the method and the apparatus is followed by a discussion of data taken in the wavelength range 5.1 to 6.1 mm (which includes the long wavelength skirt of the oxygen absorption band centered at 5 mm). These data, when compared with the theory,[1] indicate that the line-broadening constant of oxygen at atmospheric pressure is of the order of 600 mc. Some rain and fog attenuation measurements at a wavelength of 6.0 mm are included.

METHOD

The experimental setup is shown in Fig. 1. It consists of a high-gain antenna for both transmitting and receiving and a pair of spaced corner reflectors. Corner reflectors can be built to have good mechanical and electrical stability, and their reflecting properties are relatively insensitive to slight misalignments. The reflectors are mounted well above the ground to ensure free-space propagation conditions.

At the outset, the relative reflecting properties of the corner reflectors are measured by placing them side by side at a convenient distance (d_1 for example) from the antenna. By alternately covering one and the other with absorbent non-reflecting material and measuring the reflected signals, the relative effective areas are determined. The reflectors are then separated as shown and consecutive measurements are made of the signals returned from each reflector. From these measurements, knowing the distances d_1 and d_2 and the calibration of the reflectors, one determines the attenuation over the path d_2-d_1 in excess of the free-space attenuation.* This excess, in the absence of condensed water in the air, represents absorption by the atmosphere.

---

* The power received from the reflector at distance d_1 is,

$$P_1 = P_T \frac{A^2 A_1^2}{\lambda^4 d_1^4} Q(\lambda, d_1)$$

where A and A_1 are the effective areas of the antenna and corner-reflector respectively, and P_T is the transmitted power; $Q(\lambda, d_1)$ is a loss factor which accounts for atmospheric absorption. A similar relation holds for the power received from the reflector at distance d_2. The ratio of the received powers is then,

$$\frac{P_1}{P_2} = \left(\frac{A_1}{A_2}\right)^2 \left(\frac{d_2}{d_1}\right)^4 Q[\lambda, (d_2 - d_1)]$$

The accuracy of the measurements will be affected, of course, by spurious reflections in the neighborhood of the corner-reflectors. The sites for the experiment were chosen to minimize such reflections and checks were made by observing the decrease in the return signals when the corner-reflectors were covered by absorbent material. In all cases, the background reflections were at least 30 db below the signal from the corner-reflector.

The method of measuring the reflected signals is illustrated in Fig. 2. The transmitted signal is frequency modulated in a saw tooth manner with a small total frequency excursion, F. The signal reflected from the near corner-reflector is delayed with respect to the transmitted signal by a time, τ_1, equal to twice the distance to the reflector divided by the velocity of light. During a portion, $T_1 - \tau_1$, of the sawtooth cycle, there is a constant frequency difference, f, between the transmitted and received signals, $(f/F = \tau_1/T_1)$. Power at this frequency is produced by mixing the initial source signal with the delayed received signal and amplifying the difference frequency in a narrow-band amplifier centered at frequency f. The output of this amplifier is, therefore, a pulse at frequency f, of length $T_1 - \tau_1$ and repetition rate $1/T_1$.

To measure the signal returned from the far corner-reflector it is necessary merely to increase the period of the sawtooth modulation proportionate to the increase in distance. The frequency excursion, F, remains the same; hence the average power output of the transmitter is unchanged. As may be seen in Fig. 2, the frequency difference, f, between the transmitted and received signals is unchanged; thus the same amplifier and output meter can be used for the two cases. Another advantage in changing only the sawtooth repetition rate is that the delay is the same fraction of a period in both cases; therefore the duty cycle is unchanged and the intermediate frequency pulses can be detected by either an average or a peak measuring device.

Since the beat frequency, f, is not affected by slow changes in the fre-

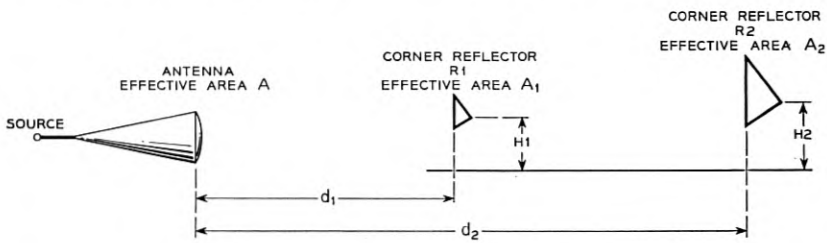


Fig. 1 — Siting arrangement for the atmospheric absorption measurements.

quency of the transmitter, the bandwidth of the intermediate frequency amplifier need be only wide enough to take care of non-linearity in the sawtooth modulation. A signal-to-noise advantage is obtained by the use of the narrow-band amplifier.

Table I gives the distances, heights and effective areas of the reflectors as well as the sawtooth repetition rates that were used in the experiment. The frequency excursion of the sawtooth modulation was 5.8 mc.

It will be noted that three reflectors were used; this was done to provide a long path (comparison of reflections from $R1$ and $R3$) for wavelengths at which the absorption was relatively low, and a short path (comparison of $R1$ and $R2$) for wavelengths at which the absorption was high. The small reflector, $R1$, was one foot on a side; the large reflectors, $R2$ and $R3$, were about 5.6 feet on a side. Fig. 3 is a set of side-by-side measurements showing the reflecting properties of the large reflectors relative to the small one for the wavelengths at which they were used.

APPARATUS

A schematic diagram of the waveguide and electronic apparatus is shown in Fig. 4; Fig. 5 is a photograph of the waveguide equipment so mounted that it moves as a unit with the horn antenna. The antenna is adjusted in azimuth and elevation by means of the milling vise at the bottom of the photograph. The box at the left contains the transmitting tube, a low voltage reflex klystron* which has an average power output of about 12 milliwatts over its 5.1- to 6.1-mm tuning range. About 2 milliwatts of the klystron output is fed through a 6-db directional coupler to a balanced converter that contains two wafer-type millimeter rectifier units.† The remainder of the power proceeds into a 3-db coupler which

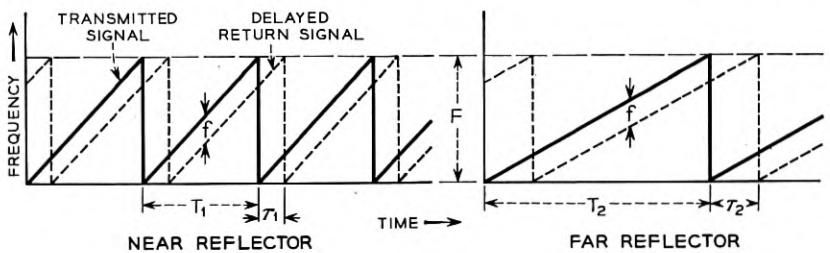


Fig. 2 — Transmitted and reflected frequency-modulated signals.

* This klystron was developed by E. D. Reed, Electron Tube Development Department, Murray Hill Laboratory.

† These millimeter-wave rectifiers were developed by W. M. Sharpless, Radio Research Department, at the Holmdel Laboratory.

TABLE I

Reflector	Distance	Height	Effective Area (Average)	Sawtooth Rep. Rate	Intermediate Frequency-f
	km	*m*	m^2	*kc*	*kc*
R1	$d_1 = 0.59$	6.7	0.05	33	750
R2	$d_2 = 1.36$	21.5	0.67	14.4	750
R3	$d_3 = 2.87$	75	0.79	6.8	750

has the antenna on one arm and an impedance composed of an adjustable attenuator and shorting plunger on another arm. This impedance is adjusted to balance out reflections from the antenna so that a negligible amount of the power flowing toward the antenna enters the converter which is on the remaining arm of the coupler. The delayed energy that re-enters the antenna after reflection from a corner reflector passes through the 3-db coupler to the converter.

The intermediate frequency amplifier shown in Fig. 4 operates with a bandwidth of 300 kc centered at $f=750$ kc. The output of the amplifier is fed to a square law detector and meter for accurate measurement and to an oscilloscope for checking operation of the equipment. Oscillograms of the pulses obtained from the three corner reflectors are shown in Fig. 6; these are all on the same time scale. The gap between the pulses is the delay, τ, shown schematically in Fig. 2.

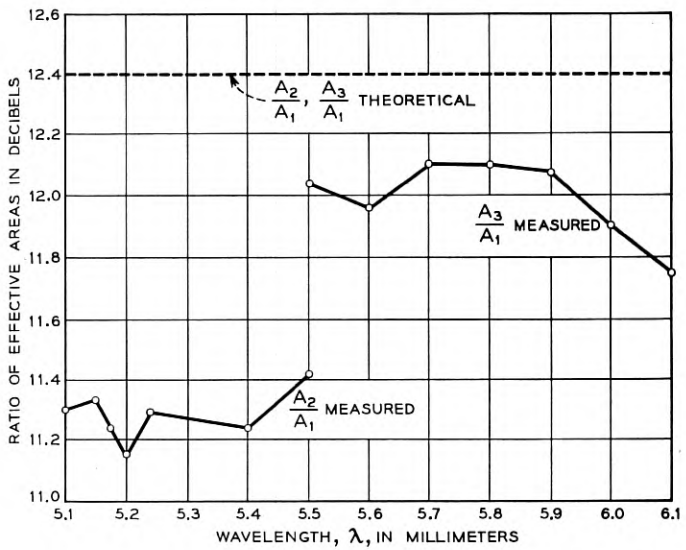


Fig. 3 — Calibration of corner-reflectors R2 and R3 using R1 as a standard.

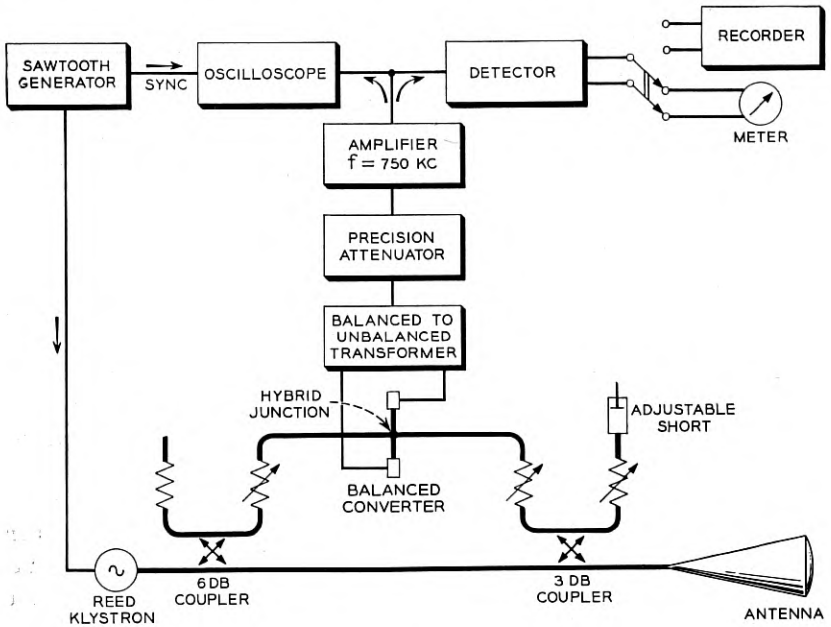


Fig. 4 — Schematic diagram of frequency-modulation radar.

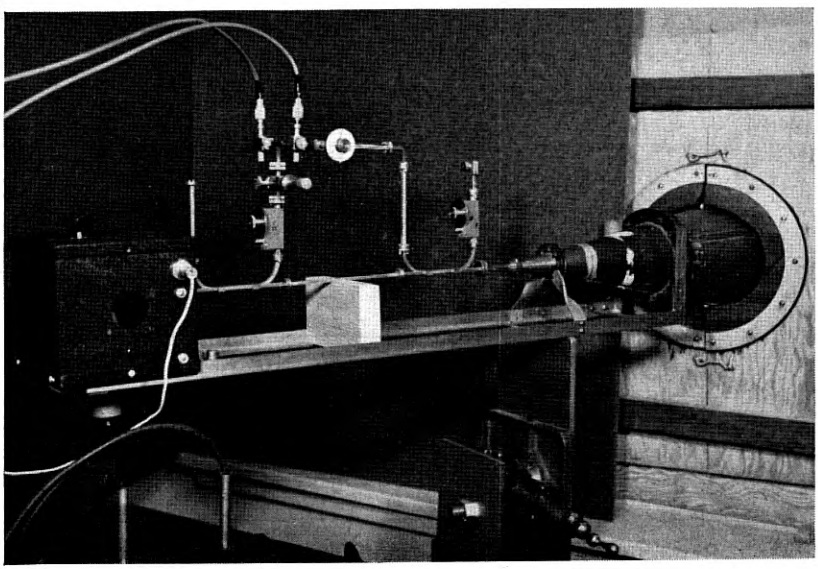


Fig. 5 — Waveguide apparatus and antenna.

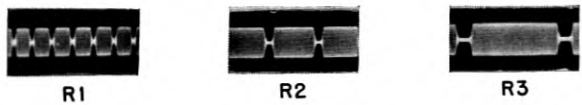


Fig. 6 — 750-kc pulses corresponding to the data in Table I.

Fig. 7 shows the conical horn-lens antenna supported by two bearings to allow adjustment of azimuth and elevation angles. The aperture of the antenna is fitted with a polyethylene lens 30 inches in diameter. The antenna has a gain of about 51 db and a beam width of about 0.5 degrees in the middle of the 5- to 6-mm wave band. This narrow beam, together with well-elevated reflectors, essentially eliminated ground reflections from the measurements.

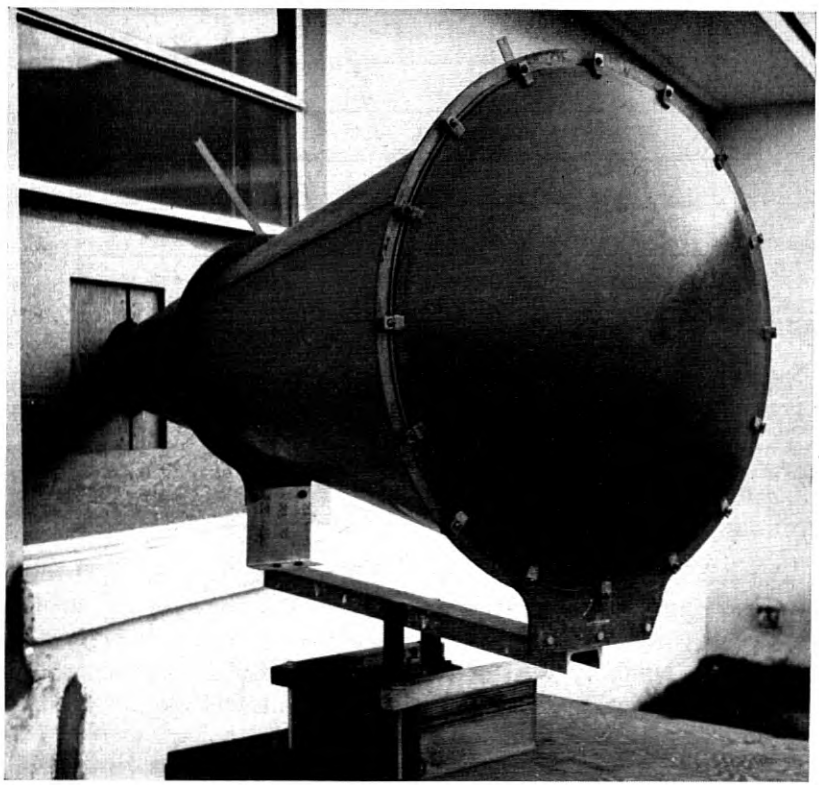


Fig. 7 — Conical horn-lens antenna and mount.

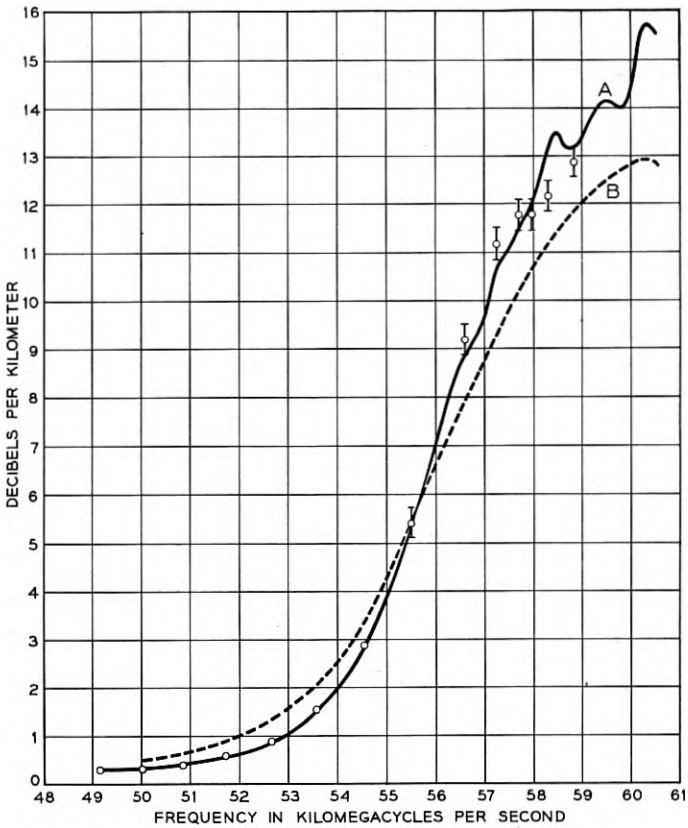


Fig. 8 — Calculated and measured absorption by air at sea level. The dots represent the experimental data; the vertical lines indicate the spread in the measured values. Curves A and B are calculated curves of oxygen absorption using line-breadth constants of 600 and 1200 mc, respectively, and a temperature of 293° K. (Courtesy of T. F. Rogers, Air Force Cambridge Research Center.)

RESULTS

The data to be discussed are shown in Fig. 8; they were taken at Holmdel, N. J., during the months of December, 1954, and January, 1955, on days when the temperature was between 25 and 40 degrees Fahrenheit; the absolute humidity was less than 5 grams/meter^3 during the measurements. It is believed, therefore, that the resonance of the oxygen molecule is the main contributor to the absorption.

The spread in the measurements is indicated by vertical lines through the average values. Each point represents an average of six or more measurements taken on different days. In the range 49 to 54.5 kmc, (5.5 to

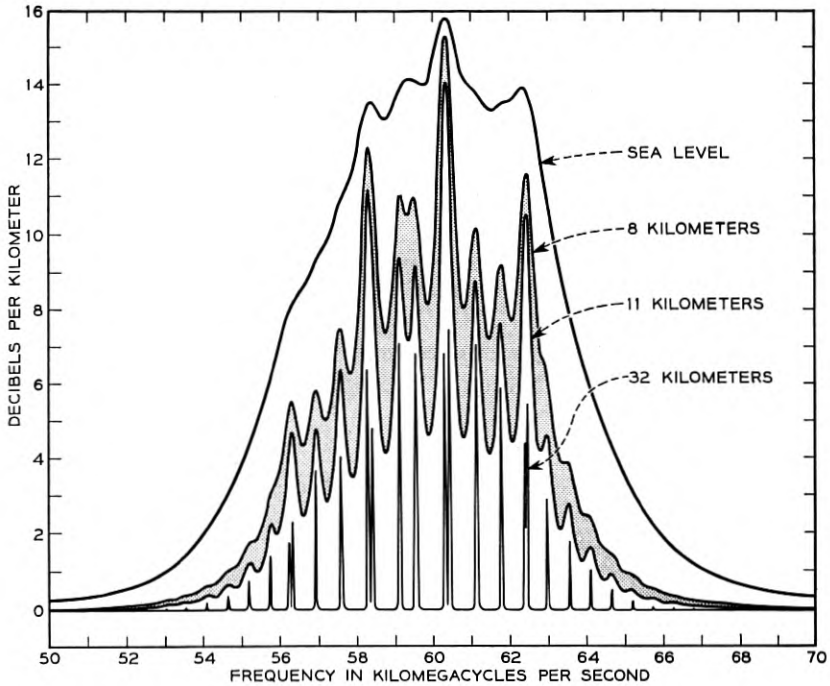


Fig. 9 — Calculated curves of oxygen absorption at various altitudes for a line-breadth constant of 600 megacycles and a temperature of 293° K. (Courtesy of T. F. Rogers, Air Force Cambridge Research Center.)

6.1 mm) the measurements were highly consistent, due mainly to the longer path that was used. Errors in the absolute values of the absorption are estimated not to exceed ±0.05 db/km in the 49 to 54.5 kmc region, ±0.25 db/km in the 55.5 to 59 kmc region. The errors in absolute absorption are governed mainly by the structural and thermo-mechanical stability of the corner reflectors.

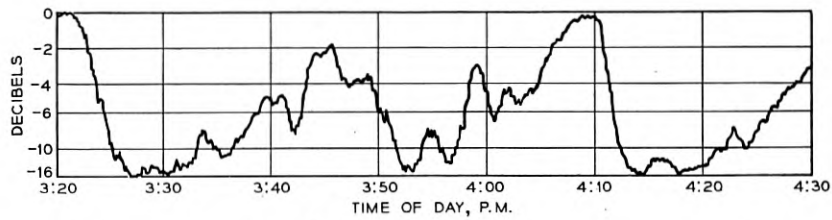


Fig. 10 — Attenuation of 6.0-mm radiation caused by a light rain.
Round-trip path length = 2.72 kilometers
Average rainfall rate = 5 millimeters per hour

TABLE II

Approximate Optical Visibility (miles)	Attenuation due to Land Fog DB/KM
1½	0.06
½	0.13
¼	0.22

In Fig. 8, measured values are compared with the theory of Van Vleck as calculated by T. F. Rogers using line-breadth constants of 600 mc and 1200 mc per atmosphere. The fit with the 600-mc curve is good from 49 to 55.5 kmc, but discrepancies are evident between 56.5 and 59 kmc. For completeness, Rogers' calculations for the absorption at higher altitudes are reproduced in Fig. 9.

A few continuous recordings of rain attenuation have been made at a wavelength of 6.0 mm; a record taken during a light rain is shown in Fig. 10. The median value of the signal is −6.7 db which corresponds to an attenuation of 2.5 db/km for this 5 mm per hour rainfall. During more intensive rainfalls, short-term attenuations in excess of 25 db/km have been observed.

On one occasion, it was possible to measure attenuation by land fog. The measurements given in Table II were made at a wavelength of 6.0 mm. No information regarding water content or drop size was available for this fog.

CONCLUSION

A frequency-modulation, two-way transmission technique has proven reliable for measurement of atmospheric attenuation at millimeter wavelengths. Prerequisite to the success of the method are corner reflectors with good mechanical, thermal and electrical stability.

The frequency-modulation method has been demonstrated by absorption measurements in the free atmosphere in the 5.1- to 6.1-mm band. The data thus obtained are in good agreement with Van Vleck's theory of oxygen absorption; the line-breadth constant appropriate to the measurements lies between 600 and 800 mc per atmosphere.

REFERENCES

1. J. H. Van Vleck, Phys. Rev., **71**, pp. 413 ff, 1947.
2. R. Beringer, Phys. Rev., **70**, p. 53, 1946. R. S. Anderson, W. V. Smith and W. Gordy, Phys. Rev. **87**, p. 561, 1952. J. O. Artman and J. P. Gordon, Phys. Rev., **96**, p. 1237, 1954.
3. R. H. Dicke, R. Beringer, R. L. Kyhl, A. B. Vane, Phys. Rev., **70**, p. 340, 1946. G. E. Mueller, Proc. I.R.E., **34**, p. 181, 1946. H. R. Lamont, Phys. Rev., **74**, p. 353, 1948.

A New Interpretation of Information Rate

By J. L. KELLY, JR.

(Manuscript received March 21, 1956)

If the input symbols to a communication channel represent the outcomes of a chance event on which bets are available at odds consistent with their probabilities (i.e., "fair" odds), a gambler can use the knowledge given him by the received symbols to cause his money to grow exponentially. The maximum exponential rate of growth of the gambler's capital is equal to the rate of transmission of information over the channel. This result is generalized to include the case of arbitrary odds.

Thus we find a situation in which the transmission rate is significant even though no coding is contemplated. Previously this quantity was given significance only by a theorem of Shannon's which asserted that, with suitable encoding, binary digits could be transmitted over the channel at this rate with an arbitrarily small probability of error.

INTRODUCTION

Shannon defines the rate of transmission over a noisy communication channel in terms of various probabilities.[1] This definition is given significance by a theorem which asserts that binary digits may be encoded and transmitted over the channel at this rate with arbitrarily small probability of error. Many workers in the field of communication theory have felt a desire to attach significance to the rate of transmission in cases where no coding was contemplated. Some have even proceeded on the assumption that such a significance did, in fact, exist. For example, in systems where no coding was desirable or even possible (such as radar), detectors have been designed by the criterion of maximum transmission rate or, what is the same thing, minimum equivocation. Without further analysis such a procedure is unjustified.

The problem then remains of attaching a value measure to a communi-

[1] C. E. Shannon, A Mathematical Theory of Communication, B.S.T.J., **27**, pp. 379–423, 623–656, Oct., 1948.

cation system in which errors are being made at a non-negligible rate, i.e., where optimum coding is not being used. In its most general formulation this problem seems to have but one solution. A cost function must be defined on pairs of symbols which tell how bad it is to receive a certain symbol when a specified signal is transmitted. Furthermore, this cost function must be such that its expected value has significance, i.e., a system must be preferable to another if its average cost is less. The utility theory of Von Neumann[2] shows us one way to obtain such a cost function. Generally this cost function would depend on things external to the system and not on the probabilities which describe the system, so that its average value could not be identified with the rate as defined by Shannon.

The cost function approach is, of course, not limited to studies of communication systems, but can actually be used to analyze nearly any branch of human endeavor. The author believes that it is too general to shed any light on the specific problems of communication theory. The distinguishing feature of a communication system is that the ultimate receiver (thought of here as a person) is in a position to profit from any knowledge of the input symbols or even from a better estimate of their probabilities. A cost function, if it is supposed to apply to a communication system, must somehow reflect this feature. The point here is that an arbitrary combination of a statistical transducer (i.e., a channel) and a cost function does not necessarily constitute a communication system. In fact (not knowing the exact definition of a communication system on which the above statements are tacitly based) the author would not know how to test such an arbitrary combination to see if it were a communication system.

What can be done, however, is to take some real-life situation which seems to possess the essential features of a communication problem, and to analyze it without the introduction of an arbitrary cost function. The situation which will be chosen here is one in which a gambler uses knowledge of the received symbols of a communication channel in order to make profitable bets on the transmitted symbols.

THE GAMBLER WITH A PRIVATE WIRE

Let us consider a communication channel which is used to transmit the results of a chance situation before those results become common knowledge, so that a gambler may still place bets at the original odds. Consider first the case of a noiseless binary channel, which might be

[2] Von Neumann and Morgenstein, Theory of Games and Economic Behavior, Princeton Univ. Press, 2nd Edition, 1947.

used, for example, to transmit the results of a series of baseball games between two equally matched teams. The gambler could obtain even money bets even though he already knew the result of each game. The amount of money he could make would depend only on how much he chose to bet. How much would he bet? Probably all he had since he would win with certainty. In this case his capital would grow exponentially and after N bets he would have 2^N times his original bankroll. This exponential growth of capital is not uncommon in economics. In fact, if the binary digits in the above channel were arriving at the rate of one per week, the sequence of bets would have the value of an investment paying 100 per cent interest per week compounded weekly. We will make use of a quantity G called the exponential rate of growth of the gambler's capital, where

$$G = \operatorname*{Lim}_{N\to\infty} \frac{1}{N} \log \frac{V_N}{V_0}$$

where V_N is the gambler's capital after N bets, V_0 is his starting capital, and the logarithm is to the base two. In the above example $G = 1$.

Consider the case now of a noisy binary channel, where each transmitted symbol has probability, p, or error and q of correct transmission. Now the gambler could still bet his entire capital each time, and, in fact, this would maximize the expected value of his capital, $\langle V_N \rangle$, which in this case would be given by

$$\langle V_N \rangle = (2q)^N V_0$$

This would be little comfort, however, since when N was large he would probably be broke and, in fact, would be broke with probability one if he continued indefinitely. Let us, instead, assume that he bets a fraction, ℓ, of his capital each time. Then

$$V_N = (1 + \ell)^W (1 - \ell)^L V_0$$

where W and L are the number of wins and losses in the N bets. Then

$$\begin{aligned} G &= \operatorname*{Lim}_{N\to\infty} \left[\frac{W}{N} \log (1 + \ell) + \frac{L}{N} \log (1 - \ell) \right] \\ &= q \log (1 + \ell) + p \log (1 - \ell) \text{ with probability one} \end{aligned}$$

Let us maximize G with respect to ℓ. The maximum value with respect to the Y_i of a quantity of the form $Z = \sum X_i \log Y_i$, subject to the constraint $\sum Y_i = Y$, is obtained by putting

$$Y_i = \frac{Y}{X} X_i ,$$

where $X = \sum X_i$. This may be shown directly from the convexity of the logarithm.

Thus we put

$$(1 + \ell) = 2q$$
$$(1 - \ell) = 2p$$

and

$$G_{\max} = 1 + p \log p + q \log q$$
$$= R$$

which is the rate of transmission as defined by Shannon.

One might still argue that the gambler should bet all his money (make $\ell = 1$) in order to maximize his expected win after N times. It is surely true that if the game were to be stopped after N bets the answer to this question would depend on the relative values (to the gambler) of being broke or possessing a fortune. If we compare the fates of two gamblers, however, playing a nonterminating game, the one which uses the value ℓ found above will, with probability one, eventually get ahead and stay ahead of one using any other ℓ. At any rate, we will assume that the gambler will always bet so as to maximize G.

THE GENERAL CASE

Let us now consider the case in which the channel has several input symbols, not necessarily equally likely, which represent the outcome of chance events. We will use the following notation:

- $p(s)$ the probability that the transmitted symbol is the s'th one.
- $p(r/s)$ the conditional probability that the received symbol is the r'th on the hypothesis that the transmitted symbol is the s'th one.
- $p(s, r)$ the joint probability of the s'th transmitted and r'th received symbol.
- $q(r)$ received symbol probability.
- $q(s/r)$ conditional probability of transmitted symbol on hypothesis of received symbol.
- α_s the odds paid on the occurrence of the s'th transmitted symbol, i.e., α_s is the number of dollars returned for a one-dollar bet (including that one dollar).
- $a(s/r)$ the fraction of the gambler's capital that he decides to bet on the occurrence of the s'th transmitted symbol *after* observing the r'th *received* symbol

Only the case of independent transmitted symbols and noise will be considered. We will consider first the case of "fair" odds, i.e.,

$$\alpha_s = \frac{1}{p(s)}$$

In any sort of parimutual betting there is a tendency for the odds to be fair (ignoring the "track take"). To see this first note that if there is no "track take"

$$\sum \frac{1}{\alpha_s} = 1$$

since all the money collected is paid out to the winner. Next note that if

$$\alpha_s > \frac{1}{p(s)}$$

for some s a bettor could insure a profit by making repeated bets on the s^{th} outcome. The extra betting which would result would lower α_s. The same feedback mechanism probably takes place in more complicated betting situations, such as stock market speculation.

There is no loss in generality in assuming that

$$\sum_s a(s/r) = 1$$

i.e., the gambler bets his total capital regardless of the received symbol. Since

$$\sum \frac{1}{\alpha_s} = 1$$

he can effectively hold back money by placing canceling bets. Now

$$V_N = \prod_{r,s} [a(s/r)\alpha_s]^{W_{sr}} V_0$$

where W_{sr} is the number of times that the transmitted symbol is s and the received symbol is r.

$$\begin{aligned} \operatorname{Log} \frac{V_N}{V_0} &= \sum_{rs} W_{sr} \log \alpha_s a(s/r) \\ G = \operatorname*{Lim}_{N\to\infty} \frac{1}{N} \log \frac{V_N}{V_0} &= \sum_{rs} p(s, r) \log \alpha_s a(s/r) \end{aligned} \tag{1}$$

with probability one. Since

$$\alpha_s = \frac{1}{p(s)}$$

here

$$G = \sum_{rs} p(s, r) \log \frac{a(s/r)}{p(s)}$$
$$= \sum_{rs} p(s, r) \log a(s/r) + H(X)$$

where $H(X)$ is the source rate as defined by Shannon. The first term is maximized by putting

$$a(s/r) = \frac{p(s, r)}{\Sigma_k p(k, r)} = \frac{p(s, r)}{q(r)} = q(s/r)$$

Then $G_{\max} = H(X) - H(X/Y)$, which is the rate of transmission defined by Shannon.

WHEN THE ODDS ARE NOT FAIR

Consider the case where there is no track take, i.e.,

$$\sum \frac{1}{\alpha_s} = 1$$

but where α_s is not necessarily

$$\frac{1}{p(s)}$$

It is still permissible to set $\sum_s a(s/r) = 1$ since the gambler can effectively hold back any amount of money by betting it in proportion to the $1/\alpha_s$. Equation (1) now can be written

$$G = \sum_{rs} p(s, r) \log a(s/r) + \sum_s p(s) \log \alpha_s .$$

G is still maximized by placing $a(s/r) = q(s/r)$ and

$$G_{\max} = -H(X/Y) + \sum_s p(s) \log \alpha_s$$
$$= H(\alpha) - H(X/Y)$$

where

$$H(\alpha) = \sum_s p(s) \log \alpha_s$$

Several interesting facts emerge here

(a) In this case G is maximized as before by putting $a(s/r) = q(s/r)$. That is, *the gambler ignores the posted odds* in placing his bets!

(b) Since the minimum value of $H(\alpha)$ subject to

$$\sum_s \frac{1}{\alpha_s} = 1$$

obtains when

$$\alpha_s = \frac{1}{p(s)}$$

and $H(X) = H(\alpha)$, any deviation from fair odds helps the gambler.

(c) Since the gambler's exponential gain would be $H(\alpha) - H(X)$ if he had no inside information, we can interpret $R = H(X) - H(X/Y)$ as the increase of $G_{\max}$ due to the communication channel. When there is no channel, i.e., $H(X/Y) = H(X)$, $G_{\max}$ is *minimized* (at zero) by setting

$$\alpha_s = \frac{1}{p_s}$$

This gives further meaning to the concept "fair odds."

WHEN THERE IS A "TRACK TAKE"

In the case there is a "track take" the situation is more complicated. It can no longer be assumed that $\sum_s a(s/r) = 1$. The gambler cannot make canceling bets since he loses a percentage to the track. Let $b_r = 1 - \sum_s a(s/r)$, i.e., the fraction not bet when the received symbol is the r^{th} one. Then the quantity to be maximized is

$$G = \sum_{rs} p(s, r) \log [b_r + \alpha_s a(s/r)], \tag{2}$$

subject to the constraints

$$b_r + \sum_s a(s/r) = 1.$$

In maximizing (2) it is sufficient to maximize the terms involving a particular value of r and to do this separately for each value of r since both in (2) and in the associated constraints, terms involving different r's are independent. That is, we must maximize terms of the type

$$G_r = q(r) \sum_s q(s/r) \log [b_r + \alpha_s a(s/r)]$$

subject to the constraint

$$b_r + \sum_s a(s/r) = 1$$

Actually, each of these terms is the same form as that of the gambler's exponential gain where there is no channel

$$G = \sum_s p(s) \log [b + \alpha_s a(s)]. \tag{3}$$

We will maximize (3) and interpret the results either as a typical term in the general problem or as the total exponential gain in the case of no communication channel. Let us designate by λ the set of indices, s, for which $a(s) > 0$, and by λ' the set for which $a(s) = 0$. Now at the desired maximum

$$\frac{\partial G}{\partial a(s)} = \frac{p(s)\alpha_s}{b + a(s)\alpha_s} \log e = k \qquad \text{for } s\varepsilon\lambda$$

$$\frac{\partial G}{\partial b} = \sum_s \frac{p(s)}{b + a(s)\alpha_s} \log e = k$$

$$\frac{\partial G}{\partial a(s)} = \frac{p(s)\alpha_s}{b} \log e \leqq k \qquad \text{for } s\varepsilon\lambda'$$

where k is a constant. The equations yield

$$k = \log e, \qquad b = \frac{1 - p}{1 - \sigma}$$

$$a(s) = p(s) - \frac{b}{\alpha_s} \qquad \text{for } s\varepsilon\lambda$$

where $p = \sum_\lambda p(s)$, $\sigma = \sum_\lambda (1/\alpha_s)$, and the inequalities yield

$$p(s)\alpha_s \leqq b = \frac{1 - p}{1 - \sigma} \qquad \text{for } s\varepsilon\lambda'$$

We will see that the conditions

$$\sigma < 1$$

$$p(s)\alpha_s > \frac{1 - p}{1 - \sigma} \quad s\varepsilon\lambda$$

$$p(s)\alpha_s \leqq \frac{1 - p}{1 - \sigma} \qquad \text{for } s\varepsilon\lambda'$$

completely determine λ.

If we permute indices so that

$$p(s)\alpha_s \geqq p(s + 1)\alpha_{s+1}$$

then λ must consist of all $s \leqq t$ where t is a positive integer or zero. Consider how the fraction

$$F_t = \frac{1 - p_t}{1 - \sigma_t}$$

varies with t, where

$$p_t = \sum_1^t p(s), \qquad \sigma_t = \sum_1^t \frac{1}{\alpha_s}; \qquad F_0 = 1$$

Now if $p(1)\alpha_r < 1$, F_t increases with t until $\sigma_t \geqq 1$. In this case $t = 0$ satisfies the desired conditions and λ is empty. If $p(1)\alpha_1 > 1$ F_t decreases with t until $p(t + 1)\alpha_{t+1} < F_t$ or $\sigma_t \geqq 1$. If the former occurs, i.e., $p(t + 1)\alpha_{t+1} < F_t$, then $F_{t+1} > F_t$ and the fraction increases until $\sigma_t \geqq 1$. In any case the desired value of t is the one which gives F_t its minimum positive value, or if there is more than one such value of t, the smallest. The maximizing process may be summed up as follows:

(a) Permute indices so that $p(s)\alpha_s \geqq p(s + 1)\alpha_{s+1}$

(b) Set b equal to the minimum positive value of

$$\frac{1 - p_t}{1 - \sigma_t} \qquad \text{where} \qquad p_t = \sum_1^t p\,(s), \qquad \sigma_t = \sum_1^t \frac{1}{\alpha_s}$$

(c) Set $a(s) = p(s) - b/\alpha_s$ or zero, whichever is larger. (The $a(s)$ will sum to $1 - b$.)

The desired maximum G will then be

$$G_{\max} = \sum_1^t p(s) \log p(s)\alpha_s + (1 - p_t) \log \frac{1 - p_t}{1 - \sigma_t}$$

where t is the smallest index which gives

$$\frac{1 - p_t}{1 - \sigma_t}$$

its minimum positive value.

It should be noted that if $p(s)\alpha_s < 1$ for all s no bets are placed, but if the largest $p(s)\alpha_s > 1$ some bets might be made for which $p(s)\alpha_s < 1$, i.e., the expected gain is negative. This violates the criterion of the classical gambler who never bets on such an event.

CONCLUSION

The gambler introduced here follows an essentially different criterion from the classical gambler. At every bet he maximizes the expected value of the logarithm of his capital. The reason has nothing to do with

the value function which he attached to his money, but merely with the fact that it is the logarithm which is additive in repeated bets and to which the law of large numbers applies. Suppose the situation were different; for example, suppose the gambler's wife allowed him to bet one dollar each week but not to reinvest his winnings. He should then maximize his expectation (expected value of capital) on each bet. He would bet all his available capital (one dollar) on the event yielding the highest expectation. With probability one he would get ahead of anyone dividing his money differently.

It should be noted that we have only shown that our gambler's capital will surpass, with probability one, that of any gambler apportioning his money differently from ours but still in a fixed way for each received symbol, independent of time or past events. Theorems remain to be proved showing in what sense, if any, our strategy is superior to others involving $a(s/r)$ which are not constant.

Although the model adopted here is drawn from the real-life situation of gambling it is possible that it could apply to certain other economic situations. The essential requirements for the validity of the theory are the possibility of reinvestment of profits and the ability to control or vary the amount of money invested or bet in different categories. The "channel" of the theory might correspond to a real communication channel or simply to the totality of inside information available to the investor.

Let us summarize briefly the results of this paper. If a gambler places bets on the input symbol to a communication channel and bets his money in the same proportion each time a particular symbol is received his, capital will grow (or shrink) exponentially. If the odds are consistent with the probabilities of occurrence of the transmitted symbols (i.e., equal to their reciprocals), the maximum value of this exponential rate of growth will be equal to the rate of transmission of information. If the odds are not fair, i.e., not consistent with the transmitted symbol probabilities but consistent with some other set of probabilities, the maximum exponential rate of growth will be larger than it would have been with no channel by an amount equal to the rate of transmission of information. In case there is a "track take" similar results are obtained, but the formulae involved are more complex and have less direct information theoretic interpretations.

ACNOWLEDGMENTS

I am indebted to R. E. Graham and C. E. Shannon for their assistance in the preparation of this paper.

Automatic Testing of Transmission and Operational Functions of Intertoll Trunks

By H. H. FELDER, A. J. PASCARELLA and H. F. SHOFFSTALL

(Manuscript received October 19, 1955)

Conditions brought about by nationwide dialing increase intertoll trunk maintenance problems substantially. Under this switching plan with full automatic alternate routing there is a considerable increase in the amount of multiswitched business, and as many as eight intertoll trunks in tandem are permissible. In addition, operator checks of transmission on the connections are lost on most calls. These factors impose more severe limitations on transmission loss variations in the individual trunks and throw on the maintenance forces additional burdens of detecting defects in the distance dialing network.

New methods of analyzing transmission performance to locate the points where maintenance effort will be most effective continue to be studied. The automatic testing arrangements described in this paper enable the maintenance forces to collect over-all transmission loss data quickly and with a minimum of effort. They also facilitate the collection of such data on groups of trunks in a form to make statistical analyses easier. The use of these testing arrangements will permit the maintenance forces to keep a closer watch on intertoll trunk performance and will assist in disclosing trouble patterns.

INTRODUCTION

The advent of nationwide dialing, especially with full automatic alternate routing, has presented additional problems in the maintenance of intertoll trunks. Transmission requirements are more rigorous, the intertoll trunk connections are more complex, and certain irregularities in the performance of the distance dialing network are difficult to detect. Automatic test equipment has been provided to aid and increase the efficiency of over-all testing. This equipment is capable of automatically

testing the operational (signaling and supervisory) functions of dial-type intertoll trunks, and of making two-way transmission loss measurements and a noise check at each end. The test results may be recorded at the originating end by means of a Teletypewriter.

Automatic trunk testing has been used for many years in the local plant for checking the signaling and supervisory features of interoffice trunks. The automatic intertoll trunk testing equipment serves a similar function with respect to these operational features of the intertoll trunks. Because published material is available on automatic operational testing,* these features will not be discussed in detail in this paper; more emphasis is given to the transmission testing features which are new.

MAINTENANCE ARRANGEMENTS FOR INTERTOLL TRUNKS

Except in the very small offices, intertoll trunks usually have a test jack appearance in the toll testboard for maintenance purposes. Cord ended testing equipment in the toll testboard positions enables the attendants to perform various operational tests and to make transmission loss, balance, noise or crosstalk measurements. Facilities are provided for communication with distant offices and with intermediate points where carrier or repeater equipment may be located. Testing of carrier or repeater equipment as individual components or systems is an important aspect of the trunk maintenance problem but is beyond the scope of the present paper.

The maintenance of intertoll trunk net losses close to their specified values is currently a most important transmission problem. Various aspects of the problem are discussed in a companion paper.†

Although the manual testing equipment mentioned above is vital to trunk net loss maintenance, the need for reduction in time and effort required to make measurements has led to the provision of semi-automatic testing arrangements. These arrangements permit a testboard attendant to check transmission in the incoming direction by dialing code 102 over a trunk. The trunk is connected to a source of one milliwatt test power at the far end and a measurement of the received power indicates the net loss. The equivalent of a semi-automatic two-way test may be obtained by making a code 102 test in each direction. If complete information on the test results is desired by one testboard attendant, the attendant at the other end of the trunk must report back his results.

* R. C. Nance, Automatic Intertoll Trunk Testing, Bell Labs. Record, Dec., 1954.

† H. H. Felder and E. N. Little, Intertoll Trunk Net Loss Maintenance Under Operator Distance and Direct Distance Dialing, page 955 of this issue.

In both the manual and semi-automatic methods of measurement, the results must be recorded manually. For statistical analysis of trunk transmission performance in terms of "bias" and "distribution grade", as discussed in the companion paper,* deviations of the measured losses from the respective specified losses must be computed and summarized manually.

The automatic testing equipment described in this paper has been developed as an additional maintenance tool. It will not supplant existing arrangements discussed above but rather is intended to increase the capabilities of plant personnel to do an effective maintenance job. The following features of the equipment contribute particularly to this end:

1. Large numbers of trunks can be tested and the results recorded without the continuous attention of a testboard attendant.
2. The attendant is informed by an alarm whenever the loss of a trunk deviates excessively from the specified value.
3. Computation and summarizing of net loss deviations into class intervals are done automatically, thus facilitating statistical analysis of trunk performance.
4. Data can be collected quickly in large volume for indicating the performance of groups of trunks. Confusion occurring with manual measurements because of changing conditions with time is reduced.
5. Stability of an individual trunk may be checked by a series of repetitive tests.
6. Semi-automatic two-way trunk tests can be made by one attendant when required.

To do an equivalent job entirely by manual methods would require an appreciable increase in the amount of manual test equipment and in the number of test personnel. A comparison of the times required for operational and transmission tests by manual, semi-automatic and automatic methods is shown in Fig. 1. The time shown for the code 102 test does not include coordination time required if information on test results in both directions is required at one end.

GENERAL DESCRIPTION OF AUTOMATIC TESTING EQUIPMENT

Automatic intertoll trunk testing requires automatic equipment at both ends of the trunk.At the originating or control end, an automatic test circuit sets up the test call and controls the various test features. In the distant offices, test lines reached through the switching train provide appropriate automatic test terminations. The automatic equipment for

* H. H. Felder and E. N. Little, Intertoll Trunk Net Loss Maintenance Under Operator Distance and Direct Distance Dialing, page 955 of this issue.

use at the control end, adapted for transmission testing, is presently available only for No. 4 type toll switching offices.

Fig. 2 is a block schematic of the arrangement for automatic intertoll trunk testing, including transmission tests. In the originating No. 4 toll crossbar office an automatic outgoing intertoll trunk test circuit is used which consists of an automatic outgoing intertoll trunk test frame and one or more associated test connector frames. These frames have been provided in all No. 4 type offices and perform the functions of setting up

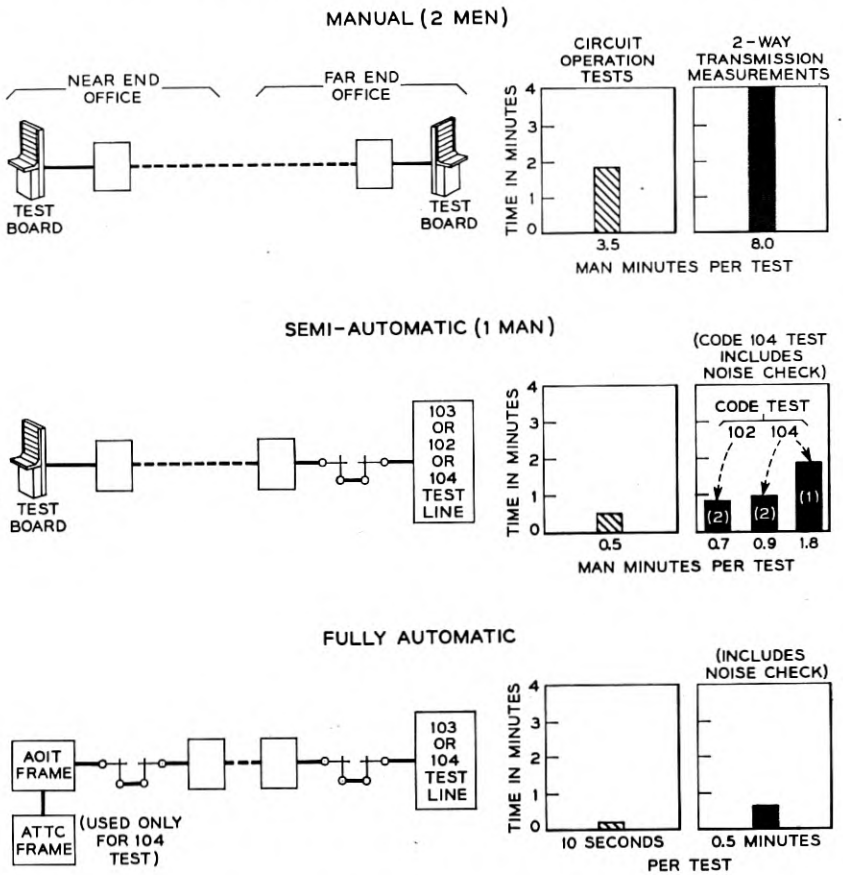


Fig. 1 — Time required for manual tests versus semi-automatic and fully automatic tests. (1) Average time per test for 52,000 field test measurements in 20 offices under normal operating test conditions (includes test preparation, time waiting to be served, testing time, and recording of results). (2) Average time per test for test measurements made in rapid sequence during light load period. The semi-automatic code 104 test includes a noise check at the far end only.

the test call and making operational tests on the intertoll trunks. For automatic transmission tests an automatic transmission test and control circuit, provided in a separate frame as an adjunct to the test frame, is brought into play. A Teletypewriter, mounted in the transmission test frame, is adapted for use with the equipment in the originating office for recording test results. The Teletypewriter is used to make a record of trunks having some defect in their operational features, busy trunks passed over without test and, during transmission tests, to record the results of the transmission measurements. The test frame and associated transmission test and control circuit and Teletypewriter are used principally by the toll test board forces and, therefore, are usually located near the toll test board. Figure 3 shows such an installation.

The intertoll trunk test connector frames in the originating office, not shown in Fig. 3, are frames of crossbar switches and there may be several such frames in a large office. Each crosspoint on the switches of the test connector frames represents an individual intertoll trunk. When a trunk is to be tested, the test frame closes the crosspoint of the test connector switches which serves that particular trunk. This extends the selecting leads (trunk sleeve and select magnet leads) of the trunk to the test frame for use in setting up the call. A class contact on the test connector crosspoint also operates one of several class relays in the test frame when the crosspoint is closed. The function of the class relay is discussed later.

The test frame has an appearance on the incoming link frame of the office switching train. The intertoll trunks to be tested appear on the outgoing link frames of the office switching train. When a trunk is to be tested, the test frame engages the office common control equipment (decoder and marker), through a connector, and requests a path between the test frame appearance on the incoming link frame and the particular intertoll trunk which is to be tested. The common control equipment is able to set up this path since the test frame has closed a test connector cross-point to bring the selecting leads of the trunk to be tested into the test frame. The common control equipment uses the select magnet lead to identify the trunk to be tested and thus is able to set up the path to that particular trunk. The test frame uses the trunk sleeve lead for busy test purposes and for controlling the test call.

In the distant offices separate groups of test lines provide automatic test terminations for operational and transmission tests, respectively. These are reached through the switching train as indicated in Fig. 2. The three digit service code 103 is reserved in toll switching offices for reaching the operational test lines and code 104 is reserved for reaching the transmission test lines. A transmission measuring and noise checking

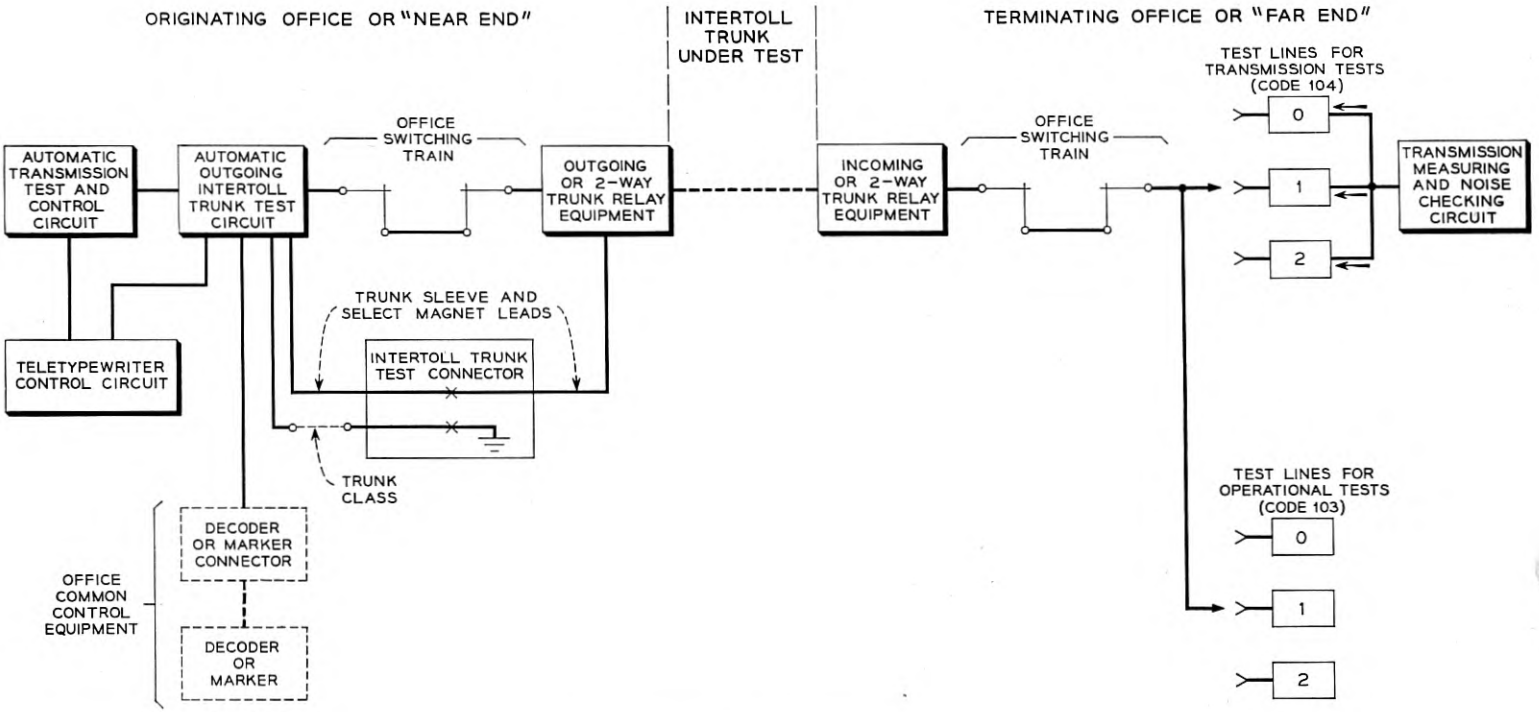


Fig. 2 — Arrangement for automatic tests.

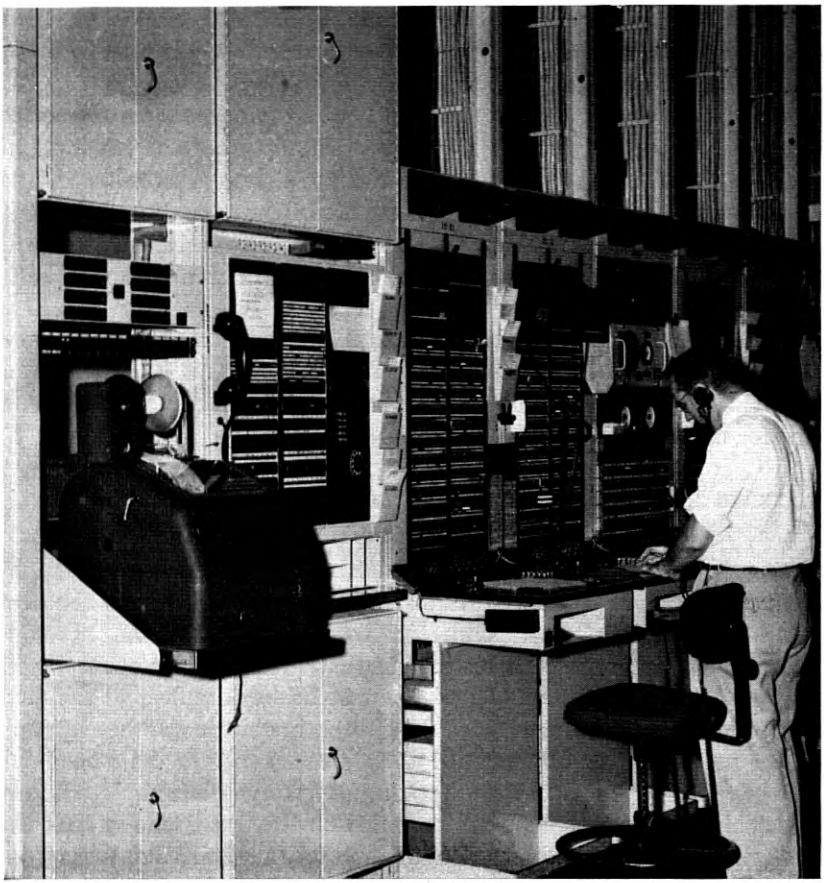


Fig. 3 — Automatic intertoll trunk testing equipment.

circuit, sometimes referred to as "far end equipment," is associated with the group of code 104 test lines for performing the transmission measurements. When simultaneous transmission test calls arrive in the distant office from different originating offices, the calls wait on the code 104 test lines and are served by the transmission measuring and noise checking circuit, one at a time, in their proper turn.

After the trunk test frame has obtained a path through the switching train in the originating office to the trunk to be tested as explained above, it pulses forward over the trunk the desired test line code, either code 103 or code 104. In response to the code, the switching equipment in the distant office sets up a path through the switching train to one of

the code 103 or code 104 test lines. A steady off-hook signal is returned over the established connection to the test frame at the originating office to indicate that the test may proceed.

The process of setting up a test call, as described above, simulates very closely the procedures followed in setting up a normal call from an incoming trunk in the originating office to a desired number in the distant office. Therefore, any irregularities in the operational features while setting up the test call will be detected by the test frame and will result in appropriate trouble indication at the test frame at the originating end.

The automatic testing arrangements will operate with offices where terminating calls are switched either on a terminal net loss (TNL) or on a via net loss (VNL) basis. This is done by including appropriate pads for use at VNL offices.

FEATURES OF INTERTOLL TRUNK TEST FRAME

Control of Test Connector

When the trunk test frame is put into operation, it closes the first crosspoint of the first test connector crossbar switch on the test connector frames to prepare for testing the first trunk in the test sequence. When the trunk test is completed, the test frame advances to the next crosspoint for testing the next trunk. This progression continues through all the test connector frames until all intertoll trunks in the office have been tested. The test frame stops when the test cycle is completed. Particular circuit selection keys are provided on the test frame so that the test connector can be directed manually to any point for testing an individual trunk or for starting a test cycle at some intermediate point in the test sequence, rather than with the first trunk. As the test frame progresses through a test cycle, it also displays on lamps a 4-digit "trunk identification number" corresponding to the test connector crosspoint which is closed. When a trouble is encountered, the attendant uses this 4-digit number to identify the trunk being tested as a particular trunk to a particular destination. The Teletypewriter prints the trunk identification number as a part of each trunk test record.

Busy Test

Before starting a test, the test frame tests the trunk sleeve lead for busy. If the trunk is busy, the test frame waits for the trunk to become idle. A "pass busy" key is provided which, when operated, cancels the waiting period and causes the test frame to immediately pass over busy

trunks to save time. The circuits are arranged so that, if desired, the Teletypewriter may print a record of busy trunks passed over without test. By means of a timing key this record can be delayed two minutes or four minutes to wait for the trunk to become idle. This is used when it is preferable to wait a reasonable time for trunks to become idle to secure tests on a larger proportion of the trunks.

Trunk Classes

A class relay, operated by a contact on the test connector crosspoint as previously mentioned, indicates to the test frame the type of trunk being tested so that it can properly handle the test call. There are 33 of these relays. A flexible cross-connection in the path of the class contact on each test connector crosspoint permits each crosspoint to be assigned to the particular one of the 33 class relays which represent the characteristics of the intertoll trunk associated with that crosspoint. Twenty-eight of the class relays are used in connection with trunks on which automatic transmission tests are made and indicate, among other things, the specified loss of the trunk being tested. These relays are provided in such a manner that, for any trunk, a class can be chosen which agrees with the specified loss of the trunk to within ± 0.1 db over the range 3.8 db to 12.1 db. The specified loss is used by the automatic transmission test and control circuit when computing the deviation of the measured loss from the specified value, as covered later.

Test Cycles

The test frame may be set up by means of control keys to perform various kinds of test cycles. Some of these test cycles are described briefly below.

Code 103 Tests. A complete check is made of all the circuit operating features including the ringing and the supervisory features while the connection is established. If this test is passed successfully, one may assume that the intertoll trunk circuit and the associated signaling channel will properly handle normal calls although this does not prove that the transmission performance is satisfactory.

Signaling Channel Tests. This is an abbreviated code 103 test to verify the integrity of the trunk and its signaling channel. It can be made quite rapidly and is useful for checking the correctness of patching after day and night circuit layout changes.

Pass Idle Test. This is a test cycle which may be run occasionally during a very light load period to detect trunks which may be falsely busy.

Repeat-2 Tests. This consists essentially of two code 103 tests on the same trunk in rapid succession. It is used to insure that a connection through the switching train in the distant office will be properly broken down when a call is completed.

Repeat Test. A "repeat test" key on the test frame cancels the advance of the intertoll trunk test connector. Since the test frame cannot then advance to the next test connector crosspoint, it tests the same trunk repeatedly. This is useful for verifying a trouble condition and for detecting an intermittent trouble, or for obtaining data on stability versus time.

Manual Tests. When the test frame is set up for making manual tests, it engages the office common control equipment to set up a path through the switching train to the trunk to be tested but does not pulse forward code 103 or 104. Instead the attendant can pulse forward the proper code to reach the toll test board at the distant end. This permits dial type trunks to distant offices, not equipped with test lines, to be tested manually.

Code 104 Tests. A 3-position key controls transmission testing. When the key is normal, the test frame makes operational tests. When the key is operated to the transmission and noise position, a two-way transmission loss measurement is made and is followed by a noise check at each end of the trunk. When the key is operated to the transmission only position the noise checks are omitted. The latter position is used when it is permissible to omit the noise checks to save time. When making code 104 tests, the test frame sets up and breaks down the test connection in the same way as when making code 103 tests and, while the connection is established, it receives supervisory signals from the far end. Thus most of the trunk operating features, except for ringing, are also checked as an incidental part of the transmission test. Irregularities in the circuit operating features can result in a trouble indication in the same way as when making operational tests.

Trouble Indications

During a test, progress lamps display the progress of the test call and, when trouble is detected, one of a number of trouble indicating lamps may also be lighted. The progress and trouble indicating lamps indicate the general nature of the trouble.

When the Teletypewriter is not in operation, a trouble indication causes the test frame to stop, to hold the trunk busy and to sound an alarm while awaiting the attention of the attendant. It is the usual practice for the attendant to make a repeat test on the same trunk to

verify the trouble condition. He notes the nature of the trouble from the progress and trouble indicating lamps and then causes the test frame to resume testing by advancing it to the next trunk with a manual advance key.

When the associated Teletypewriter is provided and operating, it is not always necessary to sound an alarm and thus interrupt the regular work of the attendant. Instead the Teletypewriter may print a trouble record. For this purpose troubles are grouped into 18 categories. When a trouble is detected, the Teletypewriter prints a record of the trunk identification number together with a letter in a separate column indicating one of these categories. The test frame then usually makes a repeat test on the same trunk to verify the trouble except when the connection must be held, as discussed later. If the second test is satisfactory, the trouble was of a transient nature and the test frame resumes testing, leaving a single line trouble record on the Teletype tape. If the trouble is still present on the second trial, a second record is printed on the next line for the same trunk.

If the nature of the trouble, as indicated by its category, is such as to render the trunk unfit for service, the test frame will stop after the second trial, hold the trunk busy, and sound an alarm to attract the immediate attention of the attendant. If, however, the trouble is of a minor nature that can be tolerated temporarily, the test frame advances automatically to the next trunk after the second record is printed, and resumes testing without sounding an alarm. By periodic inspection of the Teletype record the attendant can note those trunks needing maintenance attention by means of the double line trouble records. A test cycle can thus be completed with the minimum of supervision on the part of the attendant.

When the nature of a trouble is such that its identity is likely to be lost if the original connection is broken down, e.g., failure of a holding ground, the test frame will not attempt a second trial but stop, hold the trunk busy, and sound an alarm. Failure to complete a transmission test satisfactorily is included in this class because such failures can be due to the testing equipment itself.

AUTOMATIC TRANSMISSION TESTS

Basic Scheme of Measurement

An automatic transmission loss measurement consists essentially of adjusting the loss of a pad at the receiving end of the trunk to bring the test power level at the pad output to a fixed value. A functional diagram of the arrangement is shown in Fig. 4.

The standard one milliwatt source of test power is used at the sending end. The receiving end includes an amplifier, a set of adjustable resistance pads which are relay controlled and an amplifier-rectifier with a measuring relay (M) in its output circuit. Relay (M) is a polarized relay of a type widely used in the telephone plant.

The amplifier has a fixed gain of 19.9 db and it includes considerable negative feedback so that its gain is constant. The pad components are precision resistors to insure accuracy.

The amplifier-rectifier consists of a two-stage amplifier followed by a rectifier tube and a detector tube for controlling relay (M). This circuit is designed so that the margin between the input power which will hold relay (M) operated and the input power which will insure that relay (M) will release is less than 0.1 db. The gain is adjusted, by means of a potentiometer, so that relay (M) will operate when the test power level at the output from the receiving pads in Fig. 4 is one milliwatt or higher and so that it will release when the power level at this point is 0.1 db or more below one milliwatt. This close margin between operate and release permits relay (M) to be used as an accurate measuring device with a precision comparable with that of manual transmission measuring equipment using direct reading meters. Negative feedback, built into the amplifier portion of the amplifier-rectifier, insures gain stability and the amplifier-rectifier will maintain its gain adjustment over a long period.

When making a transmission loss measurement, the power from the sending end operates relay (M) in the amplifier-rectifier. The loss in the receiving pads is then increased, by means of control circuitry, until the power level at their output is reduced to one milliwatt. In making this adjustment, relay (M) is used as the power level indicating device. When this adjustment is finished the trunk loss will be

$$\text{Intertoll Trunk Loss} = 19.9 \text{ db} - \text{Receiving Pad Loss.}$$

Adjustment of Receiving Pads

The receiving pads, shown in Fig. 4 consist of 9 individual pads having losses of 10, 5, 4, 2, 1, 0.5, 0.4, 0.2 and 0.1 db. Each pad is inserted into the input circuit to the amplifier-rectifier by the operation of a corresponding pad control relay. Adjustment of the pad loss takes place in steps.

When relay (M) operates on arrival of the test power, the control circuit operates relay 10 to insert the 10 db pad. If this reduces the test power level at the output from receiving pads to a value below one milli-

watt, relay (M) in the amplifier-rectifier will release. The control circuit then releases relay 10 also to remove the 10 db pad before it proceeds to the next step. If the test power level remains one milliwatt or higher after the 10 db pad is inserted, relay (M) remains operated. The control circuit then locks relay 10 in its operated position to retain the 10 db pad before it proceeds to the next step. In the next step, pad control relay 5 is operated to insert the 5 db receiving pad. The 5 db receiving pad will then be rejected or retained, as described above, depending upon which position relay (M) takes after the 5 db pad is inserted. This process continues until all 9 individual receiving pads have been tried in descending order ending with the 0.1 db pad. When this process is completed, the combination of the 9 pad control relays which remain locked in the operated position determines, additively, the receiving pad loss and consequently, this combination is related directly to the trunk loss. At the originating or control end this combination of operated relays will be translated to the measured loss of the intertoll trunk being tested, when the results of the measurement are recorded. The method of transmitting the measured loss from the far end to the originating end is discussed later.

The transmitting and check pads shown in Fig. 4 are a separate set of pads also controlled by the pad control relays. At the start of the test the total loss in these pads is 19.9 db. Whenever a pad control relay operates to insert a receiving pad, it removes an equal loss from the transmitting pads. Therefore, when the receiving pad adjustment is finished, the loss remaining in the transmitting pads will be equal to the loss of the trunk. Also, the sum of the losses in the two sets of pads is always 19.9 db regardless of the trunk loss being measured, provided all pad components and all pad control relay contacts are in perfect order. This condition permits a precise accuracy check to be made, as discussed later.

Whenever the control circuit leaves pad control relay 4 or 0.4 in its operated position to retain the 4 db or 0.4 db pad, the subsequent 2 db and 1 db or 0.2 db and 0.1 db pad control relays are disabled. There will will then be no action as the control circuit passes through the 2 db and 1 db or the 0.2 db and 0.1 db steps. This limits the maximum receiving pad loss to 19.9 db, which is the maximum range of the automatic measurement. This range amply covers the range of losses of intertoll trunks in a usable condition. Loss measurements attempted outside the range of 0 to 19.9 db will cause failure of the built-in checks, mentioned later, and will result in an alarm at the control end of the trunk.

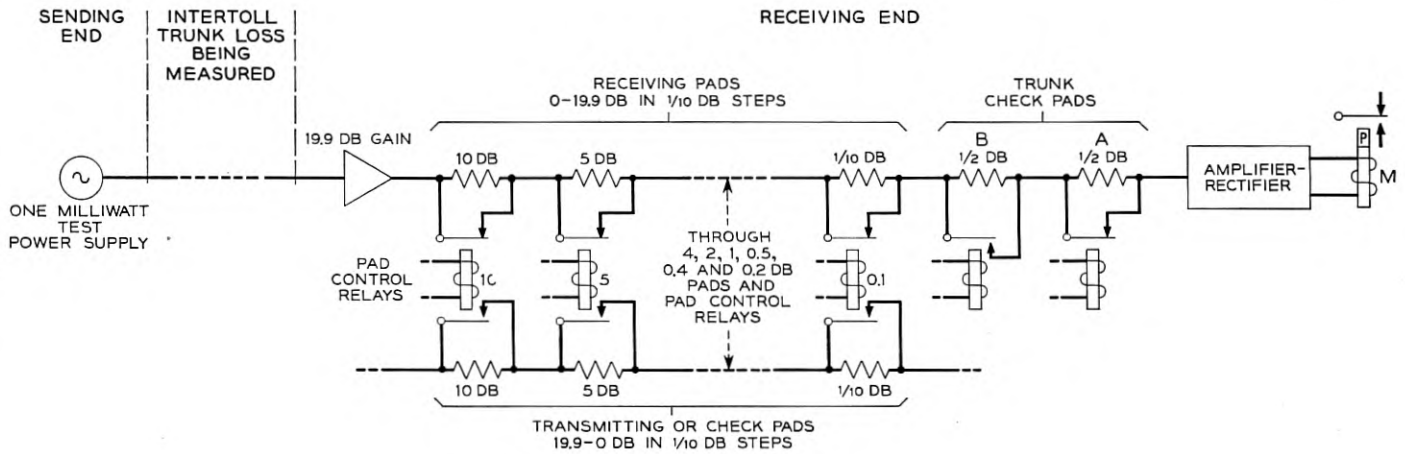


Fig. 4 — Schematic of transmission measurement.

Accuracy Checks

If the receiving pad adjustment has been successful, the power level at the pad output will be very close to one milliwatt and the measuring relay (M) will be just on the verge of moving from its front to its back contact, or vice versa. Errors may creep in, however, to prevent these things from being true. Some such sources of error are:

(1) One or more of the pad control relays might fail to lock in the operated position when they should, or fail to release when they should. It would then be impossible to adjust the total receiving pad loss to the correct value.

(2) The trunk loss might change suddenly while the pad adjustment is in progress and make it impossible, with the pads remaining to be tried, to bring the power level at the pad output to one milliwatt.

(3) The amplifier or amplifier-rectifier gains might increase or decrease due to a defective component.

(4) The milliwatt test power supply might deviate from the standard value.

(5) Defective components or faulty control relay contacts might cause the individual pad losses to be incorrect.

To detect errors of the type in items (1) and (2) a "trunk check" is made immediately after the pad adjustment is finished. Referring to Fig. 4, two 0.5 db pads, A and B, are provided in the input circuit to the amplifier-rectifier, pad A being normally out. Before the sending end removes the test power, pad A is inserted, momentarily. The resulting decrease in input power to the amplifier-rectifier should cause relay (M) to release. Both pads A and B are then cut out. The resulting increase in input power should cause relay (M) to operate. If relay (M) fails to pass either of these checks the receiving pad loss is in error by 0.5 db or more and another trial is needed to secure a more accurate adjustment. Premature removal of the test power at the sending end would, of course, cause relay (M) to fail on the second check and result in another trial.

Immediately after the trunk check and while pad B is still cut out, the receiving end rearranges its circuit locally as shown in Fig. 5 for a "loop check" to guard against errors of the types mentioned in items (3), (4) and (5) above. This rearrangement inserts a 0.3 db pad in place of the 0.5 db pad B, which is cut out. The local milliwatt supply then applies power to the amplifier-rectifier at a level about 0.2 db higher than necessary to operate relay (M). Relay (M) will fail to operate and pass this check if the combined effect of any decrease in the value of the milliwatt test power supply, any decrease in the amplifier and the amplifier-rectifier gains and cumulative errors in the receiving pads and check pads adds more than 0.2 db loss. After the above check, a 0.5 db loss is

added in the looped circuit. This reduces the input power to the amplifier-rectifier about 0.2 db below that which causes relay (M) to release. Relay (M) will remain operated and fail to pass this check if the combined effect of increases in the milliwatt test power supply or amplifier and amplifier-rectifier gains and cumulative errors in the pads exceeds 0.2 db gain. By means of the loop check the maintenance forces will be notified whenever the measuring equipment drifts more than ±0.2 db from the initially calibrated setting.

After the loop check the receiving end restores its circuit to the original connections shown in Fig. 4 and by means of relays not shown, cuts out all of the receiving pad loss. Relay (M) then reoperates. The circuit rests in this condition to await the removal of the test power at the sending end.

When the sending end removes the test power, relay (M) releases. If all accuracy checks have been passed successfully, the receiving end then prepares for the next phase of the test. If, however, the accuracy checks failed in any respect, the receiving end restores its circuit to the original condition at the start of the measurement and returns a signal to the sending end to request reconnection of the test power for another trial.

Intertoll Trunk Loss Measurement and Noise Check

An intertoll trunk loss measurement consists of two successive one-way measurements, as described above, one for each direction of transmission. The transmission test call is set up to one of the code 104 test lines in the distant office. If the transmission measuring and noise checking circuit at the far end is already engaged because another call arrived

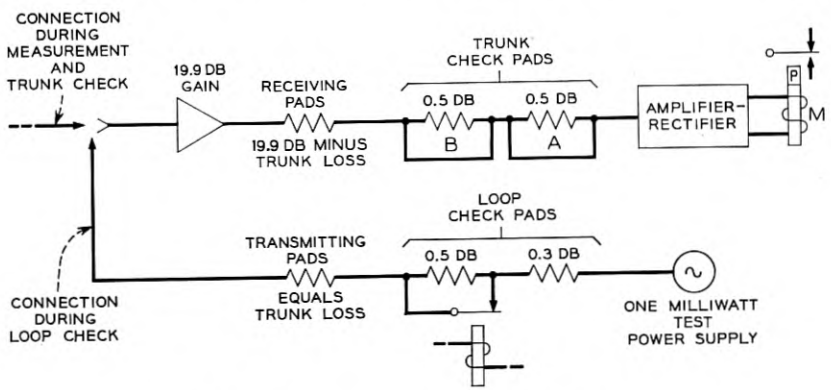


Fig. 5 — Arrangement for loop check.

just previously from some other originating office, this call waits on the test line. When the transmission measuring and noise checking circuit is ready to serve this call, it connects to the test line on which this call is waiting and then returns a steady off-hook signal to the originating end. This notifies the originating end that the transmission test may begin.

The philosophy of a two-way transmission measurement is as follows. The near end sends test power over the trunk and the far end measures the loss as previously described. In this process the loss of the transmitting pad at the far end is adjusted to a value equal to the trunk loss in the near to far direction. The far end then returns test power, first directly over the trunk and next, through the transmitting pad. The power levels received at the near end are a measure of first, the trunk loss in the far to near direction and next, the sum of the losses in the two directions. Measurement of these levels provides data for recording the loss in each direction at the near end.

The two-way transmission measurement takes place in four steps as shown in Fig. 6. These steps are described below.

Step 1

The near end sends one milliwatt and the far end adjusts its pads and checks the measurement. After about 3 seconds the near end removes the test power and then pauses for a short interval to wait for a signal denoting whether or not the accuracy tests were successful.

If they were unsuccessful, the far end will restore itself to the condition prevailing at the start of Step 1 and will also return a short (about ½ second) on-hook signal to the near end. The near end then reconnects the test power for three seconds for another trial. The test frame at the near end stops and sounds an alarm after a third unsuccessful trial.

If the far end is successful in any one of the first three trials, an on-hook signal will not be returned to the near end when the test power is removed. The near end, after the short pause, then sends a short spurt of test power which reoperates the measuring relay at the far end. This signal at the far end, after a successful Step 1, indicates to the far end that this is a full automatic test.

Step 2

For Step 2 the near end connects a far-near amplifier, a set of far-near receiving pads and an amplifier-rectifier. The far end disconnects its receiving equipment and returns one milliwatt over the trunk. The far-near receiving pads at the near end are now inserted in the proper com-

bination to reduce the power level at their output to one milliwatt. The combination of the nine far-near pad control relays remaining operated after the adjustment is finished will be translated later to measured loss of the intertoll trunk in the far to near direction. After about 3 seconds the far end removes the test power and pauses for a short interval. This completes Step 2. Each end then prepares for Step 3.

Step 3

For Step 3 the near end retains the far-near amplifier and the setting of the far-near receiving pads and adds a near-far amplifier and a set of near-far receiving pads in tandem in the input circuit to the amplifier-rectifier. The far end, after the short pause, again sends one milliwatt but this time it sends through the transmitting pad which was adjusted in Step 1 to represent the near-to-far trunk loss. The near-far receiving pads at the near end are now automatically arranged to reduce the power level at their output to one milliwatt.

In the adjustment of Step 2 the over-all loss, including the trunk in the far-to-near direction, the far-near amplifier and the far-near receiving pads, was made 0 db. Consequently, the net loss being measured in Step 3 is simply that of the transmitting pad at the far end, which is the same as the trunk loss in the near-to-far direction. Therefore the combination of the 9 near-far pad control relays remaining operated after Step 3 is finished can be translated to measured loss of the intertoll trunk in the near-to-far direction. After about 3 seconds the far end will remove the test power to complete Step 3 and it will then pause for a short interval before proceeding with Step 4.

At the near end there are also two sets of check pads, not shown, which are associated with the far-near and near-far receiving pads, respectively, as indicated in Fig. 4. During Step 2 and Step 3 the near end makes the trunk check previously described to verify the accuracy of the pad loss settings and, in addition, in Step 3, rearranges its circuit in the manner shown in Fig. 5 for the loop check. Thus at the near end the two sets of check pads, the far-near and near-far amplifiers, and the two sets of receiving pads are all connected in tandem for the loop check.

During the short pause following Step 2 and Step 3 the far end reconnects its amplifier and amplifier-rectifier as shown for Step 1 in Fig. 6. If the near end is unsuccessful in the trunk check in Step 2 or in either the trunk check or loop check in Step 3, it will restore the circuit to the original condition at the beginning of Step 2 and will also send a short spurt of test power to the far end as shown for Step 1 in Fig. 6. This reoperates the measuring relay (M) at the far end momentar-

ily. The far-end then repeats Steps 2 and 3 for another trial. The test frame at the near end will stop and sound an alarm after a third unsuccessful attempt.

If the trunk loss in the near-to-far direction exceeds 10 db, the loss in the transmitting pad at the far end will exceed 10 db. Under this condition the far end will, prior to Step 3, remove 10 db loss from the transmitting pad to increase the test power level on the trunk. This is done to improve the test power level-to-noise ratio and to reduce the error when measuring losses of intertoll trunks having apparatus whose loss is dependent on signal amplitude. The far end will also return to the near end a short on-hook signal. This on-hook signal at the near end, just prior to Step 3, is an "add 10" signal and causes the near end to add 10 db to its loss measurement in Step 3, to compensate for the loss which was removed at the far end.

Immediately after Step 3, if the transmission test control key on the test frame is in the transmission only position, the test frame will cause the teletypewriter to record the results of the measurements and will then break down the connection and advance to the next trunk. If the transmission test control key is in the transmission and noise position, the test frame will wait after Step 3 for each end to complete a noise check in Step 4.

Step 4

For Step 4 the near-end removes its near-far amplifier and the near-far and far-near receiving pads and increases the gain of the amplifier-rectifier for a noise check at the near-end. Likewise, the far-end removes the receiving pads and increases the gain of the amplifier-rectifier for a noise check at the far end. Each end rests in this condition while the amplifier-rectifier at each end integrates the noise voltage over a 5-second interval. If the integrated value of noise voltage at either end exceeds a predetermined value, the amplifier-rectifier at that end will operate measuring relay (M) in its output which causes a high noise condition to be registered at that end. If neither end registers a high noise condition, the test call proceeds to completion without a noise indication being recorded at the near end.

When the transmission measuring and noise checking circuit at the far end completes the noise check, it releases itself from the test line and is then free to serve a new call while the test line returns an on-hook signal to notify the originating end that the test is completed. This will be either a steady on-hook signal if the far end has not registered a high

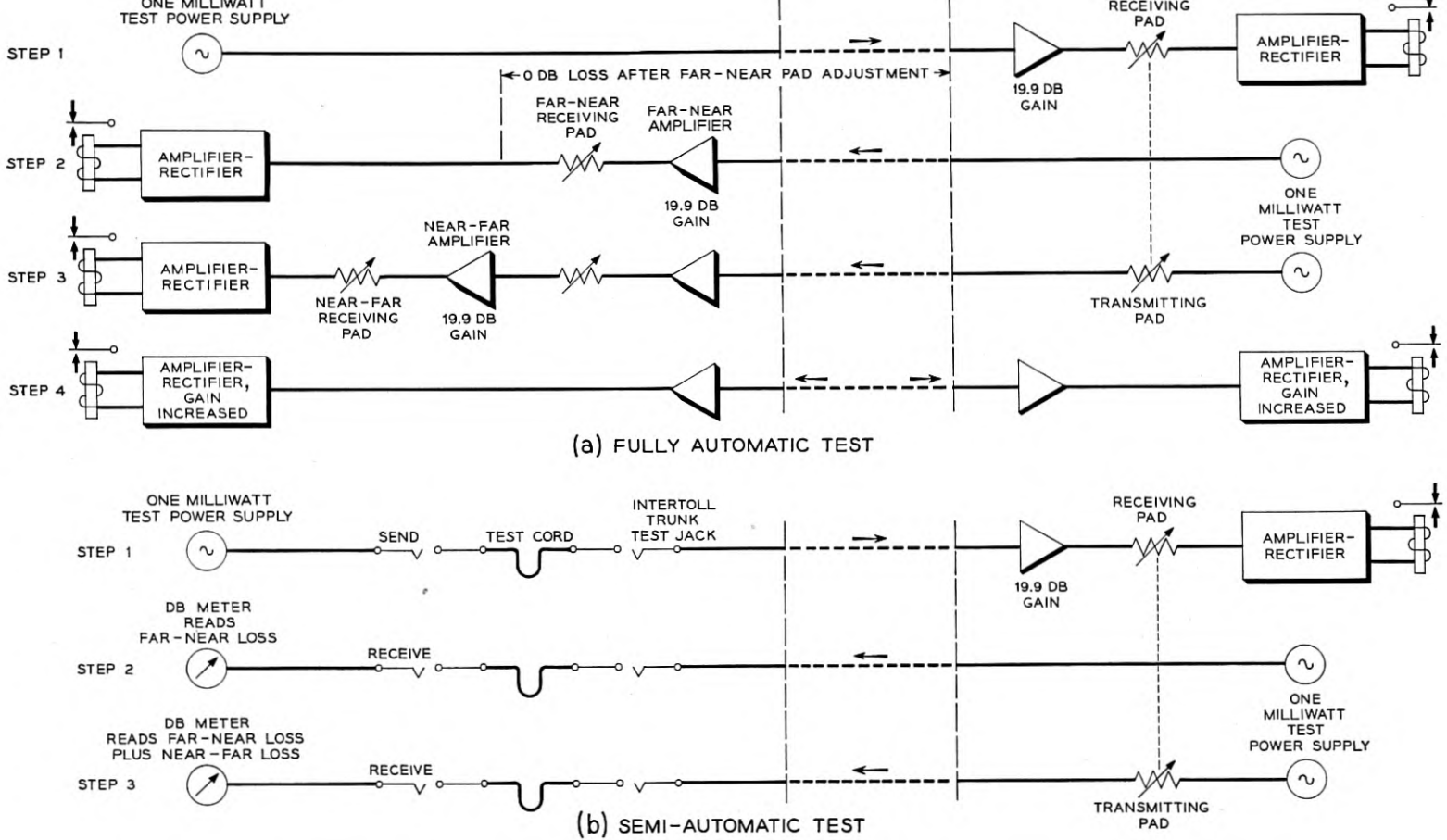


Fig. 6 — Two-way transmission test.

noise condition, or a 120 IPM flashing signal if the far-end has registered a high noise condition. The near end is thus advised of the results of the noise check at the far end. The test frame, on receipt of this signal, causes the Teletypewriter to complete the record and then breaks down the connection and advances to the next trunk.

The amplifier, which precedes the amplifier-rectifier includes a network which provides F1A noise weighting during the noise check. The amplifier-rectifier is adjusted in the noise checking condition (that is, when its gain is increased) so that a noise indication will be given when the noise exceeds about 35 or 40 or 45 dba. Since this test is intended only as a rough check to detect any abnormal noise condition, the noise rejection limit used in any given office will be governed by the types of intertoll trunk facilities terminating in that office. No correction is made for the measured loss of the trunk at the time of the noise check, hence the noise is checked at the receiving switchboard level. For the usual types of noise the results of the noise check agree roughly with those which would be obtained by an average observer using a 2-B Noise Measuring Set for a similar "go-no go" type of check.

As is evident from the previous description, each end is expected to complete the various steps of its functions within allotted time intervals. Timing intervals at the far end are controlled by a multivibrator circuit. Timing at the near end is controlled by a similar multivibrator in the intertoll trunk test frame. To insure that the test circuits always perform as they should and that the timing circuits are functioning properly, checks are built into the circuits so that anything which prevents the successful completion of a 2-way measurement on schedule causes the automatic outgoing intertoll trunk test frame at the near end to stop, hold the trunk busy and sound an alarm while awaiting attention of the attendant. The transmission measuring and noise checking circuit at the far end will, however, release itself from the test line so that it will be free to handle other calls.

Semi-Automatic Test

One-milliwatt test power supply outlets have been provided in toll offices for some time for making a one-way transmission measurement frequently referred to as a code 102 test. A test board attendant can reach the one milliwatt test power supply by pulsing forward code 102 or by requesting an operator at the distant end of a manual trunk for a connection. The test power is applied at the distant end for about 10 seconds during which time the attendant measures the loss in the receiving (far-to-near) direction. This is a fairly fast semi-automatic test but, of

course, has the disadvantage that it is a one-way test and cannot be used for all purposes.

In order to provide a semi-automatic two-way test, the far-end equipment is arranged so that a test board attendant can make a code 104 measurement unassisted. This measurement is carried out in 3 steps as shown in the lower portion of Fig. 6.

Step 1

The attendant connects a test cord to the test jack of the intertoll trunk and pulses forward code 104 using his test position dial or key set. When the far end is ready, it returns an off-hook signal which retires the test cord supervisory lamp. He then connects the other end of the cord to the one milliwatt test power supply. The far end then adjusts the receiving and transmitting pads in the same way as for a full automatic test. After about 3 seconds the attendant disconnects the test power and at that time observes the cord supervisory lamp; a single flash indicates that the far end was unsuccessful and is requesting a second trial. If the supervisory lamp remains steadily dark he connects the cord to the receive jack of his transmission measuring circuit to prepare for Step 2.

Step 2

The far end will pause about 2 seconds after the attendant removes the test power to give him time to prepare for Step 2. During this pause the far end will not receive a short spurt of test power as in the case of a full automatic test. Consequently, after the 2 second interval the far end will return one milliwatt for 10 seconds on a semiautomatic test to give the attendant time to complete a measurement. The received power is read directly on the meter of the transmission measuring circuit and is the loss in the far-to-near direction. When the far end removes the test power, the meter reading drops back to the position of no current (infinite loss) and at that time the attendant observes the cord lamp. A single flash at this time is an "add 10" signal and indicates that 10 db should be added to the next measurement. A steady dark lamp indicates that the next measurement should be recorded without correction.

Step 3

After about 2 seconds delay to give the attendant time to record the first measurement, the far end again returns 1 milliwatt, this time through the transmitting pad set up in Step 1. The meter now reads the

loss of the trunk plus the loss of the transmitting pad at the far end. Since the transmitting pad loss equals the trunk loss in the near-to-far direction, the difference between the measurements in Step 3 and Step 2 is the trunk loss in the near-to-far direction. After about 10 seconds the far end removes the test power and starts the noise check in the same way as if this were a full automatic test.

When the far end removes the test power after Step 3, the attendant leaves the connection intact until the cord supervisory lamp lights to indicate completion of the noise check at the far end. A flashing lamp indicates that the noise at the far end exceeds the prescribed limit and a steadily lighted lamp indicates the noise at the far end is below this value. A noise test at the near end may be made by the attendant if he judges, after a listening test, that a noise test is desirable. For this test he uses the standard noise measuring equipment.

PRESENTATION OF TEST RESULTS

When making operational tests and a Teletypewriter is not being used, troubles are registered by means of an audible alarm and accompanying display lamps. When making transmission loss measurements, however, a complete record of the measurements on all trunks tested, both good and bad is frequently needed. A Teletypewriter then becomes a practical necessity; otherwise the attendant would be required to supervise the automatic equipment continuously and to record, from a lamp display or similar indication, the results of each measurement as it was made. Having provided the Teletypewriter for transmission testing, its ability to print letters to represent trouble indications is utilized to avoid halting the progress of the tests when operational troubles are experienced, except when completely inoperative conditions are encountered.

Computer Circuit

As mentioned earlier intertoll trunk transmission performance is rated in terms of bias and distribution grade which are calculated from the deviations of the measured losses of the intertoll trunks from their specified values. For such calculations the maintenance forces are, therefore, more interested in the deviations than they are in actual measured losses. Accordingly, the automatic transmission test and control circuit at the near-end has a computer built into it which will compute the deviation for each measurement so that the deviation can be recorded by the Teletypewriter.

The computer is a bi-quinary relay type adder similar to those used

for other purposes in the telephone plant, for example, in the computer of the automatic message accounting system. It obtains the specified net loss of the trunk being tested from the class relay which remains operated throughout the test. When a computation is to be made of the deviation in the far-to-near direction, for example, the control circuit extends to the adder a number of leads from the contacts of the far-near pad control relays. Some combination of the 9 far-near pad control relays remains operated after the far-near pad adjustment is finished and therefore some combination of the leads extended to the adder will be closed. These leads furnish to the adder the measured loss of the intertoll trunk in the far-to-near direction. The adder then subtracts the specified loss from the measured loss and presents the answer together with the proper sign, + or −, to the teletypewriter for a printed record. The deviation in the near-to-far direction is computed in the same manner by extending corresponding leads from the near-far pad control relays to the adder at the proper time.

Deviation Registers

In determining bias and distribution grade by the method discussed in the companion article,* the deviations from specified net loss are calculated for each measurement. These deviations are grouped together in 0.5 db increments from +8 db to −8 db, all deviations exceeding +7.8 db or −7.8 db being considered as +8.0 db and −8.0 db respectively. For example, all deviations of +0.3 db to +0.7 db, inclusive are considered to be +0.5 db and are so tallied on the data, or stroke, sheet.

To assist in this work the automatic test equipment includes thirty-three manually resettable counters corresponding to the 0.5 db increments from +8.0 db to −8.0 db inclusive. Just prior to a transmission test cycle all these counters are reset to zero. At the time a deviation computation is made, the computer also causes the proper counter to register one count. After the test run on a group of trunks, the counter readings can be transcribed directly as the final tally on the stroke sheet and may be used to determine the bias and distribution grade. A "total tests" counter keeps a tally of all the computations. At the end of the test run the total count serves as a check of the total count of the other 33 counters.

Check for Excessive Deviations

In addition to obtaining data for the calculation of bias and distribution grade, the maintenance forces would also like to know promptly

* H. H. Felder and E. N. Little, Intertoll Net Loss Maintenance Under Operator Distance and Direct Distance Dialing, page 955 of this issue.

when the loss of an intertoll trunk deviates an abnormal amount from its specified value. The maintenance practices currently require that, whenever an intertoll trunk is found to have a deviation of ±5 db or more in either direction, the trunk should be removed from service immediately and the cause of the abnormal deviation corrected. Accordingly, the computer circuit includes an alarm feature which sounds an alarm to attract the immediate attention of the attendant whenever the computed deviation is ±5.0 db or greater.

The maintenance forces may also like to know promptly about trunks with wide deviations but which are not so bad as to require immediate removal from service. For this purpose the computer also includes a limit checking feature. This can be set, by means of optional wiring, to detect deviations in excess of ±3.0 db, ±4.0 db or ±5.0 db. Whenever a deviation exceeds the limit for which the computer is wired, this feature performs as follows:

(1) When the Teletypewriter is not in operation the test frame stops and sounds an alarm.

(2) When the Teletypewriter is recording all measurements, the letter U is added in a separate column at the end of the test record. The letter stands out on the record to permit quick spotting of trunks with abnormal deviations.

(3) By means of a control key, a transmission test record can be printed only for those trunks whose deviation exceeds the computer checking limit or which are "noisy" at either end.

Teletypewriter Record

The Teletypewriter is put into operation by means of a key on the test frame. When this key is normal, no records are printed. Under this condition a trouble causes the test frame to stop and sound an alarm. When the Teletypewriter is operating it prints various records and a minor operational trouble may result only in a record, without an alarm. Each record occupies a separate line on the tape. Each line starts with the four-digit trunk identification number in the first column. Fig. 7 shows a short specimen of the the Teletypewriter record.

When the pass busy key on the test frame is in its nonoperated position, the Teletypewriter will print the trunk identification number, followed by the letter B, for each trunk passed over without test because it was busy. This is done on both operational and transmission test cycles. When the pass busy key is operated no record is made of busy trunks passed without test.

During operational tests no record is printed for trunks which are

satisfactory. Troubles during either operational or transmission test cycles result in a record of the trunk identification number followed by a cue letter in a separate column denoting the nature of the trouble. This may be a single line record or a double line record for a repeat test on the same trunk as previously discussed. For example, in Fig. 7, the letter Y in the second line indicates that on trunk 1267 the far end was unable to complete its transmission measurement successfully. The letter A in lines 3 and 4 indicates that the test frame was unable to establish a connection over trunk 1293 on either its first or second attempt. The record of transmission tests is printed in several columns. Reading from left to right (see Fig. 7) these are (1) trunk identification number, (2) specified net loss, (3) deviation in the far-to-near direction together with the sign, and (4) deviation in the near-to-far direction together with the sign. In columns 2, 3, and 4 the decimal points are omitted and the ten's digits are omitted when they are zero (0). Column 5 will contain an N if the far end is "noisy" or the letter U if the deviation in the far-to-near direction exceeds the computer check limits, preference being given to N if both conditions occur on the same trunk. Likewise column 6 contains an N if the near end is "noisy" or a U if the deviation in the near-to-far direction exceeds the computer check limits. Transmission test cycles will, of course, include a trouble record whenever an operational trouble is encountered or whenever the transmission test cannot be completed successfully.

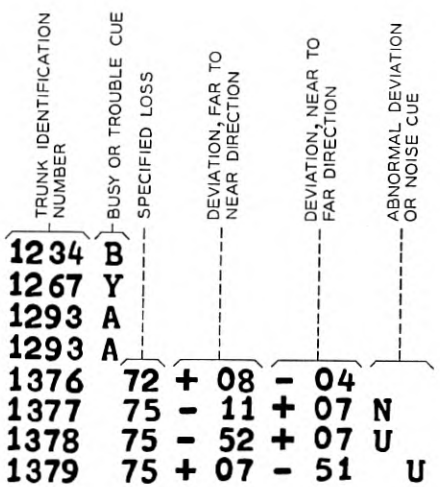


Fig. 7 — Teletypewriter test record.

APPLICATION

Maximum benefits can be derived from the automatic testing equipment by locating it at points having a considerable number of intertoll trunks. This suggests that the near-end installations be placed in offices in larger cities and far-end installations be placed at points with enough trunks available to near-end equipment to justify the far-end equipment. As has been indicated the far-end equipment can operate with either an automatic transmission test and control circuit or with a test board attendant at the opposite end. Therefore, once an office has been supplied with far-end test equipment, all incoming and two-way dial type intertoll trunks from offices provided with near end equipment can be tested on a fully automatic basis and all incoming and two-way dial type intertoll

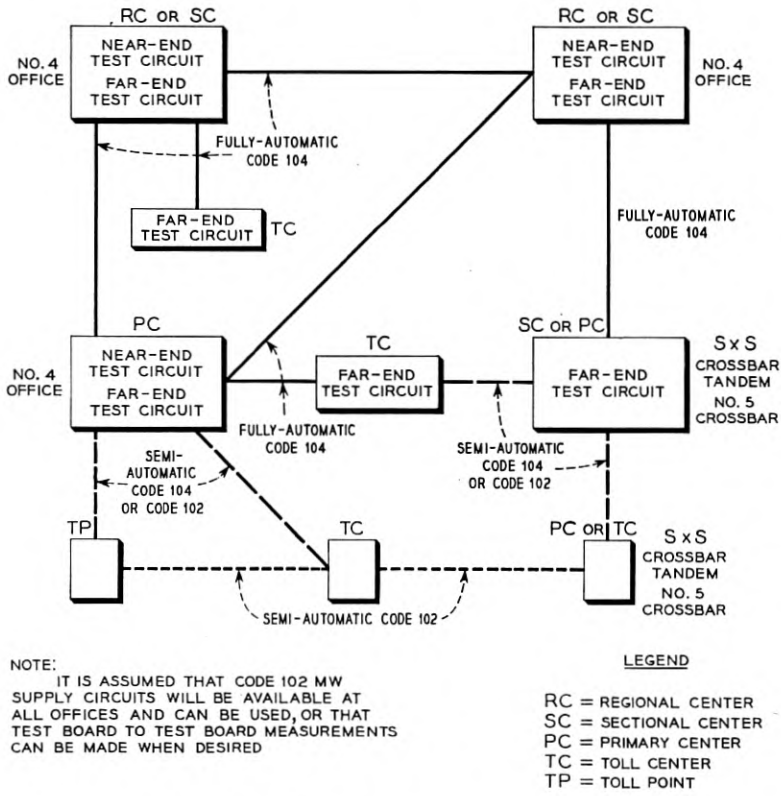


Fig. 8 — Typical layout for automatic testing.

trunks from other toll offices can be tested on a semi-automatic basis from the toll test board in the distant office.

Fig. 8 shows a possible application of automatic test circuits. In such an application, all No. 4 type toll crossbar offices would have both near-end and far-end equipments. Other offices would have far-end equipment only when they have a sufficient number of direct trunks to No. 4 type offices to justify its use. The several types of tests which would be possible are indicated in the illustration.

It can be seen that a well distributed number of near-end and far-end test circuits will make it possible to test automatically a large percentage of the intertoll trunks throughout the country. This is particularly true in the more populous sections, where the concentration of trunks results in the probability of toll centers having trunks to more than one office furnished with near-end equipment.

ACKNOWLEDGMENTS

Automatic intertoll trunk testing arrangements, including transmission tests, are the result of the ideas, efforts and experiences of many people concerned with intertoll switching and maintenance problems throughout the Bell System. Mr. L. L. Glezen and Mr. L. F. Howard deserve particular mention in this regard. Specific credit should also be given to Mr. B. McKim and Mr. T. H. Neely for the basic scheme of two-way transmission measurements and accuracy checks and to Mr. C. C. Fleming for the design of the amplifier and amplifier-rectifier. Appreciation is given to various departments of the American Telephone and Telegraph Company for their assistance during the development and trial of this equipment. Mention should also be made of the hearty cooperation and aid given by the A.T. & T. and Associated Company plant forces during the field trial of automatic transmission testing.

Intertoll Trunk Net Loss Maintenance under Operator Distance and Direct Distance Dialing

By H. H. FELDER and E. N. LITTLE

(Manuscript received March 15, 1956)

Nearly all of the components of an intertoll trunk contribute in some degree to its variations in transmission loss. Automatic transmission regulating devices in carrier systems and in many voice-frequency systems control inherent variations in the intertoll trunk plant. These variations in transmission come mainly from unavoidable causes such as temperature changes. The success of these devices depends on how precisely the trunk is lined up and the manner in which the maintenance adjustments are made. When the nationwide dialing plan with automatic alternate routing is in full swing, maintenance requirements will be more severe because of the material increase in switched business and the number of possible links in tandem, and because operator checks will not be obtained on most calls. Therefore, the maintenance forces will have to keep closer watch on intertoll trunk transmission performance and insure that the necessary adjustments are made in the right places. This article discusses some of the maintenance techniques now used and suggests fields for further study.

TABLE OF CONTENTS

INTRODUCTION

Currently there are over 230,000,000 long distance calls made in the Bell System per month. They range from relatively simple connections involving a single intercity trunk to complex connections involving several intercity trunks in tandem, perhaps totaling 4,000 miles in length. In each case there is a toll connecting trunk at each end. Almost half of this traffic involves distances over 30 miles. The transmission engineer's problem is how to provide uniformly good and dependable transmission so that every one of these calls will be satisfactory to the customers involved. To accomplish this requires among other things that:

1. The design loss of every trunk must be the lowest permissible from the standpoint of echo, singing, crosstalk and noise.
2. The actual loss of every trunk must be kept close to the design loss at all times.

Meeting the first requirement is a matter of system design and circuit layout engineering. The factors involved have been covered in a previous article.[1] Meeting the second requirement is an important function of the maintenance forces and is discussed in this article.

THE PROBLEM OF NET LOSS MAINTENANCE

The transition from manual operation under the "general toll switching plan"[2] to dial operation under the "nationwide dialing plan"[3, 4] is requiring material changes in intertoll trunk design and also in techniques for maintaining these trunks. While precise maintenance is becoming increasingly necessary, it is also becoming more difficult to achieve. There are three important reasons for this.

First, the nationwide dialing plan increases both the possible number of trunks used in tandem for a given call and the variety of the connections in which any particular trunk may be used. This increases the chances of impairment due to deviations from assigned loss in individual trunks since these deviations may combine unfavorably in multiswitched connections. To minimize this, the transmission stability of the individual trunk links must be better than under the old plan.

Second, more and more of the trunks are being put on carrier because

---

[1] H. R. Huntley, Transmission Design of Intertoll Telephone Trunks, B.S.T.J., Sept. 1953.
[2] H. S. Osborne, A General Switching Plan for Telephone Toll Service, B.S.T.J., July, 1930.
[3] A. B. Clark and H. S. Osborne, Automatic Switching for Nationwide Telephone Service, B.S.T.J., Sept., 1952.
[4] J. J. Pilliod, Fundamental Plans for Toll Telephone Plant, B.S.T.J., Sept. 1952.

it is the best solution to the transmission and economic problems. However, carrier involves many more variable elements and requires higher precision of adjustment than voice-frequency systems need. These increase the difficulty of maintaining trunk losses close to design values on a day-by-day basis.

Third, as operator distance and direct distance dialing grow, there is constantly diminishing opportunity for operators to detect and change unsatisfactory connections or to report unsatisfactory transmission conditions to the appropriate testboards for action.

Thus the maintenance problem is in two parts:

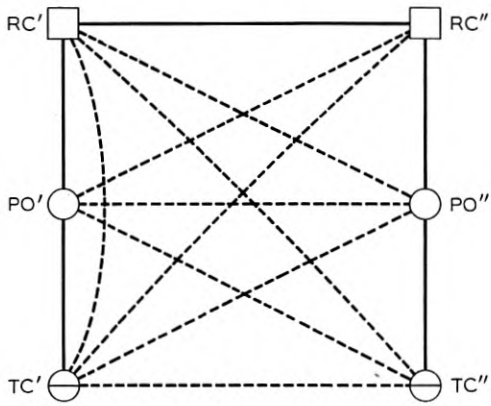
1. How can we reduce departures from design standards even in the face of increasing complexity of plant?
2. What substitute can we find for operator detection of troubles, and can we find even better means of detection?

The ways in which switching plans and the use of carrier reflect upon the problem of trunk net loss maintenance is discussed in more detail in the following sections.

EFFECT OF SWITCHING PLANS

Manual Operation

For many years long distance traffic has been handled on a manual basis under the "general toll switching plan" illustrated in Fig. 1. Between two points indicated by toll centers, TC′ and TC″, it was theoretically possible to get as many as five trunks in tandem. This rarely occurred be-



TC = Toll Center PO = Primary Outlet RC = Regional Center

Fig. 1 — General toll switching plan — manual operation.

Thus under dial operation there is a need for better trunk stability. Therefore, a greater burden is placed on the plant forces to locate unsatisfactory trunks so that proper maintenance action can be taken before customers experience difficulty.

EFFECT OF CARRIER OPERATION

Carrier is the principal transmission instrumentality which makes it possible to go ahead with nationwide dialing with assurance that people can talk satisfactorily over the complex connections set up by the switching systems. But it brings with it formidable problems of maintenance. The high attenuation per mile of the line conductors at carrier frequencies increases the number of variable elements as well as the precision with which they must be adjusted. The interrelation between the elements adds to the complication.

Table I illustrates this by giving some figures comparing 100 miles of a voice-frequency cable trunk with 100 miles of a typical trunk on K carrier, which is widely used on cable facilities. The figures apply in both cases to one direction of transmission.

The ten-to-one ratio in the number of electron tubes represents a greater chance of trouble developing in the carrier trunk due to aging or failure of electron tubes. In the carrier trunks there are more automatic adjustable features. For instance, in a typical K2 carrier system there are five flat gain regulators and one twist regulator in one twist section of approximately 100 miles, against a single regulator in a voice-frequency trunk 100 miles long. These regulators are depended upon to keep the loss variations to tolerable amounts. Any malfunction can have a serious effect on trunk loss. Furthermore, they must be adjusted to the desired regulating range and therefore they are points at which maladjustments may be made.

The channels of any one carrier system or of a 12-channel group are commonly routed by the circuit layout engineers to a number of terminal

TABLE I

	V-f Trunk	K2 Carrier Trunk
Total Conductor Loss -db	35	378
Gain Required to Reduce to Via Net Loss -db	31	377.4
Percentage of Line Loss Represented by a 2 db Variation	5.7	0.53
Number of Electron Tubes	3	28
Number of Amplifiers	3	7
Number of Automatic Regulators	1	6

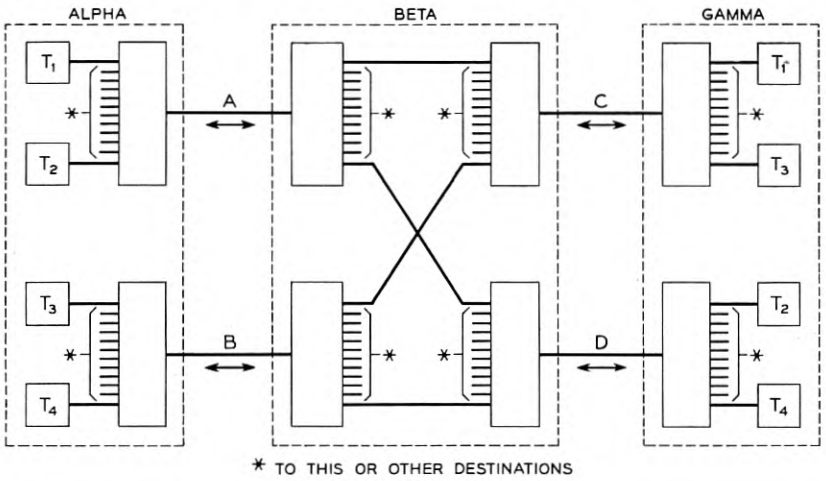


Fig. 3 — Typical carrier channel assignments.

points even though circuit requirements to a given point are sufficient to utilize 12 or more channels. This is done to minimize the chances that all of the trunks between two points will be interrupted by a system failure. A simple case is illustrated by Fig. 3 which shows trunks between Alpha and Gamma connected at an intermediate point, Beta, in such a manner that a failure in any one of the systems A, B, C, or D will affect only half the trunks.

This routing problem, however, complicates the maintenance problem. For example, if trunk T1 were found to have excess loss in the Alpha-Gamma direction it could be corrected by raising the channel gain at Gamma. On the other hand, a correct diagnosis might have disclosed that the trouble was due to a repeater in system A. If this were the case, merely compensating for the excess loss in T1 by changing the channel gain would still leave all other trunks associated with system A in trouble. Later on, if the repeater difficulty were corrected, and no further action were taken, the net loss of T1 would then be too low.

Thus, the flexibility which is so desirable to minimize interruptions of whole circuit groups leads to a difficult problem in the administration of trunk loss adjustment and maintenance. Furthermore, because of the larger numbers and greater dispersion of trunks and terminal points, the situation in the actual telephone plant is much more complex than in the above example. Also, the diagnosis of trouble conditions is made more difficult by the normal variations of channel losses in the carrier systems and consequently of the trunk losses about their design values. This can

be better appreciated when some quantitative aspects of the problem are considered.

QUANTITATIVE ASPECTS OF THE PROBLEM

When the nationwide dial switching plan began to take shape some 8 or 10 years ago, intensive study of the transmission maintenance problem was undertaken. The existing situation was examined to determine whether or not the plant would continue to be satisfactory under the changed conditions. This was done by analyzing the results of many thousands of transmission measurements which had been made on a routine basis in toll test rooms all over the Bell System. Both the measured and the assigned losses were available so the differences between them could be derived and analyzed statistically.

Although the distribution of differences expressed in db for an office did not necessarily follow precisely a normal probability law, the distributions were close enough to normal law so that they could be treated as normal. The results were similar throughout the System. The differences within an office were random as also were the means of the differences from office to office. However, the means tended to be biased in the direction of excess loss. The performance of trunks in multi-link connections which would be set up by the switching machines could therefore be estimated with reasonable accuracy. In the statistical analysis of measurements on the group of trunks, the performance was expressed in terms of "distribution grade" and "bias." In telephone transmission maintenance terminology, bias is the algebraic average of the measured transmission departures in db from individual specified net losses for the group of trunks. The distribution grade is the standard deviation of the differences between measured and specified trunk losses about this bias value. The distribution grades found in these studies were about as follows:

For trunks under 500 miles — about 1.8 db.

For longer trunks — about 2.5 db.

Table II illustrates the effects of the distribution grades on connections involving various combinations of these trunk links, assuming that bias can be neglected.

The design loss objective for a 4-link connection, say 1,000 miles long, is about 7 or 8 db (including 2 db of connecting trunk or pad loss at each end). Table II shows that, in an appreciable percentage of the 4-link connections involving the above type of plant, the variations can be expected to exceed the design loss. Variations of this magnitude can result in transmission impairment due to echo, hollowness, singing, crosstalk,

TABLE II

Number of Intertoll Trunks in the Connection	2	4*	6*	8*
Distribution Grade in db	2.5	4.4	5.0	5.6
Per cent of Connections Departing from Average				
±2 db or more	42	65	69	73
±4 db or more	11	36	42	47
±8 db or more	0.2	7	11	15

* Includes two trunks over 500 miles long.

noise or low volume. Furthermore, undesirable contrast may be encountered on successive calls between the same two telephones.

The results of the study as well as experience with the beginning of automatic alternate routing show that the performance of the existing trunk plant must be improved. Three immediate objectives have been set:

1. Reduction of distribution grades to about ½ of the values mentioned above, i.e., about 1.0 db.
2. Maintenance of office bias within ±0.25 db.
3. Removal from service of individual trunks differing widely from their design losses (in the order of 4 or 5 db).

To achieve these objectives requires effort along four lines. First, systems should be designed to have sufficient stability once they are adjusted. This involves the inclusion of stable circuit elements and the provision of automatic regulating devices to compensate for unavoidable transmission variations arising from natural causes. These features have been applied to existing systems within limits imposed by economic considerations and the state of the art. Further extension of these features will be required in the future in order to meet the above objectives.

Second, before a trunk is placed in service, each of its component parts and the over-all trunk should be adjusted to give the correct loss. From the transmission maintenance point of view, it is extremely important for each trunk to start out with all of its adjustments correctly made.

Third, existing and incipient troubles, and deterioration or maladjustment of components, must be detected and corrected by routine maintenance of individual systems used in making up trunks. Such activity must make up for the inability to design systems to have the desired stability.

Fourth, significant departures from trunk design losses must be detected by over-all transmission measurements, and must be corrected be-

fore service reactions occur. Such measurements will also be of aid in determining the effectiveness of efforts along the first and third lines.

As discussed earlier, the presence of the operator on every call was of material assistance in the detection of unsatisfactory trunks. On operator or direct distance dailed calls, there will be little or no operator conversation over the intertoll trunk connection. As a substitute, the maintenance forces may need to make more frequent checks of the transmission performance of the trunks unless the stability of individual systems and components of systems can be improved. Manual methods have been used by the maintenance forces in the past to measure trunk losses. Semiautomatic measuring methods have been developed to reduce the time and effort required. In many cases the necessary number of measurements will be economical only when made by automatic devices. One form of such gear is described in a companion paper.[6]

The ability to measure over-all trunk losses simply and frequently is of direct aid in detecting when loss deviations exceed maximum tolerances. Such measurements in themselves, however, are insufficient to detect incipient troubles or to indicate the component part responsible for unsatisfactory transmission. An attempt has been made to achieve these objectives by using statistical analysis of the measured data as an aid to diagnosis. The following sections discuss the application of such analysis.

Use of Transmission Loss Data

It has been shown that considerable variation can be expected in trunk losses even in the absence of trouble conditions. For any given group of trunks selected for analysis, the performance is described by the distribution grade and the bias. If a group of trunks is found to have bias, it is usually an indication of some assignable cause. One such cause might be a change in gain of an amplifier common to the group. Another cause might be improper gain adjustment for channel units of a carrier terminal associated with the group.

If a group of trunks is found to have a greater distribution grade than the distribution grade for all the trunks in the office, this may indicate excessive instability in a component part common to the trunks in the group. If analysis of all the trunks terminating in an office shows a higher distribution grade than is usually found in similar offices, the fault may be due to maintenance routines being inadequately or improperly applied.

[6] H. H. Felder, A. J. Pascarella and H. F. Shoffstall, Automatic Testing of Transmission and Operational Functions of Intertoll Trunks, page 927 of this issue.

Statistical analyses must thus be made of data for small groups as well as for large groups of trunks. Furthermore, the groups which are studied must have elements or factors in common in order for the statistics to have significance. Analyses of periodic measurements of losses for the same trunk or groups of similar trunks can likewise indicate significant changes in performance.

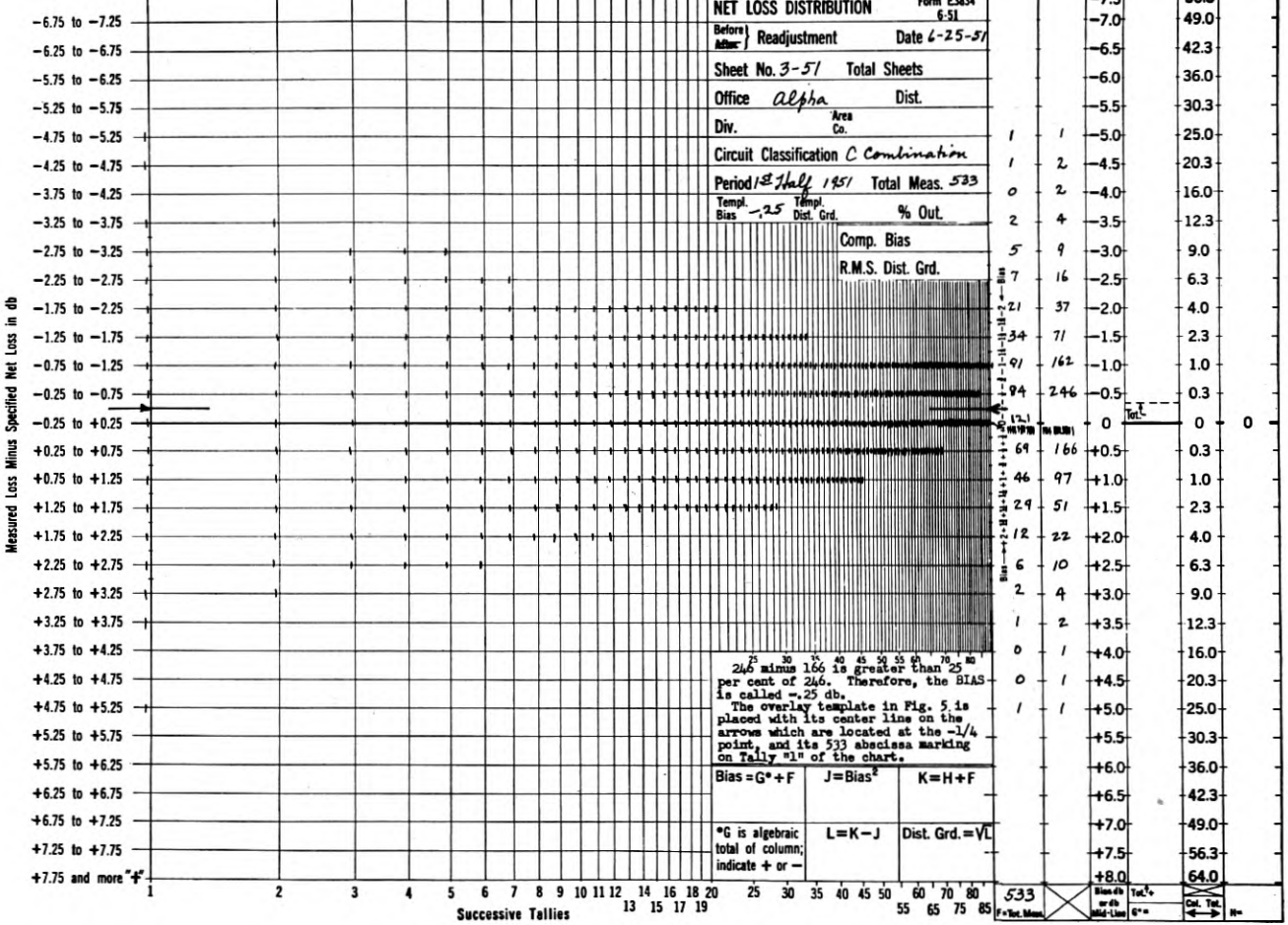
As yet, the problems of properly selecting the trunks to be analyzed and of correlating the results of the analyses with particular system elements needing maintenance attention have been solved only partially. In addition to the need for proper procedures, there is the need for thorough training of maintenance personnel. The complexity of the telephone plant today is increasing the importance of all maintenance personnel having a thorough knowledge of how individual systems function and how the performance of the various system elements reacts upon overall trunk performance.

Procedure for Analyzing Measurements

In an effort to facilitate the application of statistical analysis of trunk performance by plant personnel, a special data sheet and associated templates have been devised. These are shown in Figs. 4, 5, and 6. The method of analysis gives only approximate results but has been found to be sufficiently accurate for reasonably large amounts of data. It is simple, rapid and easily comprehended by the plant personnel. The procedure to be followed consists first of subtracting the specified loss from the measured loss for each of the trunks under study. A stroke is placed on the chart for each of the resulting deviations at the intersection of the appropriate classification and tally lines. For example, the first deviation between −3.25 db and −3.75 db would be stroked on the horizontal line for that band, just to the left of the vertical line for tally 1 (See Fig. 4). The second deviation in that band would be stroked just to the left of the tally 2 line. This is continued until all the deviations have been recorded.

The last stroke in each ½ db band indicates the number of deviations found having values within that band. As shown on Fig. 4, for the analysis by the template method this value is written in the first column, marked "Line Tots. (A)." These values are added and should equal the total number of measurements in the study (533 in the example).

Next, the column "Cum. to ½" is filled out. Beginning at the top line, totals are accumulated to the point where adding the next line total will result in a value exceeding ½ the grand total of measurements (266 in the example). Similarly a value is obtained accumulating the totals from the bottom. In Fig. 4 these values are 246 and 166, respectively.


NET LOSS DISTRIBUTION
Form E-3454 6-51
Before / After } Readjustment
Date 6-25-51
Sheet No. 3-51 Total Sheets
Office alpha Dist.
Div. Area Co.
Circuit Classification C Combination
Period 1st Half 1951 Total Meas. 533
Templ. Bias −.25 Templ. Dist. Grd. % Out.
Comp. Bias
R.M.S. Dist. Grd.
Measured Loss Minus Specified Net Loss in db
Successive Tallies
246 minus 166 is greater than 25 per cent of 246. Therefore, the BIAS is called −.25 db.
The overlay template in Fig. 5 is placed with its center line on the arrows which are located at the −1/4 point, and its 533 abscissa marking on Tally "1" of the chart.
Bias = G* + F
J = Bias²
K = H + F
*G is algebraic total of column; indicate + or −
L = K − J
Dist. Grd. = √L
533
Bias-db or db Mid-Line
Tot.
Cat. Tot.
N=

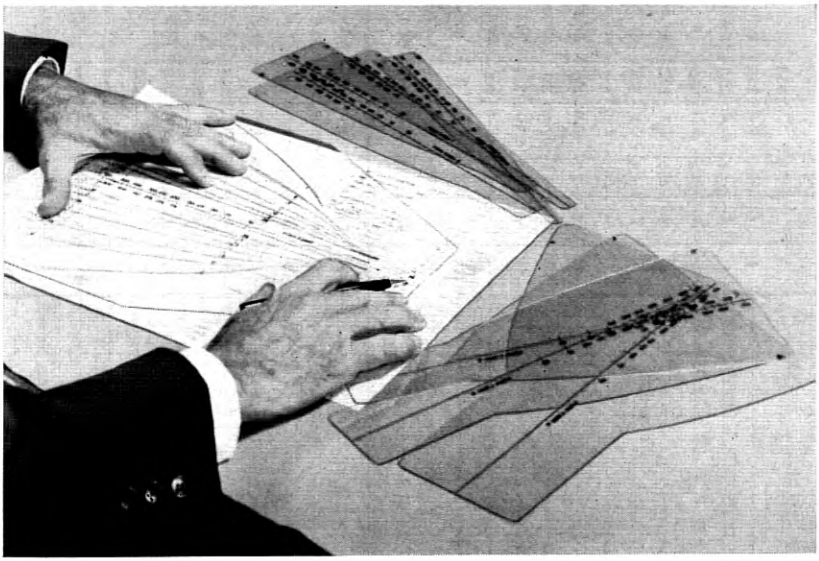


Fig. 5 — Combined template on stroke chart.

By use of this information the approximate bias is determined. The scale of the bias values on the stroke sheet is shown in ¼ db steps along the left-hand edge of the "Line Tots. (A)" column, and bias is determined to the nearest ¼ db. If the two cumulative totals differ from each other by less than 25 per cent of the larger value, an arrow indicating the bias is placed midway between the two class lines representing these cumulative totals. Its value is read on the bias scale. If the two cumulative totals differ from each other by 25 per cent or more of the larger value, an arrow indicating the bias is placed ¾ of the distance between the two class lines representing these cumulative totals and nearer the larger value. In Fig. 4, since 246 minus 166 (80) is greater than 25 per cent of 246 (61.5), the arrow is placed ¾ of the way from the line representing 166 toward the line representing 246; i.e., at the −¼ db point on the bias scale. A second arrow is placed at the corresponding point on the tally 1 line.

As shown in Fig. 5, a combined template is then placed over the chart so that the center line of the template coincides with the two arrows. Along the center line of the template there is a scale indicating numbers of measurements from 50 through 700. The template is moved horizontally so that the point on the scale corresponding to the grand total of measurements (533 in the example) is placed on the 1 tally line. Envelope

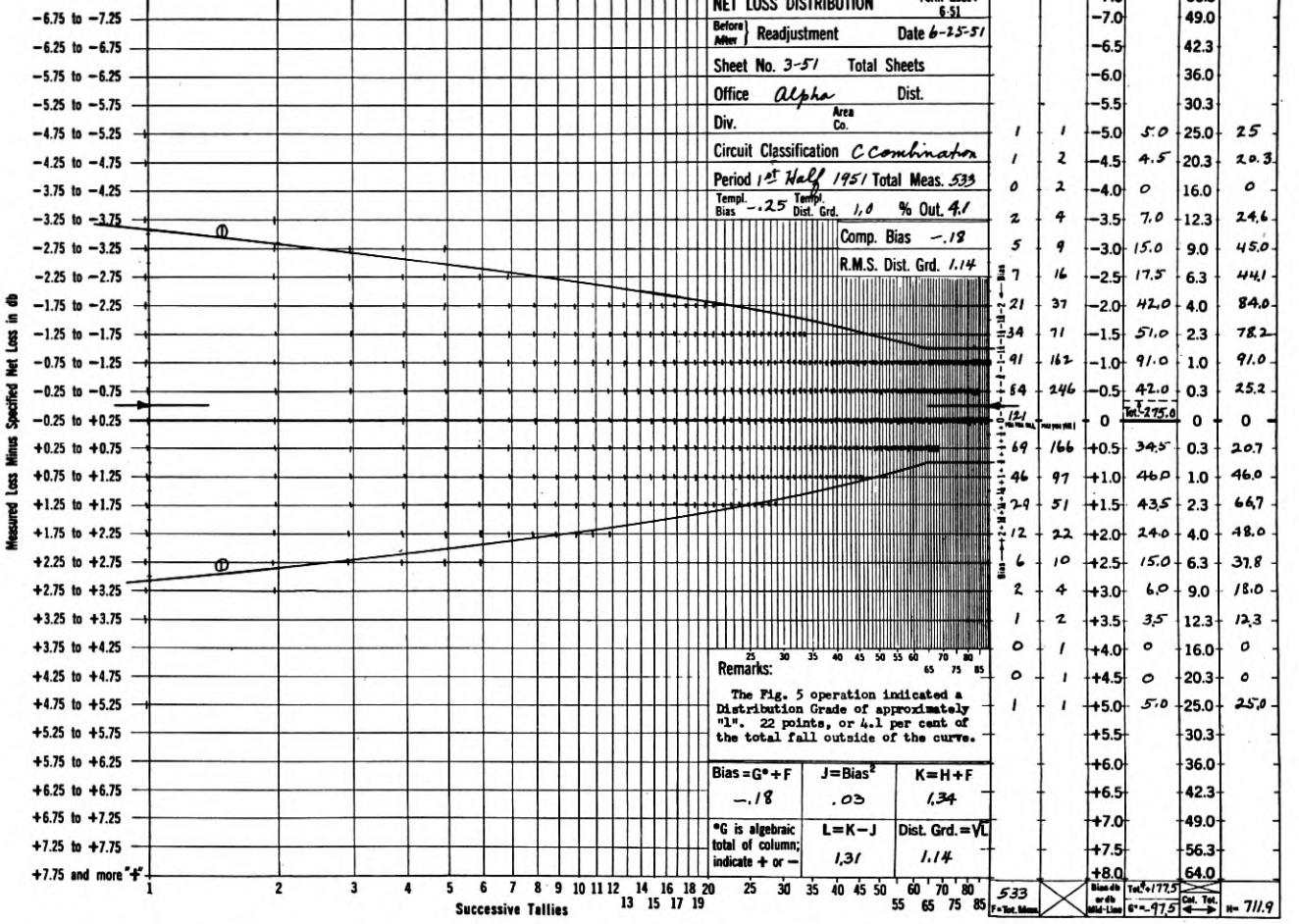


Fig. 6 — Analysis of actual net loss data — results of template and RMS methods shown.

curves for distribution grades from 0.38 db to 3 db are shown on the template. The smallest envelope having not over 8 per cent of the grand total of measurements outside of the envelope represents the approximate distribution grade. In the example, this is the 1 db curve, for which 22 points or 4.1 per cent of the total fall outside of the curve. Using a cut-out template corresponding to the distribution grade, a trace is placed on the stroke sheet, as shown on Fig. 6.

In cases where the small number of measurements or the character of the dispersion makes it difficult to fit the data with any of the envelope curves of the template, RMS methods of determining the distribution grade and bias afford a better estimate. In the example on Fig. 6, the bias is thus found to be −0.18 db and the distribution grade is found to be 1.14 db.

When the automatic transmission test and control circuit described in the companion paper is used for measuring net losses, the bias and distribution grade can be determined more quickly and easily. This circuit measures the transmission in terms of deviations from the specified loss and records these by a teletypewriter. In addition, registers indicate the total number of measurements and the number of deviations falling in the ½ db bands shown on the stroke chart. The final strokes for each band can thus be placed on the chart directly without the need for stroking each measurement. From this point on, the analysis and the final tracing of the envelope curve which is selected are the same as in the case illustrated by Fig. 6.

EFFECTIVENESS OF OVER-ALL TRUNK TESTS AND ANALYSES

Simple Layouts

With simple trunk layouts particularly those involving one voice-frequency or carrier link, plant forces have been able to use over-all trunk measurements and analyses as a direct aid in maintenance. Early field trials of the stroke chart method were made at two operating telephone company offices. The testers made up stroke sheets from their routine measurements and interpreted the results to find clues as to what to investigate. Stroke sheets made at successive routine testing periods also showed them what improvements they were obtaining in the operation of the trunks.

Both offices started with distribution grades of about 1.8 db and with biases of about ¾ db. The trunk plant was then given a thorough cleanup and realignment more rigorous than that called for in the maintenance practices at the time. Similar rigorous circuit order tests were followed

as circuit order changes were made. Many small troubles were found and cleared. As the result of such rigorous circuit order work and the use of statistical analyses, the distribution grades at the end of the trial were reduced to about 1 db and the biases were brought close to zero.

The maintenance activities were conducted by the regular test forces during normally available maintenance time. Although the initial work involved in cleaning up the trunks necessitated some slippage in the periodic maintenance tests, troubles requiring realignment were eventually reduced to the point where it became possible to carry on the periodic testing work concurrently with the more rigorous circuit order work.

Complex Layouts

During the field trial of the automatic transmission test and control circuit discussed in a companion paper, there was an opportunity for studying transmission data taken on intertoll trunks of greater length and complexity of layout. These trunks were composed of two or more carrier links and connected Washington, D.C. to several outlying points; namely, Atlanta, Georgia; Boston, Massachusetts; Hempstead, New York; New York, New York; Oakland, California; and Richmond, Virginia. A total of 231 trunks were in the groups. When the trial began, without preliminary rigorous circuit order work on the trunks involved, the distribution grade was 2.26 db and the bias was +0.35 db. Maintenance investigation was initiated only when trunks were found to have departed more than a prescribed amount from their specified net losses. Initially this value was 4 db and later it was reduced to 3.5 db.

As many of these wide deviations were investigated and corrected as available manpower permitted. The layouts were so complex, however, that it was found impracticable to give prompt attention to all of them; and in many cases it was impossible to check carrier systems that were suspected of being the source of some of the deviations. At the end of the trial the distribution grade had been reduced from the original 2.26 db to a range of about 1.8 to 2 db. The bias had not been changed significantly from the original +0.35 db.

These results indicated very little improvement from the limited readjustments found practicable during the tests. Analysis of the test results has shown that transmission maintenance methods must be improved in some respects. An example of this was a case where the data indicated several trunks to be affected by excessive variation from some common cause. This was traced to a group pilot being out of limits. If routine maintenance methods had indicated this difficulty earlier, the amount of time in which service could have been affected by these trunks would have been reduced. This is important because of the difficulty of finding evidence of common trouble sources, with complex layouts.

The scope of the trial was then limited to a smaller group of intertoll trunks which could be given close attention. The 42 trunk group between Washington, D.C., and Atlanta, Ga., was selected and these trunks were put through rigorous circuit order tests and adjustments approaching the completeness of initial line-up tests. A test cycle composed of transmission loss measurements made on the 42 trunks in both directions was performed four times daily for a period of about five months. During the period covered by this phase of the trial, adjustments were made only as indicated by carrier pilot variations, by deviations from specified net loss large enough to operate the limit feature of the automatic transmission test and control circuit, or with other trouble clearance.

The tests for each day were analyzed as a group. On the first day the distribution grade of the deviations from specified net loss for the group was 0.8 db and the bias was +0.5 db. On the last day the distribution grade was 1.2 db and the bias was −0.25 db. For the entire group of measurements (584 test cycles), the distribution grade of the deviations was 1.26 db and the bias was −0.08 db. This represented a substantial improvement over the results obtained in the first phase of the trial. It showed that a great deal can be accomplished by improving the circuit order procedures and increasing the thoroughness with which they are carried out.

It was found that combination carrier trunks composed of permanently connected links, thus not having the benefit of control by terminal-to-terminal pilots, have more variability than individual trunks having over-all pilots. Adjustment of such combination trunks requires coordinated action at the various pilot terminals through which the trunk passes, in order that readjustment of the over-all trunk loss can be made at the point in the system responsible for the deviation. In the case of many route junctions, the complexity of the layout makes it difficult to coordinate the necessary measurements at several points so that the proper point for adjustment can be determined.

NEED FOR EDUCATION

The complexity of carrier system layout as indicated above, has imposed a difficult task on the plant transmission maintenance forces. Although our present transmission maintenance practices seem to be adequate for systems in simple layouts, some expansion appears needed for the more complex layouts. This will require further study.

It is important to keep in mind, however, that the provision of good practices and training of personnel in following the detailed steps therein are not in themselves sufficient to assure good transmission maintenance. There is an additional need for *education* of plant personnel in fundamental considerations affecting operation of carrier systems. This must in-

clude over-all objectives, inherent capabilities and limitations, and the interrelation of functions of the many basic blocks comprising carrier systems. Personnel so educated can approach the problems of transmission maintenance with understanding and avoid the maladjustments and troubles due to "man-failure" which are potential hazards in any complex systems.

SUMMARY AND CONCLUSIONS

In summary, the problem of maintaining satisfactory transmission over trunks under distance dialing involves, primarily:

1. Improving the over-all trunk net loss stability so that the distribution grade does not exceed 1 db, as an initial objective.
2. Reducing trunk loss bias for individual offices to less than ± 0.25 db.
3. Removing from operation those trunks having excessive loss deviations before unfavorable service reactions occur.

To do these things in the face of the increasing complexity of our plant and the absence of operator surveillance will require that:

1. Individual systems have adequate short term stability to keep day-to-day variations small.
2. Routine tests and adjustments be made on individual systems and components to correct for long-term deterioration.
3. Frequent over-all trunk tests be made to locate trunks whose performance is beyond acceptable limits and, as a quality control measure, to monitor the performance of the trunk plant.
4. Trunk trouble-shooting be performed on a well coordinated basis to locate and correct the source of trouble. (Compensating maladjustments must be avoided.)

Although facilities are available and methods are known for doing some of these things, considerable effort is required as follows:

1. Study of performance of individual systems to determine capabilities of present design and major sources contributing to over-all trunk instability.
2. Study of transmission maintenance procedures, both routine and trouble-shooting, to determine the proper test intervals and how best the procedures can be carried out on a coordinated basis.
3. Development of improvements in systems and test facilities as indicated by the above studies. Convenience is an important factor in test arrangements.
4. Thorough education of personnel in the over-all make-up, function and interrelation of systems within the trunk plant, and in the significance of transmission maintenance in providing uniformly good and dependable transmission.

Bell System Technical Papers Not Published in This Journal

ABBOTT, L. E.,[1] and POMEROY, A. F.[1]

How To Get More Range From An Air Gage, Am. Machinist, **100,** pp. 113–115, Feb. 27, 1956.

AHEARN, A. J.,[1] and LAW, J. T.[1]

Russell Effect in Silicon and Germanium, J. Chem. Phys., Letter to the Editor, **24,** pp. 633–634, Mar., 1956.

ANDERSON, P. W., see Clogston, A. M.

ARLT, H. G.[1]

Standardization of Materials, Standards Engineering, **8,** pp. 6–7, Mar., 1956.

BASHKOW, T. R.[1]

DC Graphical Analysis of Junction Transistor Flip-Flops, A.I.E.E. Commun. and Electronics., **23,** pp. 1–6, Mar., 1956.

BECKER, J. A.,[1] and BRANDES, R. G.[1]

A Favorable Condition for Seeing Simple Molecules in a Field Emission Microscope, J. Appl. Phys., **27,** pp. 221–223, Mar., 1956.

BENNETT, W. R.[1]

Characteristics and Origins of Noise — Part I., Electronics, **29,** pp. 154–160, Mar., 1956.

BENNETT, W. R.[1]

Electrical Noise — Part II: Noise Generating Equipment, Electronics, **29,** pp. 134–137, Apr., 1956.

BENNETT, W. R.[1]

Synthesis of Active Networks, Proc. Symp. Modern Network Synthesis, MRI Symposia Series, **5,** pp. 45–61, 1956.

[1] Bell Telephone Laboratories, Inc.

BLECHER, F. H.[1]

A Junction Transistor Integrator, Proc. National Electronics Conference, **11,** pp. 415–430, Mar. 1, 1956.

BOMMEL, H. E.,[1] MASON, W. P.,[1] and WAINER, A. W.[1]

Dislocations, Relaxations and Anelasticity of Crystal Quartz, Phys. Rev., **102,** pp. 64–71, Apr. 1, 1956.

BOZORTH, R. M.[1]

Quelques Propriétés Magnétiques, Électrioques Et Optiques Des Films Obtenus Par Électrolyse Et Par Évaporation Thermique, Le J. De Physique Et Le Radium, **17,** pp. 256–262, Mar., 1956.

BOYET, H., see Weisbaum, S.

BRADY, G. W.[1]

X-Ray Study of Tillurium Oxide Gas, J. Chem. Phys., Letter to the Editor, **24,** p. 477, Feb., 1956.

BRANDES, R. G., see Becker, J. A.

BRATTAIN, W. H., see Garrett, C. G. B.

BRAUN, F. A.[1]

Mounting Scheme for Large Cathodes, Rev. Sci. Instr., Lab. and Shop Notes Section, **27,** p. 113, Feb., 1956.

CLOGSTON, A. M.,[1] SUHL, H.,[1] WALKER, L. R.,[1] and ANDERSON, P. W.[1]

Possible Source of Line Width in Ferromagnetic Resonance, Phys. Rev., Letter to the Editor, **101,** pp. 903–905, Jan. 15, 1956.

DE LEEUW, K.,[1] MOORE, E. F.,[1] SHANNON, C. E.,[1] and SHAPIRO, N.[1]

Computability by Probabilistic Machines, Automata Studies, (Princeton Univ. Press), pp. 183–212, Apr., 1956.

EIGLER, J. H., see Sullivan, M. V.

FOX, A. G.[1]

Wave Coupling by Warped Normal Modes, I.R.E. Trans., **PGMTT, 3,** pp. 2–6, Dec., 1955.

[1] Bell Telephone Laboratories, Inc.

FRANCOIS, E. E., see LAW, J. T.

GARDNER, M. B.[1]

Speech We May See, Volta Review, 58, pp. 149–155, Apr., 1956.

GARRETT, C. G. B.,[1] and BRATTAIN, W. H.[1]

Some Experiments on, and a Theory of, Surface Breakdown, J. Appl. Phys., **27,** pp. 299–306, Mar., 1956.

HAYNES, J. R.,[1] and WESTPHAL, W. C.[1]

Radiation Resulting fron Recombination of Holes and Electrons in Silicon, Phys. Rev., **101,** pp. 1676–1678, Mar. 15, 1956.

HERRMANN, D. B., see Williams, J. C.

KELLY, M. J.[1]

Contributions of Research to Telephony — A Look at Past and Glance into Future, Franklin Inst. J., **261,** pp. 189–200, Feb., 1956.

KLEIMACK, J. J., see Wahl, A. J.

LAW, J. T.,[1] and FRANCOIS, E. E.[1]

Adsorption of Gases on Silicon Surface, J. Chem. Phys., **60,** pp. 353–358, Mar., 1956.

LAW, J. T., see Ahearn, A. J.

LLOYD, S. P.,[1] and MCMILLAN, B.[1]

Linear Least Squares Filtering and Prediction of Sampled Signals, Proc. Symp., **PIB, V,** pp. 221–247, Apr., 1955.

LOGAN, R. A.[1]

Thermally Induced Acceptors in Germanium, Phys. Rev., **101,** pp. 1455–1459, Mar. 1, 1956.

MASON, W. P.[1]

Comments on Weertman's Dislocation Relaxation Mechanism, Phys. Rev., Letter to the Editor, **101,** p. 1430, Feb., 15 1956.

MASON, W. P., see Bommel, H. E.

MCMILLAN, B., see Lloyd, S. P.

[1] Bell Telephone Laboratories, Inc.

MENDEL, J. T.[1]

Microwave Detector, Proc. I.R.E., **44,** pp. 503–508, Apr., 1956.

MERZ, W. J., see Remeika, J. P.

MOORE, E. F.[1]

Gedanken-Experiments on Sequential Machines, Automata Studies (Princeton Univ. Press), pp. 129–153, Apr., 1956.

MOORE, E. F., see de Leeuw, K.

POMEROY, A. F., see ABBOTT, L. E.

REMEIKA, J. P.,[1] and MERZ, W. J.[1]

Guanidine Vanadium Sulfate Hexahydrate: A New Ferroelectric Material, Phys. Rev., Letter to the Editor, **102,** p. 295, Apr. 1, 1956.

ROBERTSON, S. D.[1]

Ultra-Bandwidth Finline Coupler, I.R.E. Trans., **PGMTT, 3,** pp. 45–48, Dec., 1955.

ROSE, D. J.[1]

On the Magnification and Resolution of the Field Emission Electron Microscope, J. Appl. Phys., **27,** pp. 215–220, Mar., 1956.

SHANNON, C. E., see de Leeuw, K.

SHAPIRO, N., see de Leeuw, K.

SUHL, H.[1]

Subsidiary Absorption Peaks in Ferromagnetic Resonance at High Signal Levels, Phys. Rev., Letter to the Editor, **101,** pp. 1437–8, Feb. 15, 1956.

SUHL, H., see Clogston, A. M.

SULLIVAN, M. V.,[1] and EIGLER, J. H.[1]

Five Metal Hydrides as Alloying Agents on Silicon, J. Electrochem. Soc., **103,** pp. 218–220, Apr., 1956.

SULLIVAN, M. V.,[1] and EIGLER, J. H.[1]

Electrolytic Stream Etching of Germanium, J. Electrochem. Soc., **103,** pp. 132–134, Feb., 1956.

[1] Bell Telephone Laboratories, Inc.

TRENT, R. L.[1]

Design Principles of Junction Transistor Audio Amplifiers, I.R.E. Trans., **PQA, 3,** pp. 143–161, Sept.–Oct., 1955.

TURNER, D. R.[1]

The Anode Behavior of Germanium in Aqueous Solutions, J. Electrochem. Soc., **103,** pp. 252–256, Apr., 1956.

UHLIR, A., JR.[1]

High-Frequency Shot Noise in PN Junctions, Proc. I.R.E., Correspondence, **44,** pp. 557–558, Apr., 1956.

VAN HASTE, W.[1]

Statistical Techniques for a Transmission System, A.I.E.E. Commun. and Electronics, **23,** pp. 50–54, Mar., 1956.

VAN HASTE, W.[1]

Component Reliability in a Transmission System, Elec. Engg., **75,** p. 413, May, 1956.

VAN ROOSBROECK, W.[1]

Theory of the Photomagnetoelectric Effect in Semiconductors, Phys. Rev., **101,** pp. 1713–1724, Mar. 15, 1956.

VAN UITERT, L. G.[1]

High Resistivity Nickel Ferrites — The Effects of Minor Additions of Manganese or Cobalt, J. Chem. Phys., **24,** p. 306, Feb., 1956.

WAHL, A. J.,[1] and KLEIMACK, J. J.[1]

Factors Affecting Reliability of Alloy Junction Transistors, Proc. I.R.E., **44,** pp. 494–502, Apr., 1956.

WAINER, A. W., see Bommel, H. E.

WALKER, L. R., see Clogston, A. M.

WEISBAUM, S.,[1] and BOYET, H.[1]

A Double-Slab Ferrite Field Displacement Isolator at 11 KMC, Proc. I.R.E., **44,** pp. 554–555, Apr., 1956.

WESTPHAL, W. C., see Haynes, J. R.

[1] Bell Telephone Laboratories, Inc.

WILLIAMS, J. C.,[1] and HERRMANN, D. B.[1]

Surface Resistivity of Nonporous Ceramic and Organic Insulating Materials at High Humidity with Observations of Associated Silver Migration, I.R.E. Trans., **PGRQC, 6,** pp. 11–20, Feb., 1956.

WOLONTIS, V. M.[1]

A Complete Floating-Decimal Interpretive System for the IBM 650 Magnetic Drum Calculator, IBM Technical Newsletter, **11,** Mar., 1956.

[1] Bell Telephone Laboratories, Inc.

Recent Monographs of Bell System Technical Papers Not Published in This Journal*

ARNOLD, W. O., and HOEFLE, R. R.

A System Plan for Air Traffic Control, Monograph 2483.

BABCOCK, W. C., RENTROP, E., and THAELER, C. S.

Crosstalk on Open-Wire Lines, Monograph 2520.

BECK, A. C., and MANDEVILLE, G. D.

Microwave Traveling-Wave Tube Millimicrosecond Pulse Generators, Monograph 2551.

BOZORTH, R. M., WILLIAMS, H. J., and WALSH, DOROTHY E.

Magnetic Properties of Some Orthoferrites and Cyanides at Low Temperatures, Monograph 2591.

BRIDGERS, H. E.

A Modern Semiconductor — Single-Crystal Germanium, Monograph 2552.

CETLIN, B. B., see Galt, J. K.

CHYNOWETH, A. G.

Measuring the Pyroelectric Effect with Special Reference to Barium Titanate, Monograph 2545.

CORENZWIT, E., see Matthias, B. T.

CUTLER, C. C.

Spurious Modulation of Electron Beams, Monograph 2543.

DAIL, H. W., JR., see Galt, J. K.

* Copies of these monographs may be obtained on request to the Publication Department, Bell Telephone Laboratories, Inc., 463 West Street, New York 14. N. Y. The numbers of the monographs should be given in all requests.

DAVIS, H. M., see Wernick, J. H.

DESOER, C. A.

Iterative Solution of Networks of Resistors and Ideal Diodes, Monograph 2583.

DUNCAN, R. S., and STONE, H. A., JR.

A Survey of the Application of Ferrites to Inductor Design, Monograph 2579.

FELDMAN, W. L., see Pearson, G. L.

FEWER, D. R., see Kircher, R. J.

FRY, THORNTON C.

Mathematics as a Profession Today in Industry, Monograph 2585.

GALT, J. K., YAGER, W. A., MERRITT, F. R., CETLIN, B. B., and DAIL, H. W., JR.

Cyclotron Resonance in Metals: Bismuth, Monograph 2535.

GEBALLE, T. H., see Hrostowski, H. J.

GIANOLA, U. F.

Photovoltaic Noise in Silicon Broad Area p-n Junctions, Monograph 2546.

GOSS, A. J., see Hassion, F. X.

GYORGY, E. M., see Heinz, O.

HAGELBARGER, D. W., see Pfann, W. G.; also Shannon, C. E.

HARKER, K. J.

Periodic Focusing of Beams from Partially Shielded Cathodes, Monograph 2553.

HASSION, F. X., THURMOND, D. C., TRUMBORE, F. A., and GOSS, A. J.

Germanium: on the Melting Point; on the Silicon Phase Diagram, Monograph 2489.

HEIDENREICH, R. D., see Williams, H. J.

HEINZ, O., GYORGY, E. M., and OHL, R. S.

Solid-State Detector for Low-Energy Ions, Monograph 2568.

HERRMANN, D. B., see Williams, J. C.

HOEFLE, R. R., see Arnold, W. O.

HROSTOWSKI, H. J., MORIN, F. J., GEBALLE, T. H., and WHEATLEY, G. H.

Hall Effect and Conductivity of InSb, Monograph 2586.

INGRAM, S. B.

The Graduate Engineer — His Training and Utilization in Industry, Monograph 2554.

KELLY, M. J.

Contributions of Research to Telephony, Monograph 2590.

KETCHLEDGE, R. W.

Distortion in Feedback Amplifiers, Monograph 2488.

KIRCHER, R. J., TRENT, R. L., and FEWER, D. R.

Audio Amplifier Applications of Junction Transistors, Monograph 2484.

KUH, E. S.

Special Synthesis Techniques for Driving Point Impedance Functions, Monograph 2581.

LEE, C. Y.

Similarity Principle with Boundary Conditions for Pseudo-Analytic Functions, Monograph 2587.

MANDEVILLE, G. D., see Beck, A. C.

MATTHIAS, B. T., and CORENZWIT, E.

Superconductivity of Zirconium Alloys, Monograph 2526.

MERRITT, F. R., see Galt, J. K.

MILLER, L. E.

Negative Resistance Regions in Collector Characteristics of Point-Contact Transistor, Monograph 2574.

MOLL, J. L., and ROSS, I. M.

Dependence of Transistor Parameters on Distribution of Base Layer Resistivity, Monograph 2575.

MONTGOMERY, H. C., see PEARSON, G. L.

MORIN, F. J., see HROSTOWSKI, H. J.

NESBITT, E. A., see WILLIAMS, H. J.

OHL, R. S., see HEINZ, O.

OWENS, C. D.

Stability Characteristics of Molybdenum Permalloy Powder Cores, Monograph 2576.

PEARSON, G. L., MONTGOMERY, H. C., and FELDMANN, W. L.

Noise in Silicon p-n Junction Photocells, Monograph 2555.

PEDERSEN, L.

Aluminum Die Castings for Carrier Telephone Systems, Monograph 2593.

PEDERSON, D. O.

Regeneration Analysis of Junction Transistor Multivibrators, Monograph 2452.

PFANN, W. G., and HAGELBARGER, D. W.

Electromagnetic Suspension of a Molten Zone, Monograph 2556.

PRINCE, M. B.

High-Frequency Silicon-Aluminum Alloy Junction Diodes, Monograph 2557.

RENTROP, E., see BABCOCK, W. C.

ROSS, I. M., see MOLL, J. L.

SCHAWLOW, A. L.

Structure of the Intermediate State in Superconductors, Monograph 2569.

SHANNON, C. E., and HAGELBARGER, D. W.

Concavity of Resistance Functions, Monograph 2547.

SIMKINS, Q. W., and VOGELSONG, J. H.

Transistor Amplifiers for Use in a Digital Computer, Monograph 2548.

SNOKE, L. R.

Specific Studies on Soil-Block Procedure for Bioassay of Wood Preservatives, Monograph 2577.

SOUTHWORTH, G. C.

Early History of Radio Astronomy, Monograph 2544.

STONE, H. A., JR., see Duncan, R. S.

TANNER, T. L.

Current and Voltage-Metering Magnetic Amplifiers, Monograph 2582.

THAELER, C. S., see Babcock, W. C.

THURMOND, D. C., see Hassion, F. X.

TRENT, R. L., see Kircher, R. J.

TRUMBORE, F. A., see Hassion, F. X.

ULRICH, W., see Yokelson, B. J.

VOGELSONG, J. H., see Simkins, Q. W.

WALSH, DOROTHY E., see Bozorth, R. M.

WERNICK, J. H., and DAVIS, H. M.

Preparation and Inspection of High-Purity Copper Single Crystals, Monograph 2571.

WHEATLEY, G. H., see Hrostowski, H. J.

WILLIAMS, H. J., see Bozorth, R. M.

WILLIAMS, H. J., HEIDENREICH, R. D., and NESBITT, E. A.

How Cobalt Ferrite Heat Treats in a Magnetic Field, Monograph 2558.

WILLIAMS, J. C., and HERRMANN, D. B.

Surface Resistivity of Non-Porous Ceramic and Organic Insulating Materials, Monograph 2560.

YAGER, W. A., see Galt, J. K.

YOKELSON, B. J., and ULRICH, W.

Engineering Multistage Diode Logic Circuits, Monograph 2592.

Contributors to This Issue

ARTHUR B. CRAWFORD, B.S.E.E. 1928, Ohio State University; Bell Telephone Laboratories 1928–. Mr. Crawford has been engaged in radio research since he joined the Laboratories. He has worked on ultra short wave apparatus, measuring techniques and propagation; microwave apparatus, measuring techniques and radar, and microwave propagation studies and microwave antenna research. He is author or co-author of articles which appeared in The Bell System Technical Journal, Proceedings of the I.R.E., Nature, and the Bulletin of the American Meteorological Society. He is a Fellow of the I.R.E. and a member of Sigma Xi, Tau Beta Pi, Eta Kappa Nu, and Pi Mu Epsilon.

C. CHAPIN CUTLER, B.S. 1937, Worcester Polytechnic Institute. Bell Telephone Laboratories 1937–. Mr. Cutler's early work was in research related to the problems of the short wave multiplex radio transmitter. During World War II he was engaged in research on the proximity fuse and microwave antennas for radar use. Since the war he has been concerned with research on the microwave amplifier and the traveling wave tube. Mr. Cutler is a member of the I.R.E. and Sigma Xi.

HARRY H. FELDER, B.S. in Electrical and Mechanical Engineering, Clemson A. and M., 1918. After some months in the U. S. Signal Corps he joined the Engineering Department of the American Telephone and Telegraph Company in 1919. He joined the Laboratories in 1934. He has been engaged in general transmission problems in connection with telephone repeater development and toll circuit layout and switching. During World War II, Mr. Felder assisted in the development of a method of laying telephone wires from airplanes. Since that time he has continued to work on the transmission aspects of intertoll trunk design, switching, maintenance and loading. He was also associated with adapting of cable carrier circuits for radio broadcast networks. Mr. Felder is a member of Tau Beta Pi.

J. H. FORSTER, B.A. 1944, M.A. 1946, University of British Columbia; Ph.D. 1953, Purdue University; Bell Laboratories 1953–. Since joining the Laboratories, Dr. Forster has been engaged in research on semi-

conductor devices including point-contact transistor development, transistor reliability studies and the development of low-noise alloy transistors. He also served as instructor of semiconductor electronics in the Laboratories Communications Development Training program. At present he is engaged in surface studies and semiconductor device reliability. Member of Sigma Pi Sigma and Sigma Xi.

DAVID C. HOGG, B.SC., University of Western Ontario, 1949; M.Sc. and Ph.D., McGill University, 1950 and 1953. Dr. Hogg joined Bell Telephone Laboratories in July 1953 and has worked at the Holmdel Laboratory. He has been engaged in studies of artificial dielectrics for microwaves, antenna problems, and over-the-horizon and millimeter wave propagation as a member of the Radio Research Department. During World War II Dr. Hogg was in the Canadian Army and spent five years in Europe. From 1950 to 1951 he was engaged in research for the Defense Research Board of Canada. He is a member of Sigma Xi.

JOHN L. KELLY, JR., B.A. in 1950, M.A. in 1952, and Ph.D. in 1953, all in Physics at the University of Texas. Dr. Kelly joined Bell Telephone Laboratories in 1953 as a member of the Television Research Department at the Murray Hill Laboratory. He has been engaged in experimental work on the nature of television pictures as well as theoretical investigations pertaining to applications of the Information Theory to television. In 1944 he was commissioned a Navy pilot and served three years.

ARCHIE P. KING, B.S. California Institute of Technology, 1927. After three years with the Seismological Laboratory of the Carnegie Institution of Washington, Mr. King joined Bell Telephone Laboratories in 1930. Since then he has been engaged in ultra-high-frequency radio research at the Holmdel Laboratory, particularly with waveguides. For the last ten years Mr. King has concentrated his efforts on waveguide transmission and waveguide transducers and components for low-loss circular electric wave transmission. He holds at least a score of patents in the waveguide field. Mr. King was cited by the Navy for his World War II radar contributions. He is a Senior Member of the I.R.E. and is a Member of the American Physical Society.

J. G. LINVILL, A.B., William Jewell College, 1941; S.B. in 1943, S.M. in 1945 and Sc.D. in 1949, all in electrical engineering at Massachusetts Institute of Technology. Dr. Linvill served at M. I. T. as assistant pro-

fessor in electrical engineering from 1949 to 1951 and was a consultant to Sylvania Electrical Products. He joined Bell Telephone Laboratories in 1951 and worked on active network problems involving applications of transistors as the active element. In March, 1955, he became Associate Professor of Electrical Engineering at Stanford University. He is a member of the American Institute of Electrical Engineers, Institute of Radio Engineers, Sigma Xi, and Eta Kappa Nu.

Edward N. Little, A.B., Yale, 1916; S.B., Massachusetts Institute of Technology, 1919; Signal Corps and Air Service Radio Officer training, World War I. Joined Long Lines Department of A. T. & T. in 1919 to work on transmission studies. Transferred to Transmission Section of O. & E. Department in 1922 in work dealing with telephone repeaters. Nine years later joined the group working on transmission maintenance, and since then has worked principally on various phases of voice-frequency toll transmission maintenance. For the last eight years he has been working on the problems of intertoll trunk transmission maintenance posed by the advent of nationwide intertoll dialing with full automatic alternate routing. One angle of this work has been the development and application of statistical analyses as tools for helping to attain the required reduction in net loss variations.

Enrique A. J. Marcatili, University of Cordoba, Argentina. Mr. Marcatili was awarded the Argentine title of Aeronautical Engineer in 1947 and the title of Electrical Engineer in 1948. He received a Gold Medal from the University of Cordoba for the highest scholastic record. He joined Bell Telephone Laboratories in 1954 after studies of Cherenkov radiation in Cordoba, and has been engaged in waveguide research at Holmdel. Specifically, Mr. Marcatili has been concerned with the theory and design of filters in the millimeter region to separate channels in waveguides. He has published technical articles in Argentina and belongs to the A. F. A. (Physical Association of Argentina).

Lewis E. Miller, B.S. in Engineering Physics, Lafayette College, 1949; General Aniline and Film Corp., 1949–1952; Bell Telephone Laboratories, 1952–. Since joining the Laboratories Mr. Miller has specialized in the development of transistors. His early work was on the development for manufacture of the point-contact transistor. From 1954 to May 1956 he was concerned with surface problems and the development of germanium alloy transistors. At present he is concentrating on diffused silicon transistors. Mr. Miller is a member of the American Physical Society.

A. J. PASCARELLA, E.E., Columbia University, 1916. After his graduation he entered the student course of the General Electric Company at Schenectady. Shortly after our entrance into World War I, Mr. Pascarella joined the U. S. Navy and was put in charge of the electrical laboratory of the Gas Engine School at Columbia. In 1921 he joined the Western Electric Company and in 1925 the Technical Staff of the Laboratories. Here with the Systems Department he was concerned with the development of toll testboards, toll signaling, telegraph, carrier and miscellaneous testing equipment. Later his work consisted of formulating maintenance requirements for the over-all testing of toll lines and the detecting and location of faults on toll cables. During World War II he was concerned with developing high level auditory systems for use in psychological warfare. He also acted as editor of repair manuals used by the Armed Services. At the present time he is working on military projects. Licensed Professional Engineer, New York State.

L. G. SCHIMPF, B.E.E., Ohio State University, 1937; Bell Telephone Laboratories, 1937–. From 1937 to 1940 Mr. Schimpf was engaged in research on the application of electronic devices to switching functions, with particular emphasis on cold cathode tubes. With the outbreak of World War II, he turned his attention to research and development work on military projects. For six years after the war he specialized in transmission research studies of local subscriber station circuits and acoustics. Since 1952 he has been engaged in transistor circuit research. In this field he has concentrated particularly on the high frequency operation of transistors in transmission circuits. Senior Member of I.R.E., member of Acoustical Society of America, Eta Kappa Nu, and Tau Beta Pi.

H. F. SHOFFSTALL, B.E.E., Ohio State University, 1916; American Telephone and Telegraph Company, 1916–35; Bell Telephone Laboratories, 1935–. Mr. Shoffstall worked on the development of telephone repeaters and on toll equipment for central offices until he came to the Laboratories in 1935. Since then he has been associated with the switching development group engaged in the design of toll-switching circuits. Member of the American Institute of Electrical Engineers.

HAROLD SEIDEL, B.E.E., College of the City of New York, 1943; M.E.E., D.E.E., Polytechnic Institute of Brooklyn, 1947 and 1954. Dr. Seidel joined Bell Telephone Laboratories in 1953 after employment with the Microwave Research Institute of the Polytechnic Institute of Brooklyn, the Arma Corporation and the Federal Telecommunications Labora-

tories. His work at the Laboratories has been concerned with general electromagnetic problems, especially regarding waveguide applications, and with analysis of microwave ferrite devices. Dr. Seidel is a member of Sigma Xi and the I.R.E.

S. WEISBAUM, B.A., M.S. and Ph.D., New York University, 1947, 1948 and 1953; instructor in physics, New York University, 1950–53; Bell Telephone Laboratories, 1953–. Since joining the Laboratories, Dr. Weisbaum has specialized in the development of microwave ferrite devices, such as isolators and circulators. He is a member of the American Physical Society and Sigma Xi.