THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XXXV MARCH 1956 NUMBER 2

An Experimental Remote Controlled Line Concentrator

By. A. E. JOEL, JR.

(Manuscript received June 30, 1955)

Concentration, which is the process of connecting a number of telephone lines to a smaller number of switching paths, has always been a fundamental function in switching systems. By performing this function remotely from the central office, a new balance between outside plant and switching costs may be obtained which shows promise of providing service more economically in some situations.

The broad concept of remote line concentrators is not new. However, its solution with the new devices and techniques now available has made the possibilities of decentralization of the means for switching telephone connections very promising.

Three models of an experimental equipment have been designed and constructed for service. The models have included equipment to enable the evaluation of new procedures required by the introduction of remote line concentrators into the telephone plant. The paper discusses the philosophy, devices, and techniques.

CONTENTS

1. INTRODUCTION

The equipment which provides for the switching of telephone connections has always been located in what have been commonly called "central offices". These offices provide a means for the accumulation of all switching equipment required to handle the telephone needs of a community or a section of the community. The telephone building in which one or more central offices are located is sometimes referred to as the "wire center" because, like the spokes of a wheel, the wires which serve local telephones radiate in all directions to the telephones of the community.

A new development, made possible largely by the application of devices and techniques new to the telephone switching field, has recently been tried out in the telephone plant and promises to change much of the present conception of "central" offices and "wire" centers. It is known as a "line concentrator" and provides a means for reducing the amount of outside plant cables, poles, etc., serving a telephone central office by dispersing the switching equipment in the outside plant. It is not a new concept to reduce outside plant by bringing the switching equipment closer to the telephone customer but the technical difficulties of maintaining complex switching equipment and the cost of controlling such equipment at a distance have in the past been formidable obstacles to the development of line concentrators. With the invention of low power, small-sized, long-life devices such as transistors, gas tubes, and sealed relays, and their application to line concentrators, and with the development of new local switching systems with greater flexibility, it has been possible to make the progress described herein.

2. OBJECTIVES

Within the telephone offices the first switching equipment through which dial lines originate calls concentrates the traffic to the remaining equipment which is engineered to handle the peak busy hour load with the appropriate grade of service.[1] This concentration stage is different for different switching systems. In the step-by-step system[2] it is the line finder, and in the crossbar systems it is the primary line switch.[3] Proposals for the application of remote line concentrators in the step-by-step system date back over 50 years.[4] Continuing studies over the years have not indicated that any appreciable savings could be realized when such equipment is used within the local area served by a switching center.

When telephone customers move from one location to another within a local service area, it is desirable to retain the same telephone numbers. The step-by-step switching system in general is a unilateral arrangement where each line has two appearances in the switching equipment, one for originating call concentration (the line finder) and one for selection of the line on terminating calls (the connector). The connector fixes the line number and telephone numbers cannot be readily reassigned when moving these switching stages to out-of-office locations.

Common-control systems[5] have been designed with flexibility so that the line number assignments on the switching equipment are independent of the telephone numbers. Furthermore, the first switching stage in the office is bilateral, handling both originating and terminating calls through the same facilities. The most recent common-control switching system in use in the Bell System, the No. 5 crossbar,[3] has the further advantage of universal control circuitry for handling originating and terminating calls through the line switches. For these reasons, the No. 5 crossbar system was chosen for the first attempt to employ new techniques of achieving an economical remote line concentrator.

A number of assumptions were made in setting the design requirements. Some of these are influenced by the characteristics of the No. 5 crossbar system. These assumptions are as follows:

1. No change in customer station apparatus. Standard dial telephones to be used with present impedance levels, transmission characteristics, dial pulsing, party identification, superimposed ac-dc ringing,[6] and signaling and talking ranges.

2. Individual and two-party (full or semi-selective ringing) stations to be served but not coin or PBX lines.

3. Low cost could best be obtained by minimizing the per line equipment in the central office. AMA[7] charging facilities could be used but to avoid per station equipment in the central office no message register operation would be provided.

4. Each concentrator would serve up to 50 lines with the central office control circuits common to a number of concentrators. (Experimental equipment described herein was designed for 60 lines to provide additional facilities for field trial purposes.) No extensive change would be made in central office equipment not associated with the line switches nor should concentrator design decrease call carrying capacities of existing central office equipment.

5. To provide data to evaluate service performance, automatic traffic recording facilities to be integrated with the design.

6. Remote equipment designed for pole or wall mounting as an addition to existing outside plant. Therefore, terminal distribution facilities would not be provided in the same cabinet.

7. Power to be supplied from the central office to insure continuity of telephone service in the event of a local power failure.

8. Concentrators to operate over existing types of exchange area facilities without change and with no decrease in station to central office service range.

9. Maintenance effort to be facilitated by plug-in unit design using the most reliable devices obtainable.

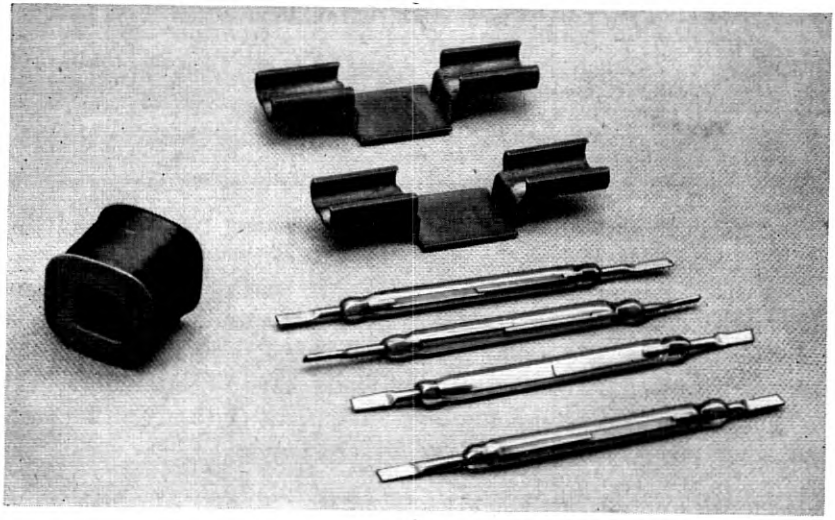
3. NEW DEVICES EMPLOYED

Numerous products of research and development were available for this new approach. Only those chosen will be described.

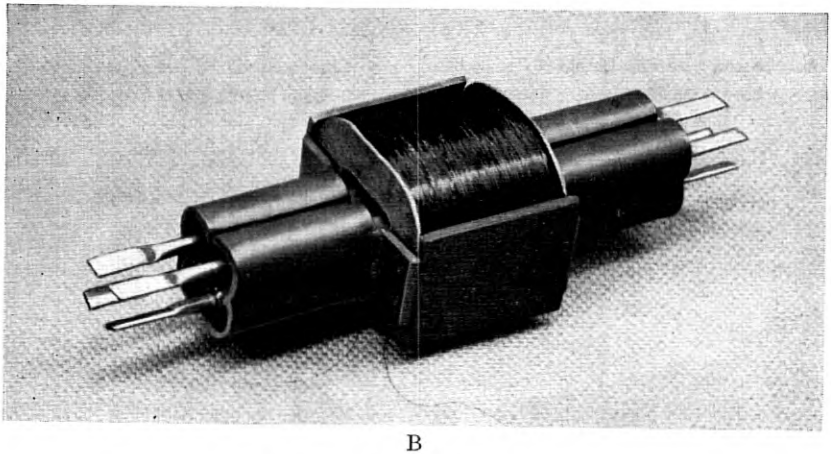
For the switching or "crosspoint" element itself, the sealed reed switch was chosen, primarily because of its imperviousness to dirt.[8] A short coil magnet with magnetic shield for increasing sensitivity of the reed switches were used to form a relay per crosspoint (see Fig. 1).

A number of switching applications[9,10] for crosspoint control using small gas diodes have been proposed by E. Bruce of our Switching Research Department. They are particularly advantageous when used in an "end marking" arrangement with reed relay crosspoints. Also, these diodes have long life and are low in cost. One gas diode is employed for operating each crosspoint (see Fig. 6). Its breakdown voltage is 125v ± 10v. A different tube is used in the concentrator for detecting marking potentials when termination occurs. Its breakdown potential is 100v ± 10v. One of these tubes is used on each connection.

Signaling between the remote concentrator and the central office control circuits is performed on a sequential basis with pulses indicative of the various line conditions being transmitted at a 500 cycle rate. This frequency encounters relatively low attenuation on existing exchange area wire facilities and yet is high enough to transmit and receive information at a rate which will not decrease call carrying capacity of the



A



B

Fig. 1 — Reed switch relay.

central office equipment. To accomplish this signaling and to process the information economically transistors appear most promising.

Germanium alloy junction transistors were chosen because of their improved characteristics, reliability, low power requirements, and margins, particularly when used to operate with relays.[13] Both N-P-N and P-N-P transistors are used. High temperature characteristics are particularly important because of the ambient conditions which obtain on pole mounted equipment. As the trials of this equipment have progressed,

TABLE I—TRANSISTOR CHARACTERISTICS

Code No.	Type and Filling	Alpha	Max. I_{co} at 28V and 65°C	Emitter Zener Voltage at 20μa
M1868	p-n-p Oxygen	0.9–1.0	150 μa	>735
M1887	n-p-n Vacuum	0.5– .75	100 μa	>735

considerable progress has been made in improving transistors of this type. Table I summarizes the characteristics of these transistors.

For directing and analyzing the pulses, the control employs semiconductor diode gate circuits.[11] The semiconductor diodes used in these circuits are of the silicon alloy junction type.[15] Except for a few diodes operating in the gas tube circuits most diodes have a breakdown voltage requirement of 27v, a minimum forward current of 15 ma at 2v and a maximum reverse current at 22v of 2×10^{-8} amp.

4. NEW TECHNIQUES EMPLOYED

The concentrator represents the first field application in Bell System telephone switching systems which departs from current practices and techniques. These include:

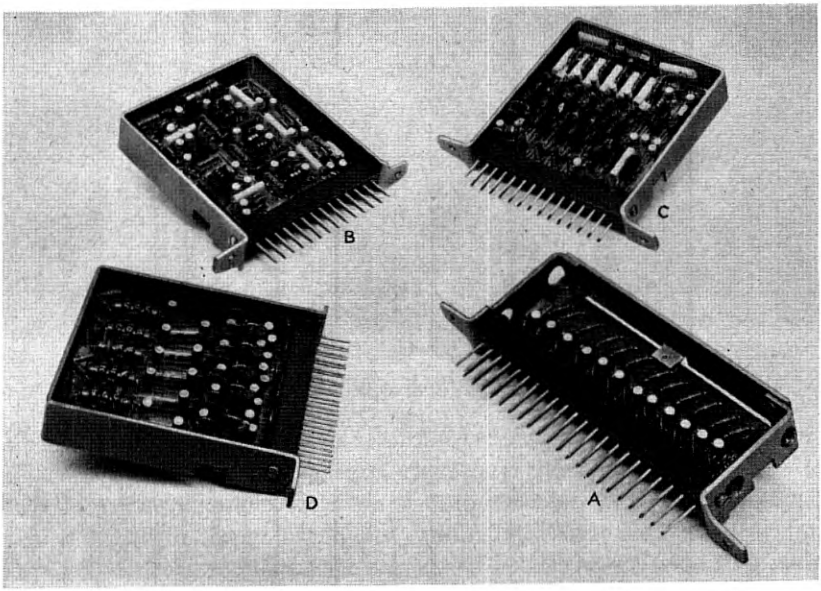


Fig. 2 — Transistor packages. (a) Diode unit. (b) Transistor counter. (c) Transistor amplifiers and bi-stable circuits. (d) Five trunk unit.

1. High speed pulsing (500 pulses per second) of information between switching units.

2. The use of plug-in packages employing printed wiring and encapsulation. (Fig. 2 shows a representative group of these units.)

3. Line scanning for supervision with a passive line circuit. In present systems each line is equipped with a relay circuit for detecting call originations (service requests) and another relay (or switch magnet) for indicating the busy or idle condition of the line, as shown in Fig. 3(a). The line concentrator utilizes a circuit consisting of resistors and semiconductor diodes in pulse gates to provide these same indications. This circuit is shown in Fig. 3(b). Its operation is described later. The pulses for each line appear at a different time with respect to one another. These pulses are said to represent "time slots." Thus a different line is examined each .002 second for a total cycle time (for 60 lines) of .120 second. This process is known as "line scanning" and the portion of the circuit which produces these pulses is known as the scanner. Each of the circuits perform the same functions, viz., to indicate to the central office equipment when the customer originates a call and for terminating calls to indicate if the line is busy.

4. The lines are divided for control and identification purposes into twelve groups of five lines each. Each group of five lines has a different pattern of access to the trunks which connect to the central office. The ten trunks to the central office are divided into two groups as shown in Fig. 4. One trunk group, called the random access group, is arranged in a random multiple fashion, so that each of these trunks is available to approximately one-half of the lines. The other group, consisting of two trunks, is available to all lines and is therefore called the full access group. The control circuitry is arranged to first select a trunk of the random access group which is idle and available to the particular line to which a connection is to be made. If all of the trunks of this random access group are busy to a line to which a connection is desired, an attempt is then made to select a trunk of the full access group. The preference order for selecting cross-points in the random access group is different for each line group, as shown in the table on Fig. 4. By this means, each trunk serves a number of lines on a different priority basis. Random access is used to reduce by 40 per cent the number of individual reed relay crosspoints which would otherwise be needed to maintain the quality of service desired, as indicated by a theory presented some years ago.[12]

5. Built-in magnetic tape means for recording usage data and making call delay measurements. The gathering of this data is greatly facilitated by the line scanning technique.

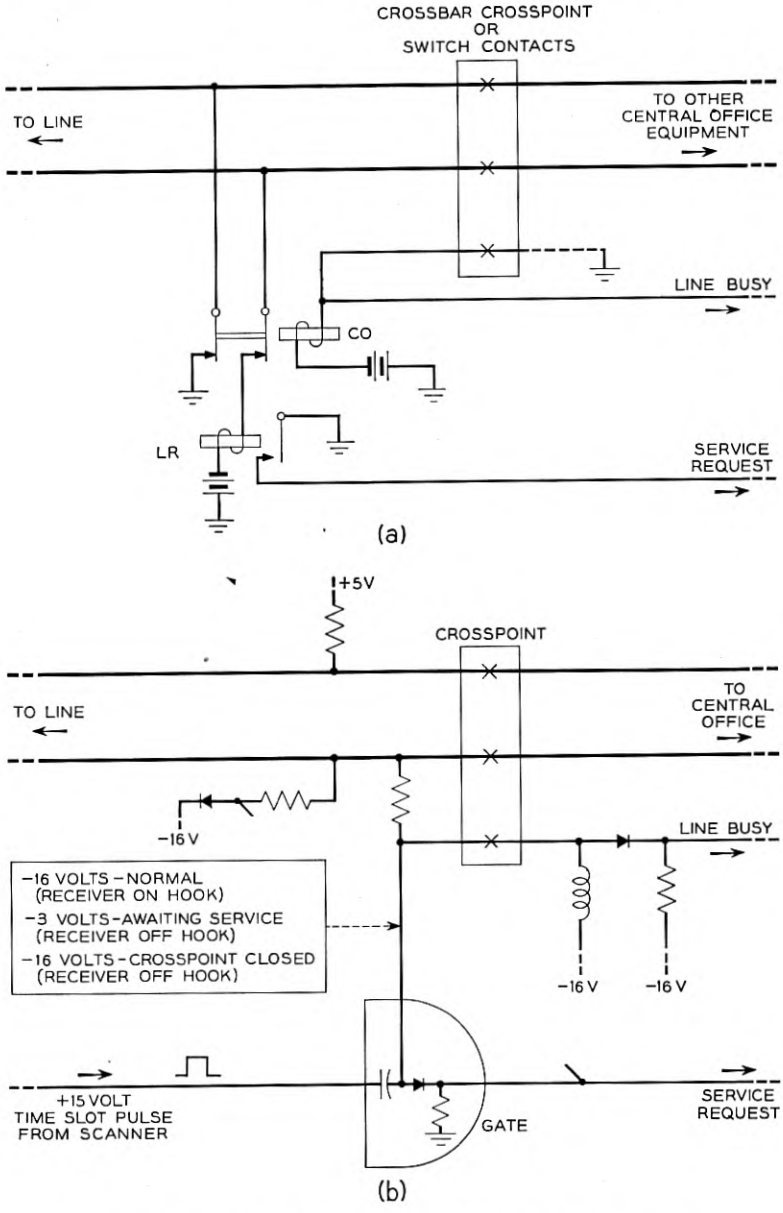


Fig. 3 — (a) Relay line circuit. (b) Passive line circuit.

5. SWITCHING PLAN

The plan for serving lines directly terminating in a No. 5 Crossbar office is shown in Fig. 5(a). Each line has access through a primary line switch to 10 line links. The line links couple the primary and secondary switches together so that each line has access to all of the 100 junctors to the trunk link switching stage. Each primary line switch group accommodates from 19 to 59 lines (one line terminal being reserved for no-test calls). A line link frame contains 10 groups of primary line switches.[14]

The remote concentrator plan merely extends these line links as trunks to the remote location. However, an extra crossbar switching stage is introduced in the central office to connect the links to the secondary line switches with the concentrator trunks as shown in Fig. 5(b). Since each line does not have full access to the trunks, the path chosen by the marker to complete calls through the trunk link frame may then be independent of the selection of a concentrator trunk with access to the line. This arrangement minimizes call blocking, simplifies the selection of a matched path by the marker, and the additional crossbar switch hold magnet serves also as a supervisory relay to initiate the transmission of disconnect signals over the trunk.

In addition to the 10 concentrator trunks used for talking paths, 2 additional cable pairs are provided from each concentrator to the central office for signaling and power supply purposes. The use of these two pairs of control conductors is described in detail in Section 6g.

The concentrator acts as a slave unit under complete control of the central office. The line busy and service request signals originate at the

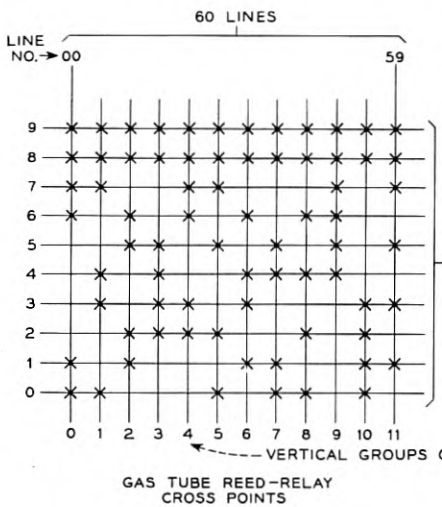


Fig 4. — Concentrator trunk to line crosspoint pattern and preference order

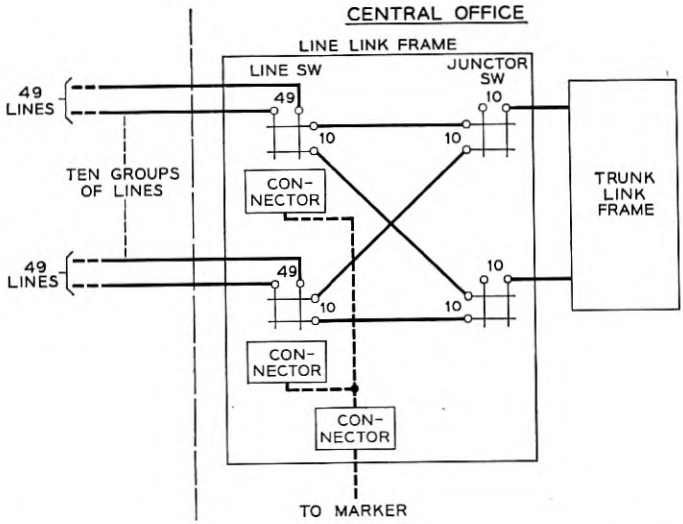


Fig. 5(a) — No. 5 crossbar system subscriber lines connected to line link frame.

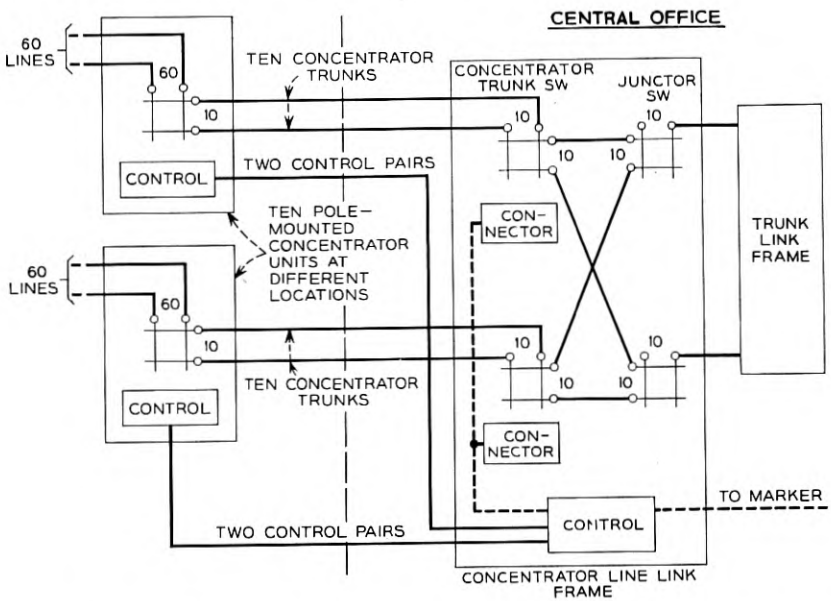


Fig. 5(b) — No. 5 crossbar system subscriber lines connected to remote line concentrators.

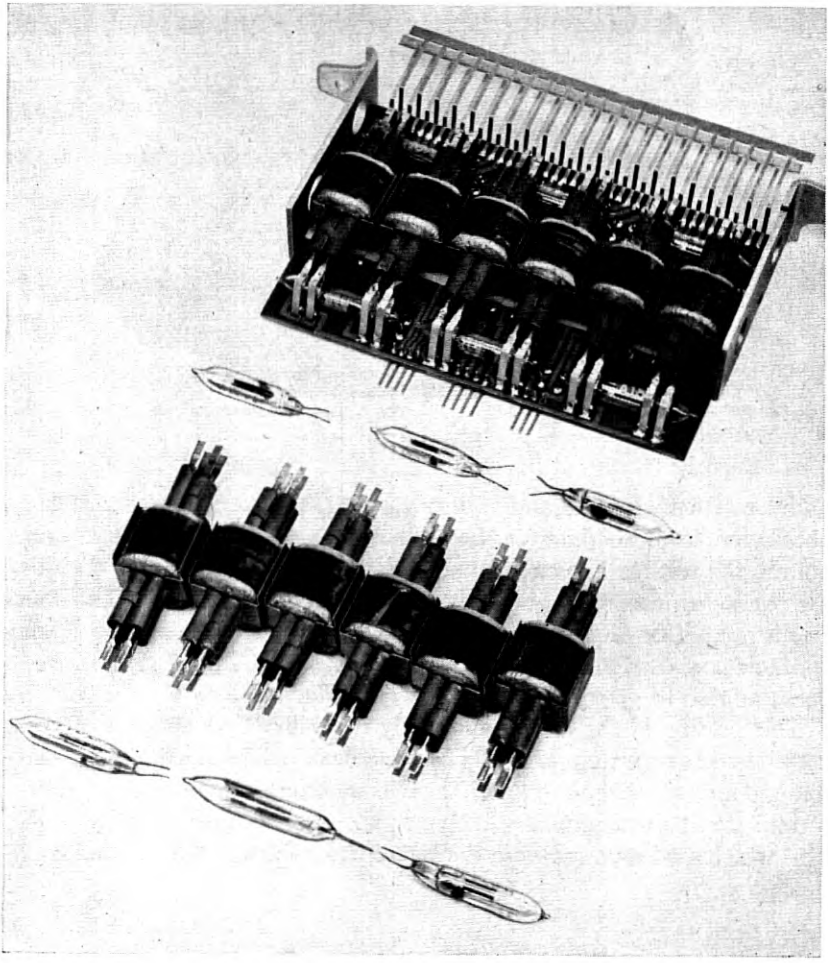


Fig. 6 — Line unit construction.

concentrator only in response to a pulse in the associated time slot or when a crosspoint operates (a line busy pulse is generated under this condition as a crosspoint closure check). The control circuit in the central office is designed to serve 10 remote line concentrators connected to a single line link frame. In this way the marker deals with a concentrator line link frame as it would with a regular line link frame and the marker modifications are minimized.

The traffic loading of the concentrator is accomplished by fixing the

Fig. 7(a) — Line unit.

number of trunks at 10 and equipping or reassigning lines as needed to obtain the trunk loading for the desired grade of service. The six crosspoints, the passive line circuit and scanner gates individual to each line are packaged in one plug-in unit to facilitate administration. The crosspoints are placed on a printed wiring board together with a comb of plug contacts as shown in Fig. 6. The entire unit is then dipped in rubber and encapsulated in epoxy resin, as shown in Fig. 7(a).

This portion of the unit is extremely reliable and therefore it may be considered as expendable, should a rare case of trouble occur. The passive line circuit and scanner gate circuit elements are mounted on a smaller second printed wiring plate (known as the "line scanner" plate, see Fig. 7(b) which fits into a recess in the top of the encapsulated line unit. Cir-

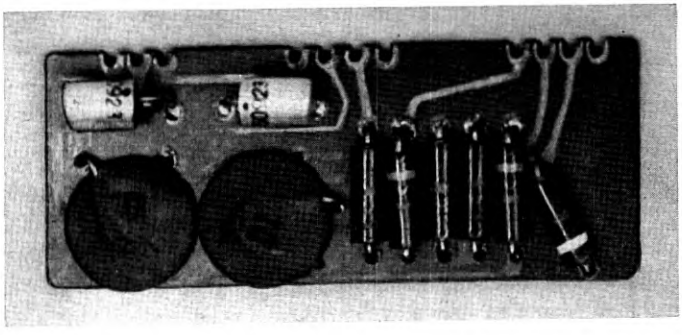


Fig. 7(b) — Scanner plate of the line unit shown in Fig. 7 (a).

cuit connection between printed wiring plates is through pins which appear in the recess and to which the smaller plate is soldered.

6. BASIC CIRCUITS

a. *Diode Gates*

All high speed signaling is on a pulse basis. Each pulse is positive and approximately 15 volts in amplitude. There is one basic type of diode gate circuit used in this equipment. By using the two resistors, one condenser and one silicon alloy junction diode in the gate configuration shown in Fig. 8, the equivalents of opened or closed contacts in relay circuits are obtained. These configurations are known respectively as enabling and inhibiting gates and are shown with their relay equivalents in Figs. 8(a) and 8(b).

In the enabling gate the diode is normally back biased by more than the pulse voltage. Therefore pulses are not transmitted. To enable or

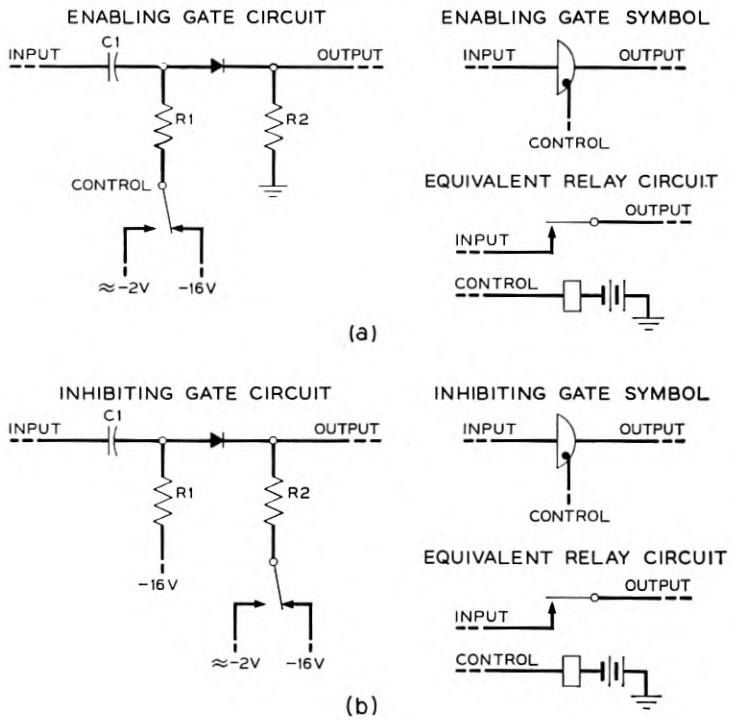


Fig. 8 — Gates and relay equivalents.

open the gate the back bias is reduced to a small reverse voltage which is more than overcome by the signal pulse amplitude of the pulse. The pulse thus forward biases the diode and is transmitted to the output.

The inhibiting gate has its diode normally in the conducting state so that a pulse is readily transmitted from input to output. When the bias is changed the diode is heavily back biased so that the pulse amplitude is insufficient to overcome this bias.

The elements of 12 gates are mounted on a single printed wiring board with plug-in terminals and a metal enclosure as shown in Fig. 2(a). All elements are mounted in one side of the board so that the opposite side may be solder dipped. After soldering the entire unit (except the plug) is dipped in a silicone varnish for moisture protection.

b. *Transistor Bistable Circuit*

Transistors are inherently well adapted to switching circuits using but two states, on (saturated) or off.[16] In these circuits with a current gain greater than unity a negative resistance collector characteristic can be obtained which will enable the transistor to remain locked in its conducting state (high collector current flowing) until turned off (no collector current) by an unlocking pulse. At the time the concentrator development started only point contact transistors were available in quantity. Point contact transistors have inherently high current gains (>1) but the collector current flowing when in the normal or unlocked condition (I_{co}) was so great that at high ambient temperatures a relay once operated in the collector circuit would not release.

Junction transistors are capable of a much greater ratio of on to off current in the collector circuit. Furthermore their characteristics are amenable to theoretical design consideration.[13] However, the alpha of a simple junction transitor is less than unity. To utilize them as one would a point contact transitor in a negative resistance switching circuit, a combination of n-p-n and p-n-p junction transistors may be employed, see Fig. 9(b). Two transistors combined in this manner constitute a "hooked junction conjugate pairs." This form of bi-stable circuit was used because it requires fewer components and uses less power than an Eccles-Jordan bistable circuit arrangement. It has the disadvantage of a single output but this was not found to be a shortcoming in the design of circuits employing pulse gates of the type described. In what follows the electrodes of the transistor will be considered as their equivalents shown in Fig. 9(b).

The basic bi-stable circuit employed is shown in Fig. 10. The set

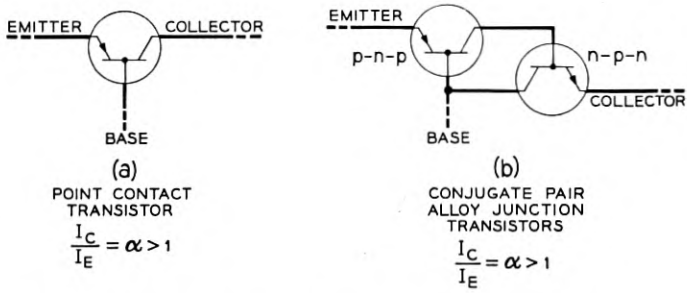


Fig. 9 — Point contact versus hooked conjugate pair.

pulse is fed into the emitter (of the pair) causing the emitter diode to conduct. The base potential is increased thus increasing the current flowing in the collector circuit. When the input pulse is turned off the base is left at about −2 volts thus maintaining the emitter diode conducting and continuing the increased current flow in the collector circuit. The diode in the collector circuit prevents the collector from going positive and thereby limits the current in the collector circuit. To reset, a positive pulse is fed into the base through a pulse gate. The driving of the base positive returns the transistor pair to the off condition.

c. *Transistor Pulse Amplifier*

This circuit (Fig. 11) is formed by making a bi-stable self resetting circuit. It is used to produce a pulse of fixed duration in response to a

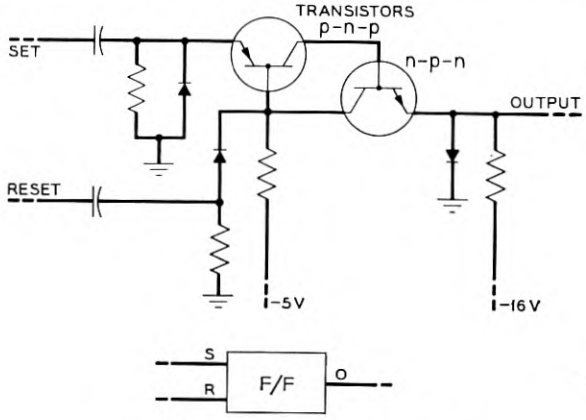


Fig. 10 — Transistor bi-stable circuit.

pulse of variable width (within limits) on the input. Normally the emitter is held slightly negative with respect to the base. The potential difference determines the sensitivity of the amplifier. When a positive input pulse is received, the emitter diode conducts causing an increase in collector current. The change in bias of the diode in the emitter circuit permits it to conduct and charge the condenser. With the removal of the input pulse the discharge of the condenser holds the transistor pair on. The time constant of the circuit determines the on time. When the emitter potential falls below the base potential, the transistor pair is turned off.

The amplifiers and bi-stable circuits or flip-flops, as they are called more frequently, are mounted together in plug-in packages. Each package contains 8 basic circuits divided 7-1, 6-2, or 2-6, between amplifiers and flip-flops. Fig. 2(c) shows one of these packages. They are smaller than the gate or line unit packages, having only 28 terminals instead of 42.

The transistors for the field trial model were plugged into small hearing aid sockets mounted on the printed wiring boards. For a production model it would be expected that the transistors would be soldered in.

d. *Transistor Ring Counter*

By combining bi-stable transistor and diode pulse gate circuits together in the manner shown in Fig. 12 a ring counter may be made, with

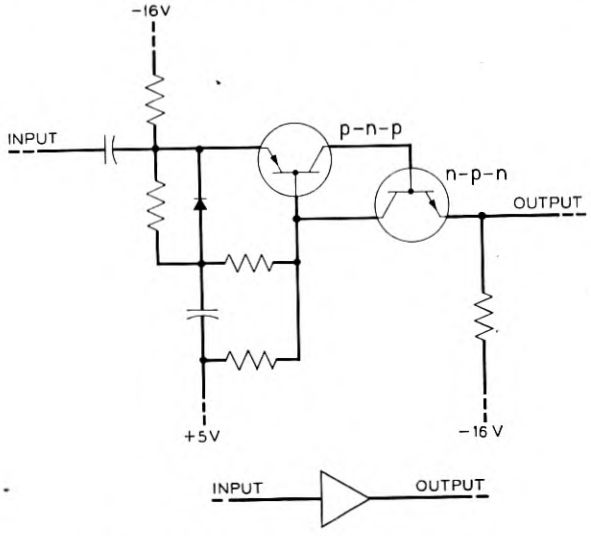


Fig. 11 — Transistor pulse amplifier.

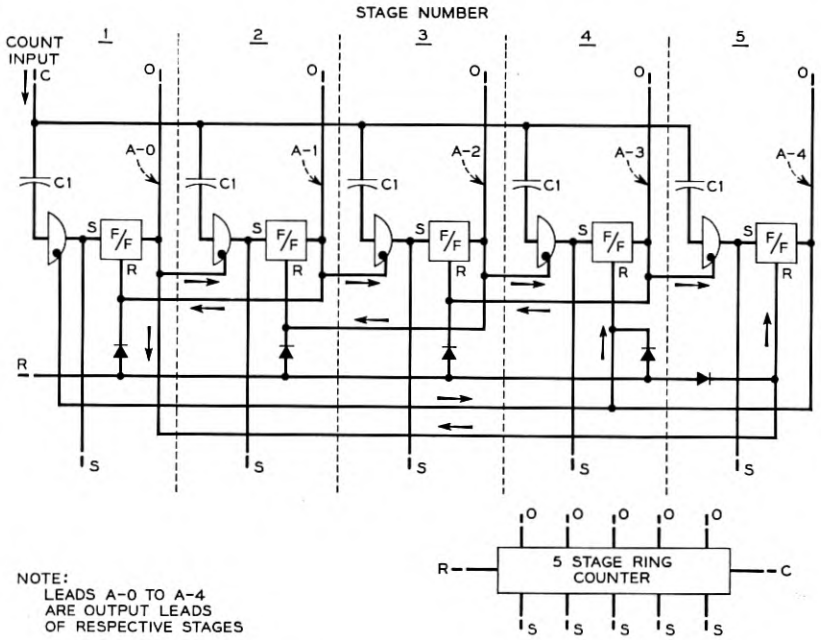


Fig. 12 — Ring counter schematic.

a bi-stable circuit per stage. The enabling gate for a stage is controlled by the preceding stage allowing it to be set by an input advance pulse. The output signal from a stage is fed back to the preceding stage to turn it off. An additional diode is connected to the base of each stage for resetting when returning the counter to a fixed reference stage.

A basic package of 5 ring counter stages is made up in the same framework and with the same size plug as the flip-flop and amplifier packages, see Fig. 2(b). A four stage ring counter is also used and is the same package with the components for one stage omitted. The input and output terminals of all stages are available on the plug terminals so that the stages may be connected in any combination and form rings of more than 5 stages. The reset lead is connected to all but the one stage which is considered the first or normal stage.

Other transistor circuits such as binary counters and square wave generators are used in small quantity in the central office equipment. They will not be described.

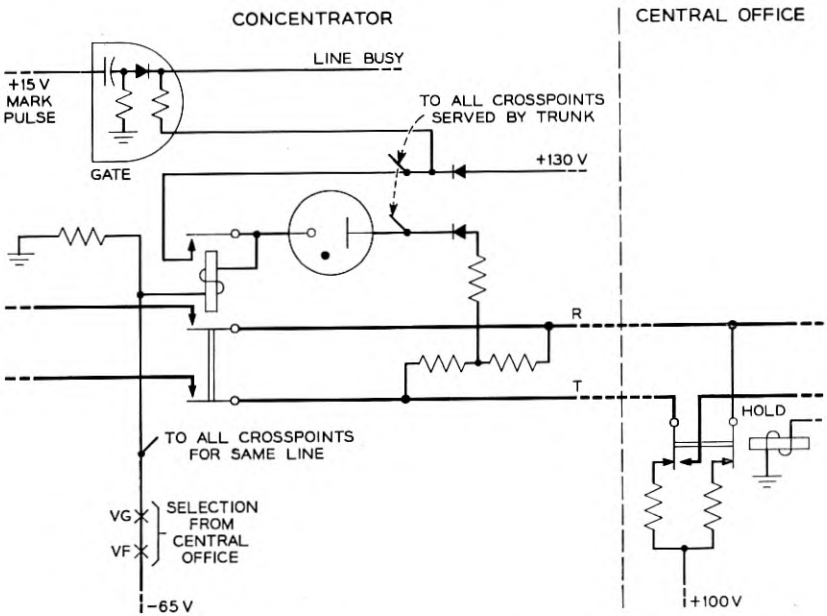


Fig. 13 — Crosspoint operating circuit.

e. *Crosspoint Operating Circuit*

The crosspoint consists of a reed relay with 4 reed switches and a gas diode (Fig. 1). The selection of a crosspoint is accomplished by marking with a negative potential (−65 volts) all crosspoints associated with a line, and marking with a positive potential (+100 volts) all crosspoints associated with a trunk (Fig. 13). The line is marked through a relay circuit set by signals sent over the control pair from the central office. The trunk is marked by a simplex circuit connected through the break contacts of the hold magnet of the crossbar switch associated with the trunk in the central office. Only one crosspoint at a time is exposed to 165 volts which is necessary and sufficient to break down the gas diode to its conducting state. The reed relay operates in series with the gas diode. A contact on the relay shunts out the gas diode. When the marking potentials are removed the relay remains energized in a local 30-volt circuit at the concentrator. The holding current is approximately 2.5 ma.

This circuit is designed so that ringing signals in the presence or absence of line marks will not falsely fire a crosspoint diode. Furthermore,

a line or trunk mark alone should not be able to fire a crosspoint diode on a busy line or trunk.

When the crosspoint operates, a gate which has been inhibiting pulses is forward biased by the −65 volt signal through the crosspoint relay winding. The pulse which initiates the mark operations at the concentrator then passes through the gate to return a line busy signal to the central office over this control pairs which is interpreted as a crosspoint closure check signal.

f. *Crosspoint Release Circuit*

The hold magnet of the central office crossbar switch operates, removing the +100-volt operate mark signal after the crosspoint check signal is received. A slow release relay per trunk is operated directly by the hold magnet. When the central office connection in the No. 5 crossbar system releases, the hold magnet is released. As shown in Fig. 14, with the hold magnet released and the slow release relay still operated, a −130-volt signal is applied in a simplex circuit to the trunk to break down a gas tube provided in the trunk circuit at the concentrator. This tube in

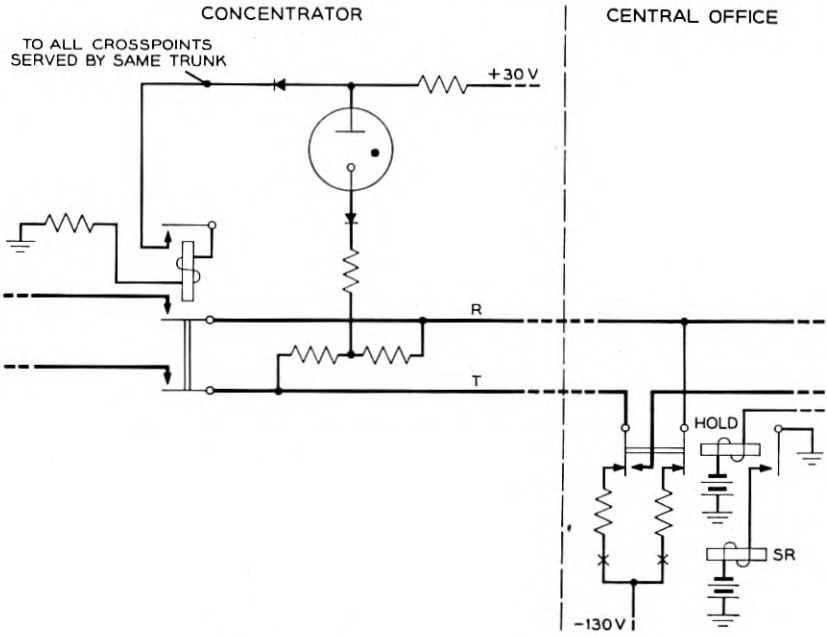


Fig. 14 — Crosspoint release circuit.

breaking down shunts the local holding circuit of the crosspoint causing it to release. The −130-volt disconnect signal is applied during the release time of the slow release relay which is long enough to insure the release of the crosspoint relay at the concentrator.

The release circuit is individual to the trunk and independent of the signal sent over the control pairs.

g. *Pulse Signalling Circuits*

To control the concentrator four distinct pulse signals are transmitted from the central office. Two of these at times must be transmitted simultaneously, but these and the other two are transmitted mutually exclusively. In addition, service request and line busy signals are transmitted from the concentrator to the central office. The two way transmission of information is accomplished on each pair by sending signals in each direction at different times and inhibiting the receipt of signals when others are being transmitted.

To transmit four signals over two such pairs, both positive and nega-

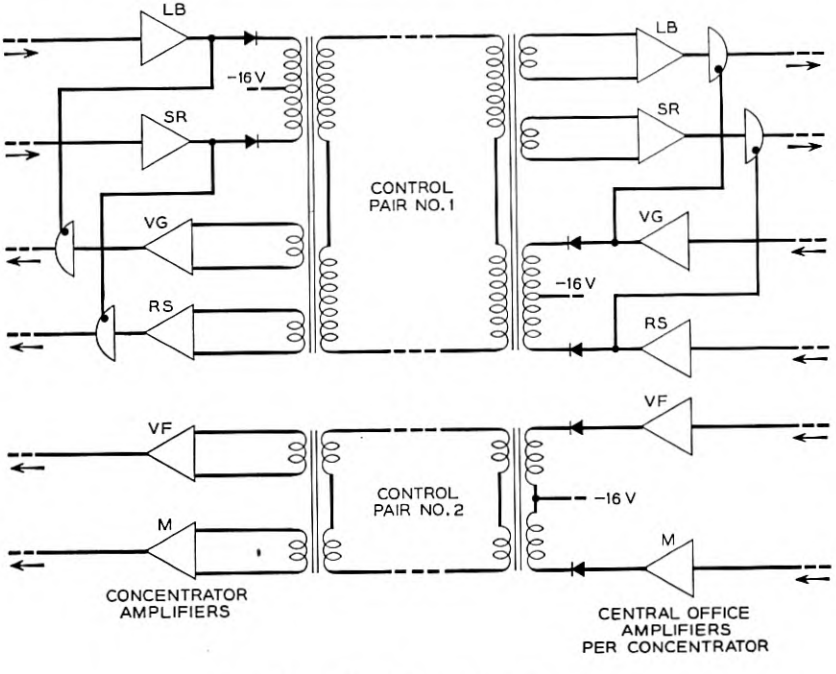


Fig. 15. — Signal transmission circuit.

tive pulses are employed. Diodes are placed in the legs of a center tapped transformer, as shown in Fig. 15, to select the polarity of the transmitted pulses. At the receiving end the desired polarity is detected by taking the signal as a positive pulse from a properly poled winding of a transformer. The amplifier, as described in Section 6c responds only to positive pulses. If pulses of the same polarity are transmitted in the other direction over the same pair, as for control pair No. 1, the outputs of the receiving amplifier for the same polarity pulse are inhibited whenever a pulse is transmitted.

As shown in Fig. 15, the service request and line busy signals are transmitted from the concentrator to the central office over one pair of conductors as positive and negative pulses respectively. The transmission of these pulses gates the outputs of two of the receiving amplifiers at the concentrator to permit the receipt of the polarized signals from the central office. This prevents the pulses from being used at the sending end. A similar gating arrangement is used with respect to the signals when sent over this control pair from the central office. The pulses designated VG or RS never occur when a pulse designated SR or LB is sent in the opposite direction. The transmission of the VF pulse over control pair No. 2 is processed by the concentrator circuit and becomes the SR or LB pulses. In section 7 the purpose of these pulses is described.

The signaling range objective is 1,200 ohms over regular exchange area cable including loaded facilities from station to central office.

h. *Power Supply*

Alternating current is supplied to the concentrator from a continuous service bus in the central office. The power supply path is a phantom circuit on the two control pairs as shown in Fig. 16. The power transformer has four secondary windings used for deriving from bridge rectifiers four basic dc voltages. These voltages and their uses are as follows: −16 volts (regulated) for transistor collector circuits and gate biases, +5 volts (regulated) for transistor base biases, +30 volts (regulated) for crosspoints holding circuits and −65 volts for the marking and operating of the line crosspoints. For this latter function a reference to the central office applied +100 volt trunk mark is necessary. The reference ground for the concentrator is derived from ground applied to a simplex circuit on the power supply phantom circuit. Series transistors and shunt silicon diodes with fixed reference breakdown voltages are used to regulate dc voltages.

Total power consumption of the concentrator is between 5 and 8 watts depending upon the number of connections being held.

7. CONCENTRATOR OPERATION

a. *Line Scanning*

The sixty lines are divided into 12 groups of 5 lines each. These groupings are designated VG and VF respectively corresponding to the vertical group and file designations used in the No. 5 crossbar system. Each concentrator corresponds to a horizontal group in that system.

To scan the lines two transistor ring counters, one of 12 stages and one of 5 stages, are employed as shown in Fig. 17. These counters are driven from pulses supplied from the central office control circuits and only one stage in each is on at any one time. The steps and combinations of these counters correspond to the group and file designation of a particular line. Each 0.002 second the five stage counter (VF) takes a step and between the fifth and sixth pulse the 12-stage counter (VG) is stepped. Thus the 5-stage counter receives 60 pulses or re-cycles 12 times in 120 milliseconds while the 12-stage counter cycles but once.

Each line is provided with a scanner gate. The collector output of each each stage of the VG counter biases this gate to enable pulses which are generated by the collector circuit of the 5-stage counter to pass on

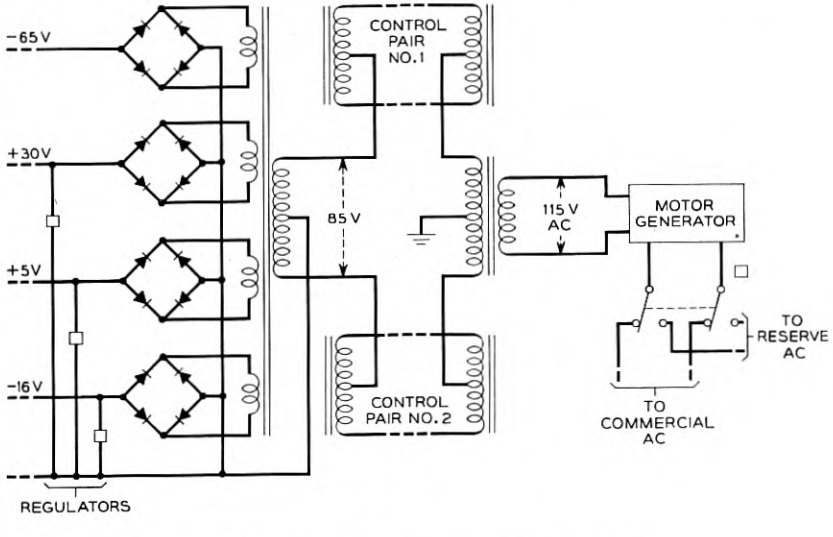


Fig. 16 — Power supply transmission circuit.

to the gate of the passive line circuit, Fig. 3(b). If the line is idle the pulses are inhibited. If the receiver is off-hook requesting service (no crosspoint closed) then the gate is enabled, the pulse passes to the service request amplifier and back to the central office in the same time slot as the pulse which stepped the VF counter. If the line has a receiver off-hook and is connected to a trunk the pulse passes through a contact of the crosspoint relay to the line busy amplifier and then to the central office in the same time slot.

At the end of each complete cycle a reset pulse is sent from the central office. This pulse instead of the VG pulse places the 12-stage counter in its first position. It also repulses the 5 stage VF counter to its fifth stage so that the next VF pulse will turn on its first stage to start the next cycle. The reset pulse insures that, in event of a lost pulse or defect in a counter stage, the concentrator will attempt to give continuous service without dependence on maintaining synchronism with the central office scanner pulse generator. Fig. 18(a) shows the normal sequence of line scanning pulses.

When a service request pulse is generated, the central office circuits

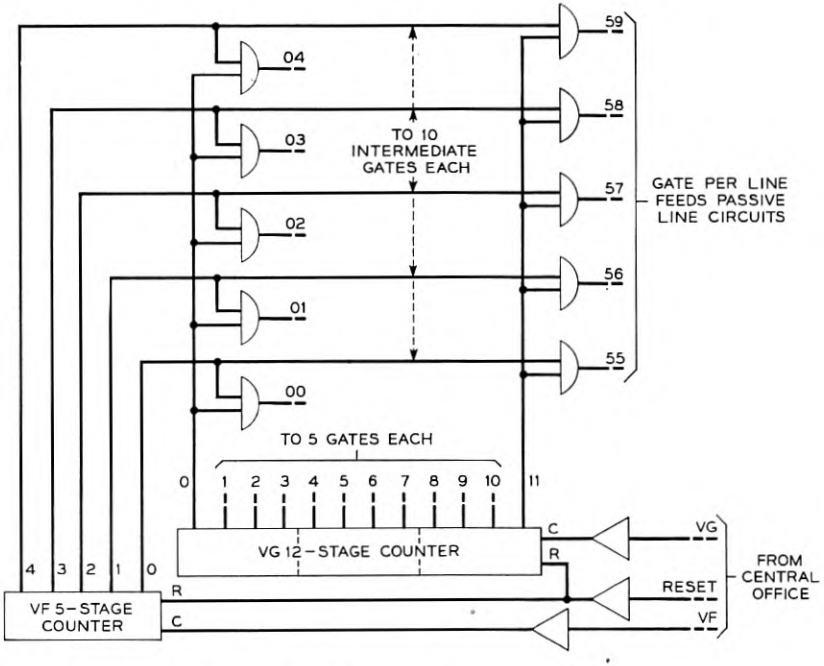


Fig. 17 — Diode matrix for scanning lines.

common to 10 concentrators interrupt the further transmission of the vertical group pulse so that the line scanning is confined to the 5 lines in the vertical group in which the call originated. In this way the central office will receive a service request pulse at least every 0.010 sec as a check that the call has not been abandoned while awaiting service. Fig. 18(b) shows the detection of a call origination and the several short scan cycles for abandoned call detection.

b. *Line Selection*

When the central office is ready to establish a connection at the concentrator a reset pulse is sent to return the counters to normal. In general, the vertical group and vertical file pulses are sent simultaneously to reduce holding time of the central office equipment and to minimize marker delays caused by this operation. For this reason the VG and VF pulses are each transmitted over different control pairs from the central office. The same polarity is used.

On originating calls it is desirable to make one last check that the call has not been abandoned, while on terminating calls it is necessary

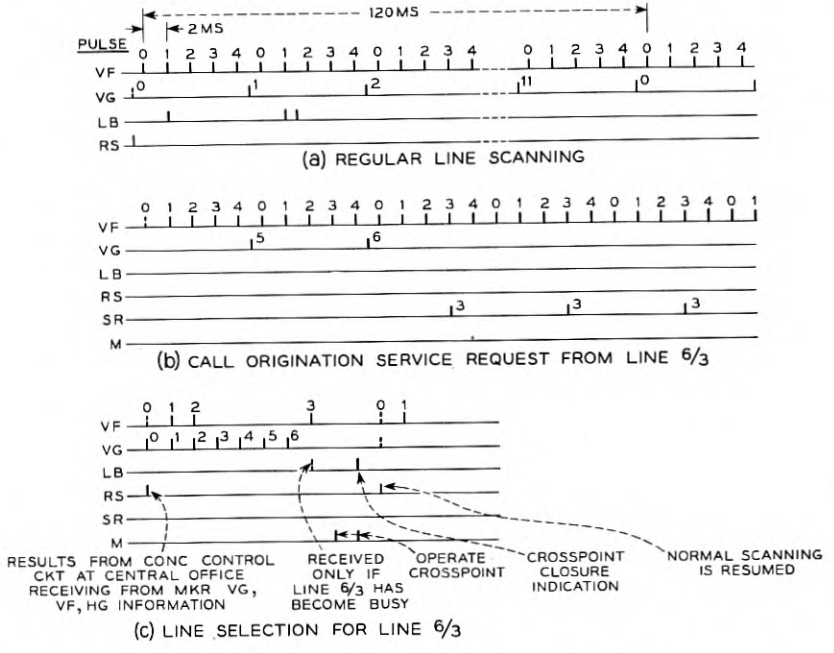


Fig. 18 — Pulse sequences. (a) Regular. (b) Call origination. (c) Line selection.

to determine if the line is busy or idle. These conditions are determined in the same manner as described for line scanning since a service request condition would still prevail on the line if the call was not abandoned. If the line was busy, a line busy condition would be detected. However to detect these conditions a VF pulse must be the last pulse transmitted since the stepping of the VF counter generates the pulse which is transmitted through an enabled line selection and passive line circuit gates. Fig. 18(c) shows a typical line selection where the number of VF pulses is equal to or less than the number of VG pulses. In all other cases there is no conflict and the sending of the last VF pulse need not be delayed. On terminating calls, the line busy indication is returned to the central office within 0.002 sec after the selection is complete. During selections the central office circuits are gated to ignore any extraneous service request or line busy pulses produced as a result of steps of the VF counter prior to its last step.

c. *Crosspoint Operation and Check*

Associated with each concentrator transistor counter stage is a reed relay. These relays are connected to the transistor collector circuits through diodes of the counter stages when relay M operates. The contacts of these reed relays are arranged in a selection circuit as shown in Fig. 19 and apply the −65 volt mark potential to the crosspoint relays of the selected line.

After a selection is made as described above a "mark" pulse is sent from the central office. This pulse is transmitted as a pulse of a different polarity over the same control pair as the VF pulses. The received pulse after amplification actuates a transistor bistable circuit which has the M reed relay permanently connected in its collector circuit. The bi-stable circuit holds the M relay operated during the crosspoint operation to maintain one VF and one VG relay operated, thereby applying −65 volts to mark and operate one of the 6 crosspoint relays of the selected line as described in section 6e, and shown on Fig. 13.

The operation and locking of the crosspoint relay with the marking potentials still applied enables a pulse gate associated with the holding circuit of the crosspoint relays in each trunk circuit. The mark pulses are sent out continuously. This does not affect the bi-stable transistor circuit once it has triggered but the mark pulse is transmitted through the enabled crosspoint closure check gate shown in Fig. 20 and back to the central office as a line busy signal.

With the receipt of the crosspoint closure check signal the sending

of the mark pulses is stopped and a reset pulse is sent to the concentrator to return the mark bi-stable circuit, counters and all operated selector relays to normal. The concentrator remains in this condition until it is resynchronized with the regular line scanning cycle.

A complete functional schematic of the concentrator integrating the circuits described above is shown in Fig. 21. Fig. 22(a) and (b) show an experimental concentrator built for field tests.

8. CENTRAL OFFICE CIRCUITS

The central office circuits for controling one or more concentrators are composed of wire spring relays as well as transistors, diode and reed

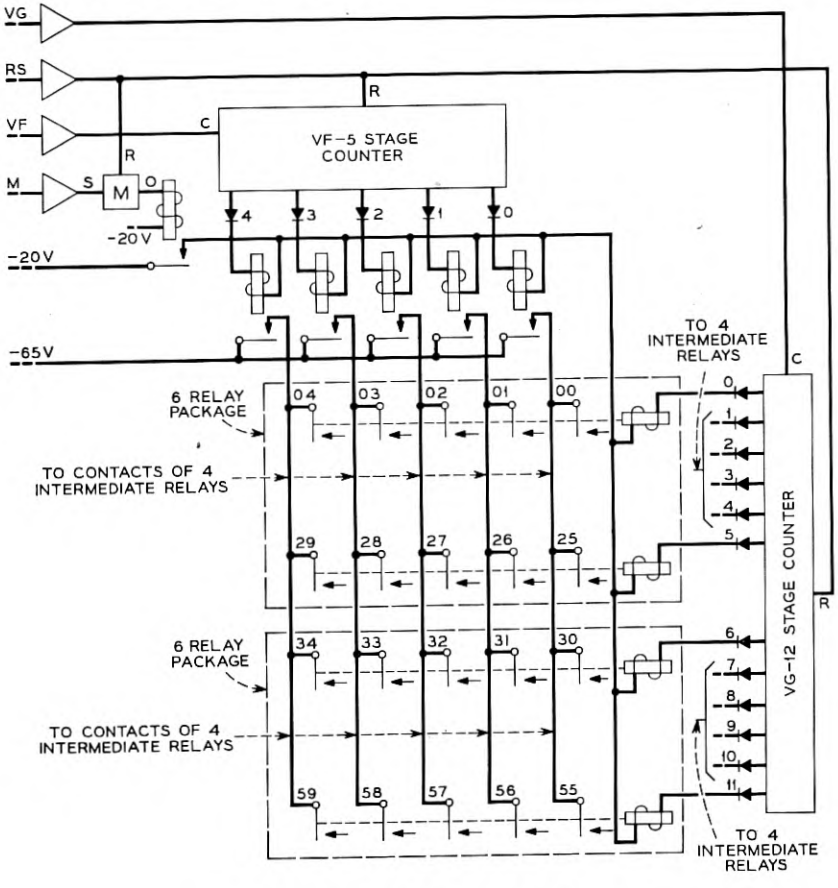


Fig. 19 — Line selection and marking.

relay packages similar to those used in the concentrator. The reed relays are energized by transistor bi-stable circuits in the same manner as described in Section 7c. The reed relay contacts in turn operate wire spring relays or send the dc signals directly to the regular No. 5 crossbar marker and line link marker connector circuits.

Fig. 23 shows a block diagram of the central office circuits. A small amount of circuitry is provided for each concentrator. It consists of the following:

1. The trunk connecting crossbar switch and associated slow relays for disconnect control.
2. The concentrator control trunk circuits and associated pulse amplifiers.
3. An originating call detector to identify which concentrator among the ten served by the frame is calling.
4. A multicontact relay to connect the circuits individual to each concentrator with the common control circuits associated with the line link frame and markers.

The circuits associated with more than one concentrator are blocked out in the lower portion of Fig. 23. Much of this circuitry is similar to the relay circuits now provided on regular line link frames in the No. 5 crossbar system.[3] Only those portions of these blocks which employ the new techniques will be covered in more detail. These portions consist of the following:

1. The scanner pulse generator.
2. The originating line number register.

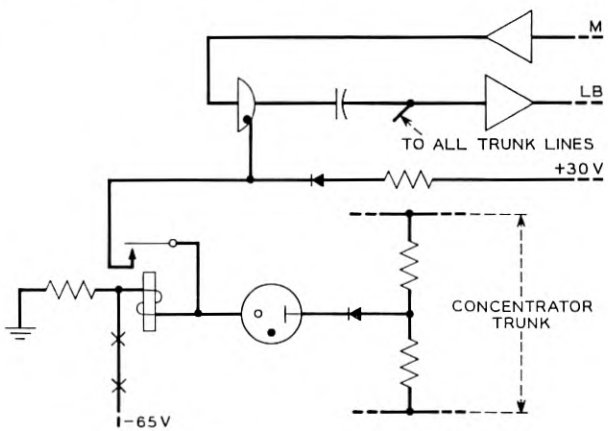
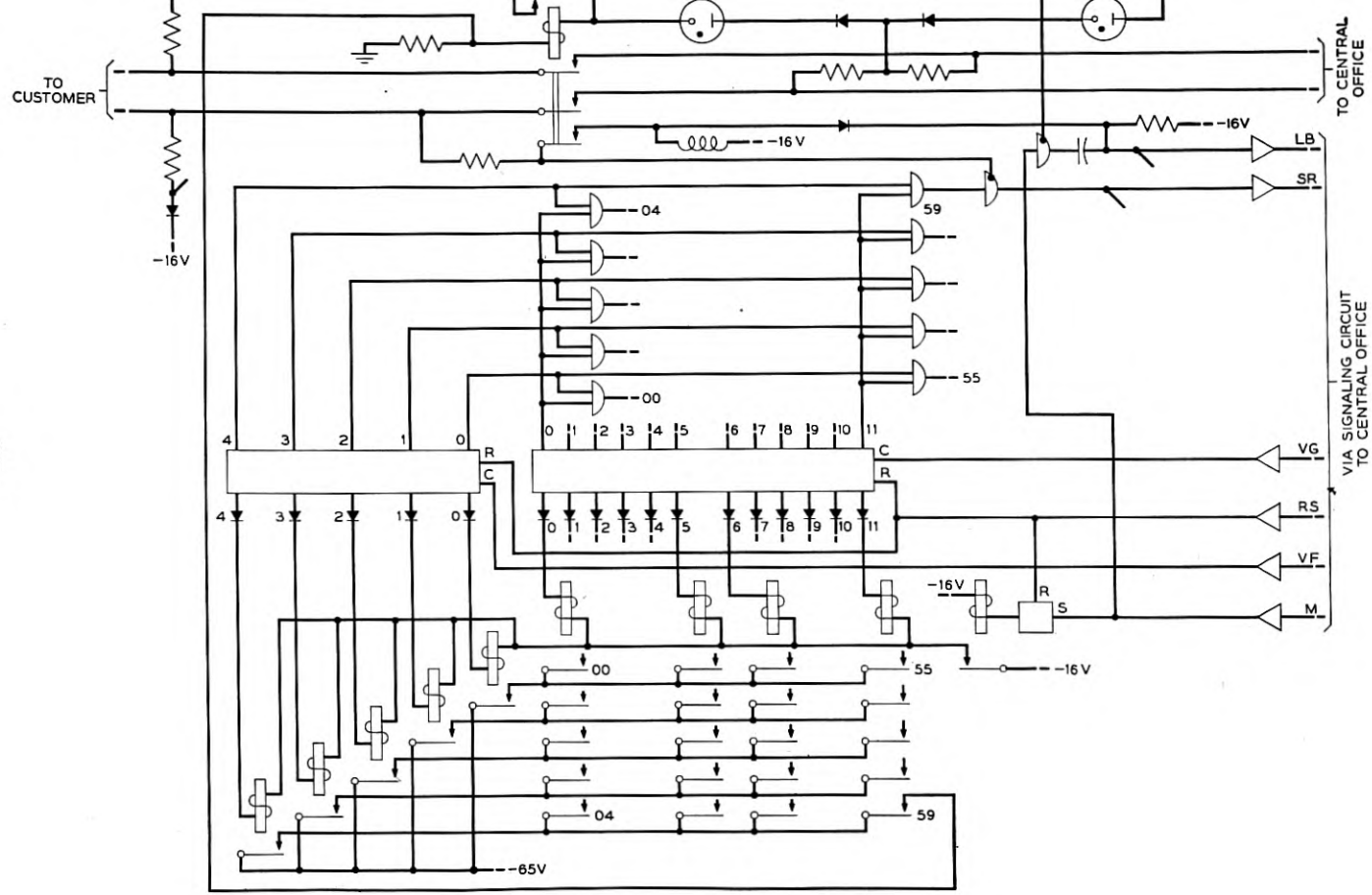


Fig. 20 — Crosspoint closure check.


TO CUSTOMER
-16V
TO CENTRAL OFFICE
-16V
LB
SR
04
59
00
55
VIA SIGNALING CIRCUIT TO CENTRAL OFFICE
VG
RS
VF
M
-16V
-16V
-65V

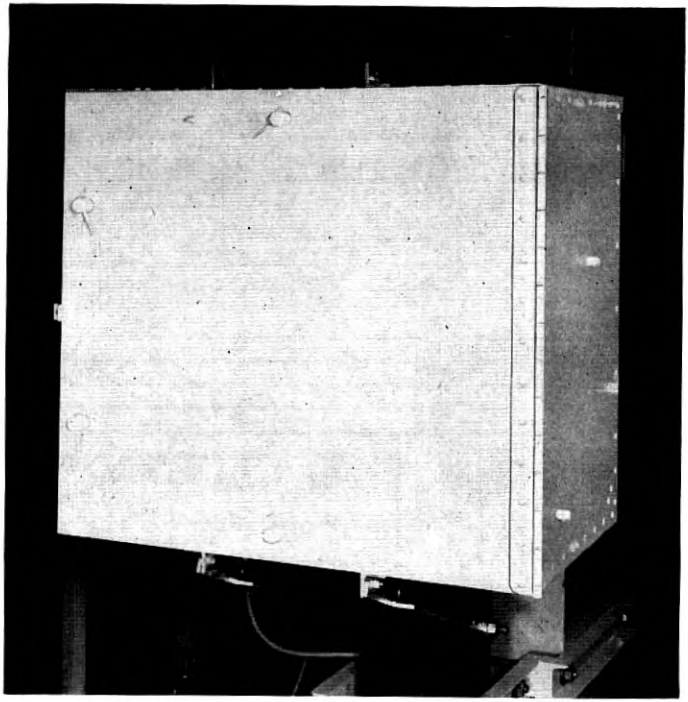


Fig. 22(a) — Complete line concentrator unit.

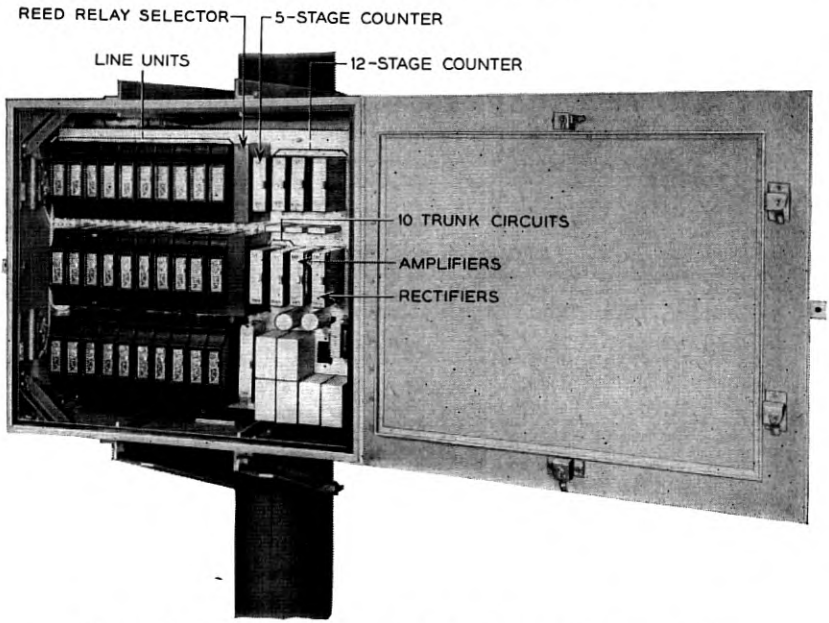


Fig. 22(b) — Identification of units within the line concentrator.

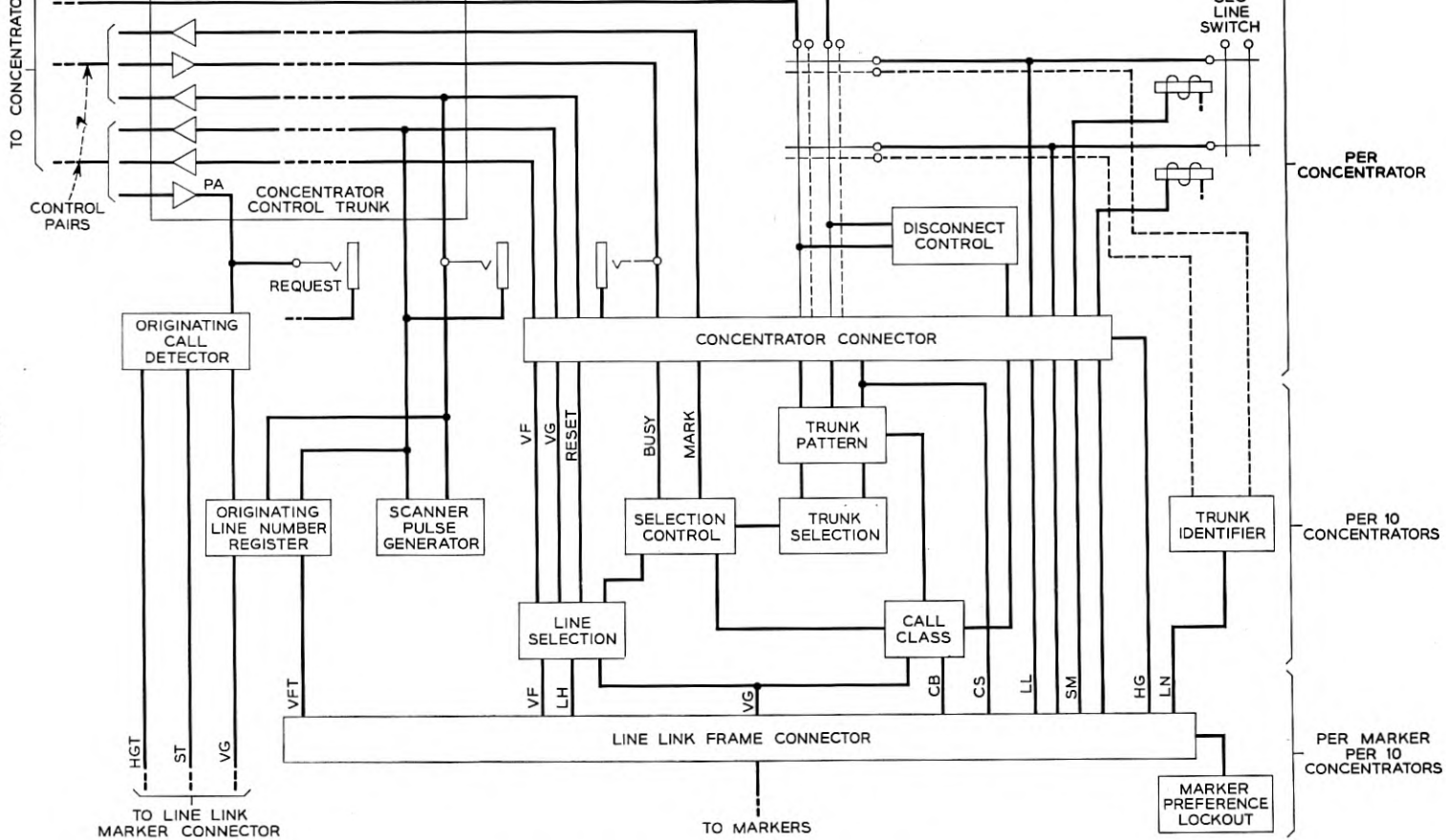


Fig. 23 — Block diagram of central office circuits.

3. The line selection circuit.
4. The trunk identifier and selection relay circuits.

(For an understanding of how these frame circuits work through the line link marker connector and markers in the No. 5 system, the reader should consult the references.)

The common central office circuits will be described first.

a. *Scanner Pulse Generator*

The scanner pulse generator, shown in Fig. 24, produces continuously the combination of VG, VF and RS or reset pulses, described in connection with Fig. 18(a), required to drive the scanners for a number of concentrators. The primary pulse source is a 1,000-cycle transistor oscillator. This oscillator drives a transistor bi-stable circuit arranged as a binary counter such that on each cycle of the oscillator output it alternately assumes one of its states. Pulses produced by one state drive a 5-stage counter. Pulses produced by the other state through gates drive a 12-stage counter.

The pulses which drive the 5-stage counter are the same pulses which are used for the VF pulses to drive scanners. Each time the first stage of the 5-stage counter is on, a gate is opened to allow a pulse to drive the 12-stage counter. The pulses which drive the 12-stage counter are also the pulses used as the VG pulses for driving the scanners. They are out of phase with the VF pulses.

When the last stage of the 12-stage counter is on, the gate which

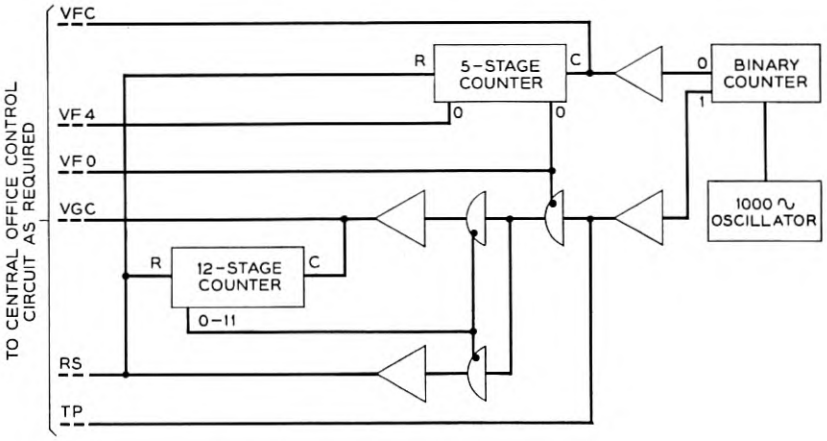


Fig. 24 — Scanner pulse generator.

transmits pulses to the 12-stage counter is closed and another gate is opened which produces the reset pulse. The reset pulse is thereby transmitted to the scanners in place of the first vertical group pulse. At the same time the 5 and 12-stage counters in the scanner pulse generator are reset to enable the starting of a new cycle.

In the central office control circuits, out of phase pulses on lead TP similar to those which drive the VG counters at the concentrator are used for various gating operations.

b. *The Originating Call Detection and Line Number Registration*

The originating call detector (Fig. 25) and the originating line number register (Fig. 26) together receive the information from the line concentrator used to identify the number of the line making a service request. The receipt of the service request pulse from a concentrator in a particular time slot will set a transistor bi-stable circuit HGT of Fig. 25 associated with that concentrator if no other originating call is being served by the frame circuits at this time.

The originating line number register consists of a 5 and 12-stage counter. These counters are normally driven through gates in synchronism with the scanning counters at concentrators with pulses supplied from the scanner pulse generator. When a service request pulse is received from any of the concentrators served by a line link frame, a pulse is sent to the originating line number register which operates a bi-stable circuit over a lead RH in Fig. 26. This bi-stable circuit then closes the gates through which the 5- and 12-stage counters are being driven, and also closes a gate which prevents them from being reset.

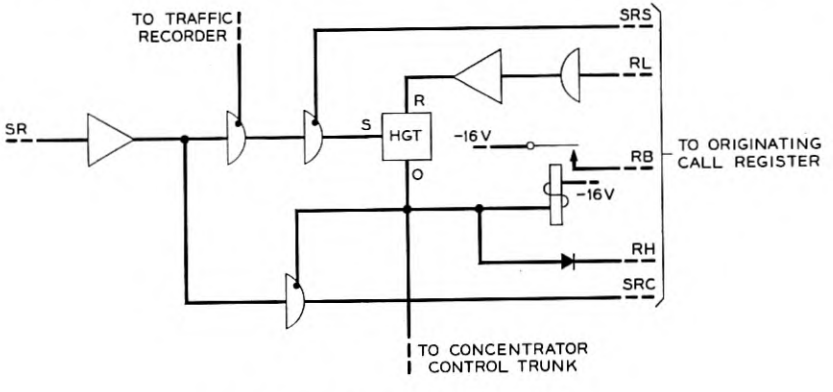


Fig. 25 — Originating call detector.

In this way, the number of the line which originated a service request is locked into these counters until the bi-stable circuit is restored to normal.

The HGT bi-stable circuit of Fig. 25 indicates which particular concentrator has originated a service request. A relay in the collector circuit has contacts which pass this information on to the other central office control circuits to indicate the number of the concentrator on the frame which is requesting service. This is the same as a horizontal group on a regular line link frame and hence the horizontal group designation is used to identify a concentrator.

With the operation of this relay, relays associated with the counters of the originating line number register are operated. These relays indicate to the other central office circuits the vertical file and vertical group identification of the calling line. Contacts on the vertical group relays are used to set a bi-stable circuit associated with lead RL of Fig. 25 each time the scanner pulse generator generates a pulse corresponding to the vertical file of the calling line number registered.

The operation of the HGT bi-stable circuit inhibits in the concentrator control trunk circuit (Fig. 27) the transmission of further VG and

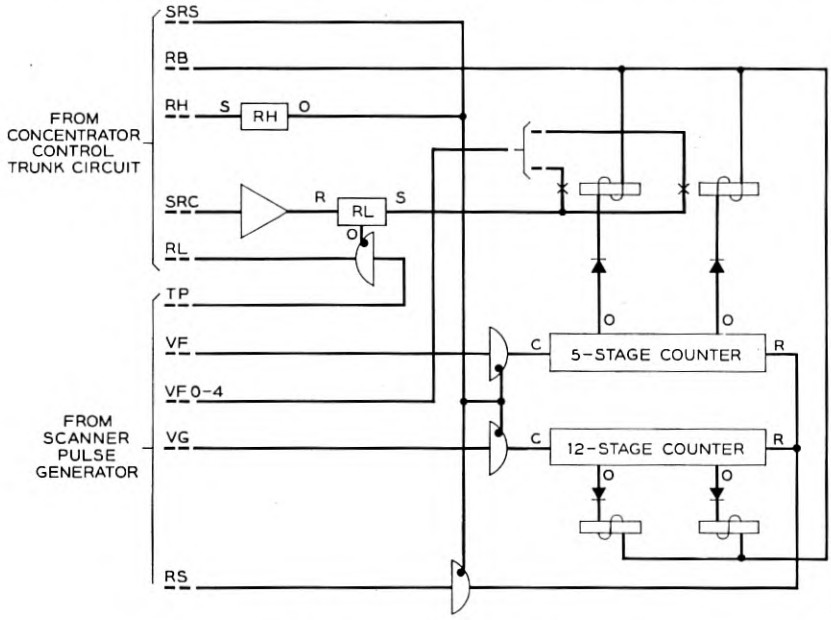


Fig. 26 — Originating line number register.

reset pulses to the concentrator so that, as described in Section 7a, only the VF counter continues to step once each 0.010 sec. So long as the line continues to request service this service request pulse is gated to reset the RL bi-stable circuit within the same time slot that it was set. If, however, a request for service is abandoned the RL bi-stable circuit of Fig. 26 will remain on and permit a TP pulse from the scanner pulse generator to reset the HGT bi-stable circuit which initiated the service request action.

Whenever the RH bi-stable circuit of Fig. 26 is energized it closes a gate over lead SRS for each concentrator to prevent any further service request pulses from being recognized until the originating call which has been registered is served. The resetting of the RH bi-stable circuit occurs once the call has been served. When more than one line concentrator is being served it is possible that the HGT bi-stable circuit of more than one concentrator will be set simultaneously as a result of coincidence in service requests from correspondingly numbered lines in these concentrators. The decision as to which concentrator is to be served is left to the marker, as it would normally decide which horizontal group to serve.

c. *Line Selection*

On all calls, originating and terminating, the marker transmits to the frame circuits the complete identity of the line which it will serve. In the case of originating calls it has received this information in the manner described in Section 8b. In either case, it operates wire spring relays VGO-11 and VFO-4, which enable gates so that the information may be stored in the 5- and 12-stage counters of the line selection circuit shown in Fig. 28.

The process of reading into the line selection counters starts when selection information has been received by the actuation of the HGS bi-stable circuit in the concentrator control trunk circuit of Fig. 27. This action stops the regular transmission of scanner pulses if they have not been stopped as a result of a call origination. At the same time it enables gates for transmission of information from the line selection circuit, Fig. 28.

The ST bi-stable circuit of the line selection circuit is also enabled to start the process of setting the line selection counters. The next TP pulse sets the R1 bi-stable circuit. This bi-stable circuit enables a gate which permits the next TP pulse to set the counters and transmit a reset pulse to the concentrator through pulse amplifier R1A. At the same time bi-stable circuit ST is reset to prevent the further read-in or reset

pulses and to permit pulses through amplifier OPA to start the outpulsing of line selections. These pulses pass to the VGP and VFP leads as long as the VG and VF line selection counters have not reached their first and last stages respectively. The output pulses to the concentrator are also fed into the drive leads of these counters so that, as the counters in the concentrator are stepped up, the counters in the central office line selection circuit are stepped down. When the first stage of the VF counter goes on, the VF pulses are no longer transmitted until the first stage of the VG counter goes on. This insures that a VF pulse is the last to be transmitted. Also this pulse is not transmitted until the other frame circuits have successfully completed selections of an idle concentrator trunk. Then bi-stable circuit VFLD is energized,

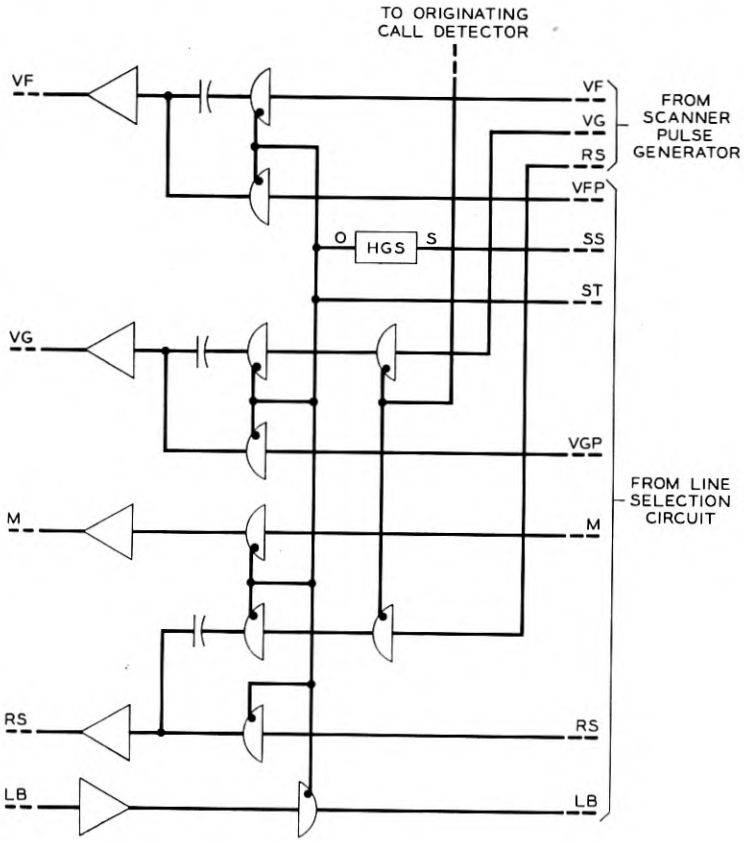


Fig. 27 — Concentrator control trunk circuit.

producing, during its transition, the last VF pulse for transmission to the concentrator.

d. *Trunk Selection and Identification*

The process of selecting an idle concentrator trunk to which the line has access utilizes familar relay circuit techniques.[19] This circuit, in Fig. 29, will not be described in detail. One trunk selection relay, TS, is operated indicating the preferred idle trunk serving a line in the particular vertical group being selected as indicated by the VG relay which has been operated by the marker.

The TS4 and TS5 relays select trunks 8 and 9 which are available to each line while the 4 trunks available to only half of the lines are selected by relays TS0-TS3. The busy or idle condition of each trunk is indicated by a contact on the hold magnet associated with each trunk through

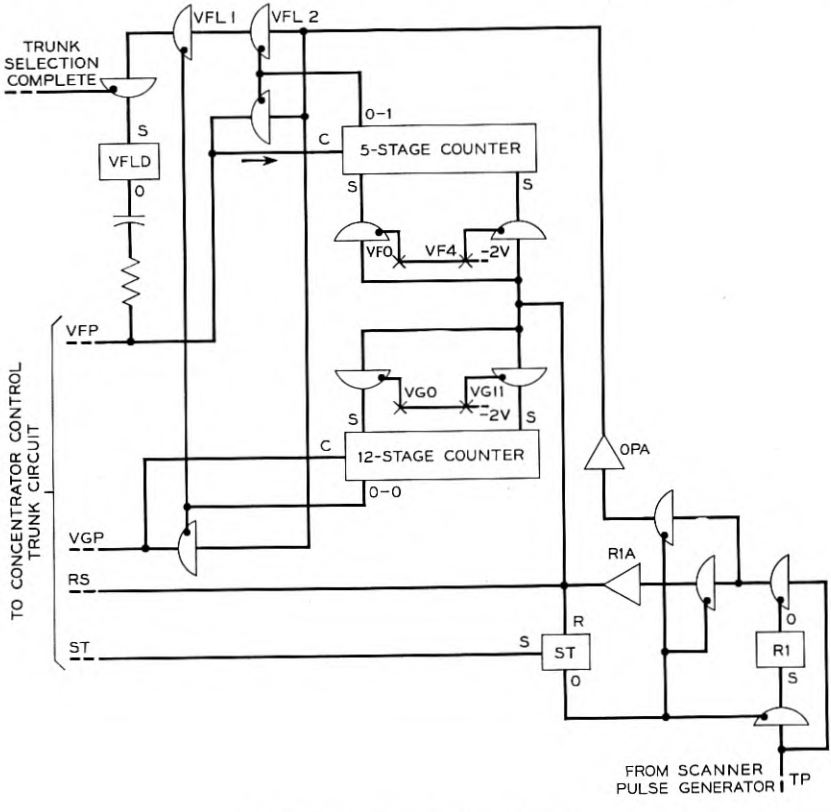


Fig. 28 — Line selection circuit.

relay HG which operates on all originating and terminating calls to the particular concentrator served by these trunks. The end chain relay TC of the lockout trunk selection circuit[19] connects battery from the SR relay windings of idle trunks to the windings of the TS relays to permit one of the latter relays to operate and to steer circuits, not shown on Fig. 29, to the hold magnet of the trunk and to the tip-and-ring conductors of the trunk to apply the selection voltages shown on Figs. 13 and 14.

The path for operating the hold magnet originates in the marker. The path looks like that which the marker uses on the line hold magnet when setting up a call on a regular line link frame. For this reason and other similar reasons this concentrator line link frame concept has been nicknamed the "fool-the-marker" scheme.

Should a hold magnet release while a new call is being served the ground from the TC relay normal or the TS relay winding holds relay

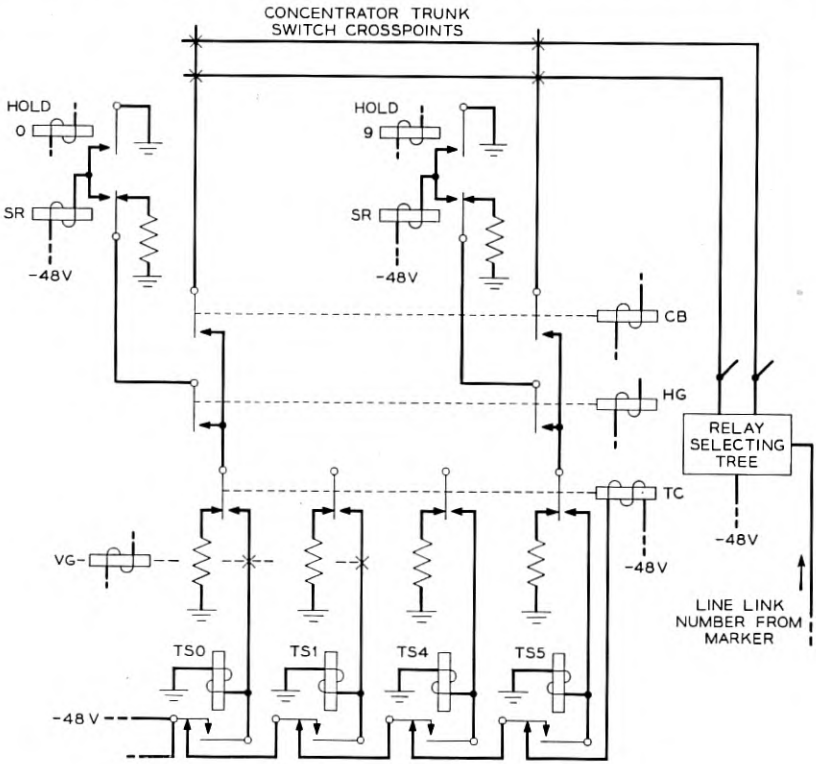


Fig. 29 — Trunk selection and identification.

SR operated through its own contact until the new call has been set up. This prevents interference of disconnect pulses applied to the trunk when a selection is being made and insures that a disconnect pulse is transmitted before the trunk is reused.

A characteristic of the No. 5 crossbar system is that the originating connection to a call register including the line hold magnet is released and a new connection, known as the "call back connection", is established to connect the line to a trunk circuit after dialing is completed.

With concentrator operation the concentrator trunk switch connection is released but the disconnect signal is not sent to the concentrator as a result of holding the SR relay as described above. However, the marker does not know to which trunk the call back connection is to be established. For this reason the frame circuits include an identification process for determining the number of the concentrator trunk to be used on call back prior to the release of the originating register connection.

Identification is accomplished by the marker transmitting to the frame circuits the number of the link being used on the call. This information is already available in the No. 5 system. The link being used is marked with -48 volts by a relay selecting tree[20] to operate the TS relay associated with the trunk to which the call back connection is to be established. Relay CB (Fig. 29) is operated on this type of call instead of relay HG. The circuits for reoperating the proper hold magnet are already available on the TS relay which was operated, thereby reselecting the trunk to which the customer is connected. The concentrator connection is not released when the hold magnet releases and again the marker operates as it would on a regular line link frame call.

9. FIELD TRIALS

Three sets of the experimental equipment described here have been constructed and placed in service in various locations. The equipment for these trials is the forerunner of a design for production which will incorporate device, circuit and equipment design changes based on the trial experiences. Fig. 30 shows the cabinet mounted central office trial equipment with the designation of appropriate parts.

For the field trials described, the line links on a particular horizontal level of existing line link frames were extended to a separate cross-bar switch provided for this purpose in the trial equipment. The regular line link connector circuits were modified to work with the trial control circuits whenever a call was originated or terminated on this level. No lines were terminated in the regular primary line switches for this level.

10. MISCELLANEOUS FEATURES OF TRIAL EQUIPMENT

There are a number of auxiliary circuits provided with the trial equipment to aid in the solutions of problems brought about by the concepts of concentrator service. One of the purposes of the trials was to determine the way in which the various traffic, plant and commercial ad-

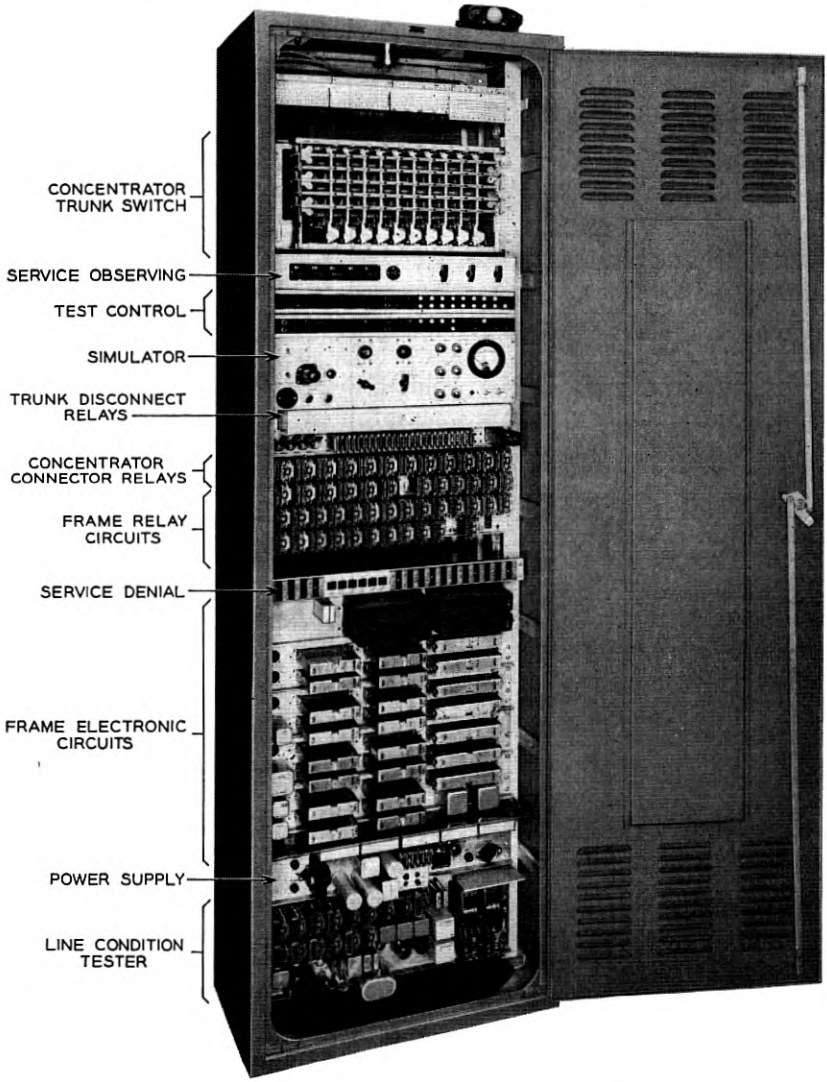


Fig. 30 — Trial central office equipment.

ministrative functions could be economically performed when concentrators become common telephone plant facilities. The more important of these miscellaneous features are discussed under the following headings:

a. *Traffic Recording*

To measure the amount and characteristics of the traffic handled by the concentrator a magnetic tape recorder, Fig. 31, was provided for each trial. The number of the lines and trunks in use each 15 seconds during programmed periods of each day were recorded in coded form with polarized pulses on the 3-track magnetic tape moving at a speed of 7½″ per second. Combinations of these pulses designate trunks busy on intra-concentrator connections and reverting calls.

The line busy indications were derived directly from the line busy information received during regular scanning at the concentrator. During one cycle in each 15 seconds new service requests were delayed to insure that a complete scan cycle would be recorded. Terminating calls were not delayed since marker holding time is involved. Trunk conditions are derived for a trunk scanner provided in the recorder.

In addition to recording the line and trunk usage, recordings were made on the tape for each service request detected during a programmed period to measure the speed with which each call received dial tone and the manner in which the call was served. In this type of operation the length of the recording for each request made at a tape speed of only ¼″ per second is a measure of service delay time.

As may be observed from Fig. 31 the traffic recorder equipment was built with vacuum tubes and hence required a rather large power supply. It is expected that a transistorized version of this traffic recorder serving all concentrators in a central office will be included in the standard model of the line concentrator equipment. With this equipment, traffic engineers will know more precisely the degree to which each concentrator may be loaded and hence insure maximum utilization of the concentrator equipment.

b. *Line Condition Tester*

It has been a practice in more modern central office equipment to include automatic line testing equipment.[21] An attempt has been made to include similar features with the concentrator trial equipment. The line condition tester (see Fig. 30) provides a means for automatically connecting a test circuit to each line in turn once a test cycle has been

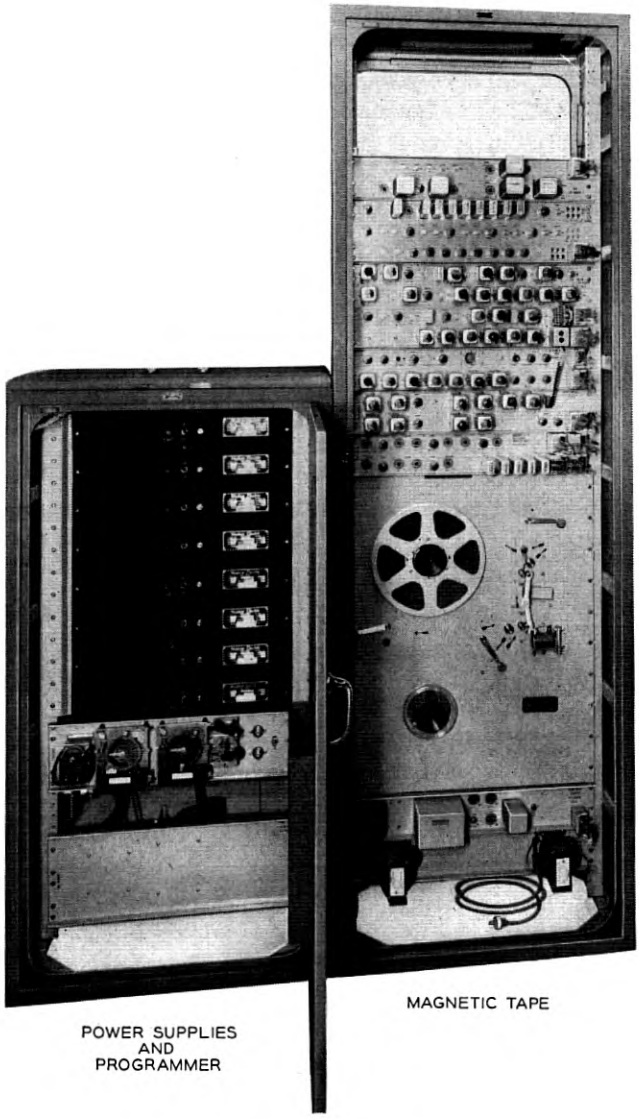


Fig. 31 — Traffic recorder.

manually initiated. This test is set up on the basis of the known concentrator passive line circuit capabilities. Should a line fail to pass this test, the test circuit stops its progress and brings in an alarm to summon central office maintenance personnel. The facilities of the line tester are also used to establish, under manual control, calls to individual lines as required to carry out routine tests.

c. *Simulator*

As the central office sends out scanner control pulses either no signal, a line busy or service request pulse is returned to the central office in each time slot. The simulator test equipment, shown in Fig. 30, was designed to place pulses in a specific time slot to simulate a line under test at the concentrator.

In addition to transmitting the equivalent of concentrator output pulses the simulator can receive the regular line selection pulses transmitted to the concentrator for purposes of checking central office operations. It is possible by combined use of the line tester and simulator to observe the operation of the concentrator and to determine the probable cause when a fault occurs.

d. *Service Observing*

The removal of the line terminals from the central office poses a number of problems in conjunction with the administration of central office equipment. One of these is service observing.

To maintain a check on the quality of service being rendered by the telephone system, service observing taps are made periodically on telephone lines. This is normally done by placing special connector shoes on line terminations in the central office.

To place such shoes at the remote concentrator point would lead to administrative difficulties and added expense. Therefore, a method was devised to permit service observing equipment to be connected to concentrator trunks on calls from specific lines which were to be observed. This method consisted of manual switches on which were set the number of the line to be observed in terms of vertical group and vertical file. Whenever this line originated a call and the call could be placed over the first preferred trunk, automatic connection was made to the service observing desk in the same manner as would occur for a line terminated directly in the central office.

In addition, facilities were provided for trying a new service observing technique where calls originating over a particular concentrator

trunk would be observed without knowledge of the originating line number. For this purpose a regular line observing shoe was connected to one of the ten concentrator trunk switch verticals in the trial equipment and from here connected to the service observing desk in the usual manner.

The basic service observing requirements in connection with line concentrator operation have not as yet been fully determined. However, it appears at this time that the trunk observing arrangement may be preferable.

e. *Service Denial*

In most systems denial of originating service for non-payment of telephone service charges, for trouble interception and for permanent signals caused by cable failures or prolonged receiver-off-hook conditions may be treated by the plant forces at the line terminals or by blocking the line relay. To avoid concentrator visits and to enable the prompt clearing of trouble conditions which tie up concentrator trunks, a service denial feature has been included in the design of the central office circuits.

This feature consists of a patch-panel with special gate cords which respond to particular time slots and inhibit service request signals produced by a concentrator during this period. In this way service requests can be ignored and prevent originating call service on particular lines until a trouble locating or other administrative procedure has been invoked.

f. *Display Circuit*

A special electronic switch was developed for an oscilloscope. This arrangement permited the positioning of line busy and service request pulses in fixed positions representing each of the 60 lines served. Line busy pulses were shown as positive and service request pulses as negative. This plug connected portable aid, see Fig. 32, was useful in tracing calls and identifying lines to which service may be denied, due to the existence of permanent signals.

Other circuits and features, too detailed to be covered in this paper, have been designed and used in the field trials of remote line concentrators. Much has been learned from the construction and use of this equipment which will aid in making the production design smaller, lighter, economical, serviceable and reliable.

Results from the field trials have encouraged the prompt undertaking

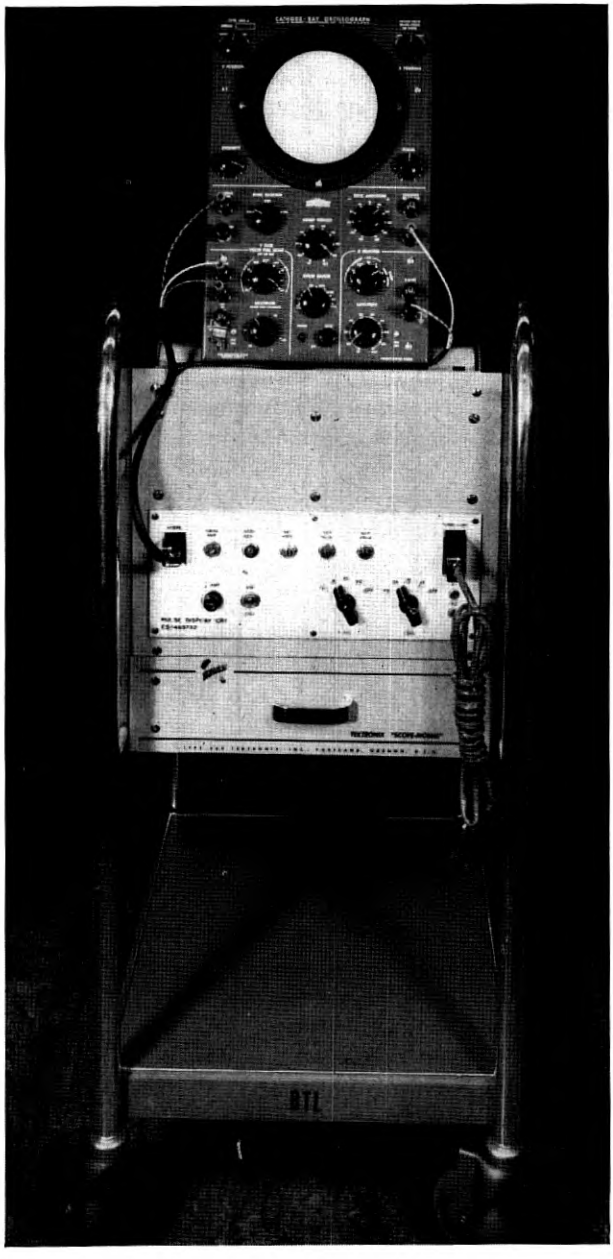


Fig. 32 — Pulse display oscilloscope.

of development of a remote line concentrator for quantity production. The cost of remote line concentrator equipment will determine the ultimate demand. In the meantime, an effort is being made to take advantage of the field trial experiences to reduce costs commensurate with insuring reliable service.

The author wishes to express his appreciation to his many colleagues at Bell Telephone Laboratories whose patience and hard work have been responsible for this new adventure in exploratory switching development. An article on line concentrators would not be complete without mention of C. E. Brooks who has encouraged this development and under whose direction the engineering studies were made.

BIBLIOGRAPHY

1. E. C. Molina, The Theory of Probabilities Applied to Telephone Trunking Problems, B.S.T.J., **1,** pp. 69–81, Nov., 1922.
2. Strowger Step-by-Step System, Chapter 3, Vol. 3, Telephone Theory and Practice by K. B. Miller. McGraw-Hill 1933.
3. F. A. Korn and J. G. Ferguson, Number 5 Crossbar Dial Telephone Switching System, Elec. Engg., **69,** pp. 679–684, Aug., 1950.
4. U. S. Patent 1,125,965.
5. O. Myers, Common Control Telephone Switching Systems, B.S.T.J., **31,** pp. 1086–1120, Nov., 1952.
6. L. J. Stacy, Calling Subscribers to the Telephone, Bell Labs. Record, **8,** pp. 113–119, Nov., 1929.
7. J. Meszar, Fundamentals of the Automatic Telephone Message Accounting System, A. I. E. E. Trans., **69,** pp. 255–268, (Part 1), 1950.
8. O. M. Hovgaard and G. E. Perreault, Development of Reed Switches and Relays, B.S.T.J., **34,** pp. 309–332, Mar., 1955.
9. W. A. Malthaner and H. E. Vaughan, Experimental Electronically Controlled Automatic Switching System, B. S.T.J., **31,** pp. 443–468, May, 1952.
10. S. T. Brewer and G. Hecht, A Telephone Switching Network and its Electronic Controls, B.S.T.J., **34,** pp. 361–402, Mar., 1955.
11. L. W. Hussey, Semiconductor Diode Gates, B.S.T.J., **32,** pp. 1137–54, Sept., 1953.
12. U. S. Patent 1,528,982.
13. J. J. Ebers and S. L. Miller, Design of Alloyed Junction Germanium Transistor for High-Speed Switching, B.S.T.J., **34,** pp. 761–781, July, 1955.
14. W. B. Graupner, Trunking Plan for No. 5 Crossbar System, Bell Labs. Record, **27,** pp. 360–365, Oct., 1949.
15. G. L. Pearson and B. Sawyer, Silicon p-n Junction Alloy Diodes, I.R.E. Proc., **42,** pp. 1348–1351, Nov., 1952.
16. A. E. Anderson, Transistors in Switching Circuits, B.S.T.J., **31,** pp. 1207–1249, Nov., 1952.
17. J. J. Ebers and J. L. Moll, Large-Signal Behavior of Junction Transistors, I. R. E. Proc., **42,** pp. 1761–1784, Dec., 1954.
18. J. J. Ebers, Four-Terminal p-n-p-n Transistors, I. R. E. Proc., **42,** pp. 1361–1364, Nov., 1952.
19. A. E. Joel, Relay Preference Lockout Circuits in Telephone Switching, Trans. A. I. E. E., **67,** pp. 720–725, 1948.
20. S. H. Washburn, Relay "Trees" and Symmetric Circuits, Trans. A. I. E. E., **68,** pp. 571–597, 1949.
21. J. W. Dehn and R. W. Burns, Automatic Line Insulation Testing Equipment for Local Crossbar Systems, B.S.T.J., **32,** pp. 627–646, 1953.

Transistor Circuits for Analog and Digital Systems*

By FRANKLIN H. BLECHER

(Manuscript received November 17, 1955)

This paper describes the application of junction transistors to precision circuits for use in analog computers and the input and output circuits of digital systems. The three basic circuits are a summing amplifier, an integrator, and a voltage comparator. The transistor circuits are combined into a voltage encoder for translating analog voltages into equivalent time intervals.

1.0. INTRODUCTION

Transistors, because of their reliability, small power consumption, and small size find a natural field of application in electronic computers and data transmission systems. These advantages have already been realized by using point contact transistors in high speed digital computers.[1] This paper describes the application of junction transistors to precision circuits which are used in dc analog computers and in the input and output circuits of digital systems. The three basic circuits which are used in these applications are a summing amplifier, an integrator, and a voltage comparator. A general procedure for designing these transistor circuits is given with particular emphasis placed on new design methods that are necessitated by the properties of junction transistors. The design principles are illustrated by specific circuits. The fundamental considerations in the design of transistor operational amplifiers are discussed in Section 2.0. In Section 3.0 an illustrative summing amplifier is described, which has a dc accuracy of better than one part in 5,000 throughout an operating temperature range of 0 to 50°C. The feedback in this amplifier is maintained over a broad enough frequency band so that full accuracy is attained in about 100 microseconds.

The design of a specific transistor integrator is presented in Section

* Submitted in partial fulfillment of the requirements for the degree of Doctor of Electrical Engineering at the Polytechnic Institute of Brooklyn.

4.0. The integrator can be used to generate a voltage ramp which is linear to within one part in 8,000. By means of an automatic zero set (AZS) circuit which uses a magnetic detector, the slope of the voltage ramp is maintained constant to within one part in 8,000 throughout a temperature range of 20°C to 40°C.

The voltage comparator, described in Section 5.0, is an electrical device which indicates the instant of time an input voltage waveform passes through a predetermined reference level. By taking advantage of the properties of semiconductor devices, the comparator can be designed to have an accuracy of ± 5 millivolts throughout a temperature range of 20°C to 40°C.

In Section 6.0, the system application of the transistor circuits is demonstrated by assembling the summing amplifier, the integrator, and the voltage comparator into a voltage encoder. The encoder can be used to translate an analog input voltage into an equivalent time interval with an accuracy of one part in 4,000. This accuracy is realized throughout a temperature range of 20°C to 40°C for the particular circuits described.

2.0. FUNDAMENTAL CONSIDERATIONS IN THE DESIGN OF OPERATIONAL AMPLIFIERS

The basic active circuit used in dc analog computers is a direct coupled negative feedback amplifier. With appropriate input and feedback networks, the amplifier can be used for multiplication by a constant coefficient, addition, integration, or differentiation as shown in Figure 1.[2] The accuracy of an operational amplifier depends only on the passive components used in the input and feedback circuits provided that there is sufficient negative feedback (usually greater than 60 db). The time that is required for the amplifier to perform a calculation is an inverse function of the bandwidth over which the feedback is maintained. Thus a fundamental problem in the design of an operational amplifier is the development of sufficient negative feedback over a reasonably broad frequency range. The associated problem is the realization of satisfactory stability margins. Finally there is the problem of reducing the drift which is inherent in direct coupled amplifiers and particularly troublesome for transistors because of the variation in their characteristics with temperature.

The first step in the design is the blocking out of the configuration for the forward gain circuit (designated A in Fig. 1). Three primary requirements must be satisfied:

(1) Stages must be direct coupled.

(2) Amplifier must provide one net phase reversal.

(3) Amplifier must have enough current gain to meet accuracy requirements.

Three possible transistor connections are available:[3] (a) the common base connection which may be considered analogous to the common grid vacuum tube connection; (b) the common emitter connection which is analogous to the common cathode connection; and (c) the common collector connection which is analogous to the cathode follower connection. These three configurations together with their approximate equivalent circuits are shown in Fig. 2. It has been shown[4] that for most junction transistors the circuit element a is given by the expression

$$a = \operatorname{sech}\left[\frac{W}{L_m}(1 + p\tau_m)^{1/2}\right] \qquad (1)^*$$

where W is the thickness of the transistor base region, L_m is the diffusion length and τ_m the lifetime of minority charge carriers in the base region,

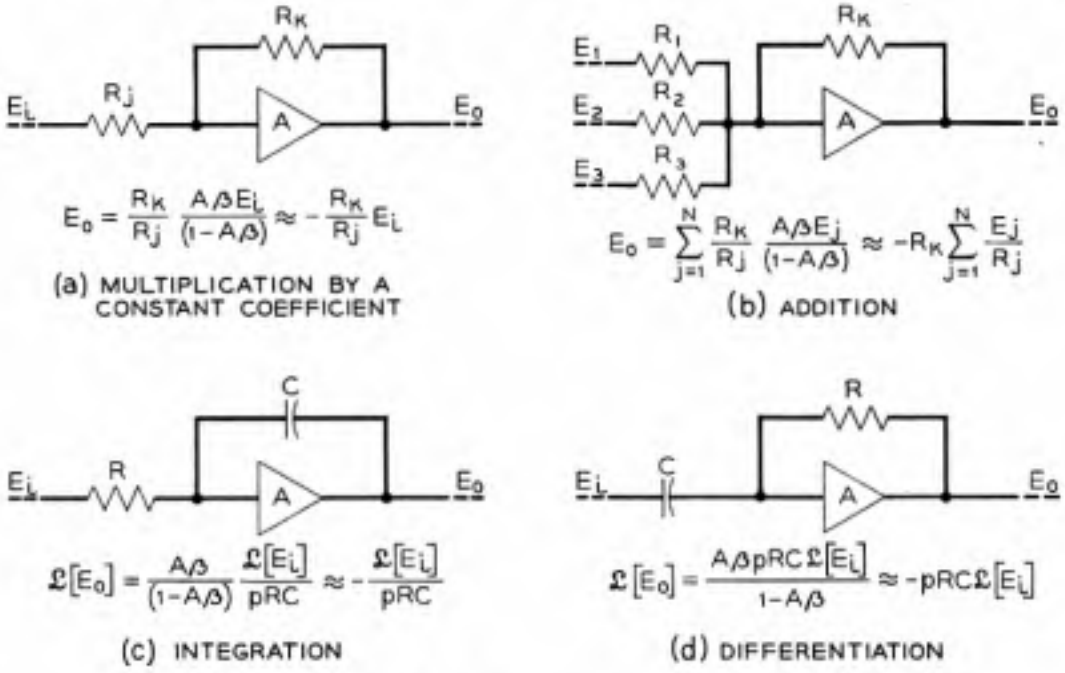


Fig. 1 — Summary of operational amplifiers.

* This expression assumes that the injection factor γ and the collector efficiency α_i are both unity. This is a good approximation for all alloy junction transistors and most grown junction transistors.

and $p = j\omega$. At frequencies less than $\omega_a/2\pi$, (1) can be approximated by

$$a = \frac{a_0}{1 + \frac{p}{\omega_a}} \qquad (2)^5$$

where a_0 is the low frequency value of

$$a \approx 1 - \frac{1}{2}\left(\frac{W}{L_m}\right)^2, \qquad \text{and} \qquad \omega_a = \frac{2.4 D_m}{W^2}$$

(D_m is the diffusion constant for the minority charge carriers in the base region). A readily measured parameter called alpha (α), the short circuit current gain of a junction transistor in the common base connec-

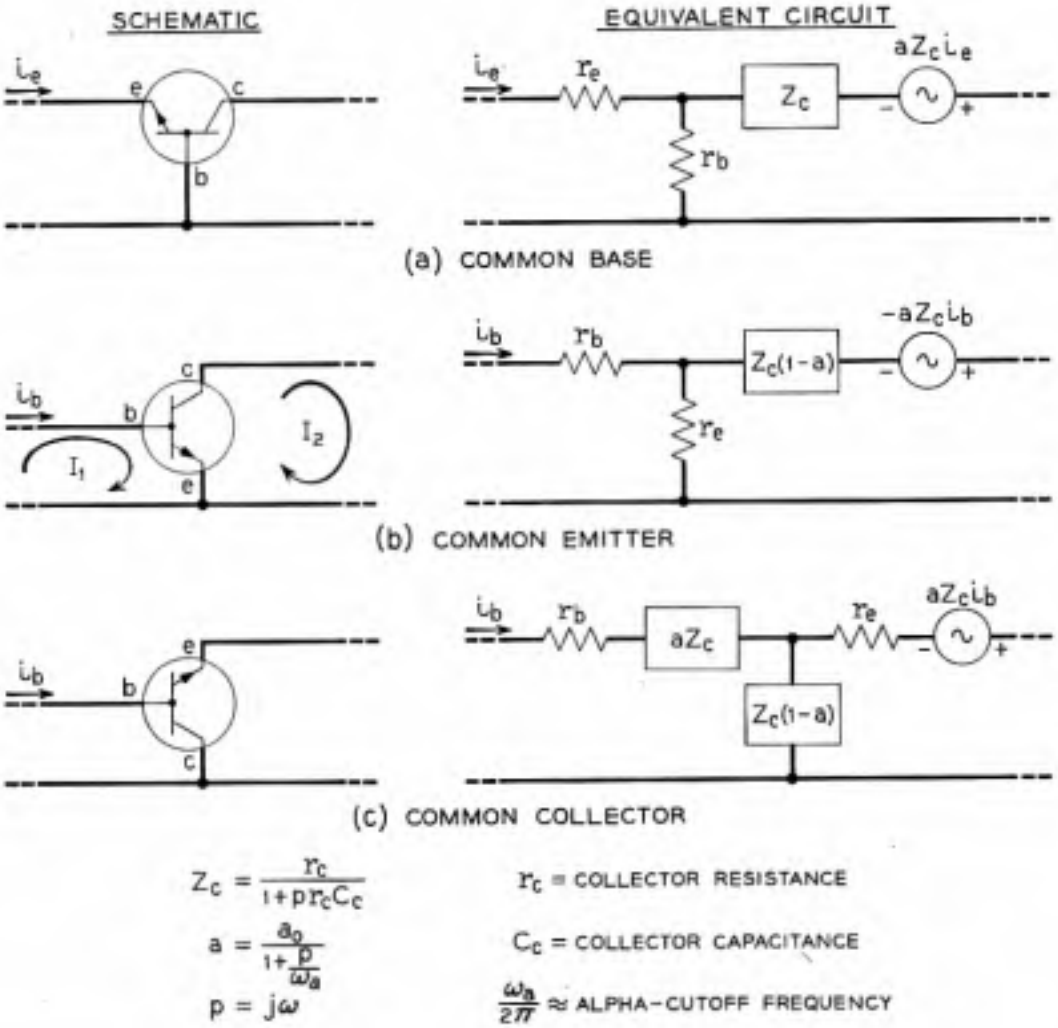


Fig. 2 — Basic transistor connections.

tion, is related to a by the equation

$$\alpha = \frac{aZ_c + r_b}{Z_c + r_b} \tag{3}$$

For most junction transistors the base resistance, r_b, is much smaller than the collector impedance $| Z_c |$, at frequencies less than $\omega_a/2\pi$. Therefore, $\alpha \approx a$ and $\omega_a/2\pi$ is very nearly equal to the alpha-cutoff frequency, the frequency at which $| \alpha |$ is down by 3 db.

The transistor parameters r_e and r_b are actually frequency sensitive and should be represented as impedances.[6] However, good agreement between theory and experiment is obtained at frequencies less than $\omega_a/2\pi$ with r_e and r_b assumed constant.

The choice of an appropriate transistor connection for a direct coupled, negative feedback amplifier, is based on the following reasoning. The common base connection may be ruled out immediately because this connection does not provide current gain unless a transformer interstage is used. The common emitter connection provides short circuit current gain and a phase reversal for each stage. Thus if the amplifier is composed of an odd number of common emitter stages, all three requirements previously listed, are satisfied. A common emitter cascade has the additional practical advantage, that by alternating n-p-n and p-n-p types of transistors, the stages can be direct coupled with practically zero interstage loss.[7]

The common collector connection provides short circuit current gain but no phase reversal. Consequently, the dc amplifier cannot consist entirely of common collector stages and operate as a negative feedback amplifier. This paper will consider only the common emitter connection since, in general, for the same number of transistor stages, the common emitter cascade provides more current gain than a cascade composed of both common collector and common emitter stages.

2.1 *Evaluation of External Voltage Gain*

Since the equivalent circuit of the junction transistor is current activated, it is convenient to treat feedback in a single loop transistor amplifier as a loop current transmission (refer to Appendix I) instead of as a loop voltage transmission which is commonly used for single loop vacuum tube amplifiers.[8] Fig. 3 shows a single loop feedback amplifier in which a fraction of the output current is fed back to the input. A is defined as the short circuit current gain of the amplifier without feedback, and β is defined as the fraction of the short circuit output current (or Norton

equivalent circuit current) fed back to the input summing node. With these definitions,

$$I_{SC} = AI_{IN}' \tag{4}$$

$$I_\beta = \beta I_{SC} \tag{5}$$

where I_{SC} is the Norton equivalent short circuit current.

From Kirchhoff's first law

$$I_{IN}' = I_{IN} + I_\beta \tag{6}$$

Combining relations (4) to (6) yields

$$\frac{I_{SC}}{I_{IN}} = \frac{A}{1 - A\beta} \tag{7}$$

Expression (7) provides a convenient method for evaluating the external

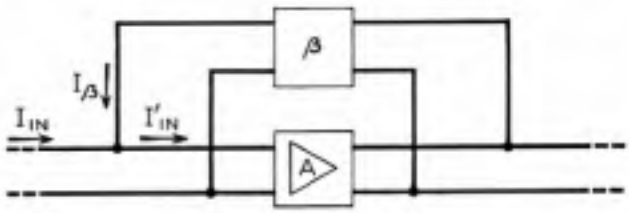


Fig. 3 — Single loop feedback amplifier.

voltage gain of an operational amplifier. Fig. 4 shows a generalized operational amplifier with N inputs. With this configuration,

$$I_{IN} = \sum_{j=1}^{N} \left[\frac{E_j - \frac{I_{SC}}{A} Z_{IN}'}{Z_j} \right] \tag{8}$$

where E_j, $j = 1, 2, \cdots, N$, are the N input voltages referred to the ground node.

Z_j, $j = 1, 2, \cdots, N$, are the N input impedances

Z_{IN}' is the input impedance of the amplifier measured at the summing node with the feedback loop opened.

$$I_\beta = \frac{E_{OUT} - Z_{IN}' \frac{I_{SC}}{A}}{Z_K} \tag{9}$$

$$E_{OUT} = \frac{I_{SC} - I_\beta}{\frac{1}{R_L} + \frac{1}{Z_{OUT}'}} \tag{10}$$

where Z_{OUT}' is the output impedance of the amplifier measured with the feedback loop opened. The expression for the output voltage is obtained by combining (7), (8), (9), and (10).

$$E_{OUT} = \sum_{j=1}^{N} E_j \frac{Z_K}{Z_j} \left[\frac{A\beta + \frac{Z_{IN}'}{Z_K}}{1 - A\beta + \sum_{j=1}^{N} \frac{Z_{IN}'}{Z_j}} \right] \tag{11}*$$

where

$$A\beta = A \left[\frac{1 - \frac{Z_{IN}'}{A} \left(\frac{1}{R_L} + \frac{1}{Z_{OUT}'} \right)}{1 + \frac{Z_K}{R_L} + \frac{Z_K}{Z_{OUT}'}} \right]$$

$A\beta$ is equal to the current returned to the summing node when a unit

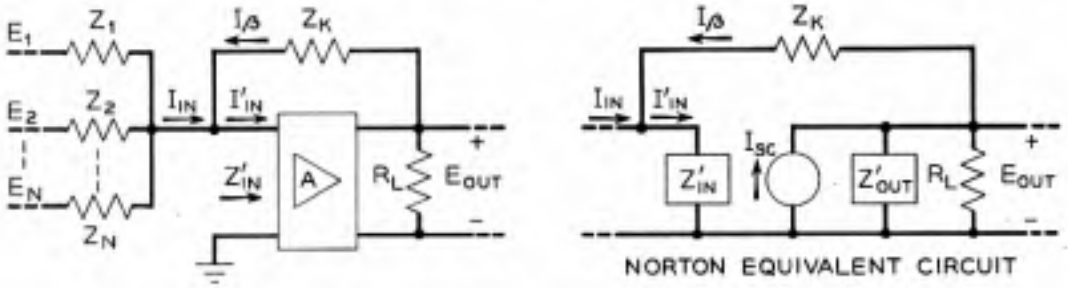


Fig. 4 — Generalized operational amplifier.

current is placed into the base of the first transistor stage ($I_{IN}' = 1$). If $|A\beta|$ is much greater than $|Z_{IN}'/Z_K|$ and

$$1 + \sum_{j=1}^{N} \left| \frac{Z_{IN}'}{Z_j} \right|,$$

then

$$E_{OUT} = -\sum_{j=1}^{N} E_j \frac{Z_K}{Z_j} \tag{12}$$

The accuracy of the operational amplifier depends on the magnitude of $A\beta$ and the precision of the components used in the input and feedback networks as can be seen from (11). There is negligible interaction between the input voltages because the input impedance at the summing node is equal to Z_{IN}' divided by $(1 - A\beta)$.[9] This impedance is usually negligibly small compared to the impedances used in the input circuit.

* In general, E_j and E_{OUT} are the Laplace transforms of the input and output voltages, respectively.

2.2. *Methods Used to Shape the Loop Current Transmission*

An essential consideration in the design of a feedback amplifier is the provision of adequate margins against instability. In order to accomplish this objective, it is necessary to choose a criterion of stability. In Appendix I it is shown that it is convenient and valid to base the stability of single loop transistor feedback amplifiers on the loop current transmission. In order to calculate the loop current transmission of the dc amplifier, the feedback loop is opened at a convenient point in the circuit, usually at the base of one of the transistors, and a unit current is injected into the base (refer to Fig. 24). The other side of the opened loop is connected to ground through a resistance $(r_e + r_b)$ and voltage $r_e I_4$. In many instances, the voltage $r_e I_4$ can be neglected. If $| Z_K |$ and

$$\frac{1}{\left| \sum_{j=1}^{N} \frac{1}{Z_j} \right|}$$

are much greater than $| Z_{IN}' |$, then $A\beta$ is very nearly equal to the loop current transmission. For absolute stability[10] the amplitude of the loop current transmission must be less than unity before the phase shift (from the low frequency value) exceeds 180°. Consequently, this characteristic must be controlled or properly shaped over a wide frequency

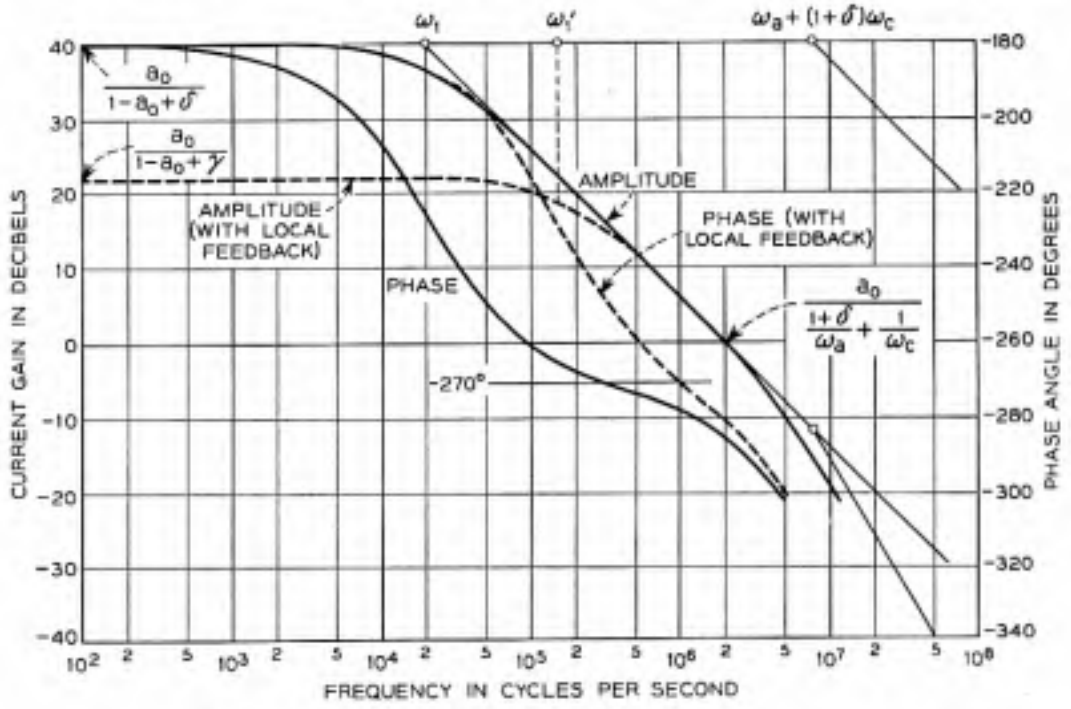


Fig. 5 — Current transmission of a common emitter stage.

band. In addition, it is desirable that the feedback fall off at a rate equal to or less than 9 db per octave in order to insure that the dc amplifier has a satisfactory transient response.

Three methods of shaping are described in this paper; local feedback shaping, interstage network shaping, and β circuit shaping. Local feedback shaping will be described first. The analysis starts by considering the current transmission of a common emitter stage, equivalent circuit shown in Fig. 2(b). If the stage operates into a load resistance R_L, then to a good approximation the current transmission is given by

$$G_I = \frac{I_2}{I_1} = \frac{-\dfrac{a_0}{1 - a_0 + \delta}}{1 + \dfrac{p}{\omega_1} + \dfrac{p^2}{\omega_a \omega_c (1 - a_0 + \delta)}} \qquad (13)^*$$

where

$$\delta = \frac{R_L + r_e}{r_c}$$

$$\omega_1 = \frac{(1 - a_0 + \delta)}{\dfrac{1 + \delta}{\omega_a} + \dfrac{1}{\omega_c}}$$

$$\frac{\omega_a}{2\pi} \approx \text{alpha-cutoff frequency}$$

$$\omega_c = \frac{1}{(R_L + r_e)C_c}$$

It is apparent from expression (13) that if $(1 - a_0 + \delta)$ is less than 0.1, then the current gain of the common emitter stage falls off at a rate of 6 db per octave with a corner frequency at ω_1 .† A second 6 db per octave cutoff with a corner frequency at $[\omega_a + (1 + \delta)\omega_c]$ is introduced by the p^2 term in the denominator of (13). A typical transmission characteristic is shown in Fig. 5. The current gain of the common emitter stage is unity at a frequency equal to

$$\frac{a_0}{\dfrac{1 + \delta}{\omega_a} + \dfrac{1}{\omega_c}}$$

* Expressions (13) and (14) are poor approximations at frequencies above $\omega_a/2\pi$.

† Strictly speaking the corner frequency is equal to $\omega_1/2\pi$. However, for simplicity, corner frequencies will be expressed as radian frequencies.

Since the phase crossover of $A\beta$* is usually placed below this frequency, the principal effect of the second cutoff is to introduce excess phase. This excess phase can be minimized by operating the stage into the smallest load resistance possible, thus maximizing ω_c .

An undesirable property of the common emitter transmission characteristic is that the corner frequency ω_1 occurs at a relatively low frequency. However, the corner frequency can be increased by using local feedback as shown in Fig. 6(a). Shunt feedback is used in order to provide a low input impedance for the preceding stage to operate into. The amplitude and phase of the current transmission is controlled principally by the impedances Z_1 and Z_2 . If $| A\beta |$ is much greater than one, and if $\beta \approx Z_1/Z_2$, then from (7) the current transmission of the stage is approximately equal to $-Z_2/Z_1$. Because of the relatively small size of $A\beta$ for a single stage, this approximation is only valid for a very limited range of values of Z_1 and Z_2 . If Z_1 and Z_2 are represented as resistances R_1 and R_2 , then the current transmission of the circuit is given to a good approximation by

$$G_I = \frac{I_2}{I_1} = -\frac{R_2}{(R_2 + r_b)} \cdot \frac{\dfrac{a_0}{1 - a_0 + \gamma}}{\left[1 + \dfrac{p}{\omega_1'} + \dfrac{p^2}{\omega_a \omega_c (1 - a_0 + \gamma)}\right]} \tag{14}$$

where

$$\gamma = \frac{R_1 + r_e}{\dfrac{(R_2 + r_b) r_c}{R_2 + r_b + r_c}} \approx \frac{R_1 + r_e}{R_2 + r_b}$$

$$\omega_1' = \frac{(1 + a_0 + \gamma)}{\dfrac{1 + \gamma}{\omega_a} + \dfrac{1}{\omega_c}}$$

$$\omega_c = \frac{1}{(R_1 + r_e) C_e}$$

By comparing (14) with (13), it is evident that the negative feedback has reduced the low-frequency current gain from $a_0/(1 - a_0)$ (δ may usually be neglected) to

$$\left(\frac{R_2}{R_2 + r_b}\right)\left(\frac{a_0}{1 - a_0 + \gamma}\right) \approx \frac{R_2}{R_1 + r_e} \qquad (\text{if } \gamma > 1 - a_0)$$

* The phase crossover of Aβ is equal to the frequency at which the phase shift of Aβ from its low-frequency value is 180°.

The half power frequency, however, has been increased from

$$\frac{1 - a_0}{\frac{1}{\omega_a} + \frac{1}{\omega_c}} \quad \text{to} \quad \frac{1 - a_0 + \gamma}{\frac{1 + \gamma}{\omega_a} + \frac{1}{\omega_c}}$$

as shown by the dashed curves in Fig. 5.*

The bandwidth of the common emitter stage can be increased without reducing the current gain at dc and low-frequencies by representing Z_1 by a resistance R_1, and Z_2 by a resistance R_2 in series with a condenser C_2. If $1/R_2C_2$ is much smaller than ω_1', then the current transmission of the stage is given by (14) multiplied by the factor

$$\frac{\left(1 + \frac{\omega_2}{p}\right)}{\left(1 + \frac{\omega_4}{p}\right)} \tag{15}$$

where

$$\omega_2 = \frac{1}{R_2C_2}$$

$$\omega_4 = \frac{1 - a_0 + \frac{R_1 + r_e}{r_c}}{C_2(R_2 + r_b)(1 - a_0 + \gamma)}$$

The current transmission for this case is plotted in Fig. 6(b). The condenser C_2 introduces a rising 6 db per octave asymptote with a corner frequency at ω_2. At dc the current gain is equal to

$$\frac{a_0}{1 - a_0 + \delta}$$

A second method of shaping the loop current transmission characteristic of a feedback amplifier is by means of interstage networks. These networks are usually used for reducing the loop current gain at relatively low frequencies while introducing negligible phase lag near the gain† and phase crossover frequencies. Interstage networks should be designed to take advantage of the variable transistor input impedance. The input impedance of a transistor in the common emitter connection

* In Figs. 5 and 6(b), the factor $R_2/(R_2 + r_b)$ is assumed equal to unity. This is a good approximation since in practice R_2 is equal to several thousand ohms while r_b is equal to about 100 ohms.

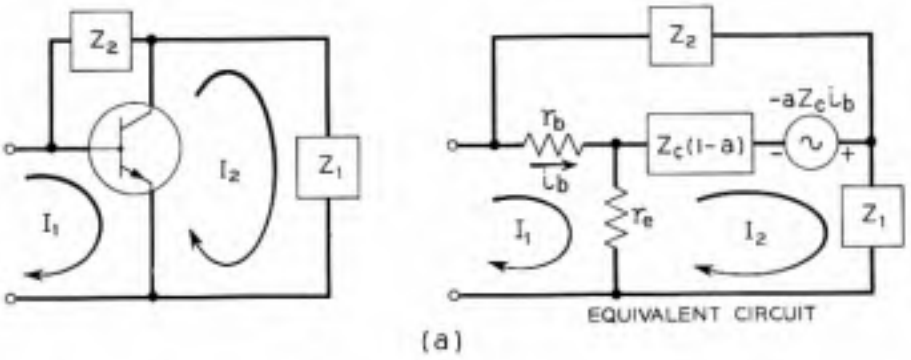
† The gain crossover frequency is equal to the frequency at which the magnitude of $A\beta$ is unity.

is given by the expression

$$Z_{\text{INPUT}} = r_b + r_e(1 - G_I) \tag{16}$$

where G_I is the current transmission given by (13). If G_I at dc is much greater than 1, then the input impedance and the current transmission of the common emitter stage fall off at about the same rate and with approximately the same corner frequency (ω_1). The input impedance finally reaches a limiting value equal to $r_e + r_b$.

A particularly useful interstage network is shown in Fig. 7(a). This network is analyzed in Appendix II and Fig. 7(b) shows a plot of the



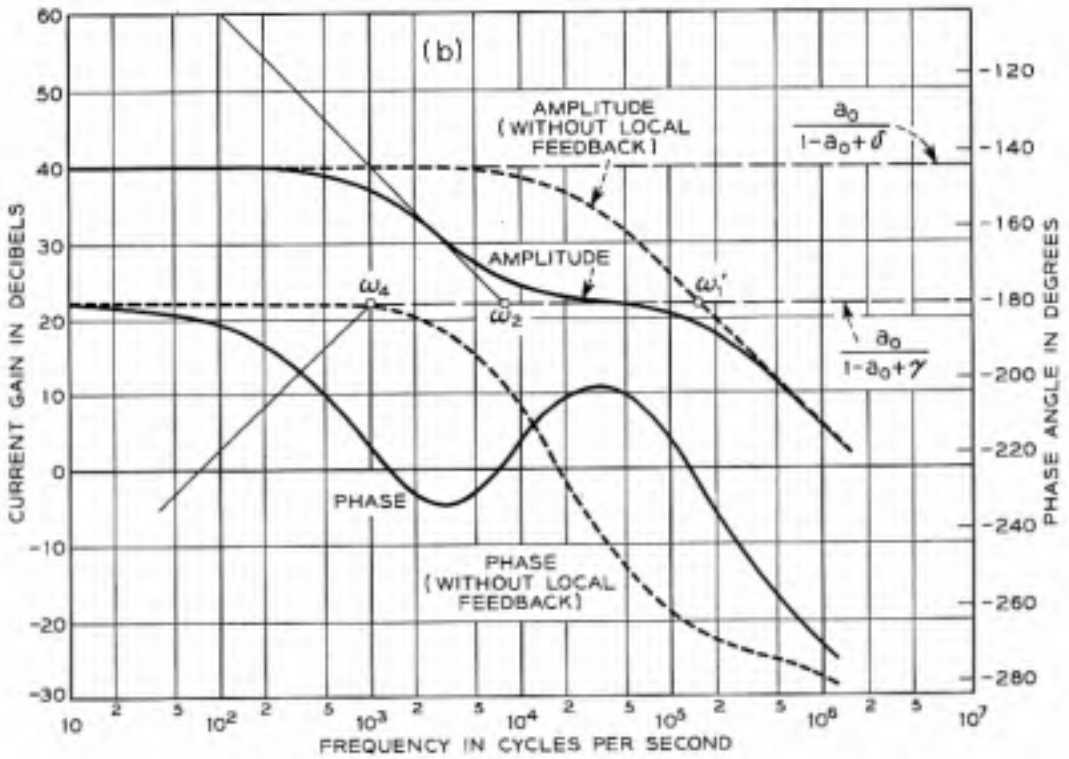


Fig. 6 — Negative feedback applied to a common emitter stage.

resulting current transmission. The amplitude of the transmission falls off at a rate of 6 db per octave with the corner frequency ω_5 determined by C_3 and the low frequency value of the transistor input impedance. The inductance L_3 introduces a 12 db per octave rising asymptote with a corner frequency at $\omega_3 = 1/\sqrt{L_3C_3}$. The corner frequencies ω_3 and ω_5 are selected in order to obtain a desirable loop current transmission characteristic (specific transmission characteristics are presented in Sections 3.0 and 4.0). The half power frequency of the current transmission of the transistor, ω_1, does not appear directly in the transmission characteristic of the circuit because of the variation in the transistor input impedance with frequency.

The overall β circuit of the feedback amplifier can also be used for

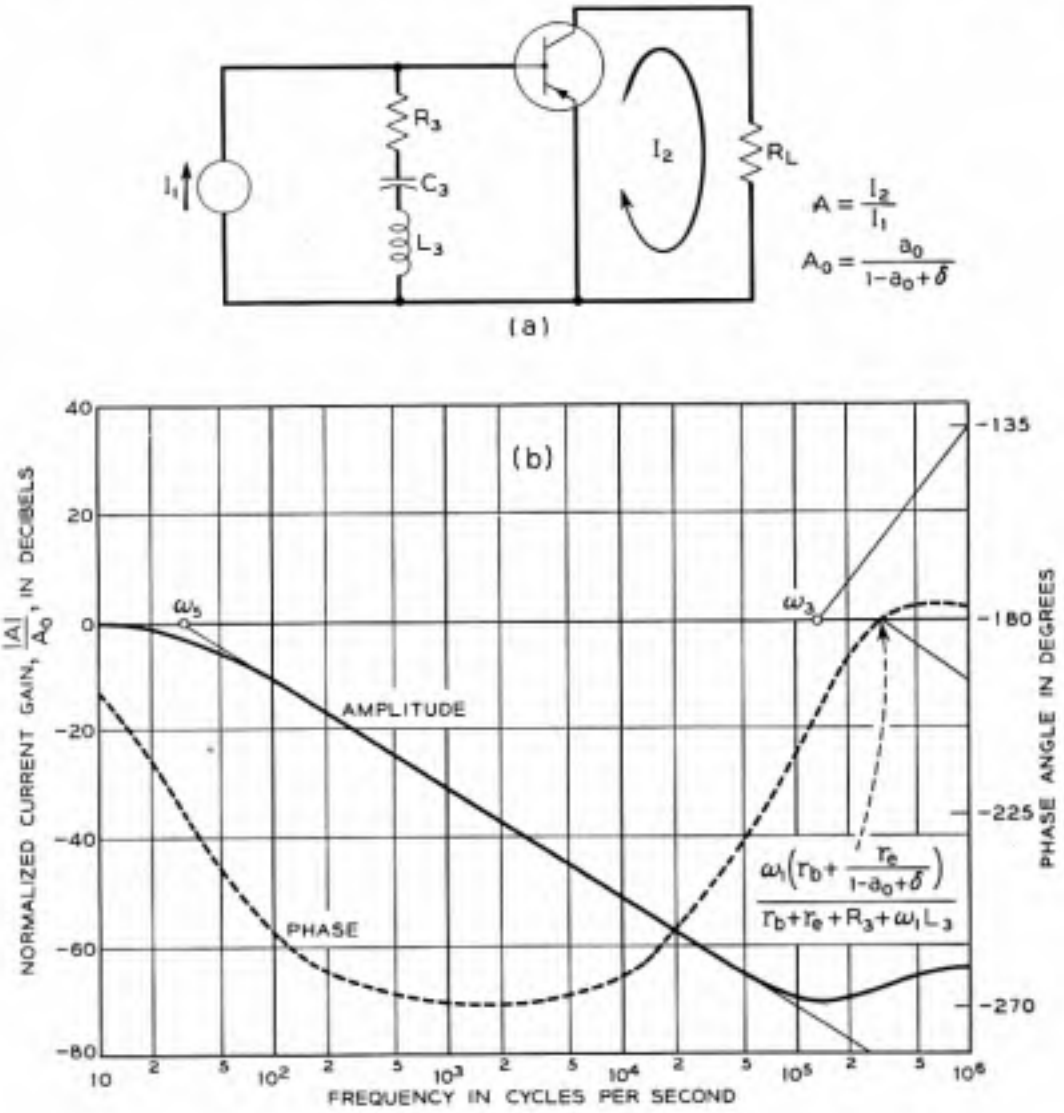


Fig. 7 — Interstage shaping network.

shaping the loop current transmission. If the feedback impedance Z_K (Fig. 4) consists of a resistance R_K and condenser C_K in parallel, then the loop current transmission is modified by the factor

$$\frac{\left(1 + \frac{p}{\omega_7}\right)}{\left(1 + \frac{p}{\omega_8}\right)} \tag{17}$$

where

$$\omega_7 = \frac{1}{R_K C_K}$$

$$\omega_8 = \frac{(R_L + R_K)}{R_L R_K C_K}$$

Since Z_K affects the external voltage gain of the operational amplifier, (11), the corner frequency ω_7 must be located outside of the useful frequency band. Usually it is placed near the gain crossover frequency in order to improve the phase margin and the transient response of the amplifier.

In Sections 3.0 and 4.0, the above shaping techniques are used in the design of specific operational amplifiers.

3.0. THE SUMMING AMPLIFIER

3.1. *Circuit Arrangement*

The schematic diagram of a dc summing amplifier is shown in Fig. 8. From the discussion in Section 2.0 it is apparent that each common emitter stage will contribute more than 90 degrees of high-frequency phase lag. Consequently, while the magnitude of the low-frequency feedback increases with the number of stages, this is at the expense of the bandwidth over which the negative feedback can be maintained. It is possible to develop 80 db of negative feedback at dc with three common emitter stages. This corresponds to a dc accuracy of one part in 10,000. In addition, the feedback can be maintained over a broad enough band in order to permit full accuracy to be attained in about 100 microseconds. Thus it is evident that the choice of three stages represents a satisfactory compromise between accuracy and bandwidth objectives.

The output stage of the amplifier is designed for a maximum power dissipation of 75 milliwats and maximum voltage swing of ±25 volts

when operating into an external load resistance equal to or greater than 50,000 ohms. A p-n-p transistor is used in the second stage and n-p-n transistors are used in the first and third stages. This circuit arrangement makes it possible to connect the collector of one transistor directly to the base of the following transistor without introducing appreciable interstage loss. "Shot" noise[11] and dc drift are minimized by operating the first stage at the relatively low collector current of 0.25 milliamperes. The 110,000-ohm resistor provides the collector current for the first stage, and the 4,700-ohm resistor provides 3.8 milliamperes of collector current for the second stage. The series 6,800-ohm resistor between the second and third stages, reduces the collector to emitter potential of the second stage to about 4.5 volts.

The loop current transmission is shaped by use of local feedback applied to the second stage, by an interstage network connected between the second and third stages, and by the overall β circuit. The 200-ohm resistor in the collector circuit of the second stage is, with reference to Fig. 6(a), Z_1 . The impedance of the interstage network can be neglected since it is small compared to 200 ohms at all frequencies for which the local feedback is effective. The interstage network is connected between the second and third stages in order to minimize the output noise voltage. With this circuit arrangement, practically all of the output noise voltage

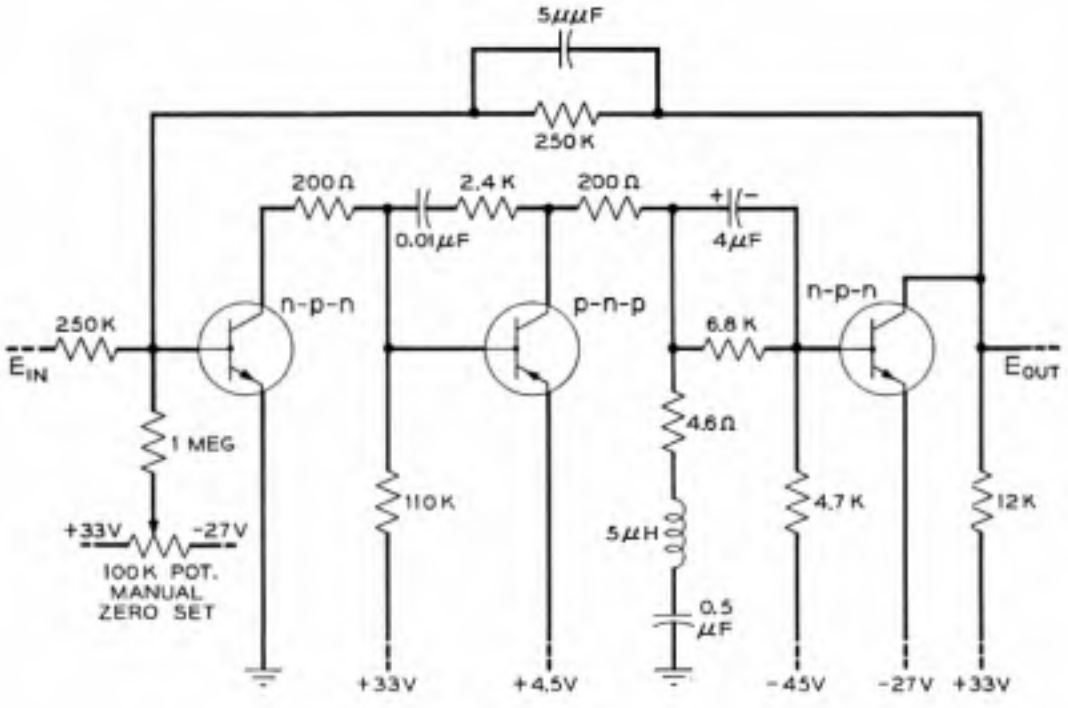


Fig. 8 — DC summing amplifier.

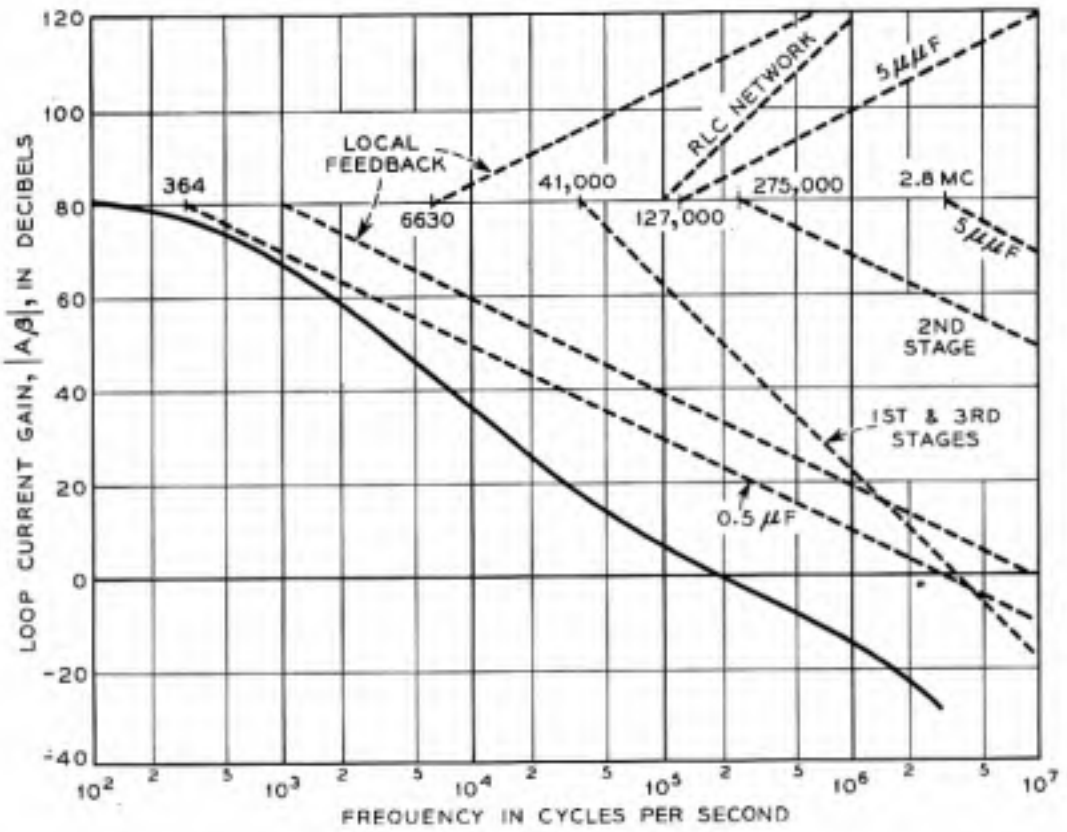


Fig. 9 — Gain-frequency asymptotes for summing amplifier.

is generated in the first transistor stage. If the transistor in the first stage has a noise figure less than 10 db at 1,000 cycles per second, then the RMS output noise voltage is less than 0.5 millivolts.

Fig. 9 shows a plot of the gain-frequency asymptotes for the summing amplifier determined from (13), (14), (15), (17), and (A6) under the assumption that the alphas and alpha-cutoff frequencies of the transistors are 0.985 and 3 mc, respectively. The corner frequencies introduced by the 0.5 microfarad condenser in the interstage network, the local feedback circuit, and the cutoff of the first and third stages are so located that the current transmission falls off at an initial rate of about 9 db per octave. This slope is joined to the final asymptote of the loop transmission by means of a step-type of transition.[12] The transition is provided by 3 rising asymptotes due to the interstage shaping network, and the overall β circuit. An especially large phase margin is used in order to insure a good transient performance.

Fig. 10 shows the amplitude and phase of the loop current transmission. When the amplitude of the transmission is 0 db, the phase angle is $-292°$, and when the phase angle is $-360°$, the amplitude is 27.5 db

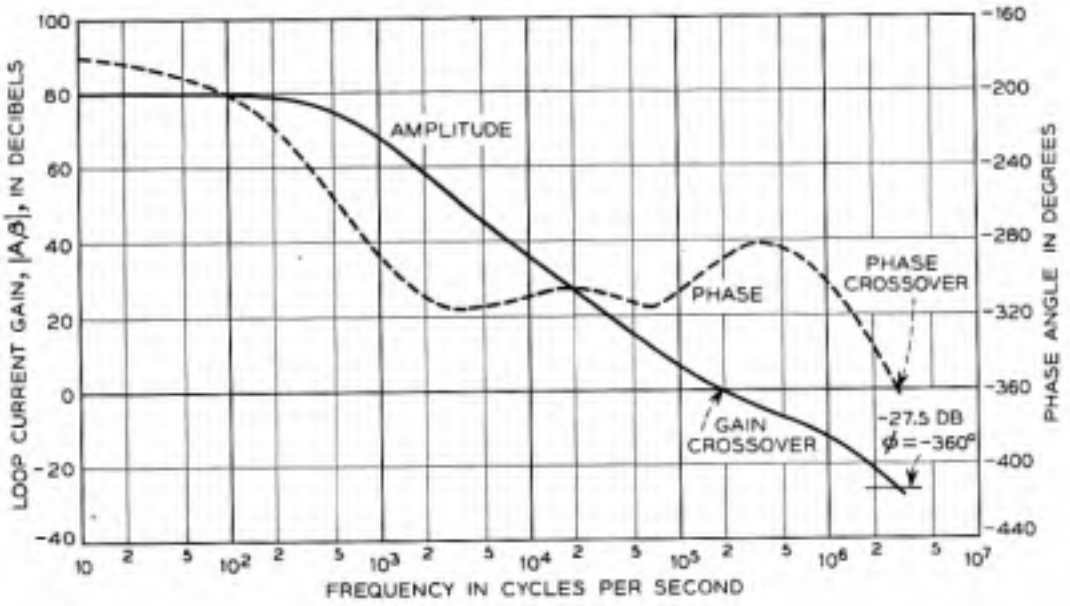


Fig. 10 — Loop current transmission of the summing amplifier.

below 0 db. The amplifier has a 68° phase margin and 27.5 db gain margin. In order to insure sufficient feedback at dc and adequate margins against instability, the transistors used in the amplifier should have alphas in the range 0.98 to 0.99 and alpha-cutoff frequencies equal to or greater than 2.5 mc.

3.2. *Automatic Zero Set of the dc Summing Amplifier*

The application of germanium junction transistors to dc amplifiers does not eliminate the problem of drift normally encountered in vacuum tube circuits. In fact, drift is more severe due principally to the variation of the transistor parameters alpha and saturation current with temperature variation. Even though the amplifier has 80 db of negative feedback at dc, this feedback does not eliminate the drift introduced by the first transistor stage. Because of the large amount of dc feedback, the collector current of the first stage is maintained relatively constant. The collector current of the transistor is related to the base current by the equation

$$I_c = \frac{I_{cr}}{1 - a} + \frac{a}{1 - a} I_b \tag{18}$$

The saturation current, I_{CO}, of a germanium junction transistor doubles approximately for every 11°C increase in temperature. The factor $a/(1 - a)$ increases by as much as 6 db for a 25°C increase in tempera-

ture. Consequently, the base current of the first stage, I_b, and the output voltage of the amplifier must change with temperature in order to maintain I_c constant. The drift due to the temperature variation in a can be reduced by operating the first stage at a low value of collector current. With a germanium junction transistor in the first stage operating at a collector current of 0.25 milliamperes, the output voltage of the amplifier drifts about ± 1.5 volts over a temperature range of 0°C to 50°C. It is possible to reduce the dc drift by using temperature sensitive elements in the amplifier.[13, 14] In general, temperature compensation of a transistor dc amplifier requires careful selection of transistors and critical adjustment of the dc biases. However, even with the best adjustments, temperature compensation cannot reduce the drift in the amplifier to within typical limits such as ± 5 millivolts throughout a temperature range of 0 to 50°C. In order to obtain the desired accuracy it is necessary to use an automatic zero set (AZS) circuit.

Fig. 11 shows a dc summing amplifier and a circuit arrangement for reducing any dc drift that may appear at the output of the amplifier. The output voltage is equal to the negative of the sum of the input voltages, where each input voltage is multiplied by the ratio of the feedback resistor to its input resistor. In addition, an undesirable dc drift voltage is also present in the output voltage. The total output voltage is

$$E_{\text{out}} = -\sum_{j=1}^{N} E_j \frac{R_K}{R_j} + E_{\text{drift}} \tag{19}$$

In order to isolate the drift voltage, the N input voltages and the output voltage are applied to a resistance summing network composed of resistors R_0, R_1', R_2', $\cdots$, R_N'. The voltage across R_s is equal to

$$E_s = \frac{R_s}{R_0} E_{\text{drift}} \tag{20}$$

if

$$R_s \ll R_0, R_j'; \qquad j = 1, 2, \cdots, N$$

and

$$R_0 R_j = R_K R_j'; \qquad j = 1, 2, \cdots, N$$

The voltage E_s is amplified in a relatively drift-free narrow band dc amplifier and is returned as a drift correcting voltage to the input of the dc summing amplifier. If the gain of the AZS circuit is large, the drift voltage at the output of the summing amplifier can be made very small.

Fig. 12 shows the circuit diagram of a summing amplifier which uses a mechanical chopper in the AZS circuit.[15] The AZS circuit consists of a

resistance summing network, a 400-cycle synchronous chopper, and a tuned 400-cycle amplifier. Any drift in the summing amplifier will produce a dc voltage E_s at the output of the summing network. The chopper converts the dc voltage into a 400 cycles per second waveform. The fundamental frequency in the waveform is amplified by a factor of about 400,000 by the tuned amplifier. The synchronous chopper rectifies the sinusoidal output voltage and preserves the original dc polarity of E_s. The rectified voltage is filtered and fed back to the summing amplifier as an additional input current. The loop voltage gain of the AZS circuit at dc is about 54 db. Any dc or low-frequency drift in the summing amplifier is reduced by a factor of about 500 by the AZS circuit. The drift throughout a temperature range of 0 to 50°C is reduced to ±3 millivolts.

Since the drift in the summing amplifier changes at a relatively slow rate, the loop voltage gain of the AZS circuit can be cutoff at a relatively low frequency. In this particular case the loop voltage gain is zero db at about 10 cycles per second.

4.0. THE INTEGRATOR

4.1. *Basic Design Considerations*

The design principles previously discussed are illustrated in this section by the design of a transistor integrator for application in a voltage

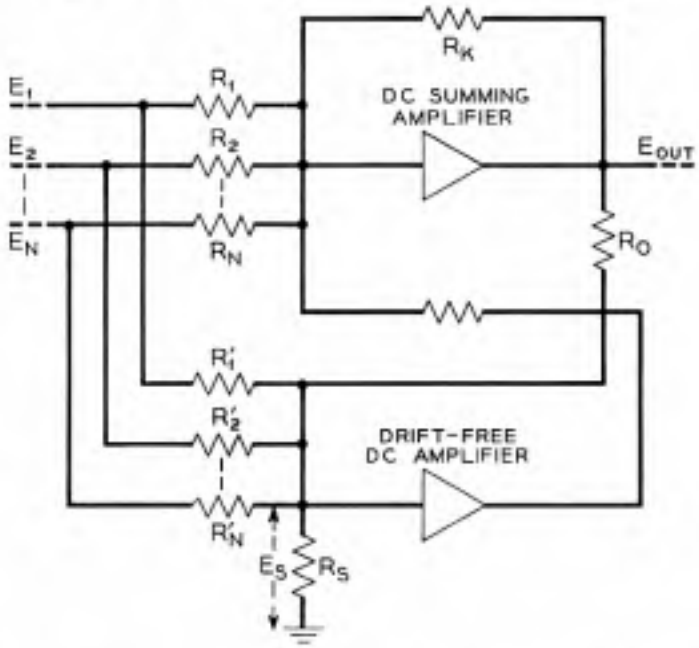


Fig. 11 — DC summing amplifier with automatic zero set.

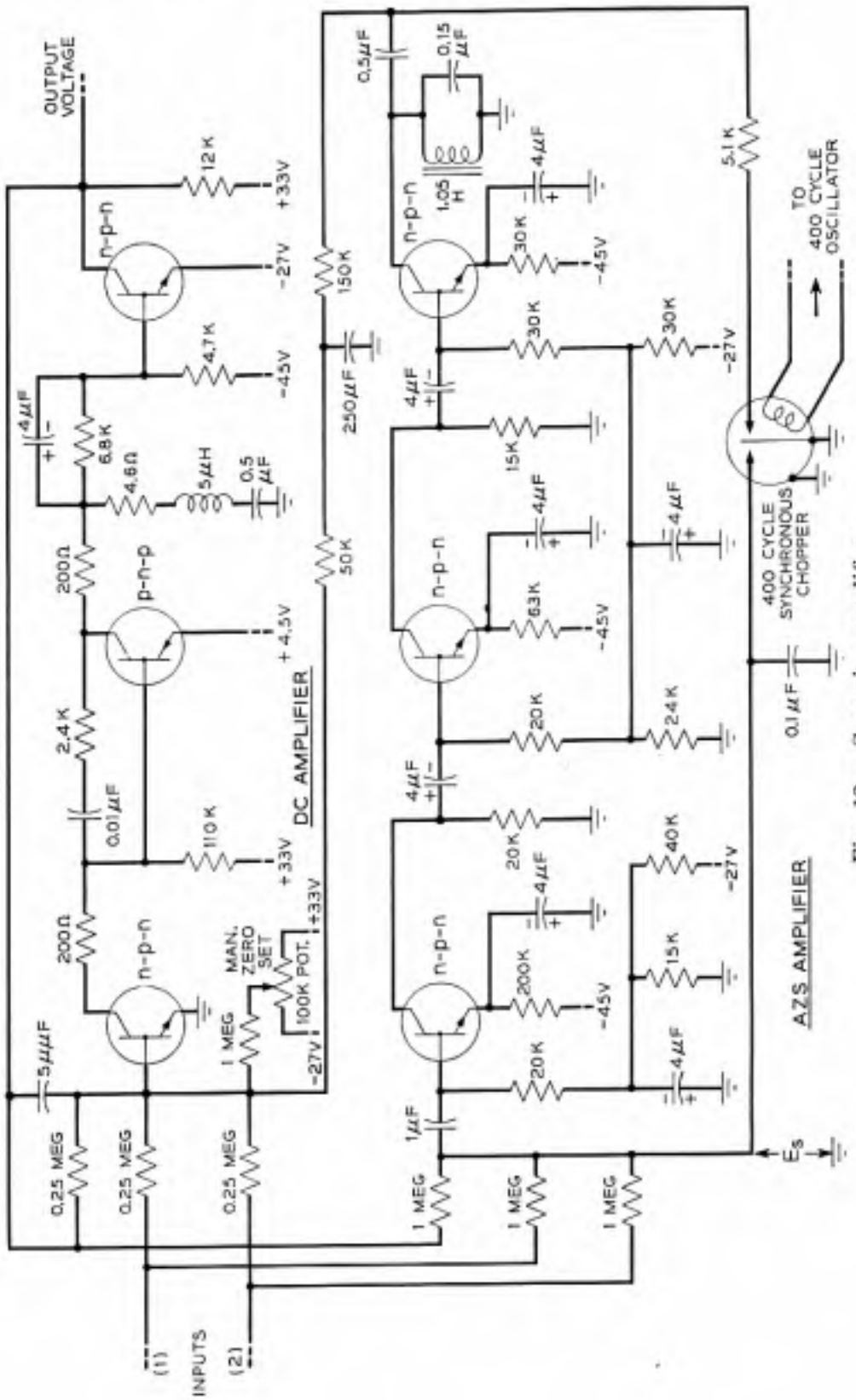


Fig. 12 — Summing amplifier.

encoder. The integrator is required to generate a 15-volt ramp which is linear and has a constant slope to within one part in 8,000. This ramp is to have a slope of 5 millivolts per microsecond for an interval of 3,000 microseconds.

The first step in the design is to determine the bandwidth over which the negative feedback must be maintained in order to realize the desired output voltage linearity. The relationship between the output and input voltage of the integrator can be obtained from expression (11) by substituting $(1/pc)$ for Z_K and R for Z_j (refer to Fig. 1).

$$\mathfrak{L}[E_{OUT}] = \frac{\mathfrak{L}[E_{IN}]}{pRC}\left[\frac{A\beta + Z_{IN}'pC}{1 - A\beta + \dfrac{Z_{IN}'}{R}}\right] \tag{21}$$

where $\mathfrak{L}[E_{OUT}]$ and $\mathfrak{L}[E_{IN}]$ are the Laplace transforms of the output and input voltages, respectively. In order to generate the voltage ramp, a step voltage of amplitude E is applied to the input of the integrator. The term Z_{IN}'/R is negligible compared to unity at all frequencies. Therefore,

$$\mathfrak{L}[E_{OUT}] = \frac{E}{p^2RC}\left[\frac{A\beta}{1 - A\beta}\right] + \frac{EZ_{IN}'}{pR}\left[\frac{1}{1 - A\beta}\right] \tag{22}$$

It will be assumed that $A\beta$ is given by the expression

$$A\beta = \frac{-K}{\left(1 + \dfrac{\omega_1}{p}\right)\left(1 + \dfrac{p}{\omega_2}\right)^2} \tag{23}$$

Expression (23) implies that $A\beta$ falls off at a rate of 6 db per octave at low frequencies and 12 db per octave at high frequencies. The output voltage of the integrator, as a function of time, is readily evaluated by substituting (23) into (22) and taking the inverse Laplace transform of the results. A good approximation for the output voltage is

$$E_{OUT} = -\frac{E}{RC}\left[t - \frac{\omega_1 t^2}{2K} - \frac{e^{-[(2\omega_2+\omega_1)t/2]}\sin\sqrt{K}\omega_2 t}{\sqrt{K}\omega_2}\right] + \frac{ER_{IN}'}{R}\left[1 - e^{-(\omega_1 t/K)} + e^{-[(2\omega_2+\omega_1)t/2]}\cos\sqrt{K}\omega_2 t\right] \tag{24}*$$

The linear voltage ramp is expressed by the term $-(Et/RC)$. The additional terms introduce nonlinearities. The voltage ramp has a slope of 5 millivolts per microsecond for $E = -21$ volts, $R = 42{,}000$ ohms,

* In evaluating E_{OUT} it was assumed that Z_{IN}' was equal to a fixed resistance R_{IN}', the low frequency input resistance to the first common emitter stage. A complete analysis indicates that this assumption makes the design conservative.

and $C = 0.1$ microfarads. For these circuit values, and $K = 10{,}000$ (corresponding to 80 db of feedback) the nonlinear terms are less than 1/8,000 of the linear term (evaluated when $t = 4 \times 10^{-3}$ seconds) if $f_1 \leqq 30$ cycles per second, $f_2 \geqq 800$ cycles per second, and if the first 1000 microseconds of the voltage ramp are not used. Consequently, 80 db of negative feedback must be maintained over a band extending from 30 to 800 cycles per second in order to realize the desired output voltage linearity.

4.2. *Detailed Circuit Arrangement*

Fig. 13 shows the circuit diagram of the integrator. The method of biasing is the same as is used in the summing amplifier. The 200,000-ohm resistor provides approximately 0.5 milliamperes of collector current for the first stage. The 40,000-ohm resistor provides approximately 0.9 milliamperes of collector current for the second stage. The output stage is designed for a maximum power dissipation of 120 milliwatts and for an output voltage swing between -5 and $+24$ volts when operating into a load resistance equal to or greater than 40,000 ohms.

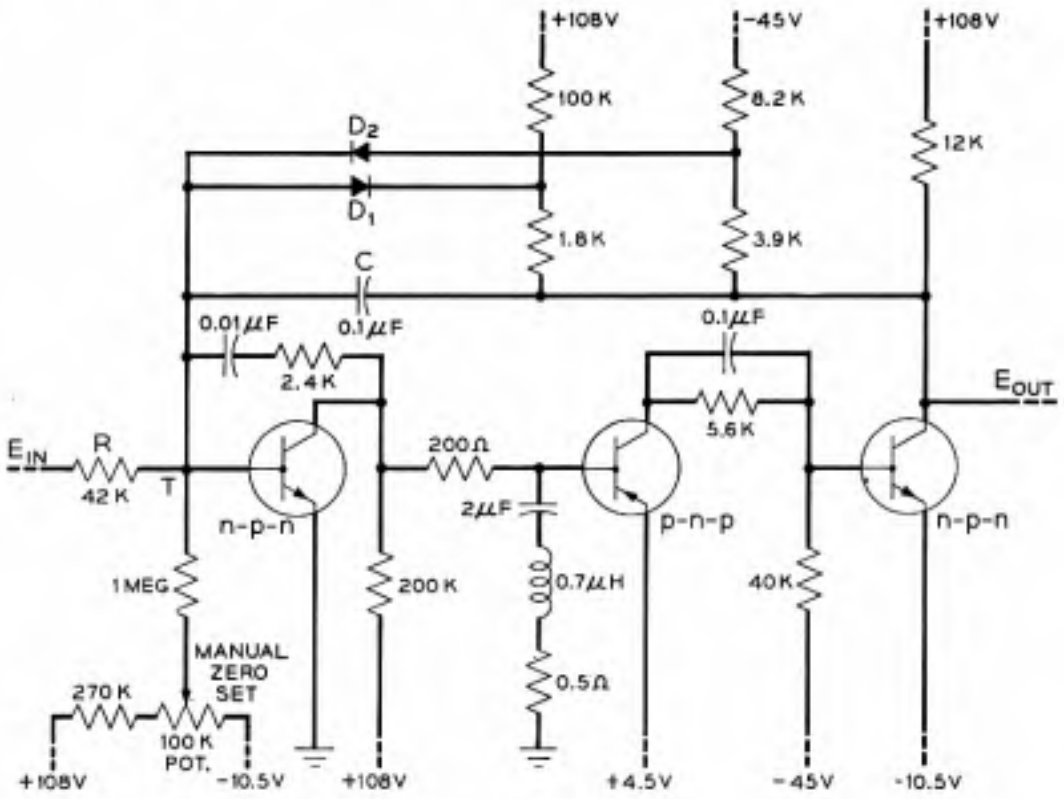


Fig. 13 — Integrator.

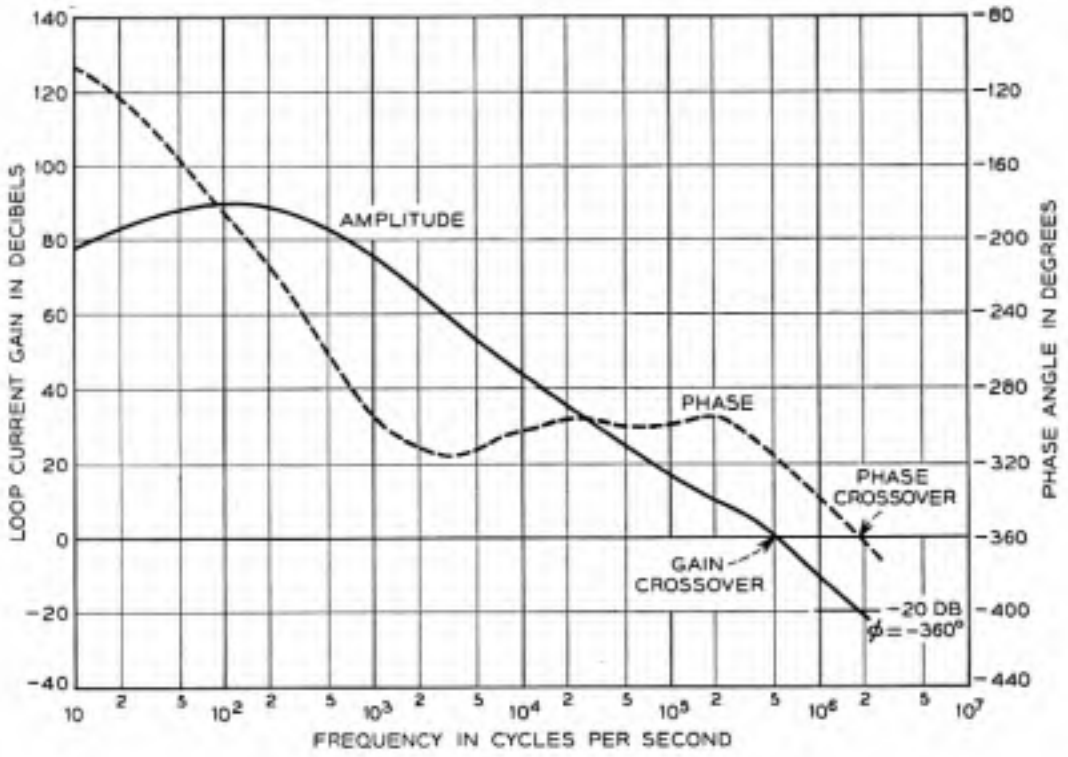


Fig. 14 — Loop current transmission of the integrator.

The negative feedback in the integrator has been shaped by means of local feedback and interstage networks as described in Section 2.2. The loop current transmission has been calculated from (13), (14), (15), and (A6) and is plotted in Fig. 14. The transmission is determined under the assumption that the alphas of the transistors are 0.985 and the alpha-cutoff frequencies are three megacycles. Since the feedback above 800 cycles per second falls off at a rate of 9 db per octave, the analysis in Section 4.1 using (23), is conservative. The integrator has a 44° phase margin and a 20 db gain margin. In order to insure sufficient feedback between 30 and 800 cycles per second and adequate margins against instability, the transistors used in the integrator should have alphas in the range 0.98 to 0.99 and alpha-cutoff frequencies equal to or greater than 2.5 megacycles.

The silicon diodes D_1 and D_2 are required in order to prevent the integrator from overloading. For output voltages between -4.0 and 21 volts the diodes are reverse biased and represent very high resistances, of the order of 10,000 megohms. If the output voltage does not lie in this range, then one of the diodes is forward biased and has a low resistance, of the order of 100 ohms. The integrator is then effectively a dc amplifier with a voltage gain of approximately 0.1. The silicon diodes affect the

linearity of the voltage ramp slightly due to their finite reverse resistances and variable shunt capacities. If the diodes have reverse resistances greater than 1000 megohms, and if the maximum shunt capacity of each diode is less than 10 micromicrofarads (capacity with minimum reverse voltage), then the diodes introduce negligible error.

As stated earlier, the integrator generates a voltage ramp in response to a voltage step. This step is applied through a transistor switch which is actuated by a square wave generator capable of driving the transistor well into current saturation. Such a switch is required because the equivalent generator impedance of the applied step voltage must be very small. A suitable circuit arrangement is shown in Fig. 15. For the particular application under discussion the switch S is closed for 5,000 microseconds. During this time, the voltage $E = -21V$ appears at the input of the integrator. At the end of this time interval, the transistor switch is opened and a reverse current is applied to the feedback condenser C, returning the output voltage to -4.0 volts in about 2500 microseconds. An alternate way of specifying a low impedance switch is to say that the voltage across it be close to zero. For the transistor switch, connected as shown in Fig. 15, this means that its collector voltage be within

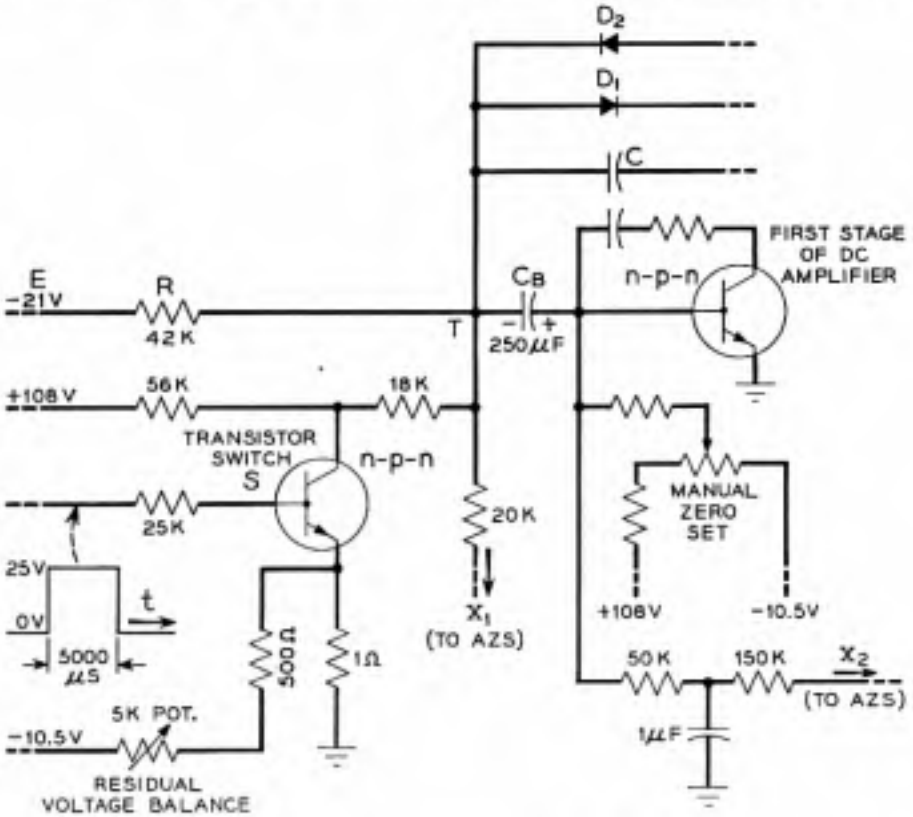


Fig. 15 — Input circuit arrangement of the integrator.

one millivolt of ground potential during the time the transistor is in saturation. Now, it has been shown[16] that when a junction transistor in the common emitter connection is driven into current saturation, the minimum voltage between collector and emitter is theoretically equal to

$$\frac{kT}{q} \ell n \frac{1}{\alpha_i} \tag{25}$$

where k is the Boltzmann constant, T is the absolute temperature, q is the charge of an electron ($(kT/q) = 26$ millivolts at room temperature), and α_i is the inverse alpha of the transistor, i.e., the alpha with the emitter and collector interchanged. There is an additional voltage drop across the transistor due to the bulk resistance of the collector and emitter regions (including the ohmic contacts). A symmetrical alloy junction transistor with an alpha close to unity is an excellent switch because both the collector to emitter voltage and the collector and emitter resistances are very small.

At the present time, a reasonable value for the residual voltage* between the collector and emitter is 5 to 10 millivolts. This voltage can be eliminated by returning the emitter of the transistor switch to a small negative potential. This method of balancing is practical because the voltage between the collector and emitter of the transistor does not change by more than 1.0 millivolt over a temperature range of 0°C to 50°C.

4.3. *Automatic Zero Set of the Integrator*

A serious problem associated with the transistor integrator is drift. The drift is introduced by two sources; variations in the base current of the first transistor stage and variations in the base to emitter potential of the first stage with temperature. In order to reduce the drift, the input resistor R and the feedback condenser C must be dissociated from the base current and base to emitter potential of the first transistor stage. This is accomplished by placing a blocking condenser C_B between point T and the base of the first transistor as shown in Fig. 15. An automatic zero set circuit is required to maintain the voltage at point T equal to zero volts. This AZS circuit uses a magnetic modulator known as a "magnettor."[17]

A block diagram of the AZS circuit is shown in Fig. 16. The dc drift current at the input of the amplifier is applied to the magnettor. The carrier current required by the magnettor is supplied by a local transistor

* The inverse alphas of the transistors used in this application were greater than 0.95.

oscillator. The useful output of the magnettor is the second harmonic of the carrier frequency. The amplitude of the second harmonic signal is proportional to the magnitude of the dc input current and the phase of the second harmonic signal is determined by the polarity of the dc input current. The output voltage of the magnettor is applied to an active filter which is tuned to the second harmonic frequency. The signal is then amplified in a tuned amplifier and applied to a diode gating circuit. Depending on the polarity of the dc input current, the gating circuit passes either the positive or negative half cycle of the second harmonic signal. In order to accomplish this, a square wave at a repetition rate equal to that of the second harmonic signal is derived from the carrier oscillator and actuates the gating circuit.

A circuit diagram of the AZS circuit is shown in Figs. 17(a) and 17(b). The various sections of the circuit are identified with the blocks shown in Fig. 16. The active filter is adjusted for a Q of about 300, and the gain of the active filter and tuned amplifier is approximately 1000. The AZS circuit provides ± 1.0 volt of dc output voltage for ± 0.05 microamperes of dc input current. The maximum sensitivity of the circuit is limited to ± 0.005 microamperes because of residual second harmonic generation in the magnettor with zero input current.

When the transistor integrator is used together with the magnettor AZS circuit, the slope of the voltage ramp is maintained constant to within one part in 8,000 over a temperature range of 20°C to 40°C.

5.0. *The Voltage Comparator*

The voltage comparator is one of the most important circuits used in analog to digital converters. The comparator indicates the exact time that an input waveform passes through a predetermined reference level. It has been common practice to use a vacuum tube blocking oscillator as a voltage comparator.[18] Due to variations in the contact potential, heater voltage, and transconductance of the vacuum tube, the maximum

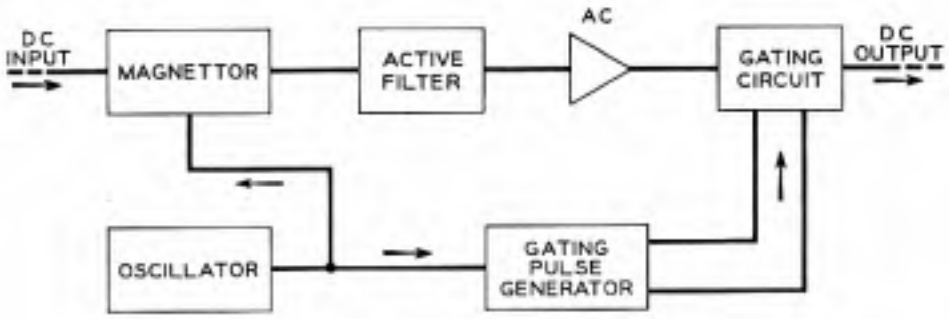


Fig. 16 — Block diagram of AZS circuit.

accuracy of the circuit is limited to about ±100 millivolts. By taking advantage of the properties of semiconductor devices, the transistor blocking oscillator comparator can be designed to have an accuracy of ±5 millivolts throughout a temperature range of 20°C to 40°C.

5.1. *General Description of the Voltage Comparator*

Fig. 18 shows a simplified circuit diagram of the voltage comparator. Except for the silicon junction diode D_1, this circuit is essentially a transistor blocking oscillator. For the purpose of analysis, assume that the reference voltage V_{cc} is set equal to zero. When the input voltage V_i is large and negative, the silicon diode D_1 is an open circuit and the junction transistor has a collector current determined by R_b and E_{bb} [Expression (18)]. The base of the transistor resides at approximately -0.2 volts. As the input voltage V_i approaches zero, the reverse bias across the diode D_1 decreases. At a critical value of V_i (a small positive potential), the dynamic resistance of the diode is small enough to permit the circuit to become unstable. The positive feedback provided by transformer T_1 forces the transistor to turn off rapidly, generating a sharp output pulse across the secondary of transformer T_2. When V_i is large and positive, the diode D_1 is a low impedance and the transistor is maintained cutoff. In order to prevent the comparator from generating more than one output pulse during the time that the circuit is unstable, the natural period of the circuit as a blocking oscillator must be properly chosen. Depending on this period, the input voltage waveform must have a certain minimum slope when passing through the reference level in order to prevent the circuit from misfiring.

The comparator has a high input impedance except during the switching interval.* When V_i is negative with respect to the reference level, the input impedance is equal to the impedance of the reverse biased silicon diode. When V_i is positive with respect to the reference level, the input impedance is equal to the impedance of the reverse biased emitter and collector junctions in parallel. This impedance is large if an alloy junction transistor is used. During the switching interval the input impedance is equal to the impedance of a forward biased silicon diode in series with the input impedance of a common emitter stage (approximately 1,000 ohms). This loading effect is not too serious since for the circuit described, the switching interval is less than 0.5 microseconds.

The voltage comparator shown in Fig. 18 operates accurately on voltage waveforms with positive slopes. The voltage comparator will operate accurately on waveforms with negative slopes if the diode and

* The switching interval is the time required for the transistor to turn off.

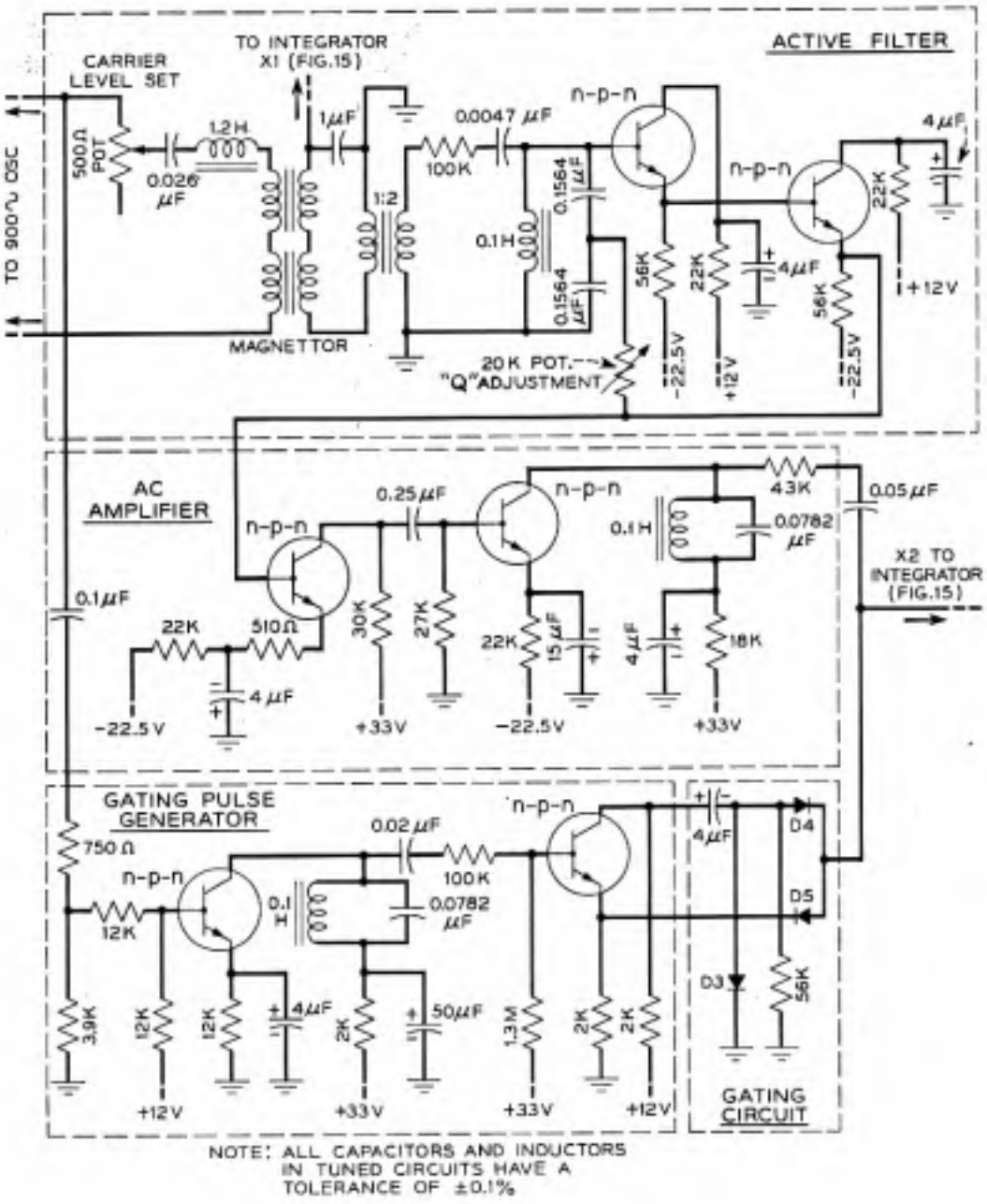


Fig. 17(a) — AZS circuit.

battery potentials are reversed and if an n-p-n junction transistor is used.

5.2. *Factors Determining the Accuracy of the Voltage Comparator*

Fig. 19 shows the ac equivalent circuit of the voltage comparator. In the equivalent circuit R_1 is the dynamic resistance of the diode D_1, Rg is the source resistance of the input voltage, and R_2 is the impedance of

the load R_L as it appears at the primary of the transformer T_2. R_1 is a function of the dc voltage across the diode D_1. At a prescribed value of R_1, the comparator circuit becomes unstable and switches. The relationship between this critical value of R_1 and the transistor and circuit parameters is obtained by evaluating the characteristic equation for the circuit and by determining the relationship which the coefficients of the equation must satisfy in order to have a root of the equation lie in the right hand half of the complex frequency plane. To a good approximation, the critical value of R_1 is given by the expression

$$R_1 + R_g + r_b = \frac{Ma_0}{R_2C_e + \frac{a_0}{\omega_a}} \tag{26}$$

where M is the mutual inductance of transformer T_1 and $R_2 = N'^2R_L$. Since the transistor parameters which appear in expression (26) have only a small variation with temperature, the critical value of R_1 is independent of temperature (to a first approximation).

It will now be shown that the comparator can be designed for an accuracy of ± 5 millivolts throughout a temperature range of 20°C to 40°C. In order to establish this accuracy it will be assumed that the critical value of R_1 is equal to 30,000 ohms. This assumption is based on the

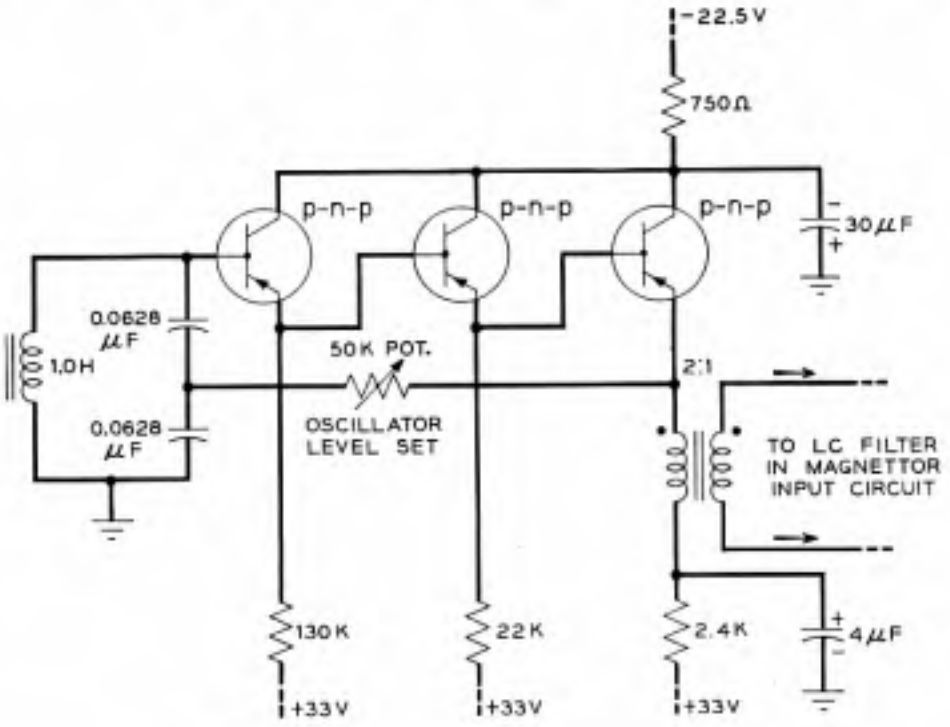


Fig. 17(b), 900-cycle carrier oscillator.

data displayed in Fig. 20 which gives the volt-ampere characteristics of a silicon diode measured at 20°C and 40°C. Throughout this temperature range, the diode voltage corresponding to the critical resistance of 30,000 ohms changes by about 30 millivolts. Fortunately, part of this voltage variation with temperature is compensated for by the variation in voltage V_{b-e} between the base and emitter of the junction transistor. From Fig. 18,

$$V_i = V_D - V_{b-e} + V_{ee} \tag{27}$$

For perfect compensation (V_i independent of temperature), V_{b-e} should have the same temperature variation as the diode voltage V_D . Experi-

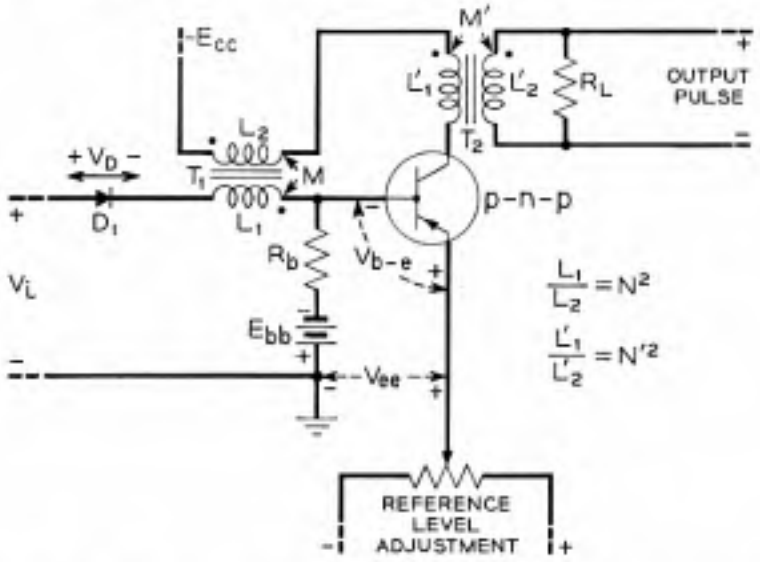


Fig. 18 — Simplified circuit diagram of voltage comparator.

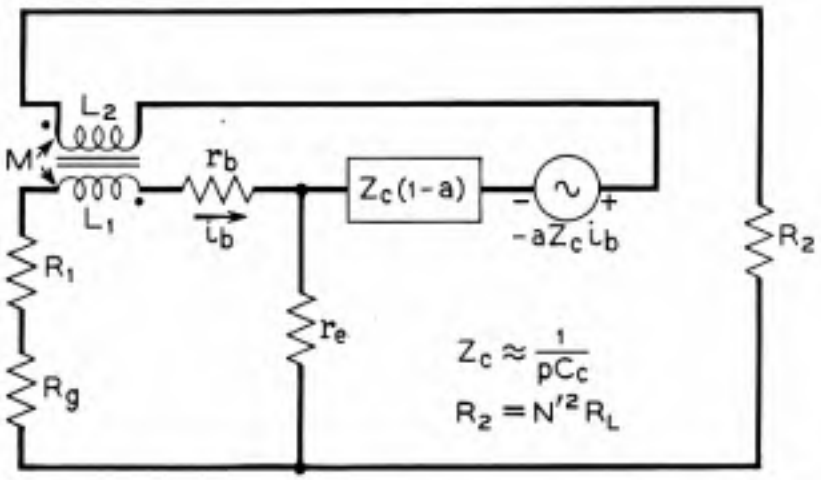


Fig. 19 — Equivalent circuit of voltage comparator.

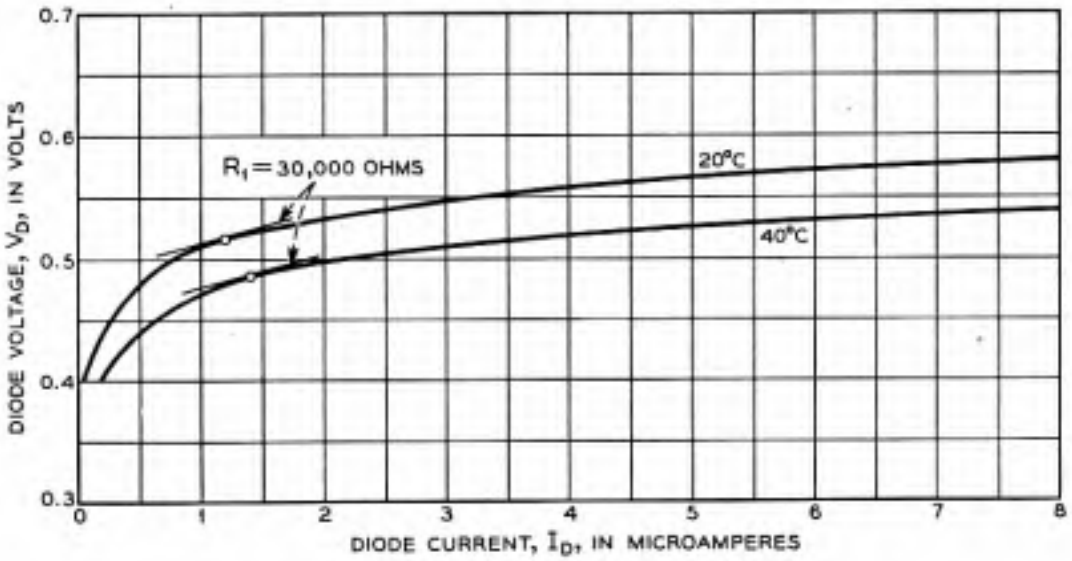


Fig. 20 — Volt-ampere characteristic of a silicon junction diode.

mentally it is found that V_{b-e} for germanium junction transistors varies by about 20 millivolts throughout the temperature range of 20°C to 40°C. Consequently, the variation in V_i at which the circuit switches is ± 5 millivolts.

It is apparent from Fig. 20 that the accuracy of the comparator increases slightly for critical values of R_1 greater than 30,000 ohms, but decreases for smaller values. For example, the accuracy of the comparator is ± 10 millivolts for a critical value of R_1 equal to 5,000 ohms. In general, the critical value of R_1 should be chosen between 5,000 and 100,000 ohms.

5.3. *A Practical Voltage Comparator*

Fig. 21 shows the complete circuit diagram of a voltage comparator. The circuit is designed to generate a sharp output pulse* when the input voltage waveform passes through the reference level (set by V_{cc}) with a positive slope. The pulse is generated by the transistor switching from the "on" state to the "off" state. To a first approximation the amplitude of the output pulse is proportional to the transistor collector current during the "on" state. When the input voltage waveform passes through the reference level with a negative slope an undesirable negative pulse is generated. This pulse is eliminated by the point contact diode D_2.

The voltage comparator is an unstable circuit and has the properties

* For the circuit values shown in Fig. 21, the output pulse has a peak amplitude of about 6 volts, a rise time of 0.5 microseconds, and a pulse width of about 2.0 microseconds.

of a free running blocking oscillator after the input voltage V_i passes through the reference level. After a period of time the transistor will return to the "on" state unless the voltage V_i is sufficiently large at this time to prevent switching. In order to minimize the required slope of the input waveform the time interval between the instant V_i passes through the reference level and the instant the transistor would naturally switch to the "on" state must be maximized. This time interval can be controlled by connecting a diode D_3 across the secondary winding of transformer T_1. When the transistor turns off, the current which was flowing through the secondary of transformer $T_1(I_c)$ continues to flow through the diode D_3 so that L_2 and D_3 form an inductive discharge circuit. The point contact diode D_3 has a forward dynamic resistance of less than 10 ohms and a forward voltage drop of 0.3 volt. If the small forward resistance of the diode is neglected, the time required for the current in the circuit to fall to zero is

$$T = \frac{I_c L_2}{0.3} \tag{28}$$

During the inductive transient, 0.3 volt is induced into the primary of transformer T_1 (since $N = 1$) maintaining the transistor cutoff. The duration of the inductive transient can be made as long as desired by increasing L_2. However, there is the practical limitation that increasing L_2 also increases the leakage inductance of transformer T_1, and in turn,

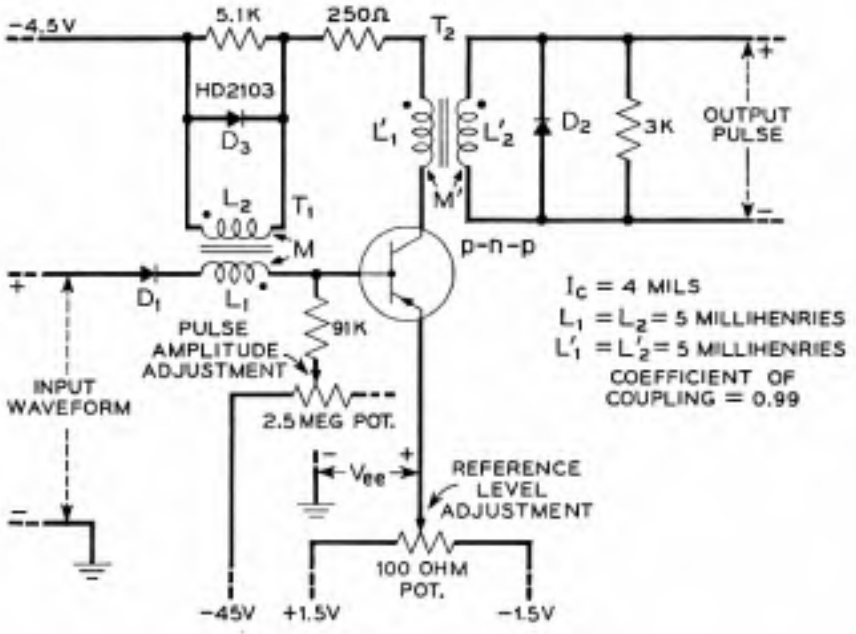


Fig. 21 — Voltage comparator.

increases the switching time. The circuit shown in Figure 21 does not misfire when used with voltage waveforms having slopes as small as 25 millivolts per microsecond, at the reference level.

6.0. A TRANSISTOR VOLTAGE ENCODER

6.1. *Circuit Arrangement*

The transistor circuits previously described can be assembled into a voltage encoder for translating analog voltages into equivalent time intervals. This encoder is especially useful for converting analog information (in the form of a dc potential) into the digital code for processing in a digital system. Fig. 22 shows a simplified block diagram of the encoder. The voltage ramp generated by the integrator is applied to amplitude selector number one and to one input of a summing amplifier. The amplitude selector is a dc amplifier which amplifies the voltage ramp in the vicinity of zero volts. Voltage comparator number one, which follows the amplitude selector, generates a sharp output pulse at the exact instant of time that the voltage ramp passes through zero volts.

The analog input voltage, which has a value between 0 and −15 volts,* is applied to the second input of the summing amplifier. The output voltage of the summing amplifier is zero whenever the ramp

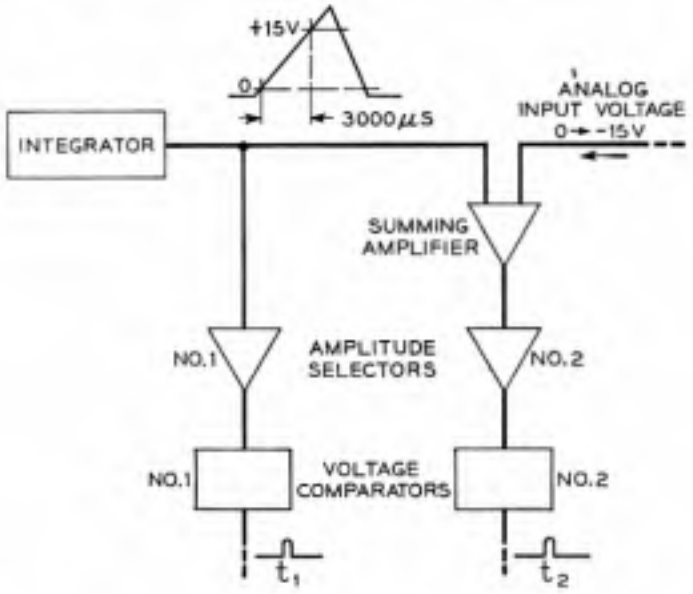


Fig. 22 — Simplified block diagram of voltage encoder.

* If the analog input voltage does not lie in this range, then the voltage gain of the summing amplifier must be set so that the analog voltage at the output of the summing amplifier lies in the voltage range between 0 and +15 volts.

voltage is equal to the negative of the input analog voltage. At this instant of time the second voltage comparator generates a sharp output pulse. The time interval between the two output pulses is proportional to the analog input voltage if the voltage ramp is linear and has a constant slope at all times.

6.2. *The Amplitude Selector*

The amplitude selector increases the slope of the input voltage waveform (in the vicinity of zero volts) sufficiently for proper operation of the voltage comparator. The amplitude selector consists of a limiter and a dc feedback amplifier as shown in Fig. 23. The two oppositely poled silicon diodes D_1 and D_2, limit the input voltage of the dc amplifier to about ± 0.65 volts. The dc amplifier has a voltage gain of thirty, and so the maximum output voltage of the amplitude selector is limited to about ± 19.5 volts. The net voltage gain between the input and output of the amplitude selector is ten.

The principal requirement placed on the dc amplifier is that the input current and the output voltage be zero when the input voltage is zero. This is accomplished by placing a blocking condenser C_B between point T and the base of the first transistor stage, and by using an AZS circuit to maintain point T at zero volts. The dc and AZS amplifiers are identical in configuration to the amplifiers shown in Fig. 12. The dc amplifier is

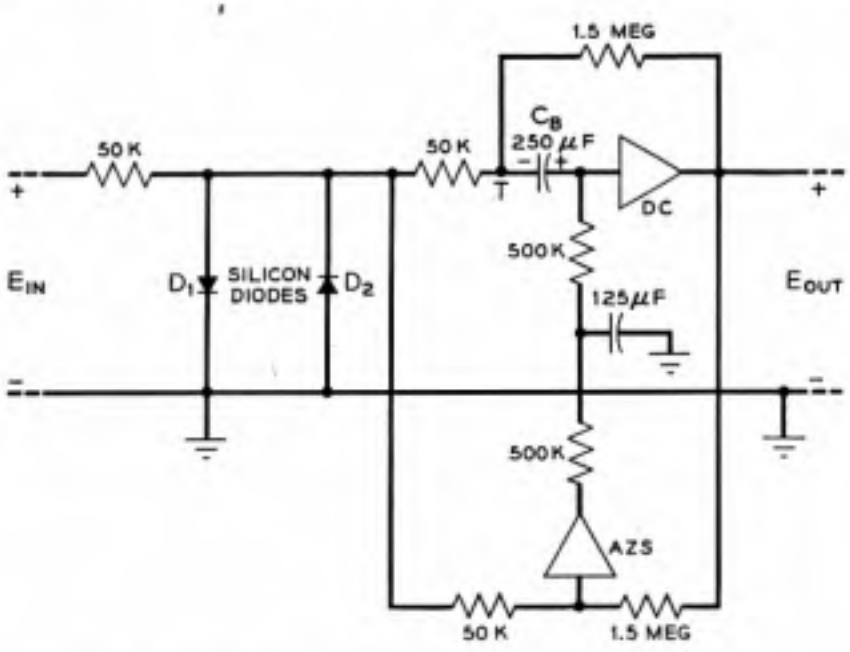


Fig. 23 — Block diagram of the amplitude selector.

designed to have about 15.6 db less feedback than that shown in Fig. 10 since this amount is adequate for the present purpose.

The bandwidth of the dc amplifier is only of secondary importance because the phase shifts introduced by the two amplitude selectors in the voltage encoder tend to compensate each other.

6.3. *Experimental Results*

The accuracy of the voltage encoder is determined by applying a precisely measured voltage to the input of the summing amplifier and by measuring the time interval between the two output pulses. The maximum error due to nonlinearities in the summing amplifier and the voltage ramp is less than ± 0.5 microseconds for a maximum encoding time of 3,000 microseconds. An additional error is introduced by the noise voltage generated in the first transistor stage of the summing amplifier. The RMS noise voltage at the output of the summing amplifier is less than 0.5 millivolts. This noise voltage produces an RMS jitter of 0.25 microseconds in the position of the second voltage comparator output pulse. The over-all accuracy of the voltage encoder is one part in 4,000 throughout a temperature range of 20°C to 40°C.

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to T. R. Finch for the advice and encouragement received in the course of this work. D. W. Grant and W. B. Harris designed and constructed the magnettor used in the AZS circuit of the integrator.

Appendix I

RELATIONSHIP BETWEEN RETURN DIFFERENCE AND LOOP CURRENT TRANSMISSION

In order to place the stability analysis of the transistor feedback amplifier on a sound basis, it is desirable to use the concept of return difference.[8] It will be shown that a measurable quantity, called the loop current transmission, can be related to the return difference of aZ_e with reference r_e.*, † In Fig. 24, N represents the complete transistor network exclusive of the transistor under consideration. The feedback loop is broken at the input to the transistor by connecting all of the feedback paths to

* In this appendix it is assumed that the transistor under consideration is in the common emitter connection. The discussion can be readily extended to the other transistor connections.

† This fact was pointed out by F. H. Tendick, Jr.

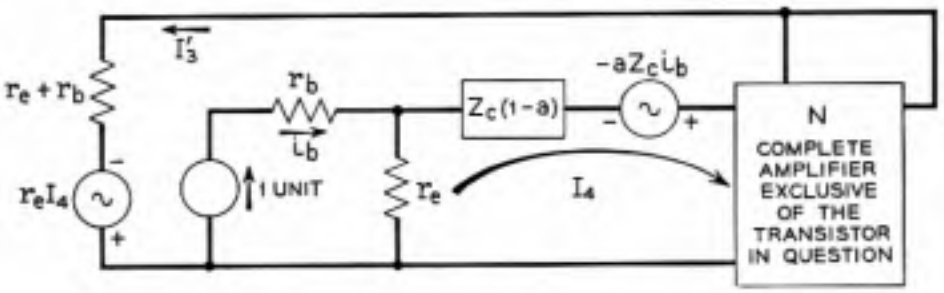


Fig. 24 — Measurement of loop current transmission.

ground through a resistance $(r_e + r_b)$ and a voltage $r_e I_4$. Using the nomenclature given in Reference 8, the input of the complete circuit is designated as the first mesh and the output of the complete circuit is designated as the second mesh. The input and output meshes of the transistor under consideration are designated 3 and 4, respectively. The loop current transmission is equal to I_3', the total returned current when a unit input current is applied to the base of the transistor.

The return difference for reference r_e is equal to the algebraic difference* between the unit input current and the returned current I_3'. I_3' is evaluated by multiplying the open circuit voltage in mesh 4 (produced by the unit base current) by the backward transmission from mesh 4 to mesh 3 with zero forward transmission through the transistor under consideration. The open circuit voltage in mesh 4 is equal to $(r_e - aZ_c)$. The backward transmission is determined with the element aZ_c, in the fourth row, third column of the circuit determinant, set equal to r_e. Hence, the return difference is expressed as

$$F_{r'_e} = 1 + (aZ_c - r_e)\frac{\Delta_{43}}{\Delta^{r'_e}} \qquad \text{(A1)}\dagger$$

$$F_{r'_e} = \frac{\Delta^{r'_e} + (aZ_c - r_e)\Delta_{43}}{\Delta^{r'_e}} \qquad \text{(A2)}$$

$$F_{r'_e} = \frac{\Delta}{\Delta^{r'_e}} = 1 + T_{r'_e} \qquad \text{(A3)}$$

The relative return ratio $T_{r'_e}$ is equal to the negative of the loop current transmission and can be measured as shown in Fig. 24. The voltage $r_e I_4$ takes into account the fact that the junction transistor is not perfectly

* The positive direction for the returned current is chosen so that if the original circuit is restored, the returned current flows in the same direction as the input current.

† $\Delta^{r'_e}$ is the network determinant with the element aZ_c in the fourth row, third column of the circuit determinant set equal to r_e.

unilateral. Fortunately, in many applications, this voltage can be neglected even at the gain and phase crossover frequencies.

In the case of single loop feedback amplifiers, $\Delta^{r'_e}$ will not have any zeros in the right hand half of the complex frequency plane. A study of the stability of the amplifier can then be based on $F_{r'_e}$ or $T_{r'_e}$.

Appendix II

Interstage Network Shaping

This appendix presents the analysis of the circuit shown in Fig. 7(a). The input impedance of the common emitter connected junction transistor is given by the expression

$$Z_{\text{INPUT}} = r_b + r_e(1 - G_I) \tag{A4}$$

where G_I is the current transmission of the common emitter stage, expression (13). The current transmission A of the complete circuit is equal to

$$A = \frac{I_2}{I_1} = \frac{Z_3}{Z_3 + Z_{\text{INPUT}}} \cdot G_I \tag{A5}$$

where $Z_3 = R_3 + pL_3 + (1/pC_3)$. Combining (13), (A4), and (A5) yields

$$A = \frac{-\dfrac{a_0}{1 - a_0 + \delta}\left[\left(1 + \dfrac{p}{\omega_3}\right)^2 + p\left(R_3C_3 - \dfrac{2}{\omega_3}\right)\right]}{\left(1 + \dfrac{p}{\omega_5}\right)\left\{1 + p\,\dfrac{C_3\omega_5}{\omega_1}(r_b + r_e + R_3 + \omega_1 L_3) + p^2\left[\dfrac{\omega_5}{\omega_1\omega_3^2} + \dfrac{C_3\omega_5(r_b + r_e + R_3)}{\omega_a\omega_c(1 - a_0 + \delta)}\right] + \dfrac{p^3\omega_5}{\omega_3^2\omega_a\omega_c(1 - a_0 + \delta)}\right\}} \tag{A6}$$

where

$$\delta = \frac{R_L + r_e}{r_c}$$

$$\omega_1 = \frac{(1 - a_0 + \delta)}{\dfrac{1 + \delta}{\omega_a} + \dfrac{1}{\omega_c}}$$

$$\omega_c = \frac{1}{(R_L + r_e)C_c}$$

$$\omega_3 = \frac{1}{\sqrt{L_3C_3}}$$

$$\omega_5 = \frac{1}{\left[r_b + \dfrac{r_e}{(1 - a_0 + \delta)}\right]C_3}$$

Expression (A6) is valid if $1/\omega_3 \gg 1/\omega_1 + R_3C_3$. The denominator of the expression indicates a falling 6 db per octave asymptote with a corner frequency at ω_3. The second factor in the denominator can be approximated by a falling 6 db per octave asymptote with a corner frequency at

$$\frac{\omega_1\left[r_b + \dfrac{r_e}{(1 - a_0 + \delta)}\right]}{r_b + r_e + R_3 + \omega_1 L_3}$$

plus additional phase and amplitude contributions at higher frequencies due to the p^2 and p^3 terms. If

$$\frac{1}{\omega_3 C_3 R_3} = \tfrac{1}{2}$$

then the circuit has a rising 12 db per octave asymptote with a corner frequency at ω_3. Fig. 7(b) shows the amplitude and phase of the current transmission.

REFERENCES

1. Felker, J. H., Regenerative Amplifier for Digital Computer Applications, Proc. I.R.E., pp. 1584–1596, Nov., 1952.
2. Korn, G. A., and Korn, T. M., Electronic Analog Computers, McGraw-Hill Book Company, pp. 9–19.
3. Wallace, R. L. and Pietenpol, W. J., Some Circuit Properties and Applications of n-p-n Transistors, B. S.T.J., **30**, pp. 530–563, July, 1951.
4. Shockley, W., Sparks, M. and Teal, G. K., The p-n Junction Transistor, Physical Review, **83**, pp. 151–162, July, 1951.
5. Pritchard, R. L., Frequency Variation of Current-Amplification for Junction Transistors, Proc. I.R.E., pp. 1476–1481, Nov., 1952.
6. Early, J. M., Design Theory of Junction Transistors, B.S.T.J., **32**, pp. 1271–1312, Nov., 1953.
7. Sziklai, G. C., Symmetrical Properties of Transistors and Their Applications, Proc. I.R.E., pp. 717–724, June, 1953.
8. Bode, H. W., Network Analysis and Feedback Amplifier Design, Van Nostrand Co., Inc., Chapter IV.
9. Bode, H. W., Op. Cit., pp. 66–69.
10. Bode, H. W., Op. Cit., pp. 162–164.
11. Bargellini, P. M. and Herscher, M. B., Investigation of Noise in Audio Frequency Amplifiers Using Junction Transistors, Proc. I.R.E., pp. 217–226, Feb., 1955.
12. Bode, H. W., Op. Cit., pp. 464–468, and pp. 471–473.
13. Keonjian, E., Temperature Compensated DC Transistor Amplifier, Proc. I.R.E., pp. 661–671, April, 1954.
14. Kretzmer, E. R., An Amplitude Stabilized Transistor Oscillator, Proc. I.R.E., pp. 391–401, Feb., 1954.
15. Goldberg, E. A., Stabilization of Wide-Band Direct-Current Amplifiers for Zero and Gain, R.C.A. Review, June, 1950.
16. Ebers, J. J. and Moll, J. L., Large Signal Behavior of Junction Transistors, Proc. I.R.E., pp. 1761–1772, Dec., 1954.
17. Manley, J. M., Some General Properties of Magnetic Amplifiers, Proc. I.R.E., March, 1951.
18. M.I.T., Waveforms, Volume 19 of the Radiation Laboratories Series. McGraw-Hill Book Company, pp. 342–344.

Electrolytic Shaping of Germanium and Silicon

By A. UHLIR, JR.

(Manuscript received November 9, 1955)

Properties of electrolyte-semiconductor barriers are described, with emphasis on germanium. The use of these barriers in localizing electrolytic etching is discussed. Other localization techniques are mentioned. Electrolytes for etching germanium and silicon are given.

INTRODUCTION

Mechanical shaping techniques, such as abrasive cutting, leave the surface of a semiconductor in a damaged condition which adversely affects the electrical properties of p-n junctions in or near the damaged material. Such damaged material may be removed by electrolytic etching. Alternatively, all of the shaping may be done electrolytically, so that no damaged material is produced. Electrolytic shaping is particularly well suited to making devices with small dimensions.

A discussion of electrolytic etching can conveniently be divided into two topics — the choice of electrolyte and the method of localizing the etching action to produce a desired shape. It is usually possible to find an electrolyte in which the rate at which material is removed is accurately proportional to the current. For semiconductors, just as for metals, the choice of electrolyte is a specific problem for each material; satisfactory electrolytes for germanium and silicon will be described.

The principles of localization are the same, whatever the electrolyte used. Electrolytic etching takes place where current flows from the semiconductor to the electrolyte. Current flow may be concentrated at certain areas of the semiconductor-electrolyte interface by controlling the flow of current in the electrolyte or in the semiconductor.

LOCALIZATION IN ELECTROLYTE

Localization techniques involving the electrolytic current are applicable to both metals and semiconductors. In some of these techniques,

the localization is so effective that the barrier effects found with n-type semiconductors can be ignored; if not, the barrier can be overcome by light or heat, as will be described below.

If part of the work is coated with an insulating varnish, electrolytic etching will take place only on the uncoated surfaces. This technique, often called "masking," has the limitation that the etching undercuts the masking if any considerable amount of material is removed. The same limitation applies to photoengraving, in which the insulating coating is formed by the action of light.

The cathode of the electrolytic cell may be limited in size and placed close to the work (which is the anode). Then the etching rate will be greatest at parts of the work that are nearest the cathode. Various shapes can be produced by moving the cathode with respect to the work, or by using a shaped cathode. For example, a cathode in the form of a wire has been used to slice germanium.[1]

Instead of a true metallic cathode, a "virtual cathode" may be used to localize electrolysis.[2] In this technique, the anode and true cathode are separated from each other by a nonconducting partition, except for a small opening in the partition. As far as localization of current to the anode is concerned, the small opening acts like a cathode of equal size and so is called a virtual cathode. The nonconducting partition may include a glass tube drawn down to a tip as small as one micron diameter but nevertheless open to the flow of electrolytic current. With such a tip as a virtual cathode, micromachining can be conducted on a scale comparable to the wavelength of visible light. A general advantage of the virtual cathode technique is that the cathode reaction (usually hydrogen evolution) does not interfere with the localizing action nor with observation of the process.

In the jet-etching technique, a jet of electrolyte impinges on the work.[3, 4] The free streamlines that bound the flowing electrolyte are governed primarily by momentum and energy considerations. In turn, the shape of the electrolyte stream determines the localization of etching. A stream of electrolyte guided by wires has been used to etch semiconductor devices.[5] Surface tension has an important influence on the free streamlines in this case.

PROPERTIES OF ELECTROLYTE-SEMICONDUCTOR BARRIERS

The most distinctive feature of electrolytic etching of semiconductors is the occurrence of rectifying barriers. Barrier effects for germanium will be described; those for silicon are qualitatively similar.

The voltage-current curves for anodic n-type and p-type germanium

in 10 per cent KOH are shown in Fig. 1. The concentration of KOH is not critical and other electrolytes give similar results. The voltage drop for the p-type specimen is small. For anodic n-type germanium, however, the barrier is in the reverse or blocking direction as evidenced by a large voltage drop. The fact that n-type germanium differs from p-type germanium only by very small amounts of impurities suggests that the barrier is a semiconductor phenomenon and not an electrochemical one. This is confirmed by the light sensitivity of the n-type voltage-current characteristic. Fig. 2 is a schematic diagram of the arrangement for obtaining voltage-current curves. A mercury-mercuric oxide-10 per cent KOH reference electrode was used at first, but a gold wire was found equally satisfactory. At zero current, a voltage V_0 exists between the germanium and the reference electrode; this voltage is not included in Fig. 1.

The saturation current I_s, measured for the n-type barrier at a *moderate* reverse voltage (see Fig. 1), is plotted as a function of temperature in Fig. 3. The saturation current increases about 9 per cent per degree, just as for a germanium p-n junction, which indicates that the

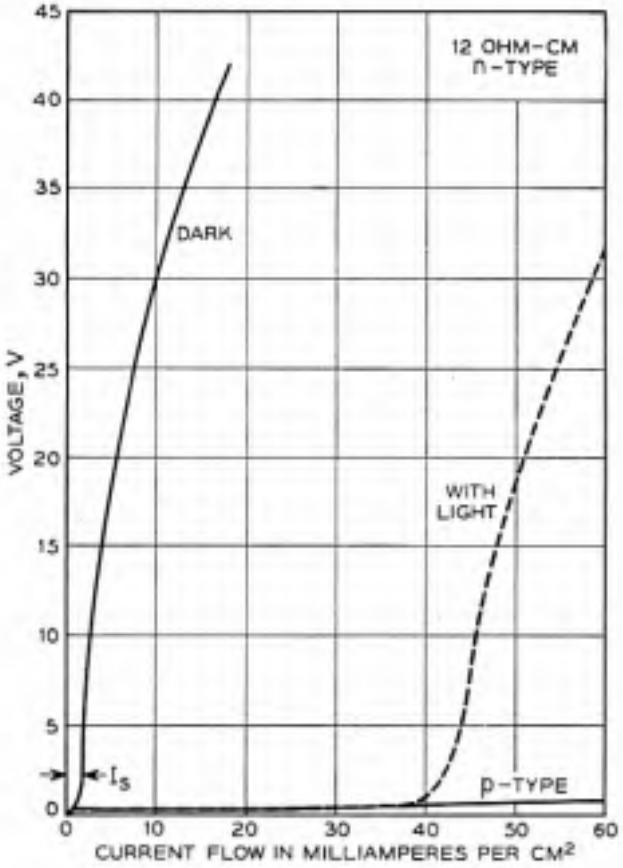


Fig. 1 — Anodic voltage-current characteristics of germanium.

current is proportional to the equilibrium density of minority carriers (holes). The same conclusion may be drawn from Fig. 4, which shows that the saturation current is higher, the higher the resistivity of the n-type germanium. But the breakdown voltages are variable and usually much lower than one would expect for planar p-n junctions made, for example, by alloying indium into the same n-type germanium.

Breakdown in bulk junctions is attributed to an avalanche multiplication of carriers in high fields.[6] The same mechanism may be responsible for breakdown of the germanium-electrolyte barrier; low and variable breakdown voltages may be caused by the pits described below.

The electrolyte-germanium barrier exhibits a kind of current multiplication that differs from high-field multiplication in two respects: it occurs at much lower reverse voltages and does not vary much with voltage.[7] This effect can be demonstrated very simply by comparison with a metal-germanium barrier, on the assumption that the latter has a current multiplication factor of unity. This assumption is supported by experiments which indicate that current flows almost entirely by hole flow, for good metal-germanium barriers.[8]

The experimental arrangement is indicated in Fig. 5(a) and (b). The voltage-current curves for an electrolyte barrier and a plated barrier on the same slice of germanium are shown in Fig. 5(c).* The curves for the

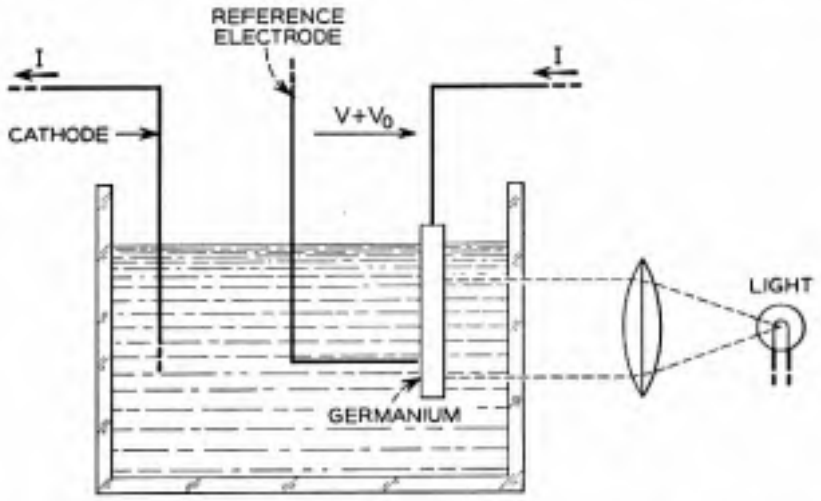


Fig. 2 — Arrangement for obtaining voltage-current characteristics.

* In Fig. 5 the dark current for the plated barrier is much larger than can be explained on the basis of hole current; it is even higher than the dark current for the electrolyte barrier, which should be at least 1.4 times the hole current. This excess dark current is believed to be leakage at the edges of the plated area and probably does not affect the intrinsic current multiplication of the plated barrier as a whole.

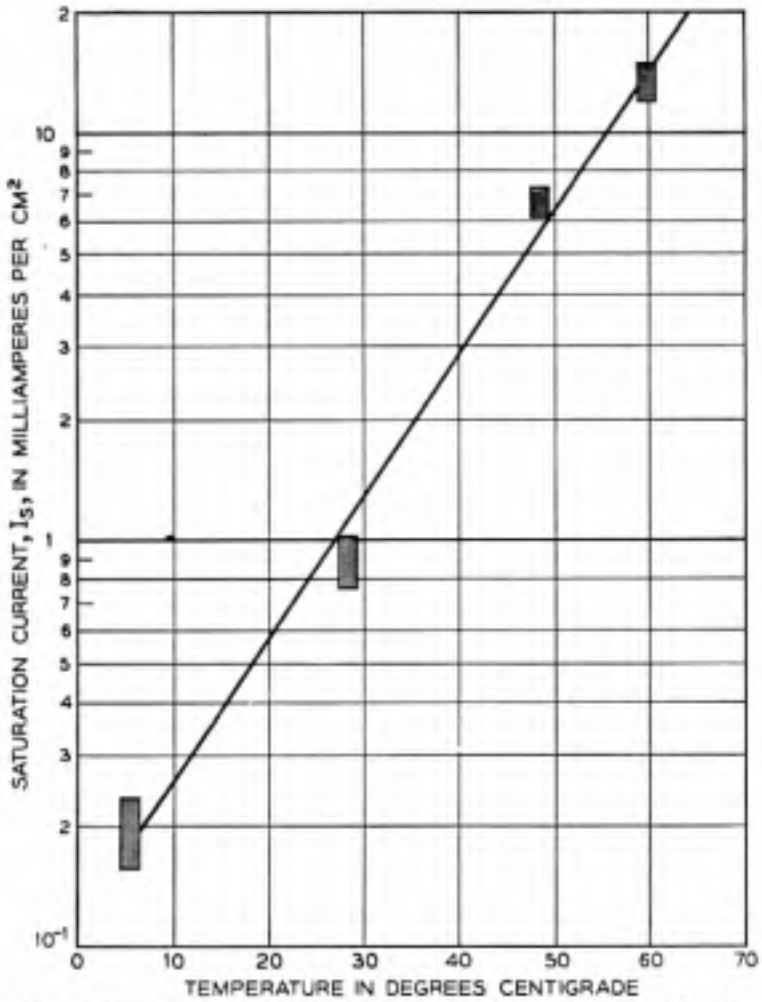


Fig. 3 — Temperature variation of the saturation current of a barrier between 5.5 ohm-cm n-type germanium and 10 per cent KOH solution.

illuminated condition were obtained by shining light on a dry face of a slice while the barriers were on the other face. The difference between the light and dark currents is larger for the electrolyte-germanium barrier than for the metal-germanium barrier, by a factor of about 1.4.

The transport of holes through the slice is probably not very different for the two barriers. Therefore, a current multiplication of 1.4 is indicated for the electrolyte barrier. About the same value was found for temperatures from 15°C to 60°C, KOH concentrations from 0.01 per cent to 10 per cent, n-type resistivities of 0.2 ohm-cm to 6 ohm-cm, light currents of 0.1 to 1.0 ma/cm^2, and for 0.1N indium sulfate.

Evidently the flow of holes to the electrolyte barrier is accompanied by a proportionate return flow of electrons, which constitutes an additional electric current. Possible mechanisms for the creation of the electrons will be discussed in a forthcoming article.[9]

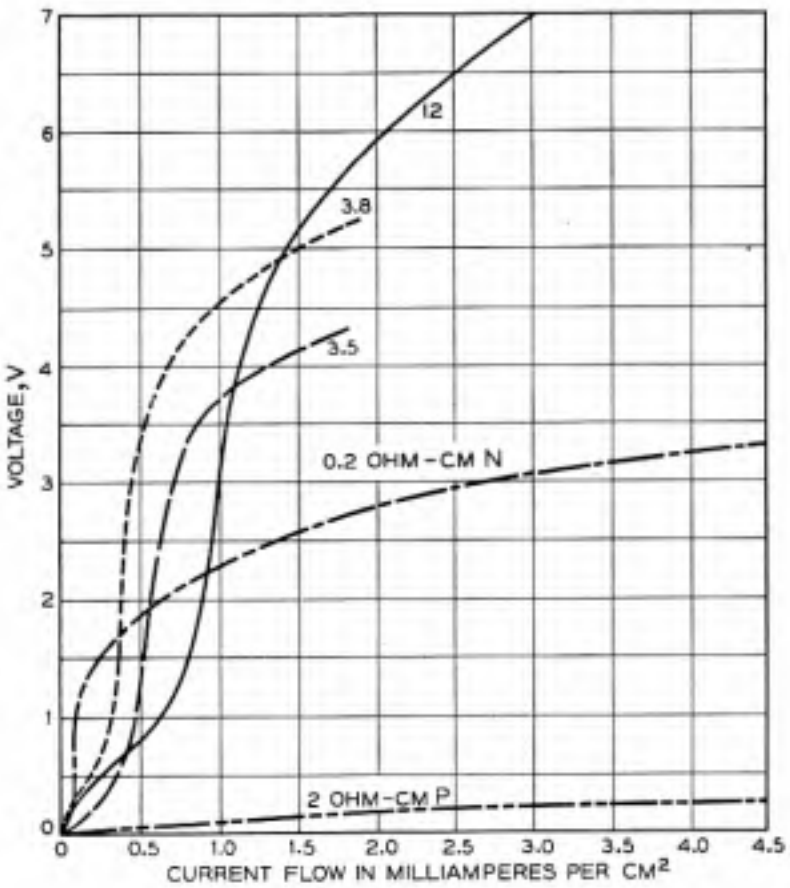


Fig. 4. — Anodic voltage-current curves for various resistivities of germanium.

SCRATCHES AND PITTING

The voltage-current curve of an electrolyte-germanium barrier is very sensitive to scratches. The curves given in the illustrations were obtained on material previously etched smooth in CP-4, a chemical etch.* [10]

If, instead, one starts with a lapped piece of n-type germanium, the electrolyte-germanium barrier is essentially "ohmic;" that is, the voltage drop is small and proportional to the current. A considerable reverse voltage can be attained if lapped n-type germanium is electrolytically etched long enough to remove most of the damaged germanium. However, a pitted surface results and the breakdown voltage achieved is not as high as for a smooth chemically-etched surface.

The depth of damage introduced by typical abrasive sawing and lapping was investigated by noting the voltage-current curve of the

* Five parts HNO_3, 3 parts 48 per cent HF, 3 parts glacial acetic acid, $\frac{1}{10}$ part Br_2.

electrolyte-germanium barrier after various amounts of material had been removed by *chemical* etching. After 20 to 50 microns had been removed, further chemical etching produced no change in the barrier characteristic. This amount of material had to be removed even if the lapping was followed by polishing to a mirror finish. The voltage-current curve of the electrolyte-germanium barrier will reveal localized damage. On the other hand, the photomagnetoelectric (PME) measurement of

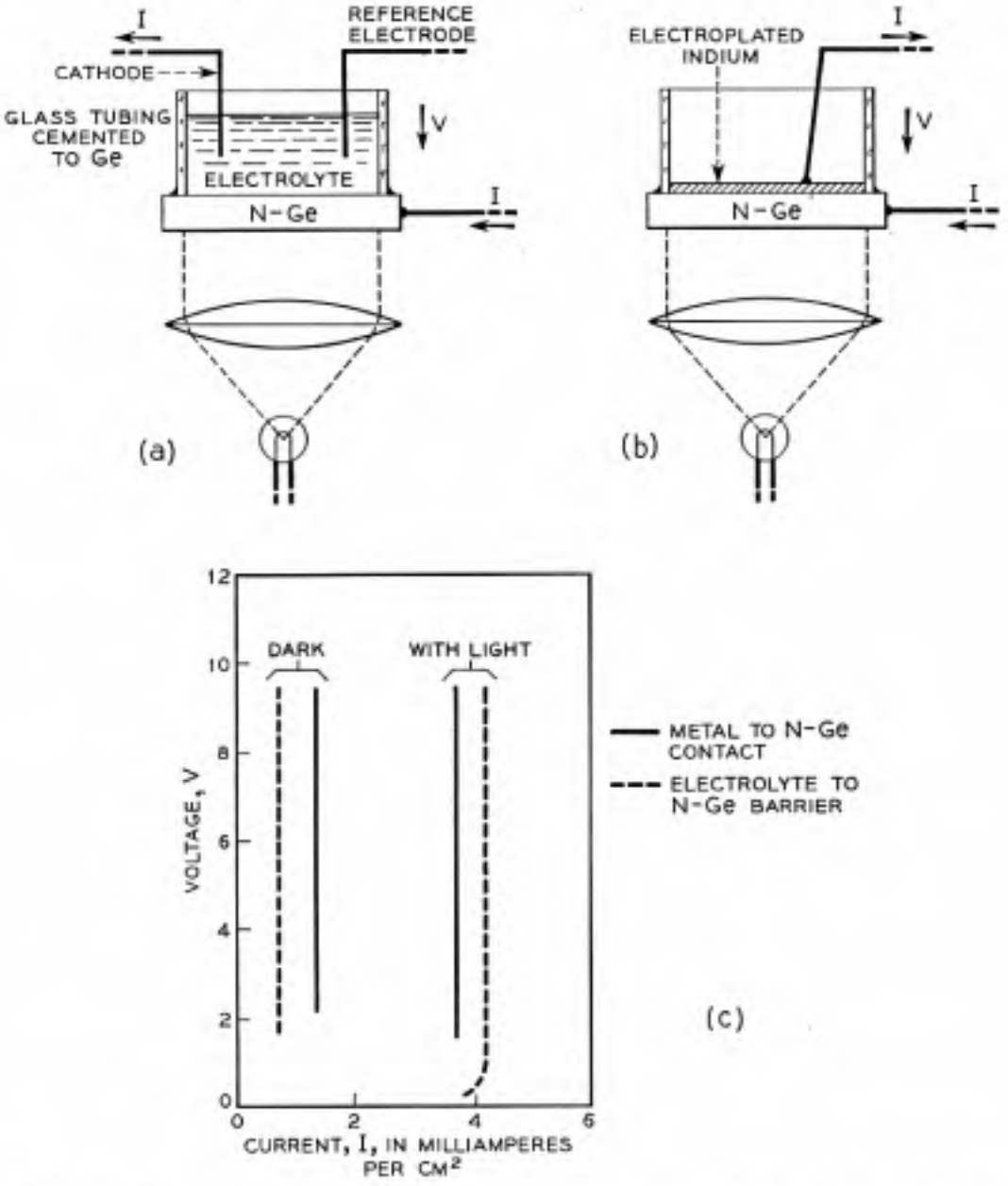


Fig. 5 — Determination of the current multiplication of the barrier between 6 ohm-cm n-type germanium and an electrolyte.

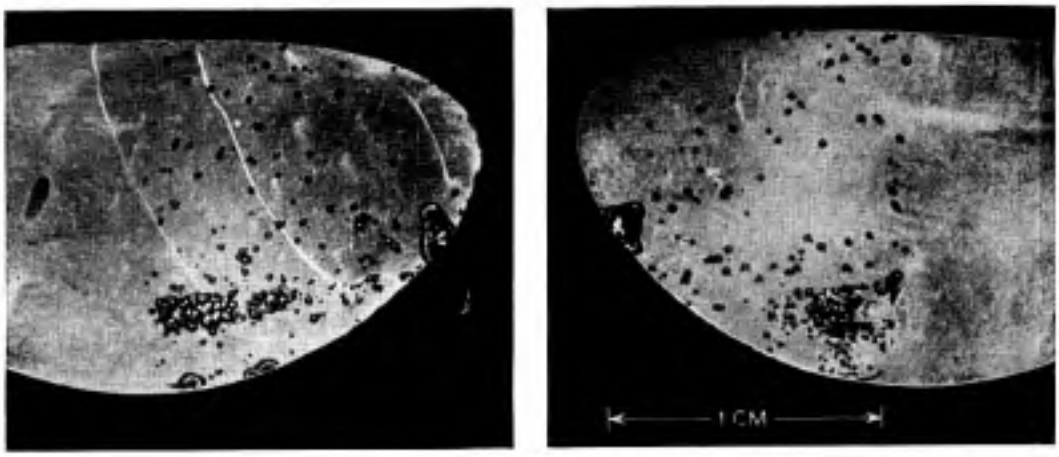


Fig. 6 — Electrolytic etch pits on two sides of 0.02-inch slice of n-type germanium. Half of the slice was in contact with the electrolyte.

surface recombination velocity gives an evaluation of the average condition of the surface.[11] A variation of the PME method has been used to study the depth of abrasion damage; the damage revealed by this method extends only to a depth comparable to the abrasive size.[12]

A scratch is *sufficient* to start a pit that increases in size without limit if anodic etching is prolonged. However, a scratch is not *necessary*. Pits are formed even when one starts with a smooth surface produced by chemical etching. A drop in the breakdown voltage of the barrier is noticed when one or more pits form. The breakdown voltage can be restored by masking the pits with polystyrene cement.

Evidence that the spontaneous pits are caused by some features of the crystal, itself, was obtained from an experiment on single-crystal n-type germanium made by an early version of the zone-leveling process. A slice of this material was electrolytically etched on both sides, after preliminary chemical etching. Photographs of the two sides of the slice are shown in Fig. 6. Only half of the slice was immersed in the electrolyte. The electrolytic etch pits are concentrated in certain regions of the slice — the same general regions on both sides of the slice. It is interesting that radioautographs and resistivity measurements indicate high donor concentrations in these regions. Improvements, including more intensive stirring, were made in the zone-leveling process, and the electrolytic etch pit distribution and the donor radioautographs have been much more uniform for subsequent material.

Several pits on a (100) face are shown in Fig. 7. The pits grow most rapidly in ⟨100⟩ directions and give the spiked effect seen in the illustration. After prolonged etching, the spikes and their branches form a complex network of caverns beneath the surface of the germanium.

High-field carrier generation may be responsible for pitting. A locally

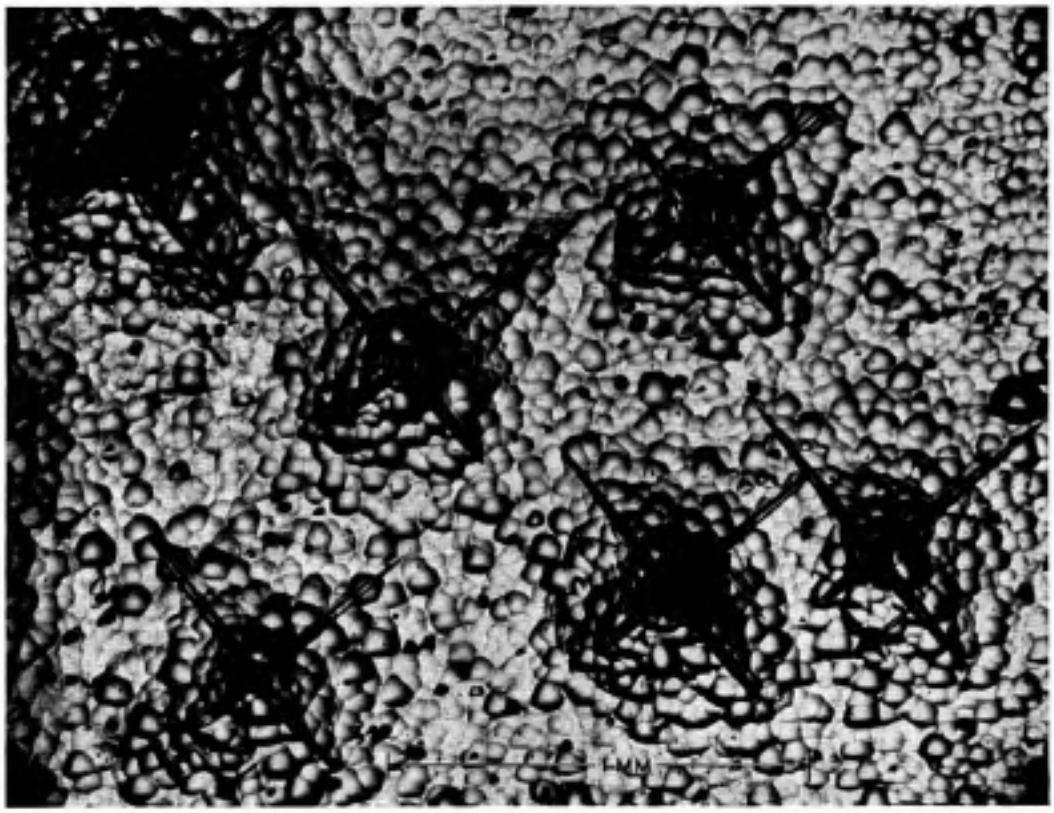


Fig. 7 — Electrolytic etch pits on n-type germanium.

high donor concentration would favor breakdown, as would any concavity of the germanium surface (which would cause a higher field for a given voltage). Very high fields must occur at the points of spikes such as those shown in Fig. 7. The continued growth of the spikes is thus favored by their geometry.

Microscopic etch pits arising from *chemical* etching have been correlated with the edge dislocations of small-angle grain boundaries.[13] A specimen of n-type germanium with chemical etch pits was photomicrographed and then etched electrolytically. The etch pits produced electrolytically could not be correlated with the chemical etch pits, most of which were still visible and essentially unchanged in appearance. Also, no correlation could be found between either kind of etch pit and the locations at which copper crystallites formed upon immersion in a copper sulfate solution. Microscopic electrolytic etch pits at dislocations in p-type germanium have been reported in a recent paper that also mentions the deep pits produced on n-type germanium.[14]

Electrolytic etch pits are observed on n-type and high-resistivity silicon. These etch pits are more nearly round than those produced in germanium.

In spite of the pitting phenomenon, electrolytic etching is success-

fully used in the fabrication of devices involving n-type semiconductors. Pitting can be reduced relative to "normal" uniform etching by any agency that increases the concentration of holes in the semiconductor. Thus, elevated temperatures, flooding with light, and injection of holes by an emitter all favor smooth etching.

SHAPING BY MEANS OF INJECTED CARRIERS

Hole-electron pairs are produced when light is absorbed by semiconductors. Light of short wavelength is absorbed in a short distance, while long wavelength light causes generation at considerable depths. The holes created by the light move by diffusion and drift and increase the current flow through an anodic electrolyte-germanium barrier at whatever point they happen to encounter the barrier. In general, more holes will diffuse to a barrier, the nearer the barrier is to the point at which the holes are created. For n-type semiconductors, the current due to the light can be orders of magnitude greater than the dark current, so that the shape resulting from etching is almost entirely determined by the light. As shown in Fig. 3, the dark current can be made very small by lowering the temperature.

An example of the shaping that can be done with light is shown in Fig. 8. A spot of light impinges on *one* side of a wafer of n-type germanium or silicon. The semiconductor is made anodic with respect to an etching electrolyte. Accurately concentric dimples are produced on *both* sides of the wafer. Two mechanisms operate to transmit the effect to the opposite side. One is that some of the light may penetrate deeply before generating a hole-electron pair. The other is that a fraction of the carriers generated near the first surface will diffuse to the opposite side. By varying the spectral content of the light and the depth within the

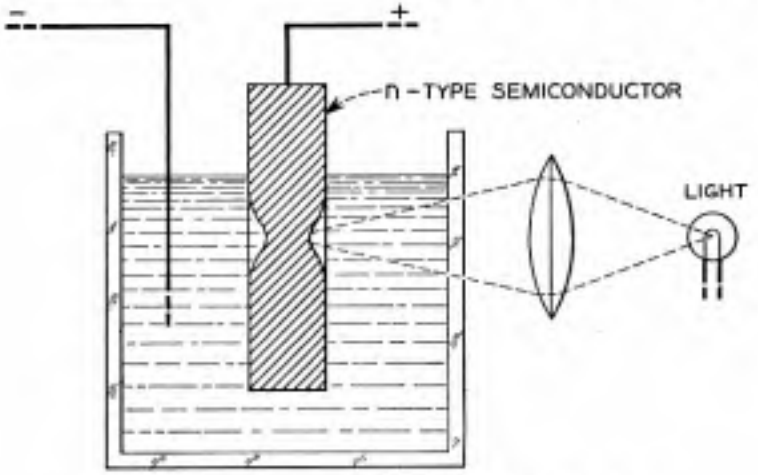


Fig. 8 — Double dimpling with light.

wafer at which the light is focused, one can produce dimples with a variety of shapes and relative sizes.

It is obvious that the double-dimpled wafer of Fig. 8 is desirable for the production of p-n-p alloy transistors. For such use, one of the most important dimensions is the thickness remaining between the bottoms of the two dimples. As has been mentioned in connection with the jet-etching process, a convenient way of monitoring this thickness to determine the endpoint of etching is to note the transmission of light of suitable wavelength.[15] There is, however, a control method that is itself automatic. It is based on the fact that at a reverse-biased p-n junction or electrolyte-semiconductor barrier there is a space-charge region that is practically free of carriers.[4] When the specimen thickness is reduced so that space-charge regions extend clear through it, current ceases to flow and etching stops in the thin regions, as long as thermally or optically generated carriers can be neglected. However, more pitting is to be expected in this method than when etching is conducted in the presence of an excess of injected carriers.

A p-n junction is a means of injecting holes into n-type semiconductors and is the basis of another method of dimpling, shown in Fig. 9. The p-n junction can be made by an alloying process such as bonding an acceptor-doped gold wire to germanium. The ohmic contact can be made by bonding a donor-doped gold wire and permits the injection of a greater excess of holes than would be possible if the current through the p-n junction were exactly equal to the etching current. Dimpling without the ohmic contact has been reported.[14]

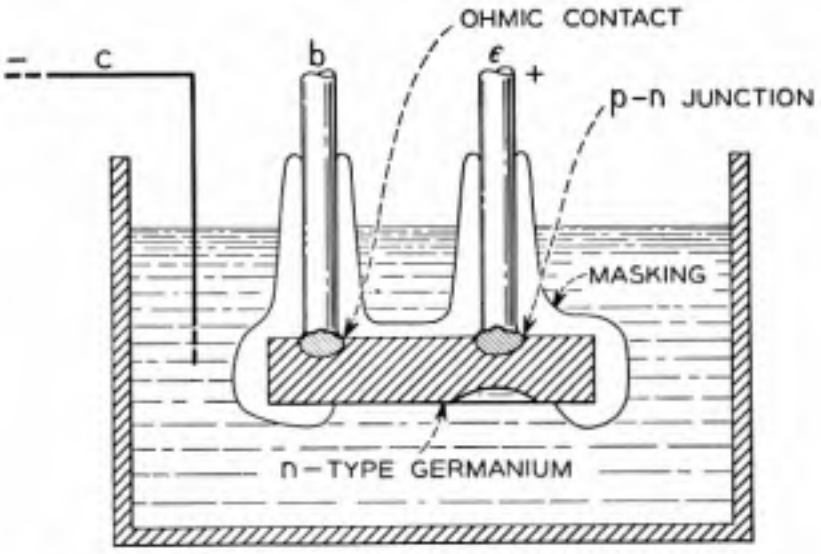


Fig. 9 — Dimpling with carriers injected by a p-n junction.

CONTROL BY OHMIC CONDUCTION

The carrier-injection shaping techniques work very well for n-type material. It is also possible to inject a significant number of holes into rather high resistivity p-type material. But what can be done about p-type material in general, short of developing cathodic etches?

The ohmic resistivity of p-type material can be used as shown in Fig. 10. More etching current flows through surfaces near the small contact than through more remote surfaces. A substantial dimpling effect is observed when the semiconductor resistivity is equal to the electrolyte resistivity, but improved dimpling is obtained on higher resistivity semiconductor. This result is just what one might expect. But the mathematical solution for ohmic flow from a point source some distance from a planar boundary between semi-infinite materials of different conductivities shows that the current density distribution does not depend on the conductivities. An important factor omitted in the mathematical solution is the small but significant barrier voltage, consisting largely of electrochemical polarization in the electrolyte. The barrier voltage is approximately proportional to the logarithm of the current density while the ohmic voltage drops are proportional to current density. Thus, high current favors localization.

ELECTROLYTES FOR ETCHING GERMANIUM AND SILICON

The electrolyte usually has two functions in the electrolytic etching of an oxidizable substance. First, it must conduct the current necessary for the oxidation. Second, it must somehow effect removal of the oxidation product from the surface of the material being etched.

The usefulness of an electrolytic etch depends upon one or both of

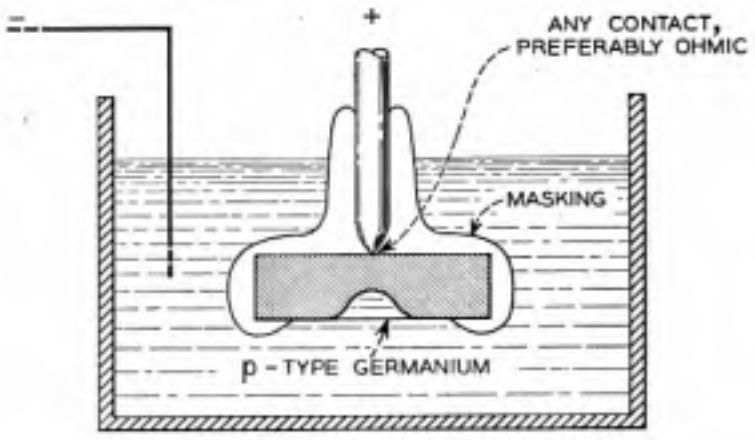


Fig. 10 — Dimpling by ohmic conduction.

the following situations — the electrolytic process accomplishes a reaction that cannot be achieved as conveniently in any other way or it permits greater control to be exercised over the reaction. Accordingly, chemical attack by the chosen electrolyte must be slight relative to the electrochemical etching.

A smooth surface is probably desirable in the neighborhood of a p-n junction, to avoid field concentrations and lowering of breakdown voltage. Therefore, a tentative requirement for an electrolyte is the production of a smooth, shiny surface on the p-type semiconductor. Such an electrolyte will give a shiny but possibly pitted surface on n-type specimens of the same semiconductor.

The effective valence of a material being electrolytically etched is defined as the number of electrons that traverse the circuit divided by the number of atoms of material removed. (The amount of material removed was determined by weighing in the experiments to be described.) If the effective valence turns out to be less than the valence one might predict from the chemistry of stable compounds, the etching is sometimes said to be "more than 100 per cent efficient." Since the anode reactions in electrolytic etching may involve unstable intermediate compounds and competing reactions, one need not be surprised at low or fractional effective valences.

Germanium can be etched in many aqueous electrolytes. A valence of almost exactly 4 is found.[16] That is, 4 electrons flow through the circuit for each atom of germanium removed. For accurate valence measurements, it is advisable to exclude oxygen by using a nitrogen atmosphere. Potassium hydroxide, indium sulfate, and sodium chloride solutions are among those that have been used. Sulfuric acid solutions are prone to yield an orange-red deposit which may be a suboxide of germanium.[16] Similar orange deposits are infrequently encountered with potassium hydroxide.

Hydrochloric acid solutions are satisfactory electrolytes. The reaction product is removed in an unusual manner when the electrolyte is about 2N hydrochloric acid. Small droplets of a clear liquid fall from the etched regions. These droplets may be germanium tetrachloride, which is denser than the electrolyte. They turn brown after a few seconds, perhaps because of hydrolysis of the tetrachloride.

Etching of germanium in sixteen different aqueous electroplating electrolytes has been mentioned.[8] Germanium can also be etched in the partly organic electrolytes described below for silicon.

One would expect that silicon could be etched by making it the anode in a cell with an aqueous hydrofluoric acid electrolyte. The seemingly

likely oxidation product, silicon dioxide, should react with the hydrofluoric acid to give silicon tetrafluoride, which could escape as a gas. In fact, a gas is formed at the anode and the silicon loses weight. But the gas is hydrogen and an effective valence of 2.0 ± 0.2 (individual determinations ranged from 1.3 to 2.7) was found instead of the value 4 that might have been expected. The quantity of hydrogen evolved is consistent with the formal reaction

$$\mathrm{Si} \rightarrow \mathrm{Si}^{+m} + me^{-} \qquad \text{(electrochemical oxidation)}$$

$$\mathrm{Si}^{+m} + (4\text{-}m)\mathrm{H}^{+} \rightarrow \mathrm{Si}^{+4} + \tfrac{1}{2}\,(4\text{-}m)\mathrm{H}_2 \qquad \text{(chemical oxidation)}$$

where m is about two. The experiments were done in 24 per cent to 48 per cent aqueous solutions of HF at current densities up to 0.5 amp/cm^2.

The suggestion that the electrochemical oxidation precedes the chemical oxidation is supported by the appearance and behavior of the etched surfaces. Instead of being shiny, the surfaces have a matte black, brown, or red deposit.

At 40× magnification, the deposit appears to consist of flakes of a resinous material, tentatively supposed to be a silicon suboxide. A remarkable reaction can be demonstrated if the silicon is rinsed briefly in water and alcohol after the electrolytic etch, dried, and stored in air for as long as a year. Upon reimmersing this silicon in water, one can observe the liberation of gas bubbles at its surface. This gas is presumed to be hydrogen. To initiate the reaction it is sometimes necessary to dip the specimen first in alcohol, as water may otherwise not wet it. The specimens also liberate hydrogen from alcohol and even from toluene.

Thus, chemical oxidation can follow electrolytic oxidation. But chemical oxidation does not proceed at a significant rate before the current is turned on.

Smooth, shiny electrolytic etching of p-type silicon has been obtained with mixtures of hydrofluoric acid and organic hydroxyl compounds, such as alcohols, glycols, and glycerine. These mixtures may be anhydrous or may contain as much as 90 per cent water. The organic additives tend to minimize the chemical oxidation of the silicon. They also permit etching at temperatures below the freezing point of aqueous solutions. They lower the conductivity of the electrolyte.

For a given electrolyte composition, there is a threshold current density, usually between 0.01 and 0.1 amps/cm^2, for smooth etching. Lower current densities give black or red surfaces with the same hydrogen-liberating capabilities as those obtained in aqueous hydrofluoric acid.

In general, smooth etching of silicon seems to result when the effective valence is nearly 4 and there is little anodic evolution of gas. The electrical properties of the smooth surface appear to be equivalent to those of smooth silicon surfaces produced by chemical etching in mixtures of nitric and hydrofluoric acids. On the other hand, the reactive surface produced at a valence of about 2, with anodic hydrogen evolution, is capable of practically shorting-out a silicon p-n junction. The electrical properties of this surface tend to change upon standing in air.

ACKNOWLEDGEMENTS

Most of the experiments mentioned in this paper were carried out by my wife, Ingeborg. An exception is the double-dimpling of germanium by light, which was done by T. C. Hall. The dimpling procedures of Figs. 9 and 10 are based on suggestions by J. M. Early. The effect of light upon electrolytic etching was called to my attention by O. Loosme. W. G. Pfann provided the germanium crystals grown with different degrees of stirring.

REFERENCES

1. J. F. Barry, I.R.E.-A.I.E.E. Semiconductor Device Research Conference, Philadelphia, June, 1955.
2. A. Uhlir, Jr., Rev. Sci. Inst., **26,** pp. 965–968, 1955.
3. W. E. Bailey, U. S. Patent No. 1,416, 929, May 23, 1922.
4. Bradley, *et al.* Proc. I.R.E., **24,** pp. 1702–1720, 1953.
5. M. V. Sullivan and J. H. Eigler, to be published.
6. S. L. Miller, Phys. Rev. **99,** p. 1234, 1955.
7. W. H. Brattain and C. G. B. Garrett, B.S.T.J., **34,** pp. 129–176, 1955.
8. E. H. Borneman, R. F. Schwarz, and J. J. Stickler, J. Appl. Phys., **26,** pp. 1021–1029, 1955.
9. D. R. Turner, to be submitted to the Journal of the Electrochemical Society.
10. R. D. Heidenreich, U. S. Patent No. 2,619,414, Nov. 25, 1952.
11. T. S. Moss, L. Pincherle, A. M. Woodward, Proc. Phys. Soc. London, **66B,** p. 743, 1953.
12. T. M. Buck and F. S. McKim, Cincinnati Meeting of the Electrochemical Society, May, 1955.
13. F. L. Vogel, W. G. Pfann, H. E. Corey, and E. E. Thomas, Phys. Rev., **90,** p. 489, 1953.
14. S. G. Ellis, Phys. Rev., **100,** pp. 1140–1141, 1955.
15. *Electronics*, **27,** No. 5, p. 194, May, 1954.
16. F. Jirsa, Z. f. Anorg. u. Allgemeine Chem., Bd. **268,** p. 84, 1952.

A Large Signal Theory of Traveling Wave Amplifiers

Including the Effects of Space Charge and Finite Coupling Between the Beam and the Circuit

By PING KING TIEN

Manuscript received October 11, 1955)

The non-linear behavior of the traveling-wave amplifier is calculated in this paper by numerically integrating the motion of the electrons in the presence of the circuit and the space charge fields. The calculation extends the earlier work by Nordsieck and the small-C theory by Tien, Walker and Wolontis, to include the space charge repulsion between the electrons and the effect of a finite coupling between the circuit and the electron beam. It however differs from Poulter's and Rowe's works in the methods of calculating the space charge and the effect of the backward wave.

The numerical work was done using 701-type I.B.M. equipment. Results of calculation covering QC from 0.1 to 0.4, b from 0.46 to 2.56 and k from 1.25 to 2.50, indicate that the saturation efficiency varies between 23 per cent and 37 per cent for C equal to 0.1 and between 33 per cent and 40 per cent for C equal to 0.15. The voltage and the phase of the circuit wave, the velocity spread of the electrons and the fundamental component of the charge-density modulation are either tabulated or presented in curves. A method of calculating the backward wave is provided and its effect fully discussed.

1. INTRODUCTION

Theoretical evaluation of the maximum efficiency attainable in a traveling-wave amplifier requires an understanding of the non-linear behavior of the device at various working conditions. The problem has been approached in many ways. Pierce,[1] and later Hess,[2] and Birdsall[3] and Caldwell[4] investigated the efficiency or the output power, using certain specific assumptions about the highly bunched electron beam. They either assume a beam in the form of short pulses of electrons, or, specify

an optimum ratio of the fundamental component of convection current to the average or d-c current. The method, although an abstract one, generally gives the right order of the magnitude. When the usual wave concept fails for a beam in which overtaking of the electrons arises, we may either overlook effects from overtaking, or, using the Boltzman's transport equation search for solutions in series form. This attack has been pursued by Parzen[5] and Kiel,[6] although their work is far from complete. The most satisfying approach to date is Nordsieck's analysis.[7] Nordsieck followed a typical set of "electrons" and calculated their velocities and positions by numerically integrating a set of equations of motion. Poulter[8] has extended Nordsieck equations to include space charge, finite C and circuit loss, although he has not perfectly taken into account the space charge and the backward wave. Recently Tien, Walker, and Wolontis[9] have published a small C theory in which "electrons" are considered in the form of uniformly charged discs and the space charge field is calculated by computing the force exerted on one disc by the others. Results extended to finite C, have been reported by Rowe,[10] and also by Tien and Walker.[11] Rowe, using a space charge expression similar to Poulter's, computed the space charge field based on the electron distribution in time instead of the distribution in space. This may lead to appreciable error in his space charge term, although its influence on the final results cannot be easily evaluated.

In the present analysis, we shall adopt the model described by Tien, Walker and Wolontis, but wish to add to it the effect of a finite beam to circuit coupling. A space charge expression is derived taking into account the fact that the a-c velocities of the electrons are no longer small compared with the average velocity. Equations are rewritten to retain terms involving C. As the backward wave becomes appreciable when C increases, a method of calculating the backward wave is provided and the effect of the backward wave is studied. Finally, results of the calculation covering useful ranges of design and operating parameters are presented and analyzed.

2. ASSUMPTIONS

To recapitulate, the major assumptions which we have made are:

1. The problem is considered to be one dimensional, in the sense that the transverse motions of the electrons are prohibited, and the current, velocity, and fields, are functions only of the distance along the tube and of the time.

2. Only the fundamental component of the current excites waves on the circuit.

3. The space charge field is computed from a model in which the helix is replaced by a conducting cylinder, and electrons are uniformly charged discs. The discs are infinitely thin, concentric with the helix and have a radius equal to the beam radius.

4. The circuit is lossfree.

These are just the assumptions of the Tien-Walker-Wolontis model. In addition, we shall assume a small signal applied at the input end of a long tube, where the beam entered unmodulated. What we are looking for are therefore the characteristics of the tube beyond the point at which the device begins to act non-linearly. Let us imagine a flow of electron discs. The motions of the discs are computed from the circuit and the space charge fields by the familiar Newton's force equation. The electrons, in turn, excite waves on the circuit according to the circuit equation[12] derived either from Brillouin's model[13] or from Pierce's equivalent circuit.[14] The force equation, the circuit equation, and the equation of conservation of charge in kinematics,[15] are the three basic equations from which the theory is derived.

3. FORWARD AND BACKWARD WAVES

In the traveling-wave amplifier, the beam excites forward and backward waves on the circuit. (We mean by "forward" wave, the wave which propagates in the direction of the electron flow, and by "backward" wave, the wave which propagates in the opposite direction.) Because of phase cancellation, the energy associated with the backward wave is small, but increases with the beam to circuit coupling. It is therefore important to compute it accurately. In the first place, the waves on the circuit must satisfy the circuit equation[12]

$$\frac{\partial^2 V(z,\,t)}{\partial t^2} - v_0^{\,2}\,\frac{\partial^2 V(z,\,t)}{\partial z^2} = v_0 Z_0\,\frac{\partial^2 \rho_\omega(z,\,t)}{\partial t^2} \tag{1}$$

Here, V is the total voltage of the waves. v_0 and Z_0 are respectively the phase velocity and the impedance of the cold circuit. z is the distance along the tube and t, the time. ρ_ω is the fundamental component of the linear charge density. V and ρ_ω are functions of z and t. The complete solution of (1) is in the form

$$\begin{aligned} V(z) &= C_1 e^{-\Gamma_0 z} + C_2 e^{\Gamma_0 z} \\ &+ e^{-\Gamma_0 z}\,\frac{\Gamma_0 v_0 Z_0}{2}\int_{z'}^{z} e^{\Gamma_0 z}\rho_\omega(z)\,dz \\ &+ e^{\Gamma_0 z}\,\frac{\Gamma_0 v_0 Z_0}{2}\int_{z'}^{z} e^{-\Gamma_0 z}\rho_\omega(z)\,dz \end{aligned} \tag{2}$$

where the common factor $e^{j\omega t}$ is omitted. $\Gamma_0 = j(\omega/v_0)$, $j = \sqrt{-1}$ and ω is the angular frequency. C_1 and C_2 are arbitrary constants which will be determined by the boundary conditions at the both ends of the beam. The first two terms are the solutions of the homogeneous equation (or the complementary functions) and are just the cold circuit waves. The third and the fourth terms are functions of electron charge density and are the particular solution of the equation.

Let us consider a long traveling-wave tube in which the beam starts from $z = 0$ and ends at $z = D$. The motion of electrons observed at any particular position is periodic in time, though it varies from point to point along the beam. To simplify the picture, we may divide the beam along the tube into small sections and consider each of them as a current element uniform in z and periodic in time. Each section of beam, or each current element excites on the circuit a pair of waves equal in amplitudes, one propagating toward the right (i.e., forward) and the other, toward the left. One may in fact imagine that these are trains of waves supported by the periodic motion of the electrons in that section of the beam. Obviously, a superposition of these waves excited by the whole beam gives the actual electromagnetic field distribution on the circuit. One may thus compute the forward traveling wave at z by summing all the waves at z which come from the left. Stated more specifically, the forward traveling energy at z is contributed by the waves excited by the current elements at the left of the point z. Similarly the backward traveling energy, (or the backward wave) at z is contributed by the waves excited by the current elements at the right of the point z. It follows obviously from this picture that there is no forward wave at $z = 0$ (except one corresponding to the input signal), and no backward wave at $z = D$. (This implies that the output circuit is matched.) With these boundary conditions, (1) is reduced to

$$V(z) = V_{\text{input}}\, e^{-\Gamma_0 z} + e^{-\Gamma_0 z}\,\frac{\Gamma_0 v_0 Z_0}{2}\int_0^z e^{\Gamma_0 z}\rho_\omega(z)\,dz + e^{\Gamma_0 z}\,\frac{\Gamma_0 v_0 Z_0}{2}\int_z^D e^{-\Gamma_0 z}\rho_\omega(z)\,dz \tag{3}$$

Equations (2) and (3) have been obtained by Poulter.[8] The first term of (3) is the wave induced by the input signal. It propagates as though the beam were not present. The second term is the voltage at z contributed by the charges between $z = 0$ and $z = z$. It is just the voltage of the forward wave described earlier. Similarly the third term which is the voltage at z contributed by the charges between $z = z$ and $z = D$ is the voltage of the backward wave at the point z. Denote F and B respec-

tively the voltages of the forward and the backward waves, we have

$$F(z) = V_{\text{input}}\, e^{-\Gamma_0 z} + e^{-\Gamma_0 z}\, \frac{\Gamma_0 v_0 Z_0}{2} \int_0^z e^{\Gamma_0 z}\, \rho_\omega(z)\, dz \tag{4}$$

$$B(z) = e^{\Gamma_0 z}\, \frac{\Gamma_0 v_0 Z_0}{2} \int_z^D e^{-\Gamma_0 z} \rho_\omega(z)\, dz \tag{5}$$

It can be shown by direct substitution that F and B satisfy respectively the differential equations

$$\frac{\partial F(z, t)}{\partial z} + \frac{1}{v_0}\, \frac{\partial F(z, t)}{\partial t} = \frac{Z_0}{2}\, \frac{\partial \rho_\omega(z, t)}{\partial t} \tag{6}$$

$$\frac{\partial B(z, t)}{\partial z} - \frac{1}{v_0}\, \frac{\partial B(z, t)}{\partial t} = -\frac{Z_0}{2}\, \frac{\partial \rho_\omega(z, t)}{\partial t} \tag{7}$$

We put (4) and (5) in the form of (6) and (7) simply because the differential equations are easier to manipulate than the integral equations. In fact, we should start the analysis from (6) and (7) if it were not for a physical picture useful to the understanding of the problem. Equations (6) and (7) have the advantage of not being restricted by the boundary conditions at $z = 0$ and D, which we have just imposed to derive (4) and (5). Actually, we can derive (6) and (7) directly from the Brillouin model[13] in the following manner. Suppose V, I and Z_0 are respectively the voltage, current and the characteristic impedance of a transmission line system in the usual sense. $(V + IZ_0)$ must then be the forward wave and $(V - IZ_0)$ must be the backward wave. If we substituted F and B in these forms into (1) of the Brillouin's paper,[13] we should obtain exactly (6) and (7).

It is obvious that the first and third terms of (2) are respectively the complementary function and the particular solution of (6), and similarly the second and the fourth terms of (2) are respectively the complementary function and the particular solution of (7). From now on, we shall overlook the complementary functions which are far from synchronism with the beam and are only useful in matching the boundary conditions. It is the particular solutions which act directly on the electron motion. With these in mind, it is convenient to put F and B in the form

$$F(z, t) = \frac{Z_0 I_0}{4C}\, [a_1(y) \cos \varphi - a_2(y) \sin \varphi] \tag{8}$$

$$B(z, t) = \frac{Z_0 I_0}{4C}\, [b_1(y) \cos \varphi - b_2(y) \sin \varphi] \tag{9}$$

where $a_1(y)$, $a_2(y)$, $b_1(y)$ and $b_2(y)$ are functions of y. y and φ are independent variables and have been used by Nordsieck to replace the variables, z and t, such as

$$y = C\frac{\omega}{u_0}z$$

$$\varphi = \omega\left(\frac{z}{v_0} - t\right)$$

Here as defined earlier, v_0 is the phase velocity of the cold circuit and u_0 the average velocity of the electrons. They are related by the parameter b defined by Pierce as

$$\frac{u_0}{v_0} = \frac{1}{(1 - bC)}$$

C is the gain parameter also defined by Pierce,

$$C^3 = \frac{Z_0 I_0}{4V_0}$$

in which, V_0 and I_0 are respectively the beam voltage and current. Adding (6) to (7), we obtain an important relation between F and B, that is,

$$\frac{\partial F(z,t)}{\partial z} + \frac{1}{v_0}\frac{\partial F(z,t)}{\partial t} = -\frac{\partial B(z,t)}{\partial z} + \frac{1}{v_0}\frac{\partial B(z,t)}{\partial t} \tag{10}$$

Substituting (8) and (9) into (10) and carrying out some algebraic manipulation, we obtain

$$\begin{aligned} b_1(y) &= -\frac{C}{2(1 + bC)}\frac{d}{dy}[a_2(y) + b_2(y)] \\ b_2(y) &= \frac{C}{2(1 + bC)}\frac{d}{dy}[a_1(y) + b_1(y)] \end{aligned} \tag{11}$$

or

$$\begin{aligned} B(z,t) = &-\frac{Z_0 I_0}{4C}\frac{C}{2(1 + bC)} \\ &\cdot\left[\frac{d(a_2(y) + b_2(y))}{dy}\cos\varphi + \frac{d(a_1(y) + b_1(y))}{dy}\sin\varphi\right] \end{aligned} \tag{12a}$$

For better understanding of the problem, we shall first solve (12a) approximately. Assuming for the moment that $b_1(y)$ and $b_2(y)$ are small compared with $a_1(y)$ and $a_2(y)$ and may be neglected in the right-hand

member of the equation, we obtain for the first order solution

$$B(z, t) \cong \frac{Z_0 I_0}{4C}\left(-\frac{C}{2(1+bC)}\left[\frac{da_1(y)}{dy}\sin\varphi + \frac{da_2(y)}{dy}\cos\varphi\right]\right) \quad (12b)$$

Of course, the solution (12b) is justified only when $b_1(y)$ and $b_2(y)$ thus obtained are small compared with $a_1(y)$ and $a_2(y)$. The exact solution[16] of B obtained by successive approximation reads

$$\begin{aligned} B(z, t) = \frac{Z_0 I_0}{4C}\Bigg(&-\frac{C}{2(1+bC)}\left[\frac{da_1(y)}{dy}\sin\varphi + \frac{da_2(y)}{dy}\cos\varphi\right] \\ &+ \frac{C^2}{4(1+bC)^2}\left[-\frac{d^2a_1(y)}{dy^2}\cos\varphi + \frac{d^2a_2(y)}{dy^2}\sin\varphi\right] + \cdots\Bigg) \end{aligned} \quad (12c)$$

It may be seen that the term involving

$$\frac{C^2}{4(1+bC)^2}$$

and the higher order terms are neglected in our approximate solution. For C equal to few tenths, the difference between (12b) and (12c) only amounts to few per cent. We thus can calculate the backward wave by (12b) or (12c) from the derivatives of the forward wave. To obtain the complete solution of the backward wave, we should add to (12b) or (12c) a solution of the homogeneous equation. We shall return to this point later.

4. WORKING EQUATIONS

With this discussion of the backward wave, we are now in a position to derive the working equations on which our calculations are based. In Nordsieck's notation, each electron is identified by its initial phase. Thus, $\varphi(y, \varphi_0)$ and $Cu_0w(y, \varphi_0)$ are respectively the phase and the ac velocity of the electron which has an initial phase φ_0. It should be remembered that y is equal to

$$C\frac{\omega}{u_0}z$$

and is used by Nordsieck as an independent variable to replace the variable z. Let us consider an electron which is at z_0 when $t = 0$ and is at z (or y) when $t = t$. Its initial phase is then

$$\varphi_0 = \frac{\omega z_0}{u_0}$$

and its phase at y is

$$\varphi(y, \varphi_0) = \omega\left(\frac{z}{v_0} - t\right)$$
$$= \omega\left(\frac{z}{u_0} - t\right) + by$$

The velocity of the electron is expressed as

$$\frac{dz}{dt} = u_0[1 + Cw(y, \varphi_0)]$$

where u_0 is the average velocity of the electrons, and, $Cu_0w(y, \varphi_0)$ as mentioned earlier, is the ac velocity of the electron when it is at the position y. The electron charge density near an electron which has an initial phase φ_0 and which is now at y, can be computed by the equation of conservation of charge,[15] it is

$$\rho(y, \varphi_0) = -\frac{I_0}{u_0}\left|\frac{d\varphi_0}{d\varphi(y, \varphi_0)}\right|\frac{1}{1 + Cw(y, \varphi_0)} \tag{13}$$

One should recall here that I_0 is the dc beam current and has been defined as a positive quantity. When several electrons with different initial phases are present at y simultaneously, a summation of

$$\left|\frac{d\varphi_0}{d(y, \varphi_0)}\right|$$

of these electrons should be used in (13). From (13), the fundamental component of the electron charge density is

$$\rho_\omega(z, t) = -\frac{1}{\pi}\frac{I_0}{u_0}\left(\sin\varphi\int_0^{2\pi} d\varphi_0\,\frac{\sin\varphi(y, \varphi_0)}{1 + Cw(y, \varphi_0)} + \cos\varphi\int_0^{2\pi} d\varphi_0\,\frac{\cos\varphi(y, \varphi_0)}{1 + Cw(y, \varphi_0)}\right) \tag{14}$$

These are important relations given by Nordsieck and should be kept in mind in connection with later work. In addition, we shall frequently use the transformation

$$\frac{d}{dt} = \frac{dz}{dt}\frac{\partial}{\partial z} = C\omega(1 + Cw(y, \varphi_0))\frac{\partial}{\partial y}$$

which is written following the motion of the electron. Let us start from the forward wave. It is computed by means of (6). After substituting (8) and (14) into (6), we obtain by equating the sin φ and the cos φ

terms

$$\frac{da_1(y)}{dy} = -\frac{2}{\pi}\int_0^{2\pi} d\varphi_0 \frac{\sin \varphi(y, \varphi_0)}{1 + Cw(y, \varphi_0)} \tag{15}$$

$$\frac{da_2(y)}{dy} = -\frac{2}{\pi}\int_0^{2\pi} d\varphi_0 \frac{\cos \varphi(y, \varphi_0)}{1 + Cw(y, \varphi_0)} \tag{16}$$

Next we shall calculate the electron motion. We express the acceleration of an electron in the form

$$\frac{d^2z}{dt^2} = C\omega u_0(1 + Cw(y, \varphi_0)) \frac{dw(y, \varphi_0)}{dy}$$

and calculate the circuit field by differentiating F in (8) and B in (12c) with respect to z. One thus obtains from Newton's law

$$\begin{aligned} 2[1 + Cw(y, \varphi_0)] \frac{dw(y, \varphi_0)}{dy} &= (1 + bC)[a_1(y) \sin \varphi + a_2(y) \cos \varphi] \\ &\quad - \frac{C}{2}\left[\frac{da_1(y)}{dy} \cos \varphi - \frac{da_2(y)}{dy} \sin \varphi\right] \\ &\quad + \frac{C^2}{4(1 + bC)}\left[\frac{d^2a_1(y)}{dy^2} \sin \varphi + \frac{d^2a_2(y)}{dy^2} \cos \varphi\right] - \frac{2e}{u_0 m\omega C^2} E_s \end{aligned} \tag{17}$$

Here E_s is the space charge field, which will be discussed in detail later. Finally a relation between $w(y, \varphi_0)$ and $\varphi(y, \varphi_0)$ is obtained by means of (13)

$$\frac{d\varphi(y, \varphi_0)}{dy} - b = \frac{w(y, \varphi_0)}{1 + Cw(y, \varphi_0)} \tag{18}$$

Equations (15), (16), (17) and (18) are the four working equations which we have derived for finite C and including space charge.

Instead of writing the equations in the above form, Rowe,[10] ignoring the backward wave, derives (15) and (16) directly from the circuit equation (1). He obtains an additional term

$$\frac{C}{2}\frac{d^2a_2}{dy^2}$$

for (15) and another term

$$\frac{C}{2}\frac{d^2a_1}{dy^2}$$

for (16). It is apparent that the backward wave, though generally a small quantity, does influence the terms involving C.

5. THE SPACE CHARGE EXPRESSION

We have mentioned earlier that the space charge field is computed from the disc-model suggested by Tien, Walker and Wolontis. In their calculation, the force excited on one disc by the other is approximated by an exponential function

$$F_s = \frac{q^2}{2\pi r_0^2 \epsilon_0} e^{-[\alpha(z'-z)/r_0]}$$

Here r_0 is the radius of the disc or the beam, q is the charge carried by each disc, and ϵ_0 is the dielectric constant of the medium. The discs are supposed to be respectively at z and z'. α is a constant and is taken equal to 2.

Consider two electrons which have their initial phases φ_0 and φ_0' and which reach the position y (or z) at times t and t' respectively. The time difference,

$$t - t' = \frac{1}{\omega}\left[\omega t - \frac{\omega}{v_0} z - \left(\omega t' - \frac{\omega}{v_0} z\right)\right] = \frac{1}{\omega}[\varphi(y, \varphi_0') - \varphi(y, \varphi_0)]$$

multiplied by the velocity of the electron $u_0[1 + Cw(y, \varphi_0')]$ is obviously the distance between the two electrons at the time t. Thus

$$(z' - z)_{t=t} = \frac{1}{\omega}[\varphi(y, \varphi_0') - \varphi(y, \varphi_0)]u_0[1 + Cw(y, \varphi_0')] \qquad (19a)$$

In this equation, we are actually taking the first term of the Taylor's expansion,

$$(z' - z)_{t=t} = \left.\frac{dz(y, \varphi_0')}{dt}\right|_{t=t} (t - t') + \frac{1}{2}\left.\frac{d^2z(y, \varphi_0')}{dt^2}\right|_{t=t} (t - t')^2 + \cdots \qquad (19b)$$

It is clear that the electrons at y may have widely different velocities after having traveled a long distance from the input end, but changes in their velocities, in the vicinity of y and in a time-period of around 2π, are relatively small. This is why we must keep the first term of (19b) and may neglect the higher order terms. From (19a) the space charge field E_s in (17) is

$$\frac{2e}{m\omega C^2 u_0} E_s = \left(\frac{\omega_p}{\omega C}\right)^2 \int_{-\infty}^{+\infty} e^{-k|\varphi(y,\varphi_0+\phi)-\varphi(y,\varphi_0)|[1+Cw(y,\varphi_0+\phi)]} \, d\phi \, \operatorname{sgn}(\varphi(\varphi_0 + \varphi) - \phi(y, \varphi_0))$$

Here, e/m is the ratio of electron charge to mass, ω_p is the electron

angular plasma frequency for a beam of infinite extent, and k is

$$k = \frac{\alpha}{\frac{\omega}{u_0} r_0} = \frac{2}{\frac{\omega}{u_0} r_0} \tag{20}$$

In the small C theory, the distribution of electrons in time or in time-phase at z is approximately the same as the distribution in z (also expressed in the unit of time-phase) at the vicinity of z. This is, however, not true when C becomes finite. The difference between the time and space distributions is the difference between unity and the factor $(1 + Cw(y, \varphi_0'))$. We can show later that the error involved in considering the time phase as the space phase can easily reach 50 per cent or more, depending on the velocity spread of the electrons.

6. NUMERICAL CALCULATIONS

Although the process of carrying out numerical computations has been discussed in Nordsieck's paper, it is desirable to recapitulate here a few essential points including the new feature added. Using the working equations (15), (16), (17) and (18),

$$\frac{da_1}{dy}, \frac{da_2}{dy}, \frac{dw}{dy} \quad \text{and} \quad \frac{d\varphi}{dy}$$

are calculable from a_1, a_2, w and φ. The distance is divided into equal intervals of Δy, and the forward integrations of a_1, a_2, w and φ are performed by a central difference formula

$$a_1(y + \Delta y) = a_1(y) + \left.\frac{da_1}{dy}\right|_{y+1/2\Delta y} \cdot \Delta y$$

In addition,

$$\frac{d^2a_1}{dy^2} \quad \text{and} \quad \frac{d^2a_2}{dy^2}$$

in (17) are computed from the second difference formula such that

$$\left.\frac{d^2a_1}{dy^2}\right|_{y=y} = \left[\left.\frac{da_1}{dy}\right|_{y+1/2\Delta y} - \left.\frac{da_1}{dy}\right|_{y-1/2\Delta y}\right] \div \Delta y$$

We thus calculate the behavior along the tube by forward integration made in steps of Δy, starting from $y = 0$. At $y = 0$ the initial conditions are determined from Pierce's linearized theory. Because of its complications in notation, this will be discussed in detail in Appendix I.

Numerical calculations were carried out using the 701-type I.B.M.

TABLE I

Case No.	QC	k	C	b	μ_1	μ_2	y(SAT.)	$A(y)$(SAT.)	$\lvert\theta(y)-\mu_2 y\rvert$ (SAT.)
1	0.1	2.5	0.05	0.455	μ_1 max. 0.795662	−0.748052	5.6	1.26	0.415
2	0.1	2.5	0.1	0.541	μ_1 max. 0.827175	−0.787624	5.2	1.24	0.482
3	0.1	2.5	0.1	1.145	$0.941\mu_1$ max. 0.778535	−1.05370	5.6	1.31	0.820
4	0.1	2.5	0.1	1.851	$0.66\mu_1$ max. 0.550736	−1.37968	7.0	1.36	1.05
5	0.1	2.5	0.2	0.720	μ_1 max. 0.900312	−0.873606	4.8	1.02	0.726
6	0.2	1.25	0.1	0.875	μ_1 max. 0.769795	−1.04078	5.9	1.22	0.570
7	0.2	1.25	0.1	1.422	$0.951\mu_1$ max. 0.724527	−1.29469	6.0	1.30	0.803
8	0.2	1.25	0.1	2.072	$0.666\mu_1$ max. 0.512528	−1.60435	7.6	1.35	1.08
9	0.2	2.5	0.05	0.765	μ_1 max. 0.731493	−0.973376	6.2	1.30	0.412
10	0.2	2.5	0.1	0.875	μ_1 max. 0.769795	−1.04078	5.8	1.22	0.490
11	0.2	2.5	0.1	1.422	$0.941\mu_1$ max. 0.724527	−1.29469	6.0	1.26	0.720
12	0.2	2.5	0.1	2.072	$0.666\mu_1$ max. 0.512528	−1.60435	7.2	1.25	0.92
13	0.2	2.5	0.1	2.401	$0.300\mu_1$ max. 0.230930	−1.76243	12.4	1.24	1.36
14	0.2	2.5	0.15	0.976	μ_1 max. 0.812900	−1.10656	5.4	1.11	0.572
15	0.2	2.5	0.15	1.549	$0.941\mu_1$ max. 0.765101	−1.37540	5.8	1.14	1.03
16	0.2	2.5	0.15	2.2311	$0.666\mu_1$ max. 0.541234	−1.70180	7.0	1.12	1.22
17	0.2	2.5	0.15	2.575	$0.300\mu_1$ max. 0.243864	−1.86844	10.8	1.04	1.34
18	0.4	2.5	0.05	1.25	μ_1 max. 0.653014	−1.36746	7.6	1.26	0.315
19	0.4	2.5	0.1	1.38	μ_1 max. 0.701470	−1.47477	6.6	1.11	0.674
20	0.4	2.5	0.1	1.874	$0.941\mu_1$ max. 0.660223	−1.71341	7.8	1.19	1.05
21	0.4	2.5	0.1	2.458	$0.666\mu_1$ max. 0.467038	−1.99840	8.6	1.09	1.25

equipment. The problem was programmed by Miss D. C. Legaus. The cases computed are listed in Table I in which μ_1 and μ_2 are respectively Pierce's x_1 and y_1, and A, $(\theta - \mu_2 y)$ and y at saturation will be discussed later. All the cases were computed with $\Delta y = 0.2$ using a model based on 24 electron discs per electronic wavelength. To estimate the error involved in the numerical work, Case (10) has been repeated for 48 electrons and Cases (10) and (19) for $\Delta y = 0.1$. The results obtained by using different numbers of electrons are almost identical and those obtained by varying the interval Δy indicate a difference in $A(y)$ less than 1 per cent for Case (10) and about 6 per cent for Case (19). As error generally increases with QC and C the cases listed in this paper are limited to $QC = 0.4$ and $C = 0.15$. For larger QC or C, a model of more electrons or a smaller interval of integration, or both should be used.

7. POWER OUTPUT AND EFFICIENCY

Define

$$\begin{aligned} A(y) &= \tfrac{1}{4}\sqrt{a_1(y)^2 + a_2(y)^2} \\ -\theta(y) &= \tan^{-1}\frac{a_2(y)}{a_1(y)} + by \end{aligned} \tag{21}$$

We have then

$$F(z, t) = \frac{Z_0 I_0}{C} A(y) \cos\left[\frac{\omega z}{u_0} - \omega t - \theta(y)\right] \tag{22}$$

The power carried by the forward wave is therefore

$$\left(\frac{F^2}{Z_0}\right)_{\text{average}} = 2CA^2 I_0 V_0 \tag{23}$$

and the efficiency is

$$\text{Eff.} = \frac{2CA^2 I_0 V_0}{I_0 V_0} = 2CA^2 \qquad \text{or} \qquad \frac{\text{Eff.}}{C} = 2CA^2 \tag{24}$$

In Table I, the values of $A(y)$, $\theta(y)$ and y at the saturation level are listed for every case computed. We mean by the saturation level, the distance along the tube or the value of y at which the voltage of the forward wave or the forward traveling power reaches its first peak. The Eff./C at the saturation level is plotted in Fig. 1 versus QC, for $k = 2.5$, b for maximum small-signal gain and C = small, 0.05, 0.1, 0.15 and 2. It is also plotted versus b in Fig. 2 for $QC = 0.2$, $k = 2.5$ and C = small, 0.1 and 0.15, and in Fig. 3 for $QC = 0.2$, $C = 0.1$ and $k = 1.25$ and 2.50. In Fig. 2 the dotted curves indicate the values of b at which

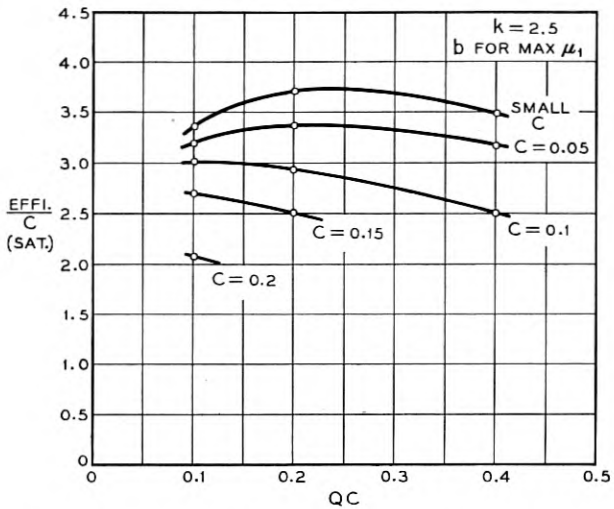


Fig. 1 — The saturation eff./C versus QC, for $k = 2.5$, b for maximum small-signal gain and C = small, 0.1, 0.15 and 0.2.

$\mu_1 = \mu_1(\max)$, 0.94 $\mu_1(\max)$, 0.67 $\mu_1(\max)$ and 0.3 $\mu_1(\max)$, respectively. It is seen that Eff./C decreases as C increases particularly when b is large. It is almost constant between $k = 1.25$ and 2.50 and decreases slowly for large values of C when QC increases.

The (Eff./C) at saturation is also plotted versus QC in Fig. 4(a) for small C, and in Fig. 4(b) for $C = 0.1$. It should be noted that for $C = 0.1$ the values of Eff./C fall inside a very narrow region say between 2.5 to 3.5, whereas for small C they vary widely.

8. VELOCITY SPREAD

In a traveling-wave amplifier, when electrons are decelerated by the circuit field, they contribute power to the circuit, and when electrons are accelerated, they gain kinetic energy at the expense of the circuit power. It is therefore of interest to plot the actual velocities of the fastest and the slowest electrons at the saturation level and find how they vary with the parameters QC, C, b and k. This is done in Fig. 5. These velocities are also plotted versus y for Case 10 in Fig. 6, in which, the $A(y)$ curve is added for reference.

9. THE BACKWARD WAVE AND THE FUNDAMENTAL COMPONENT OF THE ELECTRON CHARGE DENSITY

Our calculation of efficiency has been based on the power carried by the forward wave only. One may, however, ask about the actual power

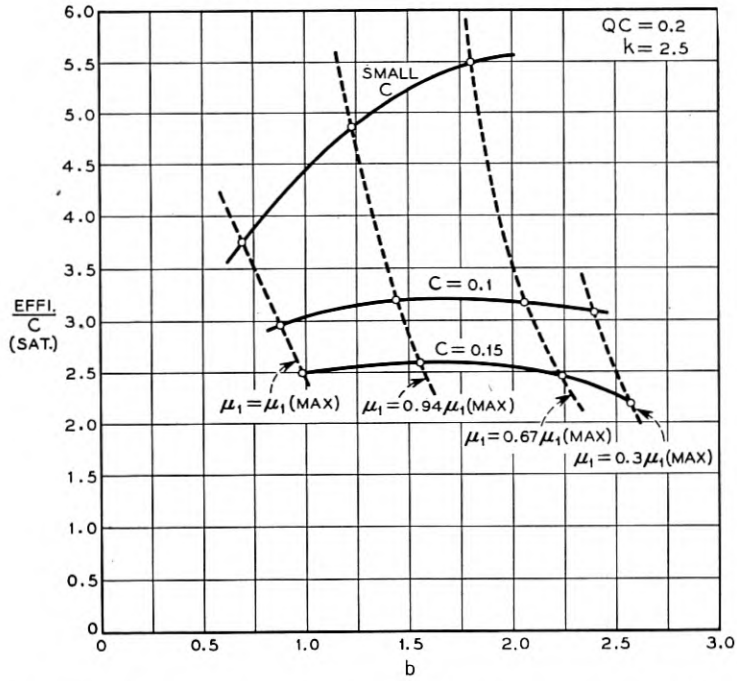


Fig. 2 — The saturation eff./C versus b, for $k = 2.5$, $QC = 0.2$, and C = small, 0.1 and 0.15. The dotted curves indicate the values of b for $\mu_1 = 1, 0.94, 0.67$, and 0.3 of μ_1(max) respectively.

output in the presence of the backward wave. For simplicity, we shall use the approximate solution (12b) which can be written in the form

$$B(z, t) \cong \text{Real Component of}$$

$$\left(\frac{Z_0 I_0}{4C}\,\frac{C}{2(1+bC)}\sqrt{\left(\frac{da_1(y)}{dy}\right)^2+\left(\frac{da_2(y)}{dy}\right)^2}\,e^{j\omega t-\Gamma_0 z-by+j\xi}\right) \tag{12d}$$

with

$$\tan\xi = \left(-\frac{da_1(y)}{dy}\right)\Big/\left(-\frac{da_2(y)}{dy}\right)$$

As mentioned earlier that the complete solution of (6) is obtained by adding to (12b) a complementary function such that

$$B(z, t) = C_1 e^{-j\omega t+\Gamma_0 z}$$

$$+\frac{Z_0 I_0}{4C}\,\frac{C}{2(1+bC)}\sqrt{\left(\frac{da_1}{dy}\right)^2+\left(\frac{da_2}{dy}\right)^2}\,e^{j\omega t-\Gamma_0 z-by+j\xi} \tag{25}$$

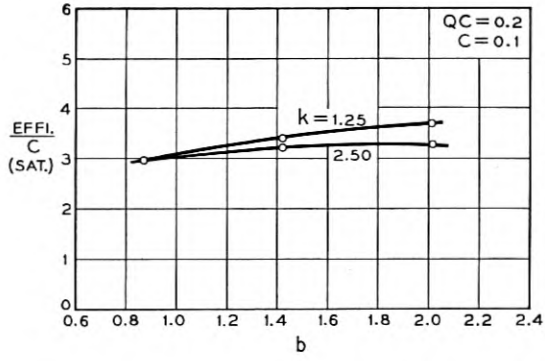


Fig. 3 — The saturation eff./C versus b, for $QC = 0.2$ $C = 0.1$ and $k = 1.25$ and 2.50.

If the output circuit is matched by cold measurements, the backward wave must be zero at the output end, $z = D$. This determines C_1, that is,

$$C_1 = -\frac{Z_0 I_0}{4C}\,\frac{C}{2(1+bC)}\sqrt{\left(\frac{da_1(y)}{dy}\right)^2_{z=D} + \left(\frac{da_2(y)}{dy}\right)^2_{z=D}}\; e^{\Gamma_0(2+bc)D+j\xi}$$

or

$$C_1 e^{j\omega t+\Gamma_0 z} = -\frac{Z_0 I_0}{4C}\,\frac{C}{2(1+bC)}\sqrt{\left(\frac{da_1(y)}{dy}\right)^2_{z=D} + \left(\frac{da_2(y)}{dy}\right)^2_{z=D}} \cdot e^{\Gamma_0(2+bC)D+j\xi} e^{j\omega t+\Gamma_0 z} \tag{26}$$

The backward wave therefore consists of two components. One compo-

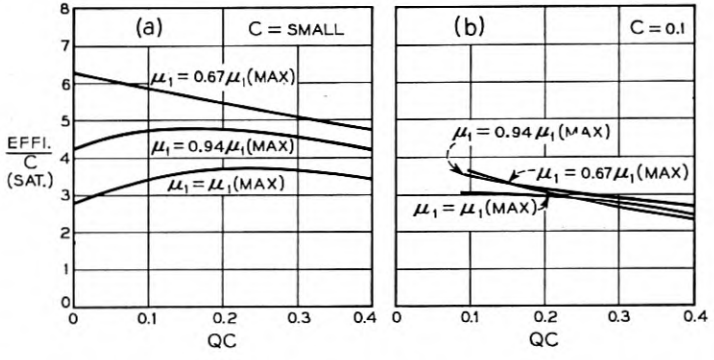


Fig. 4 — The saturation eff./C versus QC for b corresponding $\mu_1 = 1$, 0.94 and 0.67 of μ_1(max), (a) for C = small, (b) for $C = 0.1$.

nent is coupled to the beam and has an amplitude equal to

$$\frac{Z_0 I_0}{4C} \frac{C}{2(1 + bC)} \sqrt{\left(\frac{da_1}{dy}\right)^2 + \left(\frac{da_2}{dy}\right)^2}$$

which generally grows with the forward wave. It thus has a much larger amplitude at the output end than at the input end. The other component is a wave of constant amplitude, which travels in the direction opposite to the electron flow with a phase velocity equal to that of the cold circuit. At the output end, $z = D$, both components have the same amplitude but are opposite in sign. One thus realizes that there exists a reflected wave of noticeable amplitude, in the form of (26), even though the output circuit is properly matched by cold measurements. Under such circumstances, the voltage at the output end is the voltage of the forward wave and the power output is the power carried by the forward wave only. This is computed in (23).

Since (26) is a cold circuit wave it may be eliminated by properly ad-

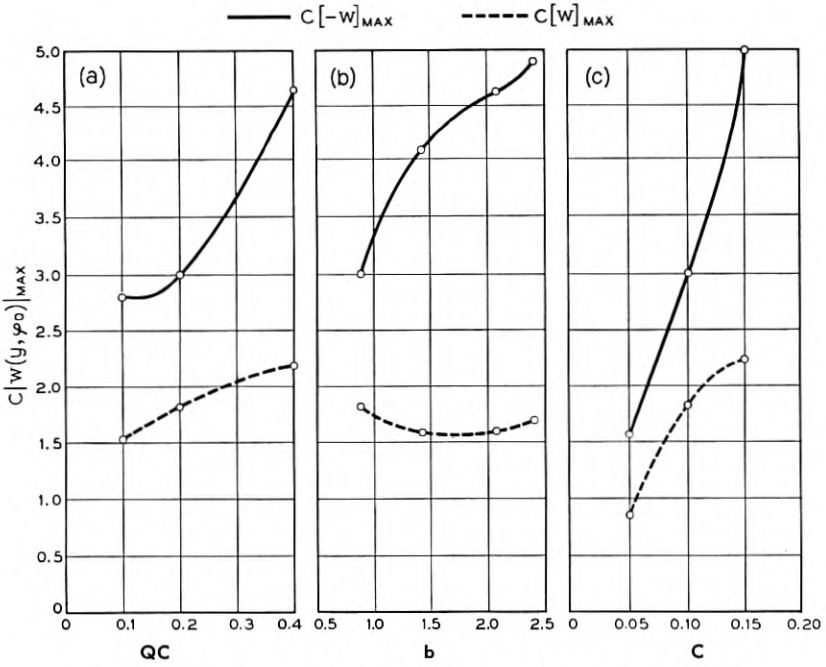


Fig. 5 — $Cw(y, \varphi_0)$ of the fast and the slowest electrons at the saturation level. (a) versus QC for $k = 2.5$, $C = 0.1$ and b for maximum small-signal gain; (b) versus b for $k = 2.50$, $C = 0.1$ and $QC = 0.2$; and (c) versus C for $k = 2.50$, $QC = 0.2$ and b for maximum small-signal gain.

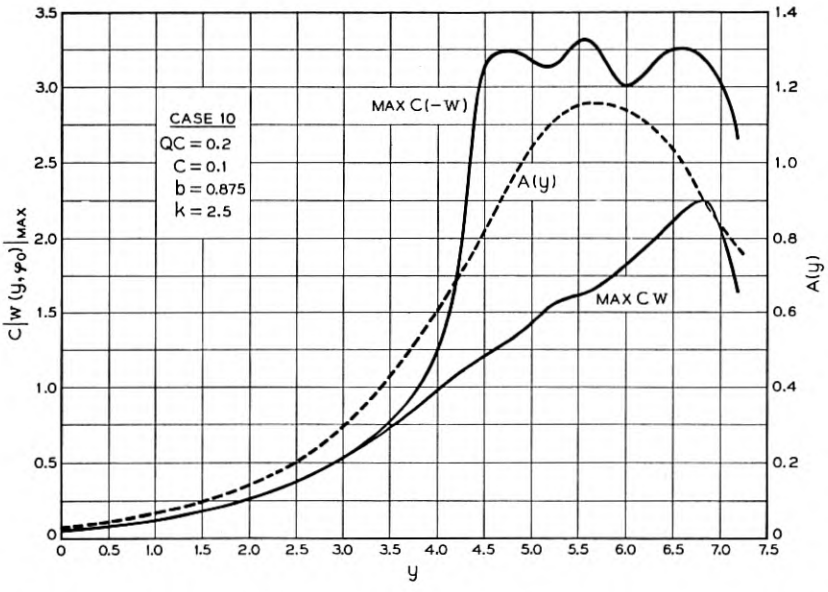


Fig. 6 — $Cw(y, \varphi_0)$ of the fast and the slowest electrons versus y for Case (10). $A(y)$ is also plotted in dotted lines for reference.

justing the impedance of the output circuit. This may be necessary in practice for the purpose of avoiding possible regenerative oscillation. In doing so, the voltage at $z = D$ is the sum of the voltage of the forward wave and that of the particular solution of the backward wave. In every case, the output power is always equal to the square of the net voltage actually at the output end divided by the impedance of the output circuit.

We find from (14), (15) and (16) that the fundamental component of electron charge density may be written as

$$\rho_\omega(z, t) = \frac{1}{2}\frac{I_0}{u_0}\left(\sin\varphi\,\frac{da_1(y)}{dy} + \cos\varphi\,\frac{da_2(y)}{dy}\right)$$

$$= \text{Real component of}\left(-\frac{1}{2}\frac{I_0}{u_0}\sqrt{\left(\frac{da_1(y)}{dy}\right)^2 + \left(\frac{da_2(y)}{dy}\right)^2}\right. \tag{26}$$

$$\left. \cdot\, e^{j\omega - \Gamma_0 z - by + j\xi}\right)$$

where $-I_0/u_0$ is the dc electron charge density, ρ_0.

If (26) is compared with (12d) or (12c), it might seem surprising that the particular solution of the backward wave is just equal to the funda-

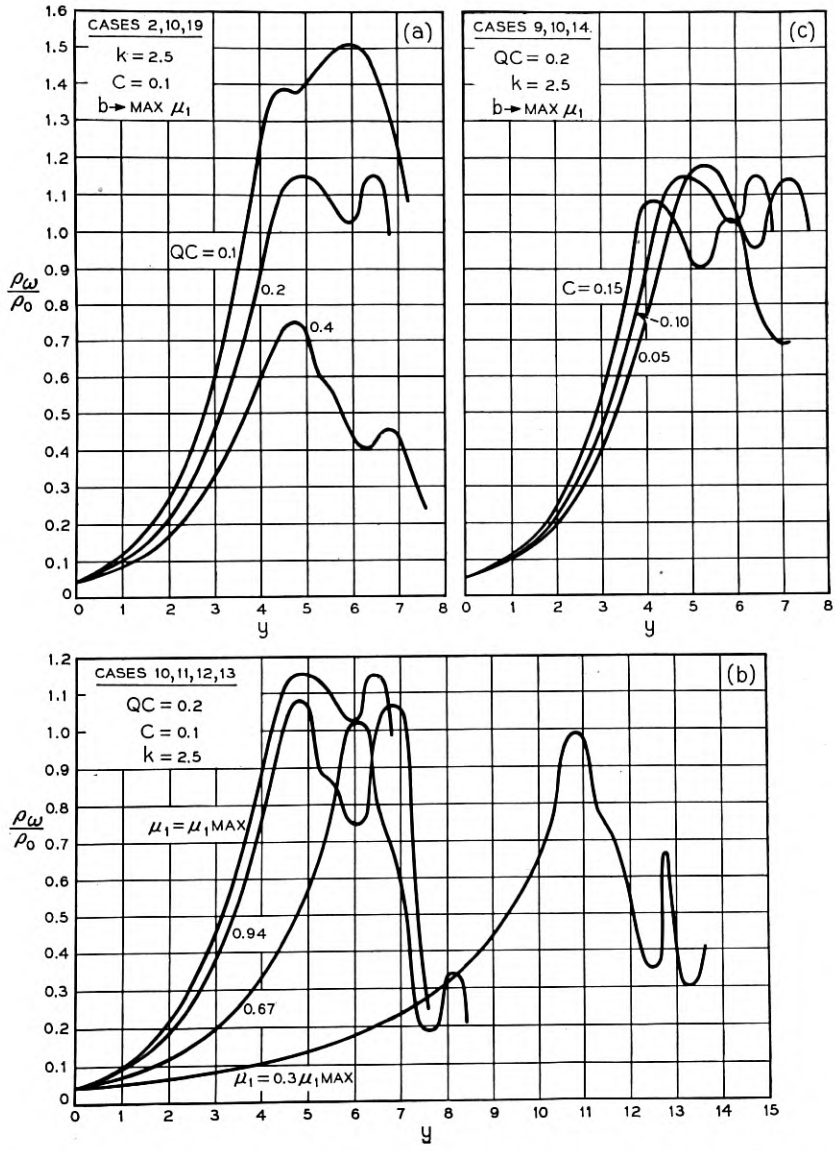


Fig. 7(a) — ρ_ω/ρ_0 versus y, (a) using QC as the parameter, for $k = 2.5$, $C = 0.1$, and b for maximum small-signal gain (Cases 2,10, and 19); (b) using b as the parameter, for $k = 2.50$, $C = 0.1$ and $QC = 0.2$ (Cases 10, 11, 12 and 13); and (c) using C as the parameter, for $k = 2.50$, $QC = 0.2$ and b for maximum small-signal gain (Cases 9, 10 and 14).

mental component of the electron charge density of the beam multiplied by a constant

$$\left(-\frac{Z_0 I_0}{4C}\,\frac{C}{2(1+bC)}\,\frac{2u_0}{I_0}\right) \tag{27}$$

The ratio of the electron charge density to the average charge density,

$$\frac{\rho_\omega(z)}{\rho_0}$$

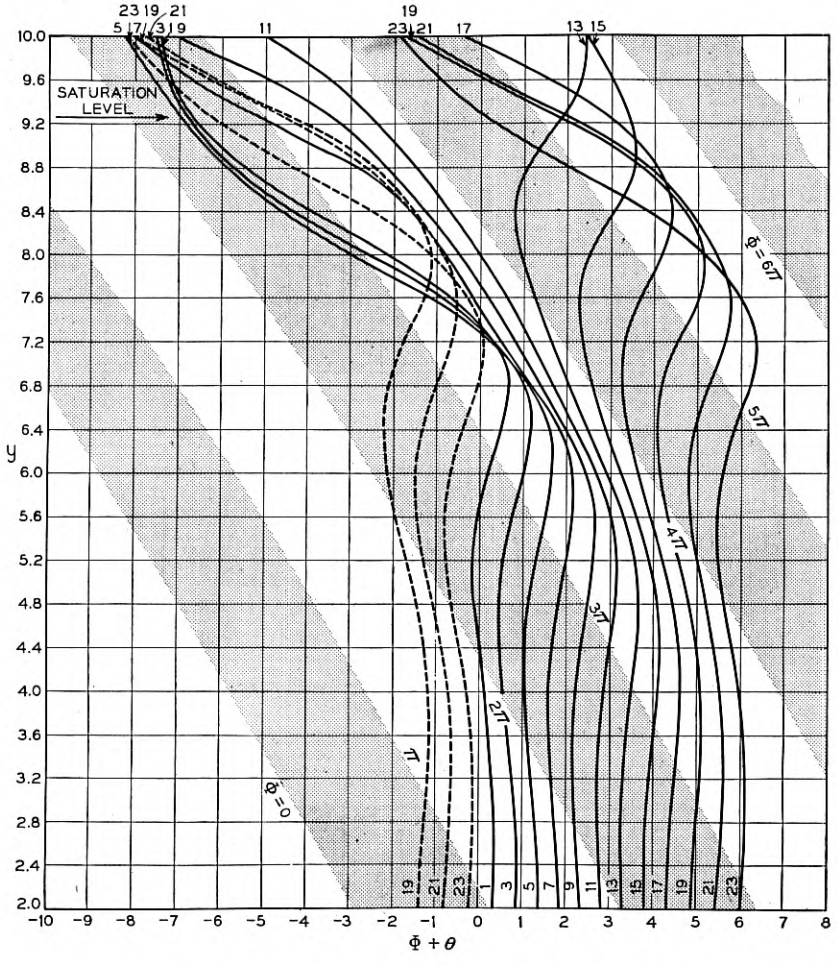


Fig. 8(a) — y versus $\varphi - by$ for $QC = 0.2$, $k = 2.5$, b for $\mu_1 = 0.67\mu_1(\max)$ and C = small.

is plotted in Fig. 7 versus y, using QC, b and C, as the parameters. They are also the curves for the backward wave (the component which is coupled to the beam) when multiplied by the proportional constant given in (27). It is interesting to see that the maximum values of ρ_ω/ρ_0 are between 1.0 and 1.2 for $QC = 0.2$ and decrease as QC increases. The peaks of the curves do not occur at the saturation values of y.

10. y VERSUS $(\varphi - by)$ DIAGRAMS

To study the effect of C, b, and QC on efficiency y versus $(\varphi - by)$ diagrams are plotted in Figs. 8(b), (c), (d) and (e) for Cases (21), (16), (10) and (21), respectively. $(\varphi - by)$ here is $(\Phi + \theta)$ in Nordsieck's notation. In these diagrams, the curves numbered from 1 to 24 correspond to the 24 electrons used in the calculation with each curve for one electron. Only odd numbered electrons are presented to avoid possible confusion arisen from too many lines. The reciprocal of the slope of the curve as

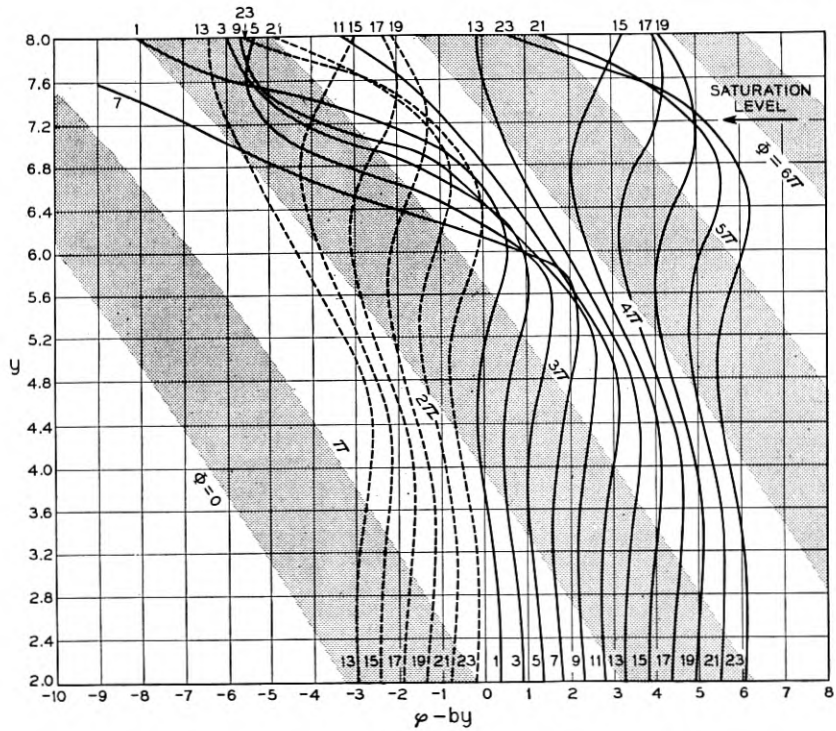


Fig. 8(b) — y versus $\varphi - by$ for $QC = 0.2$, $k = 2.5$, b for $\mu_1 = 0.67\mu_1(\max)$ and $C = 0.1$(Case 12).

given by (18) is proportional to the ac displacement of electron per unit of y. (In small-C theory it is proportional to the ac velocity of the electron.) Concentration of curves is obviously proportional to the charge-density distribution of the beam. In the shaded regions, the axially directed electric field of the circuit is negative and thus accelerates electrons in the positive z direction. Electrons are decelerated in the unshaded regions where the circuit field is positive. The boundaries of these regions are constant phase contours of the circuit wave. (They are constant Φ contours in Nordsieck's notation.)

These figures are actually the "space-time" diagrams which unfold the history of every electron from the input to the output ends. The effect of C can be clearly seen by comparing Figs. 8(a), (b) and (c). These diagrams are plotted for $QC = 0.2$, $k = 2.5$, b for $\mu_1 = 0.67\ \mu_1(\max)$ and for Fig. 8(a), C = small, for Fig. 8(b), $C = 0.1$, and for Fig. 8(c), $C = .15$. It may be seen that because of the velocity spread of the electrons, the saturation level in Fig. 8(a) is 9.3 whereas in Figs. 8(b) and (c), it is 7.2 and 7.0, respectively. It is therefore not surprising that Eff./C decreases as C increases.

The effects of b and QC may be observed by comparing Figs. 8(d) and (b), and Figs. 8(b) and (e), respectively. The details will not be described here. It is however suggested to study these diagrams with those given in the small-C theory.

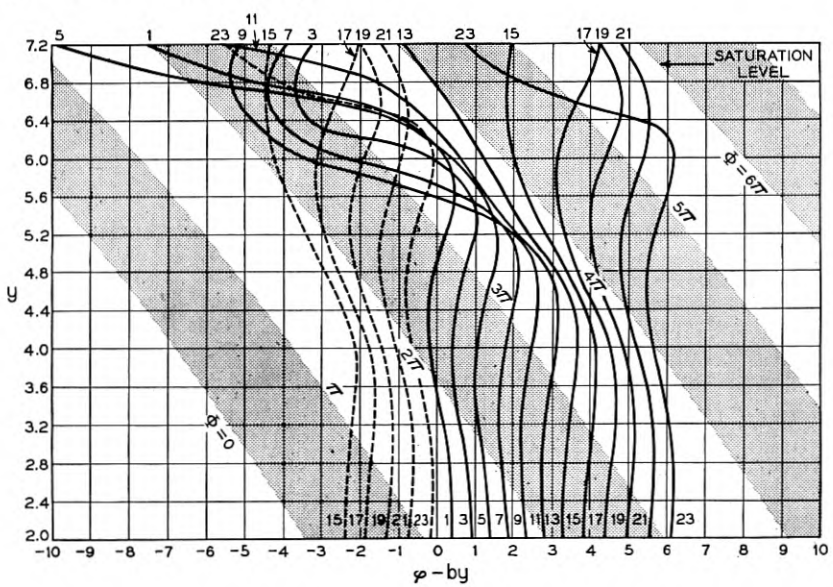


Fig. 8(c) — y versus $\varphi - by$ for $QC = 0.2$, $k = 2.5$, b for $\mu_1 = 0.67\mu_1(\max)$ and $C = 0.15$ (Case 16).

11. A QUALITATIVE PICTURE AND CONCULSIONS

We have exhibited in the previous sections the most important nonlinear characteristics of the traveling wave amplifier. Numerical computations based on a model of 24 electrons have been carried out for more than twenty cases covering useful ranges of design and operating parameters. The results obtained for the saturation Eff./C may be summarized as follows:

(1) It decreases with C particularly at large values of QC.

(2) For $C = 0.1$, it varies roughly from 3.7 for $QC = 0.1$ to 2.3 for $QC = 0.4$, and only varies slightly with b.

(3) For $C = 0.15$, it varies from 2.7 to 2.5 for QC from 0.1 to 0.2 and b corresponding to the maximum small-signal gain. It varies slightly with b for $QC = 0.2$.

(4) It is almost constant between $k = 1.25$ and 2.50.

In order to understand the traveling-wave tube better, it is important to have a simplified qualitative picture of its operation. It is obvious that to obtain higher amplification, more electrons must travel in the region where the circuit field is positive, that is, in the region where electrons

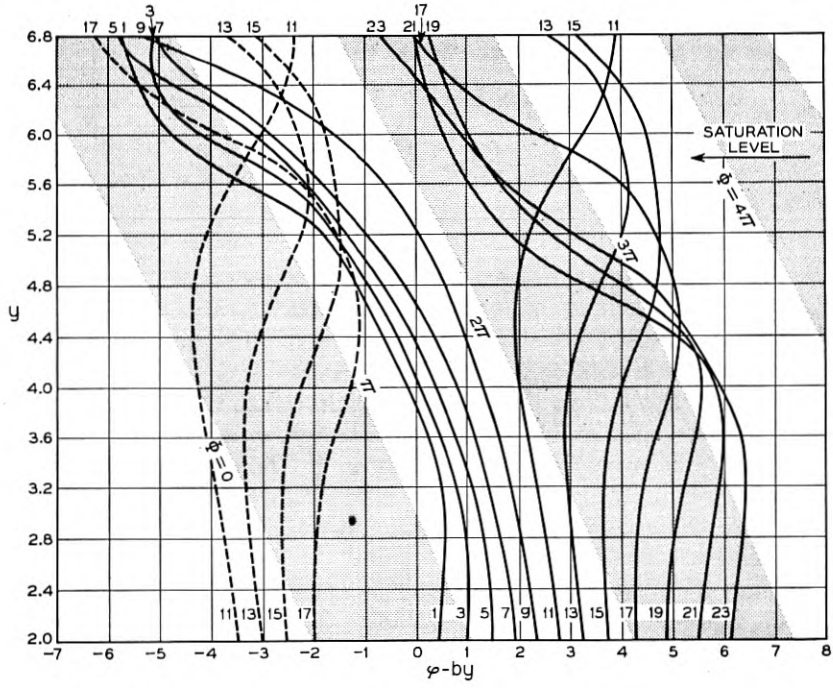


Fig. 8(d) — y versus $\varphi - by$ for $QC = 0.2$, $k = 2.5$, b for $\mu_1 = \mu_1(\max)$ and $C = 0.1$ (Case 10).

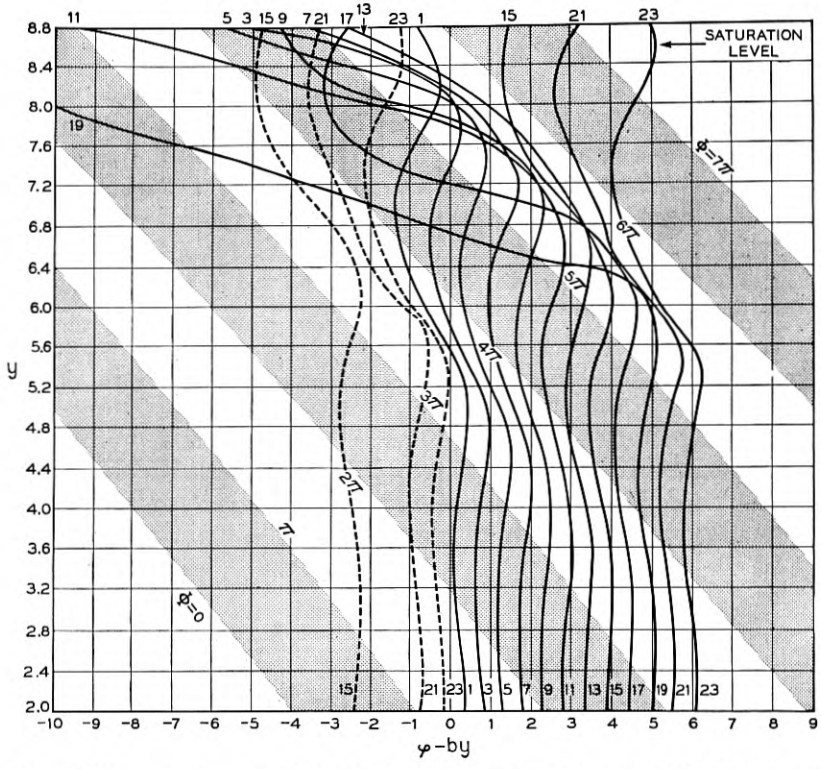


Fig. 8(e) — y versus $\varphi - by$ for $QC = 0.4$, $k = 2.5$, b for $\mu_1 = 0.67\mu_1(\max)$ and $C = 0.1$ (Case 21).

are decelerated by the circuit field. At the input end of the tube, electrons are uniformly distributed both in the accelerating and decelerating field regions. Bunching takes place when the accelerated electrons push forward and the decelerated ones press backward. The center of a bunch of electrons is located well inside the decelerating field region because the circuit wave travels slower than the electrons on the average (b is positive). The effectiveness of the amplification, or more specifically the saturation efficiency, therefore depends on (1), how tight the bunching is, and (2), how long a bunch travels inside the decelerating field region before its center crosses the boundary between the accelerating and decelerating fields.

For small-C, the ac velocities of the electrons are small compared with the dc velocity. The electron bunch stays longer with the decelerating circuit field before reaching the saturation level when b or QC is larger. On the other hand, the space charge force, or large QC or k tends to distort the bunching. As the consequence, the saturation efficiency increases with b, and decreases as k or QC increases. When C becomes finite how-

ever, the ac velocities of the electrons are no longer small as compared with their average speed. The velocity spread of the electrons becomes an important factor in determining the efficiency. Its effect is to loosen the bunching, and consequently it lowers the saturation level and reduces the limiting efficiency. It is seen from Figs. 5 and 6 that the velocity spread increases sharply with C and also steadily with b and QC. This explains the fact that in the present calculation the saturation Eff./C decreases with C and is almost constant with b whereas in the small-C theory it is constant with C and increases steadily with b.

12. ACKNOWLEDGEMENTS

The writer wishes to thank J. R. Pierce for his guidance during the course of this research, and L. R. Walker for many interesting discussions concerning the working equations and the method of calculating the backward wave. The writer is particularly grateful to Miss D. C. Leagus who, under the guidance of V. M. Wolontis, has carried out the numerical work presented with endless effort and enthusiasm.

APPENDIX

The initial conditions at $y = 0$ are computed from Pierce's linearized theory. For small-signal, we have

$$a_1(y) = 4A(y) \cos (b + \mu_2)y \tag{A-1}$$

$$a_2(y) = -4A(y) \sin (b + \mu_2)y \tag{A-2}$$

$$A(y) = \epsilon e^{\mu_1 y} \tag{A-3}$$

Here ϵ is taken equal to 0.03, a value which has been used in Tien-Walker-Wolontis' paper. Define

$$\frac{\partial X}{\partial y} = w(y, \varphi_0) \tag{A-4}$$

$$X = pe^{-j\varphi_0} + p^* e^{j\varphi_0} \tag{A-5}$$

where p^* is the conjugate of p. After substituting (A-1) to (A-5) into the working equations (15) to (18) and carrying out considerable algebraic work, we obtain exactly Pierce's equation.[16]

$$\mu^2 = \frac{(1 + jC\mu)(1 + bC)}{(j - \tfrac{1}{2}C\mu + j\tfrac{1}{2}bC)(\mu + jb)} - 4QC(1 + jC\mu)^2 \tag{A-6}$$

provided that

$$-\left(\frac{\omega_p}{\omega C}\right)^2 \int_{-0}^{+\infty} e^{-k|\varphi(y,\varphi_0+\phi)-\varphi(y,\varphi_0)|[1+Cw(y,\varphi_0+\phi)]} \cdot d\phi \operatorname{sgn} (\varphi(y, \varphi_0 + \phi) - \varphi(y, \varphi_0)) = 8\epsilon QC \cdot |(1 + jC\mu)(\mu + jb)| e^{\mu_1 y} \cos (\arg [(1 + jC\mu)(\mu + jb)] + \mu_2 y - \varphi_0) \tag{A-7}$$

Here $\mu = \mu_1 + j\mu_2$ or Pierce's $x_1 + jy_1$. From (A-7) the value of ω_p is determined for a given QC. The ac velocities of the electrons are derived from (A-4), such as,

$$w(y, \varphi_0) = -2\epsilon \left| \mu \frac{\mu + jb}{1 + jc\mu} \right| e^{\mu_1 y} \cos\left(\arg\left[\mu\left(\frac{\mu + jb}{1 + jC\mu} \right) \right] + \mu_2 y - \varphi_0 \right) \tag{A-8}$$

(A-1), (A-2), (A-7) and (A-8) are the expressions used to calculate the initial conditions at $y = 0$, when μ_1 and μ_2 are solved from Pierce's equation (A-6).

From (12c), the particular solution of the backward wave at small-signal is found to be

$$B(z, t) = -2\epsilon \left| \frac{-2jC(1 + jC\mu)(\mu + jb)}{2j - c\mu + icb} \right| e^{\mu_1 y} \cos\left[\arg\left[\frac{-2jC(1 + jC\mu)(\mu + jb)}{2j - C\mu + jcb} \right] + \mu_2 y - \varphi_0 \right]$$

which agrees with Pierce's analysis.[17]

REFERENCES

1. J. R. Pierce, Traveling-Wave Tubes, D. Van Nostrand Co., N.Y., 1950, p. 160.
2. R. L. Hess, Some Results in the Large-Signal Analysis of Traveling-Wave Tubes, Technical Report Series No. 60, Issue No. 131, Electronic Research Laboratory, University of California, Berkeley, California.
3. C. K. Birdsall, unpublished work.
4. J. J. Caldwell, unpublished work.
5. P. Parzen, Nonlinear Effects in Traveling-Wave Amplifiers, TR/AF-4, Radiation Laboratory, The Johns Hopkins University, April 27, 1954.
6. A. Kiel and P. Parzen, Non-linear Wave Propagation in Traveling-Wave Amplifiers, TR/AF-13, Radiation Laboratory, The Johns Hopkins University, March, 1955.
7. A. Nordsieck, Theory of the Large-Signal Behavior of Traveling-Wave Amplifiers, Proc. I.R.E., **41**, pp. 630–637, May, 1953.
8. H. C. Poulter, Large Signal Theory of the Traveling-Wave Tube, Tech. Report No. 73, Electronics Research Laboratory, Stanford University, California, Jan., 1954.
9. P. K. Tien, L. R. Walker and V. M. Wolontis, A Large Signal Theory of Traveling-Wave Amplifiers, Proc. I.R.E., **43**, pp. 260–277 March, 1955.
10. J. E. Rowe, A Large Signal Analysis of the Traveling-Wave Amplifier, Tech. Report No. 19, Electron Tube Laboratory, University of Michigan, Ann Arbor, April, 1955.
11. P. K. Tien and L. R. Walker, Correspondence Section, Proc. I.R.E., **43**, p. 1007, Aug., 1955.
12. Nordsieck, op. cit., equation (1).
13. L. Brillouin, The Traveling-Wave Tube (Discussion of Waves for Large Amplitudes), J. Appl. Phys., **20**, p. 1197, Dec., 1949.
14. Pierce, op. cit., p. 9.
15. Nordsieck, op. cit., equation (4).
16. Pierce, op. cit., equation (7.13).
17. J. R. Pierce, Theory of Traveling-Wave Tube, Appendix A, Proc. I.R.E. **35**, p. 121, Feb., 1947.

A Detailed Analysis of Beam Formation with Electron Guns of the Pierce Type

By W. E. DANIELSON, J. L. ROSENFELD,* and J. A. SALOOM

(Manuscript received November 10, 1955)

The theory of Cutler and Hines is extended in this paper to permit an analysis of beam-spreading in electron guns of high convergence. A lens correction for the finite size of the anode aperture is also included. The Cutler and Hines theory was not applicable to cases where the effects of thermal velocities are large compared with those of space charge and it did not include a lens correction. Gun design charts are presented which include all of these effects. These charts may be conveniently used in choosing design parameters to produce a prescribed beam.

CONTENTS

---

* Mr. Rosenfeld participated in this work while on assignment to the Laboratories as part of the M.I.T. Cooperative Program.

GLOSSARY OF SYMBOLS

$A_{1,2}$	anode designations
B, C	anode potentials
$C_{1,2}$	functions used in evaluating σ_+'
dA	increment of area
dl, dz	increments of length
e	electronic charge, base of natural logarithms
E_n	electric field normal to electron path
F	modified focal length of the anode lens
F_D	focal length of the anode lens as given by Davisson[4]
F_n	force acting normal to an electronic path
$F_{r,\sigma}$	fraction of the total current which would flow through a circle of radius r, σ
I, I_D	total beam current
I_r	beam current within a radius, r, of the center
J	current density
k	Boltzman's constant
K	a quantity proportional to gun perveance
m	electronic mass
P	gun perveance
$P(r)$	probability that a thermal electron has a radial position between r and $r + dr$
r	radial distance from beam axis
$r_{a,c}$	anode, cathode radii
r_e	distance from beam axis to path of an electron emitted with zero velocity at the edge of the cathode
r_{95}	radius of circle through which 95% of the beam current would pass
$\bar{r}$	distance from center of curvature of cathode; hence, $\bar{r}_c$ is the cathode radius of curvature and $(\bar{r}_c - \bar{r}_a)$ is the distance from cathode to anode
r_{e^+}'	slope of edge nonthermal electron path on drift side of enode lens
r_{e^-}'	slope of edge nonthermal electron path on gun side of anode lens
R	a dummy integration variable
t	time
T	cathode temperature in degrees K
u	longitudinal electron velocity
$v_{c,x,y}$	transverse electron velocities
$V, V_{a,f,x}$	beam voltages with cathode taken as ground

$V(\bar{r}, r)$, $V_c(\bar{r}, r)$, etc.	potential distributions used in the anode lens study
V'	voltage gradient
z	distance along the beam from the anode lens
$z_{\min}$	distance to the point where r_{95} is a minimum
$(-\alpha)$	Langmuir potential parameter for spherical cathode-anode gun geometry
γ	slope of an electron's path after coming into a space charge free region just beyond the anode lens
Γ	the factor which divides F_D to give the modified anode focal length
δ	dimensionless radius parameter
ϵ_0	dielectric constant of free space
ζ	dimensionless voltage parameter
θ	slope of an electron's path in the gun region
η	charge to mass ratio for the electron
μ	normalized radial position in a beam
σ	the radial position of an electron which left the cathode center with "normal" transverse velocity
σ_+'	slope of σ-electron on drift side of anode lens
σ_-'	slope of σ-electron on gun side of anode lens
ψ	electric flux

1. INTRODUCTION

During the past few years there have been several additions to the family of microwave tubes requiring long electron beams of small diameter and high current density. Due to the limited electron current which can be drawn from unit area of a cathode surface with some assurance of long cathode operating life, high density electron beams have been produced largely through the use of convergent electron guns which increase markedly the current density in the beam over that at the cathode surface.

An elegant approach to the design of convergent electron guns was provided by J. R. Pierce[1] in 1940. Electron guns designed by this method are known as *Pierce guns* and have found extensive use in the production of long, high density beams for microwave tubes.

More recent studies, reviewed in Section 2, have led to a better understanding of the influence on the electron beam of (a) the finite velocities with which electrons are emitted from the cathode surface, and (b) the defocusing electric fields associated with the transition from the accelerating region of the gun to the drift region beyond. Although these two effects have heretofore been treated separately, it is in many cases

necessary to produce electron beams under circumstances where both effects are important and so must be dealt with simultaneously and more precisely than has until now been possible. It is the purpose of this paper to provide a simple design procedure for typical Pierce guns which includes both effects. Satisfactory agreement has been obtained between measured beam contours and those predicted for several guns having perveances (i.e., ratios of beam current to the $\frac{3}{2}$ power of the anode voltage) from 0.07×10^{-6} to 0.7×10^{-6} amp $(\text{volt})^{-3/2}$.

2. PRESENT STATUS OF GUN DESIGN — LIMITATIONS

Gun design techniques of the type originally suggested by J. R. Pierce were enlarged in papers by Samuel[2] and by Field[3] in 1945 and 1946. Samuel's work did not consider the effect of thermal velocities on beam shape and, although Field pointed out the importance of thermal velocities in limiting the theoretically attainable current density, no method for predicting beam size and shape by including thermal effects was suggested. The problem of the divergent effect of the anode lens was treated in terms of the Davisson[4] electrostatic lens formula, and no corrections were applied.*

More recently, Cutler and Hines[6] and also Cutler and Saloom[7] have presented theoretical and experimental work which shows the pronounced effects of the thermal velocity distribution on the size and shape of beams produced by Pierce guns. Cutler and Saloom also point to the critical role of the beam-forming electrode in minimizing beam distortion due to improper fields in the region where the cathode and the beam-forming electrode would ideally meet. With regard to the anode lens effect, these authors also show experimental data which strongly suggest a more divergent lens than given by the Davisson formula. The Hines and Cutler thermal velocity calculations have been used[6, 7] to predict departures in current density from that which should prevail in ideal beams where thermal electrons are absent. Their theory is limited, however, by the assumption that the beam-spreading caused by thermal velocities is small compared to the nominal beam size.

In reviewing the various successes of the above mentioned papers in affording valuable tools for electron beam design, it appeared to the present authors that significant improvement could be made, in two respects, by extensions of existing theories. First, a more thorough in-

* It is in fact erroneously stated in Reference 5 that the lens action of an actual structure must be somewhat weaker than predicted by the Davisson formula so that the beam on leaving the anode hole is more convergent than would be calculated by the Davisson method. This question is discussed further in Section 3.

vestigation of the anode lens effect was called for; and second, there was a need to extend thermal velocity calculations to include cases where the percentage increase in beam size due to thermal electrons was as large as 100 per cent or 200 per cent. Some suggestions toward meeting this second need have been included in a paper by M. E. Hines.[8] They have been applied to two-dimensional beams by R. L. Schrag.[9] The particular assumptions and methods of the present paper as applied to the two needs cited above are somewhat different from those of References 8 and 9, and are fully treated in the sections which follow.

3. TREATMENT OF THE ANODE LENS PROBLEM

Using thermal velocity calculations of the type made in Reference 6, it can easily be shown that at the anode plane of a typical moderate perveance Pierce type electron gun, the average spread in radial position of those electrons which originate from the same point of the cathode is several times smaller than the beam diameter. For guns of this type, then, we may look for the effect of the anode aperture on an electron beam for the idealized case in which thermal velocities are absent and confidently apply the correction to the anode lens formula so obtained to the case of a real beam.

Several authors have been concerned with the diverging effect of a hole in an accelerating electrode where the field drops to zero in the space beyond,[10] but these treatments do not include space charge effects except as given by the Davisson formula for the focal length, F_D, of the lens:

$$F_D = -\frac{4V}{V'} \tag{1}$$

where V' would be the magnitude of the electric field at the aperture if it were gridded, and V would be the voltage there.

In attempting to describe the effect of the anode hole with more accuracy than (1) affords, we have combined analytical methods with electrolytic tank measurements in two rather different ways. The first method to be given is more rigorous than the second, but a modification of the second method is much easier to use and gives essentially the same result.

A. Superposition Approach to the Anode Lens Problem

Special techniques are required for finding electron trajectories in a space charge limited Pierce gun having a non-gridded anode. M. E.

Hines has suggested* that a fairly accurate description of the potential distribution in such guns can be obtained by a superposition method as follows:

By the usual tank methods, find suitable beam forming electrode and anode shapes for conical space charge limited flow in a diode having cathode and anode radii of curvature given by $\bar{r}_c$ and $\bar{r}_{a1}$, respectively, as shown in Fig. 1(a). Using the electrolytic tank with an insulator along the line which represents the beam edge, trace out an equipotential which intersects the insulator at a distance $\bar{r}_{a2}$ from the cathode center of curvature. Let the cathode be at ground potential and let the voltage on anode A_1 be called B. Suppose, now, that we are interested in electron trajectories in a non-gridded gun where the edge of the anode hole is a distance $\bar{r}_{a2}$ from the center of curvature of the cathode. Let the voltage, C, for this anode be chosen the same as the value of the equipotential traced out above for the case of cathode at ground potential and A_1 at potential B. If we consider the space charge limited flow from a cathode which is followed by the apertured anode, A_2, and the full anode, A_1, at potentials C and B, respectively, it is clear that a conical flow of the type which would exist between concentric spheres will result. The flow for such cases was treated by Langmuir,[14] and the associated potentials are commonly called the "Langmuir potentials."

If we operate both A_1 and A_2 at potential C, however, the electrons will pass through the aperture in anode A_2 into a nearly field-free region. If the distance, $\bar{r}_{a2} - \bar{r}_{a1}$, from A_2 to A_1 is greater than the diameter of the aperture in A_2, the flow will depend very little on the shape of A_1 and the electron trajectories and associated equipotentials will be of the type we wish to consider except in a small region near A_1. We will shortly make use of the fact that the space charge between cathode and A_2 is not changed much when the voltage on A_1 is changed from B to C, but first we will define a set of potential functions which will be needed.

In order to obtain the potential at arbitrary points in any axially symmetric gun when space charge is not neglected, we may superpose potential solutions to 3 separate problems where, in each case, the boundary condition that each electrode be an equipotential is satisfied. We will follow the usual notation in using $\bar{r}$ for the distance of a general point from the cathode center of curvature, and r for its radial distance from the axis of symmetry. Let $V_a(\bar{r}, r)$, $V_b(\bar{r}, r)$ and $V_{sc}(\bar{r}, r)$ be the three potential solutions where: (1) $V_a(\bar{r}, r)$ is the solution for the case of no space charge with A_1 and cathode at zero potential and A_2 at potential C, (2) $V_b(\bar{r}, r)$ is the solution for the case of no space charge with A_2

* Verbal disclosure.

and cathode at zero potential and A_1 at potential B, and (3) $V_{sc}(\bar{r}, r)$ is the solution when space charge is present but when A_1, A_2, and cathode are all grounded.

If the configuration of charge which contributes to $V_{sc}(\bar{r}, r)$ is that corresponding to ideal Pierce type flow, then we can use the principle of superposition to give the Langmuir potential, $V_L(\bar{r}, r)$:

$$V_L(\bar{r}, r) = V_a(\bar{r}, r) + V_b(\bar{r}, r) + V_{sc}(\bar{r}, r) \tag{2}$$

Furthermore, the potential configuration for the case where A_1 and A_2 are at potentical C can be written

$$V = V_a + \frac{C}{B} V_b + V_{(sc)'} \tag{3}$$

where the functional notation has been dropped and $V_{(sc)'}$ is the po tential due to the new space charge when A_1 and A_2 are grounded. We are now ready to use the fact that $V_{(sc)'}$ may be well approximated by V_{sc} which is easily obtained from (2). This substitution may be justified by noting that the space charge distribution in a gun using a voltage C for A_1 does not differ significantly from the corresponding distribution when A_1 is at voltage B except in the region near and beyond A_2 where the charge density is small anyway (because of the high electron velocities there). Substituting V_{sc} as given by (2) for $V_{(sc)'}$ in (3) then gives

$$V \approx V_L - \left(1 - \frac{C}{B}\right) V_b \tag{4}$$

We have thus obtained an expression, (4), for the potential at an arbi-

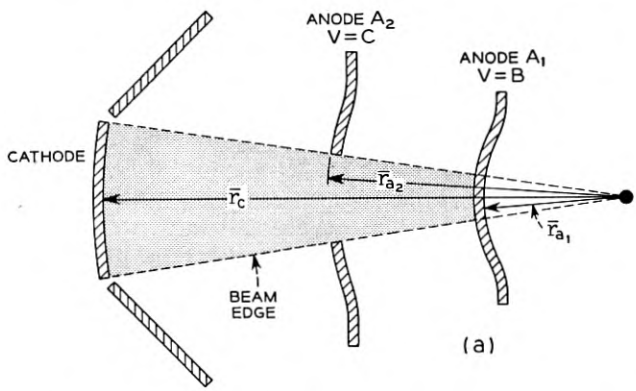


Fig. 1(a) — Electrode configuration for anode lens evaluation in Section 3A.

trary point in our gun in terms of the well known solution for space charge limited flow between two concentric spheres, V_L, and a potential distribution, V_b, which does not depend on space charge and can therefore be obtained in the electrolytic tank. Once the potential distribution is found, electron trajectories may be calculated, and an equivalent lens system found. Equation (4) is used in this way in Part C as one basis for estimating a correction to the Davisson equation. (It will be noted that (4) predicts a small but finite negative field at the cathode. This is because the space charge density associated with V_{sc} is slightly greater near the cathode than that associated with $V_{(sc)'}$, and it is this latter space charge which will make the field zero at the cathode under real space charge limited operation. Equation (4), as applied in Part C of this section, is used to give the voltage as a function of position at all points except near the cathode where the voltage curves are extended smoothly to make the field at the cathode vanish.)

B. Use of a False Cathode in Treating the Anode Lens Problem

Before evaluating the lens effect by use of (4), it will be useful to develop another approach which is a little simpler. The evaluation of the lens effect predicted by both methods will then be pursued in Part C where the separate results are compared.

In Part A we noted that no serious error is made in neglecting the difference between the two space charge configurations considered there because these differences were mainly in the very low space charge region near and beyond A_2. It similarly follows that we can, with only a small decrease in accuracy, ignore the space charge in the region near and beyond A_2 so long as we properly account for the effect of the high space charge regions closer to the cathode. To place the foregoing observations on a more quantitative basis, we may graph the Langmuir potential (for space charge limited flow between concentric spheres) versus the distance from cathode toward anode, and then superpose a plot of the potential from LaPlace's equation (concentric spheres; no space charge) which will have the same value and slope at the anode. The LaPlace curve will depart significantly from the Langmuir in the region of the cathode, but will adequately represent it farther out.[11] Our experience has shown that the representation is "adequate" until the difference between the two potentials exceeds about 2 per cent of the anode voltage. Then, since space charge is not important in the region near the anode for the case of a gridded Pierce gun, corresponding to space charge limited flow between concentric spheres, it can be expected to be similarly unimportant for cases where the grid is replaced by an aperture. Let us

therefore consider a case where electrons are emitted perpendicularly and with finite velocity from what would be an appropriate spherical equipotential between cathode and anode in a Pierce type gun. So long as (a) there is good agreement between the LaPlace and Langmuir curves at this artificial cathode and (b) the distance from this artificial cathode to the anode hole is somewhat greater than the hole diameter, we will find that the divergent effect of the anode hole will be very nearly the same in this concocted space charge free case as in the actual case where space charge is present. (The quantitative support for this last statement comes largely from the agreement between calculations based on this method and calculations by method A.) The electrode configuration is shown in Fig. 1(b), and the potential distribution in this space charge free anode region can now be easily obtained in the electrolytic tank. This potential distribution will be used in the next section to provide a second basis for estimating a correction to the Davisson equation.

C. Calculation of Anode Lens Strength by the Two Methods

The Davisson equation, (1), may be derived by assuming that none of the electric field lines which originate on charges in the cathode-anode region leave the beam before reaching the ideal anode plane where the voltage is V, and that all of these field lines leave the beam symmetrically and radially in the immediate neighborhood of the anode. Electrons are thus considered to travel in a straight line from cathode to anode, and then to receive a sudden radial impulse as they cross radially diverging electric field lines at the anode plane. A discontinuous change in

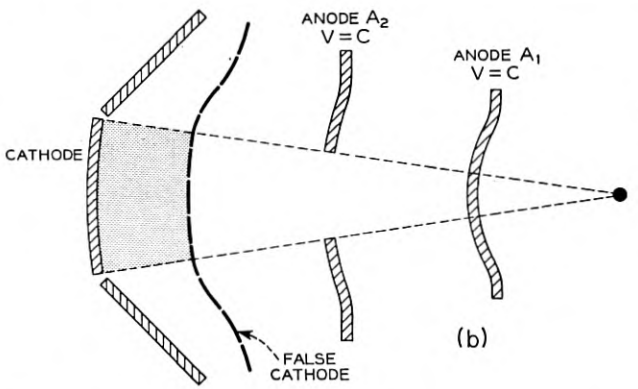


Fig. 1(b) — The introduction of a false cathode at the appropriate potential allows the effect of space charge on the potential near the anode hole to be satisfactorily approximated as discussed in Section 3*B*.

slope is therefore produced as is common to all thin lens approximations. The diverging effect of electric field lines which originate on charges which have passed the anode plane is then normally accounted for by the universal beam spread curve.[12] In our attempt to evaluate the lens effect more accurately, we will still depend upon using the universal beam spread curve in the region following the lens and on treating the equivalent anode lens as thin. Consequently our improved accuracy must come from a mathematical treatment which allows the electric field lines originating in the cathode-anode region to leave the beam gradually, rather than a treatment where all of these flux lines leave the beam at the anode plane. In practice the measured perveances, $P(= I/V^{3/2})$, of active guns of the type considered here have averaged within 1 or 2 per cent of those predicted for corresponding gridded Pierce guns. Therefore the total space charge between cathode and anode is much the same with and without the use of a grid, even though the charge distribution is not the same in the two cases. The total flux which must leave our beam is therefore the same as that which will leave the corresponding idealized beam and we may write

$$\psi = \int E_n \, dA = \pi r_a^2 V_{\text{ideal}}' \tag{5}$$

where E_n is the electric field normal to the edge of the beam, $r_a = r_c(\bar{r}_a/\bar{r}_c)$ is the beam radius at the anode lens, and V_{ideal}' is the magnitude of the field at the corresponding gridded Pierce gun anode.

To find the appropriate thin lens focal length we will now find the total integrated transverse impulse which would be given to an electron which follows a straight-line path on both sides of the lens (see Fig. 2), and we will equate this impulse to $m\Delta u$ where Δu is the transverse velocity given to the electron as it passes through the equivalent thin lens. In this connection we will restrict our attention to paraxial electrons and evaluate the transverse electric fields from (4) and from the tank plot outlined in Section B, respectively. The total transverse impulse experienced by an electron can be written

$$\int_{\text{Path}} F_n \, dt = e \int_{\text{Path}} \frac{E_n}{u} \, dl \tag{6}$$

where u is the velocity along the path and F_n is the force normal to the path.

We will usually find that the correction to (1) is less than about 20 per cent. It will therefore be worthwhile to put (6) in a form which in effect allows us to calculate *deviations* from F_D as given by (1) instead

of deriving a completely new expression for F. In accomplishing this purpose, it will be helpful to define a dimensionless function of radius, δ, by

$$\frac{r_a}{r} = 1 + \delta, \tag{7a}$$

and a dimensionless function of voltage, ζ, by

$$\sqrt{\frac{V_x}{V}} = 1 + \zeta, \tag{7b}$$

where r_a is the radius at the anode lens when the lens is considered thin, and V_x is a constant voltage to be specified later. (Note that the quantities δ and ζ are not necessarily small compared to 1.) Using $u = \sqrt{2\eta V}$, and substituting for $\sqrt{V}$ from (7b) we obtain

$$e\int \frac{E_n\, dl}{u} = \frac{4}{r_a\sqrt{2\eta V_x}}\int E_n r(1 + \zeta + \delta + \zeta\delta)\, dl \tag{8}$$

where use has also been made of (7a) in the form $1 \equiv r(1 + \delta)/r_a$. Now, as outlined above, we equate this impulse to $m\Delta u$, and we obtain

$$\Delta u = \frac{e/m}{r_a\sqrt{2\eta V_x}}\left(\int E_n r\, dl + \int E_n r(\zeta + \delta + \zeta\delta)\, dl\right) \tag{9}$$

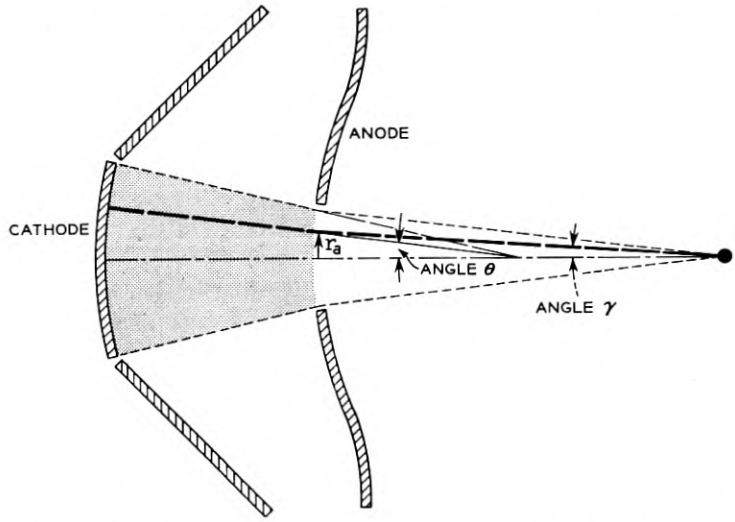


Fig. 2 — The heavy line represents an electron's path when the effect of the anode hole may be represented by a thin lens, and when space charge forces are absent in the region following the anode aperture. For paraxial electrons, the (negative) focal length is related to the indicated angles by ($\gamma = \theta + r_a/F$).

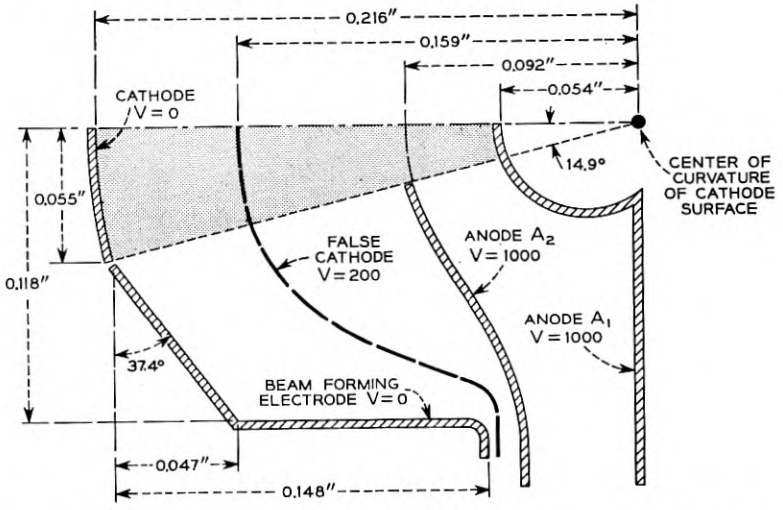


Fig. 3 — The gun parameters used in Section 3*C* for comparing two methods of evaluating the effect of the anode lens.

The first integral can be obtained from (5); hence, if we are able to choose V_x so that the second integral vanishes, we may write:

$$\Delta u = \frac{\eta}{r_a\sqrt{2\eta V_x}}\left(\frac{r_a^2 V_{\text{ideal}}'}{2}\right)$$

The reciprocal of the thin lens focal length is therefore

$$\frac{1}{F} = -\frac{\Delta u}{r_a u_f} = -\frac{V'}{4\sqrt{V_x V_f}} \tag{10}$$

where u_f and V_f are the final velocity and voltage of the electron after it leaves the lens region.

The real task, then, is to use the potential distribution in the gun as obtained by the methods of Part *A* or Part *B* above to find the value of V_x which causes the last integral in (9) to vanish: To compare the two focal lengths obtained by the methods of Part *A* and *B* respectively, a specific tank design of the type indicated in Fig. 1 was carried out. The relevant gun parameters are indicated in Fig. 3. Approximate voltages on and near the beam axis were obtained as indicated in Parts *A* and *B*, above, with the exception that in the superposition method, A, special techniques were used to subtract the effect of the space charge lying in the post-anode region (because the effect of this space charge is accounted for separately as a divergent force in the drift region*). From these data,

* See Section 4B.

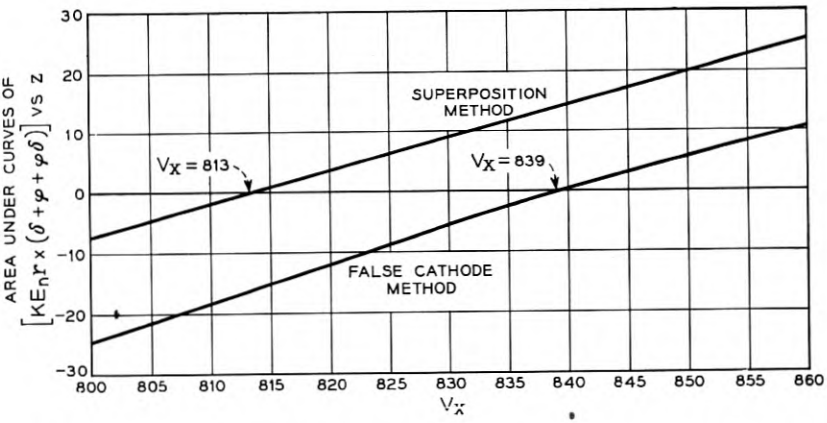


Fig. 4 — Curves for finding the value of V_x to be used in equation (10) for the set of gun parameters of Fig. 3.

both the direction and magnitude of the total electric field near the beam axis were (with much labor) determined. Once these data had been obtained, a trial value was selected for V_x, and the corresponding focal length was calculated by (10). This enabled the electron's path through the associated thin lens to be specified so that, at this point in the procedure, both r and V were known functions of ℓ, and the quantities δ and ζ were then obtained as functions of ℓ from (7). Finally the second integral in (9) was evaluated for the particular V_x chosen, and then the process was repeated for other values of V_x. Fig. 4 shows curves whose ordinates are proportional to this second integral and whose abscissae are trial values for V_x. As noted above, the appropriate value for V_x is that value which makes the ordinate vanish, so that we obtain $V_x = 813$ and 839 for methods A and B, respectively. The percentage difference in the focal lengths obtained by the two methods is thus only 1.6 per cent, and the reasonableness of making calculations as outlined in Part *B* is thus put on a more quantitative basis.

Even calculations based on the method of Part *B* are tedious, and we naturally look for simpler methods of estimating the lens effect. In this connection we have found that V_x is usually well approximated by the value of the potential at the point of intersection between the beam axis and the ideal anode sphere. The specific values of the potential at this point as obtained by the methods of Parts *A* and *B* were 814 and 827, respectively. It will be noted that these values agree remarkably well with the values obtained above. Furthermore, very little extra effort is required to obtain the potential at this intersection in the false cathode case:

Electrolytic tank measurements are normally made in the cathode-anode region to give the potential variation along the outside edge of the electron beam (for comparison with the Langmuir potential); hence, by tracing out a suitable equipotential line, the shape of the false cathode can easily be obtained. With the false cathode in place and at the proper potential, the approximate value for V_x is then obtained by a direct tank measurement of the potential at an axial point whose distance from the true cathode center is $(\bar{r}_c - \bar{r}_a)$ as outlined above. Although finite electron emission velocities typically do not much influence the trajectory of an electron at the anode, they do nevertheless significantly alter the beam in the region beyond. It is in this affected region where experimental data can be conveniently taken. We must, therefore, postpone a comparison of lens theory with experiment until the effect of thermal velocities has been treated. At that time theoretical predictions combining the effects of both thermal velocities and the anode lens can be made and compared with experiment. Such a comparison is made in Section 6.

4. TREATMENT OF BEAM SPREADING, INCLUDING THE EFFECT OF THERMAL ELECTRONS

In Section 2 the desirability of having an approach to the thermal spreading of a beam which would be applicable under a wide variety of conditions was stressed. In particular, there was a need to extend thermal velocity calculations to include the effects of thermal velocities even when electrons with high average transverse velocities perturb the beam size by as much as 100 or 200 per cent. Furthermore, a realistic mathematical description which would allow electrons to cross the axis seemed essential. The method described below is intended adequately to answer these requirements.

A. The Gun Region

The Hines-Cutler[6] method of including the effect of thermal velocities on beam size and shape leads one to conclude that, for usual anode voltages and gun perveance, the beam density profile in the plane of the anode hole is not appreciably altered by thermal velocities of emission. (This statement will be verified and put on a more quantitative basis below.) Under these conditions, the beam at the anode is adequately described by the Hines-Cutler treatment. We will therefore find it convenient to adopt their notation where possible, and it will be worthwhile to review their approach to the thermal problem.

It is assumed that electrons are emitted from the cathode of a thermionic gun with a Maxwellian distribution of transverse velocities

$$dJ_c = J_c \frac{m}{2\pi kT} e^{-(m/2kT)(v_x^2+v_y^2)}\, dv_x\, dv_y \tag{11}$$

where J_c is the cathode current density in the z direction, T is the cathode temperature, and v_x and v_y are transverse velocities. The number of electrons emitted per second with radially directed voltages between V and $V + dV$ is then

$$dJ_c = J_c e^{-(Ve/kT)}\, d\left(\frac{Ve}{kT}\right) \tag{12}$$

Now in the accelerating region of an ideal Pierce gun (and more generally in any beam exhibiting laminar flow and having constant current density over its cross section) the electric field component perpendicular to the axis of symmetry must vary linearly with radius. Consequently Hines and Cutler measure radial position in the electron beam as a fraction, μ, of the outer beam radius (r_e) at the same longitudinal position,

$$r = \mu r_e \tag{13}$$

The laminar flow assumption for constant current densities and small beam angles implies a radius of curvature for laminar electrons which also varies linearly with radius at any given cross section so that

$$\frac{d^2r}{dt^2} = \mu \frac{d^2r_e}{dt^2} \tag{14}$$

Substituting for r from (13), (14) becomes

$$\frac{d^2\mu}{dt^2} + \left(\frac{2}{r_e}\frac{dr_e}{dt}\right)\frac{d\mu}{dt} = 0 \tag{15}$$

where r_e and dr/dt can be easily obtained from the ideal Langmuir solution. Since the equation is linear in μ, we are assured that the radial position of a non-ideal electron that is emitted with finite transverse velocity from the cathode center (where $\mu = 0$) will, at any axial point, be proportional to $d\mu/dt$ at the cathode.

Let us now define a quantity "σ" such that $\mu = \sigma/r_e$ is the solution to (15) with the boundary conditions $\mu_c = 0$ and

$$\left(\frac{d\mu}{dt}\right)_c = \frac{1}{r_c}\sqrt{\frac{kT}{m}}$$

where the subscript c denotes evaluation at the cathode surface, k is

Boltzman's constant, T is the cathode temperature in degrees Kelvin, and m is mass of the electron. For the case $\mu_c = 0$, but with arbitrary initial transverse velocity, we will then have

$$\mu = \frac{\sigma}{r_c} \frac{\left(\frac{d\mu}{dt}\right)_c}{\frac{1}{r_c}\sqrt{\frac{kT}{m}}} \tag{16}$$

Hence we can express σ in terms of the thermal electron's radial position (r), and its initial transverse velocity, v_c,

$$\sigma = \frac{\mu r_c \sqrt{\frac{kT}{m}}}{\left(\frac{d(\mu r_c)}{dt}\right)} \equiv \frac{r\sqrt{\frac{kT}{m}}}{v_c} \tag{17}$$

The quantity σ can now be related to the radial spread of thermal electrons (emitted from a given point on the cathode) with respect to an electron with no initial velocity: By (11) we see that the number of electrons leaving the cathode with $d\mu/dt = v_c/r_c$ is proportional to $v_c \exp -v_c^2 m/2kT$. Suppose many experiments were conducted where all electrons except one at the cathode center had zero emission velocity, and suppose the number of times the initial transverse velocity of the single thermal electron were chosen as v_c, is proportional to $v_c \exp -v_c^2 m/2kT$. Then the probability, $P(r)$, that the thermal electron would have a radial position between r and $r + dr$ when it arrived at the transverse plane of interest would be proportional to $v_c \exp -v_c^2(m/2kT)$. Here v_c is the proper transverse velocity to cause arrival at radius r, and by (17) we have

$$v_c = \frac{r}{\sigma}\sqrt{\frac{kT}{m}}$$

so that the probability becomes

$$P(r) = J_c e^{-(r^2/2\sigma^2)}\, d\left(\frac{r^2}{2\sigma^2}\right) \tag{18}$$

We therefore identify σ with the standard deviation in a normal or Gaussian distribution of points in two dimensions. At the real cathode, thermal electrons are simultaneously being emitted from the cathode surface with a range of transverse velocities. However, if σ as defined above is small in comparison with r_c, the forces experienced by a thermal electron when other thermal electrons are present will be very nearly

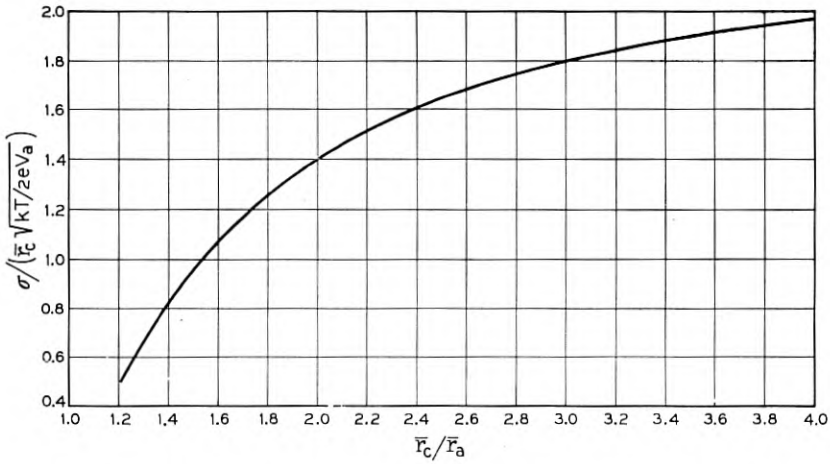


Fig. 5 — Curves useful in finding the transverse displacement of electron trajectories at the anode of Pierce-type guns.

the same as the forces involved in the equations above. Thus if $\sigma \ll r_e$, (18) may be taken as the distribution, in a transverse plane, of those electrons which were simultaneously emitted at the cathode center. Furthermore, the nature of the Pierce gun region is such that electrons emitted from any other point on the cathode will be similarly distributed with respect to the path of an electron emitted from this other point with zero transverse velocity (so long as they stay within the confines of the ideal beam). Hines and Cutler have integrated (15) with $\mu_c = 0$ and $(d\mu/dl)_c = 1$ to give $\sigma/(\bar{r}_c\sqrt{kT/2eV_a})$ at the anode as a function of $\bar{r}_c/\bar{r}_a$. This relationship is included here in graphical form as Fig. 5.

For a large class of magnetically shielded Pierce-type electron guns, including all that are now used in our traveling wave tubes, r_e/σ at the anode is indeed found to be greater than 5 (in most cases, greater than 10) so that evaluation of σ at the anode of such guns can be made with considerable accuracy by the methods outlined above. One source of error lies in the assumption that electrons which are emitted from a point at the cathode edge become normally distributed about the corresponding non-thermal (no transverse velocity of emission) electron's path, and with the same standard deviation as calculated for electrons from the cathode center. In the gun region where r_e/σ tends to be large this difference between representative σ-values for the peripheral and central parts of the beam is unimportant, but it must be re-examined in the drift region following the anode.

We have already investigated the region of the anode hole in some detail in Section 3 and have found it worth while to modify the ideal Davisson expression for focal length of an equivalent anode lens. In particular, let us define a quantity Γ by

$$F = \text{focal length} = F_D/\Gamma \tag{19}$$

where F_D is the Davisson focal length. Thus Γ represents a corrective factor to be applied to F_D to give a more accurate value for the focal length. In so far as any thin lens is capable of describing the effects of diverging fields in the anode region, we may then use the appropriate optical formulas to transfer our knowledge of the electron trajectories (calculated in the anode region as outlined above) to the start of the drift region. In particular,

$$\left(\frac{dr}{dz}\right)_2 = \left(\frac{dr}{dz}\right)_1 - \frac{r}{F} \tag{20}$$

where $(dr/dz)_1$ and $(dr/dz)_2$ are the slopes of the path just before and just after the lens, and r is the distance from the axis to the point where the ideal path crosses the lens plane.

B. The Drift Region

Although r_e/σ was found to be large at the anode plane for most guns of interest, this ratio often shrinks to 1 or less at an axial distance of only a few beam diameters from the lens. Therefore, the assumption that electron trajectories may be found by using the space charge forces which would exist in the absence of thermal velocities of emission (i.e., forces consistant with the universal beam spread curve) may lead to very appreciable error. For example, if equal normal (Gaussian) distributions of points about a central point are superposed so that the central points are equally dense throughout a circle of radius r_e, and if the standard deviation for each of the normal distributions is $\sigma = r_e$, the relative density of points in the center of the circle is only about 39 per cent of what it would be with $\sigma < (r_e/5)$.

In order to minimize errors of this type we have modified the Hines-Cutler treatment of the drift space in two ways: (1) The forces influencing the trajectories of the non-thermal electrons are calculated from a progressive estimation of the actual space charge configuration as modified by the presence of thermal electrons. (2) Some account is taken of the fact that, as the space charge density in the beam becomes less uniform as a function of radius, the spread of electrons near the center of the beam increases more rapidly than does the corresponding spread

farther out. Since item (1) is influenced by item (2), the specific assumptions involved in the latter case will be treated first.

When current density is uniform across the beam and its cross section changes slowly with distance, considerations of the type outlined above for the gun region show that those thermal electrons which remain within the beam will continue to have a Gaussian distribution with respect to a non-thermal electron emitted from the same cathode point. When current density is not uniform over the cross section, we would still like to preserve the mathematical simplicity of obtaining the current density as a function of beam radius merely by superposing Gaussian distributions which can be associated with each non-thermal electron. To lessen the error involved in this simplified approach, we will arrive at a value for the standard deviation, σ (which specifies the Gaussian distribution), in a rather special way. In particular, σ at any axial position, z, will be taken as the radial coordinate of an electron emitted from the center of the cathode with a transverse velocity of emission given by,

$$v_c = \sqrt{\frac{kT}{m}} \tag{21}$$

It is clear from (17) that for such an electron, $r = \sigma$ in the gun region. From (18), the fraction of the electrons from a common point on the cathode which will have $r \leqq \sigma$ in the gun region is

$$\text{fraction} = \int_0^{\sigma} e^{-(r^2/2\sigma^2)}\, d\,\frac{r^2}{2\sigma^2} = 1 - e^{-1/2} = 0.393 \tag{22}$$

If r_e denotes the radial position of the outermost non-thermal electron and if $\sigma > r_e$, the "σ-electron" will be moving in a region where the space charge density is significantly lower than at the axis. We could, of course, have followed the path of an electron with initial velocity equal to say 0.1 or 10 times that given in (21) and called the corresponding radius 0.1σ or 10σ. The reason for preferring (21) is that about 0.4 or nearly half of the thermal electrons emitted from a common cathode point will have wandered a distance less than σ from the path of a non-thermal electron emitted from the same cathode point, while other thermal electrons will have wandered farther from this path; consequently, the current density in the region of the σ-electron is expected to be a reasonable average on which beam spreading due to thermal velocities may be based. With this understanding of how σ is to be calculated, we can proceed to the calculation of non-thermal electron trajectories as suggested in item (1).

The non-thermal paths remain essentially laminar, and with r_e denoting the radial coordinate of the outermost non-thermal electron, we will make little error in assuming that the current density of non-thermal electrons is constant for $r < r_e$. Consequently, if equal numbers of thermal electrons are assumed to be normally distributed about the corresponding non-thermal paths, the longitudinal current density as a function of radius can be found in a straightforward way[13] by using (18). The result is

$$\frac{J_r}{J_D} = e^{-(r^2/2\sigma^2)} \int_0^{r_e/\sigma} \frac{R}{\sigma} e^{-(R^2/2\sigma^2)} I_0\left(\frac{rR}{\sigma^2}\right) d\left(\frac{R}{\sigma}\right) \tag{23}$$

where I_0 is the zero order modified Bessel function and the total current is $I_D = \pi r_e^2 J_D$. Equation (23) was integrated to give a plot of J_r/J_D versus r/σ, with r_e/σ as a parameter and is given as Fig. 6 in Reference 6. It is reproduced here as Fig. 6. Since the only forces acting on electrons in the drift region are due to space charge, we may write the equation of motion as

$$\frac{d^2r}{dt^2} = \eta E_r \tag{24}$$

where E_r is the radial electrical field acting on an electron with radial coordinate r. Since the beam is long and narrow, all electric lines of force may be considered to leave the beam radially so that E_r is simply obtained from Gauss' law. Equation (24) therefore becomes

$$\begin{aligned} \frac{d^2r}{dt^2} = \frac{\eta}{2\pi\epsilon_0 r} \int_0^r 2\pi\rho\, dr = \frac{\eta}{2\pi\epsilon_0 r} \int_0^r \frac{J(r)}{\sqrt{2\eta V_a}} 2\pi r\, dr \\ = \frac{\sqrt{\eta/(2V_a)}}{2\pi\epsilon_0 r} \int_0^r J(r) 2\pi r\, dr \end{aligned} \tag{25}$$

From (23) we note that the fraction of the total current within any radius depends only on r_e/σ and r/σ:

$$\begin{aligned} \frac{I_r}{I_D} = \frac{\int_0^r J(r)2\pi r\, dr}{\int_0^\infty J(r)2\pi r\, dr} = 2\left(\frac{\sigma}{r_e}\right)^2 \int_0^{r/\sigma} e^{-(r^2/2\sigma^2)} \\ \times \left[\int_0^{r_e/\sigma} \frac{R}{\sigma} e^{-(r^2/2\sigma^2)} I_0\left(\frac{rR}{\sigma^2}\right) d\left(\frac{R}{\sigma}\right)\right] \frac{r}{\sigma}\, d\left(\frac{r}{\sigma}\right) \equiv F\left(\frac{r}{\sigma}, \frac{r_e}{\sigma}\right) \end{aligned} \tag{26}$$

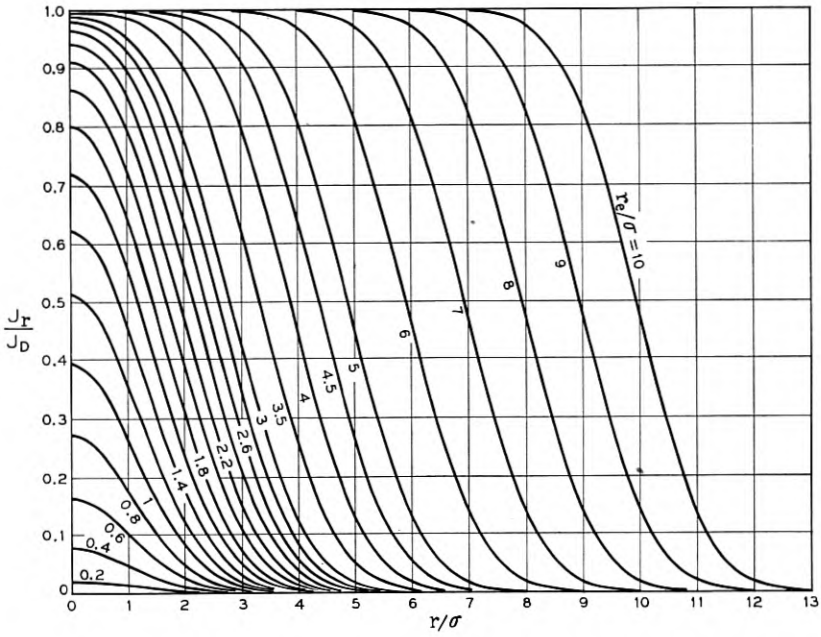


Fig. 6 — Curves showing the current density variation with radius in a beam which has been dispersed by thermal velocities. Here r_e is the nominal beam radius, r is the radius variable, and σ is the standard deviation defined in equation 17.

A family of curves with this ratio, F_r, as parameter has been reproduced from the Hines-Cutler paper and appears here as Fig. 7. Using this notation, (25) becomes

$$\frac{d^2r}{dt^2} = \frac{\sqrt{\eta/(2V_a)}}{2\pi\epsilon_0} I_D \frac{F_r}{r}$$

or

$$\frac{d^2r}{dz^2} = \frac{\eta}{2\pi\epsilon_0} \frac{I_D}{(2\eta V_a)^{3/2}} \frac{F_r}{r} \equiv K \frac{F_r}{r} \tag{27}$$

where we have made use of the dc electron drift velocity to make distance the independent variable instead of time, and have defined a quantity K which is proportional to gun perveance. We can now apply (27) to the motion of both the outer (edge) non-thermal electron and the σ-electron. From (26) we see that F_{r_e} and F_σ depend only on r_e/σ;

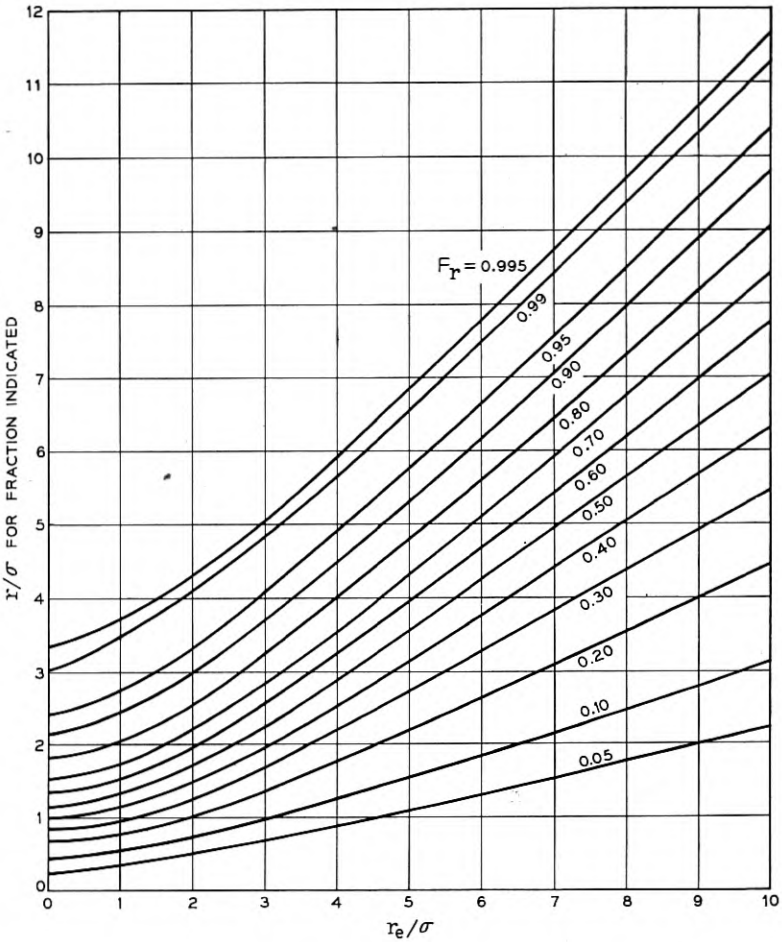


Fig. 7 — Curves showing the fraction, F_r, of the total beam current to be found within any given radius in a beam dispersed by thermal velocities as in Fig. 6.

consequently the continuous solution for r_e and r_σ $(= \sigma)$ as one moves axially along the drifting beam involves the simultaneous solution of two equations:

$$\frac{d^2 r_e}{dz^2} = KF_{r_e}/r_e$$
$$\frac{d^2 \sigma}{dz^2} = KF_\sigma/\sigma \qquad (28)$$

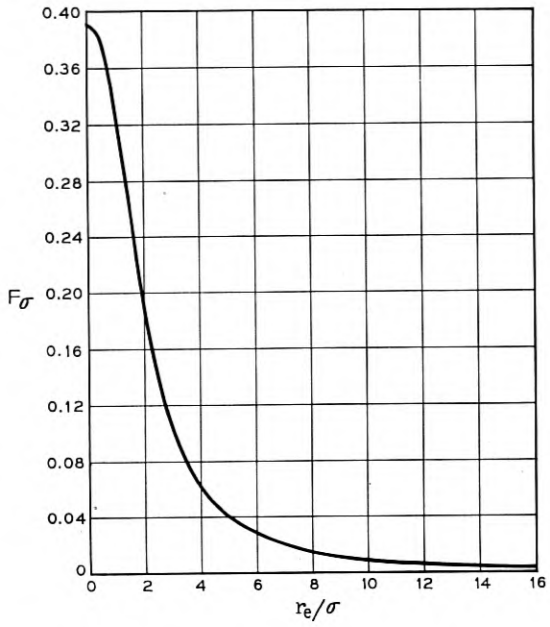


Fig. 8 — A curve showing the effect of a quantity related to the space charge force (in the drift region) on a thermal electron with standard deviation σ. (See equation 28.)

which are related by the mutual dependence of F_{r_e} and F_σ on r_e/σ. F_σ and F_{r_e}/r_e are plotted in Figs. 8 and 9.

We may summarize the treatment of the drift region, then, as follows:

(a) The input values of r_e and r_{e-}' at the entrance to the anode lens are obtained from the Pierce gun parameters r_a and θ, while the value of σ and σ_-' at the lens entrance can be obtained as mentioned above by integrating (15) from the cathode, where $\mu_c = 0$ and $(d\mu/dt)_c = 1$, to the anode plane. (The minus subscripts on r' and σ' indicate that these slopes are being evaluated on the gun side of the lens; a plus subscript will be used to indicate evaluation on the drift region side of the lens.) The values of r_e and σ on leaving the lens will of course be their entrance values in the drift region, and the effect of the lens on r_e' and σ' is simply found in terms of the anode lens correction factor Γ by use of (20). The value of σ at the anode can be obtained from (17) if μ is known there. In this regard, (15) can be integrated once to give

$$d\mu = \frac{1}{r_c}\left(\frac{dr}{dt}\right)_c \frac{dt}{(r_e/r_c)^2} \tag{29}$$

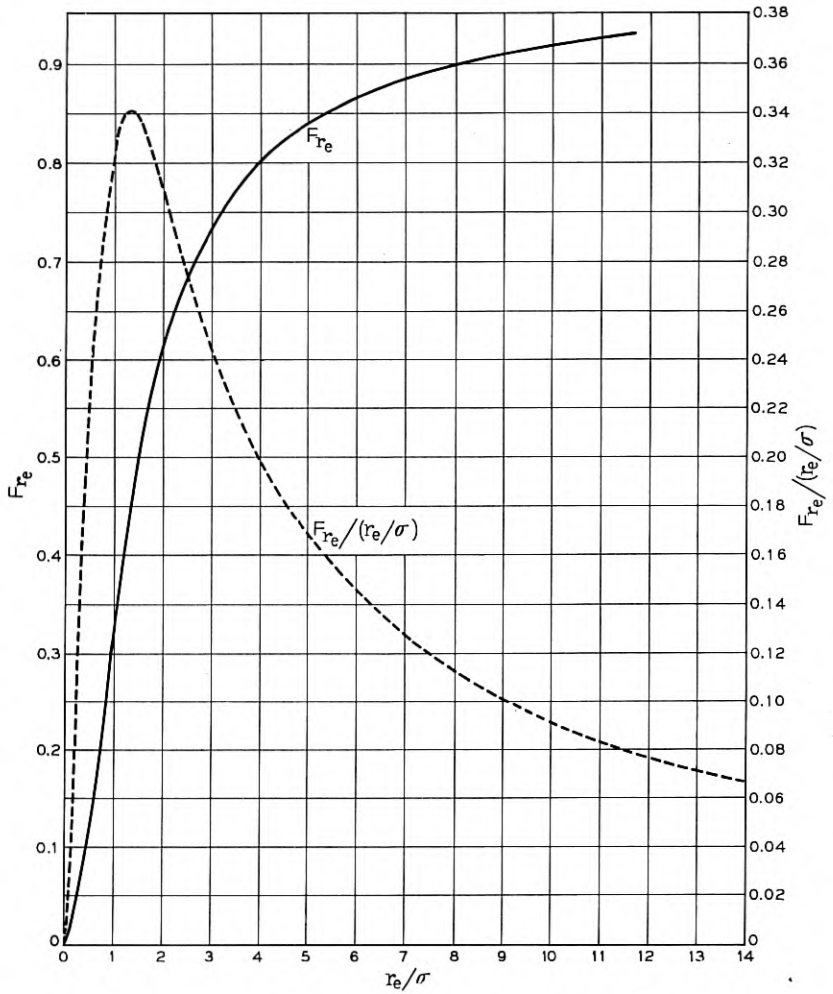


Fig. 9 — Showing quantities related to the effect of the space charge force in the drift region on the outermost non-thermal electron. (See equation 28.)

We can now substitute for transit time in terms of distance and Langmuir's well known potential function,[14] $-\alpha$. The value of this parameter, for the case of spherical cathode-anode geometry in which we are interested, depends only on the ratio $\bar{r}_c/\bar{r}$ which is equal to r_c/r_e. (Because of their frequent use in gun design, certain functions of $-\alpha$ are included here as Table I.) In terms of $-\alpha$, then, the potential in the gun region

TABLE I — TABLE OF FUNCTIONS OF $-\alpha$ OFTEN USED IN ELECTRON GUN DESIGN

$\bar{r}_c/\bar{r}$	$(-\alpha)^2$	$(-\alpha)^{4/3}$	$(-\alpha)^{2/3}$	$\int_1^{\bar{r}_c/\bar{r}} \frac{d\left(\frac{\bar{r}_c}{\bar{r}}\right)}{(-\alpha)^{2/3}}$	$\frac{d(-\alpha)^{4/3}}{d(\bar{r}_c/\bar{r})}$
1.0	0.0000	0.0000	0.0000	0.0000	
1.025	0.0006	0.0074			
1.05	0.0024	0.0179	0.134		
1.075	0.0052	0.0306	0.173		
1.10	0.0096	0.0452	0.212	1.392	0.590
1.15	0.0213	0.0768	0.277		
1.20	0.0372	0.1114	0.334	1.767	0.716
1.25	0.0571	0.1483	0.385		
1.30	0.0809	0.1870	0.432	2.031	0.790
1.35	0.1084	0.2273	0.476		
1.40	0.1396	0.2691	0.519	2.243	0.874
1.45	0.1740	0.3117	0.558		
1.50	0.2118	0.3553	0.596	2.423	0.886
1.60	0.2968	0.4450	0.667	2.583	0.915
1.70	0.394	0.5374	0.733	2.725	0.939
1.80	0.502	0.6316	0.795	2.855	0.954
1.90	0.621	0.7279	0.853	2.975	0.970
2.00	0.750	0.8255	0.908	3.087	0.982
2.10	0.888	0.9239	0.961	3.192	0.993
2.20	1.036	1.024	1.012	3.292	1.003
2.30	1.193	1.125	1.061	3.388	1.012
2.40	1.358	1.226	1.107	3.481	1.020
2.50	1.531	1.328	1.152	3.570	1.028
2.60	1.712	1.431	1.196	3.655	1.034
2.70	1.901	1.535	1.239	3.738	1.039
2.80	2.098	1.639	1.280	3.817	1.044
2.90	2.302	1.743	1.320	3.894	1.048
3.00	2.512	1.848	1.359	3.968	1.052
3.1	2.729	1.953	1.397	4.040	1.056
3.2	2.954	2.059	1.435	4.111	1.059
3.3	3.185	2.164	1.471	4.180	1.062
3.4	3.421	2.270	1.507	4.247	1.064
3.5	3.664	2.376	1.541	4.315	1.066
3.6	3.913	2.483	1.576	4.377	1.068
3.7	4.168	2.590	1.609	4.441	1.070
3.8	4.429	2.697	1.642	4.501	1.072
3.9	4.696	2.804	1.674	4.563	1.074
4.0	4.968	2.912	1.706	4.621	1.076

may be written

$$V = V_a(-\alpha)^{4/3}/(-\alpha_a)^{4/3} \tag{30}$$

$$dt = -\frac{d\bar{r}}{\sqrt{2\eta V}} = -\frac{d\bar{r}}{\sqrt{2\eta V_a}}\frac{(-\alpha_a)^{2/3}}{(-\alpha)^{2/3}} \tag{31}$$

so that upon substitution from (29) and (31), (17) becomes

$$\sigma = r_e\sqrt{\frac{T}{V_a}}\sqrt{\frac{k}{2e}}\frac{\bar{r}_c}{r_c}(-\alpha_a)^{2/3}\int_1^{\bar{r}_c/\bar{r}}(-\alpha)^{-2/3}\,d\left(\frac{\bar{r}_c}{\bar{r}}\right) \tag{32}$$

Fig. 5, which has been referred to above, shows

$$\frac{\sigma_a}{\bar{r}_c}\sqrt{\frac{2eV_a}{kT}}$$

as a function of $(\bar{r}_c/\bar{r}_a)$ as obtained from (32), and allows σ_a to be determined easily. Using (20), the value of r_{e+}' is given by

$$r_{e+}' = -\frac{r_{ea}}{F} + r_{e-}' = -\frac{r_{ea}}{F} - \theta_e = -\frac{\Gamma r_{ea}}{F_D} - \theta_e = \theta_e\left(-\frac{\Gamma\bar{r}_a}{F_D} - 1\right) \tag{33}$$

where θ_e is the half-angle of the cathode (and hence the initial angle which the path of a non-thermal edge electron makes with the axis). We may write for $1/F_D$

$$-\frac{1}{F_D} = \frac{V'}{4V} = \frac{\bar{r}_c}{4(-\alpha_a)^{4/3}\bar{r}_a^2}\left(\frac{d(-\alpha)^{4/3}}{d(\bar{r}_c/\bar{r})}\right)_a \tag{34}$$

In Fig. 10 we plot $-\bar{r}_a/F_D$ as a function of $\bar{r}_c/\bar{r}_a$ for easy evaluation of r_{e+}' in (33). Taking the first derivative of (32) with respect to z, we obtain an expression for σ_-'. Using this in conjunction with (20) and (34) we find

$$\sigma_+' = \sqrt{\frac{T}{V_a}}(\Gamma C_1 + C_2) \tag{35}$$

where

$$C_1 = \sqrt{\frac{k}{2e}}\frac{\bar{r}_c/\bar{r}_a}{4(-\alpha)^{2/3}}\left(\frac{d(-\alpha)^{4/3}}{d(\bar{r}_c/\bar{r})}\right)_a\int_1^{\bar{r}_c/\bar{r}_a}\frac{d(\bar{r}_c/\bar{r})}{(-\alpha)^{2/3}}$$

and

$$C_2 = \sqrt{\frac{k}{2e}}\left[\frac{\bar{r}_c}{\bar{r}_a} - (-\alpha_a)^{2/3}\int_1^{\bar{r}_c/\bar{r}_a}\frac{d(\bar{r}_c/\bar{r})}{(-\alpha)^{2/3}}\right]$$

C_1 and C_2 are plotted as functions of $\bar{r}_c/\bar{r}_a$ in Fig. 11.

(b) After choosing a specific value for Γ and evaluating $K = \eta I_D/$

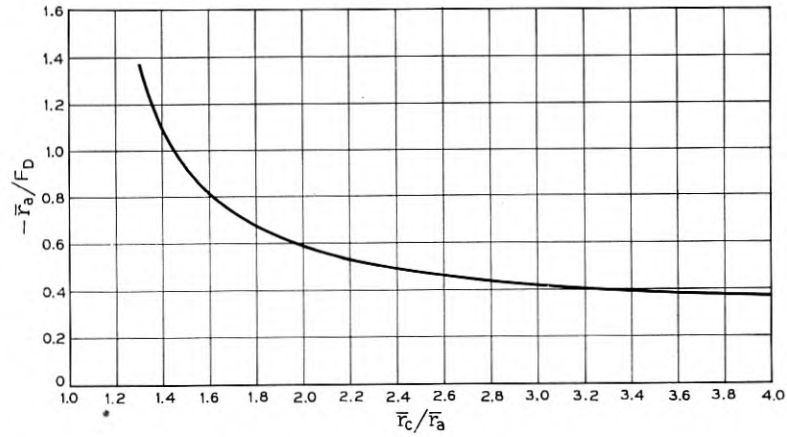


Fig. 10 — Curve used in finding r_{e+}', the direction of a nonthermal edge electron as it enters the drift region. (See equation 33.)

$(2\pi\epsilon_0(2\eta V_a)^{3/2})$, (28) is integrated numerically using the BTL analog computer to obtain σ and r_e as functions of axial distance along the beam.

(c) Knowing σ and r_e, other beam parameters such as current distribution and the radius of the circle which would encompass a given percentage of the total current can be found from Figs. 6 and 7.

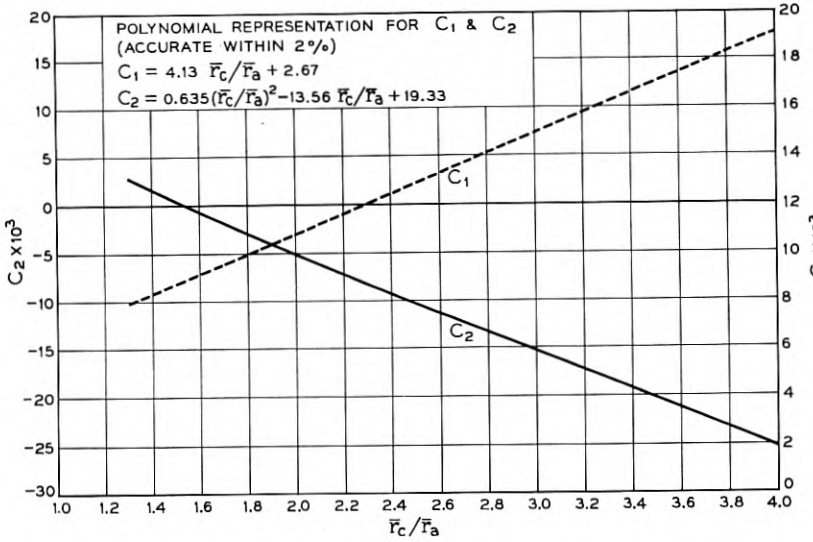


Fig. 11 — Curves used in evaluating σ_+', the slope of the trajectory of a thermal electron with standard deviation σ as it enters the drift region. (See equation 35.)

5. NUMERICAL DATA FOR ELECTRON GUN AND BEAM DESIGN

A. Choice of Variables

Except for a scaling parameter, the electrical characteristics of an ideal Pierce electron gun are completely determined when three parameters are specified, e.g., $\bar{r}_c/\bar{r}_a$, perveance, and V_a/T. Also, for the simplest case Γ is equal to 1 so that (since K depends only on gun perveance) in this case no additional parameter is needed. This implies that normalized values of r_e', σ, σ', and K at the drift side of the anode lens are not independent. If, however, the value of Γ at the anode lens is taken as an additional variable, four parameters plus simple scaling are required before complete predictions of beam characteristics can be made. In assembling analog computer data which would adequately cover values of $\bar{r}_c/\bar{r}_a$, perveance, and V_a/T which are likely to be of interest to us in designing future guns, we chose to present the major part of our data with Γ fixed at 1.1. This has seemed to be a rather typical value for Γ, and by choosing a specific value we decrease the total number of significant variables from 4 to 3. (The effect of variations in Γ on the minimum radius which contains 95 per cent of the beam is, however, included in Fig. 16 for particular values of V_a/T and perveance.) Although the boundary conditions for our mathematical description of the beam in a drift space are simplest when expressed in terms of r_e, r_e', σ and σ', we have attempted to make the results more usable by expressing all derived parameters in terms of $\bar{r}_c/\bar{r}_a$, $\sqrt{V_a/T}$, and the perveance, P.

B. Tabular Data

The rather extensive data obtained from the analog computer for the $\Gamma = 1.1$ case and for practical ranges in perveance, V_a/T, and $\bar{r}_c/\bar{r}_a$ are summarized in Tables IIA to E where the parameters r_e and σ which specify the beam cross section are given as functions of axial distance from the anode plane. Some feeling for the decrease in accuracy to be expected as the distance from the anode plane increases can be obtained by reference to Section 6B where experiment and theory are compared over a range of this axial distance parameter.

C. Graphical Data, Including Design Charts and Beam Profiles

In typical cases, the designer of Pierce electron guns is much more concerned with the beam radius at the axial position where it is smallest (and in the axial position of this minimum) than he is in the general

spreading of the beam with distance. This is true because, in microwave beam tubes, the beam from a magnetically shielded Pierce gun normally enters a strong axial magnetic field near a point where the radius is a minimum, so that magnetic focusing forces largely determine the beam's subsequent behavior. The analog computer data has therefore been reprocessed to stress the dependence of the beam's minimum diameter and the corresponding axial position of the minimum on the basic design parameters $\bar{r}_c/\bar{r}_a$, perveance, and $\sqrt{V_a/T}$. As a first step in this direction, the radius, r_{95}, of a circle which includes 95 per cent of the beam current is obtained as a function of axial position along the beam. Such data are shown graphically in Fig. 12. Finally, the curves of Fig. 12 are used in conjunction with the tabular data to obtain the "Design Curves" of Fig. 13 where all of the pertinent information relating to the beam at its minimum diameter is presented.

D. Example of Gun Design Using Design Charts

Assume that we desire an electron gun with the following properties: anode voltage $V_A = 1{,}080$ volts, cathode current $I_D = 7.1$ ma, and minimum beam diameter $2(r_{95})_{\min} = 0.015$ inches. Let us further assume a cathode temperature $T = 1080°$ Kelvin, an available cathode emission density of 190 ma per square cm, and an anode lens correction factor of $\Gamma = 1.1$. From these data we find $\sqrt{V_a/T} = 1.0$, perveance $P = 0.2 \times 10^{-6}$ amps/(volts)$^{3/2}$ and $(r_{95})_{\min}/r_c = 0.174$. Reference to the design chart, Fig. 13, now gives us the proper value for $\bar{r}_c/\bar{r}_a$: using the upper set of curves in the column for $\sqrt{V_a/T} = 1.0$ we note the point of intersection between the horizontal line for $(r_{95})_{\min}/r_c = 0.174$ and the perveance line $P = 0.2$, and read the value of $\bar{r}_c/\bar{r}_a$ $(= 2.8)$ as the corresponding abscissa. The convergence angle of the gun, θ_e, is now simply determined from the equation[15]

$$\theta_e = \cos^{-1}\left(1 - \frac{(-\alpha_a)^2 P}{14.67} \times 10^6\right) \tag{37}$$

(θ_e is found to be 13.7° in this example) and the potential distribution in the region of the cathode can be obtained from (30).

When this point has been reached, the gun design is complete except for the shapes of the beam forming electrode and the anode, which are determined with the aid of an electrolytic tank in the usual way.[16] The radius of the anode hole which will give a specified transmission can be found by obtaining $(r_e/\sigma)_a$ through the use of Fig. 5, and then choosing the anode radius from Fig. 7. In practical cases where $(r_e/\sigma)_a > 3.0$,

Table IIA — Summary of Analog Computer Data for an Anode Lens Correction of $\Gamma = 1.1$ for Perveance $= 0.05 \times 10^{-6}$ amps/(volts)$^{3/2}$

$\bar{r}_e/\bar{r}_a = 1.515$			2.0			2.5			3.0			3.5			4.0			
$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	
0	2.02	0.495	0	1.95	0.513	0	1.90	0.527	0	1.85	0.540	0	1.81	0.552	0	1.75	0.565	$\sqrt{V_a/T} = 0.5$
2	1.90	0.535	1	1.86	0.528	2	1.68	0.545	2	1.60	0.545	2	1.5	0.55	2	1.44	0.549	
4	1.76	0.579	2	1.78	0.541	4	1.45	0.560	4	1.35	0.560	4	1.19	0.548	4	1.07	0.531	
6	1.65	0.620	3	1.67	0.556	6	1.24	0.580	6	1.08	0.555	6	0.86	0.548	6	0.69	0.519	
8	1.55	0.669	4	1.60	0.570	8	1.04	0.600	8	0.82	0.565	8	0.53	0.548	8	0.27	0.505	
			6	1.44	0.601	10	0.86	0.621	10	0.58	0.575	10	0.20	0.552	10	−0.12	0.498	
			8	1.30		12	0.69	0.645	12	0.35	0.589	12	−0.10	0.558	12	−0.56	0.490	
			10	1.18	0.669	14	0.52	0.670	14	0.13	0.602	14	−0.40	0.563	14	−1.01	0.485	
			12	1.05	0.701	16	0.38	0.700	16	−0.10	0.619	16	−0.70	0.575	16	−1.48	0.482	
						18	0.25	0.725	20	−0.47	0.657	18	−1.0	0.585	18	−1.92	0.482	
						20	0.12	0.760	24	−0.81	0.702	20	−1.25	0.60	20	−2.38	0.485	
						22	0	0.790	26	−0.96	0.729	22	−1.53	0.615				
0	4.03	0.248	0	3.91	0.256	0	3.78	0.264	0	3.70	0.270	0	3.62	0.276	0	3.56	0.281	$\sqrt{V_a/T} = 1.0$
2	3.74	0.267	2	3.55	0.270	2	3.32	0.273	2	3.15	0.273	2	3.0	0.273	2	2.85	0.272	
4	3.48	0.288	4	3.20	0.285	4	2.87	0.281	4	2.62	0.277	4	2.35	0.273	4	2.11	0.265	
6	3.25	0.310	6	2.89	0.302	6	2.46	0.292	6	2.1	0.282	6	1.71	0.274	6	1.32	0.260	
8	3.06	0.332	8	2.61	0.318	8	2.08	0.308	8	1.6	0.289	8	1.06	0.280	8	0.54	0.259	
10	2.88	0.358	10	2.35	0.338	10	1.71	0.318	10	1.1	0.299	10	0.45	0.286	9	0.15	0.260	
12	2.73	0.382	12	2.11	0.358	12	1.38	0.334	12	0.67	0.311	11	0.15	0.292	10	−0.25	0.262	
14	2.59	0.408	14	1.90	0.379	13	1.23	0.344	13	0.46	0.32	12	−0.12	0.299	12	−1.0	0.27	
16	2.47	0.435	16	1.71	0.402	14	1.09	0.352	14	0.25	0.329	14	−0.63	0.314	14	−1.71	0.282	
20	2.26	0.492	20	1.40	0.456	16	0.83	0.374	16	−0.10	0.349	16	−1.1	0.332	16	−2.36	0.297	
24	2.09	0.553	24	1.14	0.516	20	0.38	0.426	18	−0.42	0.372	20	−1.86	0.380	18	−2.94	0.312	
26	2.02	0.588	28	0.95	0.582	24	0.06	0.489	20	−0.68	0.40	24	−2.43	0.433	20	−3.44	0.33	
						28	−0.20	0.561	24	−1.13	0.463	28	−2.87	0.492	24	−4.31	0.368	
									28	−1.47	0.535							

0	8.05	0.124	0	7.81	0.128	0	7.57	0.132	0	7.40	0.135	0	7.25	0.138	0	7.08	0.141	
4	6.94	0.145	4	6.4	0.142	4	5.77	0.140	4	5.3	0.138	4	4.7	0.135	4	4.2	0.132	
8	6.12	0.166	8	5.2	0.159	8	4.19	0.151	8	3.23	0.143	8	2.1	0.140	8	1.05	0.130	
12	5.5	0.191	12	4.25	0.180	12	2.81	0.168	12	1.40	0.158	10	0.9	0.148	9	0.30	0.135	
16	5.0	0.218	16	3.5	0.201	14	2.37	0.180	13	1.00	0.165	11	0.3	0.154	10	−0.32	0.140	
20	4.61	0.246	20	2.9	0.230	16	1.72	0.192	14	0.63	0.172	12	−0.1	0.161	11	−1.00	0.147	
24	4.30	0.276	24	2.4	0.265	18	1.29	0.210	15	0.30	0.182	14	−0.95	0.182	12	−1.60	0.155	$\sqrt{V_a/T} = 2.0$
28	4.02	0.310	28	2.0	0.305	20	0.90	0.232	16	0.03	0.193	16	−1.60	0.210	16	−3.25	0.201	
32	3.82	0.345	32	1.7	0.358	24	0.38	0.290	20	−0.79	0.255	20	−2.42	0.275	20	−4.25	0.255	
36	3.67	0.384	36	1.5	0.417	28	0.05	0.364	24	−1.28	0.331	24	−2.90	0.349	24	−4.85	0.31	
40	3.51	0.425	40	1.31	0.484	32	−0.18	0.453	28	−1.57	0.422	28	−3.25	0.425	28	−5.30	0.37	
			44	1.2	0.560	36	−0.31	0.552	32	−1.73	0.52	32	−3.45	0.504				
			48	1.1	0.640				36	−1.87	0.622							

TABLE IIB — SUMMARY OF ANALOG COMPUTER DATA FOR ANODE LENS CORRECTION OF $\Gamma = 1.1$ FOR PERVEANCE $= 0.1 \times 10^{-6}$ AMPS/(VOLTS)$^{3/2}$

$\bar{r}_c/\bar{r}_a = 1.515$			2.0			2.5			3.0			3.5			4.0			
$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	
0	2.85	0.350	0	2.75	0.363	0	2.69	0.372	0	2.60	0.382	0	2.56	0.390	0	2.52	0.397	
2	2.57	0.389	2	2.37	0.390	2	2.25	0.386	2	2.01	0.384	2	1.95	0.382	2	1.83	0.378	
4	2.32	0.431	4	2.04	0.421	4	1.82	0.405	4	1.56	0.391	4	1.30	0.382	4	1.05	0.365	
6	2.13	0.478	6	1.76	0.456	6	1.43	0.428	6	1.06	0.403	6	0.65	0.388	5	0.65	0.360	
8	1.95	0.525	8	1.52	0.492	8	1.07	0.452	8	0.60	0.420	7	0.35	0.392	6	0.24	0.358	
10	1.82	0.576	10	1.31	0.534	10	0.77	0.482	10	0.16	0.440	8	0.05	0.399	7	−0.15	0.357	
12	1.70	0.628	12	1.13	0.579	12	0.50	0.519	12	−0.18	0.467	9	−0.22	0.405	8	−0.56	0.357	$\sqrt{V_a/T} = 0.5$
14	1.60	0.683	14	0.97	0.625	14	0.26	0.557	14	−0.52	0.500	10	−0.51	0.415	10	−1.38	0.363	
			16	0.84	0.675	16	0.06	0.601	16	−0.82	0.537	11	−0.77	0.424	12	−2.16	0.372	
			18	0.73	0.730	18	−0.09	0.649	18	−1.07	0.575	12	−1.04	0.437	14	−2.90	0.385	
						20	−0.24	0.700	20	−1.29	0.620	14	−1.50	0.461	15	−3.23	0.392	
						22	−0.37	0.754	22	−1.48	0.668	16	−1.92	0.490	16	−3.57	0.400	
												18	−2.29	0.521				
												19	−2.46	0.537				

Table IIB (Continued)

$\bar{r}_c/\bar{r}_a = 1.515$			2.0			2.5			3.0			3.5			4.0			
$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	
0	5.72	0.175	0	5.48	0.182	0	5.37	0.186	0	5.23	0.191	0	5.13	0.195	0	5.05	0.198	$\sqrt{V_a/T} = 1.0$
1	5.39	0.183	2	4.75	0.195	2	4.49	0.191	2	4.16	0.192	2	3.90	0.190	2	3.65	0.185	
2	5.10	0.193	4	4.1	0.210	4	3.63	0.200	6	2.10	0.201	4	2.61	0.190	4	2.12	0.180	
4	4.62	0.214	5	3.83	0.219	6	2.86	0.212	8	1.17	0.215	6	1.36	0.196	5	1.34	0.180	
8	3.91	0.262	6	3.55	0.228	8	2.17	0.228	9	0.75	0.224	7	0.78	0.201	6	0.50	0.180	
12	3.41	0.314	7	3.3	0.237	10	1.55	0.248	10	0.38	0.239	8	0.22	0.210	7	−0.20	0.185	
16	3.06	0.374	8	3.09	0.248	11	1.28	0.259	12	−0.21	0.270	9	−0.27	0.220	8	−0.91	0.192	
20	2.79	0.441	10	2.66	0.270	12	1.04	0.273	14	−0.70	0.311	10	−0.72	0.234	10	−2.15	0.218	
24	2.58	0.514	12	2.31	0.294	14	0.63	0.306	16	−1.06	0.360	12	−1.5	0.266	12	−3.10	0.245	
26	2.49	0.554	16	1.75	0.357	16	0.31	0.348	18	−1.33	0.412	14	−2.08	0.304	14	−3.88	0.275	
			20	1.36	0.435	20	−0.10	0.449	20	−1.54	0.471	16	−2.51	0.349	16	−4.46	0.306	
			24	1.09	0.529	24	−0.35	0.570	22	−1.70	0.531	18	−2.85	0.392	18	−4.95	0.340	
			28	0.90	0.638	28	−0.52	0.710	24	−1.83	0.599	20	−3.1	0.442				
			32	0.77	0.760													
0	11.35	0.088	0	11.0	0.091	0	10.75	0.093	0	10.40	0.096	0	10.2	0.098	0	10.10	0.099	$\sqrt{V_a/T} = 2.0$
2	10.05	0.096	2	9.4	0.095	2	8.90	0.095	2	8.20	0.095	2	7.65	0.092	2	7.20	0.094	
4	9.15	0.106	4	8.1	0.103	4	7.20	0.101	4	6.20	0.100	4	5.12	0.091	4	4.20	0.092	
6	8.39	0.118	6	7.0	0.111	6	5.67	0.107	6	4.20	0.101	6	2.65	0.100	5	2.60	0.091	
8	7.75	0.130	8	6.08	0.120	8	4.33	0.115	8	2.40	0.110	7	1.45	0.103	6	1.00	0.083	
10	7.22	0.143	10	5.3	0.132	10	3.15	0.126	9	1.60	0.115	8	0.40	0.111	7	−0.28	0.100	
12	6.80	0.156	12	4.6	0.145	12	2.12	0.141	10	0.90	0.124	9	−0.39	0.124	8	−1.47	0.113	
16	6.15	0.186	14	4.03	0.160	13	1.72	0.151	11	0.32	0.138	10	−1.05	0.142	10	−3.05	0.149	
20	5.65	0.218	16	3.55	0.176	14	1.37	0.164	12	−0.08	0.154	12	−1.98	0.186	12	−3.95	0.189	
24	5.25	0.254	18	3.15	0.194	16	0.80	0.197	14	−0.68	0.200	14	−2.5	0.237	14	−4.62	0.230	
28	4.95	0.293	20	2.80	0.216	20	0.20	0.295	16	−1.08	0.259	16	−2.85	0.292	16	−4.95	0.274	
32	4.75	0.336	24	2.25	0.270	24	−0.05	0.427	20	−1.43	0.393	18	−3.08	0.350	18	−5.22	0.317	
34	4.65	0.357	28	1.93	0.339	28	−0.18	0.580	24	−1.65	0.548	20	−3.2	0.408	20	−5.45	0.385	
			32	1.70	0.421	32	−0.21	0.751	28	−1.73	0.710	22	−3.33	0.467				
			36	1.50	0.305													

Table IIC — Summary of Analog Computer Data for an Anode Lens Correction of $\Gamma = 1.1$ for Perveance $= 0.2 \times 10^{-6}$ amps/(volts)$^{3/2}$

$\bar{r}_c/\bar{r}_a = 1.515$			1.97			2.54			3.15			3.46			4.0			
$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	
0	4.05	0.247	0	3.9	0.258	0	3.75	0.266	0	3.63	0.276	0	3.55	0.282	0	3.48	0.287	$\sqrt{V_a/T} = 0.5$
2	3.55	0.290	2	3.2	0.290	2	2.85	0.282	2	2.52	0.278	2	2.37	0.282	2	2.05	0.270	
4	3.12	0.335	4	2.62	0.322	4	2.05	0.305	4	1.42	0.285	4	1.14	0.282	4	0.45	0.261	
6	2.77	0.380	6	2.15	0.360	6	1.35	0.331	6	0.45	0.308	5	0.55	0.285	4.5	0.05	0.265	
8	2.51	0.432	8	1.75	0.408	8	0.78	0.370	8	−0.38	0.341	6	0.00	0.296	5	−0.31	0.266	
10	2.30	0.490	10	1.45	0.459	10	0.34	0.420	10	−1.02	0.390	8	−1.00	0.325	6	−1.06	0.275	
12	2.12	0.550	12	1.21	0.520	12	0.00	0.480	12	−1.51	0.450	10	−1.78	0.370	8	−2.42	0.300	
14	1.98	0.614	14	1.01	0.580	14	−0.25	0.549	14	−1.88	0.515	12	−2.39	0.420	10	−3.52	0.330	
			16	0.87	0.655	16	−0.46	0.625	16	−2.15	0.588				12	−4.44	0.365	
			18	0.75	0.735	18	−0.61	0.710										
0	8.1	0.123	0	7.75	0.129	0	7.5	0.133	0	7.25	0.138	0	7.1	0.141	0	6.95	0.144	$\sqrt{V_a/T} = 1.0$
2	7.0	0.145	2	6.30	0.147	2	5.7	0.142	2	5.00	0.140	2	4.7	0.140	2	4.1	0.134	
4	6.15	0.166	4	5.20	0.163	4	4.1	0.153	4	2.85	0.145	4	2.25	0.140	4	0.95	0.135	
6	5.5	0.192	6	4.25	0.182	6	2.7	0.170	6	0.92	0.160	6	0.0	0.148	4.5	0.2	0.138	
8	5.0	0.216	8	3.5	0.205	8	1.59	0.195	7	0.12	0.178	8	−1.5	0.201	5	0.5	0.141	
10	4.6	0.244	10	2.95	0.234	10	0.75	0.237	8	−0.42	0.200	10	−2.35	0.264	6	1.72	0.160	
12	4.3	0.274	12	2.45	0.268	12	0.25	0.299	10	−1.27	0.261	12	−2.85	0.332	8	3.45	0.205	
14	4.02	0.306	14	2.10	0.301	14	−0.10	0.375	12	−1.76	0.339	14	−3.19	0.408	10	4.49	0.255	
16	3.8	0.34	16	1.80	0.358	16	−0.30	0.470	14	−2.08	0.423				12	5.15	0.310	
18	3.65	0.378	20	1.42	0.475	18	−0.45	0.570	16	−2.25	0.513							
20	3.5	0.42	24	1.20	0.625	20	−0.52	0.680	18	−2.40	0.610							
0	16.0	0.063	0	15.5	0.065	0	15.05	0.067	0	14.5	0.069	0	14.2	0.071	0	13.9	0.072	$\sqrt{V_a/T} = 2.0$
2	13.6	0.073	2	12.5	0.074	2	11.6	0.070	2	9.9	0.070	2	9.2	0.071	2	8.1	0.068	
4	12.1	0.084	4	10.3	0.080	4	8.3	0.075	4	5.7	0.072	4	4.4	0.071	4	1.9	0.068	
6	10.8	0.096	6	8.5	0.090	6	5.5	0.083	6	2.0	0.081	5	2.1	0.074	4.5	0.4	0.071	
8	9.8	0.110	8	7.1	0.101	8	3.4	0.095	7	0.5	0.094	6	0.2	0.085	5	−0.8	0.080	
10	9.1	0.124	10	6.0	0.115	10	1.8	0.119	8	−0.4	0.119	7	−1.0	0.108	6	−2.5	0.102	
12	8.5	0.139	12	5.1	0.131	12	0.8	0.165	10	−1.4	0.196	8	−1.8	0.140	8	−4.2	0.162	
14	8.05	0.156	14	4.45	0.151	14	0.3	0.235	12	−1.8	0.291	10	−2.58	0.220	10	−4.95	0.225	
16	7.65	0.174	16	3.9	0.175	16	0.1	0.330	14	−1.95	0.395	12	−2.9	0.305	12	−5.38	0.290	
18	7.35	0.193	20	3.2	0.236	18	−0.1	0.435	16	−2.05	0.505	14	−3.05	0.397	14	−5.60	0.360	
20	7.1	0.214	24	2.75	0.316	20	−0.1	0.558	18	−2.1	0.621	16	−3.15	0.492				
22	6.85	0.236	28	2.45	0.415	22	−0.15	0.685										
24	6.7	0.258	32	2.26	0.535	24	−0.2	0.820										

TABLE IID — SUMMARY OF ANALOG COMPUTER DATA FOR AN ANODE LENS CORRECTION OF $\Gamma = 1.1$ FOR PERVEANCE $= 0.4 \times 10^{-6}$ AMPS./(VOLTS)$^{3/2}$

$\bar{r}_c/\bar{r}_a = 1.515$			2.04			2.44			2.86			3.6			4.0			
$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	
0	5.69	0.176	0	5.45	0.184	0	5.32	0.188	0	5.19	0.193	0	5	0.200	0	4.85	0.206	
2	4.65	0.220	2	4.00	0.210	2	3.60	0.200	2	3.20	0.198	1	3.7	0.195	1	3.40	0.198	
4	3.92	0.268	4	2.95	0.242	4	2.29	0.230	4	1.50	0.220	2	2.35	0.192	2	1.86	0.192	
6	3.41	0.321	6	2.15	0.290	6	1.20	0.270	5	0.75	0.225	3	0.97	0.196	3	0.25	0.196	
8	3.09	0.382	8	1.60	0.350	8	0.49	0.341	6	0.18	0.270	4	−0.27	0.210	3.5	−0.45	0.201	
10	2.80	0.452	10	1.20	0.430	10	0.07	0.450	8	−0.65	0.360	5	−1.32	0.231	4	−1.15	0.210	$\sqrt{V_a/T} = 0.5$
12	2.59	0.526	12	0.91	0.525	12	−0.18	0.570	10	−1.15	0.460	6	−2.15	0.265	5	−2.40	0.232	
			14	0.75	0.640	14	−0.32	0.705	12	−1.41	0.575	7	−2.78	0.300	6	−3.35	0.260	
			16	0.61	0.760	16	−0.45	0.860	14	−1.60	0.715	8	−3.26	0.340	7	−4.10	0.290	
												9	−3.65	0.380	8	−4.74	0.320	
0	11.4	0.088	0	10.9	0.092	0	10.7	0.094	0	10.3	0.097	0	10.0	0.100	0	9.75	0.103	
2	9.3	0.108	2	8.1	0.106	2	7.2	0.100	2	6.6	0.099	1	7.5	0.098	1	6.8	0.097	
4	7.9	0.132	4	6.0	0.124	4	4.5	0.110	4	3.2	0.11	2	4.85	0.096	2	3.7	0.095	
6	6.9	0.156	6	4.5	0.147	6	2.4	0.135	6	0.5	0.15	3	2.1	0.099	3	0.50	0.098	
8	6.2	0.188	8	3.4	0.179	8	1.1	0.185	8	−0.6	0.25	3.5	0.80	0.104	4	−1.95	0.121	
10	5.7	0.220	10	2.6	0.220	10	0.42	0.275	10	−1.05	0.34	4	−0.32	0.115	5	−3.45	0.156	$\sqrt{V_a/T} = 1.0$
12	5.3	0.256	12	2.08	0.280	12	0.18	0.40	12	−1.3	0.54	5	−1.9	0.15	6	−4.37	0.195	
14	5.0	0.293	14	1.7	0.352	14	0.00	0.54	14	−1.4	0.70	6	−2.8	0.195	7	−4.98	0.236	
16	4.8	0.336	16	1.5	0.445	16	−0.10	0.70				8	−3.7	0.292	8	−5.4	0.278	
18	4.6	0.380	20	1.25	0.675	18	−0.15	0.88				10	−4.1	0.398	9	−5.68	0.320	
0	22.8	0.044	0	21.8	0.046	0	21.3	0.047	0	20.7	0.048	0	20	0.05	0	19.5	0.052	
2	18.5	0.052	2	17	0.050	2	14.7	0.050	2	12.5	0.050	1	14.6	0.049	1	13.4	0.050	
4	14.5	0.064	4	12.5	0.060	4	9.3	0.055	4	5.7	0.056	2	9.2	0.048	2	7.4	0.048	
6	13.7	0.078	6	9.5	0.070	6	5.2	0.070	6	1.0	0.082	3	3.8	0.049	3	1.3	0.049	
8	12.2	0.092	8	7.25	0.086	8	2.5	0.098	8	−0.50	0.174	3.5	1.3	0.052	4	−2.6	0.076	
10	11.3	0.108	10	5.6	0.110	10	1.2	0.16	10	−0.80	0.336	4	−0.60	0.062	5	−4.1	0.120	$\sqrt{V_a/T} = 2.0$
12	10.6	0.126	12	4.6	0.140	12	0.75	0.265	12	−0.95	0.52	5	−2.45	0.106	6	−4.7	0.167	
14	10	0.146	14	3.9	0.175	14	0.60	0.405	14	−1.0	0.72	6	−3.15	0.160	7	−5.1	0.215	
16	9.7	0.168	16	3.5	0.222	16	0.49	0.57				7	−3.5	0.220	8	−5.35	0.265	
18	9.3	0.190	20	2.90	0.350	18	0.45	0.75				8	−3.65	0.281	9	−5.5	0.315	
			24	2.60	0.510	20	0.40	0.94				10	−3.82	0.410				

TABLE IIE — SUMMARY OF ANALOG COMPUTER DATA FOR AN ANODE LENS CORRECTION OF $\Gamma = 1.1$ FOR PERVEANCE $= 0.8 \times 10^{-6}$ AMPS./(VOLTS)$^{3/2}$

$\bar{r}_c/\bar{r}_a = 1.515$			2.0			2.5			3.0			3.5			4.0			
$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	$\frac{z}{r_a}$	$\frac{r_e}{\sigma}$	$\frac{\sigma}{r_a}$	
0	8.00	0.125	0	7.7	0.130	0	7.42	0.135	0	7.15	0.140	0	6.85	0.146	0	6.62	0.151	$\sqrt{V_a/T} = 0.5$
2	6.1	0.165	2	5.0	0.160	2	4.05	0.153	2	2.85	0.150	1	4.3	0.140	1	3.64	0.142	
4	4.95	0.215	3	4.0	0.180	3	2.7	0.17	3	1.0	0.168	2	1.61	0.145	1.5	2.03	0.140	
6	4.21	0.275	4	3.3	0.202	4	1.55	0.196	3.5	0.20	0.184	2.5	0.40	0.155	2	0.4	0.142	
8	3.80	0.340	5	2.7	0.230	5	0.75	0.24	4	−0.30	0.205	3	−0.51	0.170	2.5	−0.95	0.153	
10	3.49	0.420	6	2.2	0.265	6	0.21	0.30	5	−1.05	0.265	4	−2.0	0.220	3	−2.15	0.17	
			8	1.6	0.360	8	−0.28	0.469	6	−1.52	0.345	5	−2.82	0.280	4	−3.85	0.214	
			10	1.2	0.490	10	−0.50	0.680	8	−2.0	0.530	6	−3.35	0.349	5	−4.95	0.260	
			12	1.0	0.645	12	−0.62	0.919	9	−2.15	0.625	7	−3.7	0.420	6	−5.71	0.310	
0	16.0	0.062	0	15.4	0.065	0	14.8	0.067	0	14.3	0.070	0	13.7	0.073	0	13.2	0.076	$\sqrt{V_a/T} = 1.0$
2	12.1	0.086	2	10	0.080	2	7.9	0.076	1	9.50	0.071	1	8.5	0.073	1	7.1	0.072	
4	9.85	0.110	4	6.6	0.100	4	3.1	0.10	2	5.6	0.075	2	3.4	0.074	1.5	4.0	0.070	
6	8.5	0.140	5	5.5	0.115	5	1.6	0.123	3	2.0	0.085	2.5	1.0	0.80	2	0.90	0.071	
8	7.65	0.175	6	4.5	0.130	6	0.70	0.170	3.5	0.60	0.100	3	−0.6	0.096	2.2	0.00	0.075	
10	7.1	0.215	7	3.9	0.154	8	0.08	0.340	4	−0.2	0.121	3.5	−1.71	0.125	2.5	−1.4	0.089	
12	6.7	0.259	8	3.3	0.180	10	−0.08	0.570	5	−1.1	0.200	4	−2.45	0.16	3.0	−3.1	0.112	
			10	2.6	0.250	12	−0.13	0.830	6	−1.5	0.295	5	−3.15	0.235	4.0	−4.7	0.170	
			12	2.15	0.345				7	−1.69	0.405	6	−3.5	0.315	5.0	−5.55	0.225	
			14	1.9	0.470				8	−1.79	0.518	7	−3.68	0.40	6.0	−6.0	0.289	
			16	1.7	0.610				9	−1.81	0.635							
0	32.1	0.031				0	29.7	0.034	0	28.5	0.035	0	27.5	0.0364	0	25.7	0.038	$\sqrt{V_a/T} = 2.0$
2	24.2	0.042				1	21.6	0.036	1	18.6	0.035	1	16.7	0.0364	1	14	0.036	
4	19.7	0.056				2	15.5	0.038	2	10.9	0.037	2	7.0	0.04	1.5	7.7	0.035	
6	17.2	0.071				3	10.5	0.042	3	3.9	0.042	2.5	2.4	0.056	2	1.6	0.035	
8	15.4	0.090				4	6.4	0.049	3.5	1.25	0.05	3	0.7	0.091	2.5	−2.1	0.050	
10	14.2	0.112				5	3.45	0.061	4	0.00	0.068	3.5	−2.1	0.131	3	−3.9	0.081	
12	13.4	0.134				6	1.7	0.089	5	−1	0.151	4	−2.6	0.221	4	−5.1	0.146	
14	12.8	0.160				7	0.9	0.145	6	−1.15	0.268	5	−3.0	0.316	5	−5.6	0.215	
						8	0.60	0.226	7	−1.25	0.395	6	−3.15	0.416	6	−5.85	0.285	
						9	0.50	0.330	8	−1.38	0.532	7	−3.25	0.520				
						10	0.48	0.447	9	−1.4	0.678	8	−3.3					
						12	0.40	0.71										

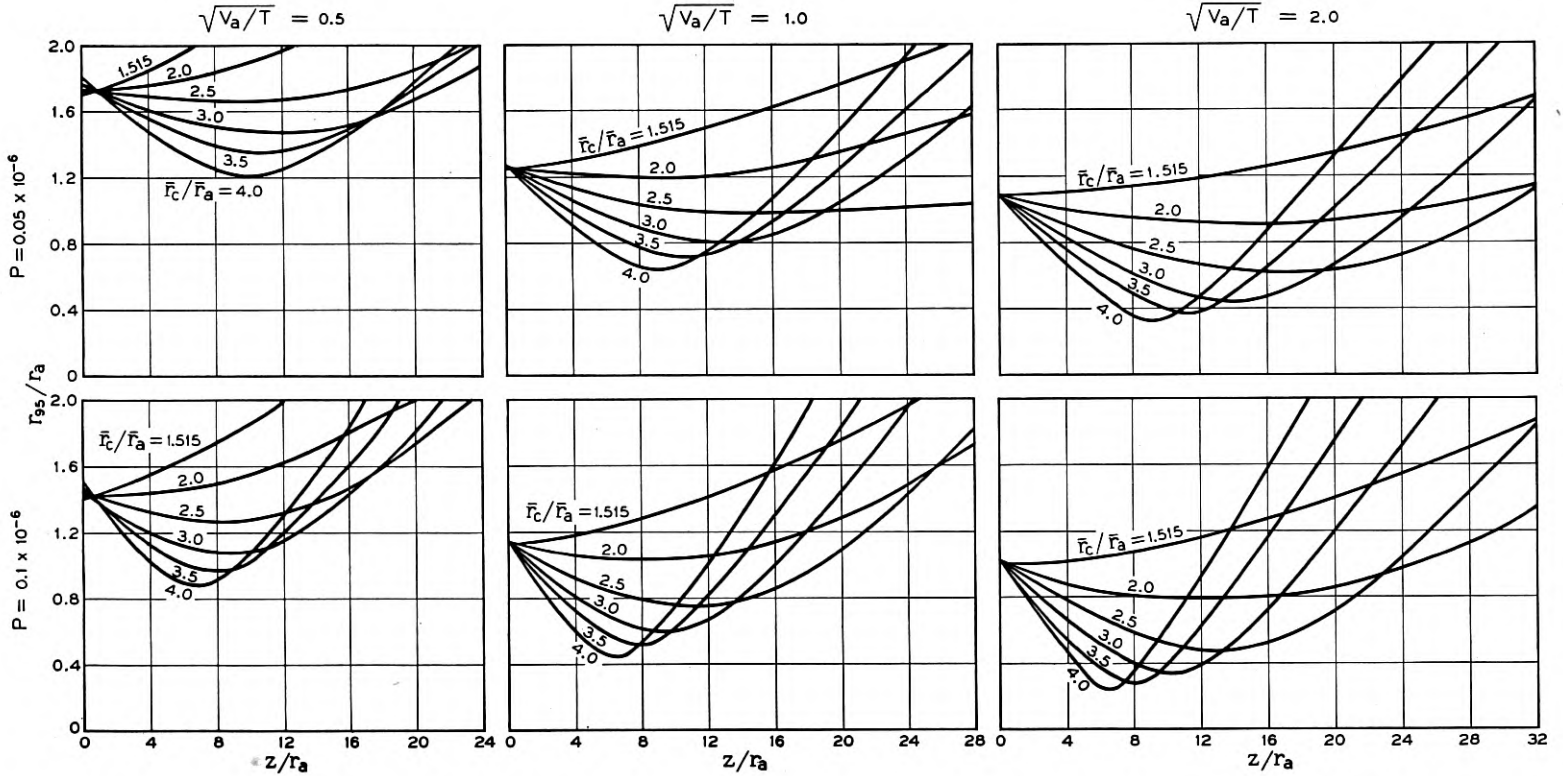


Fig. 12A — Curves showing normalized beam radius (95 per cent) versus distance from gun anode for variations in perveance (P), $\bar{r}_c/\bar{r}_a$ and $\sqrt{V_a/T}$.

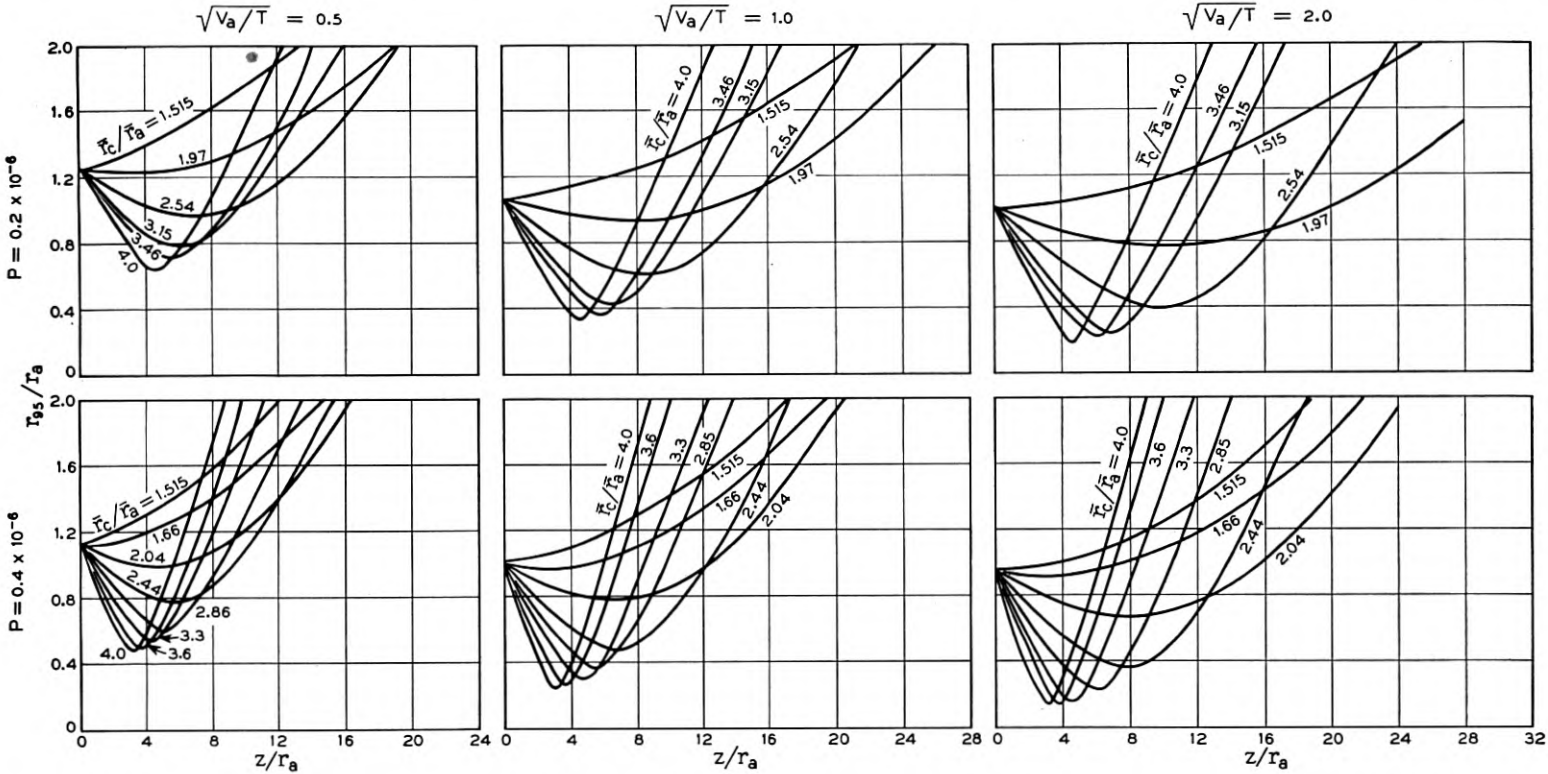


Fig. 12B — Curves showing normalized beam radius (95 per cent) versus distance from gun anode for variations in perveance (P), $\bar{r}_c/\bar{r}_a$ and $\sqrt{V_a/T}$.

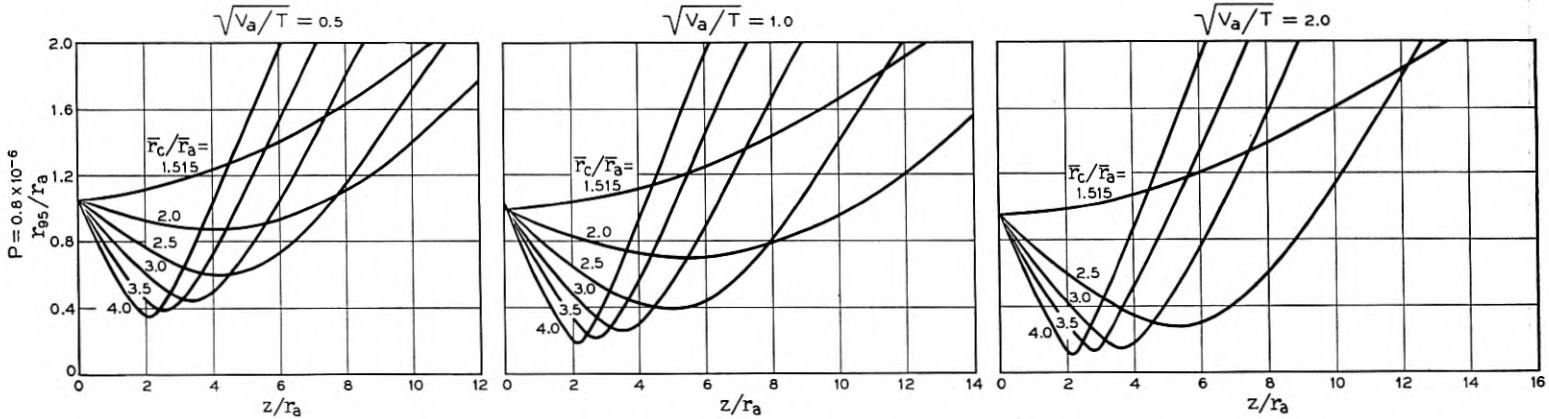


Fig. 12C — Curves showing normalized beam radius (95 per cent) versus distance from gun anode for variations in perveance (P), $\bar{r}_c/\bar{r}_a$ and $\sqrt{V_a/T}$.

we find less than 1 per cent anode interception if

$$\text{anode hole radius} = 0.93\, r_{ea} + 2\sigma_a \tag{38}$$

Additional information about the axial position of $(r_{95})_{\min}$ and the current density distribution in the corresponding transverse plane is contained in Fig. 13. The second set of curves in the $\sqrt{V_a/T} = 1$ column gives $z_{\min}/r_c = 2.42$ for this example, so that we would predict

$$z_{\min} \equiv \text{distance from anode to } (r_{95})_{\min} = 0.104''$$

The remaining 3rd and 4th sets of curves in the $\sqrt{V_a/T} = 1$ column allow us to find σ and r_e/σ at $z_{\min}$. In particular we obtain $\sigma = 0.0029''$ and $r_e/\sigma = 0.8$, and use Fig. 6 to give the current density distribution at $z_{\min}$.* Section VI contains experimental data which indicate a somewhat larger value for $z_{\min}$ than that obtained here. However the parameter of greatest importance, $(r_{95})_{\min}$, is predicted with embarrassing precision.

For those cases in which additional information is required about the beam shape at axial points other than $z_{\min}$, the curves of Fig. 12 or the data of Table II may be used.

6. COMPARISON OF THEORY WITH EXPERIMENT

In order to check the general suitability of the foregoing theory and the usefulness of the design charts obtained, several scaled-up versions of Pierce type electron guns, including the gun described in Section 5*D*, were assembled and placed in the double-aperture beam analyzer described in Reference 7.

A. Measurement of Current Densities in the Beam

Measurements of the current density distributions in several transverse planes near $z_{\min}$ were easily obtained with the aid of the beam analyzer. The resulting curve of relative current density versus radius at the experimental $z_{\min}$ is given in Fig. 14 for the gun of Section 5*D*. (This curve is further discussed in Part *C* below.) For this case, as well as for all others, special precautions were taken to see that the gun was functioning properly: In addition to careful measurement of the size and position of all gun parts, these included the determination that the distribution of transverse velocities at the center of the beam was smooth

* When $r_e/\sigma < 0.5$, the current density distribution depends almost entirely on σ, and, in only a minor way, on the ratio r_e/σ so that in such cases this ratio need not be accurately known.

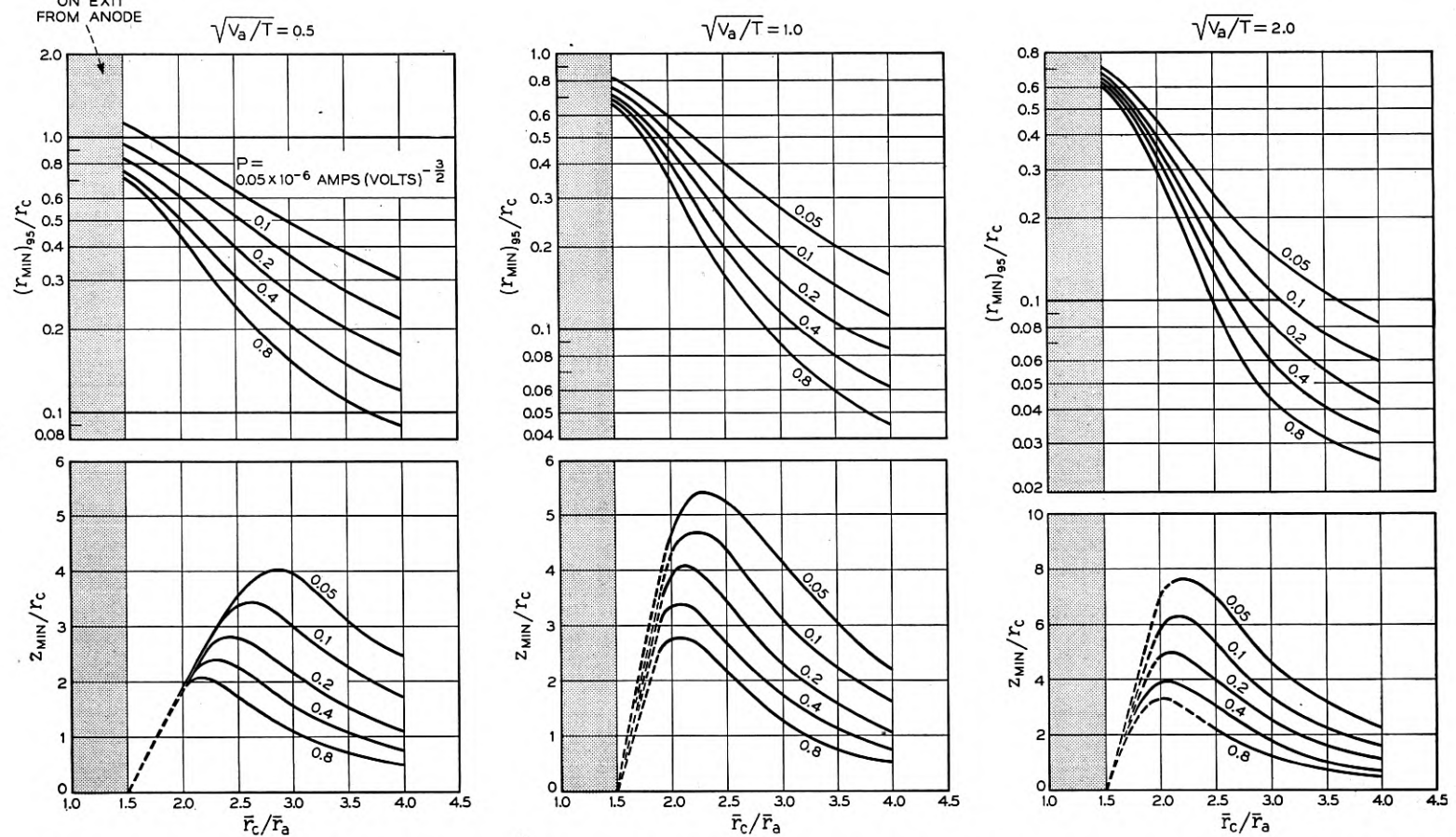
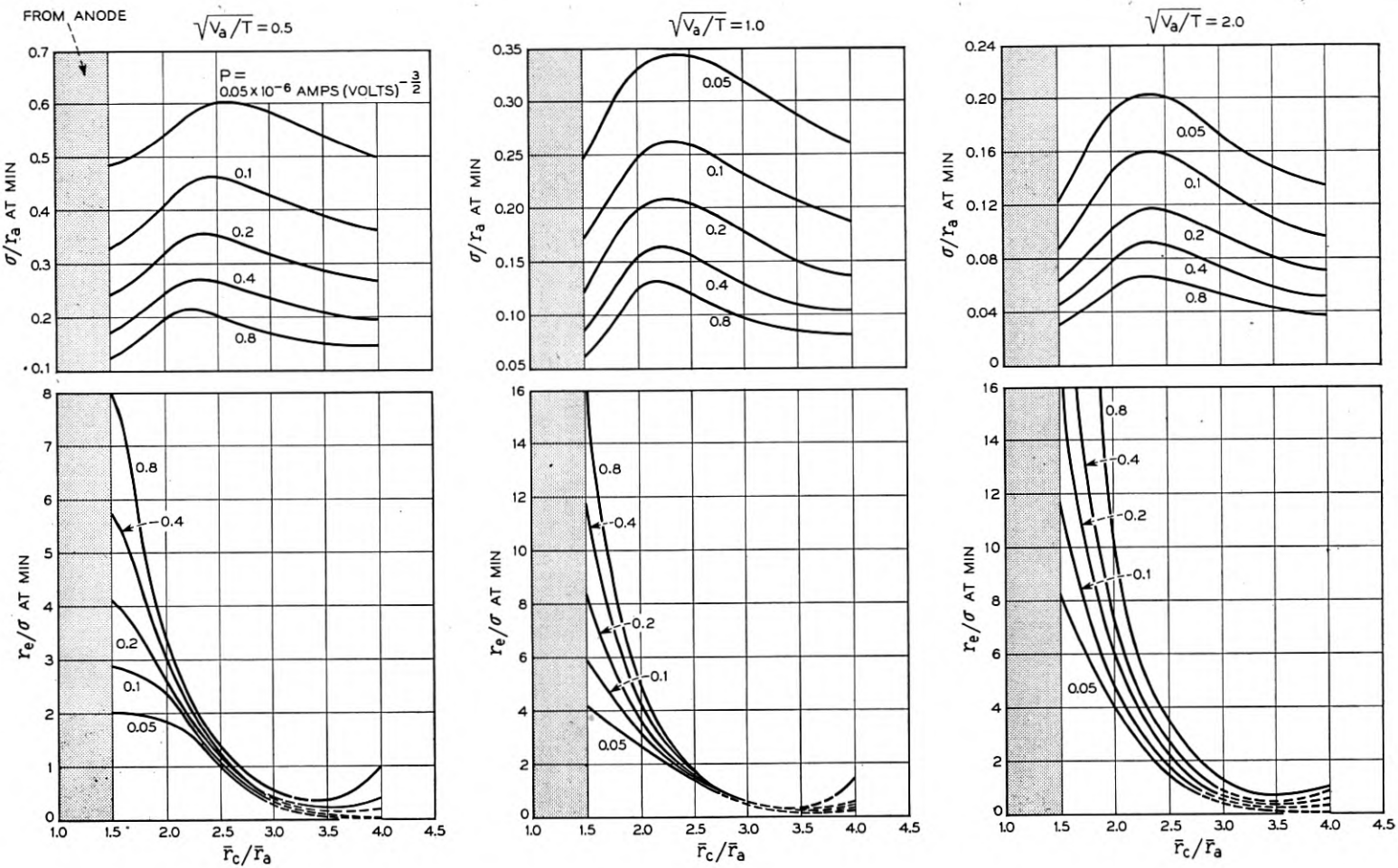


Fig. 13A — Design curves for Pierce-type electron guns considering transverse thermal velocities of electrons on emission from the cathode and an anode lens correction of $\Gamma = 1.1$. (for values of $r_e/\sigma < 0.5$, see footnote to Section 5*D*.)


FROM ANODE
$\sqrt{V_a/T} = 0.5$
$\sqrt{V_a/T} = 1.0$
$\sqrt{V_a/T} = 2.0$
P = 0.05×10^{-6} AMPS (VOLTS)$^{-\frac{3}{2}}$
σ/r_a AT MIN
r_e/σ AT MIN
$\bar{r}_c/\bar{r}_a$
0.05
0.1
0.2
0.4
0.8

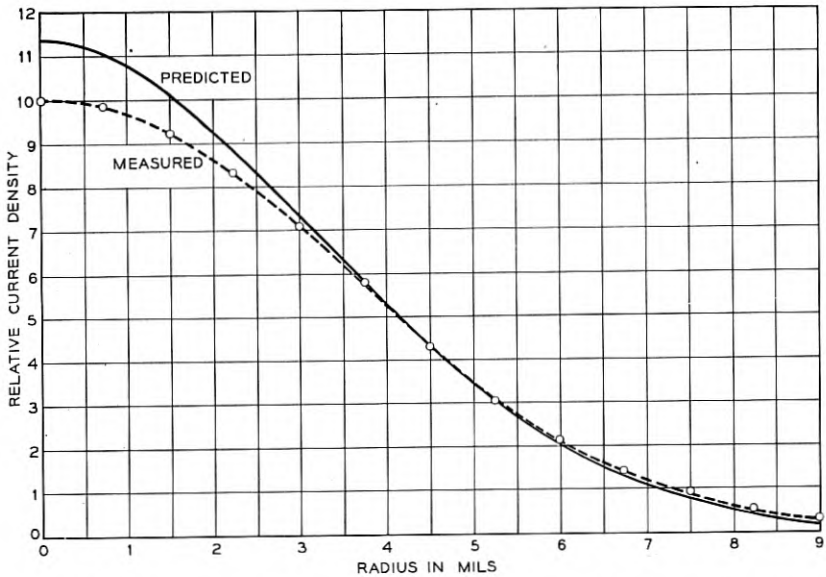


Fig. 14 — Current density distribution in a transverse plane located where the 95 per cent radius is a minimum. The predicted and measured curves are normalized to contain the same total current. (The corresponding prediction from the universal beam spread curve would show a step function with a constant relative current density of 64.2 for $r < 1.2$ mils and zero beyond.) The gun parameters are given in Section 5*D*.

and generally Gaussian in form, thereby indicating uniform cathode emission and proper boundary conditions at the edge of the beam near the cathode. The effect of positive ions on the beam shape was in every case reduced to negligible proportions, either by using special pulse techniques,[7] or by applying a small voltage gradient along the axis of the beam.

B. Comparison of the Experimentally Measured Spreading of a Beam with that Predicted Theoretically

From the experimentally obtained plots of current density versus radius at several axial positions along the beam, we have obtained at each position (by integrating to find the total current within any radius) a value for the radius, r_{95}, of that circle which encompasses 95 per cent of the beam. For brevity, we call the resulting plots of r_{95} versus axial distance, "beam profiles". The experimental profile for the gun described in Section 5*D* is shown as curve A in Fig. 15(a). Curve B shows the profile as predicted by the methods of this paper and obtained from Fig. 12. Curve C is the corresponding profile which one obtains by the Hines-Cutler method,[6] and Curve D represents r_{95} as obtained from the

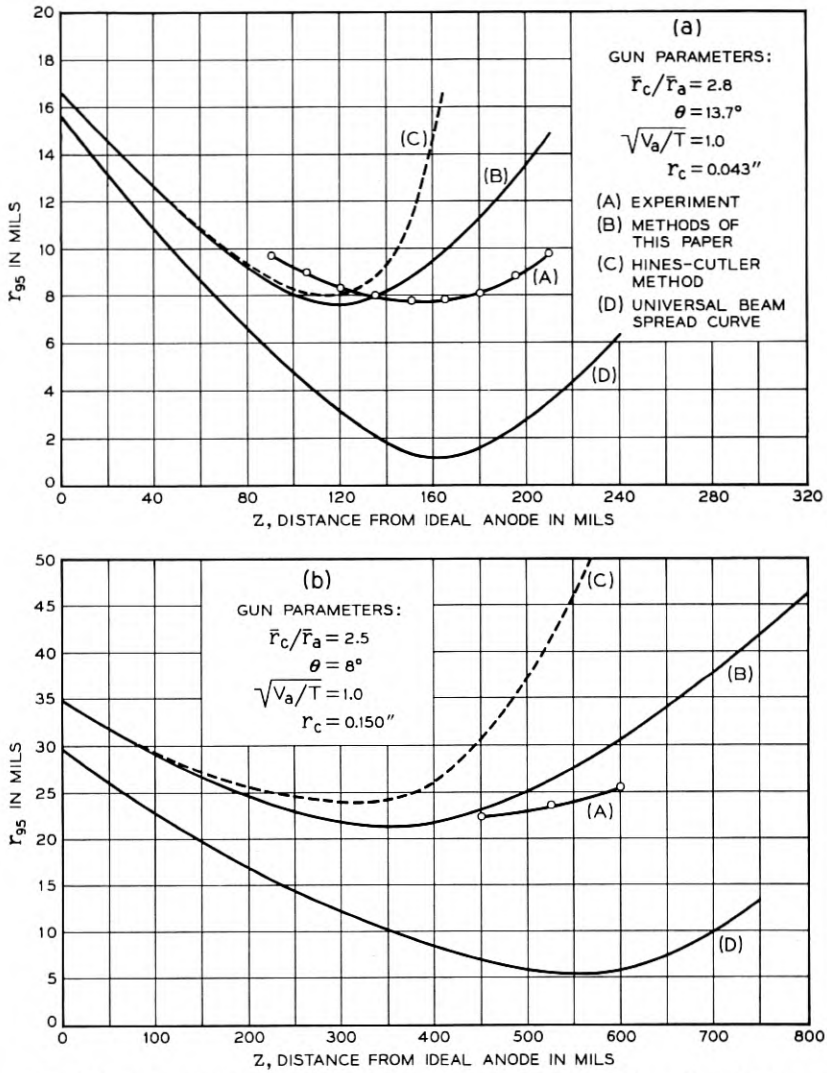


Fig. 15 — Beam profiles (using an anode lens correction of $\Gamma = 1.1$ and the gun parameters indicated) as obtained (A) from experiment, (B) by the methods of this paper, (C) Hines-Cutler method, (D) by use of the universal beam spread curve.

universal beam spread curve[12] (i.e., under the assumption of laminar flow and gradual variations of beam radius with distance). Note that in each case a value of 1.1 has been used for the correction factor, Γ, representing the excess divergence of the anode lens. The agreement in $(r_{95})_{min}$ as obtained from Curves A and B is remarkably good, but the axial position of $(r_{95})_{min}$ in Curve A definitely lies beyond the correspond-

ing minimum position in Curve B. Fortunately, in the gun design stage, one is usually more concerned with the value of $(r_{95})_{\min}$ than with its exact axial location. The principal need for knowing the axial location of the minimum is to enable the axial magnetic field to build up suddenly in this neighborhood. However, since this field is normally adjusted experimentally to produce best focusing, an approximate knowledge of $z_{\min}$ is usually adequate.

In Fig. 15b we show a similar set of experimental and theoretical beam profiles for another gun. The relative profiles are much the same as in Fig 15a, and all of several other guns measured yield experimental points similarly situated with respect to curves of Type B.

C. Comparison of Experimental and Theoretical Current Density Distributions where the Minimum Beam Diameter is Reached

In Fig. 14 we have plotted the current density distribution we would have predicted in a transverse plane at $z_{\min}$ for the example introduced in Section 5*D*. Here the experimental and theoretical curves are normalized to include the same total currents in their respective beams. The noticeable difference in predicted and measured current densities at the center of the beam does not appreciably alter the properties such a beam would have on entering a magnetic field because so little total current is actually represented by this central peak.

D. Variation of Beam Profile with Γ

All of the design charts have been based on a value of $\Gamma = 1.1$, which is typical of the values obtained by the methods of Section 3. When appreciably different values of Γ are appropriate, we can get some feeling for the errors involved, in using curves based on $\Gamma = 1.1$, by reference to Fig. 16. Here we show beam profiles as obtained by the methods of this paper for three values of Γ. The calculations are again based on the gun of Section 5*D*, and a value of just over 1.1 for Γ gives the experimentally obtained value for $(r_{95})_{\min}$.

7. SOME ADDITIONAL REMARKS ON GUN DESIGN

In previous sections we have not differentiated between the voltage on the accelerating anode of the gun and the final beam voltage. It is important, however, that the separate functions of these two voltages be kept clearly in mind: The accelerating anode determines the total current drawn and largely controls the shaping of the beam; the final beam voltage is, on the other hand, chosen to give maximum interaction between the electron beam and the electromagnetic waves traveling along the slow wave circuit. As a consequence of this separation of func-

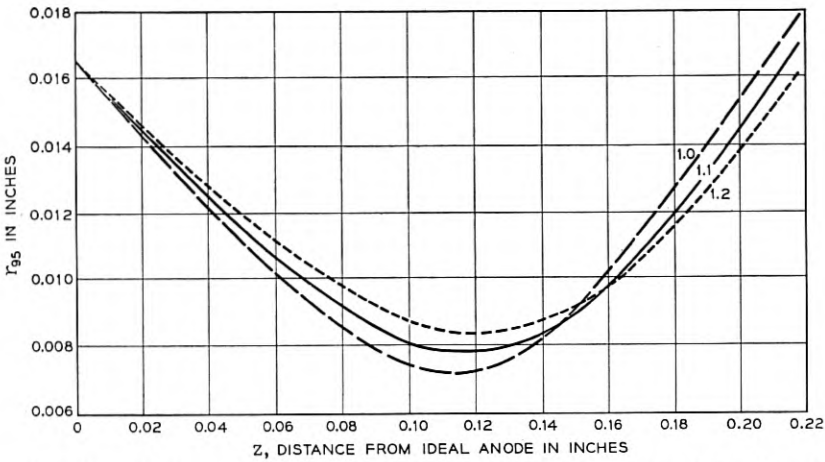


Fig. 16 — Beam profiles as obtained by the methods of this paper for the gun parameters given in Section 5*D*. Curves are shown for three values of the anode lens correction, viz. $\Gamma = 1.0$, 1.1, and 1.2.

tions, it is found that some beams which are difficult or impossible to obtain with a single Pierce-gun acceleration to final beam voltage may be obtained more easily by using a lower voltage on the gun anode. The acceleration to final beam voltage is then accomplished after the beam has entered a region of axial magnetic field.

Suppose, for example, that one wishes to produce a 2-ma, 4-kv beam with $(r_{95}/r_c) = 0.25$. If the cathode temperature is 1000°K, and the gun anode is placed at a final beam voltage of 4 kv, we have $\sqrt{V_a/T} = 2$ and $P = 0.008$. From the top set of curves under $\sqrt{V_a/T} = 2$ in Fig. 13, we find (by using a fairly crude extrapolation from the curves shown) that a ratio of $\bar{r}_c/\bar{r}_a \approx 3.5$ is required to produce such a beam. The value of (r_e/σ) at $z_{\min}$ is therefore less than about 0.2 so that there is little semblance of laminar flow here. On the other hand we might choose $V_z = 250$ volts so that $\sqrt{V_a/T} = 0.5$ and $P = 0.51$. From Fig. 13 we than obtain $\bar{r}_c/\bar{r}_a = 2.6$ and $(r_e/\sigma)_{\min} = 0.8$ for the same ratio of $r_{95}/r_c (= 0.25)$. While the flow could still hardly be called laminar, it is considerably more ordered than in the preceding case. Here we have included no correction for the (convergent) lens effect associated with the post-anode acceleration to the final beam voltage, $V = 4$ kv.

Calculations of the Hines-Cutler type will always predict, for a given set of gun parameters and a specified anode lens correction, a minimum beam size which is larger than that predicted by the methods of this paper. Nevertheless, in many cases the difference between the minimum sizes predicted by the two theories is negligible so long as the same anode lens correction is used. The extent to which the two theories agree ob-

viously depends on the magnitude of r_e/σ. When r_e/σ as calculated by the Hines-Cutler method (with a lens correction added) remains greater than about 2 throughout the range of interest, the difference between the corresponding values obtained for r_{95} will be only a few per cent. For these cases where r_e/σ does not get too small, the principal advantages of this paper are in the inclusion of a correction to the anode lens formula and in the comparative ease with which design parameters may be obtained. In other cases r_e/σ may become less than 1, and the theory presented in this paper has extended the basic Hines-Cutler approach so that one may make realistic predictions even under these less ideal conditions where the departure from a laminar-type flow is quite severe.

ACKNOWLEDGMENT

We wish to thank members of the Mathematical Department at B.T.L., particularly H. T. O'Neil and Mrs. L. R. Lee, for their help in programming the problem on the analog computer and in obtaining the large amount of computer data involved. In addition, we wish to thank J. C. Irwin for his help in the electrolytic tank work and both Mr. Irwin and W. A. L. Warne for their work on the beam analyzer.

REFERENCES

1. Pierce, J. R., Rectilinear Flow in Beams, J. App. Phys., **11**, pp. 548–554, Aug., 1940.
2. Samuel, A. L., Some Notes on the Design of Electron Guns, Proc. I.R.E., **33**, pp. 233–241, April, 1945.
3. Field, L. M., High Current Electron Guns, Rev. Mod. Phys., **18**, pp. 353–361, July, 1946.
4. Davisson, C. J., and Calbick, C. J., Electron Lenses, Phys. Rev., **42**, p. 580, Nov., 1932.
5. Helm, R., Spangenburg, K., and Field, L. M., Cathode-Design Procedure for Electron Beam Tubes, Elec. Comm., **24**, pp. 101–107, March, 1947.
6. Cutler, C. C., and Hines, M. E., Thermal Velocity Effects in Electron Guns, Proc. I.R.E., **43**, pp. 307–314, March, 1955.
7. Cutler, C. C., and Saloom, J. A., Pin-hole Camera Investigation of Electron Beams, Proc. I.R.E., **43**, pp. 299–306, March, 1955.
8. Hines, M. E., Manuscript in preparation.
9. Private communication.
10. See for example, Zworykin, V. K., et al., Electron Optics and the Electron Microscope, Chapter 13, Wiley and Sons, 1945, or Klemperer, O., Electron Optics, Chapter 4, Cambridge Univ. Press, 1953.
11. Brown, K. L., and Süsskind, C., The Effect of the Anode Aperature on Potential Distribution in a "Pierce" Electron Gun, Proc. I.R.E., **42**, p. 598, March, 1954.
12. See, for example, Pierce, J. R., Theory and Design of Electron Beams, p. 147, Van Nostrand Co., 1949.
13. See Reference 6, p. 5.
14. Langmuir, I. L., and Blodgett, K., Currents Limited by Space Charge Between Concentric Spheres, Phys. Rev., **24**, p. 53, July, 1924.
15. See Reference 12, p. 177.
16. See Reference 12, Chap. X.

Theories for Toll Traffic Engineering in the U. S. A.*

By ROGER I. WILKINSON

(Manuscript received June 2, 1955)

Present toll trunk traffic engineering practices in the United States are reviewed, and various congestion formulas compared with data obtained on long distance traffic. Customer habits upon meeting busy channels are noted and a theory developed describing the probable result of permitting subscribers to have direct dialing access to high delay toll trunk groups.

Continent-wide automatic alternate routing plans are described briefly, in which near no-delay service will permit direct customer dialing. The presence of non-random overflow traffic from high usage groups complicates the estimation of correct quantities of alternate paths. Present methods of solving graded multiple problems are reviewed and found unadaptable to the variety of trunking arrangements occurring in the toll plan.

Evidence is given that the principal fluctuation characteristics of overflow-type of non-random traffic are described by their mean and variance. An approximate probability distribution of simultaneous calls for this kind of non-random traffic is developed, and found to agree satisfactorily with theoretical overflow distributions and those seen in traffic simulations.

A method is devised using "equivalent random" traffic, which has good loss predictive ability under the "lost calls cleared" assumption, for a diverse field of alternate route trunking arrangements. Loss comparisons are made with traffic simulation results and with observations in exchanges.

Working curves are presented by which multi-alternate route trunking systems can be laid out to meet economic and grade of service criteria. Examples of their application are given.

TABLE OF CONTENTS

* Presented at the First International Congress on the Application of the Theory of Probability in Telephone Engineering and Administration, Copenhagen, June 21, 1955.

1. INTRODUCTION

It has long been the stated aim of the Bell System to make it easily and economically possible for any telephone customer in the United States to reach any other telephone in the world. The principal effort in this direction by the American Telephone and Telegraph Company and its associated operating companies is, of course, confined to interconnecting the telephones in the United States, and to providing communication channels between North America and the other countries of the world. Since the United States is some 1500 miles from north to south and 3000 miles from east to west, to realize even the aim of fast

and economical service between customers is a problem of great magnitude; it has engaged our planning engineers for many years.

There are now 52 million telephones in the United States, over 80 per cent of which are equipped with dials. Until quite recently most telephone users were limited in their direct dialing to the local or immediately surrounding areas and long distance operators were obliged to build up a circuit with the aid of a "through" operator at each switching point.

Both speed and economy dictated the automatic build-up of long toll circuits without the intervention of more than the originating toll operator. The development of the No. 4-type toll crossbar switching system with its ability to accept, translate, and pass on the necessary digits (or equivalent information) to the distant office made this method of operation possible and feasible. It was introduced during World War II, and now by means of it and allied equipment, 55 per cent of all long distance calls (over 25 miles) are completed by the originating operator.

As more elaborate switching and charge-recording arrangements were developed, particularly in metropolitan areas, the distances which customers themselves might dial measurably increased. This expansion of the local dialing area was found to be both economical and pleasing to the users. It was then not too great an effort to visualize customers dialing to all other telephones in the United States and neighboring countries, and perhaps ultimately across the sea.

The physical accomplishment of nationwide direct distance dialing which is now gradually being introduced has involved, as may well be imagined, an immense amount of advance study and fundamental planning. Adequate transmission and signalling with up to eight intertoll trunks in tandem, a nationwide uniform numbering plan simple enough to be used accurately and easily by the ordinary telephone caller, provision for automatic recording of who called whom and how long he talked, with subsequent automatic message accounting, are a few of many problems which have required solution. How they are being met is a romantic story beyond the scope of the present paper. The references given in the bibliography at the end contain much of the history as well as the plans for the future.

2. PRESENT TOLL TRAFFIC ENGINEERING PRACTICE

There are today approximately 116,000 intertoll trunks (over 25 miles in length) in the Bell System, apportioned among some 13,000 trunk groups. A small segment of the 2,600 toll centers which they interconnect is shown in Fig. 1. Most of these intertoll groups are presently traffic engineered to operate according to one of several so-called T-schedules: T-8, T-15, T-30, T-60, or T-120. The number following T (T for Toll) is

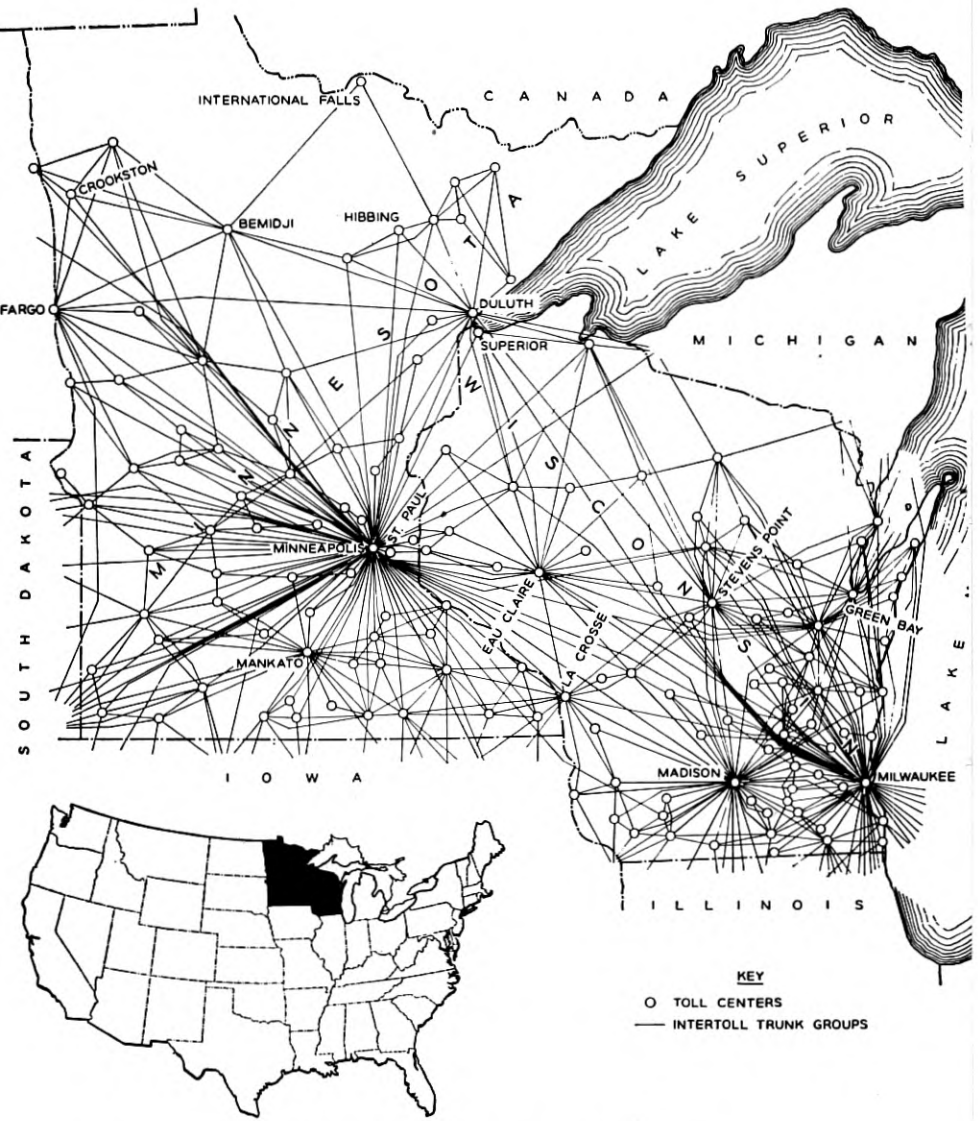


Fig. 1 — Principal intertoll trunk groups in Minnesota and Wisconsin.

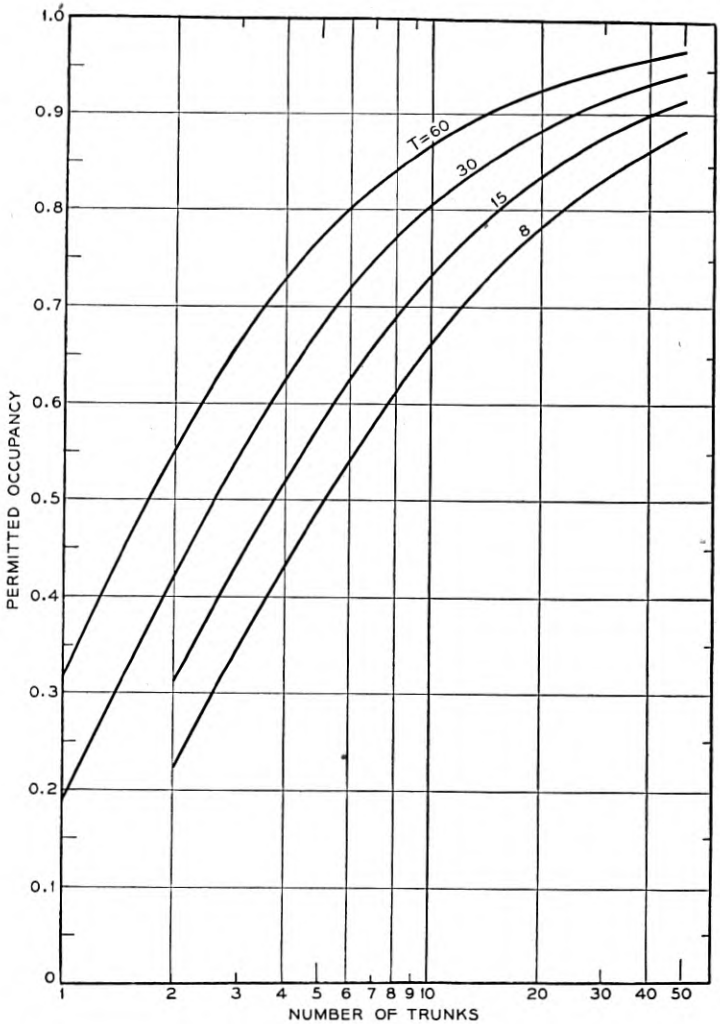


Fig. 2 — Permitted intertoll trunk occupancy for a 6.5-minute usage time per message.

the expected, or average, delay in seconds for calls to obtain an idle trunk in that group during the average Busy Season Busy Hour. In 1954 the system "average trunk speed" was approximately 30 seconds, resulting from operating the majority of the groups at a busy-hour trunking efficiency of 75 to 85 per cent in the busy season.

The T-engineering tables show permissible call minutes of use for a

wide range of group sizes, and several selections of message holding times. They were constructed following summarization of many observations of load and resultant average delays on ringdown (non-dial) intertoll trunks.[1] Fig. 2 shows the permissible occupancy (efficiency) of various trunk group sizes for 6.5 minutes of use per message, for a variety of T-schedules. It is perhaps of some interest that the best fitting curves relating average delay and load were found to be the well-known Pollaczek-Crommelin delay curves for constant holding time — this in spite of the fact that the circuit holding times were far indeed from having a constant value.

A second, and probably not uncorrelated, observation was that the per cent "No-Circuit" (NC) reported on the operators' tickets showed consistently lower values than were measured on group-busy timing devices. Although not thoroughly documented, this disparity has generally been attributed to the reluctance of an operator to admit immediately the presence of an NC condition. She exhibits a certain tolerance (very difficult to measure) before actually recording a delay which would require her to adopt a prescribed procedure for the subsequent handling of the call.* There are then two measures of the No-Circuit condition which are of some interest, the "NC encountered" by operators, and the "NC existing" as measured by timing devices.

It has long been observed that the distribution of numbers n of simultaneous calls found on T-engineered ringdown intertoll groups is in remarkable agreement with the individual probability terms of the Erlang "lost calls" formula,

$$f(n) = \frac{\dfrac{a'^n e^{-a'}}{n!}}{\sum_{n=0}^{c} \dfrac{a'^n e^{-a'}}{n!}} \tag{1}$$

where c = number of paths in the group,

a' = an enhanced average load submitted such that $a'[1 - E_{1,c}(a')] = L$, the actual load carried, and

$E_{1,c}(a') = f(c)$ = Erlang loss probability (commonly called Erlang B in America).

An example of the agreement of observations with (1) is shown in Fig. 3, where the results of switch counts made some years ago on many ringdown circuit groups of size 3 are summarized. A wide range of "sub-

* Upon finding No-Circuit, an operator is instructed to try again in 30 seconds and 60 seconds (before giving an NC report to the customer), followed by additional attempts 5 minutes and 10 minutes later if necessary.

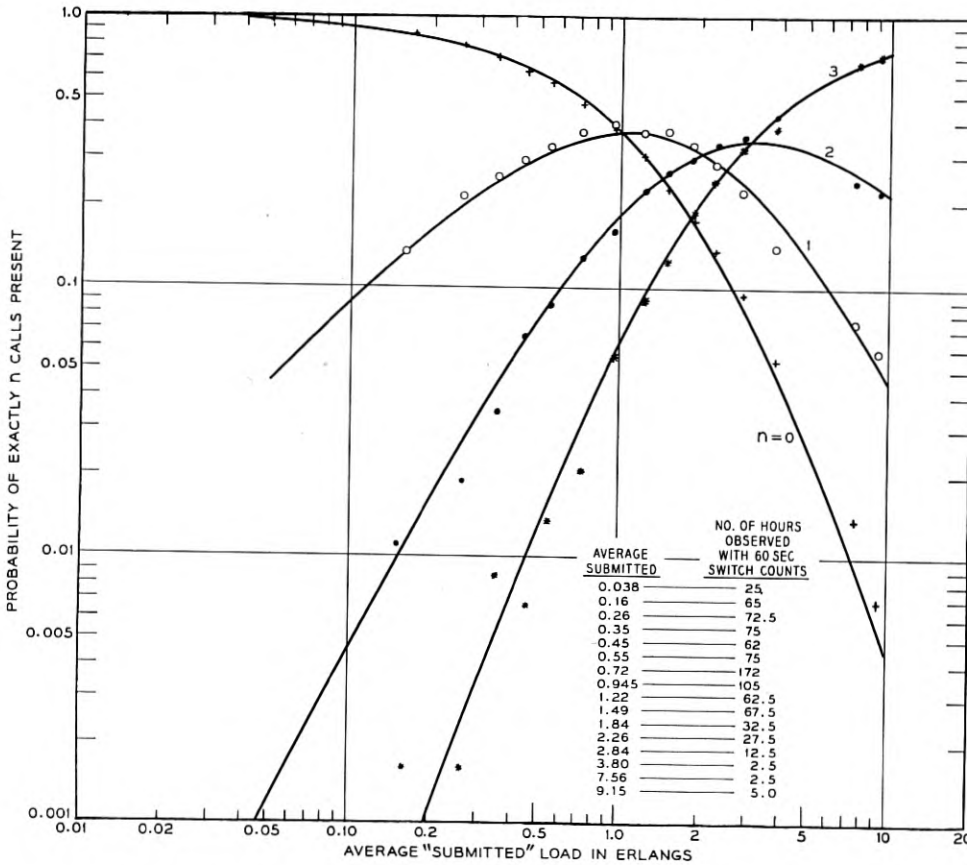


Fig. 3 — Distributions of simultaneous calls on three-trunk toll groups at Albany and Buffalo.

mitted" loads a' to produce the observed carried loads is required. On Fig. 4 are shown the corresponding comparisons of theory and observations for the proportions of time all paths are busy ("NC Existing") for 2-, 4-, 5-, 7-, and 9-circuit groups. Good agreement has also been observed for circuit groups up to 20 trunks. This has been found to be a stable relationship, in spite of the considerable variation in the actual practices in ringdown operation on the resubmission of delayed calls. Since the estimation of traffic loads and the subsequent administration of ringdown toll trunks has been performed principally by means of Group Busy Timers (which cumulate the duration of NC time), the Erlang relationship just described has been of great importance.

With the recent rapid increase in operator dialed intertoll groups, it might be expected that the above discrepancy between "% NC encountered" and "% NC existing" would disappear — for an operator now initiates each call unaware of the momentary state of the load on any particular intertoll group. By the use of peg count meters (which count calls offered) and overflow call counters, this change has in fact been observed to occur. Moreover, since the initial re-trial intervals are commonly fairly short (30 seconds) subsequent attempts tend to find some of the previous congestion still existing, so that the ratio of overflow to peg count readings now *exceeds* slightly the "% NC existing." This situation is illustrated in Fig. 5, which shows data taken on an operator-

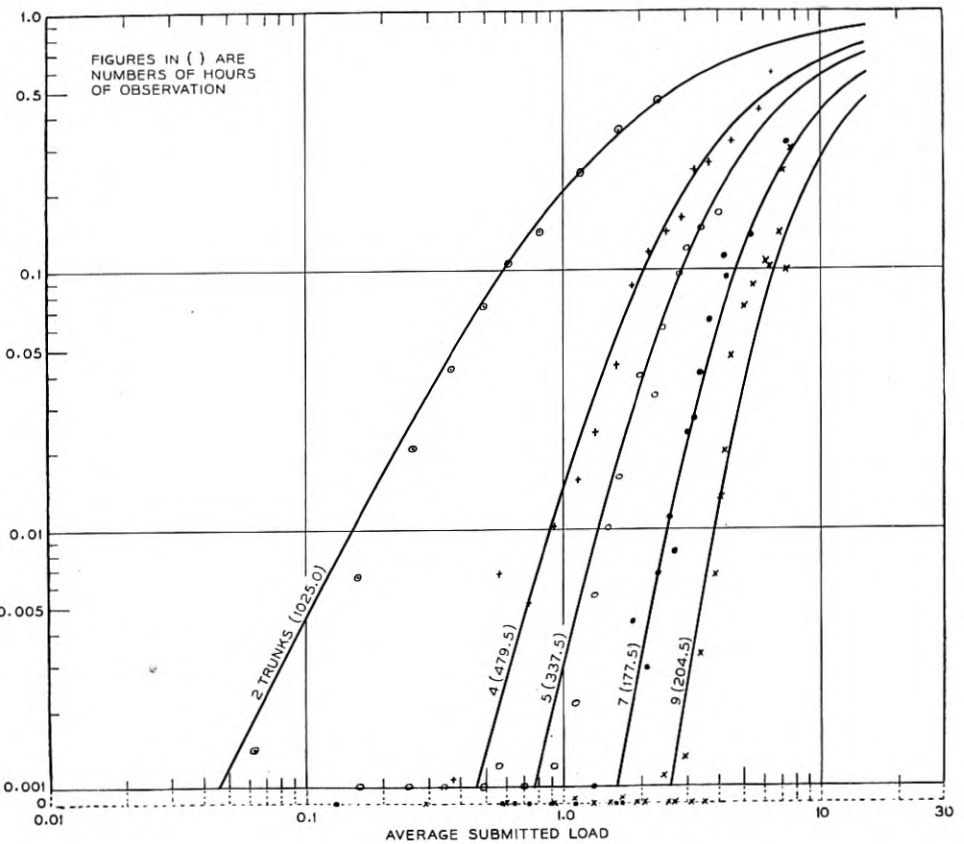


Fig. 4 — Observed proportions of time all trunks were busy on Albany and Buffalo groups of 2, 4, 5, 7, and 9 trunks.

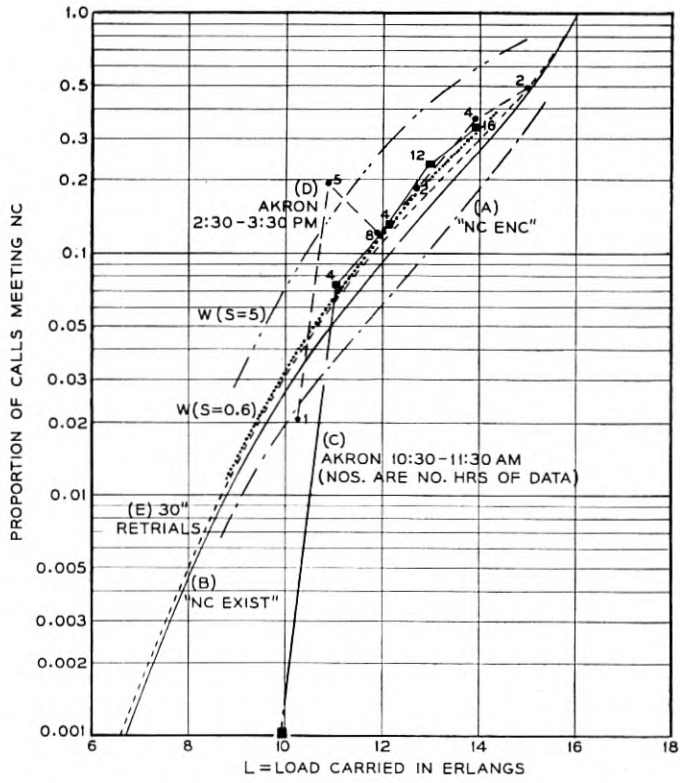


Fig. 5 — Comparison of NC data on a 16-trunk T-engineered toll group with various load versus NC theories.

dialed T-engineered group of 16 trunks between Newark, N. J., and Akron, Ohio. Curve A shows the empirically determined "NC encountered" relationship described above for ringdown operation; Curve B gives the corresponding theoretical "NC existing" values. Lines C and D give the operator-dialing results, for morning and afternoon busy hours. The observed points are now seen generally to be significantly above Curve B.*

At the same time as this change in the "NC encountered" was occurring, due to the introduction of operator toll dialing, there seems to have been little disturbance to the traditional relationship between load

* The observed point at 11 erlangs which is clearly far out of agreement with the remainder of the data was produced by a combination of high-trend hours and an hour in which an operator apparently made many re-trials in rapid succession.

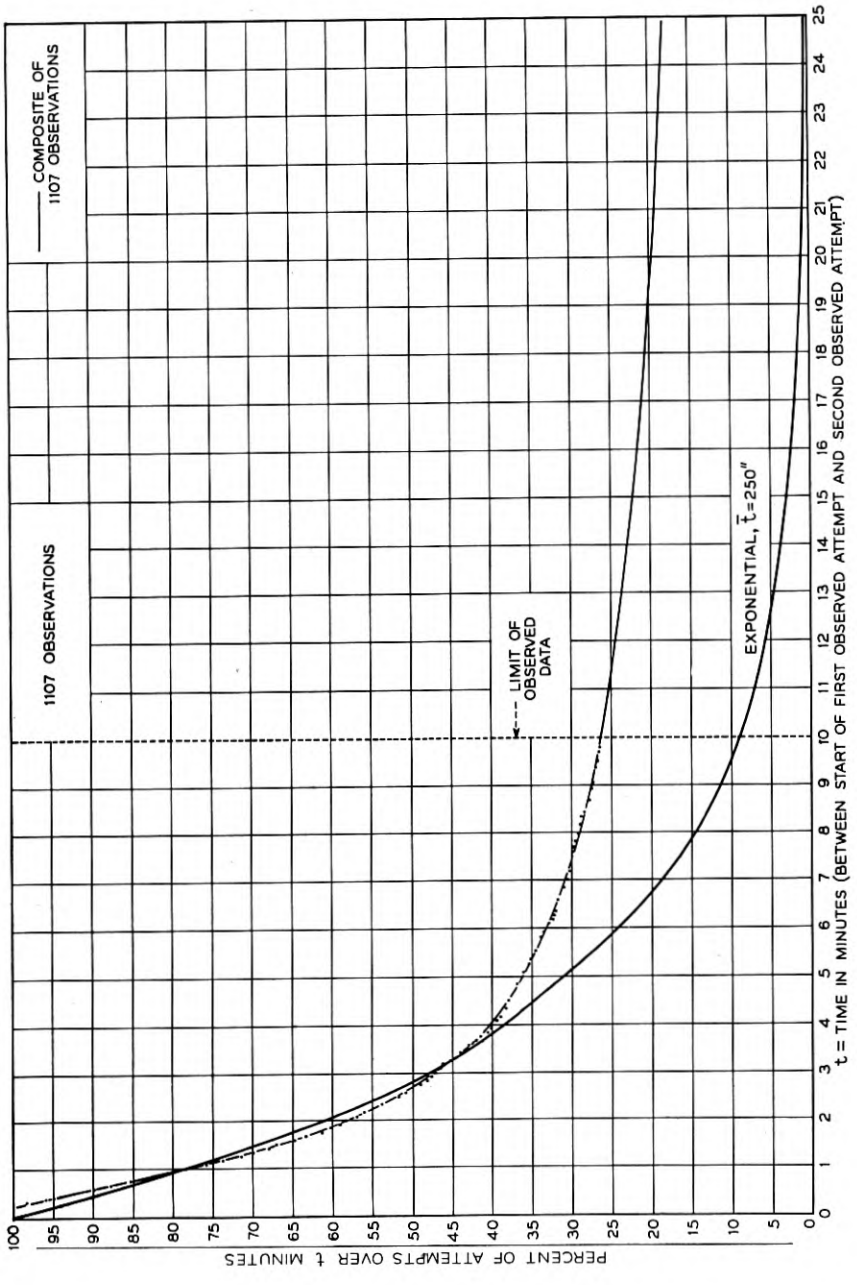


Fig. 6 — Time to second attempt after a subscriber receives a busy signal.

carried and "% NC existing." C. J. Truitt of the A.T. & T. Co. studied a number of operator-dialed T-engineered groups at Newark, New Jersey, in 1954 with a traffic usage recorder (TUR) and group-busy timers, and found the relationship of equation (1) still good. (This analysis has not been published.)

A study by Dr. L. Kosten has provided an estimate of the probability that when an NC condition has been found, it will also appear at a time τ later.[2] When this modification is made, the expected load-versus-NC relationship is shown by Curve E on Fig. 5. (The re-trial time here was taken as the operators' nominal 30 seconds; with 150-second circuit-use time the return is 0.2 holding time.) The observed NC's are seen to lie slightly above the E-curve. This could be explained either on the basis that Kosten's analysis is a lower limit, or that the operators did not strictly observe the 30-second return schedule, or, more probably, a combination of both.

3. CUSTOMERS DIALING ON GROUPS WITH CONSIDERABLE DELAY

It is not to be expected that customers could generally be persuaded to wait a designated constant or minimum re-trial time on their calls which meet the NC condition. Little actual experience has been accumulated on customers dialing long distance calls on high-delay circuits. However, it is plausible that they would follow the re-trial time distributions of customers making local calls, who encounter paths-busy or line-busy signals (between which they apparently do not usually distinguish). Some information on re-trial times was assembled in 1944 by C. Clos[3] by observing the action of customers who received the busy signal on 1,100 local calls in the City of New York. As seen in Fig. 6, the return times, after meeting "busy," exhibit a marked tendency toward the exponential distribution, after allowance for a minimum interval required for re-dialing.

An exponential distribution with average of 250 seconds has been fitted by eye on Fig. 6, to the earlier — and more critical — customer return times. This may seem an unexpectedly long wait in the light of individual experience; however it is probably a fair estimate, especially since, following the collection of the above data, it has become common practice for American operating companies in their instructional literature to advise customers receiving the busy signal to "hang up, wait a few minutes, and try again."

The mathematical representation of the situation assuming exponential return times is easily formulated. Let there be x actual trunks, and

imagine y waiting positions, where y is so large that few calls are rejected.* Assume that the offered load is a erlangs, and that the calls have exponential conversation holding times of unit average duration. Finally let the average return time for calls which have advanced to the waiting positions, be $1/s$ times that of the unit conversation time. The statistical equilibrium equation can then be written for the probability $f(m, n)$ that m calls are in progress on the x trunks and n calls are waiting on the y storage positions:

$$\begin{aligned} f(m, n) = {} & af(m-1, n)\,dt + s(n+1)f(m-1, n+1)\,dt \\ & + (m+1)f(m+1, n)\,dt + af(x, n-1)\,dt\star \\ & + [1 - (a\star\star\star + sn\star\star)\,dt - m\,dt]f(m, n) \end{aligned} \tag{2}$$

where $0 \leqq m \leqq x$, $0 \leqq n \leqq y$, and the special limiting situations are recognized by:

$\star$ Include term only when $m = x$

$\star\star$ Omit sn when $m = x$

$\star\star\star$ Omit a when $m = x$ and $n = y$

Equation (2) reduces to

$$\begin{aligned} (a\star\star\star + sn\star\star + m)f(m, n) = {} & af(m-1, n) \\ & + s(n+1)f(m-1, n+1) \\ & + (m+1)f(m+1, n) + af(x, n-1)\star \end{aligned} \tag{3}$$

Solution of (3) is most easily effected for moderate values of x and y by first setting $f(x, y) = 1.000000$ and solving for all other $f(m, n)$ in terms of $f(x, y)$. Normalizing through $\sum_{m=0}^{x}\sum_{n=0}^{y} f(m, n) = 1.0$, then gives the entire $f(m, n)$ array.

The proportion of time "NC exists," will, of course be

$$\sum_{n=0}^{y} f(x, n) \tag{4}$$

and the load carried is

$$L = \sum_{m=0}^{x}\sum_{n=0}^{y} mf(m, n) \tag{5}$$

The proportion of call attempts meeting NC, including all re-trials

* The quantity y can also be chosen so that some calls are rejected, thus roughly describing those calls abandoned after the first attempt.

will be

$$W(x, a, s) = \frac{\text{Expected overflow calls per unit time}}{\text{Expected calls offered per unit time}}$$

$$= \frac{\sum_{n=0}^{y} (a + sn) f(x, n)}{\sum_{m=0}^{x} \sum_{n=0}^{y} (a + sn) f(m, n)} = \frac{s\bar{n} + af(x, y)}{a + s\bar{n}} \tag{6}$$

in which $\bar{n} = \sum_{m=0}^{x} \sum_{n=0}^{y} nf(m, n)$. And when y is chosen so large that $f(x, y)$ is negligible, as we shall use it here,

$$L = a \tag{5'}$$

$$W(x, a, s) = \frac{s\bar{n}}{a + s\bar{n}} \tag{6'}$$

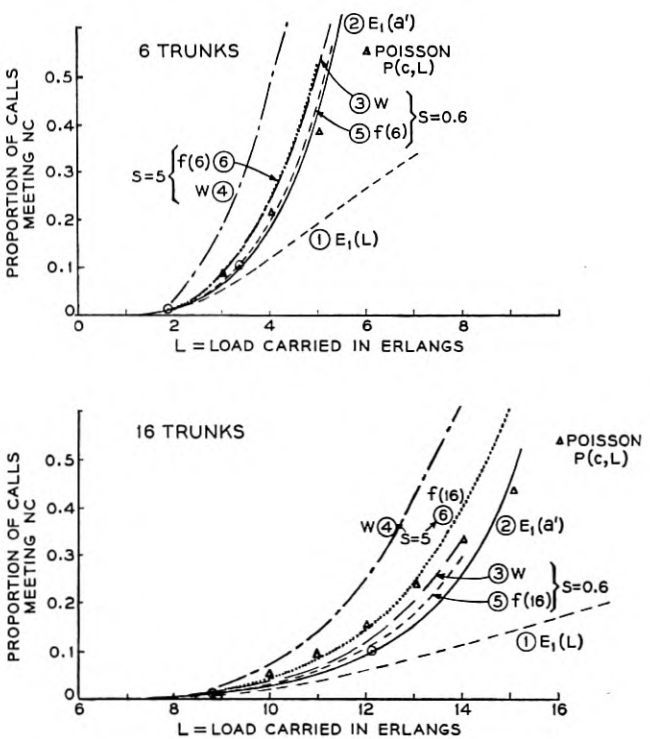


Fig. 7 — Comparison of trunking formulas.

This formula provides a means for estimating the grade of service which customers might be expected to receive if asked to dial their calls over moderate-delay or high-delay trunk groups. For a circuit use length of 150 seconds, and an average return time of 250 seconds (as on Fig. 6), both exponential, the load-versus-proportion-NC curves for 6 and 16 trunks are given as curves (3) on Fig. 7. For example with an offered (= carried) load of a = 4.15 erlangs on 6 trunks we should expect to find 27.5 per cent of the total attempts resulting in failure.

For comparison with a fixed return time of NC-calls, the W-formula curves for exponential returns of 30 seconds ($s = 5$) and 250 seconds ($s = 0.6$) averages are shown on Fig. 5. The first is far too severe an assumption for operator performance, giving NC's nearly double those actually observed (and those given by theory for a 30-second constant return time). The 250-second average return, however, lies only slightly above the 30-second constant return curve and is in good agreement with the data. Although not logically an adequate formula for interpreting Peg Count and Overflow registrations on T-engineered groups under operator dialing conditions, the W-formula apparently could be used for this purpose with suitable s-values determined empirically.

3.1. *Comparison of Some Formulas for Estimating Customers' NC Service on Congested Groups*

As has been previously observed, a large proportion of customers who receive a busy signal, return within a few minutes (on Fig. 6, 75 per cent of the customers returned within 10 minutes). It is well known too, that under adverse service conditions subscriber attempts (to reach a particular distant office for example) tend to produce an inflated estimate of the true offered load. A count of calls carried (or a direct measurement of load carried) will commonly be a closer estimate of the offered load than a count of attempts. An exception may occur when a large proportion of attempts is lost, indicating an offered load possibly in excess even of the number of paths provided. Under the latter condition it is difficult to estimate the true offered load by any method, since not all the attempts can be expected to return repeatedly until served; instead, a significant number will be abandoned somewhere through the trials. In most other circumstances, however, the carried load will prove a reasonably good estimate of the true offered load in systems not provided with alternate paths.

This is a matter of especial interest for both toll and local operation in America since principal future reliance for load measurement is ex-

pected to be placed on automatically processed TUR data, and as the TUR is a switch counting device the results will be in terms of load carried. Moreover, the quantity now obtained in many local exchanges is load carried.* Visual switch counting of line finders and selectors off-normal is widely practiced in step-by-step and panel offices; a variety of electromechanical switch counting devices is also to be found in crossbar offices. It is common to take load-carried figures as equal to load-offered when using conventional trunking tables to ascertain the proper provision of trunks or switches. Fig. 7 compares the NC predictions made by a number of the available load-loss formulas when load carried is used as the entry variable.

The lowest curves (1) on Fig. 7 are from the Erlang lost calls formula E_1 (or B) with load carried L used as the offered load a. At low losses, say 0.01 or less, either L or $a = L/[1 - E_1(a)]$ can be used indiscriminately as the entry in the E_1 formula. If however considerably larger losses are encountered and calls are not in reality "cleared" upon meeting NC, it will no longer be satisfactory to substitute L for a. In this circumstance it is common to calculate a fictitious load a' to submit to the c paths such that the load carried, $a'[1 - E_{1,c}(a')]$, equals the desired L. (This was the process used in Section 2 to obtain "% NC existing.") The curves (2) on Fig. 7 show this relation; physically it corresponds to an initially offered load of L erlangs (or L call arrivals per average holding time), whose overflow calls return again and again until successful but without disturbing the randomness of the input. Thus if the loss from this enhanced random traffic is E, then the total trials seen per holding time will be $L(1 + E + E^2 + \cdots) = L/(1 - E) = a'$, the apparent arrival rate of new calls, but actually of new calls plus return attempts.

The random resubmission of calls may provide a reasonable description of operation under certain circumstances, presumably when re-trials are not excessive. Kosten[2] has discussed the dangers here and provided upper and lower limit formulas and curves for estimating the proportions of NC's to be expected when re-trials are made at any specified fixed return time. His lower bounds (lower bound because the change in congestion character caused by the returning calls is ignored) are shown by open dots on Fig. 7 for return times of 1.67 holding times. They lie above curves (2) (although only very slightly because of the relatively long return time) since they allow for the fact that a call shortly returning

* In fact, it is difficult to see how any estimate of offered load, other than carried load, can be obtained with useful reliability.

after meeting a busy signal will have a higher probability of again finding all paths busy, than would a randomly originated call.

The curves (3) show the W-formula previously developed in this section, which contemplates exponential return times on all NC attempts. The average return time here is also taken as 1.67 holding times. These curves lie higher than Kosten's values for two reasons. First, the altered congestion due to return calls is allowed for; and second, with exponential returns nearly two-thirds of the return times are shorter than the average, and of these, the shortest ones will have a relatively high probability of failure upon re-trying. If the customers were to return with exponential times after waiting an average of only 0.2 holding time (e.g., 30 seconds wait for 150-second calls) the W-curves would rise markedly to the positions shown by (4).

Curves (5) and (6) give the proportions of time that all paths are busy (equation 4) under the W-formula assumptions corresponding to NC curves (3) and (4) respectively; their upward displacement from the random return curves (2) reflects the disturbance to the group congestion produced by the non-random return of the delayed calls. (The limiting position for these curves is, of course, given by Erlang's E_2 (or C) delay formula.) As would be expected, curve (6) is above (5) since the former contemplates exponential returns with average of 0.2 holding time, as against 1.67 for curve (5). Neither the (5)-curves nor the open dots of constant 30-second return times show a marked increase over curves (2). This appears to explain why the relationship of load carried versus "NC existing" (as charted in Figs. 3 and 4) was found so insensitive to variable operating procedures in handling subsequent attempts in toll ringdown operation, and again, why it did not appreciably change under operator dialing.

Finally, through the two fields of curves on Fig. 7 is indicated the Poisson summation $P(c, L)$ with load carried L used as the entering variable. The fact that these values approach closely the (2) and (3) sets of curves over a considerable range of NC's should reassure those who have been concerned that the Poisson engineering tables were not useful for losses larger than a few per cent.*

4. SERVICE REQUIREMENTS FOR DIRECT DISTANCE DIALING BY CUSTOMERS

As shown by the W-curves (3) on Fig. 7, the attempt failures by customers resulting from their tendency to re-try shortly following an NC

* Reference may be made also to a throwdown by C. Clos (Ref. 3) using the return times of Fig. 6; his "% NC" results agreed closely with the Poisson predictions.

would be expected to exceed slightly the values for completely random re-trials. These particular curves are based on a re-trial interval of 1.67 times the average circuit-use time. Such moderation on the part of the customer is probably attainable through instructional literature and other means if the customer believes the "NC" or "busy" to be caused by the called party's actually using his telephone (the usual case in local practice). It would be considerably more difficult, however, to dissuade the customer from re-trying at a more rapid rate if the circuit NC's should generally approach or exceed actual called-party busies, a condition of which he would sooner or later become aware. His attempts might then be more nearly described by the (4) curves on Fig. 7 corresponding to an average exponential return of only 0.2 holding time—or even higher. Such a result would not only displease the user, but also result in the requirement of increased switching control equipment to handle many more wasted attempts.

If subscribers are to be given satisfactory direct dialing access to the intertoll trunk network, it appears then that the probability of finding NC even in the busy hours must be kept to a low figure. The following engineering objective has tentatively been selected: *The calls offered to the "final" group of trunks in an alternate route system should receive no more than 3 per cent* NC(P.03) *during the network busy season busy hour.* (If there are no alternate routes, the direct group is the "final" route.)

Since in the nationwide plan there will be a final route between each of some 2,600 toll centers and its next higher center, and the majority of calls offered to high usage trunks will be carried without trying their final route (or routes), the over-all point-to-point service, while not easy to estimate, will apparently be quite satisfactory for customer dialing.

5. ECONOMICS OF TOLL ALTERNATE ROUTING

In a general study of the economics of a nationwide toll switching plan, made some years ago by engineers of the American Telephone and Telegraph Company, it was concluded that a toll line plant sufficient to give the then average level of service (about T-40) with ordinary single-route procedures could, if operated on a multi-alternate route basis, give the desired P.03 service on final routes with little, if any, increase in toll line investment.* On the other hand to attain a similar P.03 grade of service by liberalizing a typical intertoll group of 10 trunks working presently

* This, of course, does not reflect the added costs of the No. 4 switching equipment.

at a T-40 grade of service and an occupancy of 0.81 would require an increase of 43 per cent (to 14.3 trunks), with a corresponding decrease in occupancy to 0.57. The possible savings in toll lines with alternate routing are therefore considerable in a system which must provide a service level satisfactory for customer dialing.

In order to take fullest advantage of the economies of alternate routing, present plans call for five classes of toll offices. There will be a large number of so-called End Offices, a smaller number of Toll Centers, and progressively fewer Primary Centers (about 150), Sectional Centers (about 40) and Regional Centers (9), one of which will be the National Center, to be used as the "home" switching point of the other eight Regional Centers.* Primary and higher centers will be arranged to perform automatic alternate routing and are called Control Switching Points (CSP's). Each class of office will "home" on a higher class of office (not necessarily the next higher one); the toll paths between them are called "final routes." As described in Section 4, these final routes will be provided to give low delays, so that between each principal toll point and every other one there will be available a succession of approximately P.03 engineered trunk groups. Thus if the more direct and heavily loaded interconnecting paths commonly provided are busy there will still be a good chance of making immediate connection over final routes.

Fig. 8 illustrates the manner in which automatic alternate routing will operate in comparison with present-day operator routing. On a call from Syracuse, N. Y., to Miami, Florida, (a distance of some 1,250 miles), under present-day operation, the Syracuse operator signals Albany, and requests a trunk to Miami. With T-schedule operation the Syracuse-Miami traffic might be expected to encounter as much as 25 per cent NC during the busy hour, and approximately 4 per cent NC for the whole day, producing perhaps a two-minute over-all speed of service in the busy season.

With the proposed automatic alternate routing plan, all points on the chart will have automatic switching systems.† The customer (or the operator until customer dialing arrangements are completed) will dial a ten-digit code (three-digit area code 305 for Florida plus the listed Miami seven-digit telephone number) into the machine at Syracuse. The various routes which then might conceivably be tried automatically

* See the bibliography (particularly Pilliod and Truitt) for details of the general trunking plan.

† The notation used on the diagram of Fig. 8 is: Open circle — Primary Center (Syracuse, Miami); Triangle — Sectional Center (Albany, Jacksonville); Square — Regional Center (White Plains, Atlanta, St. Louis; St. Louis is also the National Center).

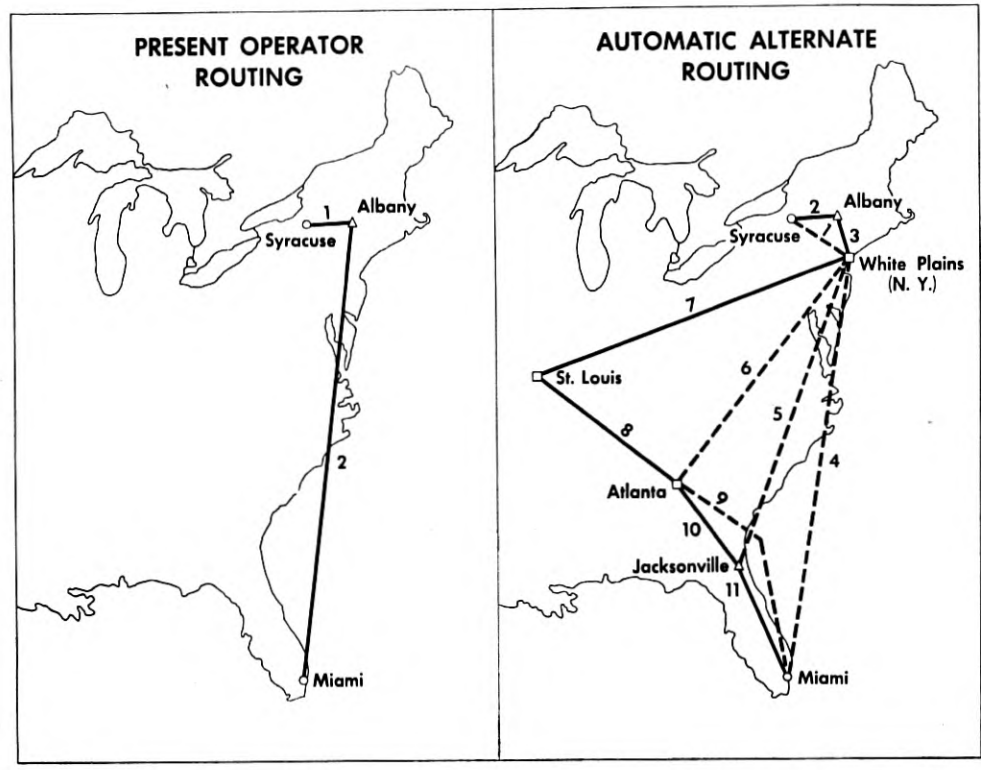


Fig. 8 — Present and proposed methods of handling a call from Syracuse, N. Y., to Miami, Florida.

are shown on the diagram numbered in the order of trial; in this particular layout shown, a maximum of eleven circuit groups could be tested for an idle path if each high usage group should be found NC. Dotted lines show the high usage routes, which if found busy will overflow to the final groups represented by solid lines. The switching equipment at each point upon finding an idle circuit passes on the required digits to the next machine.

While the routing possibilities shown are factual, only in rare instances would a call be completed over the final route via St. Louis. Even in the busy season busy hour just a small portion of the calls would be expected to be switched as many as three times. And only a fraction of one per cent of all calls in the busy hour should encounter NC. As a result the service will be fast. When calls are handled by a toll operator, the cus-

tomer will not ordinarily need to hang up when NC is obtained. When he himself dials, a second trial after a short wait following NC should have a high probability of success.

Not many situations will be as complex as shown in Fig. 8; commonly several of the links between centers will be missing, the particular ones retained having been chosen from suitable economic studies. A large number of switching arrangements will be no more involved than the illustrative one shown in Fig. 9(a), centering on the Toll Center of Bloomsburg, Pennsylvania. The dashed lines indicate high usage groups from Bloomsburg to surrounding toll centers; since Bloomsburg "homes" on Scranton this is a final route as denoted by the solid line. As an example of the operation, consider a call at Bloomsburg destined for Williamsport. Upon finding all direct trunks busy, a second trial is made via Harrisburg; and should no paths in the Harrisburg group be available, a third and final trial is made through the Scranton group.

In considering the traffic flow of a network such as illustrated at Bloomsburg it is convenient to employ the conventional form of a two-stage graded multiple having "legs" of varying sizes and traffic loads individual to each, as shown in Fig. 9(b). Here only the circuits immediately outgoing from the toll center are shown; the parcels of traffic

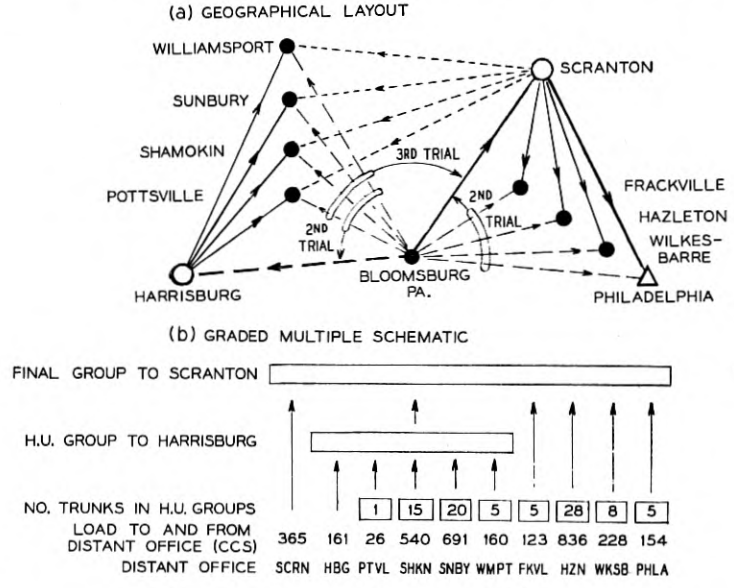


Fig. 9 — Automatic alternate routing for direct distance dialing at Bloomsburg, Pa.

calculated for each further connecting route will be recorded as part of the offered load for consideration when the next higher switching center is engineered. It is implicitly assumed that a call which has selected one of the alternate route paths will be successful in finding the necessary paths available from the distant switching point onward. This is not quite true but is believed generally to be close enough for engineering purposes, and permits ignoring the return attempt problem.

6. NEW PROBLEMS IN THE ENGINEERING AND ADMINISTRATION OF INTERTOLL GROUPS RESULTING FROM ALTERNATE ROUTING

With the greatly increased teamwork among groups of intertoll trunks which supply overflow calls to an alternate route, an unexpected increase or flurry in the offered load to any one can adversely affect the service to all. The high efficiency of the alternate route networks also reduces their overload carrying ability. Conversely, the influence of an underprovision of paths in the final alternate route may be felt by many groups which overflow to it. With non-alternate route arrangements only the single groups having these flurries would be affected.

Administratively, an alternate route trunk layout may well prove easier to monitor day by day than a large number of separate and independent intertoll groups, since a close check on the service given on the final routes only may be sufficient to insure that all customers are being served satisfactorily. When rearrangements are indicated, how-

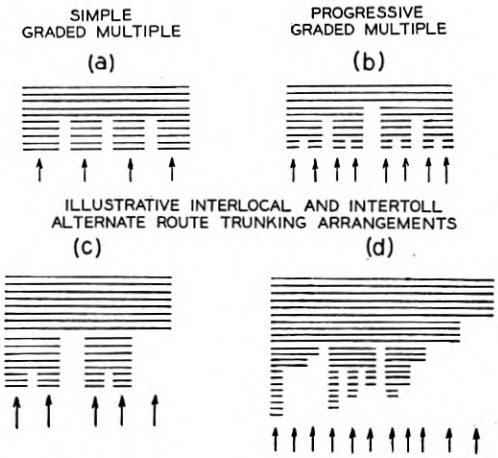


Fig. 10 — Graded multiples and alternate route trunking arrangements.

ever, the determination of the proper place to take action, and the desirable extent, may sometimes be difficult to determine. Suitable traffic measuring devices must be provided with these latter problems in mind.

For engineering purposes, it will be highly desirable:

(1) To be able to estimate the load-service relationships with any specified loads offered to a particular intertoll alternate routing network; and

(2) To know the day-to-day busy hour variations in the various groups' offered loads during the busy season, so that the general grade of service given to customers can be estimated.

The balance of this paper will review the studies which have been made in the Bell System toward a practicable method for predicting the grade of service given in an alternate route network under any given loads. Analyses of the day-to-day load variations and their effects on customer dialing service are currently being made, and will be reported upon later.

7. LOAD-SERVICE RELATIONSHIPS IN ALTERNATE ROUTE SYSTEMS

In their simplest form, alternate route systems appear as symmetrical graded multiples, as shown in Fig. 10(a) and 10(b). Patterns such as these have long been used in local automatic systems to partially overcome the trunking efficiency limitations imposed by limited access switches. The traffic capacity of these arrangements has been the subject of much study by theory and "throwdowns" (simulated traffic studies) both in the United States and abroad. Field trials have substantiated the essential accuracy of the trunking tables which have resulted.

In toll alternate route systems as contemplated in America, however, there will seldom be the symmetry of pattern found in local graded multiples, nor does maximum switch size generally produce serious limitation on the access. The "legs" or first-choice trunk groups will vary widely in size; likewise the number of such groups overflowing calls jointly to an alternate route may cover a considerable range. In all cases a given group, whether or not a link of an alternate route, will have one or more parcels of traffic for which it is the first-choice route. [See the right-hand parcel of offered traffic on Fig. 10(c).] Often this first routed traffic will be the bulk of the load offered to the group, which also serves as an alternate route for other traffic.

The simplest of the approximate formulas developed for solving the local graded multiple problems are hopelessly unwieldy when applied to such arrangements as shown in Fig. 10(d). Likewise it is impracticable

to solve more than a few of the infinite variety of arrangements by means of "throwdowns."

However, for both engineering (planning for future trunk provisions) and administration (current operating) of trunks in these multi-alternate routing systems, a rapid, simple, but reasonably accurate method is required. The basis for the method which has been evolved for Bell System use will be described in the following pages.

7.1. *The "Peaked" Character of Overflow Traffic*

The difficulty in predicting the load-service relationship in alternate route systems has lain in the non-random character of the traffic overflowing a first set of paths to which calls may have been randomly offered. This non-randomness is a well appreciated phenomenon among traffic engineers. If adequate trunks are provided for accommodating the momentary traffic peaks, the time-call level diagram may appear as in Fig. 11(a), (average level of 9.5 erlangs). If however a more limited number of trunks, say $x = 12$, is provided, the peaks of Fig. 11(a) will be clipped, and the overflow calls will either be "lost" or they may be handled on a subsequent set of paths y. The momentary loads seen on y then appear as in Fig. 11(b). It will readily be seen that a given average load on the y trunks will have quite different fluctuation characteristics than if it had been found on the x trunks. There will be more occurrences of large numbers of calls, and also longer intervals when few or no calls are present. This gives rise to the expression that overflow traffic is "peaked."

Peaked traffic requires more paths than does random traffic to operate at a specified grade of delayed or lost calls service. And the increase in paths required will depend upon the degree of peakedness of the traffic involved. A measure of peakedness of overflow traffic is then required which can be easily determined from a knowledge of the load offered and the number of trunks in the group immediately available.

In 1923, G. W. Kendrick, then with the American Telephone and Telegraph Company, undertook to solve the graded multiple problem through an application of Erlang's statistical equilibrium method. His principal contribution (in an unpublished memorandum) was to set up the equations for describing the existence of calls on a full access group of $x + y$ paths, arranged so that arriving calls always seek service first in the x-group, and then in the y-group when the x are all busy.

Let $f(m, n)$ be the probability that at a random instant m calls exist on the x paths and n calls on the y paths, when an average Poisson load

of a erlangs is submitted to the $x + y$ paths. The general state equation for all possible call arrangements, is

$$(a^* + m + n)f(m, n) = (m + 1)f(m + 1, n) + (n + 1)f(m, n + 1) + af(m - 1, n) + af(x, n - 1)‡ \quad (7)$$

in which the term marked (‡) is to be included only when $m = x$, and * indicates that the a in this term is to be omitted when $m + n = x + y$. m and n may take values only in the intervals, $0 \leqq m \leqq x; 0 \leqq n \leqq y$. As written, the equation represents the "lost calls cleared" situation.

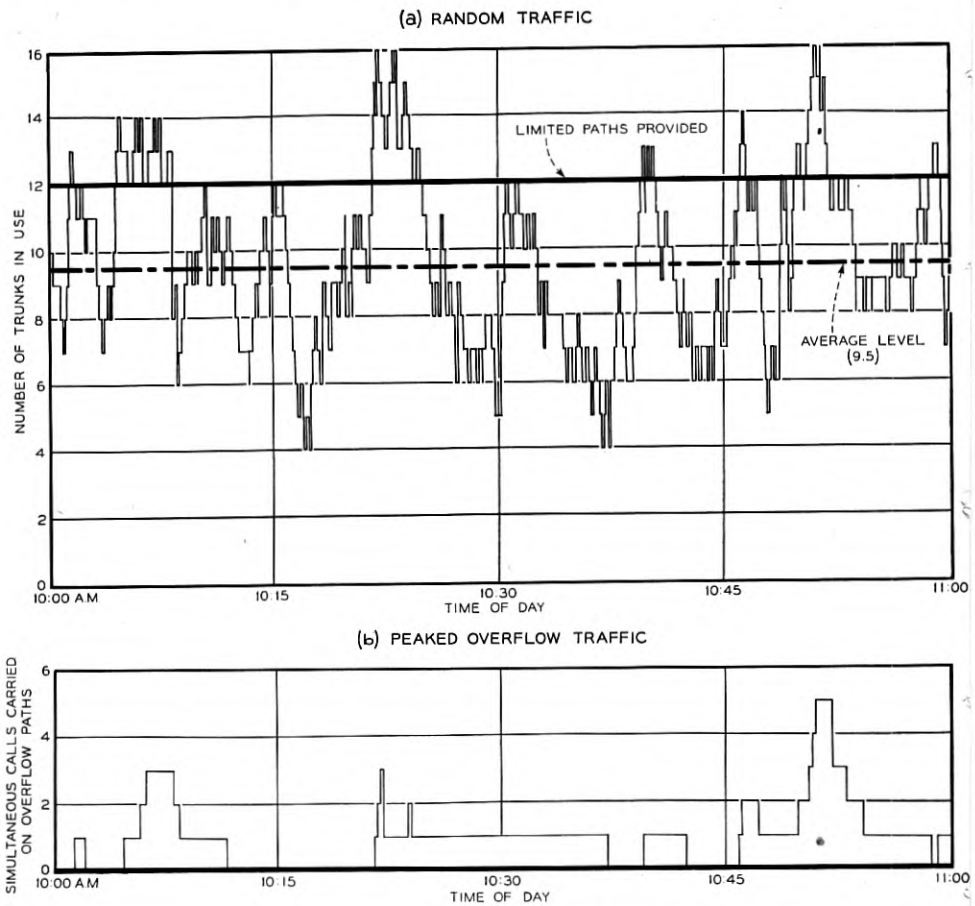


Fig. 11 — Production of peakedness in overflow traffic.

By choosing $x + y$ large compared with the submitted load a a "lost calls held" situation or infinite-overflow-trunks result can be approached as closely as desired.

Kendrick suggested solving the series of simultaneous equations (7) by determinants, and also by a method of continued fractions. However little of this numerical work was actually undertaken until several years later.

Early in 1935 Miss E. V. Wyckoff of Bell Telephone Laboratories became interested in the solution of the $(x + 1)(y + 1)$ lost calls cleared simultaneous equations leading to all terms in the $f(m, n)$ distribution. She devised an order of substituting one equation in the next which provided an entirely practical and relatively rapid means for the numerical solution of almost any set of these equations. By this method a considerable number of $f(m, n)$ distributions on x, y type multiples with varying load levels were calculated.

From the complete m, n matrix of probabilities, one easily obtains the distribution $\theta_m(n)$ of overflow calls when exactly m are present on the lower group of x trunks; or by summing on m, the $\theta(n)$ distribution without regard to m, is realized. A number of other procedures for obtaining the $f(m, n)$ values have been proposed. All involve lengthy computations, very tedious for solution by desk calculating machines, and most do not have the ready checks of the Wyckoff-method available at regular points through the calculations.

In 1937 Kosten[4] gave the following expression for $f(m, n)$:

$$f(m, n) = (-1)^n \varphi_0(x) \sum_{i=0}^{\infty} \binom{i}{n} \frac{(-a)^i}{i!} \cdot \frac{\varphi_i(m)}{\varphi_{i+1}(x)\varphi_i(x)} \tag{8}$$

where

$$\varphi_0(x) = \frac{a^x e^{-a}}{x!}$$

and for $i > 0$,

$$\varphi_i(x) = e^{-a} \sum_{j=0}^{x} \binom{i + j - 1}{j} \frac{a^{x-j}}{(x - j)!}$$

These equations, too, are laborious to calculate if the load and numbers of trunks are not small. It would, of course, be possible to program a modern automatic computer to do this work with considerable rapidity.

The corresponding application of the statistical equilibrium equations to the graded multiple problem was visualized by Kendrick who, however, went only so far as to write out the equation for the three-trunk

case consisting of two subgroups of one trunk each and one common overflow trunk.

Instead of solving the enormously elaborate system of equations describing all the calls which could simultaneously be present in a large multiple, several ingenious methods of convoluting the

$$\theta(n) = \sum_{m=0}^{x} f(m, n)$$

overflow distributions from the individual legs of a graded multiple have been devised. For example, for the multiple of Fig. 10(a), the probability of loss P_i as seen by a call entering subgroup number i, is approximately,

$$P_i = \sum_{r=0}^{y-1} \sum_{z=y}^{\infty} \theta_{x,i}(r) \cdot \psi(z - r) + \sum_{r=y}^{\infty} \theta_{x,i}(r) \tag{9}$$

in which $\psi(z - r)$ is the probability of exactly $z - r$ overflow calls being present, or wanting to be present, on the alternate route from all the subgroups except the ith, and with no regard for the numbers of calls present in these subgroups. The $\theta_{x,i}(r) = f_i(x_i, r)$ term, of course, contemplates all paths in the particular originating call's subgroup being occupied, forcing the new call arriving in subgroup i to advance to the alternate route. This corresponds to the method of solving graded multiples developed by E. C. Molina[6] but has the advantage of overcoming the artificial "no holes in the multiple" assumption which he made. Similar calculating procedures have been suggested by Kosten.* These computational methods doubtless yield useful estimates of the resulting service, and for the limited numbers of multiple arrangements which might occur in within-office switching trains (particularly ones of a symmetrical variety) such procedures might be practicable. But it would be far too laborious to obtain the individual overflow distributions $\theta(n)$, and then convolute them for the large variety of loads and multiple arrangements expected to be met in toll alternate routing.

7.2. *Approximate Description of the Character of Overflow Traffic*

It was natural that various approximate procedures should be tried in the attempt to obtain solutions to the general loss formula sufficiently accurate for engineering and study purposes. The most obvious of these is to calculate the lower moments or semi-invariants of the loads overflowing the subgroups, and from them construct approximate fitting

* Kosten gives the above approximation (9), which he calls W_b^+, as an upper limit to the blocking. He also gives a lower limit, W_b^-, in which $z = y$ throughout (References 4, 5).

distributions for $\theta(n)$ and $\theta_x(n)$. Since each such overflow is independent of the others, they may be combined additively (or convoluted), to obtain the corresponding total distribution of calls appearing before the alternate route (or common group). It may further be possible to obtain an approximate fitting distribution to the sum-distribution of the overflow calls.

The ordinary moments about the 0 point of the subgroup overflow distribution, when m of the x paths are busy, are found by

$$\mu_i'(m) = \sum_{n=0}^{y} n^i f(m, n) \tag{10}$$

When an infinite number of y-paths is assumed, the resulting expressions for the mean and variance are found to be:*

Number of x-paths busy unspecified:†

$$\text{Mean} = \alpha = a \cdot E_{1,x}(a) \tag{11}$$

$$\text{Variance} = v = \alpha[1 - \alpha + a(x + 1 + \alpha - a)^{-1}] \tag{12}$$

All x-paths occupied,

$$\text{Mean} = \alpha_x = a[x - a + 1 + aE_{1,x}(a)]^{-1} \tag{13}$$

$$\text{Variance} = v_x = \alpha_x[1 - \alpha_x + 2a(x + 2 + \alpha_x - a)^{-1}] \tag{14}$$

Equations (11) and (12) have been calculated for considerable ranges of offered load a and paths x. Figs. 12 and 13 are graphs of these results. For example when a load of 4 erlangs is submitted to 5 paths, the average overflow load is seen to be $\alpha = 0.80$ erlang, the same value, of course, as determined through a direct application of the Erlang E_1 formula. During the time that all x paths are busy, however, the overflow load will tend to exceed this general level as indicated by the value of $\alpha_x = 1.41$ erlangs calculated from (13). Similarly the variance of the overflow load will tend to increase when the x-paths are fully occupied,

* The derivation of these equations is given in Appendix I.

† The skewness factor may also be of interest:

$$\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}}$$

$$= \frac{1}{v^{3/2}}\left[\frac{a}{x+1+\alpha-a}\left\{\frac{2}{x+2}\left(\frac{(x+\alpha-a)a^2}{(x-a)^2+2(x-a)+x+2+(x+2-a)\alpha}+a\right)\right.\right. \tag{15}$$

$$\left.\left. + 3(1-\alpha)\right\} + \alpha(1-\alpha)(1-2\alpha)\right]$$

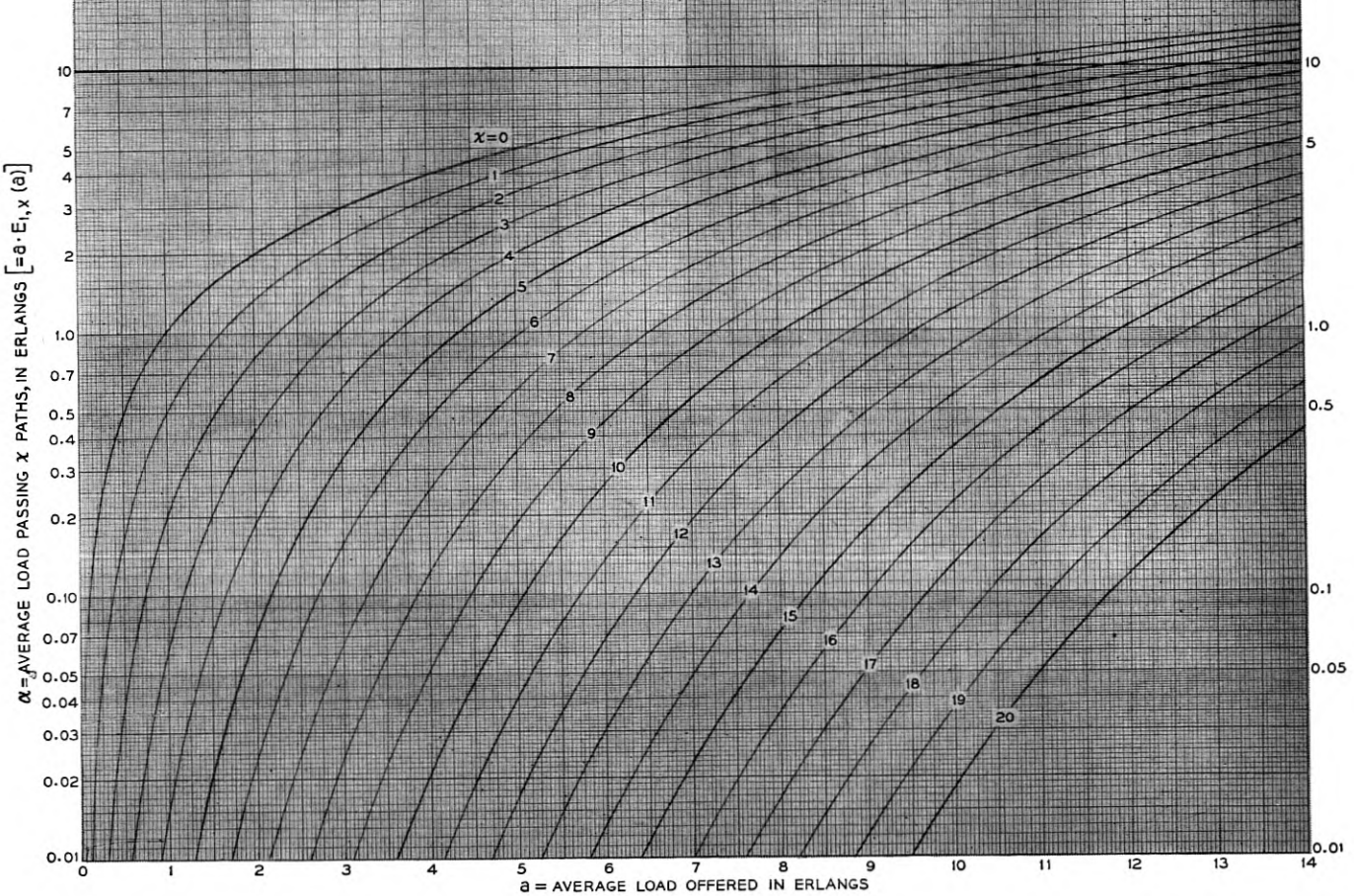


Fig. 12.1 — Average of overflow load, with 0 to 14 erlangs offered.

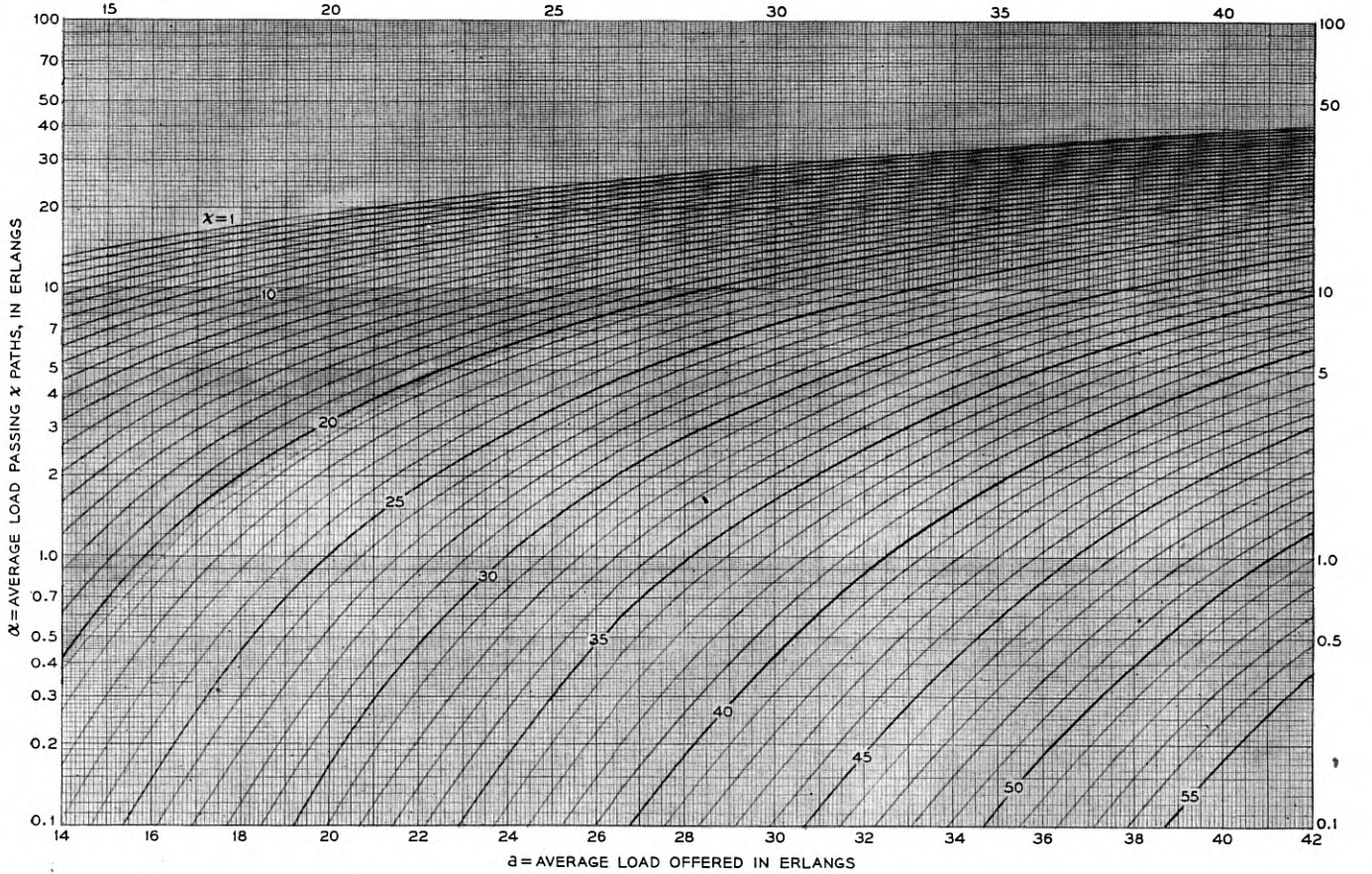


Fig. 12.2 — Average of overflow load, with 14 to 42 erlangs offered.

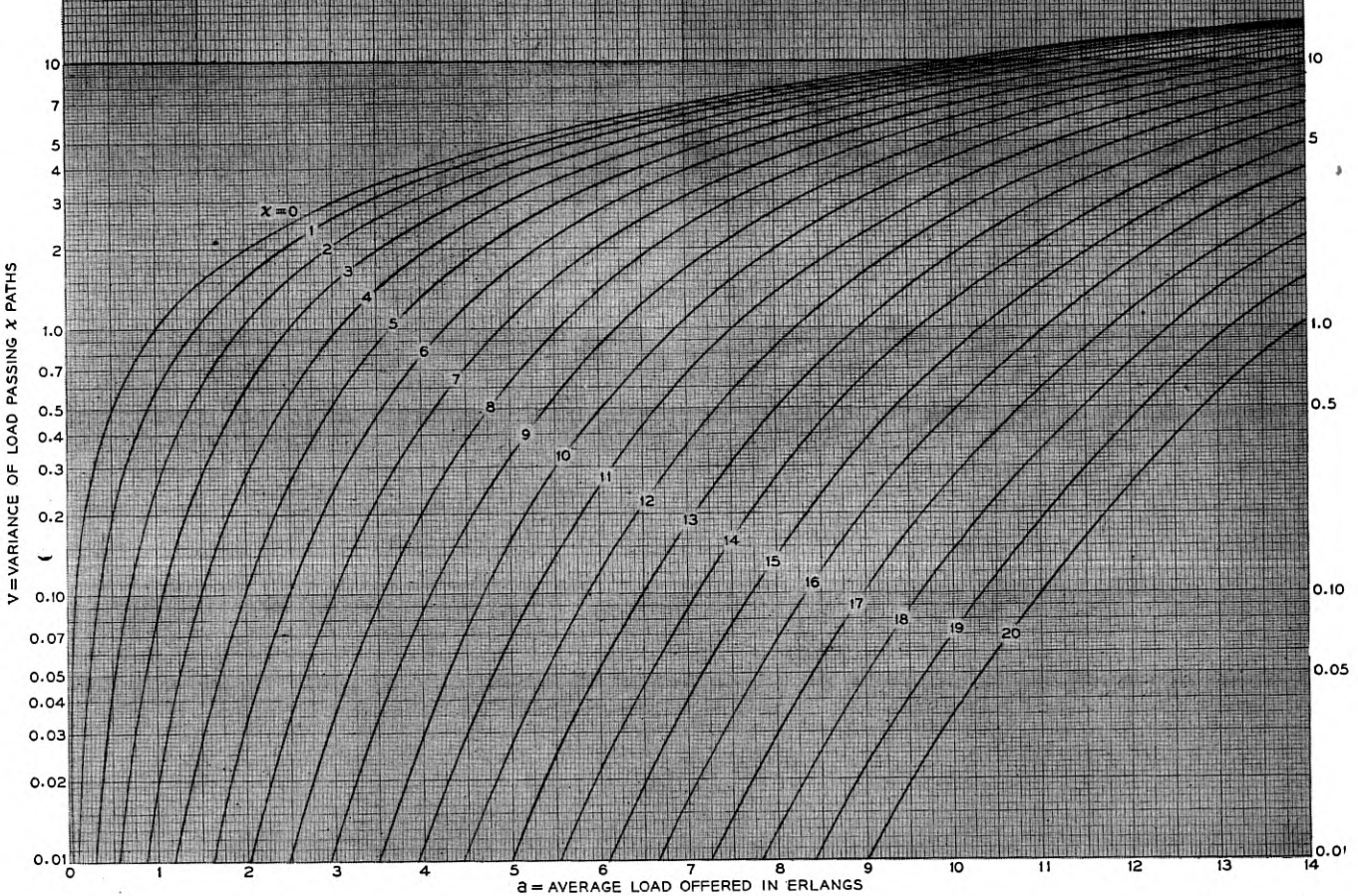


Fig. 13.1 — Variance of overflow load, with 0 to 14 erlangs offered.

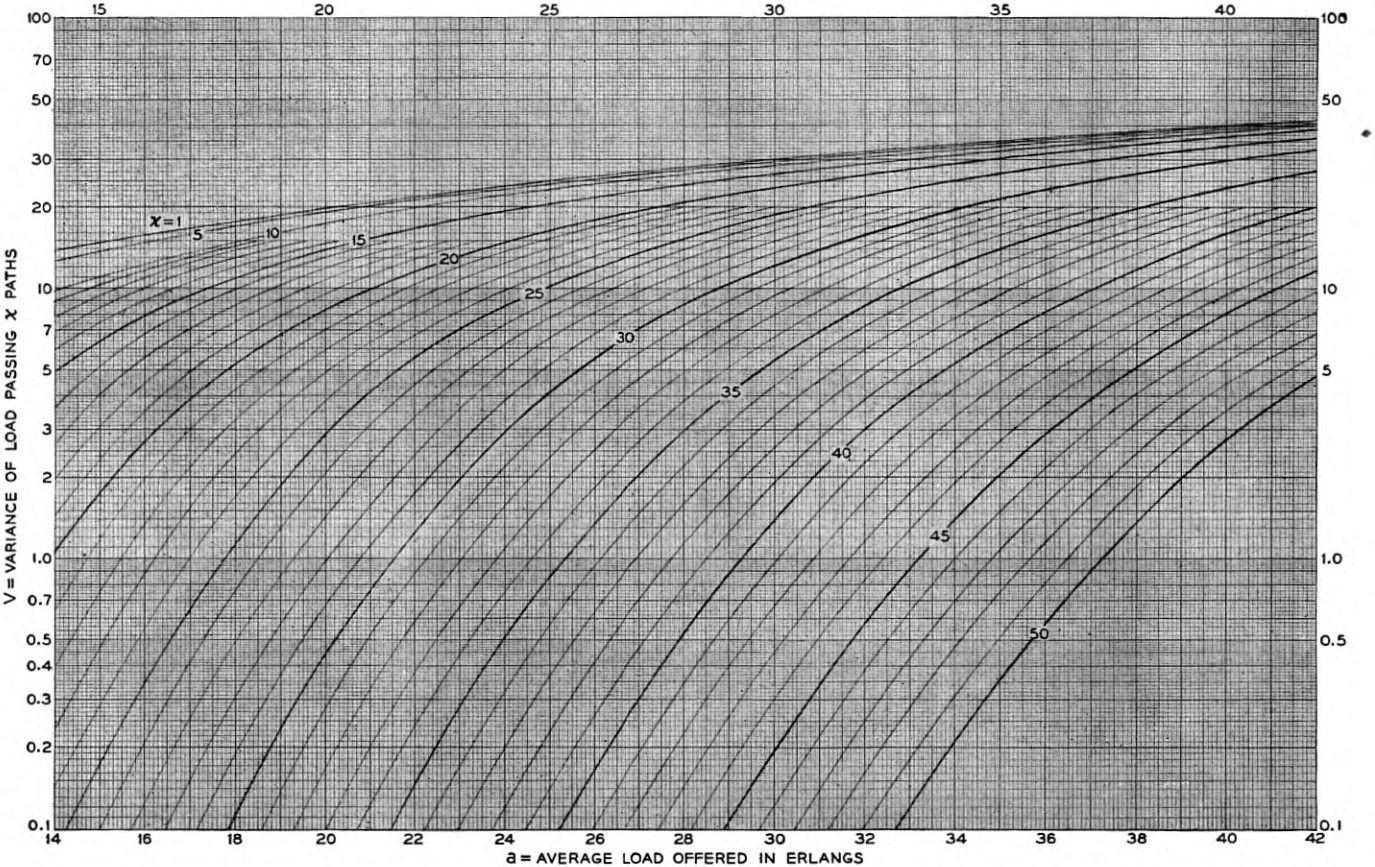


Fig. 13.2 — Variance of overflow load, with 14 to 42 erlangs offered.

as shown by $v = 1.30$, and $v_x = 1.95$. In all cases the variances v and v_x will exceed the variance of corresponding Poisson traffic (which would have variances of α and α_x respectively).

7.2.1. *A Probability Distribution for Overflow Traffic*

It would be of interest to be able, given the first several descriptive parameters of any traffic load (such as the mean and variance and skewness factors of the overflow from a group of trunks), to construct an approximate probability distribution $\theta(n)$ which would closely describe the true momentary distribution of simultaneous calls. Any proposed fitting distribution for the overflow from random traffic offered to x trunks, can, of course, be compared with

$$\theta(n) = \sum_{m=0}^{x} f(m, n)$$

determined from (7) or (8).

Suitable fitting curves should give probabilities for all possitive integral values of the variable (including zero), and have sufficient unspecified constants to accommodate the parameters selected for describing the distribution. Moreover, the higher moments of a fitting distribution should not diverge too radically from those of the true distribution; that is, the "natural shapes" of fitting and true distributions should be similar. Particularly desirable would be a fitting distribution form derived with some attention to the physical circumstances causing the ebb and flow of calls in an overflow situation. The following argument and derivation undertake to achieve these desiderata.*

A Poisson distribution of offered traffic is produced by a random arrival of calls. The assumption is made or implied that the probability of a new arrival in the next instant of time is quite independent of the number currently present in the system. When this randomness (and corresponding independence) are disturbed the resulting distribution will no longer be Poisson. The first important deviation from the Poisson would be expected to appear in a change from variance = mean, to variance $\neq$

* A two-parameter function which has the ability to fit quite well a wide variety of true overflow distributions, has the form

$$\xi(n) = K(n + 1)^b \, e^{-c(n+1)}$$

in which K is the normalizing constant. The distribution is displaced one unit from the usual discrete generalized exponential form, so that $\xi(0) \neq 0$. The expression, however, has little rationale for being selected a priori as a suitable fitting function.

mean. Corresponding changes in the higher moments would also be expected.

What would be the physical description of a cause system with a variance smaller or larger than the Poisson? If the variance is smaller, there must be forces at work which retard the call arrival rate as the number of calls recently offered exceeds a normal, or average, figure, and which increase the arrival rate when the number recently arrived falls below the normal level. Conversely, the variance will exceed the Poisson's should the tendencies of the forces be reversed.* This last is, in fact, a rough description of the incidence rates for calls overflowing a group of trunks.

Since holding times are attached to and extend from the call arrival instants, calls are enabled to project their influence into the future; that is, the presence of a considerable number of calls in a system at any instant reflects their having arrived in recent earlier time, and now can be used to modify the current rate of call arrival.

Let the probability of a call originating in a short interval of time dt be

$$P_{o,n} = [a + (n - a)\omega(n)]\, dt$$

where n = number of calls present in the system at time t,
a = base or average arrival rate of calls per unit time, and
$\omega(n)$ = an arbitrary function which regulates the modification in call origination rate as the number of calls rises above or falls below a.

Correspondingly, let the probability that one of n calls will end in the short interval of time dt be

$$P_{e,n} = n\, dt,$$

which will be satisfied in the case of exponential call holding times, with mean unity. Following the usual Erlang procedure, the general statistical equilibrium equation is

$$f(n) = f(n)[(1 - P_{o,n})(1 - P_{e,n})] + f(n - 1)P_{o,n-1}(1 - P_{e,n-1}) + f(n + 1)(1 - P_{o,n+1})P_{e,n+1} \tag{16}$$

which gives

$$(P_{o,n} + P_{e,n})f(n) = P_{o,n-1}f(n - 1) + P_{e,n+1}f(n + 1)$$

ignoring terms of order higher than the first in dt.

* The same thinking has been used by Vaulot[7] for decreasing the call arrival rate according to the number momentarily present; and by Lundquist[8] for both increasing and decreasing the arrival rate.

Or,

$$[a + (n - a)\omega(n) + n]f(n) = [a + (n - a - 1)\omega(n - 1)]f(n - 1) + (n + 1)f(n + 1) \tag{17}$$

The choice of $\omega(n)$ will determine the solution of (17). Most simply, $\omega(n) = k$, making the variation from the average call arrival rate directly proportional to the deviation in numbers of calls present from their average number. In this case, the solution for an unlimited trunk group becomes, with $a' = a(1 - k)$,

$$f(n) = \frac{\dfrac{a'(a' + k) \cdots [a' + (n - 1)k]}{n!}}{1 + a' + \dfrac{a'(a' + k)}{2!} + \dfrac{a'(a' + k)(a' + 2k)}{3!} + \cdots} \tag{18}$$

which may also be written after setting $a'' = a'/k = a(1 - k)/k$, as

$$f(n) = \frac{\dfrac{a''(a'' + 1) \cdots [a'' + (n - 1)]k^n}{n!}}{(1 - k)^{-a''}} \tag{19}$$

The generating function (g.f.) of (19) is

$$\sum_{n=0}^{\infty} f(n)T^n = \frac{(1 - kT)^{-a''}}{(1 - k)^{-a''}}$$

which is recognized as that for the negative binomial, as distinguished from the g.f.,

$$(q + pT)^N = \frac{\left(1 + \dfrac{p}{q}T\right)^N}{(1/q)^N}$$

for the positive binomial.

The first four descriptive parameters of $f(n)$ are:

Order	Moment about Mean	Descriptive Parameter	
1	$\mu_1 = 0$	Mean $= \bar{n} = a$	(20)
2	$\mu_2 =$ variance, $v = a/(1 - k)$	Std Devn, $\sigma = [a/(1 - k)]^{1/2}$	(21)
3	$\mu_3 = \dfrac{a(1 + k)}{(1 - k)^2}$	Skewness, $\sqrt{\beta_1} = \dfrac{\mu_3}{\sigma^3} = \dfrac{1 + k}{a^{1/2}(1 - k)^{1/2}}$	(22)
4	$\mu_4 = \dfrac{3a^2(1 - k) + a(k^2 + 4k + 1)}{(1 - k)^3}$	Kurtosis, $\beta_2 = \dfrac{\mu_4}{\sigma^4} = 3 + \dfrac{k^2 + 4k + 1}{a(1 - k)}$	(23)

Since only two constants, a and k, need specification in (18) or (19), the mean and variance are sufficient to fix the distribution. That is, with the mean $\bar{n}$ and variance v known,

$$a = \bar{n} \quad \text{or} \quad a' = \bar{n}(1 - k) = \bar{n}^2/v, \quad \text{or} \quad a'' = \bar{n}(1 - k)/k \tag{24}$$

$$k = 1 - a/v = 1 - \bar{n}/v. \tag{25}$$

The probability density distribution $f(n)$ is readily calculated from (19); the cumulative distribution $G(\geqq n)$ also may be found through use of the Incomplete Beta Function tables since

$$\begin{aligned} G(\geqq n) &= I_k(n - 1, a'') \\ &= I_k(n - 1, a(1 - k)/k) \end{aligned} \tag{26}$$

The goodness with which the negative binomial of (19) fits actual distributions of overflow calls requires some investigation. Perhaps a more elaborate expression for $\omega(n)$ than a constant k in (17) is required. Three comparisons appear possible: (1), comparison with a variety of $\theta_m(n)$ distributions with exactly m calls on the x trunks, or $\theta(n)$ with m unspecified, (obtained by solving the statistical equilibrium equations (7) for a divided group); (2), comparison with simulation or "throwdown" results; and (3), comparison with call distributions seen on actual trunk groups. These are most easily performed in the order listed.*

Comparison of Negative Binomial with True Overflow Distributions

Figs. 14 to 17 show various comparisons of the negative binomial distribution with true overflow distributions. Fig. 14 gives in cumulative form the cases of 5 erlangs offered to 1, 2, 5, and 10 trunks. The true

$$F(\geqq n) = \sum_{j=n}^{\infty} \theta(j)$$

distributions (shown as solid lines) are obtained by solving the difference equations (7) in the manner described in Section 7.1. The negative binomial distributions (shown dashed) are chosen to have the same mean and variance as the several $F(\geqq n)$ cases fitted. The dots shown on

* Comparison could also be made after equating means and variances respectively, between the higher moments of the overflow traffic beyond x trunks and the corresponding negative binomial moments: e.g., the skewness given by (15) can be compared with the negative binomial skewness of (22). The difficulty here is that one is unable to judge whether the disparity between the two distribution functions as described by differences in their higher parameters is significant or not for traffic engineering purposes.

the figure are for random (Poisson) traffic having the same mean values as the F distributions. The negative binomial provides excellent fits down to cumulated probabilities of 0.01, with a tendency thereafter to give somewhat larger values than the true ones. The Poisson agreement is good only for the overflow from a single trunk, as might have been anticipated, the divergence rapidly increasing thereafter.

Fig. 15 corresponds with the cases of Fig. 14 except that the true overflow $F_x(\geqq n)$ distributions for the conditional situation of all x-paths busy, are fitted. Again the negative binomial is seen to give a good agreement down to 0.01 probability, with somewhat too-high estimates for larger values of the simultaneous overflow calls n.

Fig. 16 shows additional comparisons of overflow and negative binomial distributions. As before, the agreement is quite satisfactory to 0.01 probability, the negative binomial thereafter tending to give somewhat high values.

On Fig. 17 are compared the individual $\theta(n)$ density distributions for several cases. The agreement of the negative binomial with the true distribution is seen to be uniformly good. The dots indicate the random (Poisson) individual term distribution corresponding to the $a = 9.6$ case;

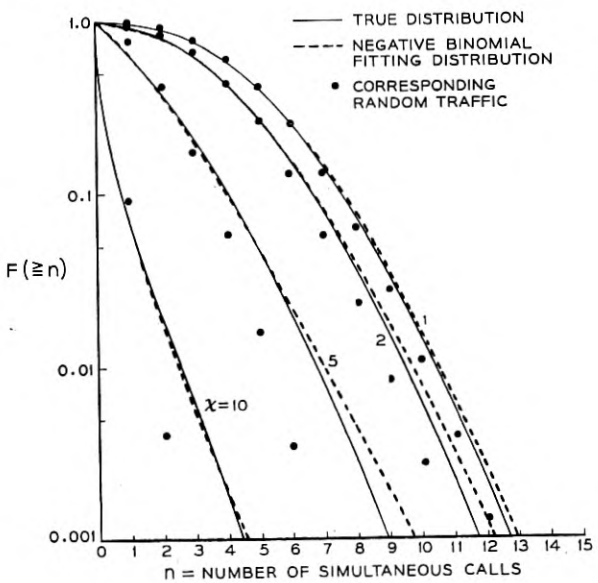


Fig. 14 — Probability distributions of overflow traffic with 5 erlangs offered to 1, 2, 5, and 10 trunks, fitted by negative binomial.

the agreement, of course, is poor since the non-randomness of the overflow here is marked, having an average of 1.88 and a variance of 3.84.

Comparison of Negative Binomial with Overflow Distributions Observed by Throwdowns and on Actual Trunk Groups

Fig. 18 shows a comparison of the negative binomial with the overflow distributions from four direct groups as seen in throwdown studies. The agreement over the range of group sizes from one to fifteen trunks is seen to be excellent. The assumption of randomness (Poisson) as shown by the dot values is clearly unsatisfactory for overflows beyond more than two or three trunks.

A number of switch counts made on the final group of an operating toll alternate routing system at Newark, New Jersey, during periods when few calls were lost, have also shown good agreement with the negative binomial distribution.

7.2.2. *A Probability Distribution for Combined Overflow Traffic Loads*

It has been shown in Section 7.2.1 that, at least for load ranges of wide interest, the negative binomial with but two parameters, chosen to agree

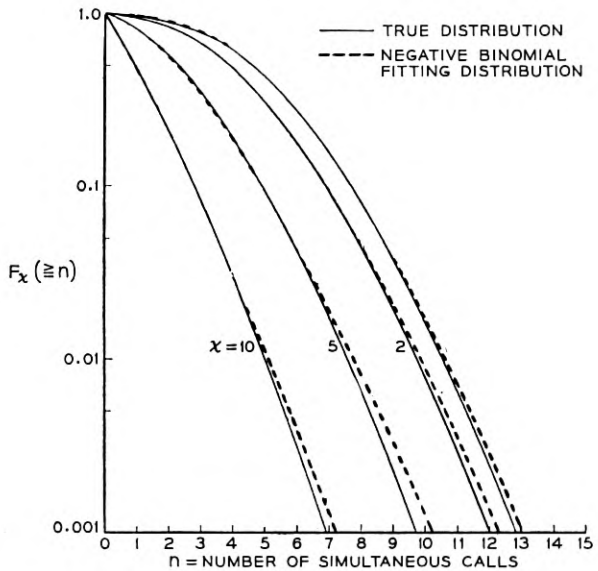


Fig. 15 — Probability distributions of overflow traffic with 5 erlangs offered to 1, 2, 5, and 10 trunks, when all trunks are busy; fitted by negative binomial.

with mean and variance, gives a satisfactory fit to the distribution of traffic overflowing a group of trunks. It is now possible, of course, to convolute the various overflows from any number of groups of varying sizes, to obtain a combined overflow distribution. This procedure, however, would be very clumsy and laborious since at each switching point in the toll alternate route system an entirely different layout of loads and high usage groups would require solution; it would be unfeasible for practical working.

We return again to the method of moments. Since the overflows of the several high usage groups will, in general, be independent of one another, the ith semi-invariants λ_i of the individual overflows can be combined to give the corresponding semi-invariants Λ_i of their total,

$$\Lambda_i = {}_1\lambda_i + {}_2\lambda_i + \cdots \tag{27}$$

Or, in terms of the overflow means and variances, the corresponding parameters of the combined loads are

$$\text{Average} = A' = \alpha_1 + \alpha_2 + \cdots \tag{28}$$

$$\text{Variance} = V' = v_1 + v_2 + \cdots \tag{29}$$

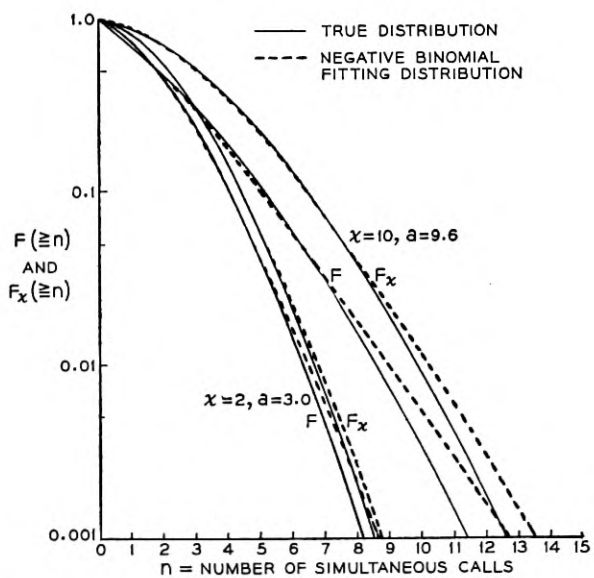


Fig. 16 — Probability distributions of overflow traffic: 3 erlangs offered to 2 trunks, and 9.6 erlangs offered to 10 trunks.

With the mean and variance of the combined overflows now determined, the negative binomial can again be employed to give an approximate description of the distribution of the simultaneous calls $\varphi(z)$ offered to the common, or alternate, group.

The acceptability of this procedure can be tested in various ways. One way is to examine whether the convolution of several negative binomials (representing overflows from individual groups) is sufficiently well fitted by another negative binomial with appropriate mean and variance, as found above.

It can easily be shown that the convolution of several negative binomials all with the same over-dispersion (variance-to-mean ratio) but not necessarily the same mean, is again a negative binomial. Shown in Table I are the distribution components and their parameters of two examples in which the over-dispersion parameters are not identical. The third and fourth semi-invariants of the fitted and fitting distributions, are seen to diverge considerably, as do the Pearsonian skewness and kurtosis factors. The test of acceptability for traffic fluctuation description comes in comparing the fitted and fitting distributions which are shown on Fig. 19. Here it is seen that, despite what might appear alarming dis-

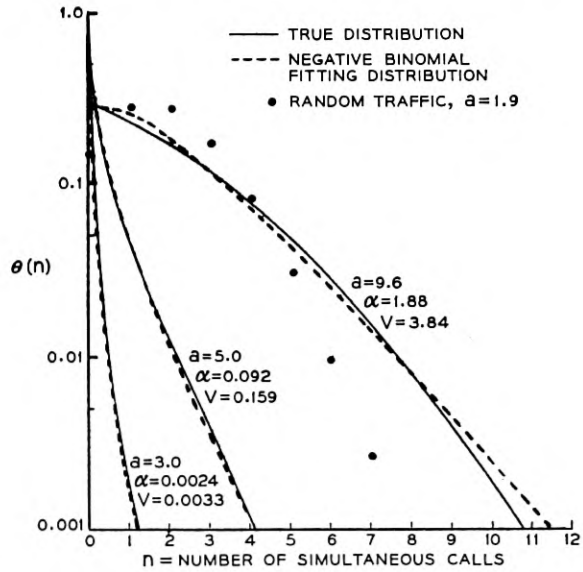


Fig. 17 — Probability density distributions of overflow traffic from 10 trunks, fitted by negative binomial.

parities in the higher semi-invariants, the agreement for practical traffic purposes is very good indeed.

Numerous throwdown checks confirm that the negative binomial employing the calculated sum-overflow mean and variance has a wide range over which the fit is quite satisfactory for traffic description purposes. Fig. 20 shows three such trunking arrangements selected from a considerable number which have been studied by the simulation method. Approximately 5,000, 3,500, and 580 calls were run through in the three examples, respectively. The overflow parameters obtained by experiment are seen to agree reasonably well with the theoretical ones from (28) and (29) when the numbers of calls processed is considered.

On Fig. 21 are shown, for the first arrangement of Fig. 20, distributions of simultaneous offered calls in each subgroup of trunks compared with the corresponding Poisson; the agreement is satisfactory as was to be expected. The sum distribution of the overflows from the eight subgroups is given at the foot of the figure. The superposed Poisson, of course, is a poor fit; the negative binomial, on the other hand, appears quite acceptable as a fitting curve.

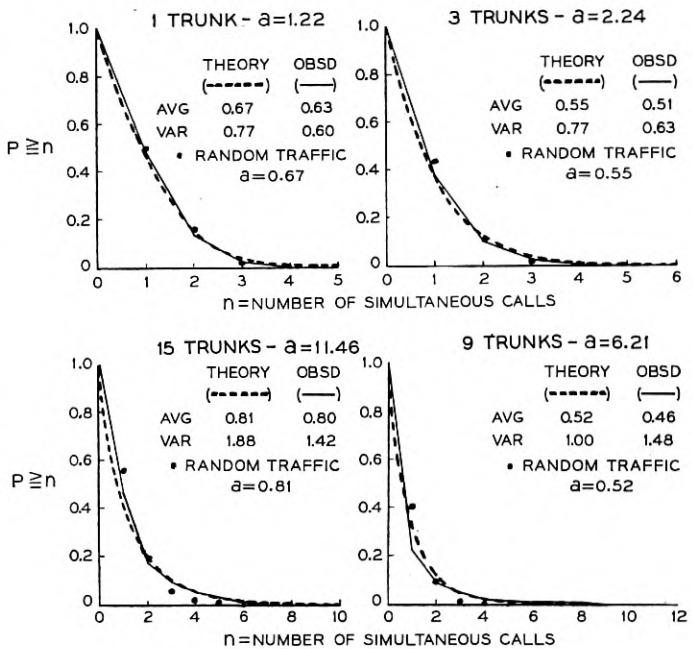


Fig. 18 — Overflow distributions from direct interoffice trunk groups; negative binomial theory versus throwdown observations.

TABLE I — COMPARISON OF PARAMETERS OF A FITTING NEGATIVE BINOMIAL TO THE CONVOLUTION OF THREE NEGATIVE BINOMIALS

Example No. 1			Example No. 2		
Component dist'n No.	Component parameters		Component dist'n No.	Component parameters	
	Mean	Variance		Mean	Variance
1	5	5	1	1	1
2	2	4	2	2	3
3	1	3	3	2	6
	8	12		5	10

Semi-Invariants Λ, Skewness $\sqrt{\beta_1}$, and Kurtosis β_2, of Sum Distributions

Parameter	Exact	Fitting	Parameter	Exact	Fitting
Λ_1	8	8	Λ_1	5	5
Λ_2	12	12	Λ_2	10	10
Λ_3	32	24	Λ_3	37	30
Λ_4	168	66	Λ_4	239.5	130
$\sqrt{\beta_1}$	0.770	0.577	$\sqrt{\beta_1}$	1.170	0.949
β_2	4.167	3.458	β_2	5.395	4.300

Fig. 22 shows the corresponding comparisons of the overflow loads in the other two trunk arrangements of Fig. 20. Again good agreement with the negative binomial is seen.

7.3. *Equivalent Random Theory for Prediction of Amount of Traffic Overflowing a Single Stage Alternate Route, and Its Character, with Lost Calls Cleared*

As discussed in Section 7.2, when random traffic is offered to a limited number of trunks x, the overflow traffic is well described (at least for traffic engineering purposes) by the two parameters, mean α and variance v. The result can readily be applied to a group divided (in one's mind) two or more times as in Fig. 23.

Employing the α and v curves of Figs. 12 and 13, and the appropriate numbers of trunks x_1, $x_1 + x_2$, and $x_1 + x_2 + x_3$, the pairs of descriptive parameters, α_1, v_1, α_2, v_2 and α_3, v_3 can be read at once. It is clear then that if at some point in a straight multiple a traffic with parameters α_1, v_1 is seen, and it is offered to x_2 paths, the overflow therefrom will have the characteristics α_2, v_2. To estimate the particular values of α_2 and v_2, one would first determine the values of the equivalent random

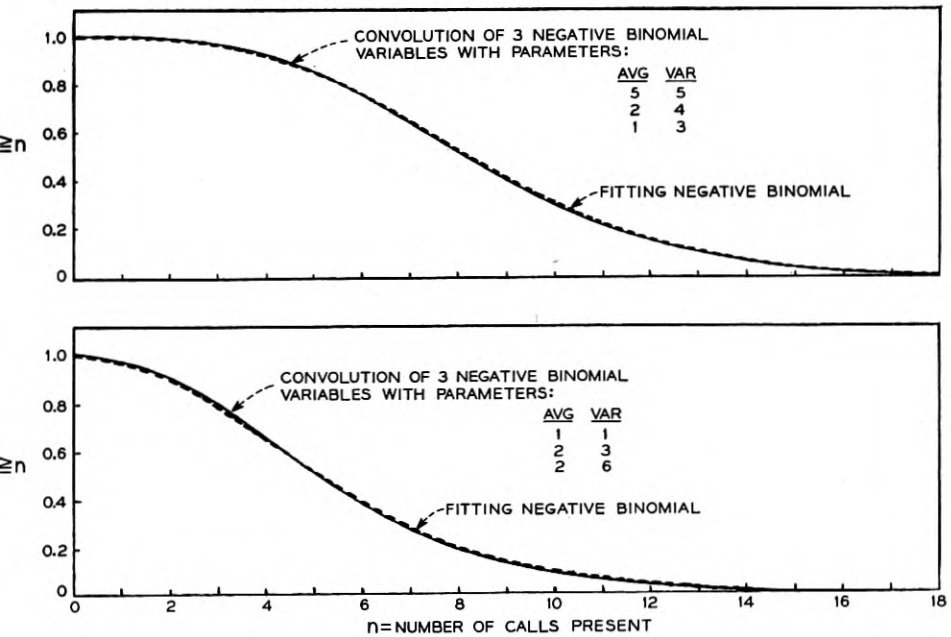


Fig. 19 — Fitting sums of negative binomial variables with a negative binomial.

traffic a and trunks x_1 which would have produced α_1 and v_1. Then proceeding in the forward direction, using a and $x_1 + x_2$, one consults the α and v charts to find α_2 and v_2. Thus, within the limitations of straight group traffic flow, the character (mean and variance) of any overflow load from x trunks can be predicted if the character (mean and variance) of the load submitted to them is known.

Curves could be constructed in the manner just described by which the overflow's α' and v' are estimated from a load, α and v, offered to x trunks. An illustrative fragment of such curves is shown in Appendix II, with an example of their application in the calculation of a straight trunk group loss by considering the successive overflows from each trunk as the offered loads to the next.

Enough, perhaps, has been shown in Section 7.2 of the generally excellent descriptions of a variety of non-random traffic loads obtainable by the use of only the two parameters α and v, to make one strongly suspect that most of the fluctuation information needed for traffic engineering purposes is contained in those two values. If this is, in fact, the case, we should then be able to predict the overflow α', v' from x trunks

with an offered load α, *v which has arisen in any manner of overflow from earlier high usage groups*, as illustrated in Fig. 24.

This is found to be the case, as will be illustrated in several studies described in the balance of this section. In the determination of the characteristics of the overflow traffic α', v' in the cases of non-full-access groups, such as Figs. 24(b) and 24(c), the equivalent straight group is visualized [Fig. 24(a)], and the Equivalent Random load A and trunks S are found.* Using A, and $S + C$, to enter the α and v curves of Figs. 12 and 13, α' and v' are readily determined. To facilitate the reading of A and S, Fig. 25† and Fig. 26† (which latter enlarges the lower left corner of Fig. 25) have been drawn. Since, in general, α and v will not have come from a simple straight group, as in Fig. 24(a), it is not to be expected that S,

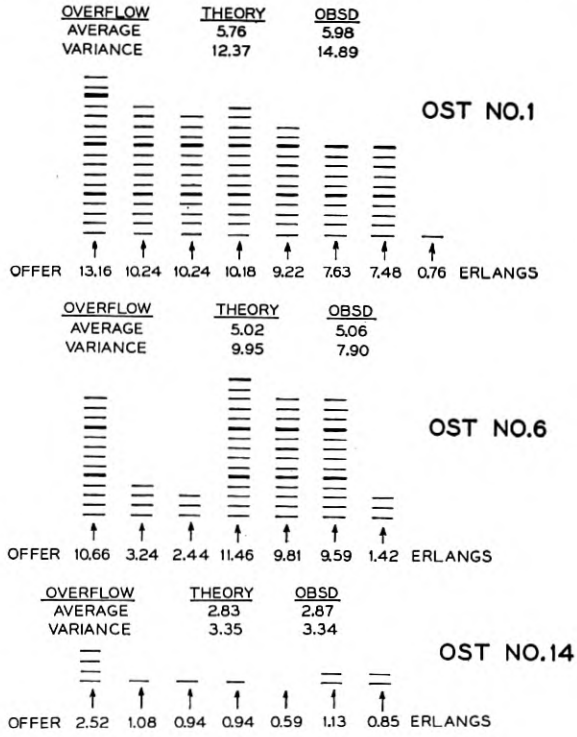


Fig. 20 — Comparison of joint-overflow parameters; theory versus throwdown.

* A somewhat similar method, commonly identified with the British Post Office, which uses one parameter, has been employed for solving symmetrical graded multiples (Ref. 9).

† Figs. 25 and 26 will be found in the envelope on the inside back cover.

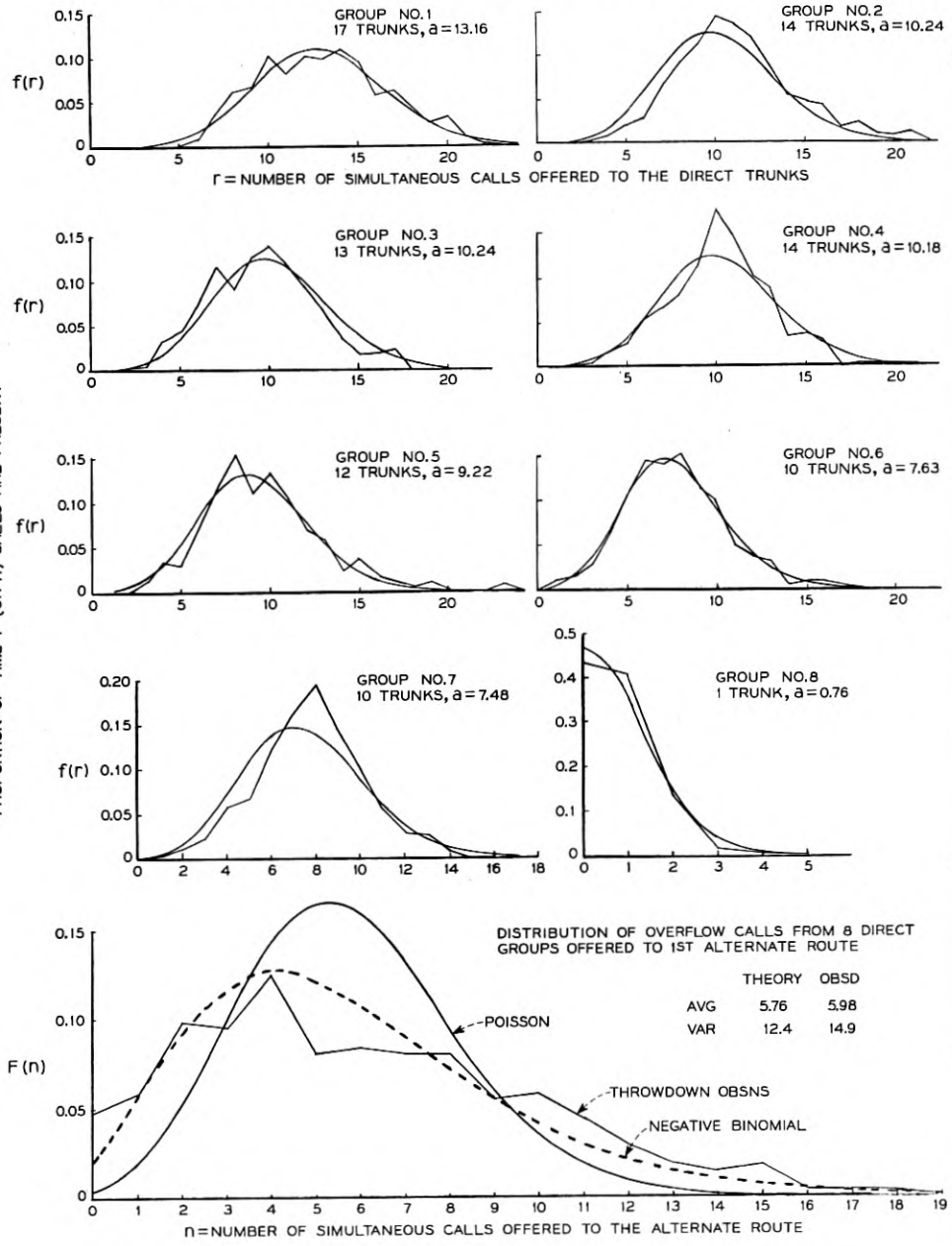


Fig. 21 — Comparison of theoretical and throwdown distributions of simultaneous calls offered to direct groups and to their first alternate route (OST No. 1).

read from Fig. 25, will be an integer. This causes no trouble and S should be carried along fractionally to the extent of the accuracy of result desired. Reading S to one-tenth of a trunk will usually be found sufficient for traffic engineering purposes.

Example 1: Suppose a simple graded multiple has three trunks in each of two subgroups, which overflow to C common trunks, where $C = 1$,

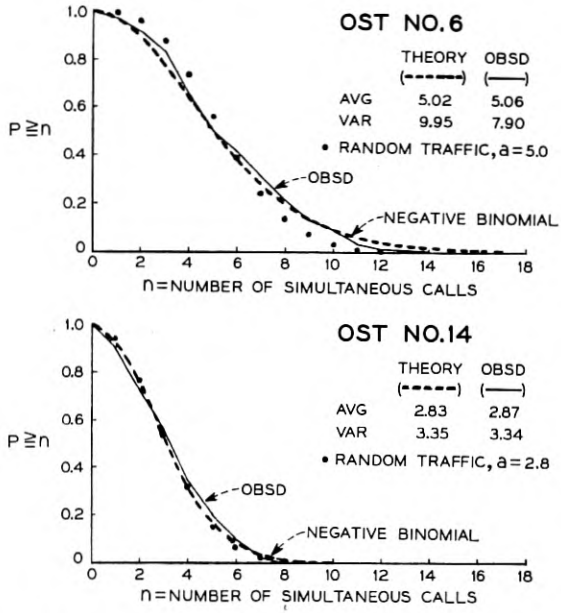


Fig. 22 — Combined overflow loads offered to alternate-route OST trunks from direct interoffice trunks; negative binomial theory vs throwdown observations.

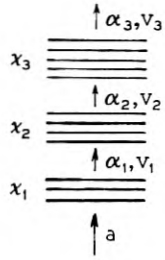


Fig. 23 — A full access group divided at several points to examine the traffic character at each point.

2 or 3. A load of a erlangs is submitted to each subgroup, a having the values 1, 2, 3, 4 or 5. What grade of service will be given?

Solution: The load overflowing each subgroup, when $a = 1$ for example, has the characteristics $\alpha = 0.0625$ and $v = 0.0790$. Then $A' = 2\alpha = 0.125$ and $V' = 2v = 0.158$. Reading on Fig. 26 gives the Equivalent Random values of $A = 1.04$ erlangs, $S = 2.55$ trunks. Reading on Fig. 12.1 with $C + S = 3.55$ when $C = 1$, and $A = 1.04$, we find $\alpha' = 0.0350$ and $\alpha'/(a_1 + a_2) = 0.0175$. We construct Table II in which loss values predicted by the Equivalent Random (ER) Theory are given in columns (3), (5) and (7). For comparison, the corresponding exact values given by Neovius* are shown in columns (2), (4) and (6). (Less exact loss

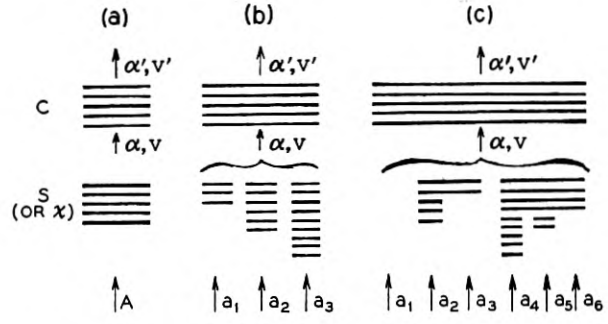


Fig. 24 — Various high usage trunk group arrangements producing the same total overflow α, v.

figures were given previously by Conny Palm[10]. The agreement is seen to be excellent for engineering needs for all values in the table.

Example 2: Suppose in Fig. 24(b) the random offered loads and paths are as given in Table III; we desire the proportion of overflow and the overflow load characteristics from an alternate route of 5 trunks.

Solution: The individual overflows α_1, v_1; α_2, v_2; and α_3, v_3 are read from Figs. 12 and 13 and recorded in columns (4) and (5) of the table. The α and v columns are totalled to obtain the sum-overflow average A' and variance V'. The Equivalent Random load A which, if submitted to S trunks would produce overflow A', V', is found from Fig. 26. Finally, with A submitted to $S + C$ trunks the characteristics α' and v', of the load overflowing the C trunks are found. The numerical values obtained

* Artificial Traffic Trials Using Digital Computers, a paper presented by G. Neovius at the First International Congress on the Application of the Theory of Probability on Telephone Engineering and Administration, Copenhagen, June, 1955.

TABLE II—CALCULATION OF LOSS IN A SIMPLE GRADED MULTIPLE
$g = 2, x_1 = x_2 = 3, \quad a_1 = a_2 = a = 1$ to 5, $\quad C = 1$ to 3

Load Submitted to each Subgroup in Erlangs a	Proportion of Each Subgroup Load which Overflows $= \alpha'/(a_1 + a_2)$					
	C = 1		C = 2		C = 3	
	True	ER	True	ER	True	ER
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.01737	0.0175	0.00396	0.0045	0.00077	0.00088
2	0.11548	0.115	0.05630	0.057	0.02438	0.024
3	0.24566	0.246	0.16399	0.163	0.10212	0.103
5	0.35935	0.363	0.27705	0.279	0.20535	0.210
5	0.44920	0.445	0.37336	0.370	0.30308	0.305

for this example are shown in the lower section of Table III. As before, of course, the "lost" calls are assumed cleared, and do not reappear in the system.

Example 3: A load of 18 erlangs is offered through four groups of 10-point selector switches to twenty-two trunks which have been designated as "high usage" paths in an alternate route plan. Which of the trunk arrangements shown in Fig. 27 is to be preferred, and to what extent?

Solution: By successive applications of the Equivalent Random method the overflow percentages for each of the three trunk arrangements are determined. The results are shown in column 2 of Table IV. The difference in percentage overflow between the three trunk plans is small; however, plan 2 is slightly superior followed by plans 3 and 1 in

TABLE III — CALCULATION OF OVERFLOWS FROM A SIMPLE ALTERNATE ROUTE TRUNK ARRANGEMENT

Subgroup Number	Offered Load in Erlangs a	Number of Trunks x	Overflow Loads	
			α	v
1	3.5	3	1.41	1.98
2	5.7	6	1.39	2.40
3	6.0	9	0.45	0.85
	15.2		3.25	5.23

Description of load offered to alternate route: $A' = 3.25, V' = 5.23$.
Equivalent straight multiple: $S = 5.8$ trunks, $A = 8.00$ erlangs (from Fig. 26).
Overflow from $C = 5$ alternate route trunks (enter Figs. 12 and 13 with $A = 8.0$ and $S + C = 10.8$: $\alpha' = 0.72, v' = 1.48$.
Proportion of load to commons which overflows = 0.72/3.25 = 0.22.
Proportion of offered load which overflows = 0.72/15.2 = 0.0475.

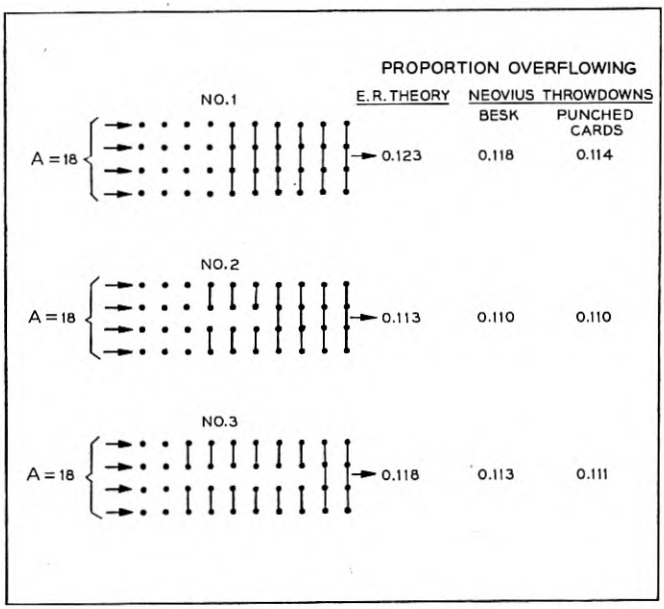


Fig. 27 — Comparison of losses on three graded arrangements of 22 trunks.

that order. The results of extensive simulations made by Neovius on the three trunk plans are available for comparison.* The values so obtained are seen to be very close to the ER theoretical ones; moreover the same order of preference among the three plans is indicated and with closely similar loss differentials between them.

7.3.1. *Throwdown Comparisons with Equivalent Random Theory on Simple Alternate Routing Arrangements with Lost Calls Cleared*

Results of manually run throwdowns on a considerable number of non-symmetrical single-stage alternate route arrangements are available. Some of these were shown in Fig. 20; they represent part of a projected multi-alternate route layout (to be described later) for outgoing calls from the local No. 1 crossbar Murray Hill–6 office in New York to all other offices in the metropolitan area. The paths hunted over initially are called direct trunks; they overflow calls to Office Selector Tandem (OST) groups, numbered from 1 to 17, which are located in widely dispersed central office buildings in the Greater New York area.

* Loc. cit.

TABLE IV—LOSS COMPARISON OF GRADED ARRANGEMENTS

Plan Number	Estimates of Percentage of Load Overflowing		
	ER Theory	Neovius Throwdowns	
		BESK Computer (262144 calls)	Punched Cards (10,000 calls)
(1)	(2)	(3)	(4)
1	12.3	11.81	11.4
2	11.3	10.98	11.0
3	11.8	11.25	11.1

TABLE V — COMPARISON OF THEORY AND THROWDOWNS FOR THE PARAMETERS OF LOADS OVERFLOWING THE COMMON TRUNKS IN SINGLE-STAGE GRADED MULTIPLES

OST (Alternate) Route Group		No. of Groups of Direct Trunks	Total No. of Trunks in Direct Groups	Total Load Offered to Direct Trunks		Total Overflow Load from OST			
Group no.	No. of trunks			Erlangs	Approximate No. of Calls (in 2.7 hours)	Theory		Throwdown	
						α'	v'	α'	v'
1	6	8	91	68.91	4950	2.00	5.50	2.36	6.52
2	3	3	45	37.49	2690	2.10	5.60	2.05	6.36
3	6	6	80	60.62	4355	1.50	4.00	1.30	5.67
4	3	6	52	38.49	2765	2.30	5.20	2.08	6.43
5	3	3	17	12.51	900	0.45	0.83	0.49	1.02
6	4	7	64	48.62	3490	2.50	5.90	2.36	4.88
7	8	12	78	57.42	4125	2.20	5.60	1.71	4.08
8	6	9	16	12.96	930	0.82	1.63	0.81	1.11
9	1	2	22	16.96	1220	1.30	2.60	1.02	1.73
10	5	6	10	9.52	685	0.78	1.40	1.05	2.07
11	8	13	16	16.43	1180	1.90	3.80	2.77	7.29
12	8	9	2	6.88	495	0.70	1.30	0.81	1.83
13	5	15	33	21.42	1540	1.75	3.30	1.16	2.01
14	2	7	11	8.05	580	1.46	2.20	1.63	2.14
15	9	15	8	11.97	860	1.60	3.25	1.55	4.12
16	11	22	34	27.46	1970	1.75	4.00	1.34	2.26
17	3	7	4	5.81	420	1.53	2.31	1.43	1.80
						26.64	58.42	25.92	61.32

In Table V are given certain descriptive data for the 17 OST trunk arrangements showing numbers of legs of direct trunks, total direct trunks, the offered erlangs and calls, and the mean and variance of the alternate routes' overflows, as obtained by the ER theory and by throwdowns.* The throwdown α' and v' values of the OST overflow

* Additional details of this simulation study are given in Section 7.4.

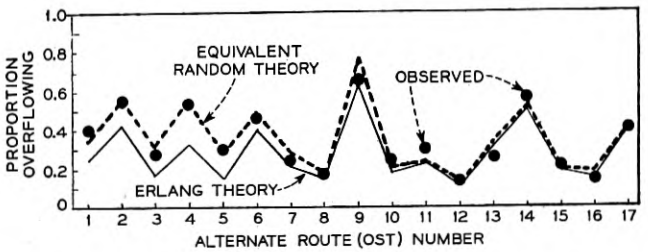


Fig. 28 — Comparison of theoretical and throwdown overflows from a number of first alternate routes.

were obtained by 36-second switch counts of those calls from each OST group which had come to rest on *subsequent* alternate routes.

On Fig. 28 is shown a summary of the observed and calculated proportions of "lost" to "offered" traffic at each OST alternate route group. As may be seen from the figure and the last four columns of Table V, the general agreement is quite good; the individual group variations are probably no more than to be expected in a simulation of this magnitude.

An assumption of randomness (which has sometimes been argued as returning when several overflows are combined) for the load offered to the OST's gives the Erlang E_1 loss curve on Fig. 28. This, as was to be expected, rather consistently understates the loss.

Since "switch-counts" were made on the calls overflowing each OST, the distributions of these overflows may be compared with those estimated by the Negative Binomial theory having the mean and variance predicted above for the overflow. Fig. 29 shows the individual and cumulative probability distributions of the overflow simultaneous calls from the first two OST alternate routes. As will be seen, the agreement is quite good even though this is traffic which has been twice "non-randomized." Comparison of the observed and calculated overflow means and variances in Table V indicates that similar agreement between observed and theoretical fitting distributions for most of the other OST's would be found.

7.3.2. *Comparison of Equivalent Random Theory with Field Results on Simple Alternate Routing Arrangements*

Data were made available to the author from certain measurements made in 1941 by his colleague C. Clos on the automatic alternate routing trunk arrangement in operation in the Murray Hill–2 central office in New York. Mr. Clos observed for one busy hour the load carried on

several of its OST alternate route groups (similar to those shown in Table V for the Murray Hill–6 office, but not identical) by means of an electromechanical switch-counter having a six-second cycle. During each hour's observation, numbers of calls offered and overflowing were also recorded.

Although the loads offered to the corresponding direct trunks which overflowed to the OST group under observation were not simultaneously measured, such measurements had been made previously for several hours so that the relative contribution from each direct group was closely known. In this way the loads offered to each direct group which produced the total arriving before each OST group could be estimated with considerable assurance. From these direct group loads the character (mean and variance) of the traffic offered to and overflowing the OST's was predicted. The observed proportion of offered traffic which overflowed is shown on Fig. 30 along with the Equivalent Random theory prediction. The general agreement is again seen to be fairly good although with some tendency for the ER theory to predict higher than observed losses in the lower loss ranges; perhaps the disparity on in-

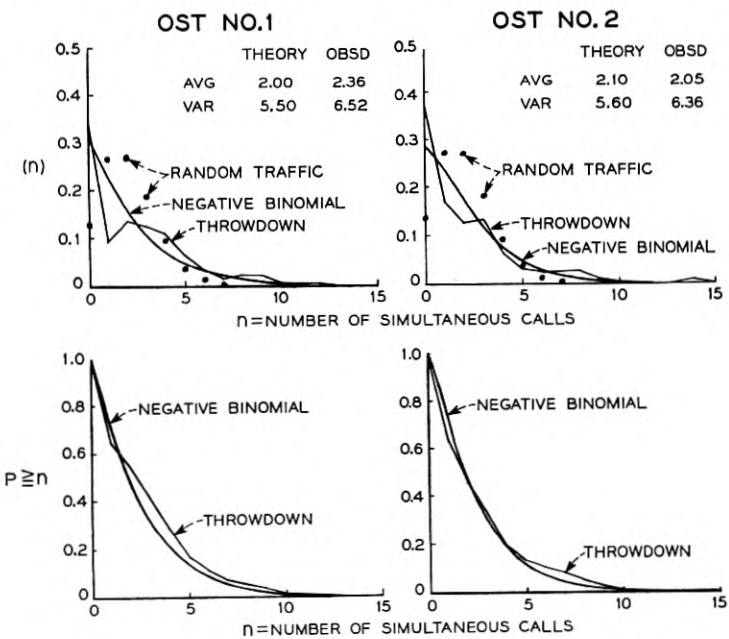


Fig. 29 — Distributions of loads overflowing from first alternate (OST) groups; negative binomial theory versus throwdown observations.

dividual OST groups is within the limits one might expect for data based on single-hour observations and for which the magnitudes of the direct group offered loads required some estimation. The assumption of random traffic offered to the OST gives, as anticipated, loss predictions (Erlang E_1) consistently below those observed.

More recently extensive field tests have been conducted on a working toll automatic alternate route system at Newark, New Jersey. High usage groups to seven distant large cities overflowed calls to the Newark-Pittsburgh alternate (final) route. Data describing the high usage groups and typical system busy hour loads are given in Table VI. (The loads, of course, varied considerably from day to day.) The size of the Pittsburgh route varied over the six weeks of the 1955 tests from 64 to 71 trunks. Altogether the system comprised some 255 intertoll trunks.

Observations were made at the Newark end of the groups by means of a Traffic Usage Recorder — making switch counts every 100 seconds — and by peg count and overflow registers. Register readings were photographically recorded by half-hourly, or more frequent, intervals. To

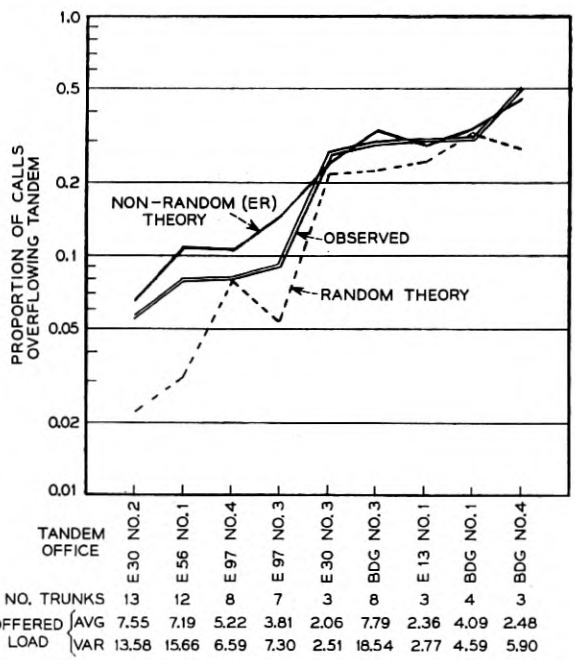


Fig. 30 — Observed tandem overflows in alternate route study at Murray Hill–2 (New York) 1940–1941.

TABLE VI—HIGH USAGE GROUPS AND TYPICAL SYSTEM BUSY HOUR LOADS

High Usage Group, Newark to:	Length of Direct Route (Air Miles)	Nominal Size of Group (Number of Trunks)	Typical Offered Load (erlangs)
Baltimore	170	18	19
Cincinnati	560	42	43
Cleveland	395	27	26
Dallas	1375	33	34
Detroit	470	37	36
Kansas City	1100	26	23
New Orleans	1170	5	4

compare theory with the observed overflow from the final route, estimates of the offered load A' and its variance V' are required. In the present case, the total load offered to the final route in each hour was estimated as

$$A' = \text{Average of Offered Load}$$

$$= \frac{\text{Peg Count of Calls Offered to Pittsburgh Group}}{\text{(Peg Count of Offered Calls)} - \text{(Peg Count of Overflow Calls)}} \times \text{Average Load Carried by Pittsburgh Group}$$

The variance V' of the total load offered to the final route was estimated for each hour as

$$V' = \text{Variance of Offered Load}$$

$$= A' - \sum_{i=1}^{7} \alpha_i + \sum_{i=1}^{7} v_i$$

where α_i and v_i are, respectively, the average and variance of the load overflowing from the ith high usage group. (The expression, $A' - \sum_{i=1}^{7} \alpha_i$, is an estimate of the average — and, therefore of the variance — of the first-routed traffic offered directly to the final route. Thus the total variance, V', is taken as the sum of the direct and overflow components.) Using A', V' and the actual number, C, of final route trunks in service, the proportion of offered calls expected to overflow was calculated for the traffic and trunk conditions seen for 25 system busy hours from February 17 to April 1, 1955 on the Pittsburgh route. The results are displayed on Fig. 31, where certain traffic data on each hour are given in the lower part of the figure. The hours are ordered — for convenience in plotting and viewing — by ascending proportions of calls overflowing the group; observed results are shown by the double line

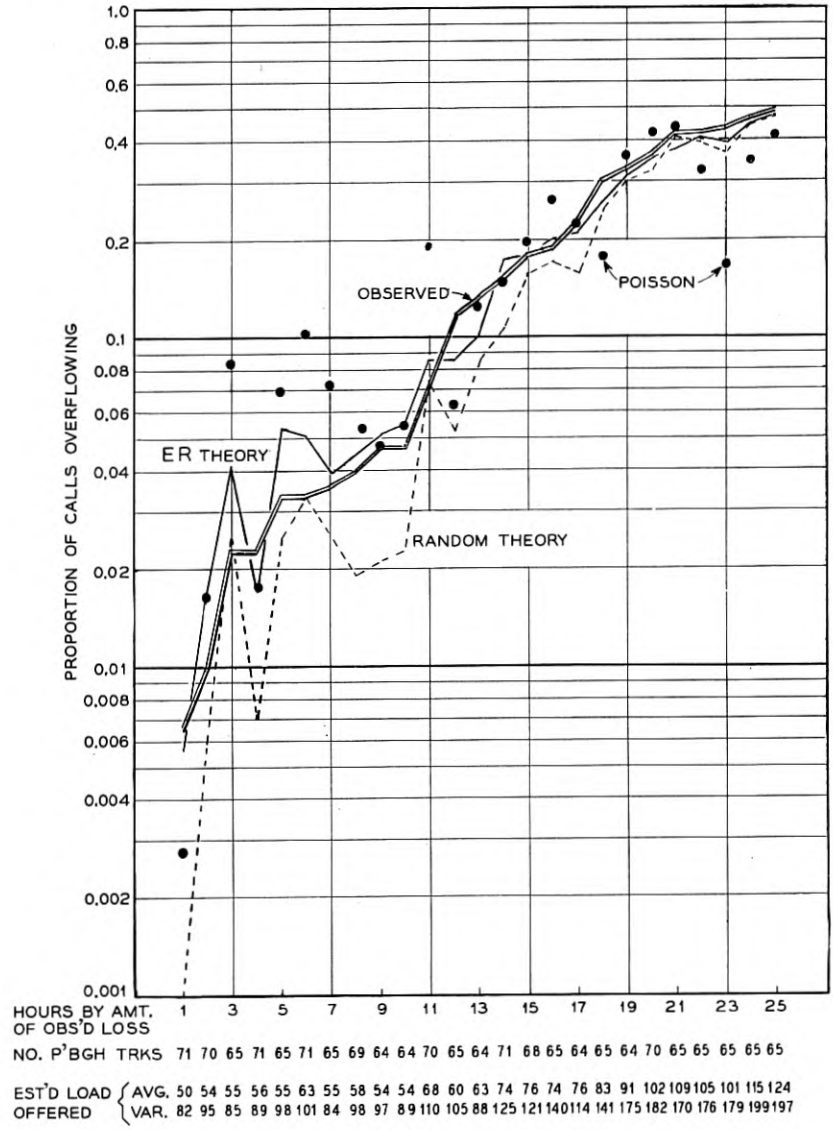


Fig. 31 — Final route (Newark-Pittsburgh) overflows in 1955 toll alternate route study.

curve. The superposed single line is the corresponding estimate by ER theory of the hour-to-hour call losses. As may be seen, theory and observation are in good agreement both point by point and on the average over the range of losses from 0.01 to 0.50. The dashed line shows the prediction of final route loss for each hour on the assumption that the offered traffic A' was random. Such an assumption gives consistently low estimates of the existing true loss.

As of interest, a series of heavy dots is included on Fig. 31. These are the result of calculating the Poisson Summation, $P(C,L)$, where L is the average load *carried* on, rather than offered to, the C trunks. It is interesting that just as in earlier studies in this paper on straight groups of intertoll trunks (for example as seen on Fig. 7), the Poisson Summation with load carried taken as the load offered parameter, gives loss values surprisingly close to those observed. Also, as before, this summation has a tendency to give too-great losses at light loadings of a group and too-small losses at the heavier loadings.

7.4 *Prediction of Traffic Passing Through a Multi-Stage Alternate Route Network*

In the contemplated American automatic toll switching plan, wide advantage is expected to be taken of the efficiency gains available in multi-alternate routing. Thus any procedure for traffic analysis and prediction needs to be adaptable for the more complex multi-stage arrangements as well as the simpler single-stage ones so far examined. Extension of the Equivalent Random theory to successive overflows is easily done since the characterizing parameters, average and variance, of the load overflowing a group of paths are always available.

Since few cases of more than single-stage automatic alternate routing are yet in operation in the American toll plant, it is not readily possible to check an extension of the theory with actual field data. Moreover collecting and analyzing observations on a large operating multi-alternate route system would be a comparatively formidable experiment.

However, in New York city's local interoffice trunking there is a very considerable development of multi-alternate routing made possible by the flexibility of the marker arrangements in the No. 1 crossbar switching system. None of these overflow arrangements has been observed as a whole, simultaneously and in detail. The Murray Hill–2 data in OST groups reviewed in Section 7.3.2 were among the partial studies which have been made.

In connection with studies made just prior to World War II on these

TABLE VII — SUM OF DIRECT GROUP OVERFLOW LOADS, OFFERED TO OST'S

	Theory	Observed
Average	86.06	87.12
Variance	129.5	127.4

local multi-alternate route systems, a throwdown was made in 1941 on a proposed trunk plan for the Murray Hill–6 office. The arrangement of trunks is shown on Fig. 32. Three successive alternate routes, Office Selector Tandems (OST), Crossbar Tandem (XBT), and Suburban Tandem (ST), are available to the large majority of the 123 direct trunk groups leading outward to 169 distant offices. (The remaining 46 parcels of traffic did not have direct trunks to distant offices but, as indicated on the diagram, offered their loads directly to a tandem group.) A total of 726 trunks is involved, carrying 475 erlangs of traffic.

A throwdown of 34,001 offered calls corresponding to 2.7 hours of traffic was run. Calls had approximate exponential holding times, averaging 135 seconds. Records were kept of numbers of calls and the load from the traffic parcels offered to each direct group, as they were carried or passed beyond the groups of paths to which they had access. Loads carried by each trunk in the system were also observed by means of a 36-second "switch-count." (The results on the 17 OST groups reported in Section 7.3.1 were part of this study.)

Comparisons of observation and theory which are of interest include the combined loads to and overflowing the 17 OST's. Observed versus calculated parameters (starting with theory from the original direct group submitted loads) are given in Table VII. The agreement is seen to be very good.

The corresponding comparison of total load from all the OST's is given in Table VIII. Again the agreement is highly satisfactory.

Not all of the overflow from the OST's was offered to the 22 crossbar tandem trunks; for economic reasons certain parcels by-passed XBT and were sent directly to Suburban Tandem.* This posed the problem of breaking off certain portions of the overflow from the OST's, to be added again to the overflow from XBT. An estimate was needed of the contribution made by each parcel of direct group traffic to any OST's overflow. These were taken as proportional to the loads *offered* the OST by each direct group (this assumes that each parcel suffers the same over-

* In the toll alternate route system by-passing of this sort will not occur.

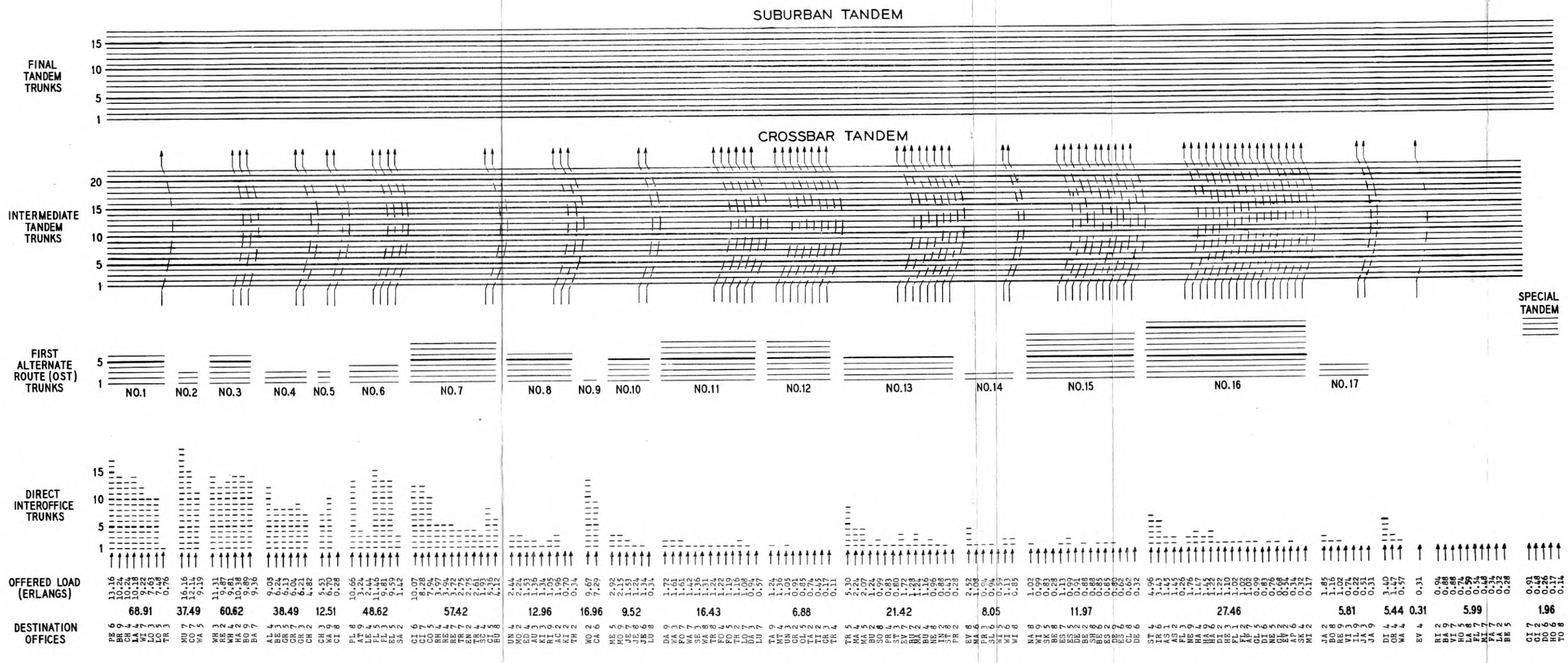


Fig. 32. — Multi-alternate route trunking arrangement at Murray Hill—6 (New York) local No. 1 crossbar office.

flow probability). The variance of this overflow portion by-passing XBT was estimated by assigning to it the same variance-to-average ratio as was found for the total load overflowing the OST. Subtracting the means and variances so estimated for all items by-passing XBT, left an approximate load for XBT from each OST. Combining these corrected overflows gave mean and variance values for offered load to XBT. Observed values

TABLE VIII — SUM OF LOADS OVERFLOWING OST'S

	Theory	Observed
Average	26.64	25.92
Variance	58.42	61.32

TABLE IX — LOAD OFFERED TO CROSSBAR TANDEM

	Theory	Observed
Average	25.18	25.51
Variance	47.67	56.10

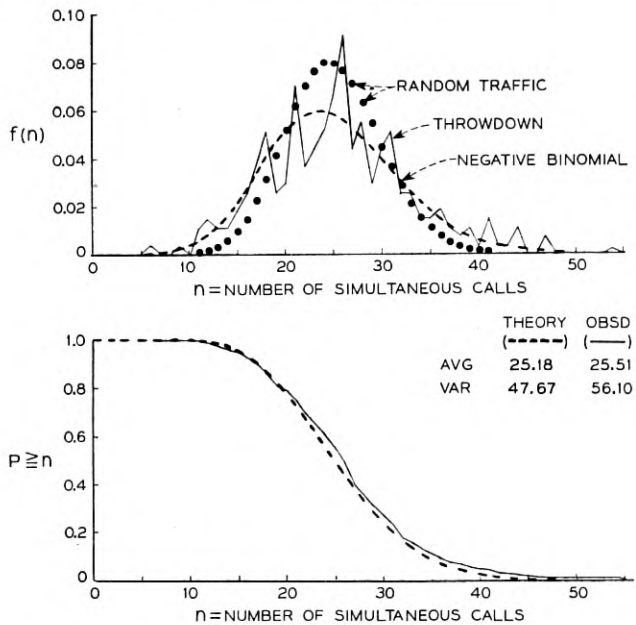


Fig. 33 — Distribution of load offered to crossbar tandem trunks; negative binomial theory versus throwdown observations.

TABLE X — LOAD OVERFLOWING CROSSBAR TANDEM

	Theory	Observed
Average	6.55	6.47
Variance	23.80	33.48

and those calculated (in the above manner) are given in Table IX. Fig. 33 shows the distribution of XBT offered loads, observed and calculated. The agreement is very satisfactory. The random traffic (Poisson) distribution, is of course, considerably too narrow.

In a manner exactly similar to previous cases, the Equivalent Random load method was applied to the XBT group to obtain estimated parameters of the traffic overflowing. Comparison of observation and theory at this point is given in Table X.

Fig. 34 shows the corresponding observed and calculated distributions

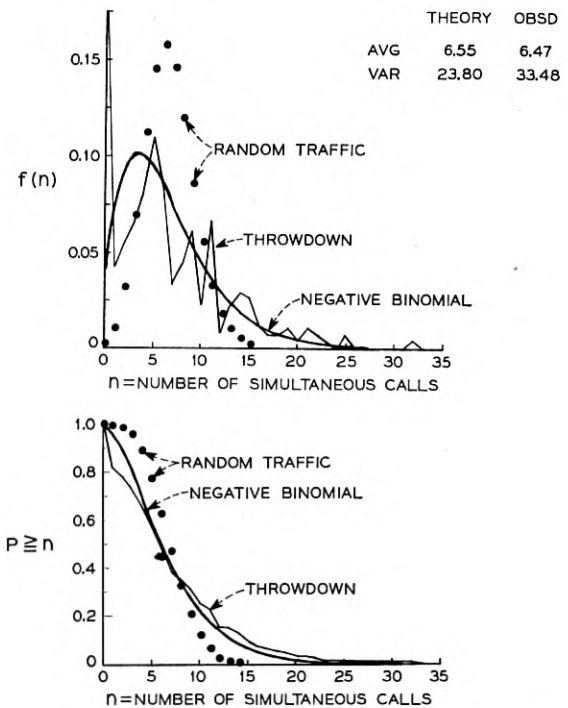


Fig. 34 — Distribution of calls from crossbar tandem trunks; negative binomial theory versus throwdown observations.

of simultaneous calls. The agreement again is reasonably good, in spite of the considerable disparity in variances.

The overflow from XBT and the load which by-passed it, as well as some other miscellaneous parcels of traffic, were now combined for final offer to the Suburban Tandem group of 17 trunks. The comparison of parameters here is again available in Table XI. On Fig. 35 are shown the observed and calculated distributions of simultaneous calls for the load offered to the ST trunks. The agreement is once again seen to be very satisfactory.

We now estimate the loss from the ST trunks for comparison with the actual *proportion of calls* which failed to find an idle path, and finally

TABLE XI — LOAD OFFERED TO SUBURBAN TANDEM

	Theory	Observed
Average	15.38	14.52
Variance	42.06	48.53

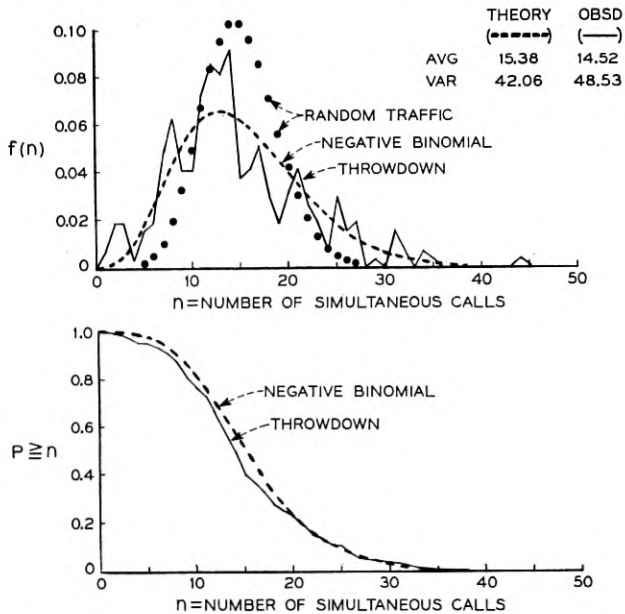


Fig. 35 — Distribution of load offered to suburban tandem trunks; negative binomial theory versus throwdown observations.

TABLE XII — GRADE OF SERVICE ON ST GROUP

	Theory	Observation	Observation	
Load submitted (erlangs)	15.38	14.52	Number of calls submitted	1057
Load overflowing (erlangs)	3.20	2.63	Number of calls overflowing	200
Proportion load overflowing	0.209	0.181	Proportion of calls overflowing	0.189

TABLE XIII — GRADE OF SERVICE ON THE SYSTEM

	Theory	Observed
Total load submitted	475 erlangs	34,001 calls
Total load overflowing	3.20 erlangs	200 calls
Proportion of load not served	0.00674	0.00588

compare the proportions of all traffic offered the system which failed to find a trunk immediately. See Tables XII and XIII.

After these several and varied combinations of offered and overflowed loads to a system of one direct and three alternate routes it is seen that the final prediction of amount of load finally lost beyond the ST trunks is gratifyingly close to that actually observed in the throwdown. The prediction of the system grade of service is, of course, correspondingly good.

It is interesting in this connection to examine also the proportions overflowing the ST group when summarized by parcels contributed from the several OST groups. The individual losses are shown on Fig. 36; they appear well in line with the variation one would expect from group to group with the moderate numbers of calls which progressed this far through the multiple.

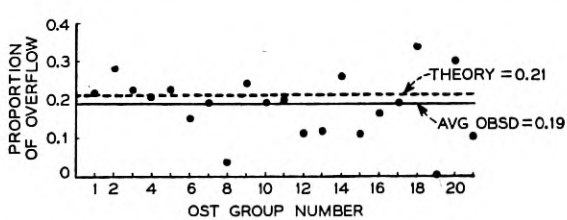


Fig. 36 — Overflow calls on third alternate (ST) route.

7.4.1 *Correlation of Loss with Peakedness of Components of Non-Random Offered Traffic*

Common sense suggests that if several non-random parcels of traffic are combined, and their joint proportion of overflow from a trunk group is P, the parcels which contain the more peaked traffic should experience overflow proportions larger than P, and the smoother traffic an overflow proportion smaller than P. It is by no means clear however, *a priori*, the extent to which this would occur. One might conjecture that if any one parcel's contribution to the total combined load is small, its loss would be caused principally by the aggregate of calls from the other parcels, and consequently its own loss would be at about the general average loss P, and hence not very much determined by its own peakedness. The Murray Hill-6 throwdown results may be examined in this respect. The mean and variance of each OST-parcel of traffic, for example, arriving at the final ST route was recorded, together with, as noted before, its own proportion of overflow from the ST trunks. The variance/mean overdispersion ratio, used as a measure of peakedness, is plotted for each parcel of traffic against its proportion of loss on Fig. 37. There is an undoubted, but only moderate, increase in proportion of overflow with increased peakedness in the offered loads.

It is quite possible, however, that by recognizing the differences between the service given various parcels of traffic, significant savings in final route trunks can be effected for certain combinations of loads and trunking arrangements. Of particular interest is the service given to a parcel of random traffic offered directly to the final route when compared

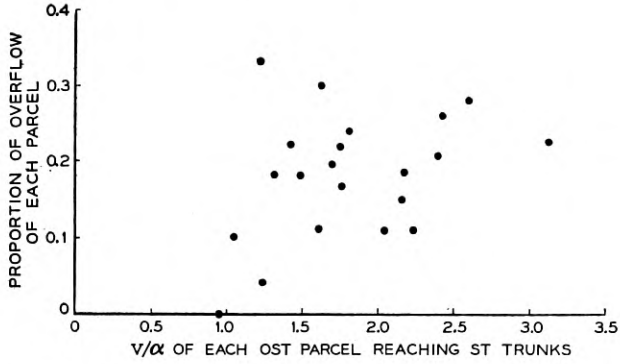


Fig. 37 — Effect of peakedness on overflow of a parcel of traffic reaching an alternate route.

with that received by non-random parcels overflowing to it from high usage groups.

7.5 Expected Loss on First Routed Traffic Offered to Final Route

The congestion experienced by the first-routed traffic offered to the final group in a complex alternate route arrangement [such as the right hand parcels in Figs. 10(c) and (d)] will be the same as encountered in a series of random tests of the final route by an independent observer, that is, it will be the proportion of time that all of the final trunks are busy. As noted before, the distribution of simultaneous calls n (and hence the congestion) on the C final trunks produced by some specific arrangement of offered load and high usage trunks can be closely simulated by that due to a single Equivalent Random load offered to a straight group of $S + C$ trunks. Then the proportion of time that the C trunks are busy in such an equivalent system provides an estimate of the corresponding time in the real system; and this proportion should be approximately the desired grade of service given the first routed traffic.

Brockmeyer[11] has given an expression (his equation 36) for the proportion of time, R_1, in a simple $S + C$ system with random offer A, and "lost calls cleared," that all C trunks are busy, independent of the condition of the S-trunks:

$$\begin{aligned} R_1 &= f(S,C,A) \\ &= E_{1,S+C}(A)\,\frac{\sigma_{C+1}(S)}{\sigma_C(S)} \end{aligned} \tag{30}$$

where

$$\sigma_C(S) = \sum_{m=0}^{S} \binom{C-1+m}{m} \frac{A^{S-m}}{(S-m)!}$$

However, $\sigma_C(S)$ is usually calculated more readily step-by-step using the formula

$$\sigma_C(S) = \sigma_C(S-1) + \sigma_{C-1}(S)\,,$$

starting with

$$\sigma_C(O) = 1 \qquad \text{and} \qquad \sigma_O(S) = A^S/S!$$

The average load carried on the C paths is clearly

$$L_C = A[E_{1,S}(A) - E_{1,S+C}(A)], \tag{31}$$

and the variance of the carried load can be shown to be*

$$V_c = AL_c \frac{\sigma_1(S)}{\sigma_2(S)} - ACE_{1,S+C}(A) + L_c - L_c^2 \qquad (32)$$

On Fig. 38, R_1 values are shown in solid line curves for several combinations of A and C over a small range of S trunks. The corresponding losses R_2 for *all* traffic offered the final group, where $R_2 = \alpha'/A'$, are shown as broken curves on the same figure. The R_2 values are always above R_1, agreeing with the common sense conclusion that a random component of traffic will receive better service than more peaked non-random components.

However, there are evidently considerable areas where the loss difference between the two R's will not be large. In the loss range of principal interest, 0.01 to 0.10, there is less proportionate difference between the R's as the $A = C$ paired values increase on Fig. 38. For example, at $R_2 = 0.05$, and $A = C = 10$, $R_2/R_1 = 0.050/0.034 = 1.47$; while for $A = C = 30$, $R_2/R_1 = 0.050/0.044 = 1.13$. Similarly for $A = 2C$, the R_2/R_1 ratios are given in Table XIV. Again the rapid decrease in the R_2/R_1 ratio is notable as A and C increase.

F. I. Tånge of the Swedish Telephone Administration has performed elaborate simulation studies on a variety of semi-symmetrical alternate route arrangements, to test the disparity between the R_1 and R_2 types of losses on the final route.† For example if g high-usage groups of 8 paths each, jointly overflow 2.0 erlangs to a final route which also serves 2.0 erlangs of first routed traffic, Tånge found the differences in losses between the two 2-erlang parcels, $R_{\text{high usage (h.u.)}} - R_1$, shown in column 9 of Table XV. The corresponding ER calculations are performed in columns 2 to 8, the last of which is comparable with the throwdown values of column 9. The agreement is not unreasonable considering the sensitiveness of determining the difference between two small probabilities of loss. A quite similar agreement was found for a variety of other loads and trunk arrangements.

---

* In terms of the first two factorial moments of n: V_c is given by

$$V_c = M_{(2)} + M_{(1)} - M_{(1)}^2, \qquad \text{where } M_{(1)} = L_c$$

General expressions $M_{(i)}$ for the factorial moments of n are derived in an unpublished memorandum by J. Riordan.

† Optimal Use of Both-Way Circuits in Cases of Unlimited Availability, a paper by F. I. Tånge, presented at the First International Congress on the Application of the Theory of Probability in Telephone Engineering and Administration, June 1955, Copenhagen.

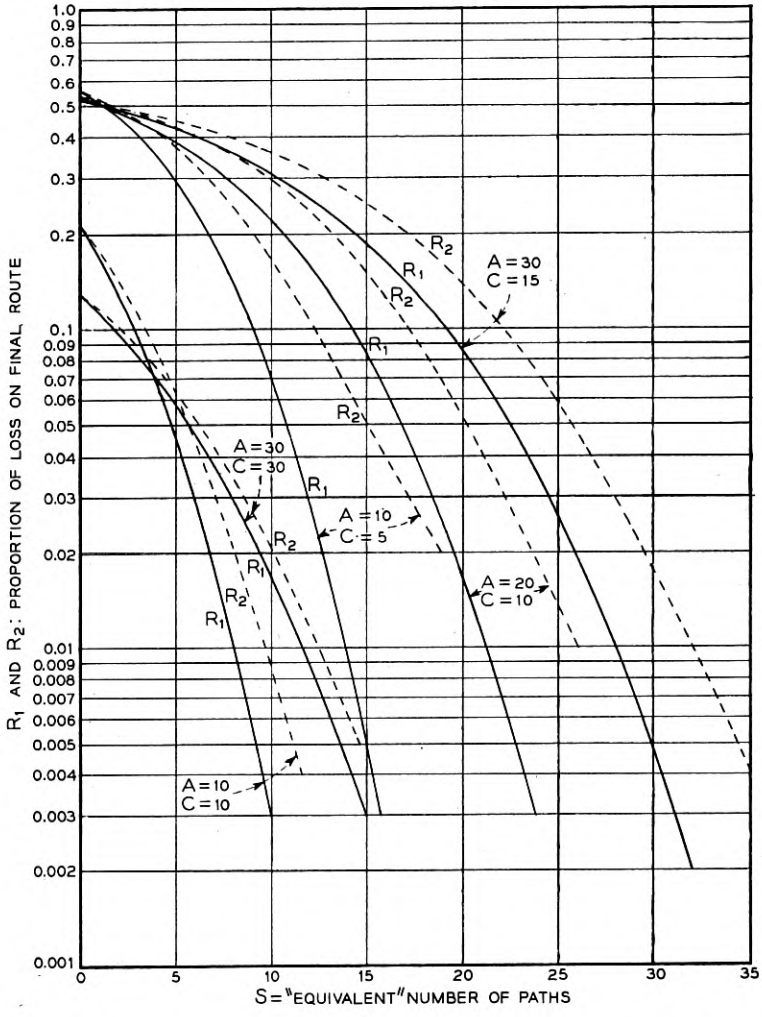


Fig. 38 — Comparison of R_1 and R_2 losses under various load and trunk conditions.

TABLE XIV—THE R_2/R_1 RATIOS FOR $A = 2C$

A	C	R_2/R_1 when $R_2 = 0.05$
10	5	10.6
20	10	3.25
30	15	2.44

TABLE XV—COMPARISON OF E.R. THEORY AND THROWDOWNS ON DISPARITY OF LOSS BETWEEN HIGH USAGE OVERFLOW AND RANDOM OFFER TO A FINAL GROUP

(8 trunks in each high usage group; 9 final trunks serving 2.0 erlangs high usage overflow and 2.0 erlangs first routed traffic.)

Number of Groups of 8 High Usage Trunks	ER Theory ($A' = 4.0$)							Tange Throwdown $R_{h.u.} - R_1$
	V'	A	S	$R_2 = \alpha'/A'$	R_1	$R_{h.u.} = 2R_2 - R_1$	$R_{h.u.} - R_1 = 2(R_2 - R_1)$	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	5.77	7.51	4.17	0.0375	0.0251	0.0499	0.0248	0.0180
2	5.80	7.50	4.25	0.0383	0.0255	0.0511	0.0256	0.0247
3	5.74	7.44	4.08	0.0369	0.0248	0.0490	0.0242	0.0286
4	5.68	7.30	3.91	0.0362	0.0247	0.0477	0.0230	0.0276
5	5.64	7.20	3.80	0.0355	0.0242	0.0468	0.0226	0.0245
6	5.58	7.06	3.64	0.0350	0.0240	0.0460	0.0220	0.0221
7	5.55	7.00	3.56	0.0345	0.0238	0.0452	0.0204	0.0202
8	5.51	6.91	3.45	0.0335	0.0236	0.0434	0.0198	0.0188
9	5.47	6.81	3.34	0.0325	0.0231	0.0419	0.0188	0.0177
10	5.45	6.76	3.29	0.0312	0.0225	0.0399	0.0174	0.0166

Limited data are available showing the disparity of R_1 and R_2 in actual operation in a range of load and trunk values well beyond those for which R_1 values have been calculated. Special peg count and overflow registers were installed for a time on the final route during the 1955 Newark alternate route tests. These gave separate readings for the calls from high usage groups, and for the first routed Newark to Pittsburgh calls. Comparative losses for 17 hours of operation over a wide range of loadings are shown on Fig. 39. The numbers at each pair of points give the per cent of final route offered traffic which was first routed (random). In general, approximately equal amounts of the two types of traffic were offered.

In 6 of the hours almost identical loss ratios were observed, in 7 hours the overflow-from-high-usage calls showed higher losses, and in 4 hours lower losses, than the corresponding first routed calls. The non-random calls clearly enjoyed practically as good service as the random calls. This result is not in disagreement with what one might expect from theory. To compare directly with the Newark-Pittsburgh case we should need curves on Fig. 38 expanded to correspond to A', V' values of (50, 85) to (120, 200). Examining the mid-range case of $C = 65$, $A' = 70$, $V' = 120$, we find $A \doteq 123$, $S \doteq 54$. Here A is approximately $2C$; extrapolating the $A = 2C$ curves of Fig. 38 to these much higher values of A and C suggests that R_2/R_1 would be but little different from unity.

It is clear from the above theory, throwdowns, and actual observation that there are certain areas where the service differences given first routed and high usage trunk overflow parcels of traffic are significant. In Section 8, where practical engineering methods are discussed, curves are presented which permit recognition of this fact in the determination of final trunk requirements.

7.6 *Load on Each Trunk, Particularly the Last Trunk, in a Non-Slipped Alternate Route*

In the engineering of alternate route systems it is necessary to determine the point at which to limit a high usage group of trunks and send the overflow traffic via an alternate route. This is an economic problem whose solution requires an estimate of the load which will be carried on

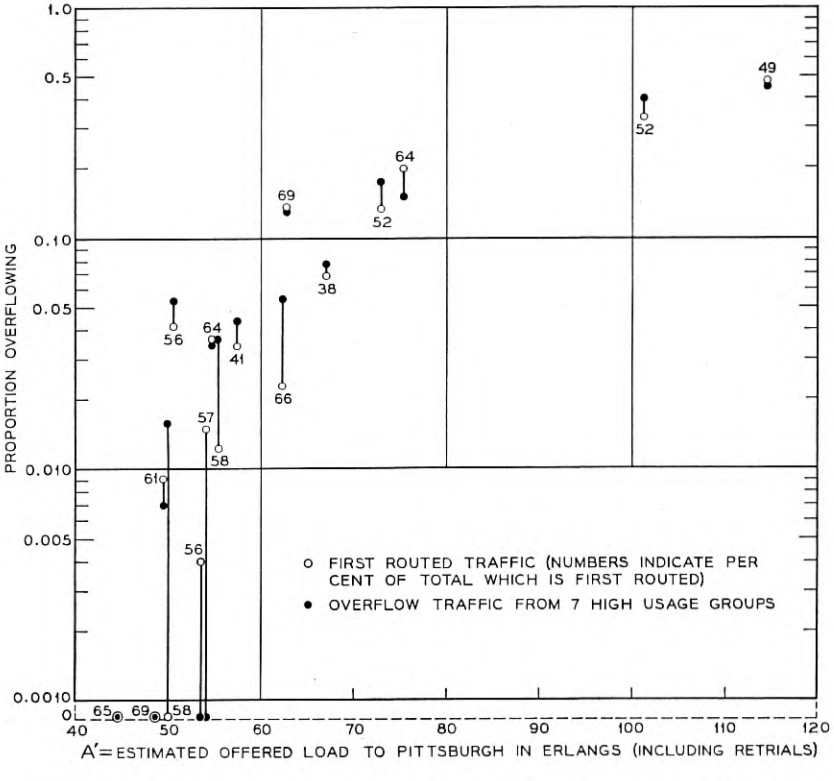


FIG. 39 — Comparison of losses on final route (Newark to Pittsburgh) for high usage overflow and first routed traffic.

the last trunk of a straight high usage group of any specified size, carrying either first or higher choice traffic or a mixture thereof.*

The Equivalent Random theory readily supplies estimates of the loads carried by any trunk in an alternate routing network. After having found the Equivalent Random load A offered to $S + C$ trunks which corresponds to the given parameters of the traffic offered to the C trunks, it is a simple matter to calculate the expected load ℓ on any one of the C trunks if they are not slipped or reversed. The load on the ith trunk in a simple straight multiple (or the $S + j$th in a divided multiple of S lower and C upper trunks), is

$$\ell_i = l_{s+j} = A[E_{1,s+j-1}(A) - E_{1,s+j}(A)] \tag{33}$$

where $E_{1,n}(A)$ is the Erlang loss formula. A moderate range of values of ℓ_i versus load A is given on Figure 40.†

Using this method, selected comparisons of theoretical versus observed loads carried on particular trunks at various points in the Murray-Hill–6 throwdown are shown in Fig. 41; these include the loads on each of the trunks of the first two OST groups of Fig. 32, and on the second and third alternate routes, crossbar and suburban tandem, respectively. The agreement is seen to be fairly good, although at the tail end of the latter two groups the observed values drop away somewhat from the theoretical ones. There seems no explanation for this beyond the possibility that the throwdown load samples here are becoming small and might by chance have deviated this far from the true values (or the arbitrary breakdown of OST overflows into parcels offered to and bypassing XBT may well have introduced errors of sufficient amount to account for this disparity). As is well known, (33) gives good estimates of the loads carried by each trunk in a high usage group to which random (Poisson) traffic is offered; this relationship has long been used for the purpose in Bell System trunk engineering.

8. PRACTICAL METHODS FOR ALTERNATE ROUTE ENGINEERING

To reduce to practical use the theory so far presented for analysis of alternate route systems, working curves are needed incorporating the

* The proper selection point will be where the circuit annual charge per erlang of traffic carried on the last trunk, is just equal to the annual charge per erlang of traffic carried by the longer (usually) alternate route enlarged to handle the overflow traffic.

† A comprehensive table of ℓ_i is given by A. Jensen as Table IV in his book "Moe's Principle," Copenhagen, 1950; coverage is for $\ell \geqq 0.001$ erlang, $i = 1(1)140$; $A = 0.1(0.1)10$, $10(1)50$, $50(4)100$. Note that $n + 1$, in Jensen's notation, equals i here.

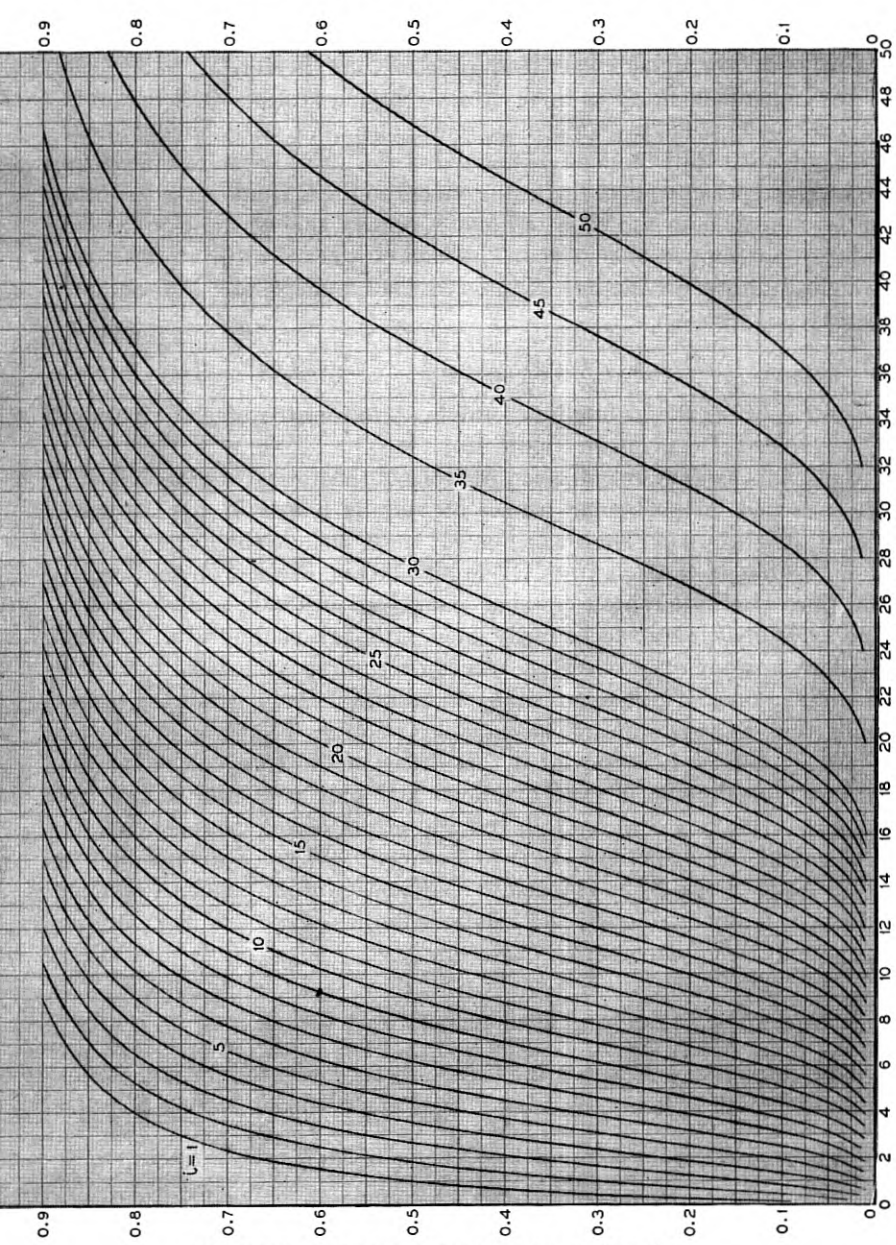


Fig. 40 — Load on each trunk of a straight multiple.

pertinent load-loss relationships. The methods so far discussed, and proposed for use, will be briefly reviewed.

The dimensioning of each high usage group of trunks is expected to be performed in the manner currently in use, as described in Section 7.6. The critical figure in this method is the load carried on the last high usage trunk, and is chosen so as to yield an economic division of the offered load between high usage and alternate route trunks. Fig. 40 is one form of load-on-each-trunk presentation suitable for choosing economic high usage group size once the permitted load on the last trunk is established.

The character (average α and variance v) of the traffic overflowing each high usage group is easily found from Figs. 12 and 13 (or equivalent

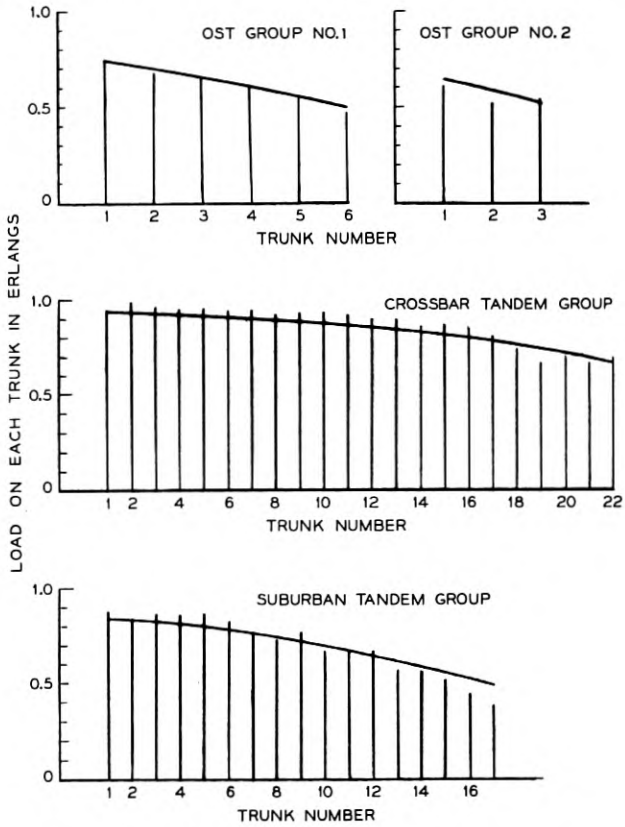


Fig. 41 — Comparison of load carried by each alternate route trunk; theory versus throwdowns.

tables). The respective sums of the overflow α's and v's, give A' and V' by (28) and (29); they provide the necessary statistical description of traffic offered to the alternate route.

According to the Equivalent Random method for estimating the alternate route trunks required to provide a specified grade of service to the overflow traffic A', one next determines a random load A which when submitted to S trunks will yield an overflow with the same character (A', V') as that derived from the complex system's high usage groups. An alternate route of C trunks beyond these S trunks is then imagined. The erlang overflow α', with random offer A, to $S + C$ trunks is found from standard E_1-formula tables or curves (such as Fig. 12).

The ratio $R_2 = \alpha'/A'$ is a first estimate of the grade of service given to each parcel of traffic offered to the alternate route. As discussed in Section 7.5, this service estimate, under certain conditions of load and trunk arrangement, may be significantly pessimistic when applied to a first routed parcel of traffic offered directly to the alternate route. An improved estimate of the overflow probability for such first routed traffic was found to be R_1 as given by (30).

8.1 *Determination of Final Group Size with First Routed Traffic Offered Directly to the Final Group*

When first routed traffic is offered directly to the final group, its service R_1 will nearly always be poorer than the *overall* service given to those other traffic parcels enjoying high usage groups. The first routed traffic's service will then be controlling in determining the final group size. Since R_1 is a function of S, C and A in the Equivalent Random solution (30), and there is a one-to-one correspondence of pairs of A and S values with A' and V' values, engineering charts can be constructed at selected service levels R_1 which show the final route trunks C required, for any given values of A' and V'. Figs. 42 to 45 show this relation at service levels of $R_1 = 0.01, 0.03, 0.05$ and 0.10, respectively.*

* On Fig. 42 (and also Figs. 46–49) the low numbered curves assume, at first sight, surprising shapes, indicating that a load with given average and variance would require fewer trunks if the average were *increased*. This arises from the sensitivity of the tails of the distribution of offered calls, to the V'/A' peakedness ratio which, of course, decreases with increases in A'. For example, with $C = 4$ trunks and fixed $V' = 0.52$, the loss rapidly decreases with increasing A':

A'	V'/A'	A	S	α'	α'/A'
0.28	1.86	6.1	10.	0.0155	0.055
0.33	1.58	3.0	5.0	0.0081	0.025
0.40	1.30	1.42	2.03	0.0036	0.009
0.52	1.00	0.52	0	0.0008	0.002

These four R_1 levels would appear to cover the most used engineering range. For example, if the traffic offered to the final route (including the first routed traffic) has parameters $A' = 12$ and $V' = 20$, reading on Fig. 43 indicates that to give $P = 0.03$ "lost calls cleared" service to the first routed traffic, $C = 19$ final route trunks should be provided. (For random traffic ($V' = A' = 12$), 17.8 trunks would be required.)

Other charts, of course, might be constructed from which R_1 could be read for specific values of A', V' and C. They would become voluminous, however, if a wide range of all three variables were required.

8.2 *Provision of Trunks Individual to First Routed Traffic to Equalize Service*

If the difference between the service R_1 given the first routed parcel of traffic and the service given all of the other parcels, is material, it may be desirable to take measures to diminish these inequities. This may readily be accomplished by setting aside a number of the otherwise full access final route trunks, for exclusive and first choice use of the first routed traffic. High usage groups are now provided for all parcels of traffic. The alternate route then services their combined overflow. The overall grade of service given the ith parcel of offered traffic in a single stage alternate route system will then be approximately

$$P_i \doteq E_{1,}x_i(a_i)R_2 = E_{1,}x_i(a_i)\frac{\alpha'}{A'}^{*} \tag{34}$$

Thus the service will tend to be uniform among the offered parcels when all send substantially identical proportions of their offered loads to the alternate route. And the natural provision of "individual" trunks for the exclusive use of the first routed traffic would be such that the same proportion should overflow as occurs in the associated high usage groups.

This procedure cannot be followed literally since high usage group size is fixed by economic considerations rather than any predetermined overflow value. The resultant overflow proportions will commonly vary over a considerable range. In this circumstance it would appear reasonable to estimate the objective overflow proportion to be used in establishing the individual group for the first routed traffic, as some weighted average $\bar{b}$ of the overflow proportions of the several high usage groups. Thus with weights g and overflow proportions b,

$$\bar{b} = \frac{g_1b_1 + g_2b_2 + \cdots}{g_1 + g_2 + \cdots} \tag{35}$$

* Although not exact, this equation can probably be accepted for most engineering purposes where high usage trunks are provided for each parcel of traffic.

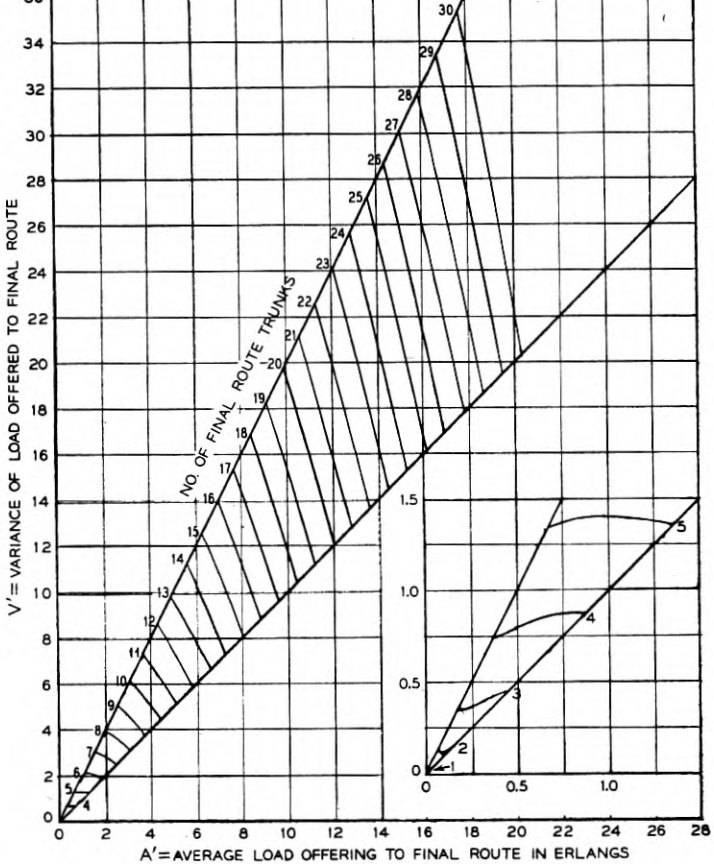


Fig. 42 — Provision of final route trunks to give first routed traffic service of $R_1 = 0.01$.

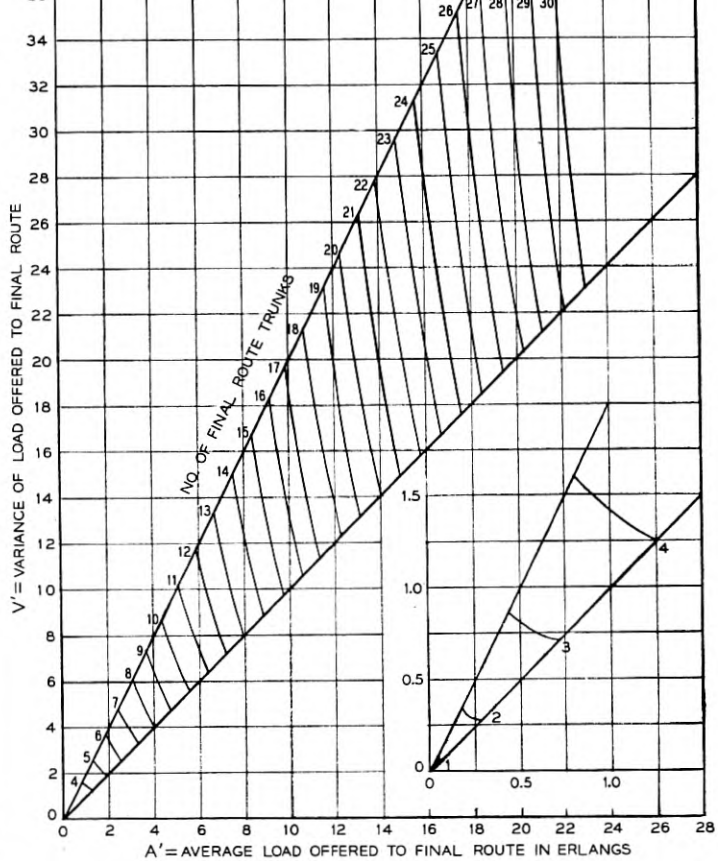


Fig. 43 — Provision of final route to trunks give first routed traffic service of $R_1 = 0.03$.

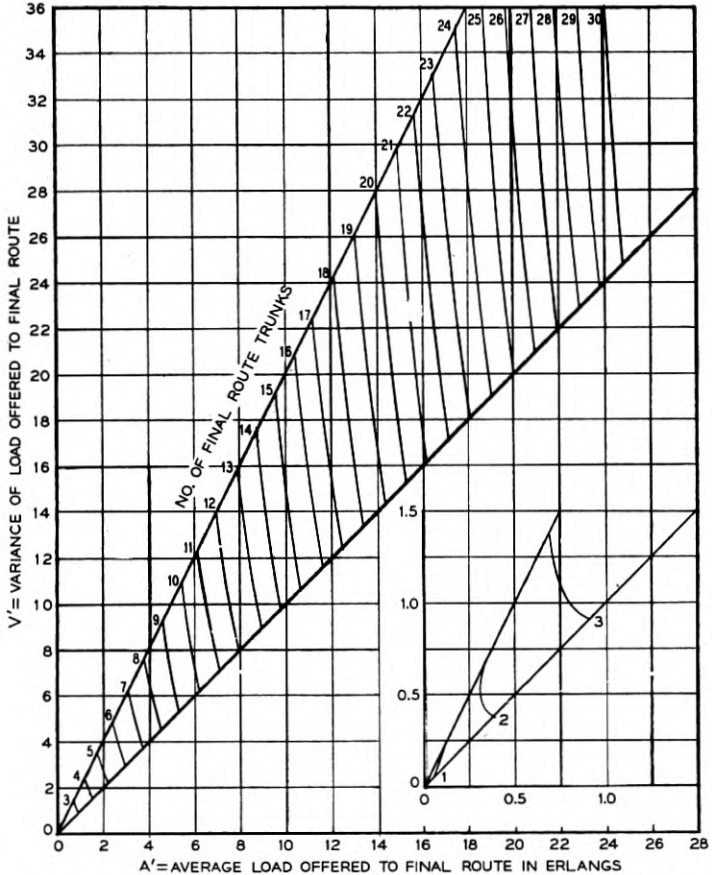


Fig. 44 — Provision of final route trunks to give first routed traffic service of $R_1 = 0.05$

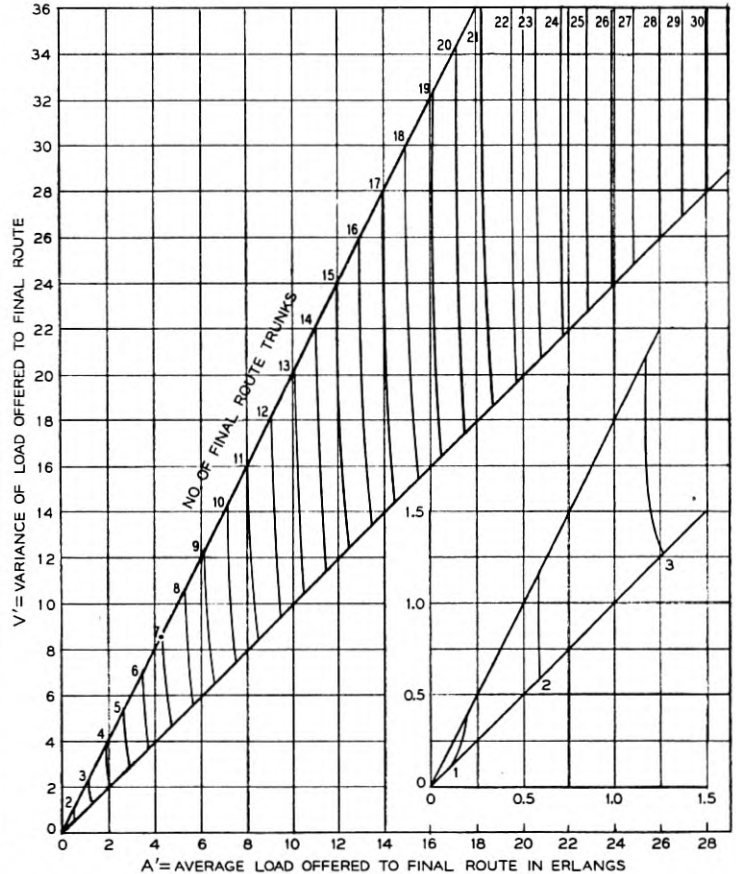


Fig. 45 — Provision of final route trunks to give first routed traffic service of $R_1 = 0.10$.

A choice of all weights g equal to unity will often be satisfactory for the present purpose. The desired high usage group size for the first routed traffic is then found from standard E_1-tables showing trunks x required, as a function of offered traffic a and proportion overflow $\bar{b}$.

Since the different parcels of traffic have varying proportions b of their loads overflowing to the final route, by equation (34) the parcel with the largest proportion will determine the permitted value of R_2. Thus

$$R_2 = P/b_{\max} \tag{36}$$

where P is the specified poorest overall service (say 0.03) for any parcel. It may be noted that on occasion some one parcel, perhaps a small one, may provide an outstandingly large $b_{\max}$ value, which will tend to give a considerably better than required service to all the major traffic parcels. Some compromise with a literal application of a fixed poorest service criterion may be indicated in such cases.

An alternative and somewhat simpler procedure here is to use an average value $\bar{b}$ in (36) instead of $b_{\max}$, with a compensating modification of P, so that substantially the same R_2 is obtained as before. The allowance in P will be influenced by the choice of weights g in (35). It will commonly be found in practice that overflow proportions to final groups for large parcels of traffic are lower than for small parcels. Choosing all weights, as unity, as opposed to weighting by traffic volumes for example, tends to insert a small element of service protection for those traffic parcels (often the smaller ones) with the higher prportionate high usage group overflows.

Having determined R_2, a ready means is needed for estimating the required number of final route trunks, Curves for this purpose are provided on Figs. 46 to 49, within whose range, $R_2 = 0.01$ to 0.10, it will usually be sufficiently accurate to interpolate for trunk engineering purposes. These R_2-curves exactly parallel the R_1-curves for use when first routed traffic is offered directly to the final group without benefit of individual high usage trunks. If R_2 is well outside the charted range a run-through of the ER calculations may be required.

8.3 *Area in Which Significant Savings in Final Route Trunks are Realized by Allowing for the Preferred Service Given a First Routed Traffic Parcel*

Considerable effort has been expended by alternate route research workers in various countries to discover and evaluate those areas where first routed (random) traffic offered to a final route enjoys a substantial service advantage over competing parcels of traffic which have over-

flowed from high usage groups. A comparison of Figs. 42 to 45, (which indicate trunk provision for meeting a first routed traffic criterion R_1) with Figs. 46 to 49 (which indicate trunk provision for meeting a composite-load-offered-to-the-final-route criterion R_2) gives a means for deciding under what conditions in practice it is important to distinguish between the two criteria. Fig. 50 shows the borders of areas, defined in terms of A' and V', the characterizing parameters of the total load offered to the final route, where a 2 and 5 per cent overprovision of final trunks would occur using R_2 for R_1 as the loss measure for first routed traffic. Thus in the alternate route examples displayed in Table XV, where $x = 8$, $g = 2$ to 10, $A' = 4.0$ and V' varies from 5.80 to 5.45, Fig. 50 shows that by failing to allow for the preferred position of the 2 erlang first routed parcel, we should at $R = 0.02$ engineered loss, provide a little over 5 per cent more final trunks than necessary. (Actually 10.2 and 9.9 versus 9.6 and 9.4 trunks for $g = 2$ and 10, respectively.)

The curves of Fig. 50 for final route loads larger than a few erlangs, are almost straight lines. At an objective engineering base of $R = 0.03$, for example, the 2 and 5 per cent trunk overprovision areas through using R_2 instead of R_1 are outlined closely by:

$$2 \text{ per cent overprovision occurs at } V'/(A' - 1) \doteq 1.4$$
$$5 \text{ per cent overprovision occurs at } V'/(A' - 1) \doteq 1.8.$$

Thus in the range of loads covered by Fig. 50, one might conclude that useful and determinable savings in final trunks can be achieved by use of the specialized R_1-curves instead of the more general R_2-curves, when the ratio $V'/(A' - 1)$ exceeds some figure in the 1.4 to 1.8 range, say 1.6. (In the examples just cited the $V'/(A' - 1)$ ratio is approximately 1.9.)

8.4. *Character of Traffic Carried on Non-Final Routes*

Telephone traffic which is carried by a non-final route will ordinarily be subjected to a peak clipping process which will depress the variance of the carried portion below that of the offered load. If this traffic terminates at the distant end of the route, its character, while conceivably affecting the toll and local switching trains in that office, will not require further consideration for intertoll trunk engineering. If, however, some or all of the route's load is to be carried on toll facilities to a more distant point (the common situation), the character of such parcels of traffic will be of interest in providing suitable subsequent paths. For this purpose it will be desirable to have etimates of the mean and variance of these carried parcels.

When a random traffic of "a" erlangs is offered to a group of "c" paths

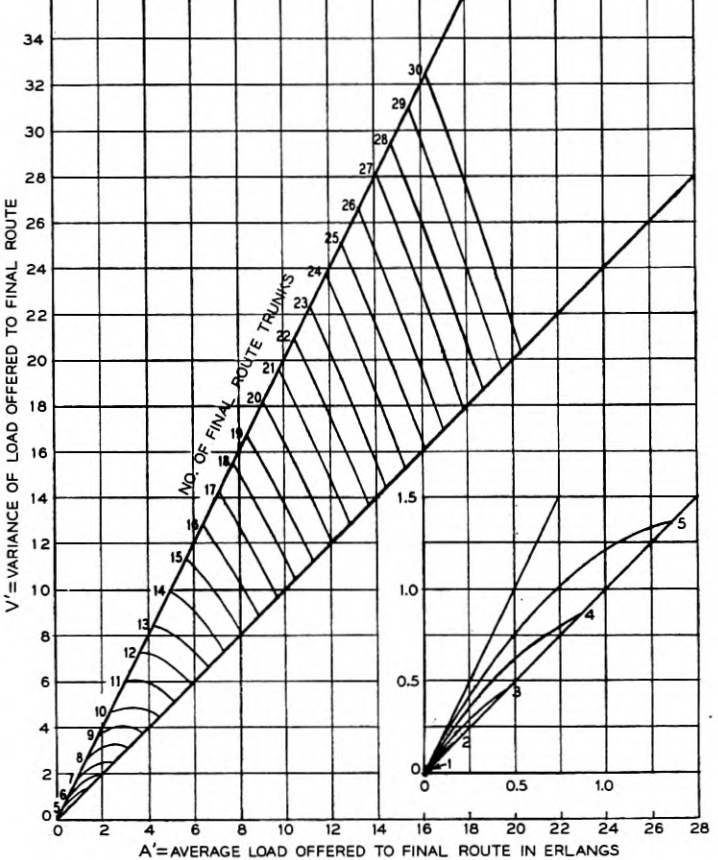


Fig. 46 — Provision of final route trunks to give combined offered load a service of $R_2 = 0.01$.

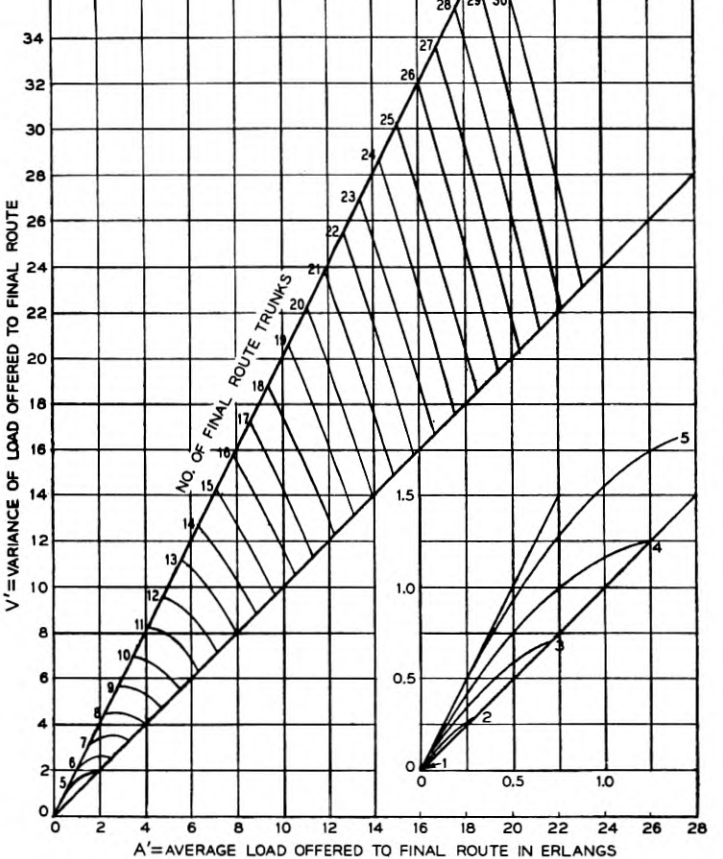


Fig. 47 — Provision of final route trunks to give combined offered load a service of $R_2 = 0.03$.

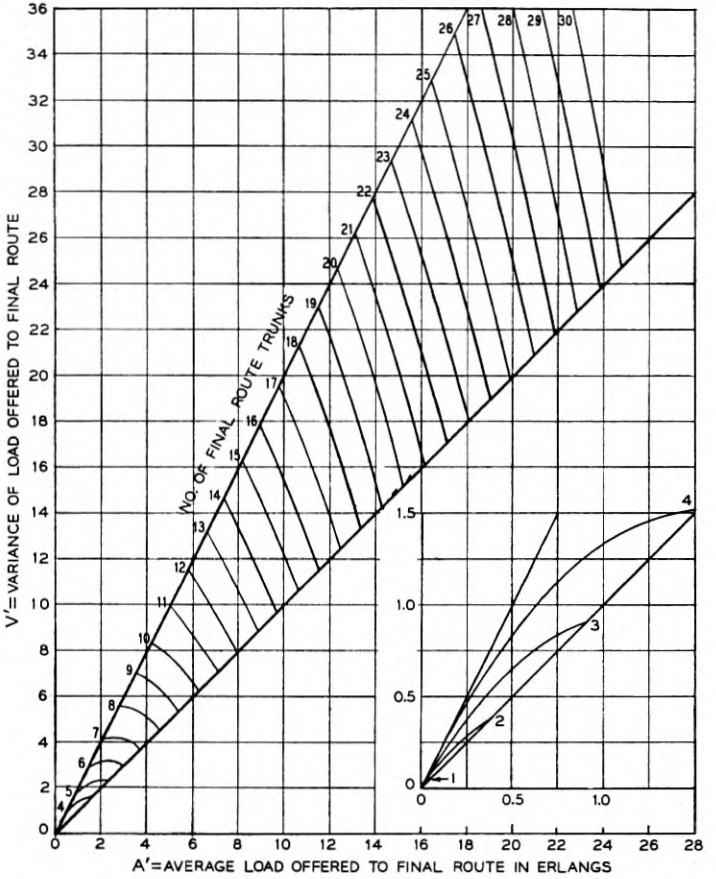


Fig. 48 — Provision of final route trunks to give combined offered load a service of $R_a = 0.05$

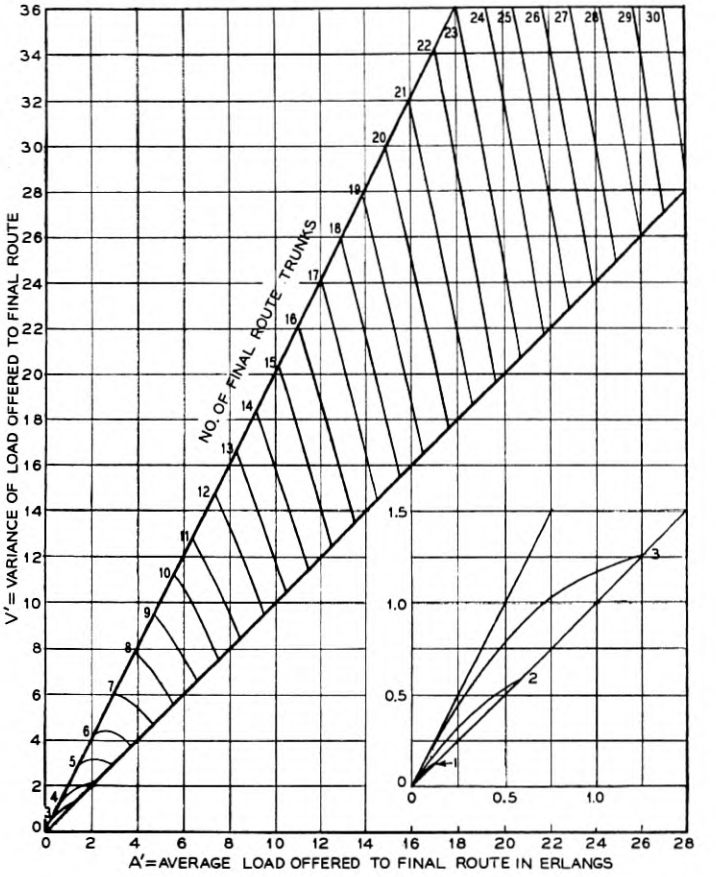


Fig. 49 — Provision of final route trunks to give combined offered load a service of $R_a = 0.10$

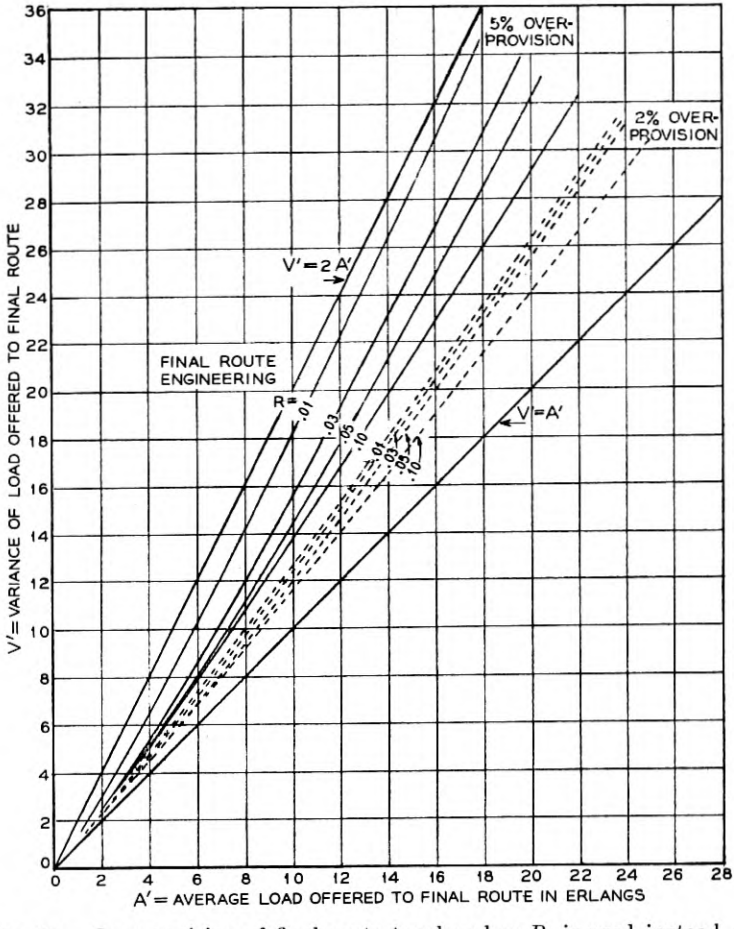


Fig. 50 — Overprovision of final route trunks when R_2 is used instead of R_1 as service to first routed traffic.

and overflowing calls do not return, the variance of the carried load is

$$V_{cd} = a[1 - E_{1,c}(a)]\,[1 + aE_{1,c}(a) - aE_{1,c-1}(a)]^* \qquad (37)$$

and the ratio of variance to average of the carried load is

$$\begin{aligned} \frac{V_{cd}}{L} &= 1 - a\,[E_{1,c-1}(a) - E_{1,c}(a)]^* \\ &= 1 - \left(\frac{c}{L} - 1\right)(a - L)^* \\ &= 1 - \ell_c \end{aligned} \qquad (38)$$

* These particular forms are due to P. J. Burke.

From (38) it is easy to see that

$$V_{cd} = L(1 - \ell_c)$$
$$= (\text{Load carried by the group})(1 - \text{load on last trunk}) \quad (39)$$

This is a convenient relationship since for high usage trunk study work, both the loads carried (in erlangs) on the group and on the last trunk will ordinarily be at hand.

If the high usage group's load is to be split in various directions at the distant point for re-offer to other groups, it would appear not unreasonable to assign a variance to each portion so as to maintain the ratio expressed in equation (38). That is, if a carried load L is divided into parts λ_1, $\lambda_2 \cdots$ where $L = \lambda_1 + \lambda_2 \cdots$, then the associated variances γ_1, $\gamma_2 \ldots$ would be

$$\begin{aligned} \gamma_1 &= \lambda_1 (1 - \ell_c) \\ \gamma_2 &= \lambda_2 (1 - \ell_c) \\ &\cdots\cdots\cdots\cdots \end{aligned} \quad (40)$$

If, however, the load offered to the group is non-random (e.g., the group is an intermediate route in a multi-alternate route system), the procedure is not quite so simple as in the random case just discussed. Equation (32) expresses the variance V_c of the carried load on a group of C paths whose offered traffic consists of the overflow from a first group of S paths to which a random load of A erlangs has been offered. V_c could of course be expressed in terms of A', V' and C, and curves or tables constructed for working purposes. However, such are not available, and in any case might be unwieldy for practical use.

A simple alternative procedure can be used which yields a conservative (too large) estimate of carried load variance. With random load offered to a divided two stage multiple of x paths followed by y paths, a positive correlation exists between the numbers m and n of calls present simultaneously on the x and y paths, respectively. Then the variance V_{m+n} of the $m + n$ distribution is greater than the sum of the individual variances of m and n,

$$V_{m+n} \geq V_m + V_n$$

or

$$V_m \leq V_{m+n} - V_n \quad (41)$$

Now n can be chosen arbitrarily, and if made very large, V_{m+n} becomes the offered load variance, and V_n the overflow load variance. Both of these are usually (or can be made) available. Their difference then, according to (41) gives an upper limit to V_m, the desired carried load

TABLE XVI—APPROXIMATE DETERMINATION OF THE VARIANCE OF CARRIED LOADS;
x lower paths, 8 upper paths; offer to upper paths = 3 erlangs

Lower Paths, x					Upper Paths, y			
No. Lower Paths x	Random offered load $A\ (= V)$	Variance of overflow V_n	Estimated variance of carried load $V - V_n$	True variance of carried load Eq (37)	Variance of offer $V'\ (= V_n)$ (Col 3)	Variance of overflow V''	Estimated variance of cd load $V' - V''$	True variance of cd load (Brockmeyer)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	3.00	3.00	0	0	3.00	0.035	2.97	2.853
3	5.399	4.05	1.35	0.60	4.05	0.121	3.93	3.664
6	7.856	4.98	2.88	1.418	4.95	0.236	4.74	4.175
12	12.882	6.22	6.66	3.538	6.22	0.520	5.70	4.790

variance. Corresponding reasoning yields the same conclusion when the offered load before the x paths is non-random.

A numerical example by Brockmeyer[11] while clearly insufficient to establish the degree of the inequality (41), indicates something as to the discrepancy introduced by this approximate procedure. Comparison with the true values is shown in Table XVI.

In the case of random offer to the 0, 3, 6, 12 "lower paths," the approximate method of equation (41) overestimates the variance of the carried load by nearly two to one (columns 4 and 5 of Table XVI). The exact procedure of (37) is then clearly desirable when it is applicable, that is when random traffic is being offered. For the 8 upper paths to which non-random load is offered (the non-randomness is suggested by comparing the variance of column 6 in Table XVI with the average offered load of 3 erlangs), the approximate formula (41) gives a not too extravagant overestimate of the true carried load variance. Until curves or tables are computed from equation (32), it would appear useful to follow the above procedure for estimating the carried load variance when non-random load is offered.

8.5. *Solution of a Typical Toll Multi-Alternate Route Trunking Arrangement: Bloomsburg, Pa.*

In Fig. 9 a typical, moderately complex, toll alternate route layout was illustrated. It is centered on the toll office at Bloomsburg, Pa. The loads to be carried between Bloomsburg and the ten surrounding cities are indicated in *CCS* (hundred call seconds per hour of traffic; 36 *CCS* = 1 erlang). The numbers of direct high usage trunks shown are assumed to have been determined by an economic study; we are asked to find

the number of trunks which should be installed on the Bloomsburg-Harrisburg route, so that the last trunk will carry approximately 18 *CCS* (0.50 erlang). Following this determination, (*a*) the number of final trunks from Bloomsburg to Scranton is desired so that the poorest service given to any of the original parcels of traffic will be no more than 3 calls in 100 meeting *NC*. Also (*b*) the modified Bloomsburg-Scranton trunk arrangement is to be determined when a high usage group is provided for the first routed traffic.

Solution (a): First Routed Traffic Offered Directly to Final Group

The offered loads in *CCS* to each distant point are shown in column (2) of Table XVII; the corresponding erlang values are in column (3). Consulting Figs. 12 and 13, the direct group overflow load parameters, average and variance, are read and entered in columns (5) and (6) respectively for the four groups overflowing to Harrisburg, and in columns (7) and (8) for the four groups directly overflowing to Scranton. The variance for the direct Bloomsburg-Harrisburg traffic equals its average; likewise for the direct Bloomsburg-Scranton traffic. They are so entered in the table. The parameters of the total load on the Harrisburg group are found by totalling, giving $A' = 11.19$, and $V' = 19.90$.

The required size C_1 of the Harrisburg group is now determined by the Equivalent Random theory. Entering Fig. 25 with A' and V' just determined, the ER values of trunks and load found are $S_1 = 13.55$, and $A_1 = 23.75$. C_1 is to be selected so that on a straight group of $S_1 + C_1$ trunks with offered load A, the last trunk will carry 0.50 erlang. Reading from Fig. 40, the load carried by the 26th trunk approximates this figure. Hence $C_1 = 26 - S_1 = 12.45$ trunks; or choose 12 trunks.

The overflow load's mean and variance from the Harrisburg group with 12 trunks, is now read from Figs. 12 and 13, entering with load $A_1 = 23.75$ and $C_1 + S_1 = 25.55$ trunks. The overflow values ($\alpha' = 2.50$ and $v' = 7.50$) are entered in columns (7) and (8) of the table. The total offered load to Scranton is now obtained by totalling columns (7) and (8), giving $A'' = 16.27$ and $V'' = 25.60$.

We desire now to know the number of trunks C_2 for the Scranton group which will provide *NC* 3 per cent of the time to the poorest service parcel of traffic, i.e., the first routed Bloomsburg-Scranton parcel. The $R_1 = 0.03$ and $R_2 = 0.03$ solutions are available, the former of course being more closely applicable. A check reference to Fig. 50 shows a difference of approximately 4 per cent in trunk provision would result from the two methods. Entering Figs. 43 and 47 with $A'' = 16.27$ and

TABLE XVII — ILLUSTRATIVE CALCULATION OF ALTERNATE ROUTE TRUNKS AT BLOOMSBURG, PA.

Distant Office	Load Between Bloomsburg and Distant Office		No. Trunks Between Bloomsburg and Distant Office x	Characteristics of Load to Harrisburg Group		Characteristics of Load to Scranton Group		Approximate Proportion of Original Offer Going To Final Route
	CCS	Erlangs		α (Fig. 12)	β (Fig. 13)			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
								$\frac{\text{Col. 5}}{\text{Col. 3}} \times \frac{2.50}{11.19}$
Pottsville	26	0.72	1	0.30	0.35			0.093
Shamokin	540	15.0	15	2.70	6.35			0.040
Sunbury	691	19.19	20	2.66	7.00	2.50	7.50	0.031
Williamsport	160	4.44	5	1.06	1.73			0.053
Harrisburg	161	4.47	C_1	4.47	4.47			0.223
				$A' = 11.19$	$V' = 19.90$			Col 7 ÷ Col 3
Frackville	123	3.42	5			0.50	0.81	0.146
Hazleton	836	23.22	28			1.28	3.90	0.055
Wilkes-Barre	228	6.33	8			0.89	1.68	0.141
Philadelphia	154	4.28	5			0.96	1.57	0.224
Scranton	365	10.14	C_2			10.14	10.14	$\bar{b} = 0.112$
						$A'' = 16.27$	$V'' = 25.60$	= unweighted average

$A_1 = 23.75$ $C_1 = 12$ $S_1 = 13.55$

Final route trunks required = 24.1. (Read from Fig. 43 using $A''' = 16.57$, $V''' = 25.90$ for 3 per cent first routed retrials.)

$V'' = 25.60$, we obtain the trunk requirements:

R_1 Method..... 23.8 trunks
R_2 Method..... 24.8 trunks

Thus the more precise method of solution here yields a reduction of 1.0 in 25 trunks, a saving of 4 per cent, as had been predicted.

The above calculation is on a Lost Calls Cleared basis. Since the overflow direct traffic calls will return to this group to obtain service, to assure their receiving no more than 3 per cent NC, the provision of the final route would theoretically need to be slightly more liberal. An estimate of the allowance required here may be made by adding the expected erlangs loss Δ for the direct traffic (most of the final route overflow calls which come from high usage routes will be carried by their respective groups on the next retrial) to both the A'' and V'' values previously obtained, and recalculating the trunks required from that point onward. (In fact this could have been included in the initial computation.) Thus:

$$\Delta = 0.03 \times 10.14 = 0.30 \text{ erlang}$$
$$A''' = 16.27 + 0.30 = 16.57 \text{ erlangs}$$
$$V''' = 25.60 + 0.30 = 25.90 \text{ erlangs}$$

Again consulting Figs. 43 and 47 gives the corresponding final trunk values

R_1 Method..... 24.1 trunks
R_2 Method..... 25.1 trunks

Of the above four figures for the number of trunks in the Scranton route, the R_1-Method with retrials, i.e., 24.1 trunks, would appear to give the best estimate of the required trunks to give 0.03 service to the poorest service parcel.

Solution (b): With High Usage Group Provided for First Routed Traffic

Following the procedure outlined in Section 8.2, we obtain an average of the proportions overflowing to the final route for all offered load parcels. The individual parcel overflow proportion estimates are shown in the last column of Table XVII; their unweighted average is 0.112. With a first routed offer to Scranton of 10.14 erlangs, a provision of 12 high usage trunks will result in an overflow of $\alpha = 1.26$ erlangs, or a *proportion* of 0.125 which is the value most closely attainable to the objective 0.112. With 12 trunks the overflow variance is found to be 2.80.

Replacing 10.14 in columns 7 and 8 of Table XVII with 1.26 and 2.80, respectively, gives new estimates characterizing the offer to the final route, $A'' = 7.39$ and $V'' = 18.26$. We now proceed to insure that the poorest service parcel obtains 0.03 service. This occurs on the Philadelphia and Harrisburg groups, which overflow to the final group approximately 0.224 of their original offered loads. The final group must then, according to equation (34) be engineered for

$$R_2 = 0.03/0.224 = 0.134 \text{ service.}$$

This value lies above the highest R_2 engineering chart (Fig. 49, $R_2 = 0.10$), so an ER calculation is indicated.

The Equivalent Random average is 28.6 erlangs, and $S = 23.5$ trunks. We determine the total trunks $S + R$ which, with 28.6 erlangs offered, will overflow $0.134(7.39) = 0.99$ erlang. From Fig. 12.2, 35.6 trunks are required. Then the final route provision should be $C = 35.6 - 23.5 = 12.1$ trunks; and a total of $12 + 12.1$ or 24.1 Scranton trunks is indicated.

Simplified Alternative Solution: In Section 8.2 a simplified approximate procedure was described using a modified probability P' for the average overall service for all parcels of traffic, instead of P for the poorest service parcel. Suppose $P' = 0.01$ is chosen as being acceptable. Then

$$R_2 = \frac{P'}{\bar{b}} = \frac{0.01}{0.112} = 0.089$$

Interpolating between the $R_2 = 0.05$ and 0.10 curves (Figs. 48 and 49) gives with $A'' = 7.39$ and $V'' = 18.26$, $C = 13.4$, the number of final trunks required. Again the same result could have been obtained by making the suitable ER computation. It may be noted that if P' had been chosen as 0.015 (one-half of P), R_2 would have become 0.134, exactly the same value found in the poorest-service-parcel method. The final trunk provision, of course, would have again been 12.1 trunks.

Disscussion

In the first solution above, 24.1 full access final trunks from Bloomsburg to Scranton were required. The service on the first routed traffic was 0.03; however, the service enjoyed by the offered traffic as a whole was markedly better than 0.03. The corresponding ER calculation shows ($A = 28.3$, $S + C = 12.3 + 24.1$) a total overflow of $\alpha'' = 0.72$ erlangs, or an overall service of $0.72/91.21 = 0.008$.

In the second solution, 12 high usage and 12.1 common final, or a total of 24.1, trunks were again required, to give 0.03 service to the poorest service parcels of offered load. The overall service here, however, was 0.99/91.21 = 0.011. Thus, with the same number of paths provided, in the second solution (high usage arrangement) the overall call loss was 40 pes cent larger than in the first solution.* However, it may well be desirable to accept such an average service penalty since by providing high usage trunks for the first routed traffic, the latter's service cannot be degraded nearly so readily should heavy overloads occur momentarily in the other parcels of traffic.

9. CONCLUSION

As direct distance dialing increases, it will be necessary to provide intertoll paths so that substantially no-delay service is given at all times. To do this economically, automatic multi-alternate routing will replace the present single route operation. Traffic engineering of these complicated trunking arrangements will be more difficult than with simple intertoll groups.

One of the new problems is to describe adequately the non-random character of overflow traffic. In the present paper this is proposed to be done by employing both mean and variance values to describe each parcel of traffic, instead of only the mean as used heretofore. Numerous comparisons are made with simulation results which indicate that adequate predictive reliability is obtained by this method for most traffic engineering and administrative purposes. Working curves are provided by which trunking arrangements of considerable complexity can readily be solved.

A second problem requiring further review is the day-to-day variation among the primary loads and their effect on the alternate route system's grade of service. A thorough study of these variations will permit a re-evaluation of the service criteria which have tentatively been adopted. A closely allied problem is that of providing the necessary kind and amounts of traffic measuring devices at suitable points in the toll alternate route systems. Requisite to the solution of both of these problems is an understanding of traffic flow character in a complex overflow-type

* The actual loss difference may be slightly greater than estimated here since in the first solution (complete access final trunks), an allowance was included for return attempts to the final route by first routed calls meeting an 0.03 loss, while in the second solution (high usage group for first routed traffic) no return attempts to the final route were considered. These would presumably be small since only 1 per cent of all calls would overflow and most of these upon retrial would be handled on their respective high usage groups.

of trunking plan, and a method for estimating quantitatively the essential fluctuation parameters at each point in such a system. The present paper has undertaken to shed some light on the former, and to provide an approximate yet sufficiently accurate method by which the latter can be accomplished. It may be expected then that these studies, as they are developed, will provide the basis for assuring an adequate direct distance dialing service at all times with a minimum investment in intertoll trunk facilities.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the technical and mathematical assistance of his associates, Mrs. Sallie P. Mead, P. J. Burke, W. J. Hall, and W. S. Hayward, in the preparation of this paper. Dr. Hall provided the material on the convolution of negative binomials leading to Fig. 19. Mr. Hayward extended Kosten's curve E on Fig. 5 to higher losses by a calculating method involving the progressive squaring of a probability matrix. The author's thanks are also due J. Riordan who has summarized some of the earlier mathematical work of H. Nyquist and E. C. Molina, as well as his own, in the study of overflow load characteristics; this appears as Appendix I.

The extensive calculations and chart constructions are principally the work of Miss C. A. Lennon.

REFERENCES

1. Rappleye, S. C., A Study of the Delays Encountered by Toll Operators in Obtaining an Idle Trunk, B.S.T.J., **25**, p. 539, Oct., 1946.
2. Kosten, L., Over de Invloed van Herhaalde Oproepen in de Theorie der Blokkeringskausen, De Ingenieur, **59**, p. E123, Nov. 21, 1947.
3. Clos, C., An Aspect of the Dialing Behavior of Subscribers and Its Effect on the Trunk Plant, B.S.T.J., **27**, p. 424, July, 1948.
4. Kosten, L., Uber Sperrungswahrscheinlichkeiten bei Staffelschaltungen, E.N.T., **14**, p. 5, Jan., 1937.
5. Kosten, L., Over Blokkeerings-en Wachtproblemen, Thesis, Delft, 1942.
6. Molina, E. C., Appendix to: Interconnection of Telephone Systems — Graded Multiples (R. I. Wilkinson), B.S.T.J., **10**, p. 531, Oct., 1931.
7. Vaulot, A. E., Application du Calcul des Probabilités à l'Exploitation Téléphonique, Revue Gen. de l'Electricité, **16**, p. 411, Sept. 13, 1924.
8. Lundquist, K., General Theory for Telephone Traffic, Ericsson Technics, **9**, p. 111, 1953.
9. Berkeley, G. S., Traffic and Trunking Principles in Automatic Telephony, 2nd revised edition, 1949, Ernest Benn, Ltd., London, Chapter V.
10. Palm, C., Calcul Exact de la Perte dans les Groupes de Circuits Échelonnés, Ericsson Technics, **3**, p. 41, 1936.
11. Brockmeyer, E., The Simple Overflow Problem in the Theory of Telephone Traffic, Teleteknik, **5**, p. 361, December, 1954.

ABRIDGED BIBLIOGRAPHY OF ARTICLES ON TOLL ALTERNATE ROUTING

Clark, A. B., and Osborne, H. S., Automatic Switching for Nationwide Telephone Service, A.I.E.E., Trans., **71**, Part I, p. 245, 1952. (Also B.S.T.J., **31**, p. 823, Sept., 1952.)
Pilliod, J. J., Fundamental Plans for Toll Telephone Plant, A.I.E.E. Trans., **71**, Part I, p. 248, 1952. (Also B.S.T.J., **31**, p. 832, Sept., 1952.)
Nunn, W. H., Nationwide Numbering Plan, A.I.E.E. Trans., **71**, Part I, p. 257, 1952. (Also B.S.T.J., **31**, p. 851, Sept., 1952.)
Clark, A. B., The Development of Telephony in the United States, A.I.E.E. Trans., **71**, Part I, p. 348, 1952.
Shipley, F. F., Automatic Toll Switching Systems, A.I.E.E. Trans., **71**, Part I, p. 261, 1952. (Also B.S.T.J., **31**, p. 860, Sept., 1952.)
Myers, O., The 4A Crossbar Toll System for Nationwide Dialing, Bell Lab. Record, **31**, p. 369, Oct., 1953.
Clos, C., Automatic Alternate Routing of Telephone Traffic, Bell Lab. Record, **32**, p. 51, Feb., 1954.
Truitt, C. J., Traffic Engineering Techniques for Determining Trunk Requirements in Alternate Routing Trunk Networks, B.S.T.J., **33**, p. 277, March, 1954.
Molnar, I., Some Recent Advances in the Economy of Routing Calls in Nationwide Dialing, A.E. Tech. Jl., **4**, p. 1, Dec., 1954.
Jacobitti, E., Automatic Alternate Routing in the 4A Crossbar System, Bell Lab. Record, **33**, p. 141, April, 1955.

Appendix I*

DERIVATION OF MOMENTS OF OVERFLOW TRAFFIC

This appendix gives a derivation of certain factorial moments of the equilibrium probabilities of congestion in a divided full-access multiple used as a basis for the calculations in the text. These moments were derived independently in unpublished memoranda (1941) by E. C. Molina (the first four) and by H. Nyquist; curiously, the method of derivation here, which uses factorial moment generating functions, employs auxiliary relations from both Molina and Nyquist. Although these factorial moments may be obtained at a glance from the probability expressions given by Kosten in 1937, if it is remembered that

$$p(x) = \sum_{k=0}^{\infty} (-1)^{k-x} \binom{k}{x} \frac{M_{(k)}}{k!}, \tag{1.1}$$

where $p(x)$ is a discrete probability and $M_{(k)}$ is the kth factorial moment of its distribution, Kosten does not so identify the moments and it may be interesting to have a direct derivation.

Starting from the equilibrium formulas of the text for $f(m, n)$, the probability of m trunks busy in the specific group of x trunks, and n in

* Prepared by J. Riordan.

the (unlimited) common group, namely

$$\begin{aligned}(a + m + n)f(m, n) - (m + 1)f(m + 1, n) \\ - (n + 1)f(m, n + 1) - af(m - 1, n) = 0 \\ (a + x + n)f(x, n) - af(x, n - 1) \\ - (n + 1)f(x, n + 1) - af(x - 1, n) = 0\end{aligned} \tag{1.2}$$

and

$$f(m, n) = 0, \quad m < 0 \quad \text{or} \quad n < 0 \quad \text{or} \quad m > x,$$

factorial moment generating function recurrences may be found and solved.

With m fixed, factorial moments of n are defined by

$$M_{(k)}(m) = \sum_{n=0}^{\infty} (n)_k f(m, n) \tag{1.3}$$

or alternatively by the factorial moment exponential generating function

$$M(m, t) = \sum_{k=0}^{\infty} M_{(k)}(m)t^k/k! = \sum_{n=0}^{\infty} (1 + t)^n f(m, n) \tag{1.4}$$

In (1.3), $(n)_k = n(n - 1) \cdots (n - k + 1)$ is the usual notation for a falling factorial.

Using (1.4) in equations (1.2), and for brevity $D = d/dt$, it is found that

$$\begin{aligned}a + m + tD)M(m, t) - (m + 1)M(m + 1, t) \\ - aM(m - 1, t) = 0 \\ (x - at + tD)M(x, t) - aM(x - 1, t) = 0\end{aligned} \tag{1.5}$$

which correspond (by equating powers of t) to the factorial moment recurrences

$$\begin{aligned}(a + m + k)M_{(k)}(m) - (m + 1)M_{(k)}(m + 1) \\ - aM_{(k)}(m - 1) = 0 \\ (x + k)M_{(k)}(x) - akM_{(k-1)}(x) - aM_{(k)}(x - 1) = 0\end{aligned} \tag{1.6}$$

Notice that the first of (1.6) is a recurrence in m, which suggests (following Molina) introducing a new generating function defined by

$$G_k(u) = \sum M_{(k)}(m)u^m \tag{1.7}$$

Using this in (1.5), it is found that

$$\left[(a + k - au + (u - 1)\frac{d}{du}\right] G_k(u) = 0 \tag{1.8}$$

Hence

$$\frac{1}{G_k(u)}\frac{dG_k(u)}{du} = a + \frac{k}{1 - u} \tag{1.9}$$

and, by easy integrations,

$$G_k(u) = ce^{au}(1 - u)^{-k}, \tag{1.10}$$

with c an arbitrary constant, which is clearly identical with $G_k(0) = M_{(k)}(0)$.

Expansion of the right-hand side of (1.10) shows that

$$M_{(k)}(m) = M_{(k)}(0)\sum_{j=0}^{m}\binom{k + j - 1}{j}\frac{a^{m-j}}{(m - j)!} = M_{(k)}(0)\sigma_k(m), \tag{1.11}$$

if

$$\sigma_0(m) = a^m/m! \quad \text{and,} \quad \sigma_k(m) = \sum_{j=0}^{m}\binom{k + j - 1}{j}\frac{a^{m-j}}{(m - j)!} \tag{1.12}$$

The notation $\sigma_k(m)$ is copied from Nyquist; the functions are closely related to the $\varphi_x^{(n)}$ used by Kosten; indeed $\sigma_k(m) = e^a\varphi_m^{(k)}$. They have the generating function

$$g_k(u) = \sum_{m=0}^{\infty}\sigma_k(m)u^m = e^{au}(1 - u)^{-k} \tag{1.13}$$

from which a number of recurrences are found readily. Thus

$$\begin{aligned} g_k(u) &= (1 - u)g_{k+1}(u) \\ u\frac{dg_k(u)}{du} &= aug_k(u) + kug_{k+1}(u) \\ &= -ag_{k-1}(u) + (a - k)g_k(u) + kg_{k-1}(u) \end{aligned}$$

(the last by use of the first) imply

$$\begin{aligned} \sigma_k(m) &= \sigma_{k+1}(m) - \sigma_{k+1}(m - 1) \\ m\sigma_k(m) &= a\sigma_k(m - 1) + k\sigma_{k+1}(m - 1) \\ &= -a\sigma_{k-1}(m) + (a - k)\sigma_k(m) + k\sigma_{k+1}(m) \end{aligned}$$

The first of these leads to

$$\sigma_k(0) + \sigma_k(1) + \cdots + \sigma_k(x) = \sigma_{k+1}(x) \tag{1.14}$$

and the last is useful in the form

$$k\sigma_{k+1}(m) = (m + k - a)\sigma_k(m) + a\sigma_{k-1}(m) \tag{1.15}$$

Also, the first along with $\sigma_0(m) = a^m/m!$ leads to a simple calculation procedure, as Kosten has noticed.

By (1.11) the factorial moments are now completely determined except for $M_{(k)}(0)$. To determine the latter, the second of (1.6) and the normalizing equation

$$\sum_{m=0}^{x} M_0(m) = 1 \tag{1.16}$$

are available.

Thus from the second of (1.6)

$$[(x + k)\sigma_k(x) - a\sigma_k(x - 1)]M_{(k)}(0) = ak\sigma_{k-1}(x)M_{(k-1)}(0) \tag{1.17}$$

Also

$$\begin{aligned}(x + k)\sigma_k(x) - a\sigma_k(x - 1) &= (x + k - a)\sigma_k(x) + a[\sigma_k(x) - \sigma_k(x - 1)] \\ &= (x + k - a)\sigma_k(x) + a\sigma_{k-1}(x) \\ &= k\sigma_{k+1}(x),\end{aligned}$$

the last step by (1.15). Hence

$$M_{(k)}(0) = a\,\frac{\sigma_{k-1}(x)}{\sigma_{k+1}(x)}\,M_{(k-1)}(0) \tag{1.18}$$

and by iteration

$$M_{(k)}(0) = a^k\,\frac{\sigma_1(x)\sigma_0(x)}{\sigma_{k+1}(x)\sigma_k(x)}\,M_0(0) \tag{1.19}$$

From (1.11) and (1.16), and in the last step (1.14),

$$\sum_{m=0}^{x} M_0(m) = \sum_{m=0}^{x} M_0(0)\sigma_0(m) = M_0(0)\sigma_1(x) = 1 \tag{1.20}$$

Hence finally

$$\begin{aligned}M_{(k)}(m) &= M_{(k)}(0)\sigma_k(m) \\ &= a^k\,\frac{\sigma_0(x)\sigma_k(m)}{\sigma_{k+1}(x)\sigma_k(x)}\end{aligned} \tag{1.21}$$

and

$$M_{(k)} = \sum_{m=0}^{x} M_{(k)}(m) = a^k \sigma_0(x)/\sigma_k(x) \tag{1.22}$$

Ordinary moments are found from the factorial moments by linear relations; thus if m_k is the kth ordinary moment (about the origin)

$$m_0 = M_{(0)} \qquad m_1 = M_{(1)} \qquad m_2 = M_{(2)} + M_{(1)}$$
$$m_3 = M_{(3)} + 3M_{(2)} + M_{(1)}$$

Thus

$$m_0(m) = \sigma_0(m)/\sigma_1(x)$$
$$m_1(m) = a\sigma_1(m)\sigma_0(x)/\sigma_1(x)\sigma_2(x)$$
$$m_2(m) = a^2\sigma_2(m)\sigma_0(x)/\sigma_2(x)\sigma_3(x) + a\sigma_1(m)\sigma_0(x)/\sigma_1(x)\sigma_2(x)$$

and, in particular, using notation of the text

$$m_0(x) = \sigma_0(x)/\sigma_1(x) = E_{1,x}(a)$$

$$\alpha_x = \frac{m_1(x)}{m_0(x)} = a\,\frac{\sigma_1(x)}{\sigma_2(x)} = \frac{a}{x - a + 1 + aE_{1,x}(a)} \tag{1.23}$$

$$v_x = \frac{m_2(x)}{m_0(x)} - \alpha_x^2 = \frac{a^2\sigma_1(x)}{\sigma_3(x)} + \alpha_x - \alpha_x^2 \tag{1.24}$$
$$= \alpha_x[1 - \alpha_x + 2a(x + 2 + \alpha_x - a)^{-1}]$$

Finally the sum moments: $m_k = \sum_0^x m_k(m)$ are

$$m_0 = 1$$
$$m_1 = \alpha = a\sigma_0(x)/\sigma_1(x) = aE_{1,x}(a) \tag{1.25}$$

$$m_2 = a^2\sigma_0(x)/\sigma_2(x) + m_1 = m_1[a(x + 1 + m_1 - a)^{-1} + 1]$$
$$v = m_2 - m_1^2 = m_1[1 - m_1 + a(x + 1 + m_1 - a)^{-1}] \tag{1.26}$$

In these, $E_{1,x}(a) = \sigma_0(x)/\sigma_1(x)$ is the familiar Erlang loss function.

Appendix II — Character of Overflow Load When Non-Random Traffic Is Offered to a Group of Trunks

It has long been recognized that it would be useful to have a method by which the character of the overflow traffic could be determined when non-random traffic is offered to a group of trunks. Excellent agreement has been found in both throwdown and field observation over ranges of considerable interest with the "equivalent random" method of describ-

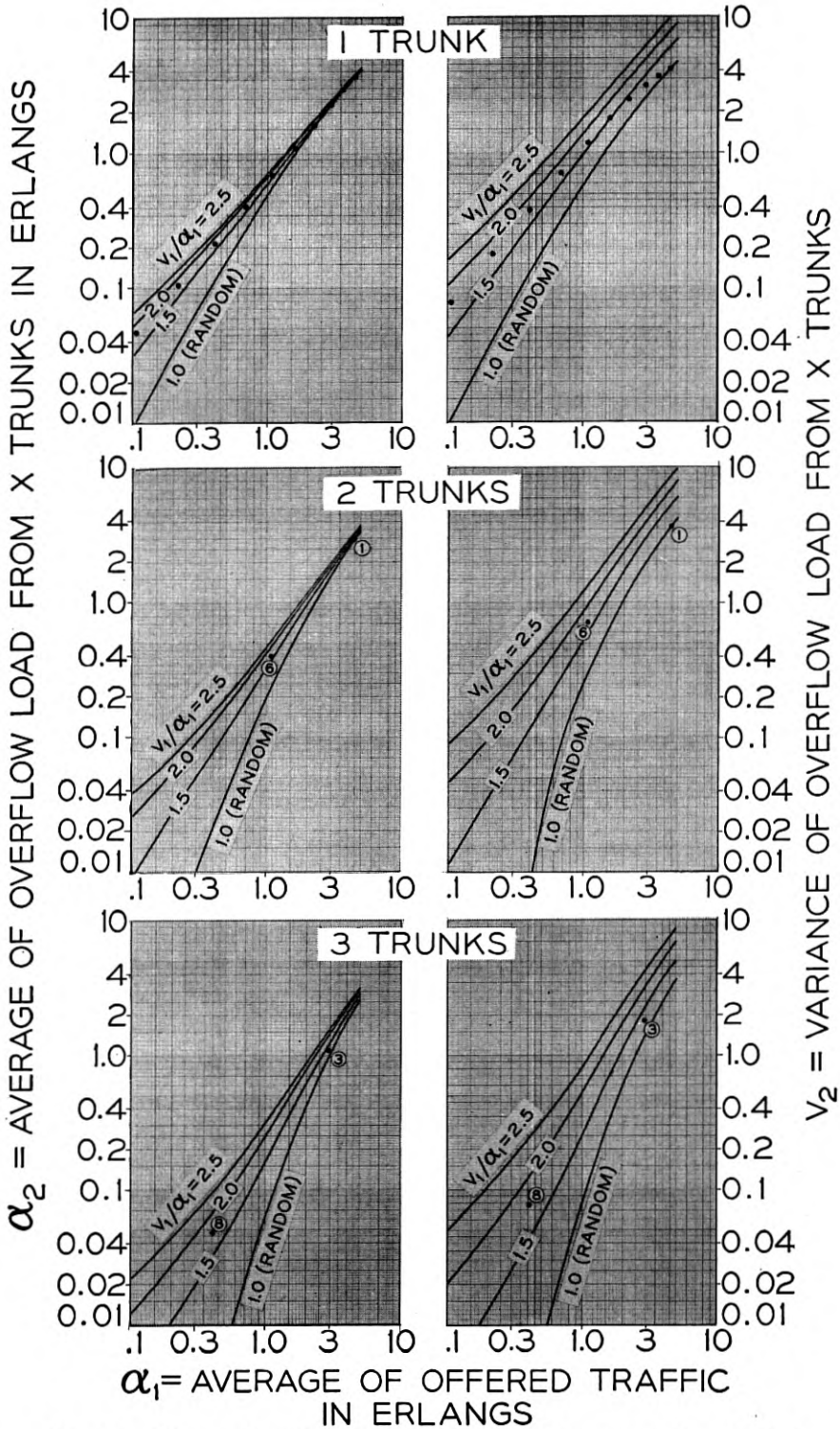
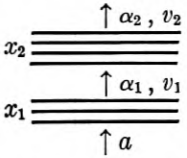


Fig. 51 — Mean and variance of overflow load when non-random traffic is offered to a group of trunks.

ing the character of non-random traffic. An approximate solution of the problem is offered based on this method.

Suppose a random traffic a is offered to a straight multiple which is divided into a lower x_1 portion and an upper x_2 portion, as follows:



From Nyquist's and Molina's work we know the mean and variance of the two overflows to be:

$$\alpha_1 = a \cdot E_{1,x_1}(a) = a \frac{\dfrac{a^{x_1}}{x_1!}}{1 + a + \dfrac{a^2}{2!} + \cdots + \dfrac{a^{x_1}}{x_1!}}$$

$$v_1 = \alpha_1 \left[1 - \alpha_1 + \frac{a}{x_1 - a + \alpha_1 + 1} \right]$$

$$\alpha_2 = a \cdot E_{1,x_1+x_2}(a)$$

$$v_2 = \alpha_2 \left[1 - \alpha_2 + \frac{a}{x_1 + x_2 - a + \alpha_2 + 1} \right]$$

Since α_1 and v_1 completely determine a and x_1, and these in turn, with x_2, determine α_2 and v_2, we may express α_2 and v_2 in terms of only α_1, v_1, and x_2. The overflow characteristics (α_2 and v_2), are then given for a non-random load (α_1 and v_1) offered to x trunks as was desired.

Fig. 51 of this Appendix has been constructed by the Equivalent Random method. The charts show the expected values of α_2 and v_2 when α_1, v_1 (or v_1/α_1), and x_2, are given. The range of α_1 is only 0 to 5 erlangs, and v/α is given only from the Poisson unity relation to a peakedness value of 2.5. Extended and more definitive curves or tables could readily, of course, be constructed.

The use of the curves can perhaps best be illustrated by the solution of a familiar example.

Example: A load of 4.5 erlangs is submitted to 10 trunks; on the "lost calls cleared" basis; what is the average load passing to overflow?

Solution: Compute the load characteristics from the first trunk when 4.5 erlangs of random traffic are submitted to it. These values are found to be $\alpha_1 = 3.68$, $v_1 = 4.15$. Now using α_1 and v_1 (or $v_1/\alpha_1 = 4.15/3.68 = 1.13$) as the offered load to the second trunk, read on the chart the parameters of the overflow from the second trunk, and so on. The successive overflow values are given in Table XVIII.

The proportion of load overflowing the group is then 0.0472/4.50 = 0.0105, which agrees, of course, with the Erlang $E_{1,10}(4.5)$ value. The successive overflow values are shown on the chart by the row of dots along the α_2 and v_2 1-trunk curves.

Instead of considering successive single-trunk overflows as in the example above, other numbers of trunks may be chosen and their overflows determined. For example suppose the 10 trunks are subdivided into 2 + 3 + 2 + 3 trunks. The loads overflowing these groups are given in Table XIX.

Again the overflow is 0.0472 erlang, or a proportion lost of 0.0105, which is, as it should be, the same as found in the previous example. The values read in this example are indicated by the row of dots marked 1, 3, 6, 8 on the 2-trunk and 3-trunk curves.

The above procedure and curves should be of use in obtaining an estimate of the character of the overflow traffic when a non-random load is offered to a group of paths.

TABLE XVIII — SUCCESSIVE NON-RANDOM OVERFLOWS

Trunk Number i	Characteristics of Load Offered to Trunk No. i (same as overflow from previous trunk)		
	Average	Variance	Ratio of variance to average
1	4.50	4.50	1.00 (Random)
2	3.68	4.15	1.13
3	2.92	3.68	1.26
4	2.22	3.11	1.40
5	1.61	2.46	1.53
6	1.09	1.80	1.64
7	0.694	1.19	1.72
8	0.406	0.709	1.75
9	0.217	0.377	1.74
10	0.106	0.180	1.70
Overflow	0.0472	0.077	1.64

TABLE XIX — SUCESSIVE NON-RANDOM OVERFLOWS

Trunk Number i	No. Trunks in Next Bundle	Offered Load Characteristics (same as overflow from previous trunk)		
		Average	Variance	Ratio of variance to average
1	2	4.50	4.50	1.00 (Random)
3	3	2.92	3.68	1.26
6	2	1.09	1.80	1.64
8	3	0.406	0.709	1.75
Overflow		0.0472	0.077	1.64

Crosstalk on Open-Wire Lines

By W. C. BABCOCK, ESTHER RENTROP, and C. S. THAELER

(Manuscript received September 29, 1955)

Crosstalk on open-wire lines results from cross-induction between the circuits due to the electric and magnetic fields surrounding the wires. The limitation of crosstalk couplings to tolerable magnitudes is achieved by systematically turning over or transposing the conductors that comprise the circuits. The fundamental theory underlying the engineering of such transposition arrangements was presented by A. G. Chapman in a paper entitled *Open-Wire Crosstalk* published in the Bell System Technical Journal in January and April, 1934.

There is now available a Monograph (No. 2520) supplementing Mr. Chapman's paper which reflects a considerable amount of experience resulting from the application of these techniques and provides a basis for the engineering of open-wire plant. The scope of the material is indicated by the following:

TRANSPOSITION PATTERNS

This describes the basic transposition types which define the number and locations of transpositions applied to the individual open-wire circuits.

TYPES OF CROSSTALK COUPLING

Crosstalk occurs both within incremental segments of line and between such segments. Furthermore, the coupling may result from cross-induction directly from a disturbing to a disturbed circuit or indirectly by way of an intervening tertiary circuit. On the disturbed circuit the crosstalk is propagated both toward the source of the original signal and toward the distant terminal. A knowledge of the relative importance of the various types of coupling is valuable in establishing certain time-saving approximations which facilitate the analysis of the total crosstalk picture.

TYPE UNBALANCE CROSSTALK

Crosstalk is measured in terms of a current ratio between the disturbing and disturbed circuits at the point of observation. Crosstalk between open-wire circuits is also generally computed in terms of a current ratio (cu) but it is also convenient to refer to it in terms of a coupling loss (db). The coupling in crosstalk units (cu) is the product of three terms: a coefficient dependent on wire configuration; a type unbalance dependent on transposition patterns; and frequency. The coefficient represents the coupling between relatively untransposed circuits of a specified length (1 mile) at a specific frequency (1 kc). The type unbalance is a measure of the inability to completely cancel out crosstalk by introducing transpositions because of interaction effects between the two halves of the exposure and because of propagation effects, primarily phase shift. Type unbalance is expressed in terms of a residual unbalance in miles and the frequency is expressed in kilocycles.

The coefficients applicable to lines built in accordance with certain standardized specifications are available in tabular form. When it is desired to obtain coefficients for other types of line, it is possible to compute approximate values which may be modified by correction factors to indicate the relationship between the computed values and measurements on carefully constructed lines.

Expressions for near-end type unbalance for certain simple types of exposures are developed and the formulas for all types of exposures are given. In addition, the values for near-end type unbalance are tabulated at 30° line angle intervals for lines where the propagation angle is 2,880° or less.

The principal component of far-end crosstalk between well transposed circuits results from compound couplings involving tertiary circuits. Again the expressions are developed for some of the exposures involving a few transpositions and the procedure for obtaining the formulas for any type of exposure is shown. Formulas are included for the types of exposures encountered in normal practice and the numerical values of far-end type unbalance are given at 30° intervals for line angles up to 2,880°.

SUMMATION OF CROSSTALK

The procedures referred to thus far evaluate the crosstalk occurring within a limited length of line known as a transposition section. In practice, however, a line is transposed as a series of sections. It is necessary, therefore, to determine how the crosstalk arising within the several

sections and that arising from interactions between the sections tend to combine. In a series of like transposition sections there is a tendency for the crosstalk to increase systematically, sometimes reaching intolerable magnitudes. This tendency can be controlled to a degree by introducing transpositions at the junctions between the sections, thus cancelling out some of the major components of the crosstalk. Complete cancellation is impossible because of interaction and propagation effects.

ABSORPTION

Since very significant couplings exist by way of tertiary circuits, it is possible for crosstalk to reappear on the disturbing circuit and thus strengthen or attenuate the original signal. This gives rise to the appearance of high attenuation known as absorption peaks in the line loss characteristic at certain critical frequencies. The evaluation of such pair-to-self coupling requires the use of coefficients which differ from those between different pairs and these are given for standard configurations.

STRUCTURAL IRREGULARITIES

It is impracticable to maintain absolute uniformity in the spacing between wires and in the spacing of transpositions. Thus there are unavoidable variations in the couplings between pairs from one transposition interval to the next. This in turn reduces the effectiveness of the measures to control the systematic or type unbalance crosstalk and produces what is known as irregularity crosstalk. Since the occurrence of structural irregularities tends to follow a random distribution, it is possible to evaluate it statistically and procedures for doing so are included. In addition to this direct effect of structural irregularities, there is a component of crosstalk resulting from the combination of systematic and random unbalances. A method is developed for estimating the magnitude of this important component of crosstalk.

EXAMPLES

In order to demonstrate how the procedures and data are used in solving practical problems, there is included the development of a transposition system to satisfy certain assumed conditions. This is carried through to the selection of transposition types for one transposition section and the selection of suitable junction transpositions.

Additional examples of transposition engineering are given in the form

of several transposition systems which have been widely used in the Bell System. These include:

Exposed Line — for voice frequency service.

C1 — for voice frequency and carrier service up to 30 kc.

J5 — for voice frequency and carrier operation up to 143 kc.

O1 — for voice frequency and compandored carrier operation up to 156 kc.

R1C — suitable for exchange lines with a limited number of carrier assignments.

Altogether, the theory, explanatory material, formulas and comprehensive data included in the Monograph make it possible to estimate open-wire crosstalk couplings and provide the necessary background for the development of new transposition systems.

Bell System Technical Papers Not Published in This Journal

ALSBERG, D. A.[1]

6-KMC Sweep Oscillator, I.R.E. Trans., **PGI-4,** pp. 32–39, Oct., 1955.

ANDERSON, J. R.,[1] BRADY, G. W.,[1] MERZ, W. J.,[1] and REMEIKA, J. P.[1]

Effects of Ambient Atmosphere on the Stability of Barium Titanate, J. Appl. Phys., Letter to the Editor, **26,** pp. 1387–1388, Nov., 1955.

ANDERSON, O. L.,[1] and ANDREATCH, P.[1]

Stress Relaxation in Gold Wire, J. Appl. Phys., **26,** pp. 1518–1519, Dec., 1955.

ANDERSON, P. W.,[1] and HASEGAWA, H.[5]

Considerations on Double Exchange, Phys. Rev., **100,** pp. 675–681, Oct. 15, 1955.

ANDERSON, P. W.[1]

Electromagnetic Theory of Cyclotron Resonance in Metals, Phys. Rev., Letter to the Editor, **100,** pp. 749–750, Oct. 15, 1955.

ANDREATCH, P., see Anderson, O. L.

AUGUSTINE, C. F., see Slocum, A.

BARSTOW, J. M.[1]

The ABC's of Color Television, Proc. I.R.E., **43,** pp. 1574–1579, Nov., 1955.

BARTLETT, C. A.[2]

Closed-Circuit Television in the Bell System, Elec. Engg., **75,** pp. 34–37, Jan., 1956.

1. Bell Telephone Laboratories, Inc.
2. American Telephone and Telegraph Company.
5. University of Tokyo, Japan.

BECKER, J. A.[1]

Adsorption on Metal Surfaces and Its Bearing on Catalysis, Advances in Catalysis, 1955, Nov., 1955.

BÖMMEL, H. E.[1]

Ultrasonic Attenuation in Superconducting and Normal-Conducting Tin at Low Temperatures, Phys. Rev., Letter to the Editor, **100,** pp. 758–759, Oct. 15, 1955.

BEMSKI, G.[1]

Lifetime of Electrons in p-Type Silicon, Phys. Rev., **100,** pp. 523–524, Oct. 15, 1955.

BENNETT, W. R.[1]

Steady State Transmission Through Networks Containing Periodically Operated Switches, Trans. I.R.E., **PGC.T., 2,** pp. 17–21, Mar., 1955.

BÖMMEL, H. E.,[1] MASON, W. P.,[1] and WARNER, A. W., JR.[1]

Experimental Evidence for Dislocation in Crystalline Quartz, Phys. Rev., Letter to the Editor, **99,** pp. 1895–1896, Sept. 15, 1955.

BRADLEY, W. W., see Compton, K. G.

BRATTAIN, W. H., see Buck, T. M., and Pearson, G. L.

BRADY, G. W., see Anderson, J. R.

BROWN, W. L.[1]

Surface Potential and Surface Charge Distribution from Semiconductor Field Effect Measurements, Phys. Rev., **100,** pp. 590–591, Oct. 15, 1955.

BUCK, T. M.,[1] and BRATTAIN, W. H.[1]

Investigations of Surface Recombination Velocities on Germanium by the Photoelectric Magnetic Method, J. Electrochem. Soc., **102,** pp. 636–640, Nov., 1955.

CETLIN, B. B., see Galt, J. K.

CHARNES, A., see Jacobson, M. J.

1. Bell Telephone Laboratories, Inc.

COMPTON, K. G.,[1] MENDIZZA, A.,[1] and BRADLEY, W. W.[1]

Atmospheric Galvanic Couple Corrosion, Corrosion, **11,** pp. 35–44, Sept., 1955.

CORENZWIT, E., see Matthias, B. T.

DAIL, H. W., JR., see Galt, J. K.

DILLON, J. F., JR.,[1] GESCHWIND, S.,[1] and JACCARINO, V.[1]

Ferromagnetic Resonance in Single Crystals of Manganese Ferrite, Phys. Rev., Letter to the Editor, **100,** pp. 750–752, Oct. 15, 1955.

DODGE, H. F.[1]

Chain Sampling Inspection Plan, Ind. Quality Control, **11,** pp. 10–13, Jan., 1955.

DODGE, H. F.[1]

Skip-lot Sampling Plan, Ind. Quality Control, **11,** pp. 3–5, Feb., 1955.

FAGEN, R. E.,[1] and RIORDAN, J.[1]

Queueing Systems for Single and Multiple Operation, J. S. Ind. Appl. Math., **3,** pp. 73–79, June, 1955.

FINE, M. E.[1]

Erratum: Elastic Constants of Germanium Between 1.7° and 80°K J. Appl. Phys., Letter to the Editor, **26,** p. 1389, Nov., 1955.

FLASCHEN, S. S.[1]

A Barium Titanate Synthesis from Titanium Esters, J. Am. Chem. Soc., **77,** p. 6194, Dec., 1955.

FLETCHER, R. C.,[1] YAGER, W. A.,[1] and MERRITT, F. R.[1]

Observation of Quantum Effects in Cyclotron Resonance, Phys. Rev., Letter to the Editor, **100,** pp. 747–748, Oct. 15, 1955.

FRANKE, H. C.[1]

Noise Measurement on Telephone Circuits, Tele-Tech., **14,** pp. 85–97, Mar., 1955.

1. Bell Telephone Laboratories, Inc.

GALT, J. K.,[1] YAGER, W. A.,[1] MERRITT, F. R.,[1] CETLIN, B. B.,[1] and DAIL, H. W., JR.[1]

Cyclotron Resonance in Metals: Bismuth, Phys. Rev., Letter to the Editor, **100,** pp. 748–749, Oct. 15, 1955.

GELLER, S.,[1] and THURMOND, C. D.[1]

On the Question of a Crystalline SiO, Am. Chem. Soc. J., **77,** pp. 5285–5287, Oct. 20, 1955.

GESCHWIND, S., see Dillon, J. F.

HARKER, K. J.[1]

Periodic Focusing of Beams from Partially Shielded Cathodes, I.R.E. Trans., **ED-2,** pp. 13–19, Oct., 1955.

HASEGAWA, H., see Anderson, P. W.

HAYNES, J. R.,[1] and HORNBECK, J. A.[1]

Trapping of Minority Carriers in Silicon II: n-type Silicon, Phys. Rev., **100,** pp. 606–615, Oct. 15, 1955.

HORNBECK, J. A., see Haynes, J. R.

ISRAEL, J. O.,[1] MECHLINE, E. B.,[1] and MERRILL, F. F.[1]

A Portable Frequency Standard for Navigation, I.R.E. Trans., **PGI-4,** pp. 116–127, Oct., 1955.

JACCARINO, V., see Dillon, J. F.

JACOBSON, M. J.,[1] CHARNES, A., and SAIBEL, E.[1]

The Complete Journal Bearing With Circumferential Oil Inlet, Trans. A.S.M.E., **77,** pp. 1179–1183, Nov., 1955.

JAMES, D. B., see Neilson, G. C.

KOHN, W.,[1] and SCHECHTER, D.[4]

Theory of Acceptor Levels in Germanium, Phys. Rev., Letter to the Editor, **99,** pp. 1903–1904, Sept. 15, 1955.

1. Bell Telephone Laboratories, Inc.
4. Carnegie Institute.

LAW, J. T.,[1] and MEIGS, P. S.[1]

The Effect of Water Vapor on Grown Germanium and Silicon n-p Junction Units, J. Appl. Phys., **26,** pp. 1265–1273, Oct., 1955.

LEWIS, H. W.[1]

Search for the Hall Effect in a Superconductor: II — Theory, Phys. Rev., **100,** pp. 641–645, Oct. 15, 1955.

LINVILL, J. G.,[1] and MATTSON, R. H.[1]

Junction Transistor Blocking Oscillators, Proc. I.R.E., **43,** pp. 1632–1639, Nov., 1955.

LOGAN, R. A.[1]

Precipitation of Copper in Germanium, Phys. Rev., **100,** pp. 615–617, Oct. 15, 1955.

LOGAN, R. A.,[1] and SCHWARTZ, M.[1]

Restoration of Resistivity and Lifetime in Heat Treated Germanium, J. Appl. Phys., **26,** pp. 1287–1289, Nov., 1955.

McCALL, D. W., see Shulman, R. G.

MASON, W. P., see Bömmel, H. E.

MATTHIAS, B. T.,[1] and CORENZWIT, E.[1]

Superconductivity of Zirconium Alloys, Phys. Rev., **100,** pp. 626–627, Oct. 15, 1955.

MATTSON, R. H., see Linvill, J. G.

MAYS, J. M., see Shulman, R. G.

MECHLINE, E. B., see Israel, J. O.

MEIGS, P. S., see Law, J. T.

MENDIZZA, A., see Compton, K. G.

MERRILL, F. F., see Israel, J. O.

MERRITT, F. R., see Fletcher, R. C., and Galt, J. K.

MERZ, W. J., see Anderson, J. R.

1. Bell Telephone Laboratories, Inc.

MOLL, J. L.[1]

Junction Transistor Electronics, Proc. I.R.E., **43,** pp. 1807–1818, Dec., 1955.

MUMFORD, W. W.,[1] and SCHAFERSMAN, R. L.[1]

Data on Temperature Dependence of X-Band Fluorescent Lamp Noise Sources, I.R.E. Trans., **PGI-4,** pp. 40–46, Oct., 1955.

NEILSON, G. C.,[6] and JAMES, D. B.[1]

Time of Flight Spectrometer for Fast Neutrons, Rev. Sci. Instr., **26,** pp. 1018–1023, Nov., 1955.

NESBITT, E. A.,[1] and WILLIAMS, H. J.[1]

New Facts Concerning the Permanent Magnet Alloy, Alnico 5, Conf. on Magnetism and Magnetic Materials, **T-78,** pp. 205–209, Oct., 1955.

NESBITT, E. A.,[1] and WILLIAMS, H. J.[1]

Shape and Crystal Anisotropy of Alnico 5, J. Appl. Phys., **26,** pp. 1217–1221, Oct., 1955.

OWNES, C. D.[1]

Stability of Molybdenum Permalloy Powder Cores, Conf. on Magnetism and Magnetic Materials, **T-78,** pp. 334–339, Oct., 1955.

PEARSON, G. L.,[1] and BRATTAIN, W. H.[1]

History of Semiconductor Research, Proc. I.R.E., **43,** pp. 1794–1806, Dec., 1955.

PEDERSON, L.[1]

Aluminum Die Castings in Carrier Telephone Systems, Modern Metals, **11,** pp. 65, 68, 70, Sept., 1955.

PRINCE, M. B.[1]

High-Frequency Silicon Aluminum Alloy Junction Diode, Trans. I.R.E., **ED-2,** pp. 8–9, Oct., 1955.

REMEIKA, J. P., see Anderson, J. R.

RIORDAN, J., see Fagen, R. E.

1. Bell Telephone Laboratories, Inc.
6. University of British Columbia, Vancouver, Canada.

SAIBEL, E., see Jacobson, M. J.

SCHAFERSMAN, R. L., see Mumford, W. W.

SCHECHTER, D., see Kohn, W.

SCHELKUNOFF, S. A.[1]

On Representation of Electromagnetic Fields in Cavities in Terms of Natural Modes of Oscillation, J. Appl. Phys., **26,** pp. 1231–1234, Oct., 1955.

SCHWARTZ, M., see Logan, R. A.

SHULMAN, R. G.,[1] MAYS, J. M.,[1] and MCCALL, D. W.[1]

Nuclear Magnetic Resonance in Semiconductors: I — Exchange Broadening in InSb and GaSb, Phys, Rev., **100,** pp. 692–699, Oct. 15, 1955.

SLOCUM, A.,[1] and AUGUSTINE, C. F.[1]

6-KMC Phase Measurement System For Traveling Wave Tube, Trans. I.R.E., **PGI-4,** pp. 145–149, Oct., 1955.

THURMOND, C. D., see Geller, S.

UHLIR, A., JR.[1]

Micromachining with Virtual Electrodes, Rev. Sci. Instr., **26,** pp. 965–968, Oct., 1955.

ULRICH, W., see Yokelson, B. J.

VAN UITERT, L. G.[1]

DC Resistivity in the Nickel and Nickel Zinc Ferrite System, J. Chem. Phys., **23,** pp. 1883–1887, Oct., 1955.

VAN UITERT, L. G.[1]

Low Magnetic Saturation Ferrites for Microwave Applications, J. Appl. Phys., **26,** pp. 1289–1290, Nov., 1955.

WANNIER, G. H.[1]

Possibility of a Zener Effect, Phys. Rev., Letter to the Editor, **100,** p. 1227, Nov., 15, 1955.

1. Bell Telephone Laboratories, Inc.

WANNIER, G. H.[1]

Threshold Law for Multiple Ionization, Phys. Rev., **100,** pp. 1180, Nov. 15, 1955.

WARNER, A. W., JR., see Bömmel, H. E.

WILLIAMS, H. J., see Nesbitt, E. A.

YAGER, W. A., see Fletcher, R. C., and Galt, J. K.

YOKELSON, B. J.,[1] and ULRICH, W.[1]

Engineering Multistage Diode Logic Circuits, Elec. Engg., **74,** p. 1079, Dec., 1955.

1. Bell Telephone Laboratories, Inc.

Recent Monographs of Bell System Technical Papers Not Published in This Journal*

ALLISON, H. W., see Moore, G. E.

BAKER, W. O., see Winslow, F. H.

BASSECHES, H., and MCLEAN, D. A.

Gassing of Liquid Dielectrics Under Electrical Stress, Monograph 2448.

BOZORTH, R. M., TILDEN, E. F., and WILLIAMS, A. J.

Anistropy and Magnetostriction of Some Ferrites, Monograph 2513.

BRADLEY, W. W., see Compton, K. G.

COMPTON, K. G., MENDIZZA, A., and BRADLEY, W. W.

Atmospheric Galvanic Couple Corrosion, Monograph 2470.

DAVIS, J. L., see Suhl, H.

FAGEN, R. E., and RIORDAN, JOHN

Queueing Systems for Single and Multiple Operation, Monograph 2506.

FINE, M. E.

Elastic Constants of Germanium Between 1.7° and 80°K, Monograph 2479.

FORSTER, J. H., see Miller, L. E.

GALT, J. K., see Yager, W. A.

GEBALLE, T. H., see Morin, F. J.

* 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.

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

Accelerated Power Aging with Lithium-Doped Point Contact Transistors, Monograph 2482.

MILLER, S. L.

Avalanche Breakdown in Germanium, Monograph 2477.

MOORE, G. E., see Wooten, L. A.

MOORE, G. E., and ALLISON, H. W.

Adsorption of Strontium and of Barium on Tungsten, Monograph 2498.

MORIN, F. J., and GEBALLE, T. H.

Electrical Conductivity and Seebeck Effect in $Ni_{0.80}$ $Fe_{2.20}O_4$, Monograph 2514.

MORRISON, J., see Wooten, L. A.

NESBITT, E. A., and WILLIAMS, H. J.

Shape and Crystal Anisotropy of Alnico 5, Monograph 2502.

OLMSTEAD, P. S.

Quality Control and Operations Research, Monograph 2530.

PEARSON, G. L., see Read, W. T., Jr.

PFANN, W. G.

Temperature Gradient Zone Melting, Monograph 2451.

POOLE, K. M.

Emission from Hollow Cathodes, Monograph 2480.

READ, W. T., JR., and PEARSON, G. L.

The Electrical Effects of Dislocations in Germanium, Monograph 2511.

RIORDAN, JOHN, see Fagen, R. E.

ROBINSON, F. N. H., see Haus, H. A.

SALOOM, J. A., see Hines, M. E.

SCHELKUNOFF, S. A.

Electromagnetic Fields in Cavities in Terms of Natural Modes of Oscillation, Monograph 2505.

SEARS, R. W.

A Regenerative Binary Storage Tube, Monograph 2527.

SLICHTER, W. P.

Proton Magnetic Resonance in Polyamides, Monograph 2490.

SUHL, H., VAN UITERT, L. G., and DAVIS, J. L.

Ferromagnetic Resonance in Magnesium-Manganese Aluminum Ferrite Between 160 and 1900 mc, Monograph 2472.

TILDEN, E. F., see Bozorth, R. M.

TREUTING, R. G.

Some Aspects of Slip in Germanium, Monograph 2485.

UHLIR, A., JR.

Micromachining with Virtual Electrodes, Monograph 2515.

VAN UITERT, L. G., see Suhl, H.

WALKER, L. R.

Power Flow in Electron Beams, Monograph 2469.

WILLIAMS, A. J., see Bozorth, R. M.

WILLIAMS, H. J., see Nesbitt, E. A.

WINSLOW, F. H., BAKER, W. O., YAGER, W. A.

Odd Electrons in Polymer Molecules, Monograph 2486.

WOOTEN, L. A., MOORE, G. E., GULDNER, W. G., and MORRISON, J.

Excess Barium in Oxide-Coated Cathodes, Monograph 2497.

YAGER, W. A., see Winslow, F. H.

YAGER, W. A., GALT, J. K., and MERRITT, F. R.

Ferromagnetic Resonance in Two Nickel-Iron Ferrites, Monograph 2478.

Contributors to This Issue

ARMAND O. ADAM,* New York Telephone Company, 1917–1920; Western Electric Company, 1920–24; Bell Telephone Laboratories; 1925–. Mr. Adam tested local dial switching systems before turning to design on the No. 1 and toll crossbar systems. From 1942 to 1945 he was associated with the Bell Laboratories School For War Training. Since then he has been concerned with the design and development of the marker for the No. 5 crossbar system. Currently he is supervising a group doing common control circuit development work for the crossbar tandem switching system.

WALLACE C. BABCOCK, A.B., Harvard University, 1919; S.B., Harvard University, 1922. U.S. Army, 1917–1919. American Telephone and Telegraph Company, 1922–1934; Bell Telephone Laboratories, 1934–. Mr. Babcock was engaged in crosstalk studies until World War II. Afterward he was concerned with radio countermeasure problems for the N.D.R.C. Since then he has been working on antenna development for mobile radio and point-to-point radio telephone systems and military projects. Member of I.R.E. and Harvard Engineering Society.

FRANKLIN H. BLECHER, B.E.E., 1949, M.E.E., 1950 and D.E.E., 1955, Brooklyn Polytechnic Institute; Polytechnic Research and Development Company, June, 1950 to July, 1952; Bell Telephone Laboratories 1952–. Dr. Blecher has been engaged in transistor network development. His principal interest has been the application of junction transistors to feedback amplifiers used in analog and digital computers. He is a member of Tau Beta Pi, Eta Kappa Nu and Sigma Xi and is an associate member of the I.R.E.

W. E. DANIELSON, B.S., 1949, M.S., 1950, Ph.D, 1952, California Institute of Technology; Bell Laboratories 1952–. Dr. Danielson has been engaged in microwave noise studies with application to traveling-wave tubes and he has been in charge of development of traveling-wave tubes

* Inadvertently, Mr. Adam's biography was omitted from the January issue of the JOURNAL in which his article, "Crossbar Tandem as a Long Distance Switching Equipment," appeared.

for use at 11,000 megacycles since June of 1954. He is the author of articles published by the Journal of Applied Physics, Proceedings of the I.R.E., and the B.S.T.J., and he is a Member of the American Physical Society, Tau Beta Pi, and Sigma Xi.

AMOS E. JOEL, JR., B.S., Massachusetts Institute of Technology, 1940; M.S., 1942; Bell Telephone Laboratories, 1940–. Mr. Joel's first assignment was in relay engineering. He then worked in the crossbar test laboratory and later conducted fundamental development studies. During World War II, he made studies of communications projects and from 1944 to 1945 designed circuits for a relay computer. Later he prepared text and taught a course in switching design. The next two years were spent designing AMA computer circuits, and since 1949 Mr. Joel has been engaged in making fundamental engineering studies and directing exploratory development of electronic switching systems. He was appointed Switching Systems Development Engineer in 1954. Member of A.I.E.E., I.R.E., Association for Computing Machinery, and Sigma Xi.

ESTHER M. RENTROP, B.S., 1926, Louisiana State Normal College. Miss Rentrop joined the transmission group of the Development and Research Department of the American Telephone and Telegraph Company in 1928, and transferred to Bell Laboratories in 1934. In both companies she has been concerned principally with control of crosstalk, both in field studies and transposition design work. During World War II, she assisted in problems of the Wire Section, Eatontown Signal Corps Laboratory at Fort Monmouth, and later she worked on other military projects at the Laboratories for the duration of the war. Miss Rentrop is presently a member of the noise and crosstalk studies group of the Outside Plant Engineering Department and is engaged in studies of interference prevention.

JACK L. ROSENFELD is a student in electrical engineering at the Massachusetts Institute of Technology. He will receive the S.M. and S.B. degrees in 1957. He has been with Bell Telephone Laboratories on cooperative assignments in microwave tube development and electronic central office during 1954 and 1955. He is a student member of the I.R.E. and a member of Tau Beta Pi and Eta Kappa Nu.

JOSEPH A. SALOOM, JR., B.S., 1948, M.S., 1949, and Ph.D., 1951, all in Electrical Engineering, University of Illinois. He joined Bell Laboratories in 1951. Mr. Saloom worked on electron tube development at

Murray Hill until 1955 with particular emphasis on electron beam studies. He is now at the Allentown, Pa., laboratory where he is engaged in the development of microwave oscillators. Member of the Institute of Radio Engineers, Sigma Xi, Eta Kappa Nu, Pi Mu Epsilon.

CHARLES S. THAELER, Moravian College, 1923–25, Lehigh University 1925–28, E.E., 1928. During the summer of 1927 he was employed by the Bell Telephone Company of Pennsylvania, returning there after graduation, where he was concerned with transmission engineering and the Toll Fundamental Plan. In 1943 he was on loan to the Operating and Engineering Department of the A.T.&T. Co., working on toll transmission studies. From 1944 to the present he has been with the Operating and Engineering Department and is currently engaged in toll circuit noise and crosstalk problems on open wire and cable systems. Mr. Thaeler is an Associate Member of A.I.E.E., and member of Phi Beta Kappa, Tau Beta Pi, and Eta Kappa Nu.

PING KING TIEN, B. S., National Central University, China, 1942; M.S., 1948, Ph.D., 1951, Stanford University; Stanford Microwave Laboratory, 1949–50; Stanford Electronics Research Laboratory, 1950–52; Bell Telephone Laboratories, 1952–. Since joining the Laboratories, Dr. Tien has been concerned with microwave tube research, particularly traveling-wave tubes. In the course of this research he has engaged in studies of space charge wave amplifiers, helix propagation, electron beam focusing, and noise. He is a member of Sigma Xi.

ARTHUR UHLIR, JR., B.S., M.S. in Ch.E., Illinois Institute of Technology, 1945, 1948; S.M. and Ph.D. in Physics, University of Chicago, 1950, 1952. Dr. Uhlir has been engaged in many phases of transistor development since joining the Laboratories in 1951, including electrochemical techniques and semiconductor device theory. Since 1952 he has participated in the Laboratories' Communications Development Training Program, giving instruction in semiconductors. Member of American Physical Society, Sigma Xi, Gamma Alpha, and the Institute of Radio Engineers.

ROGER I. WILKINSON, B.S. in E.E., 1924, Prof. E.E., 1950, Iowa State College; Northwestern Bell Telephone Company, 1920–21; American Telephone and Telegraph Company, 1924–34; Bell Telephone Laboratories, 1934–. As a member of the Development and Research Department of the A.T.&T. Co., Mr. Wilkinson specialized in the applications of the mathematical theory of probability to telephone problems.

Since transferring to Bell Telephone Laboratories in 1934, he has continued in the same field of activity and is at present Traffic Studies Engineer responsible for probability studies and traffic research. For two years during World War II, in a civilian capacity, he engaged in operations analysis studies for the Far East Air Forces in the South Pacific, for which he received the Medal for Merit. He has also served as a consultant to the Air Force, the Navy and the Air Navigation Delevopment Board. Mr. Wilkinson is a member of A.I.E.E., American Society for Engineering Education, American Statistical Association, Institute of Mathematical Statistics, Operations Research Society of America, American Society for Quality Control, Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi and Pi Mu Epislon.