ELECTRON TUBES

Volume II

(1942-1948)

ELECTRON TUBES

Volume II

(1942-1948)

Edited by

ALFRED N. GOLDSMITH

ARTHUR F. VAN DYCK

ROBERT S. BURNAP
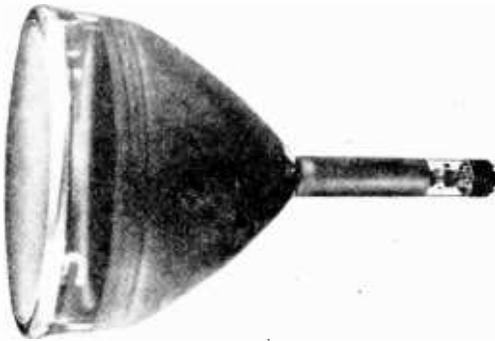
EDWARD T. DICKEY

GEORGE M. K. BAKER

MARCH, 1949

ELECTRON TUBES OF TODAY

ELECTRON TUBES

Volume II

(1942-1948)

PREFACE

ELECTRON TUBES, Volume II, is the tenth volume in the **RCA Technical Book Series** and the second on the general subject of vacuum tubes, thermionics and related subjects. This volume contains material written by RCA authors and originally published during the years 1942-1948; the companion book, ELECTRON TUBES, Volume I, covers the period 1935-1941.

The papers in this volume are presented in four sections: general; transmitting; receiving; and special. The appendices include an electron tube bibliography for the years 1942-1948 and, as an additional source of reference, a list of Application Notes. The bibliography lists all papers concerning tubes even though they relate to specific applications and are covered in other volumes of the Technical Book Series on television, facsimile, UHF, and frequency modulation. This has been done to insure that all applicable material on tubes would be available in this volume—at least in reference form.

*   *   *

RCA Review gratefully acknowledges the courtesy of the Institute of Radio Engineers *(Proc. I.R.E.)*, the American Institute of Physics *(Jour. Appl. Phys., Jour. Opt. Soc. Amer.*, and *Phys. Rev.)*, the Franklin Institute *(Jour. Frank. Inst.)*, the Society of Motion Picture Engineers *(Jour. Soc. Mot. Pic. Eng.)*, and Radio Magazines, Inc. *(Audio Eng.)* in granting to RCA Review permission to republish material by RCA authors which has appeared in their publications. The appreciation of RCA Review is also extended to all authors whose papers appear herein.

*   *   *

As outstanding as were electron tube developments from their invention until the start of the recent war, the progress in tube design and application technique during and since the war has been even more remarkable, particularly in power and miniature tubes. Still

newer work has already produced components which lend promise of replacing electron tubes for certain uses, but for the great majority of applications, electron tubes will continue to serve as the framework around which radio-electronic progress will be fashioned.

ELECTRON TUBES, Volume II, like its predecessor is being published, therefore, in the sincere hope that it will serve as a useful reference text and source of basic information to advance radio and electronics in all of its many facets.

The Manager, RCA Review

RCA Laboratories,
Princeton, New Jersey
March 19, 1949

ELECTRON TUBES

Volume II

(1942-1948)

CONTENTS

CONTENTS (Continued)

RECEIVING

SPECIAL

ANALYSIS OF RECTIFIER OPERATION*†

By

O. H. Schade

Tube Department, RCA Victor Division,
Harrison, N. J.

Summary—An analysis of rectifier operation in principal circuits is made. The introduction of linear equivalent diode resistance values permits a simplified and accurate treatment of circuits containing high-vacuum diodes and series resistance. The evaluation of these equivalent resistance values and a discussion of emission characteristics of oxide-coated cathodes precede the circuit analysis.

Generalized curve families for three principal condenser-input circuits are given to permit the rapid solution of rectifier problems in practical circuits without inaccuracies due to idealizing assumptions.

The data presented in this paper have been derived on the basis of a sinusoidal voltage source. It is apparent that the graphic analysis may be applied to circuits with nonsinusoidal voltage sources or intermittent pulse waves.

It is also permissible to consider only the wave section during conduction time and alter the remaining wave form at will. Complicated wave shapes may thus be replaced in many cases by a substantially sinusoidal voltage of higher frequency and intermittent occurrence as indicated by shape and duration of the highest voltage peak.

The applications of these principles have often explained large discrepancies from expected results as being caused by series or diode resistance and excessive peak-current demands.

Practical experience over many years has proved the correctness and accuracy of the generalized characteristics of condenser-input circuits.

INTRODUCTION

RECTIFIER circuits, especially of the condenser-input type, are extensively used in radio and television circuits to produce unidirectional current and voltages. The design of power supplies, grid-current bias circuits, peak voltmeters, detectors and many other circuits in practical equipment is often based on the assumption that rectifier- and power-source resistance are zero, this assumption resulting in serious errors. The rectifier element or diode, furthermore has certain peak-current and power ratings which should not be exceeded. These values vary considerably with the series resistance of the circuit.

General operating characteristics of practical rectifier circuits have been evaluated and used by the writer for design purposes and informa-

1

tion since early 1934, but circumstances have delayed publication. Several papers[1-4] have appeared in the meantime treating one or another part of the subject on the assumption of zero series resistance. Practical circuits have resistance and may even require insertion of additional resistance to protect the diode and input condenser against destructive currents. The equivalent diode resistance and the emission from oxide-coated cathodes are, therefore, discussed preceding the general circuit analysis. This analysis is illustrated on graphic constructions establishing a direct link with oscillograph observations on practical circuits. A detailed mathematical discussion requires much space and is dispensed with in favor of graphic solutions, supplemented by generalized operating characteristics.

I. PRINCIPLES OF RECTIFICATION

General

Rectificaiton is a process of synchronized switching. The basic rectifier circuit consists of one synchronized switch in series with a single-phase source of single frequency and a resistance load. The switch connection between load terminals and source is closed when source and load terminals have the same polarity, and is open during the time of opposite polarity. The load current consists of half-wave pulses. This simple circuit is unsuitable for most practical purposes, because it does not furnish a smooth load current.

The current may be smoothed by two methods: (a) by increasing the number of phases, and (b) by inserting reactive elements into the circuit. The phase number is limited to two for radio receivers. The circuit analysis which follows later on will treat single- and double-phase rectifier circuits with reactive circuit elements.

Switching in reactive circuits gives rise to "transients." Current and voltage cannot, therefore, be computed according to steady-state methods.

The diode functions as a self-timing electronic switch. It closes the circuit when the plate becomes positive with respect to the cathode

---

[1] M. B. Stout, "Analysis of rectifier filter circuits," *Elec. Eng. Trans. A.I.E.E.* (*Elec. Eng.*, September, 1935), vol. 54, pp. 977-984; September, 1935.

[2] N. H. Roberts, "The diode as half-wave, full-wave and voltage-doubling rectifier," *Wireless Eng.*, vol. 13, pp. 351-362; July, 1936; and pp. 423-470; August, 1936.

[3] J. C. Frommer, "The determination of operating data and allowable ratings of vacuum-tube rectifiers," *Proc. I.R.E.*, vol. 29, pp. 481-485; September, 1941.

[4] D. L. Waidelich, "The full-wave voltage-doubling rectifier circuit," *Proc. I. R. E.*, vol. 29, pp. 554-558; October, 1941.

and opens the circuit at the instant when the plate current becomes zero.

The diode has an internal resistance which is a function of current. When analyzing rectifier circuits, it is convenient to treat the internal resistance of the diode rectifier as an element, separated from the "switch action" of the diode. Fig. 1 illustrates the three circuit elements so obtained and their respective voltage-current characteristics (see Section II). The diode characteristic is the sum of these characteristics. The resistance r_d is effective only when the switch is closed, i.e., during the conduction period of the diode. The effective diode resistance must, therefore, be measured or evaluated within conduction-time limits. Consider a switch in series with a fixed resistance and any number of other circuit elements connected to a battery of fixed voltage. The direct current and root-mean-square current
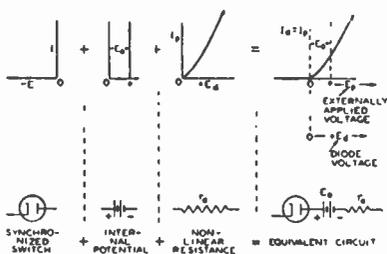
Fig. 1—Characteristics and equivalent circuit for high-vacuum diodes.
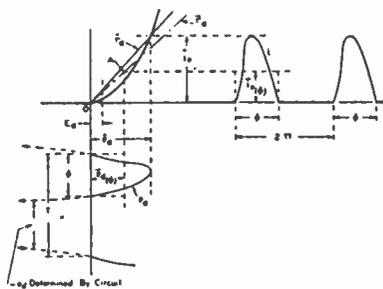
Fig. 2 — Graphic evaluation of equivalent diode resistance values.

which flow in this circuit will depend on the time intervals during which the switch is closed and open; the resistance value is not obtainable from these current values and the battery voltage. The correct value is obtained only when the current and voltage drop in the resistance are measured during the time angle ϕ (Fig. 2) when the switch is closed.

The method of analysis of rectifier circuits to be discussed in this paper is based on the principle that the nonlinear effective resistance of the diode may be replaced analytically by an equivalent fixed resistance which will give a diode current equal to that obtained with the actual nonlinear diode resistance. The correct value to be used for the equivalent fixed resistance depends upon whether we are analyzing for peak diode current, average diode current, or root-mean-square diode current.

At the outset of an analysis amplitude and wave shape of the diode current are not known and the diode resistance must, therefore, be determined by successive approximations.
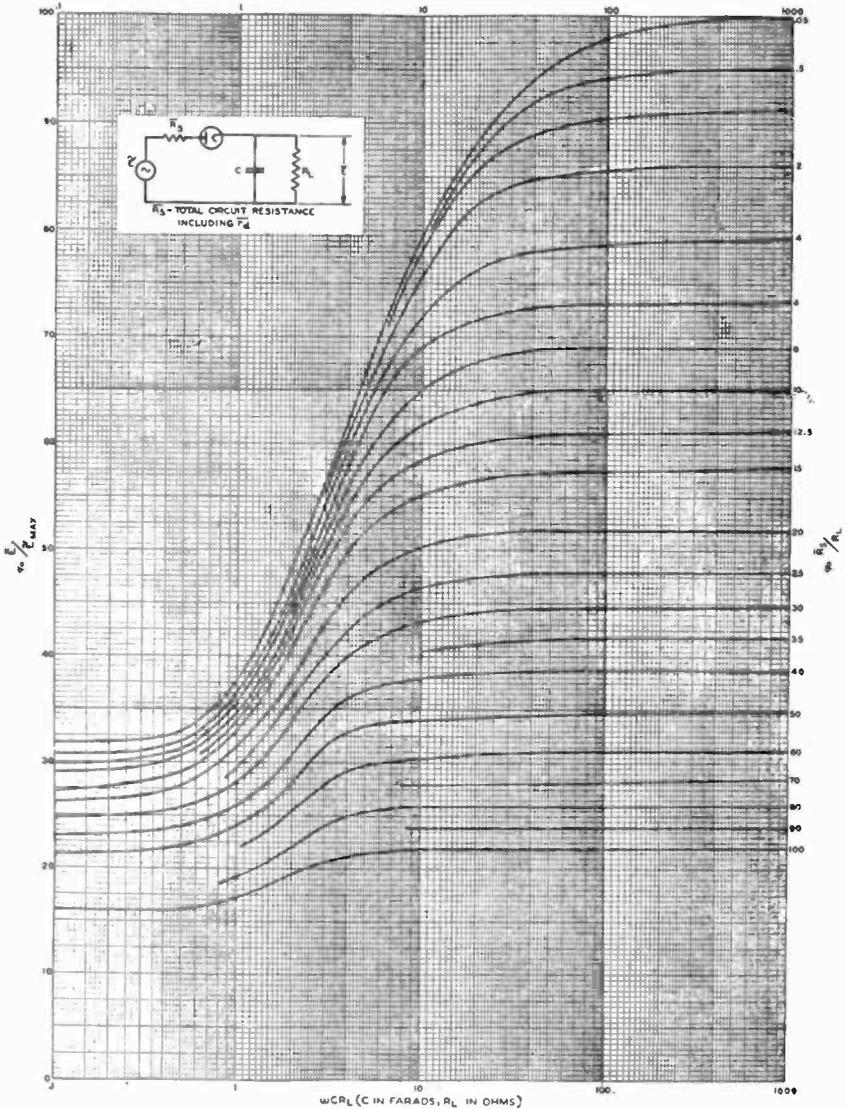
Fig. 3—Relation of applied alternating peak voltage to direct output voltage in half-wave, condenser-input circuits.

The complexity of repeated calculations, especially on condenser-input circuits, requires that the operating characteristics of the circuit be plotted generally as functions of the circuit constants including series resistance in the diode circuit as a parameter.

Data for these plots (such as Figs. 3 to 7) are to be obtained by general analysis of circuits with linear resistances.

The solution of a practical condenser-input-circuit problem requires the use of three different equivalent linear circuits and diode resistance values.

The resistance values are obtainable from the peak current alone because wave shape can be eliminated as a factor by means of a general relation given by (6). The practical analysis of condenser input circuits thus simplified, is carried out as follows:
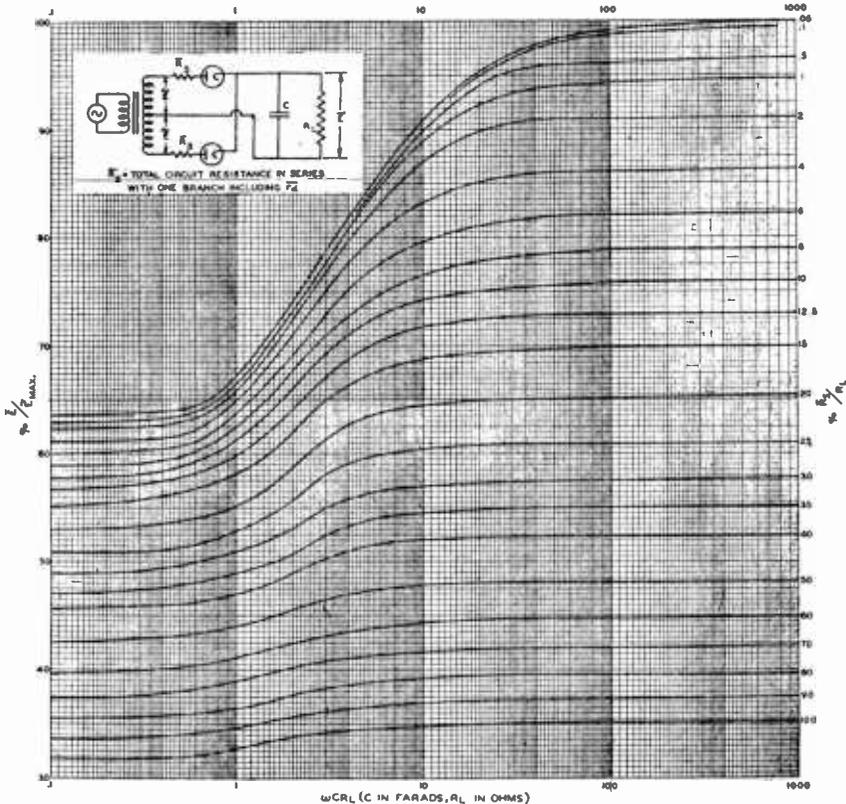
Fig. 4—Relation of applied alternating peak voltage to direct output voltage in full-wave, condenser-input circuits.

The average diode current is estimated roughly and the diode peak current is assumed to be four times the average value. The diode characteristic (Fig. 8) furnishes an initial peak-resistance value and (6) furnishes the other diode resistance values (see $\hat{R}_s$ values in Fig. 9). Direct output voltage and average current are now obtained with the equivalent average value $\overline{R}_s$ from the respective plot (Figs. 3 to 5) as a first approximation. Another chart (Fig. 6) furnishes the peak-
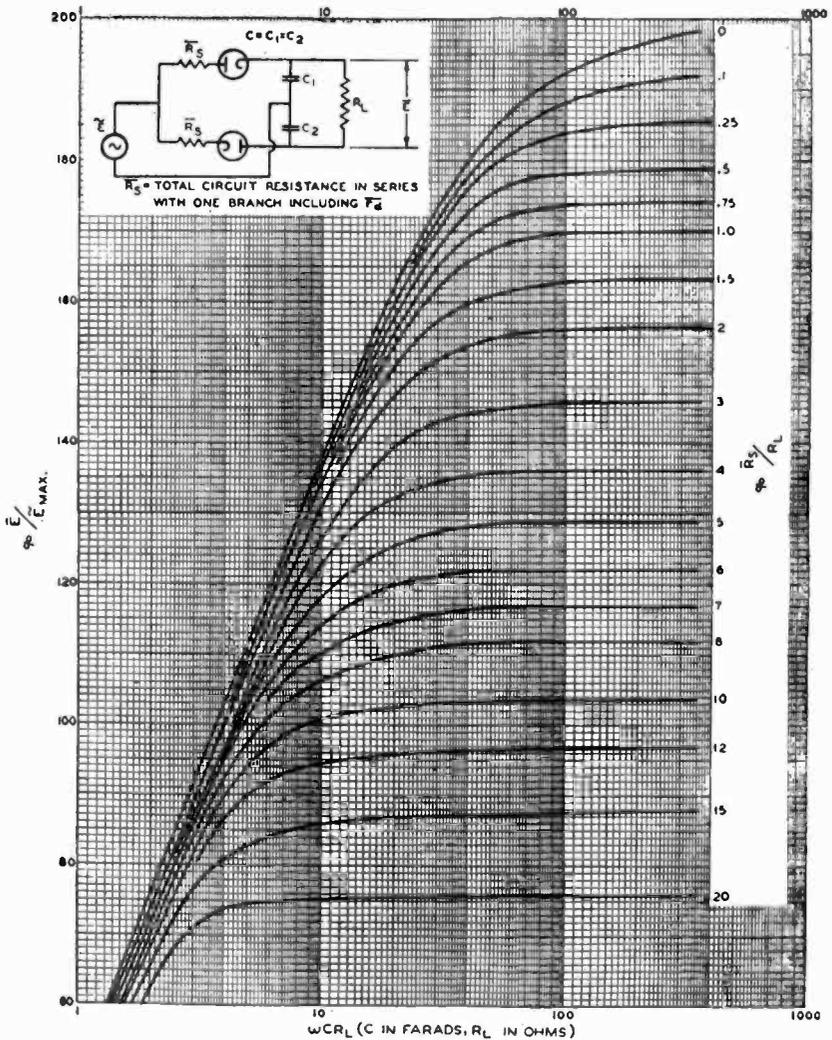
Fig. 5—Relation of applied alternating peak voltage to direct output voltage in condenser-input, voltage-doubling circuits.

to-average-diode-current ratio with the peak value $\hat{R}_s$ and thus the peak current and diode peak resistance in close approximation.

A second approximation gives usually good agreement between initial and obtained resistance values, which are then used to obtain other operating data.

A theoretical treatment of the method just described will be omitted in favor of an analysis of operating characteristics of the

rectifier tube itself. The user of tubes may welcome information on the subject of peak emission which is of vital importance in the rating and trouble-free operation of any tube with an oxide-coated cathode.

II. ANODE AND CATHODE CHARACTERISTICS OF RECTIFIER TUBES

Anode Characteristics

1. *Definitions of Resistance Values*

The *instantaneous resistance* (r_d) of a diode is the ratio of the instantaneous plate voltage e_d to the instantaneous plate current i_p
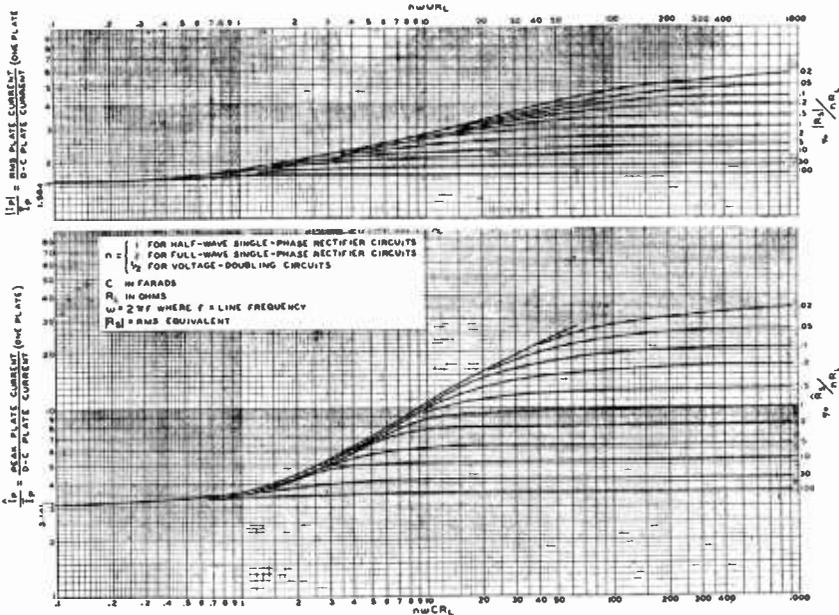
Fig. 6—Relation of peak, average, and root-mean-square diode current in condenser-input circuits.

at any point on the characteristic measured from the operating point (see Fig. 1). It is expressed by

$$r_d = \frac{e_d}{i_p}. \tag{1}$$

The *operating point* (0) of a diode is a fixed point on the characteristic, marked by beginning and end of the conduction time. It is, therefore, the cutoff point $I_d = 0$ and $E_d = 0$, as shown in Fig. 1. The operating point is independent of the wave form and of the conduction time ϕ (see Fig. 2).

The peak resistance[5] ($\hat{r}_d$) is a specific value of the instantaneous resistance and is defined as

$$\hat{r}_d = \frac{\hat{e}_d}{\hat{i}_p} \text{ (see Fig. 2)}. \tag{2}$$

Peak voltage $\hat{e}_d$ and peak current $\hat{i}p$ are measured from the operating point 0.

CIRCUIT	PARAMETER
	% $|R_s|/R_L$
HALF-WAVE	0.1
	1.0
	10
	30
VOLTAGE-DOUBLER	0.1
	10
FULL-WAVE	0.1
	1.0
	10
	30
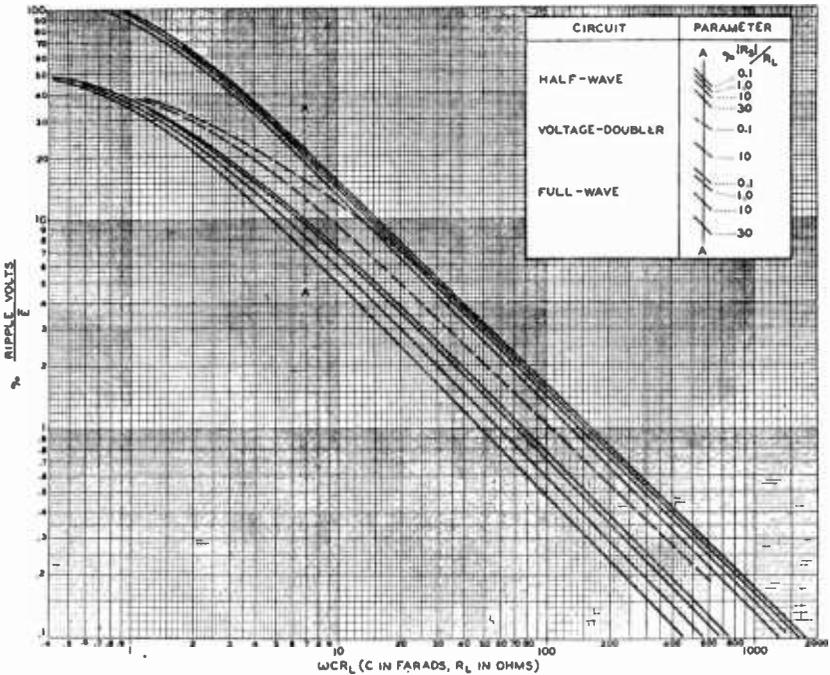
Fig. 7—Root-mean-square ripple voltage of condenser-input circuits.

The equivalent average resistance ($\bar{r}_d$) is defined on the basis of circuit performance as a resistance value determining the magnitude of the average current in the circuit. The value $\bar{r}_d$ is, therefore, the ratio of the average voltage drop $\bar{e}_{d(\phi)}$ in the diode during conduction time to the average current $\bar{i}_{p(\phi)}$ during conduction time, or

$$\bar{r}_d = \frac{\bar{e}_{d(\phi)}}{\bar{i}_{p(\phi)}}. \tag{3}$$

_____

[5] For system of symbols, see Appendix.

The curved diode characteristic is thus replaced by an equivalent linear characteristic having the slope $\bar{r}_d$ and intersecting the average point A, as shown in Fig. 2. The co-ordinates $\bar{e}_{d(\phi)}$ and $\bar{i}_{p(\phi)}$ of the average point depend on the shape of voltage and current within the time angle ϕ. The analysis of rectifier circuits shows that the shape of the current pulse in actual circuits varies considerably between different circuit types.
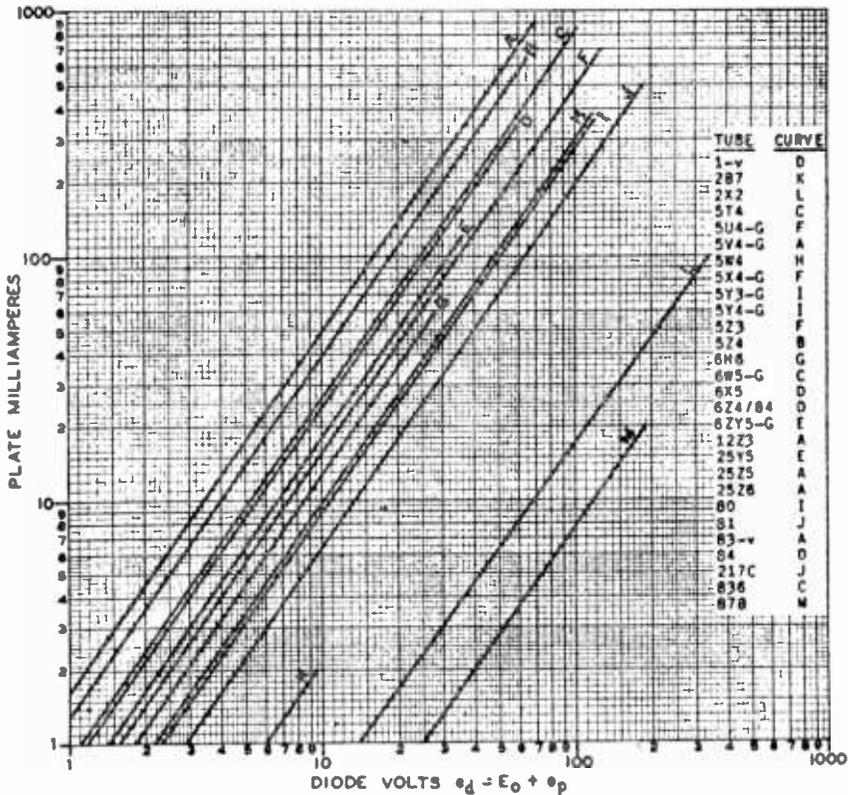
Fig. 8—Average anode characteristics of some RCA rectifier tubes.

The *equivalent root-mean-square resistance* ($|r_d|$) is defined as the resistance in which the power loss P_d is equal to the plate dissipation of the diode when the same value of root-mean-square current $|I_d|$ flows in the resistance as in the diode circuit. It is expressed by

$$| r_d | = \frac{P_d}{| I_d |^2}. \tag{4}$$

HALF-WAVE CIRCUIT

FULL-WAVE CIRCUIT

VOLTAGE-DOUBLING CIRCUIT
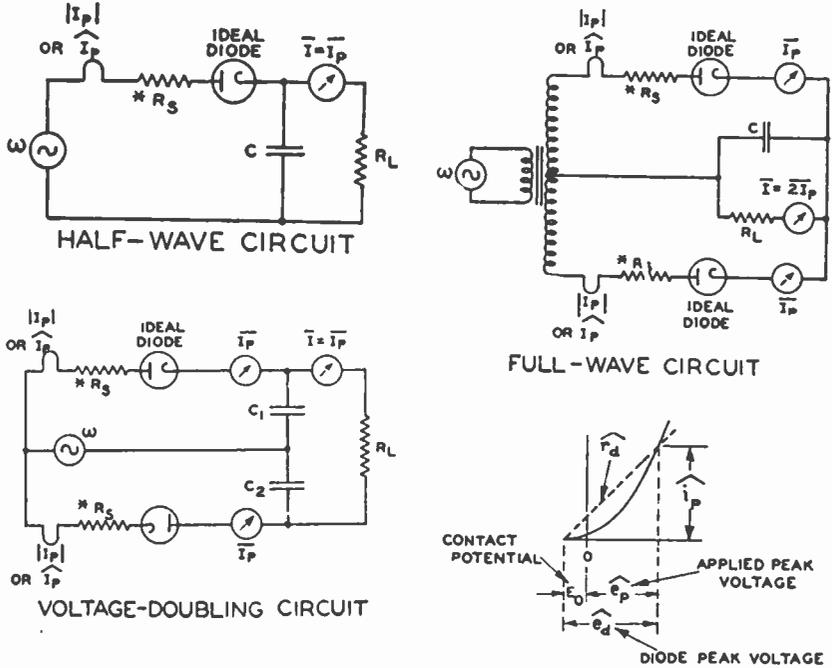
$$\widehat{r_d} = .88\,\overline{r_d} = .935\,|r_d| = \frac{\widehat{e_d}}{\widehat{i_p}}$$

$$\widehat{R_S} = R_S + \widehat{r_d}$$

$$\overline{R_S} = R_S + \overline{r_d}$$

$$|R_S| = R_S + |r_d|$$

* USE $\widehat{R_S}, \overline{R_S},$ OR $|R_S|$ AS REQUIRED

R_S = EXTERNAL RESISTANCE

$\widehat{r_d}$ = PEAK DIODE RESISTANCE

$\overline{r_d}$ = EQUIVALENT AVE. DIODE RESISTANCE

$|r_d|$ = EQUIVALENT R M S DIODE RESISTANCE

Fig. 9—Equivalent circuits and resistance values for condenser-input rectifier circuits.

2. Measurement of Equivalent Diode Resistances

The equivalent resistance values of diodes can be measured by direct substitution under actual operating conditions. The circuit arrangement is shown in Fig. 10. Because the diode under test must be replaced as a whole by an adjustable resistance of known value, a second switch (a mercury-vapor diode identified in the figure as the ideal diode) with negligible resistance must be inserted in order to preserve the switch-action in the circuit.

When a measurement is being made, the resistor R_d is varied until the particular voltage or current under observation remains unchanged for both positions of the switch S. We observe (1) that it is impossible to find one single value of R_d which will duplicate conditions of the

actual tube circuit, i.e., give the same values of peak, average, and root-mean-square current in the circuit; (2) that the ratio of these three "equivalent" resistance values of the diode varies for different combinations of circuit elements; and (3) that the resistance values are functions of the current amplitude and wave shape.

3. *Wave Forms and Equivalent Resistance Ratios for Practical Circuit Calculations*

The form of the current pulse in practical rectifier circuits is determined by the power factor of the load circuit and the phase number. Practical circuits may be divided into two main groups: (a) circuits with choke-input filter; and (b) circuits with condenser-input filter.

The diode current pulse in choke-input circuits has a rectangular form on which is superimposed one cycle of the lowest ripple frequency. In most practical circuits, this fluctuation is small as compared with the average amplitude of the wave and may be neglected when determining the equivalent diode resistances. It is apparent then that the equivalent diode resistance values are all equal and independent of the type of diode characteristics for square-wave forms. Hence, for choke-input circuits, we have

Fig. 10 — Circuit for measuring equivalent diode resistance values.

$$\hat{r}_d = \bar{r}_d = |\, r_d\, |. \tag{5}$$

The diode current pulse in condenser-input circuits is the summation of a sine-wave section and a current having an exponential decay. It varies from a triangular form for $\phi < 20$ degrees to a full half cycle ($\phi = 180$ degrees) as the other extreme. In Table I are given the ratios of voltages, currents, and resistance values during conduction time for two principal types of rectifier characteristics: the 3/2-power-law characteristic of high-vacuum diodes, and the idealized rectangular characteristic of hot-cathode, mercury-vapor diodes. In this table, the designation $|i_p|_{(\phi)}$ represents the root-mean-square value of the current during the conduction time.

It follows that the relation

$$\hat{r}_d = 0.88\, \bar{r}_d = 0.93\, |\, r_d\, | \tag{6}$$

Table I

Conduction Time Angle ∅	Wave Shape	$\dfrac{\bar{i}_{p(\phi)}}{\hat{i}_p}$	$\dfrac{\|\bar{i}_p\|_{(\phi)}}{\hat{i}_p}$	3/2-Power Rectifier Characteristic			Rectangular Characteristic		
				$\dfrac{\bar{e}_{d(\phi)}}{\hat{e}_d}$	$\dfrac{\bar{r}_d}{\hat{r}_d}$	$\dfrac{\|r_d\|}{\hat{r}_d}$	$\dfrac{\bar{e}_{d(\phi)}}{\hat{e}_d}$	$\dfrac{\bar{r}_d}{\hat{r}_d}$	$\dfrac{\|r_d\|}{\hat{r}_d}$
Condenser-Input Circuits									
Degrees ≤ 20		0.500	0.577	0.593	1.185	1.120	1.0	2.00	1.500
90 and 180		0.637	0.707	0.715	1.120	1.057	1.0	1.57	1.272
130		0.725	0.780	0.787	1.085	1.030	1.0	1.38	1.190
Choke-Input Circuits									
180		1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

is representative for the group of condenser-input circuits containing high-vacuum diodes, and holds within ± 5 per cent over the entire range of variation in wave shape. The actual error in circuit calculations is smaller as the diode resistance is only part of the total series resistance in the circuit.

CATHODE CHARACTERISTICS

Peak-Emission and Saturation of Oxide-Coated Cathodes

The normal operating range of diodes (including instantaneous peak values) is below the saturation potential because the plate dissipation rises rapidly to dangerous values if this potential is exceeded. Saturation is definitely recognized in diodes with tungsten or thoriated-tungsten cathodes as it does not depend on the time of measurement, provided the plate dissipation is not excessive. The characteristics of such diodes are single-valued even in the saturated range, i.e., the range in which the same value of current is obtained at a given voltage whether the voltage has been increased or decreased to the particular value.

Diodes with oxide-coated cathodes may have double-valued characteristics because of the coating characteristic. The cathode coating has resistance and capacitance, both of which are a function of temperature, current, and the degree of "activation."

A highly emitting monatomic layer of barium on oxygen is formed on the surface of the coating, which, when heated, supplies the electron cloud forming the space charge above the coating surface (see Fig. 11). The emission from this surface may have values as high as 100 amperes per square centimeter. The flow of such enormous currents is, however, dependent on the internal-coating impedance, and is possible only under certain conditions. Special apparatus is required to permit observation of high current values which, to prevent harm to the tube, can be maintained only over very short time intervals determined by the thermal capacity of the plate and coating. For example, an instantaneous power of 15 kilowatts must be dissipated in the close-spaced diode type 83-v at a current of 25 amperes from its cathode surface of only 1 square centimeter.

Equipment for such observations was built in June, 1937, by the author after data obtained in 1935 on a low-powered curve tracer[6] indicated the need for equipment having a power source of very low internal impedance for measurements on even relatively small diodes.
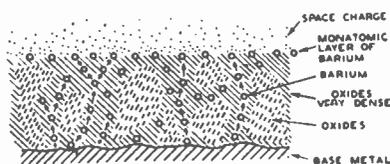
Fig. 11—Representation of cathode coating.
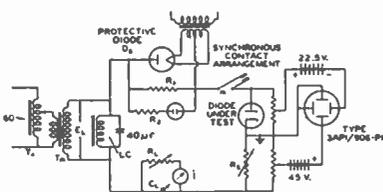
Fig. 12—Peak emission test circuit.

1. *Measurement of Diode Characteristics and Peak Emission*

The circuit principle is shown in Fig. 12. The secondary voltage of a 2-kilovolt-ampere transformer T_m is adjustable from zero to 2 kilovolts by means of an autotransformer T_A. Transformer and line reactances are eliminated for short-time surge currents by a large condenser load ($C = 20$ to 80 microfarads). The large reactive current is "tuned out" by a choke L of considerable size. The voltage is applied through a large mercury diode and a synchronous contact arrangement m to the tube under test in series with a resistance box R_s and a condenser input load C_L and R_L. This load permits adjustment of the peak-to-average current ratio. Variation of R_L changes the average current. Variation of C_L and phasing of the synchronous contact m with respect to the 60-cycle line voltage permit regulation, within wide limits, of the rate of change and duration of the current pulses.

The dynamic voltage-current characteristic of the tube under test

<hr />

[6] Demonstrated, Rochester Fall Meeting, Rochester, N. Y., November 18, 1935.

is observed on a cathode-ray oscillograph connected in the conventional manner. Calibration deflections are inserted (not shown) by other synchronous contacts to provide accurate and simultaneously visible substitution co-ordinates which may be moved to any point in the characteristic.

The motor-driven synchronous contactor closes the circuit at a desired instant of the line-voltage cycle. The circuit may then be maintained closed for approximately 30 cycles to allow decay of the starting transient (see Fig. 13). It is then opened for approximately
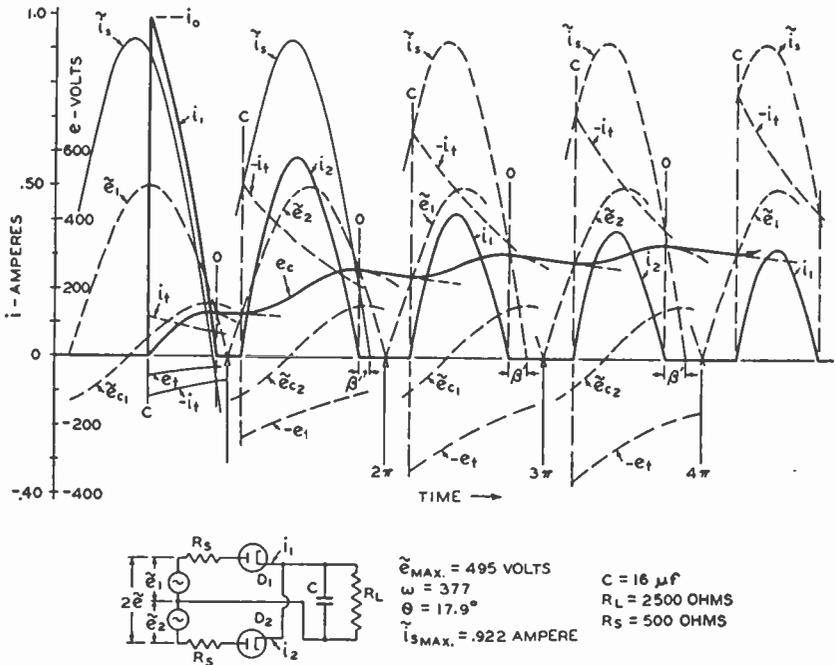
Fig. 13—Starting conditions in a full-wave, condenser-input circuit with large series resistance.

70 cycles to allow time for the discharge of condenser C_L. This cycle repeats continuously. The diode D_S in series with the tube under test protects it against damage in case it breaks down or arcs, because the diode takes up the inverse voltage if a given small reversed current determined by R_1 is exceeded. This condition is indicated by a small glow tube in shunt with D_S.

2. Coating Characteristics

A theory of electron movement and conditions in oxide coatings has been formulated after careful analysis of saturation characteristics

observed on the curve tracer. As saturated coatings produce closed reactive loops in the characteristic, it is found necessary to assume the existence of a capacitance in the diode itself. Because of its large value (see Fig. 14(c)), this capacitance requires a dielectric thickness approaching crystal spacing and, hence, must be located inside the coating. It is beyond the scope of this paper to report the many investigations which led to this particular conception.

The oxide coating is an insulator at room temperature. At increased temperatures, it becomes conductive (normal operating temperatures are between 1000 and 1100 degrees Kelvin). Electronic conduction may be thought of as occurring by relay movement of electrons under the influence of electrostatic potentials in the coating, which is a layer containing insulating oxide crystals (shaded areas in Fig. 11) interposed with metal atoms and ions (circles). These have been produced during the activation and aging processes by high ctahode temperature and electrolysis. The required potential gradients can be produced by rather small potentials because of the minute distances in the structure; the potential drop throughout the coating, therefore, is low under normal conditions.

Fig. 14—Double-valued characteristics of actual and artificial diodes showing coating saturation.

The conduction is high, when a sufficient number of relay paths not broken by oxides have been formed and when electron movement is facilitated by the loosening of the atomic structure which takes place at increased temperatures.

The coating is not necessarily a homogeneous conductor as it may consist of many sections operating in parallel but having different conductance values with individual temperature parameters. At increased plate potentials, poorly conducting sections tend to saturate, the section potential becoming more positive towards the surface. Negative-grid action of neighboring sections with higher conductivity may tend to limit emission from the surface over the poor section but the increased positive gradient towards the saturating section causes it to

draw electrons from the surrounding coating towards its surface. Further increase in current demand may then saturate the better conducting paths and may even fuse them, thus forcing current through poorer sections. Forced electron flow results in local power dissipation and temperature increases and may cause ionization and electrolysis accompanied by liberation of gas (oxygen) and formation of barium metal; i.e., it causes an accelerated activation process.

These conditions in the diode coating, therefore, should furnish a voltage-current characteristic of purely ohmic character as long as activation-gas liberation is substantially absent. Characteristics of this type are single valued. Single-valued characteristics i n d i c a t e, however, unsaturated ohmic coating conductance and limiting surface
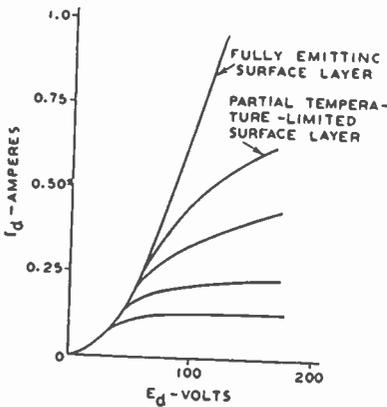
Fig. 15—Single-valued diode characteristics.
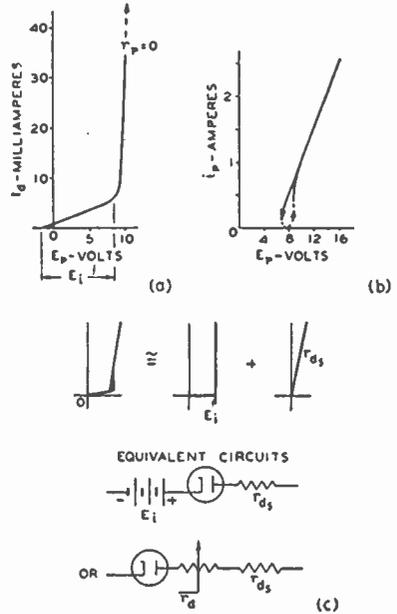
Fig. 16—Characteristics and equivalent circuit for hot-cathode, mercury-vapor diodes.

emission when moderate-current densities are involved as will be apparent from the following discussion. As cathode and coating temperatures are relatively slowly varying parameters, characteristics such as shown in Fig. 15 are observed on the cathode-ray curve tracer. The characteristic of diodes containing larger amounts of gas exhibits a discontinuity or "gas loop" (compare Fig. 16(b)) which is recognized by the fact that corresponding current values after ionization require less diode potential than before "breakdown." The characteristic, hence, is steeper than normal.

3. Transient Emission

Let us now consider the action of insulating oxides in the coating. They block many possible electron paths to sections of the surface layer which, therefore, cannot emit steady electron currents. However, electrons can be moved to the oxide surfaces and a displacement current can flow in these coating sections allowing transient-emission currents to be drawn from the corresponding surface sections.

The displacement current in the coating and the corresponding transient surface emission represent a certain fraction of the total diode current, which may permit a total emission current of short duration much in excess of the possible steady-state conduction current. The "transient-emission" current depends on the effective capacitance value of the blocking oxides, their series and shunt conductance in the coating, the emission and area of corresponding surface elements
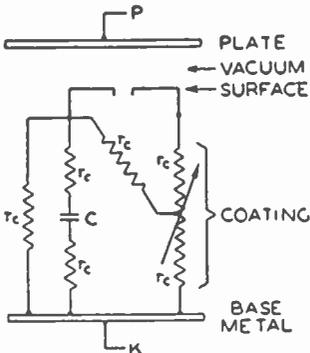
Fig. 17 — Circuit network representing the coating impedance in high-vacuum diodes.
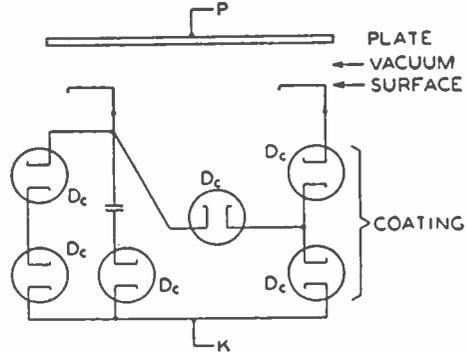
Fig. 18—Same as Fig. 17 with resistances replaced by special diodes.

to the plate as well as on the external plate-circuit impedance, and the wave form of the applied plate voltage.

For the purpose of analysis, therefore, we may draw representative networks such as shown in Figs. 17 or 18 and show the temperature-controlled coating conductances r_c as a network of "close-spaced diodes" which may conduct in two directions, each one having a single-valued characteristic which may be unsaturated or saturated depending on the assumed conditions in the coating; the conductance values of these "diodes" depend on the number of parallel or series paths they represent.

The diode contains, therefore, in its coating, a type of condenser-input load circuit, which is analyzed later on in this paper; its action explains double-valued voltage-current characteristics obtainable from the diode alone.

Consider a high plate voltage suddenly applied by means of a switch to a diode as in the circuit of Fig. 19. If the coating is not limiting, the current obtained is that at a point P on the corresponding diode characteristic. Hence, the current wave form in the circuit is as shown in Fig. 19(a). If the surface emission is assumed to be unchanged, but the coating conductance is limited, due to an insufficient number of "coating diodes" and too many nonconducting oxide groups, the wave form of Fig. 19(b) is obtained. At the instant when the switch is closed the current value $\hat{\imath}$ is demanded by E_d from the surface layer; the conduction current in the coating is limited to the value I_c by saturation of the "coating diodes." Because of the oxide capacitance, a displacement current can flow and charge up the oxides, but their charge may be limited by hypothetical series diodes.

The coating resistance is extremely low[7] below saturation, but becomes infinite when the conduction current is saturated; the charging current must then flow in the plate circuit (external) of the diode.
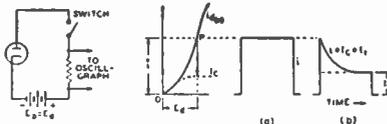
Fig. 19—Circuit for observation of peak emission transients.

The total plate current is, therefore, the sum of the conduction current I_c and a "transient-emission" current. The "coating transient" decays to zero the same as normal transients at a rate depending on the actual shunt-conductance value and the total series resistance in the circuit (Fig. 19(b)). The decay can be changed by adding external resistance in the plate circuit. When the surface emission is good, i.e., as long as the total vacuum-space plate current is space-charge-limited, the current will rise initially to the value (point P) determined by the applied potential, but will then decay to the saturation value determined by the coating conductance.

The condition of oxide-coated cathodes can, therefore, not be judged alone by their capability of furnishing high peak currents, but the time of current flow and the current wave form must also be carefully considered, because the diode characteristic may not be single-valued. Fig. 14 shows characteristics which are not single-valued. It should be noted that the characteristic loops are formed in the opposite sense as gas loops. Their extent depends on the time interval involved and the current value exceeding the unsaturated conductance current. An artificially produced characteristic of this type is also shown in Fig. 14(c). The loop size can be varied by adjusting the cathode temperature of the shunting diode. Both diodes had single-valued characteristics.

---

[7] Its magnitude depends on the number of series diodes and, hence on the barium content and thickness of the coating.

4. *Current Overload and Sputter*

The degree of activation is not stable during the life of the cathode. Coating conductance and surface emission change. Factors affecting the change are the coating substances, the evaporation rate of barium which depends on the base material, and the operating conditions to which the cathode is subjected. This life history of the cathode is the basis on which current ratings are established. Rectifier tubes especially are subject to severe operating conditions. If a diode is operated with too high a current in a rectifier circuit and its surface emission is decreased to the saturation value, then the tube-voltage drop will increase rapidly and cause excessive plate dissipation and destruction of the tube. Should the coating conductance in this diode decrease to a value which limits the demanded current, power is dissipated in the now-saturated coating with the result that the coating-voltage drop and coating temperature are raised. The votage and temperature rise in the coating may cause reactivation but also may become cumulative and melt the coating material. We may consider that good conducting paths are fused or that a dielectric breakdown of oxide capacitance occurs; in any event vapor or gas discharges result from saturated coatings. In most cases breakdown occurs during one of the following inverse voltage cycles as observed on the curve tracer. A saturation loop is first formed as shown in Fig. 14 and a certain time must be allowed for diffusion of the gas into the vacuum space. Fusion of coating material may also occur during the conduction period. These breakdowns are known as "sputter," and in usual circuits destroy the cathode.

A second type of sputter is caused by the intense electrostatic field to which projecting "high spots" on the plate or cathode are subjected. The resulting current concentration causes these spots to vaporize with the result than an arc may be started. Hundreds of scintillating small spots can be observed at first at very high applied surge potentials, but may be cleared after a relatively short time.

Transient peak currents of 25 amperes per square centimeter have been observed from well-activated oxide-coated cathodes. The stable peak emission over an extended period is usually less than one-third of this value.

5. *Hot-Cathode Mercury-Vapor Diodes*

The breakdown voltage E_i of mercury vapor for cumulative ionization is a function of the gas pressure and temperature. It is approximately 10 volts in the RCA-83 and similar tubes. A small electron current begins to flow at $E_p = 0$ (see Fig. 16), and causes ionization

of the mercury vapor. This action decreases the variational diode resistance r_p to a very low value. The ionization becomes cumulative at a certain current value ($r_p = 0$ at 40 milliamperes in Fig. 16(a)), and causes a discontinuity in the characteristic. Hence, it is not single-valued within a certain voltage range. Beyond this range (see Fig. 16(b)), the slope (r_p) of the characteristic becomes again positive until saturation of the emitter is reached.

For circuit analysis, the mercury-vapor diode may be replaced by a bucking battery having the voltage E_i and a fixed resistance as shown in Fig. 16(c); or the diode characteristic may be replaced by an ideal rectangular characteristic and its equivalent resistance values and the series resistance r_{ds} as shown.

The first representation is adequate for most practical calculations. The value r_{ds} is in the order of 4 ohms for small rectifier tubes. The low series resistance and the small constant-voltage drop E_i are distinct advantages for choke-input filters, as they cause very good regulation; the low resistance, however, will give rise to enormously high starting transients in condenser-input circuits, in case all other series resistances are also small. The destruction of the coating in mercury-vapor diodes is caused by concentration of current to small sections of the coating surface and not by heat dissipation in the coating. Mercury-vapor diodes as well as high-perveance (close-spaced), high-vacuum diodes having oxide cathodes should, therefore, be protected against transient-current overloads when they are started in low-resistance circuits to prevent destruction of the cathode coating.

6. Protective Resistance Values

Very high instantaneous peak currents may occur in noninductive condenser-input circuits when the circuit is opened long enough to discharge the condenser, but reclosed before the cathode temperature of the diode has decreased substantially. The maximum peak current i_{max} occurs when closing the circuit at peak line voltage. At the instant of switching, C is a short circuit and the current i_{max} is limited only by the series resistance (including diode) of the circuit,

$$i_{max} = \frac{\bar{e}_{max}}{\hat{R}_s} \, .$$

For a given maximum diode current i_{dmax} and the corresponding diode peak voltage $\hat{E}_{dmax}$, the minimum effective series resistance R_s in the circuit must hence be

$$R_s = \frac{\bar{e}_{max} - \hat{E}_{dmax}}{\hat{I}_{dmax}}.$$

This limiting resistance must be inserted in series with low-impedance sources (power line in transformerless sets). Commercial power transformers for radio receivers have often sufficient resistance besides some leakage reactance to limit starting currents to safe values.

III. CIRCUIT ANALYSIS

General

The rectifier diode is a switch operated in synchronism with the applied alternating-current frequency. Switching in reactive circuits causes transients. The total current in the circuit may be regarded as the sum of all steady-state currents and transient currents within the time between two switching operations. Steady-state voltages (e_s) and currents (i_s) in the particular circuit before and after switching are determined without difficulty. It is very helpful to draw them approximately to scale and with proper phase relation.

The switching time of the diode is then located on the graph. Currents change at switching time t_0 from i_1 to $i_2 = i_{s(2)} + i_t$ and voltages from e_1 to $e_2 = e_{s(2)} + e_t$. The transients i_t or e_t are zero, when the current change does not occur in an inductive circuit or when a voltage change is not required on a capacitance at the time of switching. A sudden change Δi_L or Δe_c demanded at t_0 causes transients. They initially cancel the change Δi_L or Δe_c because an inductance offers infinite impedance to an instantaneous change in total current and a capacitance offers zero impedance to an instantaneous voltage change.

The initial transient values are, therefore,

$$i_{t(0)} \text{ in } L = -\Delta i_L$$

and

$$e_{t(0)} \text{ on } C = -\Delta e_c.$$

The transients decay exponentially from their initial value.

According to the decay time of the transients, fundamental rectifier circuits may be classified into two principal groups: (1) circuits with repeating transients in which the energy stored in reactive elements decreases to zero between conduction periods of the diode; and (2) circuits with chain transients in which (a) the magnetic energy stored in the inductance of the circuit remains above zero value, and (b) the electric energy stored in the capacitance of the circuit remains

and similar to (7) except for an increase of the transient term due to the battery current $I_B = \bar{E}_B/R_s$.

Equation (7b) is valid only over a range of load or battery voltage $(\bar{E}_B)$ in which switching time and conduction period of the diodes are constant $(\phi = \pi)$. This range is shown by the solid part of curve F in Fig. 22 and ends at a particular current and voltage of the circuit characteristic marked the "critical point."

The critical point is the operating condition at which the instantaneous current i in the common branch circuit has zero value at one instant. An analysis shows that in the range $\bar{E}_B = \tilde{e}_{max}$ to $\bar{E}_B = E_B'$ each diode circuit operates independently as a half-wave rectifier circuit (battery-charger operation, curve H in Fig. 22). Current commutation begins at E_B'; the diode circuits begin to interact, but the conduction angle is still $\phi < \pi$.
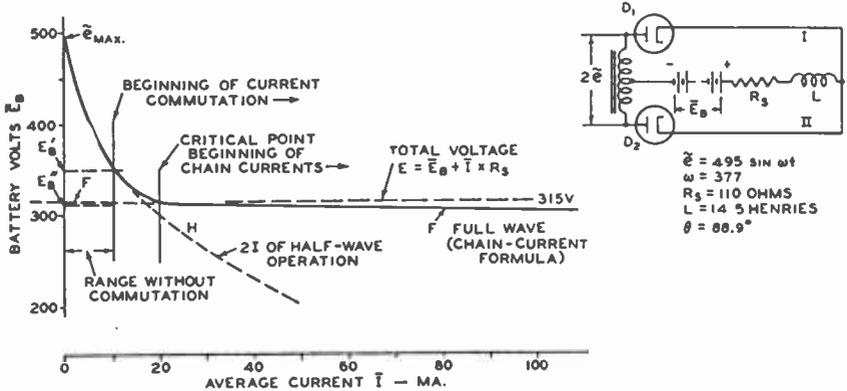
Fig. 22—Operating characteristic of a full-wave, choke-input rectifier circuit with battery load $\bar{E}_B$ or resistance load $R_L = \bar{E}_B/I$ shunted by a large capacitance.

The conduction angle increases from $\phi = 0$ at $\bar{E}_B = \tilde{e}_{max}$ to $\phi = \pi$ at the critical point E_B'' which marks the beginning of chain current operation.

The critical operating condition is obtained by solving for $i = 0$ with $\phi = \pi$ or by equating the direct current to the negative peak value of the total alternating current in L. The critical point is hence specified by a certain current or by a certain ratio K of direct-current resistance to alternating-current impedance in the circuit. With reference to the equivalent circuit treated in the following section, a relation to the fundamental alternating-current component of the rectified current (see (10)), i.e., to the impedance $Z_{(2F)}$, at double line frequency is more useful. We set, therefore,

$$\frac{(R_s + R_L)}{Z_{(2F)}} = K \tag{8}$$

and determine significant values of K for particular circuit impedance conditions.

If we neglect harmonics higher than $2F$, which contribute little to the peak value because of phase shift and increasing attenuation in L, the peak ripple current (equation (10)) becomes

$$i_{\min} = 4/3\pi \ (\bar{e}_{\max}/Z_{(2F)})$$

and setting it equal to the average current
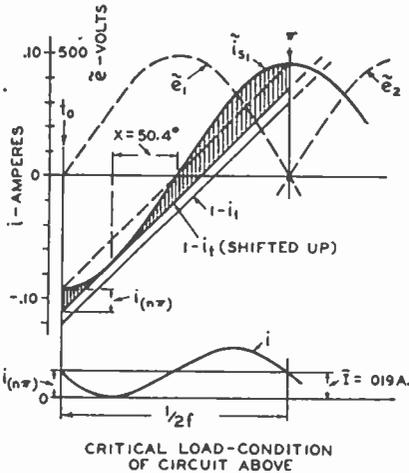
Fig. 23—Graphic solution for the critical load condition with negligible series resistance.

CRITICAL LOAD-CONDITION OF CIRCUIT ABOVE

CRITICAL LOAD-CONDITION OF CIRCUIT AS SHOWN ABOVE WITH THE FOLLOWING VALUES:
$\tilde{e} = 495 \ \text{SIN} \ 377t$
$R_s = 3770 \ \text{OHMS}$
$L = 10 \text{HENRIES}, \omega L = 3770 \ \text{OHMS}$
$\theta = 45°$
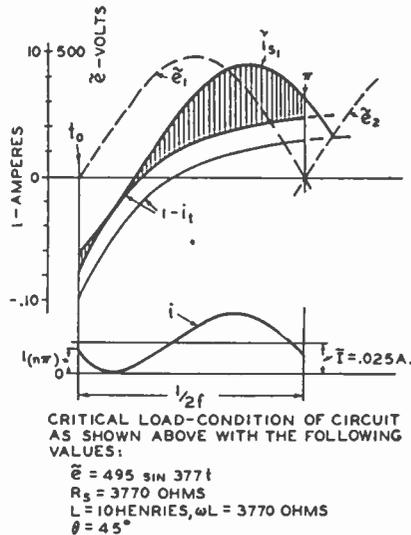
Fig. 24—Same as Fig. 23 but with large series resistance.

$$I = 2/\pi \ (\bar{e}_{\max}/(R_s + R_L))$$

we obtain $K = 1.5$ for the ratio as shown in (8).

The exact solution for the critical current can be obtained from a graphic analysis by simple reasoning for the case $R_s = 0$. The general solution will only be indicated. It is obtained by drawing the complementary curve $(1 - i_t)$ of the total transient beginning at the time $\bar{e} = 0$ (see Figs. 23 and 24) and shifting it upward until it touches the current i_s, thus solving for $i = 0$ at the point of contact. Note that $i_{(n\pi)}$ is the same at t_0 and π in both cases shown.

For $R_s = 0$, the transient section becomes a straight line having the slope $2/\pi$ and running parallel to the peak-to-peak connecting line of $\bar{i}_s$. The sine-wave slope $2/\pi = -\cos x$ gives the point of contact at $X = 50.4$ degrees (Fig. 23), and the peak ripple current is obtained from

$$i_{\min} = \bar{i}_{s\max} \left(\sin 50.4° - \frac{50.4}{90} \right) = 0.211 \bar{i}_{s\max}$$

$$= 0.211 \frac{\bar{e}_{\max}}{\omega L}.$$

Equating this value to the average current given by (10), we obtain the value $K = 1/0.211 = 1.51$ for circuits with $R_s = 0$. The graphic analysis of circuits with larger resistance (see Fig. 24) furnishes K values sufficiently close to 1.5 to justify the use of this constant for all practical purposes. For practical circuits with $2\omega L \gg 1/2\omega C$ we may further write $Z_{(2F)} \cong 2\omega L$ and obtain the *critical inductance*[8]

$$L_0 \cong (R_s = R_L)/2\omega K = (R_s + R_L)/6\pi F. \tag{9}$$

c) Equivalent circuit for the chain current operating range ($\phi = \pi$ or $(R_s + R_L) < 1.5 Z_{(2F)}$)

Inspection of (7b) shows that average and commutation current are directly proportional to the sum of the battery current I_B and a term having a constant current value "I_K" for a given circuit and constant line voltage. Equation (7b) can be changed into the form

$$I = (I_K R_s)/(R_L + R_s),$$

indicating that the secondary circuit may be replaced by an equivalent circuit without switches and energized by a voltage which contains a constant direct-current component $\bar{E} = I_K R_s$. The equivalent voltage in the circuit is the commutated sine wave resulting from the sequence of positive half cycles $+ \bar{e}_1$ and $+ \bar{e}_2$ in the range $\phi = \pi$. The equivalent circuit is shown in Fig. 25(a). The single generator may be replaced by a battery and a series of sine-wave generators (Fig. 25(b)) having amplitudes and frequencies as given by the following equation of the

---

[8] The relation $L_0 = R_L/1000$ was given on an empirical basis for $\omega = 377$ by F. S. Dellenbaugh, Jr., and R. S. Quinby, "The important first choke in high-voltage rectifier circuits," *QST*, vol. 16; pp. 14-19; February, 1932.

commutated sine wave:

$$e = \frac{2\tilde{e}_{\max}}{\pi}\left(1 - \frac{2\cos 2F}{1\cdot 3} - \frac{2\cos 4F}{3\cdot 5} - \frac{2\cos 6F}{5\cdot 7} - \cdots\right). \quad (10)$$

All current components in the circuit may now be computed separately by steady-state methods; the direct-current component is the total average voltage $\tilde{E}$ in the circuit.

Some useful relations of voltage components are: Line voltage induced in one half of the secondary winding (root-mean-square)

$$|\tilde{E}| = 1.1\tilde{E}$$

Total average voltage

$$\tilde{E} = \begin{cases} 0.90\,|\tilde{E}| \\ 0.637\tilde{e}_{\max} \end{cases}$$
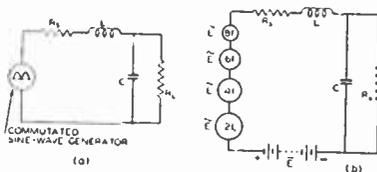
Voltage of frequency $2F$ (root-mean-square)

$$|\tilde{E}|_{2F} = \begin{cases} 0.424\,|\tilde{E}| \\ 0.471\tilde{E} \end{cases} \quad (11)$$
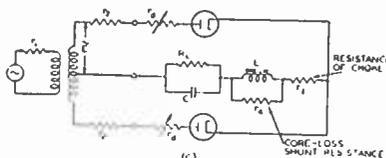
Fig. 25—Components of equivalent and practical full-wave, choke-input circuits.

Voltage of frequency $4F$ (root-mean-square)

$$|\tilde{E}|_{4F} = \begin{cases} 0.085\,|\tilde{E}| \\ 0.0945\tilde{E} \end{cases}$$

Total choke voltage (root-mean-square)

$$|E|_L = \begin{cases} \sqrt{|\tilde{E}|^2 - E^2} \\ 0.482\tilde{E} \end{cases}$$

The current components in the common circuit branch are calculated from the above voltages divided by the impedance of one branch circuit at the particular frequency. Because the current is commutated every half cycle of the line frequency from one to the other branch circuit, the average current in each diode circuit is one half of the total

average current; and root-mean-square values of currents or current components in each branch circuit are obtained by multiplying the root-mean-square current values in the common circuit branch by $1/\sqrt{2}$. The peak current in each diode circuit has the same value as in the common circuit branch.

Average load current

$$I = \frac{\bar{E}}{R_s + R_L}$$

Average plate current (per diode)

$$I_p = 0.5I \tag{12a}$$

Double-frequency current (root-mean-square) in common circuit branch

$$| I |_{2F} = \frac{| \bar{E} |_{2F}}{Z_{(2F)}}$$

Total current (root-mean-square) in common circuit branch

$$| I |_L = \sqrt{I^2 + | I |^2_{2F}}$$

Root-mean-square diode current or root-mean-square current per transformer winding

$$| I |_p = \frac{| I |_L}{\sqrt{2}} \tag{12b}$$

Peak diode current

$$\hat{i}_d = I + (| \tilde{I} |_{2F} \times \sqrt{2}$$

The total power dissipated in diode and load circuits of the practical secondary circuit shown in Fig. 25(c) is the sum of the power losses in the circuit resistances. In equation form, it is

Total power = series-resistance loss

+ choke-core loss

+ direct-current power in load.

The plate dissipation per diode is given by

$$P_d = 0.5 \mid I \mid^2_L \times \mid r_d \mid . \tag{13}$$

With reference to (5), we have

$$P_d = 0.5 \mid I \mid_L^2 \times \frac{\bar{e}_d}{I} \tag{14}$$

where $\bar{e}_d$ is the diode voltage taken from the static diode characteristic at the output-current value I.

d) Regulation

The regulation of choke-input circuits is determined by the total series resistance $\overline{R}_S$, since the voltage $\bar{E}$ in the circuit is constant in the useful chain current range for an energizing alternating voltage of constant value. Thus, the regulation curve has the slope $\overline{R}_S$ (see Fig. 26), which includes the diode resistance. The regulation curve
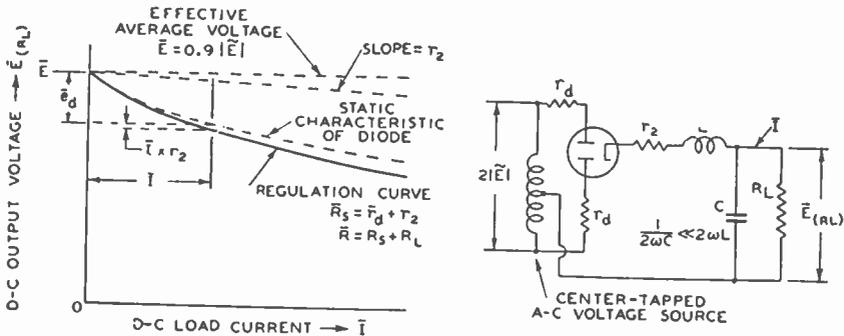
Fig. 26—Regulation characteristic of a full-wave, choke-circuit with high-vacuum diode.

for a circuit with high-vacuum diodes is the sum of the 3/2-power-law diode characteristic and the ohmic series resistance r_2 of one branch circuit as shown in Fig. 26. The curve is correct for constant voltage $\bar{e}$ and beyond the critical current value. In practical circuits, the voltage source $\bar{e}$ has a certain equivalent resistance, which must be added to r_2. The regulation curve Fig. 26 is invalid below the critical current value and must be replaced by a curve following the laws discussed for Fig. 22.

The equivalent internal resistance of the rectifier circuit as a direct-current supply source is the slope of the regulation curve at the current value under consideration. This value should be used for steady-output conditions only, since the reactances in the load circuit cause transients at the instant of sudden load changes.

2. The Condenser-Input Circuit

In rectifier circuits with shunt-condenser-input loads, the condenser is alternately charged and discharged. In the final state of operation, charge and discharge are balanced. The graphic analysis of such circuits is comparatively simple and readily followed. Formulas for the calculation of specific circuit conditions are easily derived from the constructions.

a) Circuits without series resistance

The graphic analysis of a half-wave rectifier circuit without series resistance (R_s) is illustrated in Fig. 27. Steady-state voltage $\bar{e}$ and
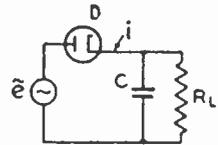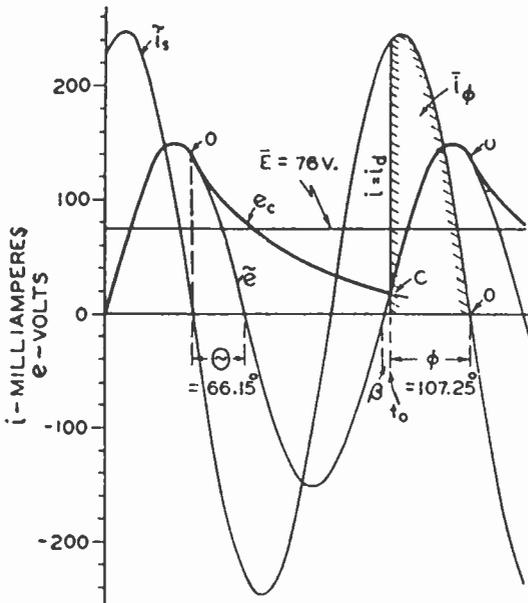
$C = 4 \ \mu f$
$R_L = 1500 \ \text{OHMS}$
$R_S = 0$
$\tilde{e} = 150 \sin 377t$

Fig. 27—Graphic solution of operation for a half-wave, condenser-input circuit without series resistance.

current $\bar{\imath}_s$ are constructed on the assumption that the diode is short-circuited. The steady-state condenser voltage $\bar{e}_c$ coincides with $\bar{e}$ because $R_s = 0$.

The diode timing is as follows:

The diode opens the circuit at point 0 when the diode current becomes zero.

Since the condenser-discharge circuit consists of C and R_L, the condenser voltage decays exponentially as shown. At point C it has become equal to the energizing voltage $\bar{e}$. The diode becomes conducting and closes the circuit. Because there is no potential difference between the steady-stage voltages $\bar{e}$ and $\bar{e}_c$, the condenser does not

receive a transient charge. The current, therefore, rises instantly to the steady-state value of the $\tilde{i}_s$ curve and follows it until zero at point 0.

The timing of the full-wave circuit in Fig. 28 is quite similar. The time for the condenser discharge through R_L is reduced since e_o meets the positive half cycle $\tilde{e}_2$ and thus closes the circuit through D_2. Point C in Fig. 28 is located at a higher value of $\tilde{e}$ than in Fig. 27. The conduction angle ϕ is consequently reduced although C, R_L, and Θ have the same values in both circuits. The average current in the full-wave circuit is, therefore, smaller than twice that of the half-wave circuit.

$C = 4 \mu f$
$R_L = 1500$ OHMS
$R_s = 0$
$\tilde{e}_1 = \tilde{e}_2 = .150 \sin 377t$

Fig. 28—Graphic solution of operation for a full-wave, condenser-input circuit without series resistance.

Some of the relations obtainable directly from Figs. 27 and 28 are

i.   the conduction angle $\phi = 180° - \Theta - \beta$.   (15)

The intersection of $\tilde{e}$ with the decaying voltage e_t furnishes for half-wave operation $(n = 1)$ and full-wave operation $(n = 2)$

ii. $\sin \beta = \sin \Theta \, \epsilon^{-(\pi+\Theta+\beta)/\omega C R_L}$ for $n = 1$

and $\sin \beta = \sin \Theta \, \epsilon^{-(\Theta+\beta)/\omega C R_L}$ for $n = 2$ (16)

where π, Θ, and β in the exponents are in radius. This equation may be solved graphically or by trial and error, varying β.

iii.   The average current during conduction time is

$$I_{(\phi)} = I_s \, (1 - \cos \phi)/\phi.$$

It is the area under a sine-wave section divided by its base. Hence, the average plate current is as shown in (iv).

iv. $$I_p = \bar{i}_{(\phi)} \, \frac{\phi}{2\pi} = \frac{t_s}{2\pi} \, (1 - \cos \phi). \tag{17}$$

v.   Average current I and voltage $\bar{E}$ in the load resistor are

$$\left. \begin{aligned} I &= I_p \quad \text{for} \quad n = 1 \\ I &= 2I_p \quad \text{for} \quad n = 2 \\ \bar{E} &= IR_L \end{aligned} \right\} . \tag{18}$$

vi. The diode peak current $\hat{i}_p$ is, obviously

$$\left. \begin{aligned} \hat{i}_p &= \hat{i}_s \qquad \text{for} \quad \phi > 90° \\ \hat{i}_p &= \hat{i}_s \sin \phi \quad \text{for} \quad \phi < 90° \end{aligned} \right\} . \tag{19}$$

and

The performance of these circuits, hence, is determined by their power factor ωCR_L and the phase number n. It will be evident from the following that the series resistance R_s of practical circuits appears as an additional parameter which cannot be neglected.

b) Circuits with series resistance

In circuits with series resistance, the steady-state condenser voltage $\bar{e}_c$ does not coincide with the supply voltage $\bar{e}$, as illustrated in Figs. 29 and 30. Phase displacement and magnitudes of current and voltage under steady-state conditions are required for analysis of the circuit and are computed in the conventional manner. The parallel circuit $C \,||\, R_L$ is converted into an equivalent series circuit to determine the angles Θ and Θ' by which $\bar{i}_s$ is leading $\bar{e}_c$ and $\bar{e}$, respectively. The steady-state condenser voltage $\bar{e}_c$ in the parallel circuit equals the voltage across the equivalent circuit as shown by the vector diagram in Fig. 30.

The diode opens the circuit at the instant $i_d = 0$. For circuit constants as in Fig. 30, the diode current i_d substantially equals $\bar{i}_s$ at the time of circuit interruption because the transient component i_t' of the current, as shown later, has decayed to a negligible value. Point 0

is thus easily located. In circuits with large series resistance, however, $i_d = 0$ does not coincide with $\tilde{i}_s = 0$ due to slow decay of the transient i_t'. In both cases the condenser voltage $e_{c(0)}$ equals the voltage $\tilde{e}_{(0)}$ at the time 0, because $i_d = 0$ and consequently there is no potential difference on R_s and transients do not occur at 0. The condenser voltage decays exponentially on R_L from its initial value at 0, as discussed for circuits with $R_s = 0$, and meets the supply voltage $\tilde{e}$ again at point C. At this instant (t_0), the diode closes the circuit. Current and voltage, however, do not rise to their steady-state values as in
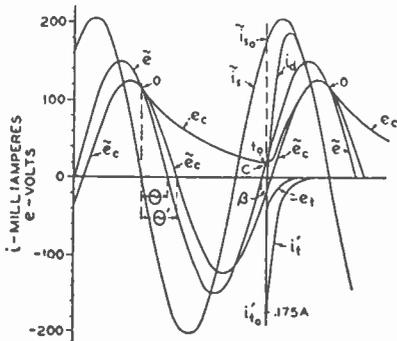
Fig. 29 (*above*)—Graphic solution of operation for a half-wave, condenser-input circuit with series resistance.
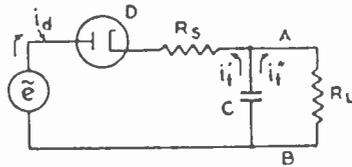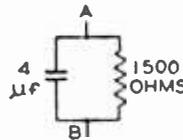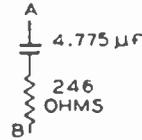
Fig. 30 (*right*)—Equivalent series circuit for the analysis of half-wave, condenser-input circuits with $R_s > 0$.
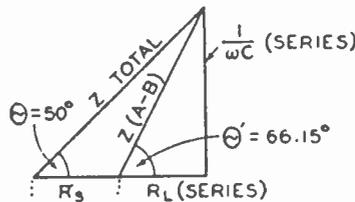
circuits with $R_s = 0$, because the steady-state voltage $\tilde{e}_{c(0)}$ differs from the line voltage $\tilde{e}_{(0)}$ by the amount $\Delta e_c = \tilde{i}_{s(0)} R_s$. A transient voltage of initial value $e_{t(0)} = -(\tilde{i}_{s(0)} R_s)$ occurs on C. It drives transient currents i_t' *and* i_t'' determined by Ohm's law through the resistances R_s and R_L respectively. (See Fig. 30).

The transients e_t and i_t' prevent voltage and current from following the steady-state wave forms, as

$$i_d = \tilde{i}_s + i_t' = \tilde{i}_s - \tilde{i}_{s(0)} \epsilon^{-t/(R_s \| R_L)C} \tag{20}$$

and

$$c_c = \tilde{e}_c + e_t = \tilde{e}_c + R_s \tilde{i}_{s(0)} \epsilon^{-t/(R_s||R_L)C} \tag{21}$$

between the time t_0 and the opening time at 0.

For small values R_S and C, the transient decay is rapid as shown in Fig. 29 and point 0 is readily determined. The oscillogram Fig. 31 was taken on the circuit Fig. 30 and checks the graphic construction.

The solution of operating conditions in circuits with large time constants requires additional steps, as e_c and i_d do not reach steady-state values before $\tilde{i}_s = 0$. The diode opens the circuit earlier at an angle β', which increases from cycle to cycle as shown for a full-wave circuit in Fig. 13. The condenser voltage e_c rises in successive conduction periods until its numerical decay over R_L equals the numerical rise during ϕ. This final condition is shown in Fig. 32(b). The graphic solution for the final operating condition is illustrated in Fig. 32(a) and is made as follows:
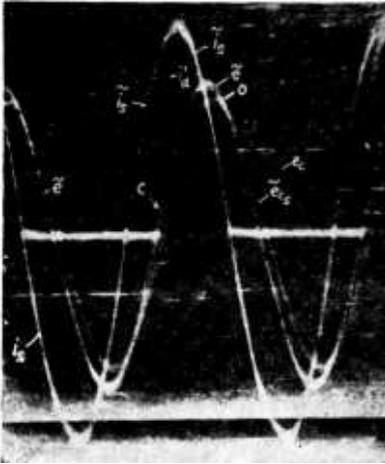
Fig. 31—Oscillograms taken with circuit of Fig. 30.

Steady-state current $\tilde{i}_s$ and voltage $\tilde{e}_s$ are drawn with proper phase relation. A closing time t_0 is assumed near the estimated average output voltage, condition A in Fig. 32(a) assumes $\tilde{i}_{s(0)} = 0.7A$ and $\tilde{e}_{(0)} = 258$ volts at t_0. The current transient i_t' is subtracted graphically from $\tilde{i}_s$. Only two points t_1 and t_2 are necessary near the intersection; t_1 gives a decay of 57.4 per cent and then t_2 gives a decay of 50 per cent from $\tilde{i}_{s(0)}$. The intersection with the $\tilde{i}_{s1}$ curve gives a solution for i_p equal to 0 and determines line 0, which gives $\tilde{e}_1 = 308$ volts which is also the volatge e_c. This voltage decays now over R_L until it intersects the following half cycle $\tilde{e}_2$ for closing time C_2 at point $A = 283$ volts which is the second closing time. As this voltage is higher than the initially assumed voltage ($\tilde{e}_{(0)} = 258$ volts), the final condition is not yet reached. A second trial marked B was made with an initial voltage $\tilde{e}_{s(0)} = 333$ volts and furnished $\tilde{e}_{(2)} = 319$ volts at C_2. The correct condition $\tilde{e}_{(0)} = \tilde{e}_{(2)}$ is obtained from the auxiliary graph in Fig. 32(a) in which the voltage pairs A and B are connected by a straight line, which intersects the 45-degree lines $\tilde{e}_{(0)}C_1 = \tilde{e}_{(0)}C_2$ at the point X, and provides the solu-

tion for the final condition $\tilde{e}_{(0)} = 306$ volts. If desired this value can be checked and corrected by exact calculation.

The final construction in Fig. 32(b) was made with this value. The shaded areas include the amplitude values i_d and e_c during ϕ which are given by (20) and (21).

The average current during ϕ is the area under the sine-wave section minus the area under the exponential curve i_t, both divided by the base. This furnishes

$$\tilde{i}_{d(\phi)} - \tilde{i}_{smax} \mid (\cos \beta' - \cos (\phi + \beta'))$$

$$- \omega CR' (1 - \epsilon^{-\phi/\omega CR'}) \sin (\theta + \beta)]/\phi \qquad (22)$$
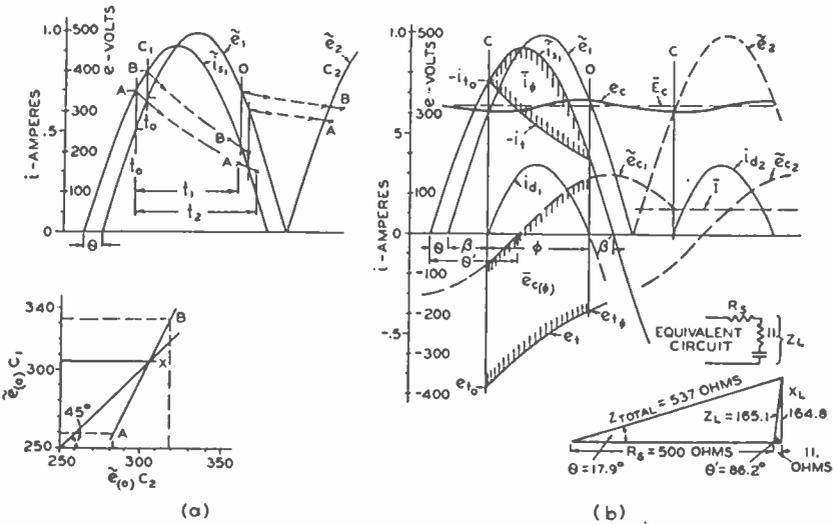
Fig. 32—Graphic solution of final operating conditions for circuit in Fig 13.

with $R' = R_S||R_L$ and ϕ, β and β' determined graphically from the construction or by trial of values. The average plate current per diode is again

$$I_p = i_{d(\phi)}'\phi°/360°$$

and the direct load current in this full-wave circuit is $I = 2I_p$. In case of large time constants, as in the example, the average condenser voltage E_c is quite accurately obtained from

$$\bar{E}_c = 0.5(e_{c(0)} + e_{c(\phi)}) \qquad (23)$$

and the load current by Ohm's law $I = \bar{E}_c/R_L$.

The root-mean-square values of ripple voltage and diode current are needed for many calculations. They may be obtained for all cases from

$$| E |_{\text{(ripple)}} = 0.321 \ (e_{c\text{max}} - e_{c\text{(min)}}) \tag{24}$$

and

$$| I_p | = 1.1 I_p \ \sqrt{\overline{\frac{360°}{\phi°}}}. \tag{25}$$

Equation (24) holds within 10 percent for wave shapes varying from a sine-wave to a saw-tooth and (25) gives better than 5 per cent accuracy for all wave shapes occurring in condenser-input circuits.

c) Generalized operation characteristic (steady-state operation)

It has been shown that the conduction angle is a function of the circuit constants in condenser-input circuits. The section of the energizing voltage $\bar{e}$ utilized during conduction time has, therefore, no fixed value as in choke-input circuits where $\phi = 180$ degrees and where the voltage $\bar{e}$ during ϕ is a half sine wave. It is, therefore, not possible to derive a general equivalent circuit for condenser-input circuits which contains a voltage source of fixed wave shape and magnitude.[9]

Steady-state conditions as well as transients are controlled by the circuit constants, which are contained in the product $\omega C R_L$. The angle ϕ depends further on the relative magnitudes of R_L and R_S and is, therefore, described in general if also the ratio R_S/R_L is known. General curve families may thus be evaluated which show the dependent variables $\bar{E}$, $\hat{\imath}$, and I in terms of ratio versus the independent variable $\omega C R_L$ for various parameter values R_S/R_L. The series resistance R_S includes the equivalent diode resistance which is evaluated by means of (6), because the current wave is periodic in the final operating state. The reasoning leading to (6) is not applicable to a single transient, as obtained for starting conditions of rectifier circuits.

Generalized characteristics have been evaluated for the three types of circuits shown in Fig. 9. The characteristics in Figs. 3, 4, and 5 show the average voltage $\bar{E}$ across the load resistance R_L as a function of $\omega C R_L$ and $\overline{R}_S$ for half-wave, full-wave, and voltage-doubling circuits. They permit the solution of the reversed problem to determine the magnitude of the applied voltage necessary to give a certain

---

[9] The equivalent voltage may be expressed by a Fourier series for each individual case as shown for the simplest case $R_s = 0$ by M. B. Stout in footnote reference 1; the method, however, is hardly suitable for practical circuit analysis.

average voltage output for a given load. The series-resistance value $\bar{R}_s$ includes the equivalent average resistance $\bar{r}_d$ of one diode and the power-transformer resistances as reflected into one secondary winding. As their complete calculation required too much time, the characteristics were plotted from accurately measured values. The measurements were made on circuits of negligible inductive reactance. Series-resistance values in these circuits were determined accurately by the method shown in Fig. 10. Table II gives a number of calculated values which show the accuracy of the curves to be approximately 5 per cent or better.

Table II

Type of Condenser-Input Circuit	$n\omega CR_L$	$\dfrac{\overline{R}_s}{n\,R_L}$	θ degrees	ϕ degrees	$\dfrac{\bar{E}}{\bar{e}_{max}}$	$\dfrac{i_a}{I_p}$	$\dfrac{|I_p|}{I_p}$
Half-Wave $n=1$	0.5	0	26.5	153.5	0.335	3.33	1.69
	1.	0	45.0	134.0	0.384	3.68	1.81
	2.	0	63.4	111.6	0.486	4.61	2.00
	2.26	0	66.15	106.4	0.503	4.91	2.02
	4.	0	75.9	87.1	0.623	6.60	2.24
	8.	0	82.9	65.1	0.742	9.86	2.60
	16.	0	86.4	48.6	0.862	13.92	3.00
	32.	0	88.2	35.3	0.930	19.90	3.51
	64.	0	89.1	25.1	0.996	27.5?	4.16
	2.	0.10	—	121.	0.434	4.48	1.9
	2.26	0.147	50.	123.	0.428	4.42	1.88
	4.	0.05	65.1	99.3	0.632	5.28	2.1
	4.	0.10	56.	108.4	0.537	5.14	2.0
Full-Wave $n=2$	1.	0	26.5	142.5	0.644	3.47	1.75
	2.	0	45.0	121.0	0.678	4.17	1.90
	4.	0	63.4	92.6	0.740	6.06	2.17
	4.52	0	66.15	86.8	0.744	6.55	2.24
	8.	0	75.9	67.0	0.816	9.30	2.55
	16.	0	83.0	49.0	0.885	13.74	3.00
	32.	0	86.4	35.6	0.945	19.70	3.50
	64.	0	88.2	25.4	0.999	27.1?	4.15
	4.	0.05	—	104.	0.671	5.43	2.05
	4.52	0.0735	50.	105.	0.636	5.35	2.04
	8.	0.05	56.	90.	0.710	6.20	2.20
	30.2	0.10	17.9	100.6	0.646	5.39	2.08

In compiling the data for the current-ratio characteristics in Fig. 6, it was found that the three rectifier-circuit types could be shown by a single family after a "charge factor" n was added to the product of the circuit constants ωCR_L and to R_s as shown in Table II.

The factor n is unity for the half-wave circuit. For the full-wave circuit, n is 2 because the condenser C is charged twice during one cycle. For the voltage-doubling circuit, n is $\frac{1}{2}$ because the two condensers require together twice the charge to deliver the same average current at double voltage. The values in the table indicate that the factor n is actually not a constant. The mean value of the current ratios does, however, not depart more than 5 per cent from the true value, the error being a maximum in the steep portion of the curves and decreasing to zero at both ends. The upper section of Fig. 6 shows the ratio of root-mean-square current to average current per diode plate. This family is of special interest in the design of power transformers and for computation of diode plate dissipation.

Fig. 7 shows the root-mean-square value of the ripple voltage across R_L in per cent of the average voltage.

The voltage-doubling circuit shown with the other two condenser-input circuits in Fig. 9 may be regarded in principle as a series connection of two half-wave rectifier circuits. Each condenser is charged separately during conduction time of one diode, but is discharged in series with the other condenser during the time of nonconduction of its associated diode. The analysis of operation is made according to the method discussed but will not be treated. The average anode characteristics of RCA rectifiers are shown in Fig. 8. The method of carrying out a practical analysis by use of these curve families has been outlined in the first section of this paper.

<center>APPENDIX</center>

System of Symbols

The number of special symbols and multiple indexing have been greatly reduced by introducing four special signs for use with any symbol.

1) The symbols in general are of standard notation, lower case letters i, r, indicate instantaneous, sectional, or variable values and capital letters I and $\dot{R}$ indicate steady values.

2) Special values
   a) *Sinusoidal voltages or currents* are indicated by a sine-wave sign above the symbol $\tilde{e}$, $\tilde{i}$, $\tilde{E}$. Their maximum values are indicated by index, $\tilde{e}_{max}$, $\tilde{E}_{max}$.
   b) *Peak values* are indicated by circumflex; $\hat{e}_c$, $\hat{i}$, $\hat{r}_d$, maximum peak values are written $\hat{i}_{max}$, etc.
   c) *Average values* are indicated by a horizontal bar; $\bar{E}$, $\bar{I}$, $\bar{R}$.

d) Root-mean-square values are indicated by vertical bars $|E|$, $|I|$, $|R_s|$.

3) An index in parenthesis specifies the time at which the symbol is valid, i.e., its numerical value. Hence, $\bar{i}_{s(\pi)}$ is the steady-state alternating-current value at the time π and $i_{t(0)}$ is the transient current at the time 0. When used with an average or root-mean-square value, the time index specifies the period over which average or root-mean-square values are taken, such as $I_{(\phi)}$, $|i_p|_{(\phi)}$. A conduction time index (ϕ) on resistance values such as $\bar{r}_d$, $\overline{R}_s$ is unnecessary. (See definition.)

SPACE-CURRENT FLOW IN VACUUM-TUBE STRUCTURES*†

By

B. J. THOMPSON

RCA Laboratories,
Princeton, N. J.

Summary—From well-known formulas for space-current in diodes and for amplification factor in triodes, interelectrode capacitance, plate current, and potential distribution in triodes and multi-grid tubes are determined through use of the concept of planes of equivalent potential. By the same means, amplification factor in multigrid tubes is derived.

INTRODUCTION

VACUUM-TUBE design is a subject which has intrigued many workers, largely, one may suspect, because it has presented many possibilities for ingenious methods of analysis. The resulting knowledge of the design factors which determine the various performance characteristics of tubes is quite complete and is expressed in terms which can readily be applied to practical tube-design problems. In spite of this state of the art, the general tendency is to make use of the more "scientific" phases of tube design to aid qualitative understanding rather than to supply specific design information. In this paper, the writer presents some of the concepts of vacuum-tube analysis which he has found informative and useful.

First, space-current flow in diodes will be discussed. Then, methods will be presented for reducing triodes and multigrid tubes to equivalent diodes. Amplification factor, interelectrode capacitance (cold), and electron transit time will be covered. The writer claims little originality and no novelty in this material. Some effort has been made to give credit to the proper sources.

A. DIODE THEORY

Ideal Case

The simplest vacuum tube is the diode. The behavior of multi-electrode tubes may be described most readily in terms of the behavior of a diode. For these reasons our treatment will start with the diode.

In the ideal diode, electrons are emitted from the cathode in un-

---

* Decimal Classification: R131.
† Reprinted from *Proc. I.R.E.*, September, 1943.

limited numbers at zero velocity and a part of these are drawn over to the anode under the influence of the positive field established by its potential.

In Figure 1, K represents the infinite plane cathode at zero potential and A the plane anode at a positive potential E_b, spaced a distance d_{kp} from the cathode. Let us suppose first that no electrons are emitted from the cathode. The potential distribution will then be as represented by the line a, the gradient at all points being E_b/d_{kp}. If now the cathode begins to emit a limited supply of electrons, all of these electrons will be drawn to the anode. The electrons move at a finite velocity and, therefore, there is a certain number of them in the space at all times. The field set up by the negative "space charge" of these electrons acts to depress the potential in the space below that of the first condition, increasing the field near the anode and decreasing it near the cathode. This condition is shown by line b.
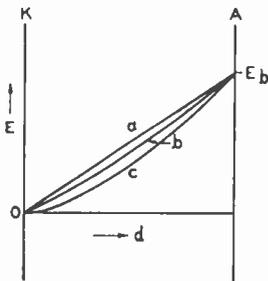
Fig. 1—Potential distribution in a diode with varying amounts of space charge.

If the rate of emission of electrons is continually increased, all of the emitted electrons will be drawn to the anode and the gradient at the cathode continually reduced until the gradient reaches zero. Since the electrons are assumed to be emitted with zero velocity, they can not move against a retarding field; therefore, there will be no increase in anode current with further increase in emission beyond this point. The condition of zero gradient at the cathode is represented by the line c in Figure 1.

The mathematical analysis of the ideal parallel-plane case is quite simple. It will be presented here as an example of this type of analysis.

Poisson's equation in rectangular co-ordinates is

$$\partial^2 E/\partial x^2 + \partial^2 E/\partial y^2 + \partial^2 E/\partial z^2 = -4\pi\rho. \tag{1}$$

Since there is no gradient in directions parallel to the cathode and anode, the equation becomes simply

$$\partial^2 E/\partial x^2 = -4\pi\rho. \tag{2}$$

We may also write that

$$I = \rho v \tag{3}$$

and $\qquad\qquad v = (2eE/m)^{1/2}$ $\qquad\qquad\qquad$ (4)

where ρ is the space-charge density, E the potential at any point a distance x from the cathode, v the velocity of the electrons at x, I the current per unit area, and e and m the charge and mass of the electron.

On combining the last three equations, we obtain

$$\frac{d^2E}{dx^2} = -4\pi \frac{I}{(2eE/m)^{1/2}}.$$ $\qquad$ (5)

If we multiply both sides by dE/dx and integrate once, we obtain

$$\frac{1}{2}\left(\frac{dE}{dx}\right)^2 = \frac{8\pi I}{(2e/m)^{1/2}} E^{1/2} + F_0^2$$ $\qquad$ (6)

where F_0 is the field at the cathode. If we let F_0 equal zero, a second integration gives us

$$E^{3/4}\,\bigg|_0^{E_b} = 3(\pi I)^{1/2}\left(\frac{m}{2e}\right)^{1/4} x\,\bigg|_0^{d_{kp}}$$ $\qquad$ (7)

or $$I = \frac{1}{9\pi}\left(\frac{2e}{m}\right)^{1/2}\frac{E_b^{3/2}}{d_{kp}^2}$$

$$= 2.33 \times 10^{-6}\ (E_b^{3/2}/d_{kp}^2).$$ $\qquad$ (8)

This is the well-known Langmuir-Child[1] equation for space-charge-limited current flow per unit area between parallel-plane electrodes. It means that for each square centimeter of cathode or anode area 2.33 microampers of current will flow with 1 volt difference in potential and a distance of 1 centimeter between cathode and anode, and that a current of 233 microamperes per square centimeter will flow if the potential be raised to a little over 30 volts or the distance reduced to 1 millimeter.

The foregoing analysis is for parallel-plane electrodes. The case of concentric cylinders, of much practical interest, is very much less simple to analyze and, therefore, only the result will be presented here. Excellent analyses are available in the literature[2].

[1] I. Langmuir and K. T. Compton, "Electrical discharges in gases—Part II," *Rev. Mod. Phys.*, Vol. 3, pp. 238-239; April, 1931.

[2] See pp. 245-249 of footnote reference 1.

The current in amperes per centimeter length of the concentric cylinders is given by the well-known Langmuir equation

$$I = 14.66 \times 10^{-6} \; (E_b^{3/2}/r_b\beta_b^2) \tag{9}$$

where r_b is the radius of the anode and β_b^2 is a function depending on the ratio of anode radius to cathode radius. Tables and curves of β have been published[3]. It will be noted that the current again depends on the 3/2 power of the anode voltage; otherwise, the expressions at first glance do not appear very similar. Part of this difference is due to the fact that one expression is for current per unit area, while the other expression is for current per unit length.

It will be interesting to put the two expressions in similar form. Let us divide equation (9) by $2\pi r_b$. Equation (9) then becomes identical with equation (8) except for the presence of the term β_b^2 in the denominator and the fact that the distance r_b is measured from the axis of the cylindrical system. When the ratio of anode diameter to cathode diameter becomes very large, β_b^2 approaches unity and, of course, the distance between cathode and anode approaches r_b as a limit. At this limit, then equations (8) and (9) become identical, and we observe the interesting fact that the anode current flow per unit area is the same in a cylindrical system with fine-wire filament as it would be in a parallel-plane system with the same distance between cathode and anode. This statement, of course, neglects the effect of initial velocity of emission.

At the other limit where the cathode and anode diameters approach each other the system is obviously essentially a parallel-plane one. The value of β_b^2 then changes rapidly and maintains such a value that $r_b^2\beta_b^2$ is equal to d_{kp}^2.

The fact that the two expressions give identical results at the two limits of ratio of anode-to-cathode diameter should not lead one to suppose that the expressions are approximately identical for intermediate ratios. Where the anode diameter is from 4 to 20 times the cathode diameter, the current calculated from (8) is in excess of that indicated by (9) by very nearly 20 per cent. This is the maximum error that would result from the use of expression (8) for cylindrical structures.

The potential distribution between cathode and anode may be calculated most usefully from the expressions for current. From (8) we may write

---

[3] See pp. 247-248 of footnote reference 1.

$$E_b^{3/2}/d_{kp}^2 = E^{3/2}/x^2$$

or

$$E = E_b(x/d_{kp})^{4/3}.$$

In other words, the potential between parallel planes varies as the four-thirds power of the distance from the cathode in the case of space-charge-limited currents.

The potential distribution between concentric cylinders is less simple. We may write from (9)

$$E_b^{3/2}/r_b\beta_b^2 = E^{3/2}/r\beta^2$$

or

$$E = E_b(r\beta^2/r_b\beta_b^2)^{2/3}$$

where β^2 is taken for the ratio r/r_k. This expression is not analytical, the values of β and β_b being obtained from curves or tables.

Effects of Velocities of Emission

Electrons are emitted from a heated surface with a random distribution of velocities in all directions. The velocities which concern us in the present analysis are those normal to the surface of the cathode. This velocity distribution may be expressed most simply as follows: $n/n_0 = \epsilon^{-Ee/kT}$ where n is the number of electrons out of the total number n_0 which has a sufficient velocity to reach a plane electrode parallel to the cathode at a negative potential of E, T is the temperature of the cathode, and k is Boltzmann's constant. Expressed in terms of current this becomes $I = I_s\epsilon^{-Ee/kT}$ where I is the current reaching the negative electrode and I_s is the total emission current from the cathode. To carry this out experimentally, it is necessary that the collector electrode be placed so close to the cathode that space-charge effects do not cause a potential minimum in space.

We initially assumed that all electrons were emitted with zero velocity and that, therefore, the field at the cathode would not be negative. In the practical case where all electrons have finite velocities normal to the cathode, all of the emitted electrons must reach a positive anode parallel to the cathode unless at some point between cathode and anode a negative potential exists.

Fig. 2 shows the potential distribution between parallel-plane cathode and anode for successively higher values of emission. Line a represents the case where there is no emission, and, therefore, no space charge, with resulting constant potential gradient between cathode and anode. Line b shows the case where there is sufficient

emission to reduce the gradient at the cathode just to zero. This is similar to the condition represented by c in Fig. 1 with the important difference that now all electrons pass over to the anode because of their finite velocities of emission.

Any further increase in cathode emission, however, will cause the potential near the cathode to become slightly negative as shown in line c. In this case all electrons having velocities less than E_m are turned back to the cathode, while those electrons having greater velocities of emission pass on to the anode. Further increases in cathode emission cause the potential minimum to become more negative with the result that a larger fraction of the emitted electrons return to the cathode. For continued increase in cathode emission, however, there will always be some slight increase in anode current.
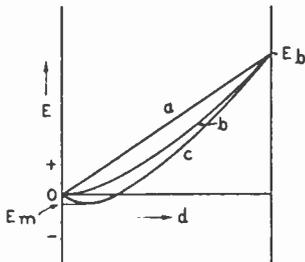
Fig. 2—Potential distribution in a diode, showing the effect of initial velocity of electron emission.

The results obtained from the simple analysis based on zero velocity of emission are obviously not applicable to this practical case if precision is desired. Since a greater maximum potential difference ($E_b + E_m$) is acting over a shorter effective distance ($d_{kp} - d_{km}$) and since the average velocity of electrons is higher because of their initial velocities and hence the sapce-charge effect of the electrons is less, it is o b v i o u s that the space-charge-limited current flow for a given anode potential is greater in the actual case than in the ideal.

Langmuir[1] has presented a complete analysis of the space-charge-limited current flow with initial velocities of emission. He has shown that a good approximation may be made by the use of (8) with a correction for the reduced effective distance and the increased effective potential. His equation is as follows:

$$I_b = 2.33 \times 10^{-6} \frac{(E_b - E_m)^{3/2}}{(d_{kp} - d_{km})^2} \times \left[1 + \frac{0.0247T^{1/2}}{(E_b - E_m)^{1/2}} \right] \qquad (10)$$

where T is the cathode temperature in degrees Kelvin. I_b is in amperes per unit area. E_m is negative in sense. The value of d_{km} in centimeters may be calculated from the approximate expression

[4] See pp. 239-244 of footnote reference 1.

$$d_{km} \approx 0.0156 \ (1/1000I_b)^{1/2} \ (T/1000)^{3/4}.$$

The value of E_m is given by $E_m = - (T/5040) \log_{10} (I_s/I_b).$

More complete results of Langmuir's analysis are too cumbersome to be presented here. The use of (10) should lead to errors not greatly in excess of 2 per cent even under extreme conditions.

It is interesting to observe from Langmuir's calculation in a practical case where the cathode temperature is 1000 degrees Kelvin, the emission density greatly in excess of the anode current, and the anode current density 1 milliampere per square centimeter, that the distance from cathode to virtual cathode is approximately 0.016 centimeter (0.006 inch). Thus, in modern close-spaced vacuum tubes the position of the virtual cathode cannot be neglected.

The error involved in using (9) as compared with the exact solution for cylindrical structures is less than in the corresponding case of parallel planes. For a discussion of the effect of initial velocities in this case, the reader is referred to Langmuir and Compton[5].

The potential distribution between parallel planes, taking into account initial velocities, may best be determined by the use of a plot presented by Langmuir and Compton[6].

B. TRIODE THEORY

Triode Mu Formulas

The earliest analysis of the electric field existing between parallel planes with a parallel-wire screen interposed is that of Maxwell[7]. In this it is assumed that the spacings between the planes and the screen are large compared with the spacings between wires and that these in turn are large compared with the wire diameter. The result expressed in vacuum-tube terminology is

$$\mu = - \frac{2\pi d_{gp}}{a \log_e (2 \sin \pi r/a)}$$

or $\mu = \dfrac{2\pi d_{gp}}{a \log_e (a/2\pi r)}$ (where $\pi r/a$ is small).

---

[5] See pp. 252-255 of footnote reference 1.

[6] See Fig. 42, p. 243 of footnote reference 1.

[7] J. C. Maxwell, "Electricity and Magnetism," third edition, 1904, Vol. 1, section 203.

In these expressions, d_{gp} is the distance from the center of the grid wires to the plate, a the spacing between grid wires ($a = 1/n$, where n is the number of wires per unit length), and r is the radius of the grid wires. It will be noted that the distance between grid and cathode does not appear.

This formula is in serious error when the spacing between grid wires is not large compared with the wire diameter, as is frequently the case. Because of this, van der Bijl developed empirically the formula $\mu = Cd_{gp} r n^2 + 1$, where C is equal to 160 for parallel planes. An obvious defect of this expression is that μ can never be less than unity.

The most generally useful and accurate formula for amplification constant which has been published is that developed by Vogdes and Elder[8]. This analysis assumes that the spacing between grid wires is small compared with the distances between the grid and the other electrodes. The development is as follows.

Fig. 3 represents the geometry of the vacuum tube. By means of a conformal transformation, this same geometry may be represented in different co-ordinates. In such a transformation, equipotential surfaces and flux lines still cross at right angles and all laws of electricity still apply.

Suppose the geometry represented in the w plane in Fig. 3 be transformed to the z plane by the transformation $z = e^{2\pi nw}$.

Since $$z = x + jy$$

and $$w = u + jv$$

then $$x + jy = e^{2\pi nu} \times e^{j2\pi nv} = \rho e^{j\theta}.$$

This transformation is represented in Fig. 4. The cathode is a point at the origin. The grid wires become a single figure intersecting the x axis as $e^{-2\pi nr}$ and $e^{2\pi nr}$. The center of the grid wires is at $x = 1$. The anode is a circle about the origin of radius equal to $e^{2\pi nd_{gp}}$.

The figure representing the grid wires is not a circle. If r is less than $a/2\pi$, however, it can be shown readily that the figure is essentially circular and it will be assumed, therefore, that such is the case. If the figure is a circle, its radius is

$$\frac{e^{2\pi nr} - e^{-2\pi nr}}{2} = \sinh 2\pi nr$$

[8] B. F. Vogdes and F. R. Elder, "Formulas for the amplification constant for three-element tubes," *Phys. Rev.*, Vol. 24, p. 683; December, 1924.

and its center is located at

$$x = \frac{e^{2\pi nr} + e^{-2\pi nr}}{2} = \cosh 2\pi nr.$$

In Fig. 3, if the anode were removed to infinity and a potential applied to the grid, the successive equipotential surfaces at greater distances from the grid would become more and more nearly planes until, at distances several times a, the surface could be regarded as essentially a plane. Therefore, under the limitations of our assumptions concerning relative spacings, the anode plane may be considered to be the equipotential surface due to the field of the grid alone. This
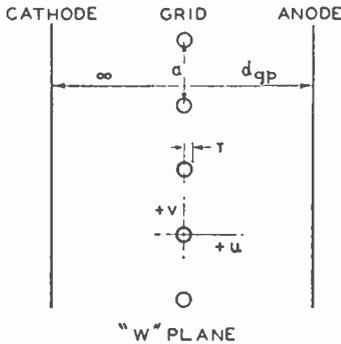
Fig. 3—Cross section of a triode in normal co-ordinates.
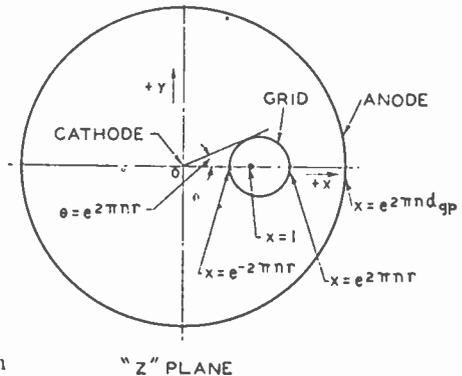
"Z" PLANE

Fig. 4—Cross section of the triode of Fig. 3 transformed from the w plane to the z plane.

is equivalent to saying that a circle of radius $e^{2\pi nd_{gp}} - \cosh 2\pi nr$ drawn about the "center" of the grid wire in Fig. 4 does not differ materially from a circle of radius $e^{2\pi nd_{gp}}$ drawn about the origin. The justification for this assumption may be checked by considering the rather extreme case where $nd_{gp} = 0:50$ and $nr = 0.03$. Then $e^{2\pi nd_{gp}}$ equals 23.1 and $\cosh 2\pi nr$ equals 1.02.

The convenient result of these assumptions is that a line charge placed at the "center" of the circular grid wire, Fig. 4, produces equipotential surfaces at the surface of the grid wires and at the anode, since the charge on the cathode located at minus infinity must be zero.

Let us place a charge $-Q$ at the "center" of the grid wire. The potentials E_k, E_g and E_a of the cathode, grid, and anode become

$$E_k = C + 2Q \log \cosh 2\pi nr$$

$$E_g = C + 2Q \log \sinh 2\pi nr$$

$$E_a = C + 2Q\, 2\pi nd_{gp}.$$

If the cathode potential be taken as zero,

$$E_g = 2Q \log \sinh 2\pi nr - 2Q \log \cosh 2\pi nr$$

$$= 2Q \log \tanh 2\pi nr$$

and $$E_a = 2Q\, 2\pi nd_{yp} - 2Q \log \cosh 2\pi nr.$$

Under these circumstances, the amplification constant may be defined as $\mu = - E_a/E_g$

whence $$\mu = \frac{\log \cosh 2\pi nr - 2\pi nd_{gp}}{\log \tanh 2\pi nr}.$$

The assumptions made in this derivation invalidate the expression for use with relatively very close spacings between electrodes. The same type of analysis as that presented by Vogdes and Elder may be made to give more rigorous results. Salzberg[9] has carried out such an analysis. It differs from that just presented chiefly in that an additional line charge is placed on the x axis, Fig. 4, outside the anode at such a position as to make the true anode cylinder an equipotential surface. Therefore, the anode may be allowed to approach much more closely to the grid. This leads to an expression accurate for cases where the spacing between the anode and grid is small compared with the wire spacing, though not when the wire spacing is small compared with the wire diameter. Salzberg's expression is

$$\mu = \frac{\log \cosh 2\pi nr - 2\pi nd_{gp}}{\log \tanh 2\pi nr - \log\,(1 - e^{-4\pi nd_{gp}} \times \cosh^2 2\pi nr)}.$$

There is no obviously useful definition of amplification factor in the purely electrostatic case (no space charge) when the charge density induced on the cathode is not uniform. It is possible by extension of the analysis described above, however, to arrive at an expression for the charge distribution on the cathode when the spacing between cathode and grid is finite. Salzberg has carried out such an analysis[10]. It departs from that of the cathode at infinity by considering the poten-

---
[9] Bernard Salzberg, "Formulas for the amplification factor of triodes," *Proc. I.R.E.*, Vol. 30, pp. 134-138; March, 1942.
[10] Not published.

tials in space produced by a line charge at the cathode in addition to the others.

The amplification-factor formulas here given may be applied to cylindrical tubes if $r_g \log(r_a/r_g)$ is substituted for d_{gp}, where r_g and r_a are the radii of the grid and anode, provided r/r_g is small.

Equivalent Potentials in Triodes

For most practical purposes in calculating the electric fields at cathode, anode, and the space between, except very near the grid, a potential may be assigned to the plane of the grid. In other words, it is assumed that an equipotential plane may be substituted for the grid without altering the electric fields. This would be true only when the grid wires are small and closely spaced in comparison with the spacings between grid and cathode and anode.
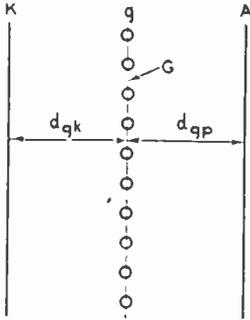
Fig. 5—Triode with equivalent plane G at grid.
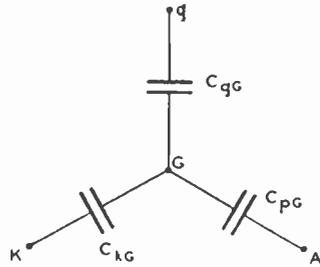
Fig. 6—Star network of the capacitances of the triode of Fig. 5.

The equivalent potential of the plane of the grid E_G may be derived in several ways. The most simple with which the writer is familiar is the following. The capacitance between anode and the equivalent plane G at the grid, Fig. 5, is $C_{pG} = 1/4\pi d_{gp}$ and the capacitance from cathode to G, $C_{kG} = 1/4\pi d_{kg}$ while, by definition $C_{gG} = \mu C_{pG}$. In the star network of capacitances, Fig. 6,

$$E_G = \frac{E_c C_{gG} + E_b C_{pG} + E_k C_{kG}}{C_{gG} + C_{kG} + C_{pG}}.$$

Let us make E_k equal to zero. Then,

$$E_G = \frac{\mu E_c + E_b}{\mu + 1 + d_{gp}/d_{gk}}$$

or

$$E_G = \frac{E_c + E_b/\mu}{1 + 1/\mu + d_{gp}/d_{gk}\mu}.$$

The physical basis for this analysis is that the anode can influence the field at the cathode only by acting through the grid plane. By definition, the grid has μ time the influence of the anode. It is obvious that this reasoning implicitly assumes that amplification factor is proportional to grid-anode spacing, for we might just as well have called the cathode the anode. The quantity $d_{gp}/d_{gk}\mu$ is simply the reciprocal of the amplification factor of the grid with respect to the cathode.

We shall find it convenient to determine another equivalent-potential plane. The equivalent potential of the grid plane depends on grid
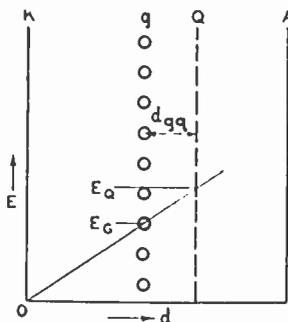
Fig. 7 — Determination of equivalent potential E_q of the Q plane.

and anode potentials and on grid-cathode and grid-anode spacings. Is there an equivalent plane the potential of which depends only on grid and anode potentials and grid-anode spacing?

In Fig. 7, E_G is the equivalent potential of the grid. If the constant potential gradient between grid and cathode extended past the grid, the potential E at any point a distance x from the grid would be

$$E = E_G \ (1 + x/d_{gk})$$

$$= \frac{(E_c + E_b/\mu)}{1 + 1/\mu + d_{gp}/d_{gk}\mu} \cdot (1 + x/d_{gk}).$$

We wish to find a potential $E = E_q$ at a distance $x = d_{gq}$ from the grid which is independent of d_{gk}. At such a point the ratio

$$\frac{1 + x/d_{gk}}{1 + 1/\mu + d_{gp}/d_{gk}\mu}$$

must be independent of d_{gk}. Obviously this means that

$$\frac{x}{d_{gk}} = \frac{d_{gp}/d_{gk}\mu}{1 + 1/\mu}$$

or
$$x = d_{gp}/(\mu + 1) = d_{gq}.$$

The potential E_q is given by

$$E_q = \frac{E_c + E_b/\mu}{1 + 1/\mu}.$$

Applications of this equivalent-potential plane will be given

Interelectrode Capacitances in Triodes without Space Charge

The direct capacitance between grid and anode, C_{gp}, may be calculated readily from the expression for E_G, the equivalent potential of the plane of the grid.

The capacitance per unit area from anode to the equivalent plane of the grid is $C_{pG} = 1/4\pi d_{gp}$.

Then
$$C_{gp} = C_{pG}\frac{dE_G}{dE_g} = \frac{1}{4\pi d_{gp}}\left(\frac{1}{1 + 1/\mu + d_{gp}/d_{gk}\mu}\right).$$

Similarly
$$C_{gk} = \frac{1}{4\pi d_{gk}}\left(\frac{1}{1 + 1/\mu + d_{gp}/d_{gk}\mu}\right).$$

By definition
$$C_{pk} = \frac{C_{gk}}{\mu} = \frac{1}{4\pi d_{gk}}\left(\frac{1}{\mu + 1 + d_{gp}/d_{gk}}\right).$$

These derivations are for the parallel-plane case. The case of cylindrical electrodes may be treated in a similar fashion.

Amplification Factor in Multigrid Tubes

The analysis of multigrid tubes may be readily carried out by use of the second expression for equivalent potential E_q.

In Fig. 8, the Q plane is to be substituted for g_2 and A. Its potential is

$$E_q = \frac{E_{c2} + E_b/\mu_{g2p}}{1 + 1/\mu_{g2p}}$$

and its distance from g_1 is $d_{g1p} = d_{g1g2} + d_{g2p}/(1 + \mu_{g2p})$. We now have a triode and can calculate its μ. The simplest expression is $\mu_{g1q} = (\mu'_{g1g2}/d_{g1g2})d_{g1q}$ where μ'_{g1g2} is the amplification factor of g_1 with respect to a plane at g_2.

Now
$$\mu_{g1g2} = \mu_{g1q}(dE_{c2}/dE_q)$$
$$= \mu_{g1q}(1 + 1/\mu_{g2p})$$
$$= \mu'_{g1g2}(d_{g1q}/d_{g1g2}(1 + 1/\mu_{g2p}).$$

On substituting the expression for d_{g1q} in this equation, one may reduce it to the following form by simple manipulation: $\mu_{g1g2} = \mu'_{g1g2} + \mu'_{g1p}/\mu_{g2p}$ where μ'_{g1p} is the value μ_{g1p} would have if g_2 were removed.

Of course $\mu_{g1p} = \mu_{g1g2}\mu_{g2p}$ whence[11] $\mu_{g1p} = \mu'_{g1g2}\mu_{g2p} + \mu'_{g1p}$. The direct capacitance between g_1 and g_2 may also be determined readily, since $C_{g1g2} = C_{g1q}(dE_q/dE_{c2})$. Also, the capacitances between grids and anode or cathode may be determined in the same manner.

By an obvious extension of the method, amplification factors and capacitances may be determined in structures containing any number of grids.
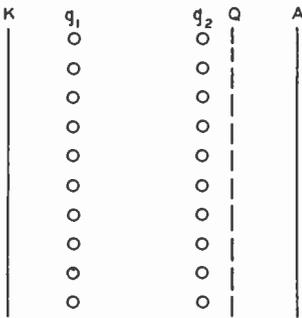
Fig. 8—Determination of amplification factor and inter-electrode capacitance in multigrid structures by use of the Q plane.
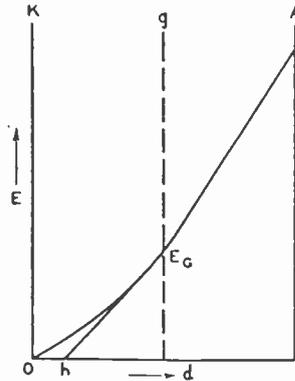
Fig. 9—Apparent location h of the cathode as seen from the grid.

Effects of Space Charge on Potential Distribution in Triodes

In the case where space charge between grid and anode may be neglected (as is usually the case in receiving tubes with negative control grids), a quite precise equivalent diode may be constructed by the use of the first expression for equivalent potential with a space-charge correction. Fig. 9 shows the potential distribution in a triode with space-charge-limited current. It is obvious that the field at the grid is the same as would exist without space charge if the cathode were at point h, determined by drawing a tangent to the potential curve at the grid. If it be assumed that the potential between cathode

---

[11] For an alternative derivation see S. Koizumi, "On the amplification constants of multi-electrode vacuum tubes," *Jour. I.E.E.* (Japan), pp. 505, 857; 1930.

and grid varies as the four-thirds power of distance, d_{gh} is three fourths of d_{gk}. Hence, we must modify the expression for E_G as follows:[12]

$$E_G = \frac{E_c + E_b/\mu}{1 + 1/\mu + (4/3)\,(d_{gp}/d_{gk}\mu)}.$$

The analysis of the current-voltage relationship of a triode may be made directly from the diode case by the use of this equivalent-diode expression. If in the equivalent diode the space current $I_b = f(E_G)$ the cathode current (equal to the plate current with negative grid) is given directly. The transconductance g_m is found by taking the derivative of $f(E_G)$ with respect to E_c. The plate conductance $1/r_p$ is found by taking the derivative[13] of $f(E_G)$ with respect to E_b.

Electron Transit Time in Negative-Grid Triodes

The electron transit time in any electrode structure may be calculated readily if the potential distribution is known. In general

$$t = \int \frac{dx}{v} = \left(\frac{m}{2e}\right)^{1/2} \int \frac{dx}{E^{1/2}}.$$

The calculation of transit time in the absence of space charge is obvious. In a parallel-plane diode with space-charge-limited current, the transit time from cathode to anode may be calculated if it be assumed that

$$E = E_b(x/d_{kp})^{4/3}$$

whence

$$t = \left(\frac{m}{2e}\right)^{1/2} \frac{d_{kp}^{2/3}}{E_b^{1/2}} \int_0^{dkp} x^{-2/3} dx$$

---

[12] B. D. H. Tellegen, "The calculation of the emitted current in a triode," *Physica*, Vol. 5e, pp. 301-315; 1925.

[13] Note added April 6, 1943: Fremlin[14] discusses several old expressions for equivalent diode potential, none of which is similar to that given above, and then derives an expression for anode current of a triode starting from the known condition of grid and anode at such potentials as to maintain the space-potential distribution of a diode undisturbed to the anode (grid at "space potential"). Unfortunately, no simple analysis of this sort is valid when there is appreciable space charge in the grid-anode space as is implicit under Fremlin's assumptions. In the case of negligible grid-anode space charge, Tellegen's expression seems satisfactory.

[14] J. H. Fremlin, "Calculations of triode constants," *Elec. Communications*, July, 1939.

$$= \left(\frac{m}{2e} \right)^{1/2} \frac{3d_{kp}}{E_b^{1/2}}$$

$$= 5.05 \times 10^{-8} \ (d_{kp}/E_b^{1/2})$$

where t is in seconds, d_{kp} in centimeters, and E_b in volts. In other words, the electron take three times as long to pass from cathode to anode as if it had traveled at the final velocity the entire distance, and half again as long as if it had been uniformly accelerated.

The cylindrical analysis is not so simple but may be carried out as presented by W. R. Ferris.[15]

In the case of electron transits between grid and anode, the integration is carried out with the initial velocity of the electron corresponding to the equivalent potential of the grid.

[15] W. R. Ferris, "Input resistance of vacuum tubes as ultra-high-frequency amplifiers," *Proc. I.R.E.*, Vol. 24, pp. 82-108; January, 1936.

THE ELECTRON MECHANICS
OF INDUCTION ACCELERATION*†

By

Jan A. Rajchman and William H. Cherry

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—The radial and axial motions of electrons in the betatron are described by means of a potential function of forces. Previously reported conditions of equilibrium, stability and damping of oscillations are derived for the region of parabolic variation of the potential. Extension of the analysis to non-parabolic regions gives an account of the injection in conventional instruments in better agreement with experiment, particularly in regard to higher voltages of injection. Space charge limitations are discussed with the help of the Laplacian of the potential of forces. By means of an additional radial electric field electrons can be introduced as in the magnetron, without any asymmetry inherent in the conventional betatron circumferential injector. The analysis of the conditions of equilibrium and stability, greatly facilitated in this case by the notion of potential, shows that no substantial improvement in space charge limitations can be expected and that the required variations between the flux linking the electron orbits and the magnetic and electric fields at the orbits are difficult to realize on account of their complexity and narrow tolerances. The X-ray output of a small experimental double yoke instrument was measured by a phototube multiplier viewing an irradiated fluorescent screen and gave evidence of multiple group electron capture.

PART I

INTRODUCTION

THE current widespread development of the betatron as a generator of high energy electrons and gamma rays has created need for a comparatively detailed knowledge of the electron dynamics of induction acceleration, both for the sake of engineering design and for possible future modifications and improvements. New points of view worked out for this purpose are applicable to other electronic devices such as the magnetron, or various electron optical lens systems of rotational symmetry. It is interesting to note that in the betatron are the longest, longest lived, and most stable, closed electron orbits known outside of the atoms themselves.

The idea of accelerating electrons by the solenoidal electric field of magnetic induction has been considered by many investigators (1, 2, 3,

---

* Decimal Classification: R138.
† Reprinted from *Jour. Frank. Inst.*, April, May, 1947.

6, 31).* In order for large energies to be gained in this manner, the acceleration must take place over very long electron paths, and while the magnetic field producing the electric field may simultaneously be used to bend the paths back on themselves, the elements of stability of these paths is of primary importance to successful operation. The basic principles concerning this stability were recognized by Steenbeck (17, 32), who published a design embodying most of the essential features of present apparatus. These and other properties of the orbits, especially the mechanisms of electron injection, were discussed by Kerst and Serber (10), and the successful operation of a small betatron was reported by Kerst in 1940 (8).

The betatron or induction accelerator consists in its essentials of an alternating current magnet, generally with a laminated iron core[1] and an air gap, and within the gap an evacuated chamber containing an electron injector and a target. The injector, an electron gun, directs a beam of electrons circumferentially at a phase of the cycle of the magnet when the magnetic field is weak but increasing. The electrons, momentarily of moderate energies, are turned by the weak magnetic field into an approximately circular path while an electric field is induced along that path by the increasing flux of the magnet, a field which in ordinary transformers is responsible for the electromotive force induced in the windings. Therefore, as the electrons revolve in their orbits, they gain energy according to the number of turns they make. Meanwhile, the magnetic field increases too, so that in spite of the increased energy of the particles, they are held close to their original paths. Evidently, since the energy gain is determined by the average field intensity inside the circular path, while the field just at the circle keeps the electrons there, some definite relation between the two must exist. It turns out that a satisfactory relation is for the average field to be just twice the orbital field. Finally when the electrons have gained an energy corresponding to a change of flux of almost a quarter cycle, this two-to-one condition is artificially upset and the orbit moves toward the target. This is bombarded by the electrons and high energy gamma rays are produced. The gamma rays leave the chamber in the forward direction in a fairly well defined beam, but while some high speed electrons scattered by the target are capable of leaving the chamber, there is at present no published description (38) of a method for extracting the bulk of the electrons in a well defined beam.

---

* All such references will be found under heading "References" on page 103.

[1] Higher frequency air core systems may have desirable characteristics but have not yet been developed successfully (1, 22).

It is intended here to develop the equations of electron motion as they are applied to induction acceleration, and with the aid of an artificial potential function to encompass the earlier work on the electron trajectories and the conditions for stability, which constitute first order calculations for the most part, considerations on wider departures from equilibrium, particularly during injection, and the general influence of space charge. The application of an auxiliary electric field of rotational symmetry is discussed in detail without materially complicating the equations. Some experimental work complementing the theory is reported.

EQUATIONS OF MOTION

The pertinent equation for electron motion in an electromagnetic field is given by

$$\frac{d}{dt}\{m\vec{v}\} = -e\,\{\vec{E} + \vec{v} \times \vec{B}\}, \tag{1}$$

wherein

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}} \text{ and } \frac{e}{m_0} = 1.76 \times 10^{11} \text{ coulombs per kilogram} \tag{2}$$

Here it is presumed at once that effects peculiar to quantum mechanical calculations are of no significance to the motion. Also, it is assumed that the acceleration of the electron is sufficiently small that associated force terms which appear in the general equation of motion are quite negligible. This assumption is substantially correct for all but the highest attainable accelerations for very high energy particles, a case which will be commented upon separately. The component equations of motion from (1), expressed in the conventional coordinates of the circular cylinder, are

$$\frac{d}{dt}\big[m\dot{r}\big] - mr\dot{\theta}^2 = -e\big[E_r + r\dot{\theta}B_z - \dot{z}B_\theta\big], \tag{3-r}$$

$$\frac{d}{dt}\big[mr^2\dot{\theta}\big] = -er\big[E_\theta + \dot{z}B_r - \dot{r}B_z\big], \tag{3-\theta}$$

$$\frac{d}{dt}\big[m\dot{z}\big] = -e\big[E_z + \dot{r}B_\theta - r\dot{\theta}B_r\big]. \tag{3-z}$$

The electric and magnetic field components expressed in these equations are of course related by the Maxwell electromagnetic field equations, and even though the sources of the field, i.e., charges and currents, may be of arbitrary distribution, two restrictive equations must be obeyed, namely:

$$\operatorname{curl} \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \quad \text{and} \quad \operatorname{div} \vec{B} = 0. \tag{4}$$

In the ideal betatron, the electric field is assumed to be generated solely by induction from the changing magnetic field, and it is without sources such as charge or polarized materials. To the lines of the magnetic field are ascribed the symmetries of an axis of revolution and a plane of reflection. With the coordinate axis and the plane of the origin coinciding with these two respectively, it is also assumed that the angular component of the magnetic field, B_θ, and the angular variation of the axial component, $\partial B_z/\partial \theta$, are everywhere zero. From these hypotheses, it can be shown through equations (4) that of the field components only three remain, B_z which is a function of r, z, and t, and B_r and E_θ, which are related to B_z by the equations

$$rB_r = -\int_0^r r \frac{\partial B_z}{\partial z} dr \quad \text{and} \quad rE_\theta = -\int_0^r r \frac{\partial B_z}{\partial t} dr. \tag{5}$$

In vacuum, the remaining Maxwell field equation requires of B_z that

$$-\frac{1}{r}\int_0^r r \frac{\partial^2 B_z}{\partial z^2} dr - \frac{\partial B_z}{\partial r} + \frac{1}{c^2}\frac{1}{r}\int_0^r r \frac{\partial^2 B_z}{\partial t^2} dr = 0. \tag{6}$$

In the construction of an actual accelerator, the symmetry of the magnetic field is unavoidably imperfect and dielectrics and conductors associated with the vacuum chamber disturb the induced electric field. The error made in the forthcoming analysis by neglecting the small forces arising from the angular dissymmetries which may be present, depends upon whether the principal forces are in unstable equilibrium and upon whether the small effects are likely to be cumulative. Some tentative judgment on these questions will become possible as a result of the analysis of the ideal case.

With the assumption of complete symmetry in the fields as stated, equations (3) reduce, with the aid of equations (5), to a comparatively simpler set.

$$\frac{d}{dt}\left[m\dot{r}\right] = mr\dot{\theta}^2 - er\dot{\theta}B_z, \tag{7-r}$$

$$\frac{d}{dt}\left[m\dot{z}\right] = -e\dot{\theta}\int_0^r r \frac{\partial B_z}{\partial z} dr, \tag{7-z}$$

$$\frac{d}{dt}\left[mr^2\dot{\theta}\right] = e\left[\int_0^r r \frac{\partial B_z}{\partial t} dr + \dot{z}\int_0^r r \frac{\partial B_z}{\partial z} dr + \dot{r}B_z\right]$$

$$= e \frac{d}{dt} \int_0^r r\, B_z dr. \tag{7-\theta}$$

Recalling the functional dependence of B_z, it is noted that

$$2\pi \int_0^r r B_z(r, z, t) dr$$

is the magnetic flux passing through the circle r, z. The integration of equation ($7\text{-}\theta$) calls for a constant, C, which describes some of the initial circumstances upon injection of the electron into the field.

$$C = \frac{m^0}{e} r_0^2 \dot{\theta}_0 - \int_0^r r B_z(r, z_0, t_0) dr, \tag{8}$$

wherein the superscript and subscripts 0 denote the initial values of the quantities. Thus the integral of equation ($7\text{-}\theta$) is

$$r\dot{\theta} = \frac{e}{m} \frac{\int_0^r r B_z dr + C}{r}. \tag{9}$$

If this result is placed in equations ($7\text{-}r$) and ($7\text{-}z$), it is found that

$$\frac{d}{dt} [m\dot{r}] = -\frac{e^2}{2m} \frac{\partial}{\partial r} \left[\frac{\int_0^r r B_z dr + C}{r} \right]^2, \tag{10-r}$$

$$\frac{d}{dt} [m\dot{z}] = -\frac{e^2}{2m} \frac{\partial}{\partial z} \left[\frac{\int_0^r r B_z dr + C}{r} \right]^2. \tag{10-z}$$

Were it not for the variance of the mass, m, the form of equations (10) would be strongly suggestive of two-dimensional potential motion. However, the mass, or the total energy of the particle, can change only through motion along the one existing component of the electric field, E_θ. Thus from equations (5) and (9), the mass-energy equivalence gives for the total variation of mass:

$$\frac{d}{dt} [mc^2] = -er\dot{\theta}E_\theta = -\frac{e^2}{2m} \frac{\partial}{\partial t} \left[\frac{\int_0^r r B_z dr + C}{r} \right]^2 \tag{11}$$

and the value of m is given directly by combining equation (9) with the formula for the mass, (2),

$$m^2c^2 = \frac{e^2 \left[\dfrac{\int_0^r rB_z dr + C}{r} \right]^2 + m_0^2 c^2}{1 - \dfrac{\dot{r}^2 + \dot{z}^2}{c^2}}.$$ (12)

the particle's kinetic energy is thus:

$$V = \frac{(m - m_0)c^2}{e} = \frac{m_0 c^2}{e} \left\{ \sqrt{\frac{\dfrac{e}{m_0} \left[\dfrac{\int_0^r rB_z dr + C}{r} \right]^2}{\dfrac{m_0 c^2}{e} \left[1 - \dfrac{\dot{r}^2 + \dot{z}^2}{c^2} \right]} + 1} - 1 \right\} \cdot$$ (13)

In virtually all operable betatrons, the number of "volts per turn" is very small in comparison to the total energy of the electron, starting at $V_0 = .512$ Mev rest energy, so that in the course of a few revolutions the proportional change in m as it appears in equations (10) is very small. The dependence of this change on the radial and axial trajectory as in equation (11) is an order of magnitude smaller still and may be neglected altogether.

The differentiation on the left of equations (10) leads to the terms $r(dm/dt)$ and $\dot{z}(dm/dt)$ having the character of viscous damping terms, because the increasing mass of the accelerated electron tends to bring about a reduction in the radial and axial velocities of its motion. Again for conventional instruments having a small number of "volts per turn," and this includes practically all low frequency magnets, these terms are very small and for most purposes may be neglected. During the early, low energy stages of the acceleration, where damping effects are of particular interest, there are other very much larger ones to be discussed, which arise through the changing of the magnetic field. Although the initially larger terms tend to vanish at relativistic speeds and leave the mass variation terms as the only damping, the latter will be ignored in most of this discussion.

Under the above provisions, equations (10) can be written exactly as if describing two-dimensional potential motion.

$$m_0 \ddot{r} = - e \frac{\partial}{\partial r} \left\{ \frac{em_0}{2m^2} \left[\frac{\int_0^r rB_z dr + C}{r} \right]^2 \right\} = - e \frac{\partial}{\partial r} V_M,$$ (14-r)

$$m_0\ddot{z} = -e\frac{\partial}{\partial z}\left\{\frac{em_0}{2m^2}\left[\frac{\int_0^r rB_z dr + C}{r}\right]^2\right\} = -e\frac{\partial}{\partial z}V_M, \quad (14\text{-}z)$$

wherein the symbol V_M, the "potential of magnetic and inertial forces," is introduced for brevity. Also for convenience let the symbol V_{M0} replace $(m^2/m_0^2)V_M$. With this notation, an approximation to equation (13), useful for electrons whose tangential component of velocity is the principal one, is

$$V \cong V_0\left[\sqrt{\frac{2V_{M0}}{V_0} + 1} - 1\right], \quad (15)$$

the limiting forms of which are, for low energies, V_{M0}, and for very high energies, $\sqrt{2V_0 V_{M0}}$.

If, in addition to the magnetic and induction fields, an electrostatic potential field V_E having the same symmetries as the magnetic field is applied to the region of the electron motion, and thus superposed on the fields which have already been considered, the angular motion equations (7-θ) and (9) remain unaltered, while to the right sides of equations (10) for the radial and axial motions are added the appropriate forces obtained from the gradient of V_E, namely, $e(\partial V_E/\partial r)$ and $e(\partial V_E/\partial z)$. In this case, however, the change in mass does not occur only through motion in the tangential or angular direction, so that in order to achieve the same simplicity as in equations (14), the new equations must be restricted to the non-relativistic region. For most purposes, this is not a severe restriction, since electrostatic potentials large enough to materially influence the motion of the higher speed electrons are hard to apply. In conventional accelerators, the additional electric field is due primarily to the space charge of the electrons in the chamber and is of importance only in the early stages of acceleration. The equations of motion corresponding to equations (14) now have the form:

$$m_0\ddot{r} = -e\frac{\partial}{\partial r}\left[V_M - V_E\right], \quad m_0\ddot{z} = -e\frac{\partial}{\partial z}\left[V_M - V_E\right]. \quad (16)$$

THE TWO DIMENSIONAL MOTION

The approximate equivalence of the radial and axial electron motion to two-dimensional classical motion in a potential field slowly changing with time offers important advantages, both for the analytical and the intuitive understanding of the processes involved. As has already been

stated, the total electron path is very long, and while it is turned back upon itself, the region in r and z which it may occupy, must of course be within the vacuum chamber. To thus confine the electron, there must be, in this region, a minimum in the potential surface V_M, or $V_M - V_E$ if the additional electric field is present, and the two-dimensional analog of the particle's kinetic energy must be insufficient to carry it beyond the potential barrier surrounding this minimum which corresponds to a stable circular orbit. With the particle kept in this bowl, the necessary properties of the trajectory are obtained. For purposes of discussion, these properties may be subdivided according to whether they depend on the momentary existence and shape of the bowl or whether they pertain to the effects of the change of this shape with time.

The elementary conditions for the existence of an instantaneous stable circular orbit, an "instantaneous equilibrium orbit" of radius r_i, may be found directly by applying the conventional relations of analytic geometry to the function $[V_M - V_E]$.

(a) $\dfrac{\partial}{\partial r}[V_M - V_E] = 0$

(b) $\dfrac{\partial^2}{\partial r^2}[V_M - V_E] > 0$

(c) $\dfrac{\partial^2}{\partial z^2}[V_M - V_E] > 0$

(d) $\dfrac{\partial}{\partial z}[V_M - V_E] = 0$

(e) $\dfrac{\partial^2[V_M - V_E]}{\partial r^2}\dfrac{\partial^2[V_M - V_E]}{\partial z^2} - \left\{\dfrac{\partial^2[V_M - V_E]}{\partial r \partial z}\right\}^2 > 0.$ (17)

On account of the symmetry which the electric and magnetic fields possess with respect to the plane of $z = 0$, the fourth condition is satisfied everywhere on this "equilibrium orbit plane." It then follows that on this plane the last condition is satisfied if the second and third are, so that only the discussions of the first three conditions, called the radial equilibrium, the radial and the axial focusing, conditions, respectively, are not trivial. If V_E is taken to be absent, the radial equilibrium condition applied to V_M gives the alternative relations

$$\frac{\int_0^r rB_z dr + C}{r} = 0, \qquad (18)$$

$$B_z - \frac{\int_0^r rB_z dr + C}{r^2} = 0. \qquad (19)$$

The first of these refers to a condition for the electron at rest, see equation (9), and is of no interest here. The second alternative is the appropriate condition for equilibrium in the radial direction. If it is included in the formulas when the radial and the axial focusing conditions are applied to V_M, the following two inequalities are obtained, respectively (32):

$$\frac{r}{B_z} \frac{\partial B_z}{\partial r} > -1, \qquad (20)$$

$$B_z \int_0^r r \frac{\partial^2 B_z}{\partial z^2} dr > 0, \qquad (21)$$

wherein $\partial B_z/\partial z$ is zero for $z = 0$. Taking into account the quasi-static nature of the magnetic fields with which one must deal in practice, the time derivative term in equation (6) may be neglected and the result used for substitution in relation (21) which may be written with the same functional expression as (20). Combining the two:

$$1 > \left[-\frac{r}{B_z} \frac{\partial B_z}{\partial r} \right] > 0. \qquad (22)$$

This means simply that the lines of the magnetic field must be convex outward and that the field intensity diminishes with increasing radius, while the corresponding diminishing in centripetal magnetic force must occur more slowly than the decrease in centrifugal force at constant electron velocity. The quantity in brackets is a more general form of the "n" of Kerst and Serber who assumed that the magnetic field was proportional to r^{-n}. During the acceleration cycle the importance of the constant C becomes less and less relative to the time increasing $\int_0^r rB_z dr$. Equation (19) approaches the condition,

$$B_z = \frac{\int_0^r rB_z dr}{r^2}, \qquad (23)$$

wherein the right side is obviously $\frac{1}{2}$ the average value of B_z within the circle. To have a stable orbit, condition (23), together with condition (22), must be satisfied for some value of r. In conventional instruments, the magnetic field intensities at all points vary synchronously throughout the greater portion of the acceleration cycle, and so this value of r, called the principal equilibrium orbit, is a constant, r_c. Then, as may be noted by applying the present conditions to the formula from (5), this position is also the minimum for E_θ or the equivalent, volts per turn divided by radius. Kerst and others have used this fact for measuring the position of the equilibrium orbit.

Induction accelerators are usually built so that condition (21) is satisfied for all positions within the vacuum chamber, and since the equilibrium orbit plane $z = 0$ is one of symmetry, the axial sections of the potential surfaces V_M have substantially similar characteristics whatever the other circumstances may be. They are parabola-like in shape, symmetrical and increasing outward monotonically on either side of $z = 0$. The rate of the increase depends on how negative is $\dfrac{r}{B_z}\dfrac{\partial B_z}{\partial r}$, and the adjustment of this parameter determines whether the walls of the potential bowl, in the axial direction, are sufficient to keep the electrons from the corresponding material walls or other obstructions of the vacuum chamber. While the radial sections of the potential surfaces are of no greater intrinsic importance than the axial ones, the appearance of the constant C in the formula for V_M makes their characteristics critically dependent on the initial conditions. For each value of C, the radial sections are, for the most part, of the same general shape for different values of z. A representative set of such radial sections of V_M for different values of C for the plane $z = 0$, and which thus includes the equilibrium orbit position, is given in Figure 2. The instantaneous magnetic field distribution from which these curves were computed is shown in Figure 1, which includes an outline of the pole pieces generating this field.

A large variety of injection phenomena can be understood by consideration of the V_M radial section curves only, since usually the axial sections may be taken for granted. While the curves of Figure 2 are entirely representative of all values of C, the particular ones drawn were chosen with the position $r = 18$ in mind as the initial location of the electrons. In such an instance, these particular values of C correspond to tangential velocities of injection (neglecting relativity corrections) differing by a simple factor, one plus the coefficient of P_{18}, from the velocity of the "normal" electrons, whose C value is zero and hence
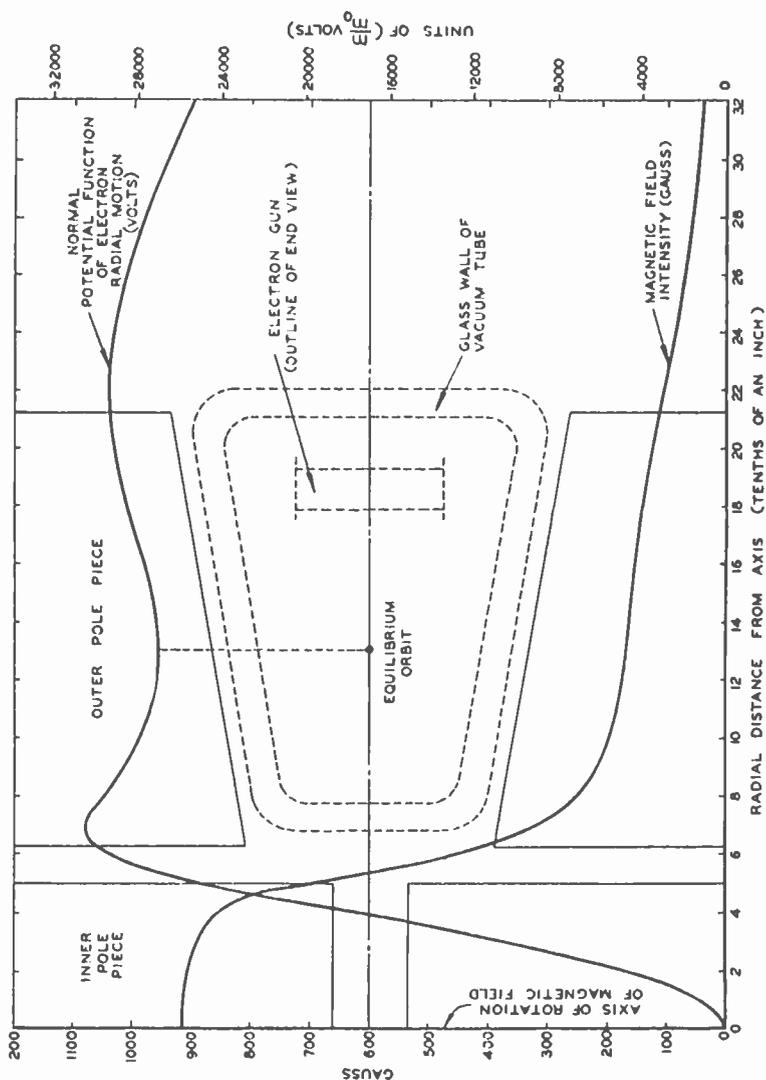
Fig. 1—Radial Variation of magnetic field.

whose equilibrium orbit radius is unchanged during the acceleration. With reference to the concept of two-dimensional conservative motion, it is a matter of inspection of Figure 2, to see that for electrons injected at $r = 18$, even for those without radial and axial velocities, only those with tangential velocities corresponding to C values within the interval $-.020 P_{18}$ to $+.125 P_{18}$, approximately, can be captured initially within the potential bowls. For those with small radial velocities in addition, the range is evidently more restricted and can be found immediately by
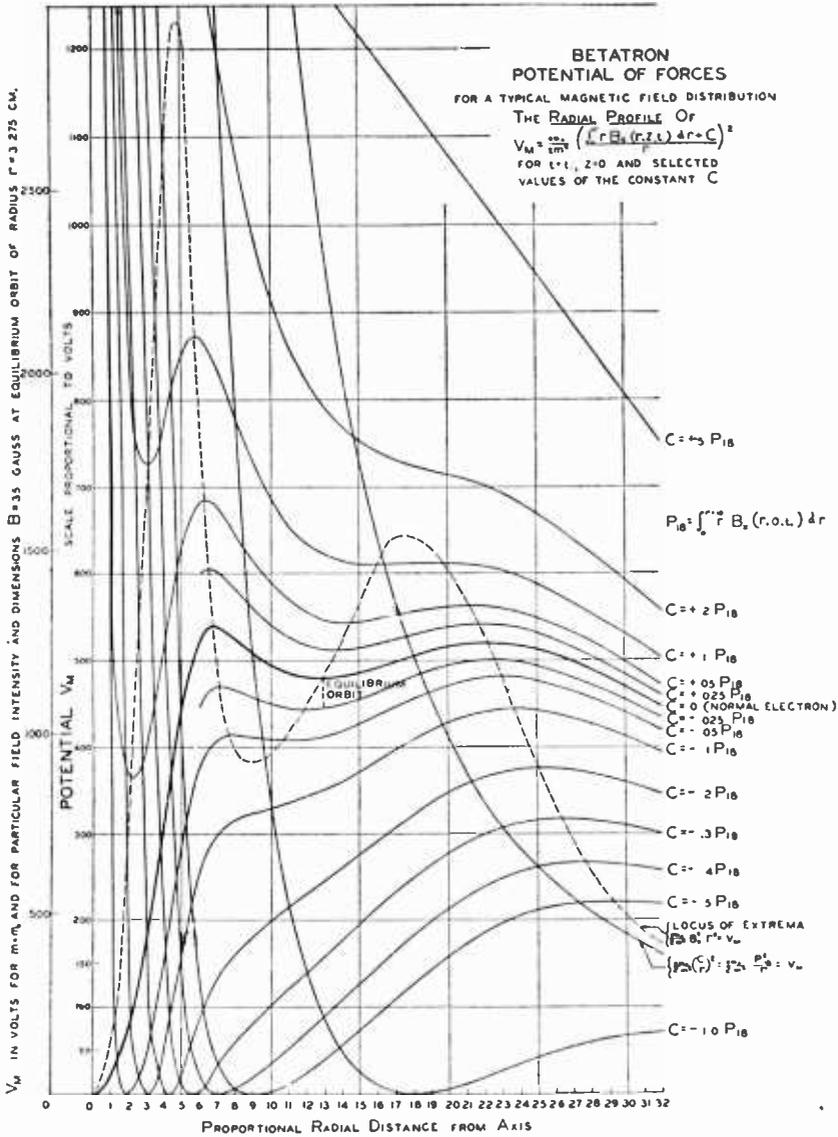
Fig. 2—The radial profiles of the potential of magnetic and inertial forces.

recognizing the additional barrier necessary to contain the additional "two-dimensional kinetic energy." While the potential minimum for which V_M is zero has been excluded from consideration on the conventional betatron because the associated acceleration is small, it repre-

sents the situation of zero tangential velocity obtained when electrons are emitted directly into the chamber from a hot cathode located at the zero position, such as $r = 18$ for $C = 1.0\,P_{18}$. With the addition of the electrostatic field V_E from a concentric anode, this case describes a static magnetron with a non-uniform magnetic field and is considered in detail below.

If an electron in a conventional betatron is confined near to a stable circular orbit in accordance with the existence of a minimum in the appropriate surface V_M, there are associated with its motion three frequencies or periods which are more or less unique, depending on whether the radial and axial oscillations are of small or large amplitude. In the limit of very small such amplitudes the periods of rotation and radial, and axial oscillation are, respectively:

$$T_\theta = 2\pi \left[\frac{e}{m} B_z \right]^{-1}, \qquad\qquad \text{at } r = r_i \qquad (24)$$

$$T_r = 2\pi \left[\frac{e}{m} B_z \right]^{-1} \left[1 + \frac{r}{B_z}\frac{\partial B_z}{\partial r} \right]^{-\frac{1}{2}}, \quad \text{at } r = r_i \qquad (25)$$

$$T_z = 2\pi \left[\frac{e}{m} B_z \right]^{-1} \left[-\frac{r}{B_z}\frac{\partial B_z}{\partial r} \right]^{-\frac{1}{2}}. \quad \text{at } r = r_i \qquad (26)$$

Provided the particle remains within the potential bowl for large amplitudes, only the formula for T_r needs major correction: to be multiplied by a factor much greater than unity because the particle approaches the rounded crests of the surrounding potential barrier. In any event, since $\left. \dfrac{r}{B_z}\dfrac{\partial B_z}{\partial r}\right|_{r=ri}$ obeys condition (22), the periods T_r and T_z are always greater than T_θ.

Very small angular dissymmetries of magnetic and induced electric field strengths may be present in an accelerator. If it is presumed that the electron motions are very nearly those of the ideally symmetrical case, the potential notion may still be applied, and the small dissymmetries can be accounted for by small, additional radial and axial forces, which recur with a period of T_θ. Since T_r and T_z are greater than T_θ, it follows that no direct resonance effect is likely to occur and that the disturbances on the ideal motion are not likely to accumulate beyond what are produced in the first few revolutions after injection. Thus, if the electron is well contained in the potential bowl in the first place, the motion will be in accordance with the formulas developed for symmetrical fields.

SPACE CHARGE

Whether an electrostatic field is applied from electrodes or not, a large space charge will contribute to V_E and should be taken into account. Since the charge distribution is not known and in general will depend on the manner of injection, the stability conditions (17), which still must be satisfied, are not capable of direct verification. They lead, however, to a definite limitation on the supportable space charge density, for by summing the first three of conditions (17), [(17a) being multiplied by $1/r$], one finds that the Laplacian of $[V_M - V_E]$ must exceed zero. Since V_E itself must obey Poisson's equation, it follows that:

$$-\frac{\rho}{\epsilon_0} < \nabla^2 V_N = \frac{em_0}{m^2} \left\{ \left[B_z - \frac{\int_0^r rB_z dr + C}{r^2} \right]^2 \right.$$
$$\left. + \left[\frac{\int_0^r rB_z dr + C}{r^2} \right]^2 + \left[\frac{\int_0^r r\frac{\partial B_z}{\partial z} dr}{r} \right]^2 \right\}. \tag{27}$$

The limiting charge density given by this relation may be found in charge distributions in complete axial and radial equilibrium, wherein the repulsive forces between particles are just compensated by the retaining forces of the applied fields. Consideration of an optimum of such distributions will establish the upper limit of charge supporable in the betatron, whatever the injection mechanism. The magnetic forces of the space charge current are assumed negligible at the time of injection.

Supposing at first that all electrons present have a common value for the integration constant C, the surface of $[V_M - V_E]$ which originally possessed a minimum before the electrons were inserted and space charge forces contributed to V_E, now has the shape of a bowl whose bottom is flat over the region of the charge distribution. Here the charge density is given by the right side of relation (27) and as more charge is added, this value is not exceeded but the distribution extends itself over a larger space. Eventually, with additional charge, the repulsive forces can no longer be counterbalanced. This is shown in the surrounding potential barrier, where, at least at one point on the edge of the flat bottom, the barrier is pressed down, as the surface curvature changes from concave upward to concave downward. Then no further charge can be supported for it spills out by this route. With the usual field shapes, this "leak" will occur in a radial direction in the $z = 0$ plane provided the vacuum chamber walls and other obstructions have

not interfered beforehand, because the potential surface goes to relatively greater height in the axial direction. It seems fairly certain that in this case of equilibrium, there is supported the greatest possible number of electrons with the chosen value of C. The configuration is one of minimum total energy as compared to non-equilibrium distributions of the same amount of charge, and these distributions in sufficient time must surely rearrange themselves so that some of the particles will surmount the surrounding barrier by virtue of their extra energy, even if, as is unlikely, the initial arrangements succeed in reducing the space charge repulsions relative to the retaining forces.

If groups of electrons of different values of the integration constant C are present and in complete equilibrium (without orbital oscillations), they cannot mingle in the same spatial region because these values appear explicitly in the radial equilibrium condition (17a) which cannot be satisfied for the same space coordinates but different C's. It follows from the continuous dependence of the equilibrium relations on C that although one group may be space charge limited with its potential bowl depressed at some point, in general another group may be found of slightly higher or lower C whose potential surface $[V_M - V_E]$ still has a minimum and can receive more charge. Assuming the injector is able to provide it, the additional space charge will cause some of the previous group to be lost, but the equilibrium position of the new group, as given by (17a) with the new value of C, is necessarily well removed from the position of the "leak" in the old group. Then on members of the old group at this position, the proportional repulsion, or force per unit charge, of the new group is less by virtue of greater distance than was that of the displaced charge, and there is a net gain in the total charge supported. If the injection mechanism is able to meet the demand, this displacement of groups may go on over a continuous range of C values until that group is reached which simultaneously depresses its potential barrier in two places of different radii, both radially outward and inward for conventional instruments. Now no further group of slightly different C can have a minimum in its potential surface and the space charge limited condition of the last named group, in equilibrium and standing alone, is the final configuration.

For the above reasons, it is thought that the optimum space charge saturated condition of the betatron, which can be obtained at the time of injection, consists of the support in radial and axial equilibrium of the full complement of one group of electrons of a single value of C, or in most cases of a single tangential velocity and time of injection. Since the influence of the space charge field depends not only on the

charge distribution but also on the electrostatic boundary conditions which must be satisfied, the exact computation of the value of the integration constant C which applies to the select electron group is difficult, if not impossible. Fortunately, there is a wide latitude in the choice of this C which will lead to substantially the same total for the supported space charge, so that one may assume a value corresponding in the case of a conventional betatron to the radial profile curve of V_M whose two maxima are of the same height, on the grounds that there will be a rough symmetry about the minimum position. In Figure 2, this value is about $C = - .008 \, P_{18}$. It should be emphasized, however, that the injection mechanism may be quite unable to establish this optimum electron group and it may be necessary to consider another group of very different value of C, which can support much less charge.

To find the upper limit of the capturable charge itself, the space charge density given by equation (27) with a proper choice of C, must be integrated within the limits of the charge distribution. In estimating these limits, electrostatic boundary conditions must be taken into account. If the induced space charge image fields are assumed to be negligible, the radial curvature of the negative space charge field alone will be negative throughout the distribution. Thus, by radial focusing condition (17b), the charge distribution cannot extend beyond the points of inflection of the radial profile curve of V_M and probably not quite as far. A fair guess of the limits can be made by taking the lower inflection point for one limit and its horizontal projection to the other side of the bowl of V_M for the other. In Figure 2, these limits are, for $C = .008 \, P_{18}$, $r_2 = 16$ and $r_1 = 9$, and for $C = + .1 \, P_{18}$, $r_2 = 16.5$ and $r_1 = 15.5$. The relative axial to radial extents of the charge distribution is determined to a first approximation by the square root of the inverse ratio of curvatures of V_M. Evaluated at the instantaneous equilibrium orbit $r = r_i$, these values are

$$\frac{\partial^2 V_M}{\partial z^2} = - \frac{em_0}{m^2} B_z{}^2 \frac{r}{B_z} \frac{\partial B_z}{\partial r}, \quad \text{at } r = r_i \tag{28-z}$$

$$\frac{\partial^2 V_M}{\partial r^2} = \frac{em_0}{m^2} B_z{}^2 \left[1 + \frac{r}{B_z} \frac{\partial B_z}{\partial r} \right]. \quad \text{at } r = r_i \tag{28-r}$$

The experimental adjustment of the parameter $\dfrac{r}{B_z} \dfrac{\partial B_z}{\partial r}$ to from $-\frac{3}{4}$ to $-\frac{1}{2}$ in accordance with condition (22) so as to have the electron configuration roughly conform with the shape of the vacuum chamber

leads in present instruments to an axial extent from $\frac{1}{2}$ to 1 times the radial extent for values of r_i close to the main equilibrium orbit, but very much smaller ratios for other electron groups. The radial-axial distribution having an approximately elliptical boundary, one finds the total charge, Q, supported in this toroidal ring is the product of the volume and the average charge density of (27) which is roughly given by its value at r_i.

$$Q = \pi^2 \epsilon_0 r \left\{ \frac{[r_2 - r_1]^2}{r^2} \frac{\left[1 + \frac{r}{B_z} \frac{\partial B_z}{\partial r} \right]^{\frac{1}{2}}}{\left[-\frac{r}{B_z} \frac{\partial B_z}{\partial r} \right]^{\frac{1}{2}}} \right\} \left\{ \frac{e m_0}{2 m^2} \left[\frac{\int_0^r r B_z dr + C}{r} \right]^2 \right\} \Bigg|_{r=r_i} \quad (29)$$

For low energies of injection, the second quantity in braces is by equation (15), the voltage of injection. The first braced quantity is a geometrical factor which depends both on the field shape of the betatron and the choice of C. From Figures 1 and 2, the value of this factor for the optimum C of $-.008 P_{18}$ is approximately $\frac{1}{4}$ but for a C of $+.1 P_{18}$ it is about 1/100. Excepting for the geometrical factor, equation (29) is the same as that of Kerst (9).

The maximum supportable charge, and hence the maximum space charge limited output of the betatron, is, other things being equal, directly proportional to the injection voltage. It has been quite generally observed that the output increases with injection voltage, but the total output has been of the order of ten per cent, or less, of that computed by the old formula or what is practically the same thing, formula (29) with the optimum geometrical factor. The injection voltages currently in use seem too high for electron scattering by residual gas to enter into this effect.

The implication of these experimental results appears to be that while the betatron may at injection approach a space charge limited condition, the injected electron groups which can be captured in stable orbits are those of tangential velocities or C values departing considerably from the optimum.

MOTION IN THE POTENTIAL OF FORCES

The increase of the magnetic field with time brings about the acceleration of the electrons and the change in magnitude and shape of

the appropriate potential of force $[V_M - V_E]$. While the energy given by equation (15) in combination with (19) increases at first with the square of $r B_z$ at the equilibrium orbit, and for relativistic velocities directly with $r B_z$, the radial and axial oscillations about the orbit are altered by the change in the potential surface. The existence of a minimum determines the possibility of stable electron orbits, but during the acceleration period the limits and velocities of the oscillations are changed. On this account, the electrons might alternatively strike the walls of the vacuum chamber and injection apparatus or be focused into a narrow beam.

The radial and axial motion constitutes two-dimensional motion in a potential field changing with time, but most of the essential characteristics can be found in a corresponding one-dimensional motion. This is particularly true in the conventional betatron because it can be shown by a Taylor expansion of the potential V_M near the main equilibrium orbit that the motions in the r and z directions are independent up to fourth order terms. The potential surface changes but little in the period of one oscillation of the particle. An approximation in the analysis will be made based on the smallness of the change in energy per oscillation as compared to the energy itself, namely, an adiabatic approximation.

Consider at first the equation of a constant mass point in a potential of forces W, of the form $W(x, t) = G(x)F(t)$ where $G(x)$ is a non-negative function of space only and $F(t)$ a slowly varying function of time only:

$$\ddot{x} = -\frac{\partial W}{\partial x}. \tag{30}$$

Without ultimate loss in generality, it may be assumed for simplicity that $G(0) = 0$. An energy integral can be obtained by integrating both sides of equation (30) multiplied by $\dot{x}$:

$$\tfrac{1}{2}(\dot{x}^2 - \dot{x}_0^2) = -W + W(x_0, t_0) + \int_{t_0}^{t} \frac{\partial W}{\partial t} dt, \tag{31}$$

where $x = x_0$ and $\dot{x} = \dot{x}_0$ at $t = t_0$. The value of the integral in this equation can be estimated by the law of the mean since $G > 0$.

$$\int_{t_0}^{t} \frac{\partial W}{\partial t} dt = \int_{t_0}^{t} G \frac{dF}{dt} dt = \left[\frac{1}{F}\frac{dF}{dt}\right]_{t_1} \int_{t_0}^{t} W(x, t)dt,$$
$$\text{where } t_0 < t_1 < t. \tag{32}$$

By substituting W from equation (31) into the last integral, $\int_{t_0}^{t}\frac{\partial W}{\partial t}\,dt$

can be written:

$$\int_{t_0}^{t}\frac{\partial W}{\partial t}\,dt = (1 + \epsilon)\left[\frac{1}{F}\frac{dF}{dt}\right]_{t_1}\int_{t_0}^{t}\left[W(x_0, t_0) + \frac{\dot{x}_0^2 - \dot{x}^2}{2}\right]dt, \quad (33)$$

where

$$|\epsilon| < \tfrac{1}{2}\left[\frac{1}{F}\frac{dF}{dt}\right]_{max}(t - t_0)\,\frac{W'_{max}}{W_0 + \dfrac{\dot{x}_0^2}{2} - \dfrac{1}{t - t_0}\displaystyle\int_{t_0}^{t}\frac{\dot{x}^2}{2}dt} \quad (34)$$

The value of $\dfrac{1}{F}\dfrac{dF}{dt}(t - t_0)$ is the proportional change of energy of oscillations per period and is assumed very small in the adiabatic approximation. Therefore, the value of ϵ is also small since the ratio of the maximum value of the potential in the interval to the difference between the initial energy and the average kinetic energy is of the order of unity. In the application of this idealized problem to the present day betatrons, ϵ is less than one per cent and will be neglected. By substituting into equation (31) the value from equation (33) where $\dfrac{1}{F}\dfrac{dF}{dt}$ at t_1 is assumed equal to its value at t_0, again according to the adiabatic approximation, the total energy of oscillation E of the mass point becomes:

$$E = W + \frac{\dot{x}^2}{2} = W(x_0, t_0) + \frac{\dot{x}_0^2}{2}$$
$$+ \frac{1}{F(t_0)}\frac{dF}{dt}\left[\left(W(x_0, t_0) + \frac{\dot{x}_0^2}{2}\right)(t - t_0) - \int_{t_0}^{t}\frac{\dot{x}^2}{2}dt\right] \quad (35)$$

Let us assume an oscillatory motion such that at $t = t_0$, $\dot{x} = 0$ and $x = a$ and again $\dot{x} = 0$ when $t = t_0 + T$ at which time $x = a + \Delta a$. The increase of energy ΔE during the period T is:

$$\Delta E = \frac{1}{F(t_0)}\frac{dF}{dt}\left[W(x_0, t_0)T - \int_{t_0}^{t_0+T}\frac{\dot{x}^2}{2}dt\right]$$
$$= \Delta W - \frac{\Delta W}{W(x_0, t_0)}\frac{1}{T}\int_{t_0}^{t_0+T}\frac{\dot{x}^2}{2}dt,$$
$$\Delta E = \Delta W - \delta E =$$

$$\Delta W \left(1 - \frac{\text{Average kinetic energy in one period}}{\text{Maximum kinetic energy in period}} \right) \quad (36)$$

where ΔW is the increase of W at the point $x = a$ and δE is what may be called the energy increment deficiency because it is the difference between the most energy which the mass point would gain were it immobilized at $x = a$ and its actual gain of energy while oscillating in the increasing potential trough.

The change of amplitude of oscillation, Δa, is evidently related to the energy increment deficiency by the approximate relation:

$$\Delta a = \frac{\delta E}{\left. \dfrac{\partial W}{\partial x} \right|_{x=a}} = - \frac{\text{Average kinetic energy}}{\text{Maximum kinetic energy}} \frac{\Delta W}{\left. \dfrac{\partial W}{\partial x} \right|_{x=a}}. \quad (37)$$

showing that it is proportional to the rate of change of the potential, the ration of the average to maximum kinetic energy and inversely proportional to the gradient of the potential at the turning point of zero kinetic energy.

In the case of harmonic motion in a parabolic potential, the average kinetic energy is half of the maximum and the gradient at the extremum is twice the average gradient from the minimum to the extremum, so that:

$$\frac{\Delta a}{a} = - \tfrac{1}{4} \frac{\Delta W}{W}. \quad (38)$$

These considerations can be applied to the actual case of the electron in the V_M potential surface by writing:

$$W = \frac{e}{m_0} \left[V_M(r, t) - V_M(r_e, t) \right]. \quad (39)$$

Relation (38) can then be written, for the normal electrons, $C = 0$, as:

$$\frac{\Delta a}{a} = - \tfrac{1}{4} \frac{\Delta V_M}{V_M} = - \tfrac{1}{2} \frac{\Delta B}{B}, \quad (40)$$

relations derived by Kerst and Serber (10). These relations are valid even when the terms involving dm/dt which were omitted in deriving equation (14) are taken into account. This relation (40) shows that the amplitude of oscillation is proportional to $B^{-\frac{1}{2}}$.

In the case of the parabolic potential, the integrated equation of motion itself of the mass point can be obtained and a closer estimate

can be made of the error due to the adiabatic approximation than was possible in the general derivation. Equation (30) takes the form $\ddot{x} = -x(\partial^2 W/\partial x^2)$ where $\partial^2 W/\partial x^2$ is a function of time only, and can be solved approximately by the substitution $x = u\,e^v$ which leads to:

$$x = A\left(\frac{\partial^2 W}{\partial x^2}\right)^{-\frac{1}{4}} \cos\left(\int_{t_0}^{t}\sqrt{\frac{\partial^2 W}{\partial x^2}}\,dt + \alpha\right),\qquad (41)$$

where the amplitude of oscillation is seen to vary as found before and the instantaneous angular frequency of oscillation is $\sqrt[2]{\dfrac{\partial^2 W}{\partial x^2}}$. A differential equation can now be written which is exactly satisfied by equation (41) whose form is $\ddot{x} + x\dfrac{\partial^2 W}{\partial x^2}(1 + \epsilon_1) = 0$. The proportional error ϵ_1 with respect to equation (30) can be computed:

$$\epsilon_1 = \left(\frac{\partial^2 W}{\partial x^2}\right)^{-2}\left[\frac{1}{4}\frac{\partial^2}{\partial t^2}\left(\frac{\partial^2 W}{\partial x^2}\right) - \frac{5}{16}\left(\frac{\partial^2 W}{\partial x^2}\right)^{-1}\left(\frac{\partial}{\partial t}\frac{\partial^2 W}{\partial x^2}\right)^2\right]. \qquad (42)$$

For the case of non-relativistic speeds of the corresponding electron motion, if $v = \dfrac{e}{m}B_c$ is the instantaneous Larmor frequency corresponding to the field B_c, this error is less than:

$$\epsilon_1 < \frac{\omega^2}{v^2}\frac{B^2_m}{B^2_e}\frac{1}{1 - n}\left(\tfrac{1}{2} + \tfrac{1}{4}\cos^2\omega t\right), \qquad (43)$$

assuming the magnetic field to vary sinusoidally as $B_e = B_m \sin \omega t$. The error is most at injection time but is extremely small, of the order of 10^{-10} to 10^{-8} for practical instruments, since it depends on the square of the ratio of the magnet and electron Larmor frequencies as mentioned by Kerst and Serber (10).

In general, the potential function V_M or $|V_M - V_E|$ cannot be represented as a product of separate space and time functions, particularly because C appears in V_M. The position of the minimum changes with time but the oscillations of the electrons will take place around the instantaneous position in a reference system moving with it, just as if the minimum position were stationary. This is because the rate of acceleration of that position is negligible in all practical cases. Since equation (19) implicitly expresses the radial position of this minimum, r_i, for the conventional betatron, it may be used to give r_i through differentiation.

$$\dot{r}_i = \frac{\dfrac{\displaystyle\int_0^r r \dfrac{\partial B_z}{\partial t}\, dr}{r^2} - \dfrac{\partial B_z}{\partial t}}{\dfrac{B_z}{r}\left[\dfrac{r}{B_z}\dfrac{\partial B_z}{\partial r} + 1\right]}. \quad \text{at } r = r_i. \tag{44}$$

Assuming the magnetic flux and field to increase synchronously, this equation may be approximated for very small values of the difference $r_i - r_e = x_i$ and written for a small time increment Δt.

$$\frac{\Delta x_i}{x_i} = -\left.\frac{\Delta B_z}{B_z}\right|_{r=r_e}. \tag{45}$$

For non-relativistic energies, this is equivalent by equation (15) to a relation obtained by Kerst and Serber (10).

$$\frac{\Delta x_i}{x_i} = -\tfrac{1}{2}\frac{\Delta V}{V}. \tag{46}$$

Thus the instantaneous equilibrium orbit shifts asymptotically toward the main equilibrium orbit so that the separation is at first inversely proportional to the magnetic field. This shift is in contrast to the damping of the amplitude of oscillation which is inversely proportional to the square root of the magnetic field.

An important departure of the electron motion from what has already been described may take place during the latter part of the acceleration period from force terms which have hitherto been neglected in the equations of motion. Iwanenko and Pomeranchuk (19) have pointed out that in consequence of the high centripetal acceleration of the electron in its orbit, the charged particle will radiate energy at a rate which for large values of its own kinetic energy (of the order of 100 Mev) may be a considerable proportion of the rate of energy gain from the induction field. The results of this will be first a change in the position of the equilibrium orbit and ultimately a limit on the attainable kinetic energy. This limitation can be offset, in principle, by building high energy accelerators with relatively smaller centripetal accelerations, that is, with smaller field strengths and larger orbital radii, or with shorter acceleration periods, that is, higher frequency (12). If the amount of energy radiated were known, the effects on the electron path could be computed from present formulas with the aid of equation (15) by introducing an artificial time variation to the constant C. However, it is at present not clear whether the radiation formulas

for a single electron are immediately applicable to the case of the multitude of circulating electrons in the betatron. References have been made in the literature to a paper by J. S. Schwinger (39) which is shortly to appear and which presumably will give a definitive treatment of the radiation problem. See also ref. 28.

INJECTION IN A PURELY MAGNETIC BETATRON

The betatron has been discussed thus far under the assumption of perfect symmetry of mechanical arrangement, and electrons were considered present with appropriate energies without regard to any actual injection mechanism. In the conventional magnetic betatron electrons are injected circumferentially from a gun of finite dimensions, which is necessarily located in the region of acceleration of the particles. The gun presents, therefore, a possible obstruction to their manifeld revolutions. In fact, it is found experimentally that the output of the betatron is very much less than the upper limit imposed by space charge.

Upon injection, the electrons oscillate, as has been shown, around instantaneous equilibrium orbits shifting radially in the plane $z = 0$, with axial and radial amplitudes which are damped as the energy increases. Since the extremum positions of oscillations recede in the radial direction on account of both the change in amplitude and the shift of equilibrium orbit, while axially only the change in amplitude is effective, it is more favorable to locate the center of the gun at some radial distance x_0 from the main equilibrium orbit, r_e rather than away from the plane $z = 0$. An electron will return to the gun, or to a region nearby where the electrostatic leakage field may cause it to strike the gun, at some time T after emission. This time depends on the relation between the period of revolution T_θ and the period of radial oscillation T_r, in connection with the angular and radial extents of the gun. For an exactly tangential direction of injection, a particular electron will succeed in missing the gun if the total recession of the path δr, in the times T of the first or any other subsequent passage very near the gun, is greater than the distance d between the point of injection and the edge of the region near the gun where collision can occur. If, at injection, the electron has a small radial velocity in addition, the situation is less favorable but still depends on the magnitude of the recession. This total recession δr is evidently the sum of the decrements $(\Delta r_i + \Delta a)$ occurring in time T, and is obtained from equations (37) and (45) where ΔV, the energy gain in this time, is approximately the product of the number of revolutions N and the energy gain per revolution, w "volts per turn."

If the radial section of the potential trough is approximated by a parabola, which is a consideration essentially the same as the first order treatment of Kerst and Serber (10), the total recession δr immediately after injection is given by equations (38) and (46).

$$\delta r = -\tfrac{1}{2}(x_{i0} + \tfrac{1}{2}a_0)\frac{Nw}{V_i}. \tag{47}$$

Since the sum of the initial instantaneous orbit position x_{i0} and the initial amplitude a_0 is the total initial distance x_0 from the main equilibrium orbit. the condition for avoiding the gun requires at least that

$$\tfrac{1}{2}Nwx_0 > V_i d. \tag{48}$$

The value of N can be estimated from the sinusoidal nature of the trajectory by considering the ratio T_r/T_θ from equations (24) and (25) in relation to the mechanical dimensions of the gun. For an experimental arrangement used by the present authors, the value of N varied from 2 to 6, depending on the injection constant C but always within the approximation that the potential functions are merely first order departures from the main equilibrium radius. In general practice it is found experimentally that much higher values of the product $V_i d$ than those of the above inequality (48) can be used. Indeed, the output of the betatron apparently increases with the voltage of injection V_i. It may be expected, however, that the actual non-parabolic nature of the potentials for some injection parameters will materially alter the factors contributing to the recession δr, these factors being the damping of the amplitude of oscillation Δa, the equilibrium orbit shift Δx_i, and the period T of consecutive passages very near the gun.

As has already been pointed out, the range of C values of electrons which can be captured in stable orbits depends first on the existence of a minimum in the potential function and second, that the walls of the bowl are everywhere as high as at the point of injection. In conventional practice, all electrons are injected at the same voltage throughout an interval of time, so that the range of C values comes about through differences in the time of injection, and hence in the value of the magnetic field integral appearing in C, equation (8).

For values of C toward the ends of this range of capturable electrons, the radial sections of the potential surfaces depart significantly from the parabolic approximation. When the gun is located further from the main equilibrium orbit than the extrema of the locus of extrema ($r_0 < 8.5$ and $r_0 > 17.5$ for the example of Figure 2), there is a

value C_1 of C for which the potential V_M has a maximum at the gun. The rate of change of the amplitude of oscillations of electrons injected at times just following that corresponding to C_1 (when $r_0 > 17.5$, or just preceding when $r_0 < 8.5$), is very large since it is inversely proportional to the value of $\partial V_M/\partial r$ at the injection point in accordance with equation (37). Since this rate becomes very large for injection times corresponding to C_1, the gun clearance may be improved by a large factor of the order of a hundred over that of the parabolic case, provided of course that the gun is so located that the electrons are not lost in their first swing across the equilibrium orbit on the side opposite to the gun. ($7 < r_0 < 8.5$ or $17.5 < r_0 < 22.5$, in the example of Figure 2.) The recession will depend on the product of the rate of change of amplitude by the time T of consecutive passages near the gun, and therefore on the number N of revolutions between dangerously close approaches. The number N varies with C and passes through small and large values depending on the ratio of T_r/T_θ. Hence there is the possibility of capturing electrons in several distinct groups corresponding to the large values of N. The capture of electrons through the large damping of amplitude will occur in a range of C values close to C_1 (smaller than C_1 for an external gun, greater for an internal one), which is a comparatively small part of the total range corresponding to potentials with a minimum and in general the potential trough will be shallower than the normal curve ($C = 0$) with hence less space charge capacity.

Consider the gun located at or near the limit at which the potential can have a maximum, that is, at or near one of the extrema of the locus of extrema of the potential function ($r_0 = 8.5$ and $r_0 = 17.5$ in Figure 2). At these points there will be a value C_2 of C for which the potential V_M will have an inflection point. Near such a point the rate of shifting of the instantaneous equilibrium orbit, which is inversely proportional to $\left(1 + \dfrac{r}{B_z}\dfrac{\partial B_z}{\partial r} \right)$ as shown by equation (44), becomes very large. Since it approaches infinite values at the point itself, here again the position of the electrons may recede in the time T, which is at least equal to T_θ, an amount δr greater than d even if the amplitude damping were not taken into account. The range of C values for which the rate of equilibrium orbit shift is sufficient is only a small part of the range for which equilibrium orbits exist. Furthermore, the potential troughs are very shallow and it is probable that the number of electrons captured in them approaches being space charge limited, even for small injecting currents.

Finally, consider the case of an injector located at a position such that for some value C_3 of C the initial potential curve has a maximum whose value is equal to that at the injector position but which is located on the opposite side of the potential minimum. It is apparent that for C values approaching C_3, greater for an external gun and smaller for an internal one, the period of oscillation T_r around the equilibrium orbit tends to infinity since the radial velocity of the electron near the maximum of V is very small. Consequently the time T becomes very large and for a finite rate of shifting of the equilibrium orbit given by equation (44), the recession δr of the electrons may be sufficient for clearing the gun, provided the shift of the orbit is away and not toward the gun. This occurs only for particular gun positions and C values (for example $r_0 = 8$, $C = -.004$) which are small compared to the total range having minima in the potential.

These considerations show that for high injection voltages the principal mechanism of capture is the damping of amplitude of oscillation when the potential function has a maximum near the gun. The cases of rapid shift of the equilibrium orbit and of long periods of radial oscillation may account for similar capture but are less probable mechanisms.

To estimate the interval of time over which electrons can be captured, consider a synchronously varying magnetic field, or one which can be expressed as the product of separate functions of space and time:

$$B(r, z, t) = b(r, z) \cdot f(t). \tag{49}$$

The potential V can then be written as:

$$V_M = \frac{em_0}{2m^2} f^2(t) \left[\frac{\int_0^r brdr + \dfrac{f(t_0)}{f(t)} D}{r} \right]^2, \tag{50}$$

where t_0 is the injection time and $D = C/f(t_0)$. This relation shows explicitly as has been implicitly assumed so far, that the scale factor of the potential surface after emission time increases in the ratio $f^2(t)/f^2(t_0)$ while the shape of the surface changes with respect to one derivable from an unchanging magnetic field just as if the constant C decreased in ratio $f(t_0)/f(t)$. The interval of time t_1 to t_2 in which the C values correspond to potentials with a minimum is therefore, for the case illustrated in Figure 2:

$$\frac{f(t_2)}{f(t_1)} = \frac{P_{18} + C_{max}}{P_{18} + C_{min}} = 1.15, \tag{51}$$

corresponding approximately to 15 per cent. of the phase angle of the injection period characteristic of a sinusoidally varying magnetic field, an angle proportional to the square root of the injection voltage. The range of C values corresponding to initial functions having the special properties necessary for capture is a small part of the total range of C values characterizing a minimum of the potentials V_M and can be estimated on the basis of the size and location of the gun and is unlikely to exceed 10 per cent. in practical cases. The effective phase in which electrons can be captured is therefore only a small percentage, about 1 per cent., of the injection phase itself, and the corresponding effective time of capture is quite short (about 10^{-7} second for 4,000 volts injection voltage and other parameters used in the example of this paper). In spite of this, it is found that in present day betatrons the current emission capabilities of conventional guns are ample to provide a charge equal to the limit given by space charge, equation (29).

The mechanism of gun clearing was accounted for by ignoring the effects of space charge, while the upper limit of capturable charge was estimated in a preceding section by ignoring the obstruction of the gun. The actual effects are only partially described by these two limiting, idealized cases. However, it is probable that even if the capturing potential troughs were space charge limited, this fact would not alter the existence of limiting requirements on values for the damping of the amplitude of oscillation, the rate of equilibrium orbit shift or the period T of consecutive passages near the gun, which are primarily responsible for the gun clearing mechanism. It is even possible that space charge has an enhancing effect in damping oscillations for if one supposes formula (37) to be entended to such a case, the principal alteration would be a reduction in $\partial V_M/\partial r$, with a consequent considerable increase in Δa. On the other hand, it is very likely that the particularities required from the potential function to avoid the gun obstruction are inconsistent with large capturable charges and are responsible for the low percentages of the upper space charge limit which are usually observed in actual experiments.

PART II

BETATRON WITH AN ADDITIONAL ELECTRIC FIELD

The injection of electrons from a cylindrical cathode coaxial with the magnetic field with the help of an additional radial electric field has tempted many investigators. A priori, it seems that large space charge could be captured by this means since electrons are injected in initially intense fields, a condition which is known to yield large rotating charges

in magnetrons. In addition, the problems of an asymmetrically located gun are eliminated. The following analysis shows the requirements in the variations of the fields necessary to obtain equilibrium and stability. These turn out to be more complex and to have narrower tolerances than in the case of a purely magnetic betatron. It is shown also that no substantial gain in space charge can actually be realized.

The analysis is restricted to a toroidal region whose radial and axial limits are small compared to the average radius and also, to non-relativistic speeds. It is convenient to use a power expansion of the total potential of forces $V_t = V_{M0} - V_E$ around a position $z = 0$ and $r = r_e$, or $x = 0$ when $r - r_e$ is abbreviated by x, where r_e is the asymptotic equilibrium orbit of the purely magnetic instrument with synchronous field variation defined by equation (23).

The expansion, limited to cubic terms:

$$V_t = V_t(0, 0) + x \frac{\partial V_t}{\partial x} + \tfrac{1}{2}x^2 \frac{\partial^2 V_t}{\partial x^2} + \tfrac{1}{2}z^2 \frac{\partial^2 V_t}{\partial z^2} + \tfrac{1}{6}x^3 \frac{\partial^3 V_t}{\partial x^3} + \tfrac{1}{2}xz^2 \frac{\partial^3 V_t}{\partial x \partial z^2}, \quad (52)$$

where the derivatives are understood to be for $x = z = 0$, has no terms with odd powers of z because of the assumed symmetry about the $z = 0$ plane. It is convenient also to break up this expansion into separate ones for V_{M0} and V_E and to introduce for the former two new symbols, B_e and B_q, defined as follows:

$$B_e = B_z(r_e, 0, t), \quad (53)$$

$$B_q = \frac{1}{r_e^2}\left[\int_0^{r_e} B_z \big|_{z=0} r\,dr - \int_0^{r_0} B_z \big|_{\substack{z=z_0 \\ t=t_0}} r\,dr + \frac{m_0}{e} r_0^2 \theta_0 \right]$$

$$= \frac{1}{r_e^2} \int_0^{r_e} B_z \big|_{z=0} r\,dr + \frac{C}{r_e^2}. \quad (54)$$

It is seen that B_e is the magnetic field at the main equilibrium orbit and B_q is proportional to the angular momentum of the electrons at the position $r = r_e$ and $z = 0$ and is also equal to $\tfrac{1}{2}\pi r_e^2$ times the flux linking the circle $r = r_e$, $z = 0$, at any time, plus a constant which depends on the initial configuration of the magnetic field within the initial circle $r = r_0$, $z = z_0$ as well as the initial position and angular momentum of the electron.

The expansion of V_{M0}, expressed in terms of B_e and B_q with the relation between the spatial derivatives of B taken into account, takes the

form:

$$V_{M0} = \tfrac{1}{2}\frac{e}{m_0} r_e^2 B_q^2 + x\frac{e}{m_0} r_e B_q [B_e - B_q]$$

$$+ \tfrac{1}{2}x^2\frac{e}{m_0}\left[B_e^2 + 3B_q^2 - (3 + n)B_e B_q\right] + \tfrac{1}{2}z^2\frac{e}{m_0} nB_e B_q$$

$$+ \tfrac{1}{6}x^3\frac{e}{m_0 r_e}\frac{1}{}\left[-(3 + 3n)B_e^2 - 12B_q^2 + (12 + 4n + n_2)B_e B_q\right]$$

$$+ \tfrac{1}{2}xz^2\frac{e}{m_0 r_e}\frac{1}{}\left[nB_e^2 - (n + n_2)B_e B_q\right], \tag{55}$$

where the numerical coefficients n and n_2 are defined as:

$$n = -\left.\frac{r}{B_z}\frac{\partial B_z}{\partial r}\right|_{\substack{r=r_e\\z=0}}, \quad n_2 = \left.\frac{r^2}{B_z}\frac{\partial^2 B_z}{\partial r^2}\right|_{\substack{r=r_e\\z=0}}. \tag{56}$$

If the magnetic field varies synchronously in the region of validity of the expansion, n and n_2 are constants and the coefficients of the expansion are second order polynomials of the two variables B_e and B_q. These variables must be particular functions of time for particular modes of operation. To analyze all possible modes, it is convenient to assume at first that B_e and B_q vary independently and later to consider the effect of their actual functional dependence. The assumption that B_e and B_q are arbitrary functions of time requires that the flux linking the circle $r = r_e$ and $z = 0$ and the field at that circle be controlled independently, and implies that the flux exterior to and linking the toroid is independent of the field within the toroid since some of the flux linking r_e is controlled by B_e. A large degree of independent control was actually obtained in an experimental double yoke magnetic circuit.

The analysis of the variation of the potential functions V_t is greatly facilitated by a diagram having B_e as abscissa, B_q as ordinate and showing curves for which the values of the coefficients of the expansions V_t, V_{M0} or V_E remain constant. Figure 3 is such a diagram for a particular choice of geometrical parameters. The lines of constant coefficients of V_{M0} are conic sections and each coefficient is represented by a family of concentric and homothetic conics. For example, the lines of constant $\partial V_{M0}/\partial x$ are hyperbolas whose asymptotes are the B_e axis and the $45°$ line $B_q = B_e$. The asymptotes themselves are lines of $\partial V_{M0}/\partial x = 0$ and they devide the $B_q - B_e$ plane into positive and negative regions of $\partial V_{M0}/\partial x$.

Fig. 3—Betatron focusing diagram.

In the expansion of the electrostatic potential V_E, the coefficients are related by Laplace's equation, assuming the space charge to be negligible. Assuming also the spatial distribution of the electrostatic potential to remain similar to itself as the potential varies in time, that is, synchronous variation of the electrostatic field, the ratios of $\partial^2 V_E / \partial x^2$ and $\partial^2 V_E / \partial z^2$ to $\partial V_E / \partial x$ remain constant, and it is possible to express them in terms of a numerical factor k:

$$\frac{\partial^2 V_E}{\partial z^2} = \frac{k}{r_e}\frac{\partial V_E}{\partial x} \quad \text{and} \quad \frac{\partial^2 V_E}{\partial x^2} = -\frac{k+1}{r_e}\frac{\partial V_E}{\partial x}. \tag{57}$$

The expansion of V_E limited to cubic terms is therefore:

$$V_E = V_E(0, 0) + x\frac{\partial V_E}{\partial x} - \tfrac{1}{2}x^2\frac{k+1}{r_e}\frac{\partial V_E}{\partial x} + \tfrac{1}{2}z^2\frac{k}{r_e}\frac{\partial V_E}{\partial x} + \tfrac{1}{6}x^3\frac{\partial^3 V_E}{\partial x^3} + \tfrac{1}{2}xz^2\frac{\partial^3 V_E}{\partial x\partial z^2}. \tag{58}$$

The conditions (17) for stable equilibrium, or the existence of a minimum of the function $V_t = |V_M - V_E|$, can now be conveniently analyzed. In general, the capture of charge can occur as long as the minimum is within the toroidal region but we will consider at first only the case of electrons whose equilibrium orbit is always in the center of the toroidal region $x = 0$, $z = 0$, and later compare other captured electrons to these "normal" electrons. The first order condition (17a) requires that at all times:

$$\frac{\partial V_{M0}}{\partial x} = \frac{\partial V_E}{\partial x} = \frac{e}{m_0}r_e B_q[B_e - B_q]. \tag{59}$$

The lines of constant $\partial V_{M0}/\partial x$, identical with $\partial V_E/\partial x$, on the focusing diagram show what value of electric field must be applied for a given magnetic field B_e, and a given equivalent change of flux B_q in order to balance the electric, Lorentz and centrifugal forces at every instant.

Consider the location of the representative point on the $B_q - B_e$ diagram for the instant of emission. (θ_0 is zero, neglecting the thermal velocities of emission.) The value of B_q can be expressed approximately in terms of B_e by expanding the integrals of equation (54):

$$B_q = sB_e$$
$$= \left[-\frac{x_0}{r_e} + \frac{1-n}{2}\left(\frac{x_0}{r_e}\right)^2 - \frac{n}{3}\left(\frac{x_0}{r_e}\right)^3 - \frac{n}{2}\left(\frac{z_0}{r_e}\right)^2\frac{r_0}{r_e} \right]B_e. \tag{60}$$

The constant s is hence approximately equal to $- (x_0/r_e)$, x_0 being the initial value of x. The representative point at the emission time is therefore on the $B_q = sB_e$ line shown on the figure for two values of $s = \pm 1/16$. Subsequent to emission, the value of B_q will increase in an actual induction accelerator since the kinetic energy is an increasing monotonic function of B_q. No natter how B_e is related to the increase of B_q, eventually the operating point (B_q, B_e) must end on or near the asymptote $B_q = B_e$ for otherwise the values of $\partial V_E/\partial x$ would have to become unpractically large to balance the magnetic and inertial forces which increase rapidly with speed. Therefore, the focusing problem consists essentially in analyzing the conditions at points between the lines $B_q = sB_e$ and $B_q = B_e$ and on the line $B_q = B_e$ itself. It is interesting to observe that these limiting lines are representative of the field conditions in a static magnetron and a purely magnetic betatron respectively. The electric field at the emission time

$$\frac{\partial V_E}{\partial x}\bigg|_{t=t_0} = \frac{e}{m_0} r_e s(1 - s)B_{e0}^2, \tag{61}$$

where B_{c0} is the value of B_e at $t = t_0$, is approximately the value of the cut-off electric field of the magnetron, while on the line $B_q = B_e$, the electric field is zero, the characteristic of a conventional purely magnetic betatron. The betatron with radial injection by an auxiliary electric field operates toward the end of the acceleration cycle, as a purely magnetic instrument. Thus, it may be considered as the superposition on a magnetic betatron with its main sinusoidal power source and synchronously varying magnetic field of auxiliary means to alter the field and flux by the incremental values B_{ec} and B_{qc} necessary to obtain equilibrium in combination with the electric field.

Consider now the conditions imposed on the variations of the magnetic and electric fields by the second order conditions (17b and c) for the existence of a potential minimum at $x = z = 0$. Since the values of $\partial^2 V_E/\partial x^2$ and $\partial^2 V_E/\partial z^2$ are proportional to $\partial V_E/\partial x$, equations (57), they can be expressed in terms of B_c and B_q if it is assumed that the electric field $\partial V_E/\partial x$ is actually adjusted at every instant in accordance with the first order minimum condition (59). The second derivatives of V_t are thus:

$$\frac{\partial^2 V_t}{\partial x^2} = \frac{\partial^2 V_{M0}}{\partial x^2} - \frac{\partial^2 V_E}{\partial x^2}$$

$$= \frac{e}{m_0} \left[B_c^2 + 3B_q^2 - (3 + n)B_e B_q + (k + 1)B_q(B_e - B_q) \right], \tag{62}$$

$$\frac{\partial^2 V_t}{\partial z^2} = \frac{\partial^2 V_{M0}}{\partial z^2} - \frac{\partial^2 V_E}{\partial z^2} = \frac{e}{m_0} \left[nB_e B_q - kB_q(B_e - B_q) \right]. \quad (63)$$

The lines of constant curvatures $\partial^2 V_t/\partial x^2$ and $\partial^2 V_t/\partial z^2$ are conic sections depending on the parameters n and k characteristics of the geometrical configurations of the magnetic and electric fields. In order for these curvatures to be positive as the operating point moves from the line $B_q = sB_e$ to and on the line $B_q = B_e$, these parameters must have appropriate values. The value of n must satisfy the relation, $0 < n < 1$, as in condition (22), of the limiting magnetic betatron. This can be seen by making $B_q = B_e$ in equations (62) and (63).

The permissible limits of k can be found by considering at first the lines of constant $\partial^2 V_t/\partial z^2$ which form a family of concentric and homothetic hyperbolas whose asymptotes, or lines of zero value, are the lines $B_q = 0$ and $B_q = B_e(1 - n/k)$. Obviously, the operating line must not cross the asymptotes, so that for the starting point (B_{q0}, B_{e0}) on the line $B_q = sB_e$, B_{q0} must be positive or $sB_{e0} > 0$. This means that the initial magnetic field must be positive with an internal cathode $(s > 0)$ and negative for an external cathode $(s < 0)$ as illustrated on the focusing diagram. In order for the asymptote $B_q = B_e(1 - n/k)$ to be outside of the operating angle, it is easy to show that $k < (n/1 - s)$ for $s > 0$ and $k > (n/1 - s)$ for $s < 0$.

The lines of constant $\partial^2 V_t/\partial x^2$ are ellipses if the discriminant of the polynomial in B_e and B_q is negative, a condition occurring when:

$$n - 2\sqrt{1 - n} < k < n + 2\sqrt{1 - n}. \quad (64)$$

If k lies outside of the limits of this inequality the equivalue lines $\partial^2 V_t/\partial x^2$ are hyperbolas and there are points in the B_q, B_e plane for which $\partial^2 V_t/\partial x^2 < 0$. It can be shown that, if $0 < n < 1$ and the inequalities required by $\partial^2 V_t/\partial z^2 > 0$ are satisfied, some of these negative points are necessarily in the operating angle, so that inequalities (64) must be satisfied.

The focusing conditions in the radial and axial directions within the operating ranges are therefore:

$$0 < n < 1,$$

$$n - 2\sqrt{1 - n} < k < \frac{n}{1 - s} \quad \text{for internal cathode } s > 0,$$

$$\frac{n}{1 - s} < k < n + 2\sqrt{1 - n} \quad \text{for external cathode } s < 0.$$

$$(65)$$

The parameters used in drawing the diagram of Figure 3 are those of an experiment performed by the authors. The magnetic field was so adjusted that the measured values were $r_e = 1.6''$, $n = .565$ and $n_2 \cong 0$. The two cylindrical electrodes both carried electron emissive areas so that either could be made the cathode. ($x_0 = \pm .1''$, making $s = \mp \tfrac{1}{16}$.) See Figure 6. A definite curvature of electric field was obtained by adjusting the potential of the end shields or "hats." Measurements in an electrolytic tank of a scale model showed that when the hats were connected to the cathode $k = -.6$ for an internal cathode and $k = 1.15$ for an external cathode. These values of s, k and n satisfy the inequalities (65).

With the help of the focusing diagram, feasible modes of operation can be predicted by considering what variations of magnetic field increment B_{ec}, flux increment B_{qc} and electric field $\partial V_E/\partial x$ produce focusing and among them, which are most easily obtained in practice. Linear variations of the flux B_q and field B_e are particularly simple if the corresponding lines in the diagram are parallel to the asymptotes of the equivalues of $\partial V_{M0}/\partial x$ because the required variation of $\partial V_E/\partial x$ is then also linearly related to the field or flux. For short enough phase angles of a sinusoidal drive, these variations are also linear with respect to time. Two such operating lines have been shown on the focusing diagram, one for an internal and the other for an external cathode, and the corresponding time variations in Figure 4. It can be observed in the case of an internal cathode, that B_q takes two different values for some values of B_e so that B_e must necessarily be non-monotonic and consequently both field and flux must be controlled by the increments B_{ec} and B_{qe}. On the other hand, control of the flux alone is sufficient with an external cathode. In this case, at the emission time, the magnetic field is negative and therefore defocusing in the axial direction. The electric field overcompensates that effect and is also responsible for giving stability to the electron orbits at the instant when there is no magnetic field at all. This shows, incidentally, that the B_q axis illustrates the conditions of a purely electrostatic system.

The focusing diagram showing the conditions for a minimum at $x = z = 0$ is convenient for finding further conditions for capturing the charge. It is evident that a finite range of electron emission times must be considered, for infinite current densities would be necessary if all the emission were to occur at one particular instant. For every emission time there is a different potential function V_{M0} because the constant C involving the initial flux is different. Since the electrostatic potential V_E is necessarily the same for all electrons, the minima

of $V_t = V_{M0} - V_E$, if existing, will be, in general, at different instantaneous positions x_i and have different curvatures $\partial^2 V_t/\partial x^2$ and $\partial^2 V_t/\partial z^2$ for each electron. The electrons will be captured for such shifts of instantaneous orbits x_i and such changes in the depth of the potential trough which will effectively prevent them from hitting the bounding electrodes. It is obvious that the initial position x_{i0} of the instantaneous

Fig. 4—Time variation of the magnetic field, magnetic flux and electric field.

equilibrium orbit x_i, must be closer to the cathode than the anode;

$$0 < \frac{x_{i0}}{x_0} < 1. \tag{66}$$

Furthermore, if at no time the orbit is closer to the cathode or anode than it was originally to the cathode, or

$$|x_i| < |x_{i0}|, \tag{67}$$

capture will be insured as long as $\partial^2 V_t/\partial x^2$ and $\partial^2 V_t/\partial z^2$ remain no smaller than their initial values.

We shall restrict the analysis to the case, described above, of linear field variations providing an equilibrium at $x = 0$ for so-called normal electrons for which the variables $B_q = B_{qn}$, B_e and $\partial V_E/\partial x$ satisfy equation (59) and B_{qn} is given on the successive segments of the operating line by:

$$B_{qn} = B_e - (1 - s)B_{e0n}, \ B_{qn} = \text{constant}, \ B_{qn} = B_e, \quad (68)$$

while $\partial V_e/\partial x$ can be expressed on the first of these by:

$$\frac{\partial V_E}{\partial x} = \frac{e}{m_0} r_e(1 - s)B_{e0N}(B_e - (1 - s)B_{e0N}), \quad (69)$$

where B_{e0N} is the initial value, B_{e0}, of B_e at the emission time of the normal electron. The potential V_{M0} of other electrons finding themselves in the same magnetic field will depend on B_e and a value B_q differing from B_{qn} by a constant amount ΔB_q characteristic of the emission time. The position of the minimum of the potential $V_t = V_{M0} - V_E$ will be at a position x_i determined for times subsequent to normal emission and approximately equal to:

$$x_i = \frac{-\dfrac{\partial V_t}{\partial x}}{\dfrac{\partial^2 V_t}{\partial x^2}} = -r_e\frac{\Delta B_q(B_e - 2B_{qN})}{B_e{}^2 + 3B_q{}^2 - (3+n)B_eB_q + (k+1)B_{qN}(B_e - B_{qN})}. \quad (70)$$

The initial equilibrium orbit position x_{i0} of an electron emitted subsequently to the normal electron is seen from equation (70) to have the same sign as B_{e0}, therefore opposite to x_0, in violation of condition (66), an understandable result since for late electrons the electric field is stronger than cut-off of an analogous magnetron.

Therefore, only early electrons are capturable. Whether this will actually occur depends on the electric field, which, prior to the normal emission, is not specified by the normal equilibrium equations (59) and (69). It will be assumed to vary linearly with B_e at a rate p times faster than immediately subsequent to normal emission. With this assumption, both V_{M0} and V_E are determined for early electrons and it is possible to compute the instantaneous position of x_i of the potential minimum of $V_t = V_{M0} - V_E$ for each instant t and each emission time t_e in the emission period extending from the instant t_0 at which the electric field is zero to the instant t_n of the emission of the normal electron. It is convenient to introduce the numerical factors $q = t_e/(t_n - t_0)$ characteristic of the emission time and $\lambda = t/(t_n - t_e)$ characteristic of any instant in the period between the emission of any electron and the normal electron. With these abbreviations, the result of the computation is:

$$x_i = x_0(1 - q)\left(1 - \lambda + \frac{\lambda}{p}\right), \quad (71)$$

when terms in $(x_0/r_e)^2$ are neglected with respect to terms in (x_0/r_e). Therefore the initial position x_{i0} of the position minimum $(\lambda = 0)$ and the position x_{in} at the time of the normal emission $(\lambda = 1)$ are:

$$x_{i0} = x_0(1 - q), \tag{72}$$

$$x_{in} = x_0 \frac{(1 - q)}{p}. \tag{73}$$

Equation (72) shows that the initial equilibrium orbit is at the cathode for the earliest emission time $(q = 0)$ and in the center of the tube at the latest $(q = 1)$. The amplitude of oscillation at the emission time of any electron is $(x_0 - x_{i0}) = |qx_0|$ and becomes slightly smaller at the normal emission time in virtue of the increase of $\partial^2 V_t/\partial x^2$ in the interim.

At the instant of emission of the normal electron, there are already in the tube a series of electrons all oscillating around different equilibrium orbits with different amplitudes. In order for (x_{in}) to be less than (x_{i0}) in accordance with relation (67), p must be greater than one. At the limiting value when the rate of change of the electric field at the emission period is equal to the rate subsequent to normal emission time, or $p = 1$, the instantaneous equilibrium orbit of any electron does not change from the instant of emission to the normal emission time, $x_i = x_{i0} = x_{in}$, and the electrons all oscillate so as to come back to the cathode at each swing.

After the emission of the normal electrons, the position x_i of the equilibrium orbits is given by equation (70). The constants ΔB_q can be eliminated by comparing the instantaneous equilibrium positions x_i. to those, x_{in}, the electrons had at the time t_n of the emission of the normal electron:

$$\frac{x_i}{x_{i_n}} \cong \frac{B_e - 2B_{qN}}{B_{e0N} - 2B_{q0N}} \frac{\left. \frac{\partial^2 V_t}{\partial x^2} \right|_{t_n}}{\left. \frac{\partial^2 V_t}{\partial x^2} \right|_t} = f_n. \tag{74}$$

The function f_n is determined for every point of the operating line on the B_q–B_e diagram. For the lines shown in Figure 3, for an external cathode f_n varies from 1, passes through a maximum of 2 and eventually tends asymptotically to zero, while for an internal cathode f_n varies from one, passes through zero to a minimum value of -2.5 and eventually also tends to zero.

The position x_i of the orbit at any time subsequent to the emission

period can be compared now to its initial position by combining equations (72), (73) and (74).

$$\frac{x_i}{x_{i0}} = \frac{f_n}{p},\tag{75}$$

showing that x_i/x_{i0} is independent of the emission time q. In order to satisfy condition (67):

$$p \geqslant |f_n|_{max}.\tag{76}$$

This shows that the rate of increase of the electric field during emission time must be greater than subsequently by a factor p which depends on the particular subsequent mode of operation. In our example $p > 2$ for external cathode and $p > 2.5$ for internal cathode. Higher rates than necessary insure more positively that the electrons will miss striking the bounding electrodes but the emission period is shortened so that high cathode emission is necessary to inject a given charge. However, for most practical parameters, as in the examples illustrated, the emissivity of the cathode is no limiting factor.

The schematic history of the instantaneous equilibrium position and amplitude of oscillation of Figure 5 summarizes graphically the above considerations. When the electric field becomes zero, the conditions are those of a purely magnetic betatron and the instantaneous orbits tend towards the center while the amplitude of oscillation diminishes, as explained before.

The analysis of the emission period made above on the basis of the motion of the electrons in the radial direction only, were based on the two first terms of the expansion of the potential of forces. Actually, the two dimensional motion in the r-z plane should be considered by taking into account the effects of the actual boundaries, such as the cathode, for which the second order terms approximation breaks down. Such a detailed analysis was made for a specific example by using the cubic terms for V_{M0} and an actual electrolytic plot for V_E and confirmed essentially the conclusions of the simpler analysis.

The magnetic field B_e, the flux linking the circle $r = r_e$, $z = 0$, as characterized by B_q, and the electrostatic field $\partial V_E/\partial x$ must be related through equation (59) at all times. It is significant to consider the tolerances on the variations of these variables by computing the differential changes in the time variables which cause a shift Δx in the equilibrium orbit:

$$\Delta \left(\frac{\partial V_E}{\partial x}\right) = -\Delta x \frac{\partial^2 V_t}{\partial x^2},\tag{77}$$

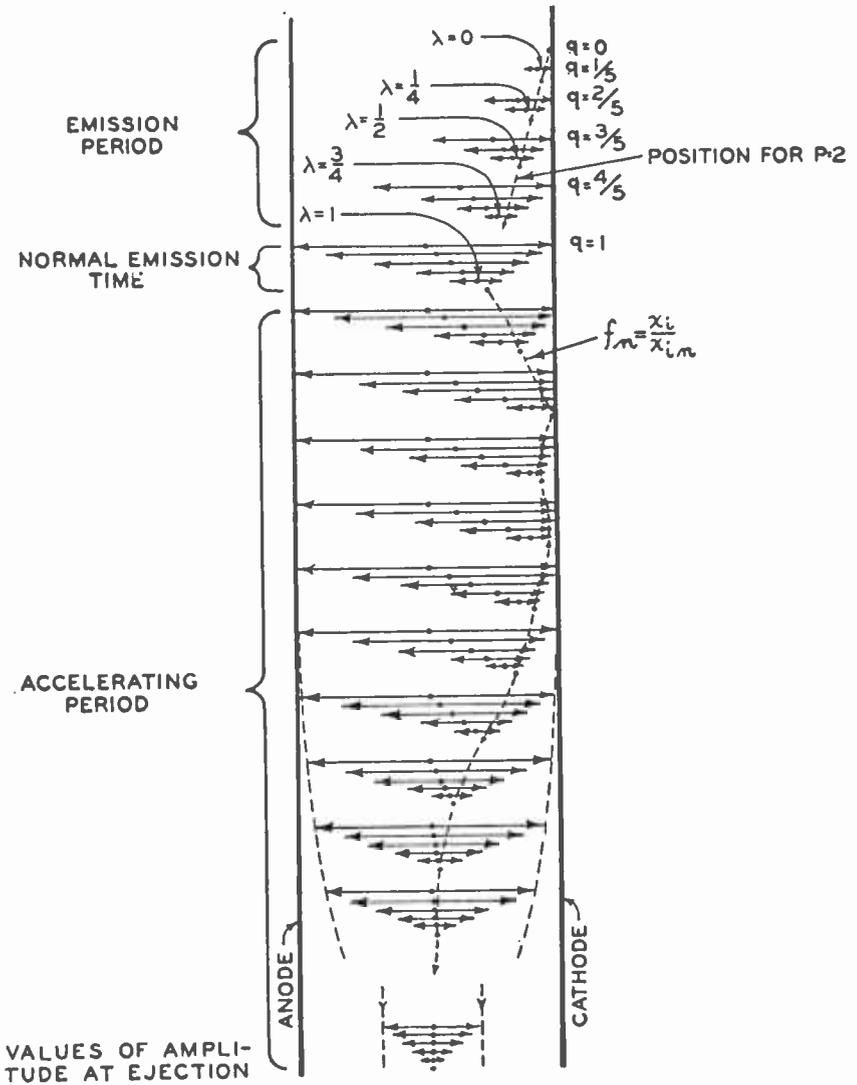Fig. 5—Schematic history of oscillations.

$$\Delta B_q = -\Delta x \frac{\dfrac{\partial^2 V_t}{\partial x^2}}{\dfrac{e}{m_0} r_e (B_e - 2B_q)}, \tag{78}$$

$$\Delta B_e = - \Delta x \frac{\dfrac{\partial^2 V_i}{\partial x^2}}{\dfrac{e}{m_0} r_e B_q}. \tag{79}$$

The maximum permissible value of Δx is the space between the point of maximum oscillating displacement and the cathode or anode, which for times subsequent to the emission period is:

$$= (1 - q)\left(\pm 1 - \frac{f_n}{p} \right) x_0. \tag{80}$$

It is apparent that there is no tolerance for the normal electron $(q = 1)$, and the maximum tolerance for the earliest emitted electron $(q = 0)$. If half of the initial emission is sacrificed a reasonable value for Δx is $\pm x_0/4$ for practical values of f_n and p.

The tolerances for the operating lines shown on the focusing diagram are as follows: Since $\partial^2 V_i/\partial x^2$ remains approximately constant when $\partial V_E/\partial x \neq 0$, the absolute tolerance of $\partial V_E/\partial x$ remains constant and tends to a minimum relative error of .6 per cent at maximum voltage. The tolerance ΔB_q is expressible as

$$\Delta B_q = \frac{\Delta x}{x_0} \cdot \frac{s(1 - s)}{f_n} B_{eos},$$

so that the least relative tolerance in the control flux B_q, $\Delta B_q/B_{qc} = \Delta B_q/B_{eos}$ is about .5 per cent for the maximum value of f_n. Similarly, the relative tolerance in B_e is of the order of .7 per cent, well within the approximate linearity of a sine function at the phase angles of the injection period. These tolerances of a few tenths of a per cent on the variation of the electric field and magnetic flux constitute very close requirements for electronic circuits controlling appreciable power at reasonably high frequencies.

The ejection of the accelerated charge in a betatron with a system of injection using a radial electric field can of course be obtained by a non-synchronous change of flux or field as is done in the purely magnetic betatron. It could be achieved also, in principle, by applying an ejecting electric field on the existing injecting electrode structure. However, the required gradients would be impractically high.

It was shown that space charge determines the ultimate limit to the charge which can be captured and accelerated. In the purely magnetic betatron this limit increases with the energy of the electrons so that

far more charge is potentially supportable near ejection time than can possibly be introduced at injection time. As mentioned before, the addition of an electrostatic field might be expected to provide a means for taking some advantage of this potentiality since the electrons could be injected into a much stronger initial magnetic field, as is the case in static magnetrons which are known to have high current densities. Actually the gain of supportable charge can be only very small, for the following reasons:

The Laplacian of the magnetic potential was shown to represent the limiting space charge density of stationary electrons at any point. This limit ρ at the minimum of V_M can be found by introducing the equilibrium condition $\dfrac{\partial V_E}{\partial r} = \dfrac{\partial V_M}{\partial r}$ into equation (27).

$$\rho = \epsilon_0 \left\{ \frac{\left(\dfrac{\partial V_E}{\partial r}\right)^2}{2V_M} + \frac{2V_M}{r^2} \right\} \tag{81}$$

For a given electric field $\partial V_E/\partial r$, ρ is a minimum when $V_M = \dfrac{r}{2}\left|\dfrac{\partial V_E}{\partial r}\right|$, a value which V_M must assume necessarily at some time since it passes continuously from nearly zero at emission time to values much larger than $\dfrac{r}{2}\left|\dfrac{\partial V_E}{\partial r}\right|$. Therefore, the upper limit of capturable space charge density at the minimum of the potential for a given largest value of the electric field is:

$$\rho = \epsilon_0 \frac{2}{r}\frac{\partial V_E}{\partial r}, \tag{82}$$

whatever the particular variation of the magnetic flux and fields and electric field chosen to satisfy the equilibrium conditions. The similar limit for a purely magnetic instrument was found to be equation (27), $\rho = \epsilon_0 \dfrac{2\,V_M\text{ injection}}{r^2}$, and therefore the relative gain in space charge density obtained from the electric field can be at most $\dfrac{\partial V_E}{\partial r}\dfrac{r}{V_M}$. If the maximum voltage V_E applied across the tube is equal to the injection voltage, this ratio becomes approximately $r/\Delta r$, assuming the gradient at r to be equal to the average gradient. The ratio of capturable charges themselves will be also $r/\Delta r$ if the volumes occupied by the

charges are the same in both cases. In practice, however, the region of acceleration is smaller for radial injection, while the limit imposed by breakdown on the applicable voltage is likely to be higher. Therefore, the gain in space charge limit, if any, is not very important.

In the example used in the preceding analysis, the actual limit of capturable charge can be estimated by applying Gauss' theorem over the bounding electrodes.

$$Q = \epsilon_0 \int\int_s \mathrm{grad}_n \ V_t ds \cong \epsilon_0 4\pi r_e \left[z_0 \frac{\partial V_t}{\partial x}\bigg|_{\substack{x=x_0 \\ z=0}} + x_0 \frac{\partial V_t}{\partial z}\bigg|_{\substack{x=0 \\ z=z_0}} \right] \cdot \ (83)$$

The value of $\dfrac{\partial V_t}{\partial x}\bigg|_{\substack{x=x_0 \\ z=0}}$ can be approximated by assuming a parabolic variation of V_t in the radial direction, while the value of $\dfrac{\partial V_t}{\partial z}\bigg|_{\substack{x=0 \\ z=z_0}}$, which is very sensitive to the sharp boundary conditions due to the hats, can be conveniently assumed equal to $\dfrac{x_0}{z_0} \cdot \dfrac{\partial V_t}{\partial x}\bigg|_{\substack{x=0 \\ z=0}}$. Hence Q is approximately:

$$Q = \epsilon_0 4\pi r_e \frac{\partial^2 V_t}{\partial x^2} x_0^2 \left[\frac{z_0}{x_0} + \frac{x_0}{z_0} \right]. \tag{84}$$

Since Q is proportional to $\partial^2 V_t/\partial x^2$, it will be most for a given $\partial V_E/\partial x$ at a time when $\dfrac{\partial V_E}{\partial x} \bigg/ \dfrac{\partial^2 V_t}{\partial x^2}$ is extremum or, on the focusing diagram when an elliptical equivalue of $\partial^2 V_t/\partial x^2$ is tangent to a hyperbolic equivalue of $\partial V_E/\partial x$. For this point the discriminant of the quadratic equation for B_q^2, which can be obtained by substituting B_c from equation (63) in equation (62), is zero, a condition occurring when:

$$\frac{\partial V_E}{\partial r} = \frac{\partial^2 V_t}{\partial x^2} r_e \frac{(k-n) \pm 2\sqrt{1-n}}{(k-n) - 4(1-n)}, \tag{85}$$

in which the plus sign refers to an external and the minus to an internal cathode. Therefore Q:

$$Q = 4\pi\epsilon_0 x_0^2 \frac{(k-n) - 4(1-n)}{(k-n) \pm 2\sqrt{1-n}} \left[\frac{z_0}{x_0} + \frac{x_0}{z_0} \right] \frac{\partial V_E}{\partial x}$$

$$= 4\pi\epsilon_0 x_0^2 h \frac{\partial V_E}{\partial x} \simeq 2\pi\epsilon_0 x_0 h V_{E\,\text{max}}, \quad (86)$$

is seen to be proportional to the maximum applied voltage, the linear dimensions, and a factor h which depends on the geometry of the boundaries of the fields. Actually, this optimum ration of $\partial V_E/\partial x$ to $\partial^2 V_E/\partial x^2$ is not realized if the three equilibrium variables, magnetic flux and field and electric field vary linearly with respect to each other. Taking this factor into account, for the parameters indicated on the diagram and a maximum applied voltage of 4,000 volts, this maximum charge turns out to be 5×10^{-11} coulomb.

EXPERIMENTAL

In order to investigate the injection mechanisms and current limitations the present authors built a small but special betatron with which some relevant experimental results were obtained.

The primary feature of the magnetic circuit was the independent control of the field at—and the flux linking—the main equilibrium orbit. The pole pieces were composed of two coaxial sets, an outer set with a return link in the form of a C carrying the field exciting coils, and an inner set relying on leakage for a return path and having its flux coils inside of the hollow outer pieces. This shape of the outer poles minimized the leakage between the inner and outer circuits so that only one-fifth of the variation on one was perceptible in the other. A coil on the outer circuit in series with a coil on the inner, tuned by a bank of condensers to the driving frequency of 420 cycles, was coupled through a transformer to a 5 KW generator. This main circuit was adjusted to produce the conventional magnetic betatron field with its 1:2 condition for the equilibrium orbit. Another system of coils comprising one on the outer field link in antiseries with another on the inner, flux pole piece was adjusted to have no coupling with the main circuit. The excitation of this auxiliary, isolated circuit produced a differential in the flux and field and could be used for both ejection and injection in experiments either with conventional injection or with an additional electric field. Figure 6 shows a cross section of the pole pieces.

The ends of the outer pieces were made of laminations stacked radially in a bakelite holder and forming a 78° cone of about $4\frac{3}{8}''$ diameter which produced the radial variation of field shown in Figure 1. The remainder of the pole pieces consisted of a sheet of steel wound with proper insulation in a spiral to form an approximate cylinder $3\frac{3}{4}''$
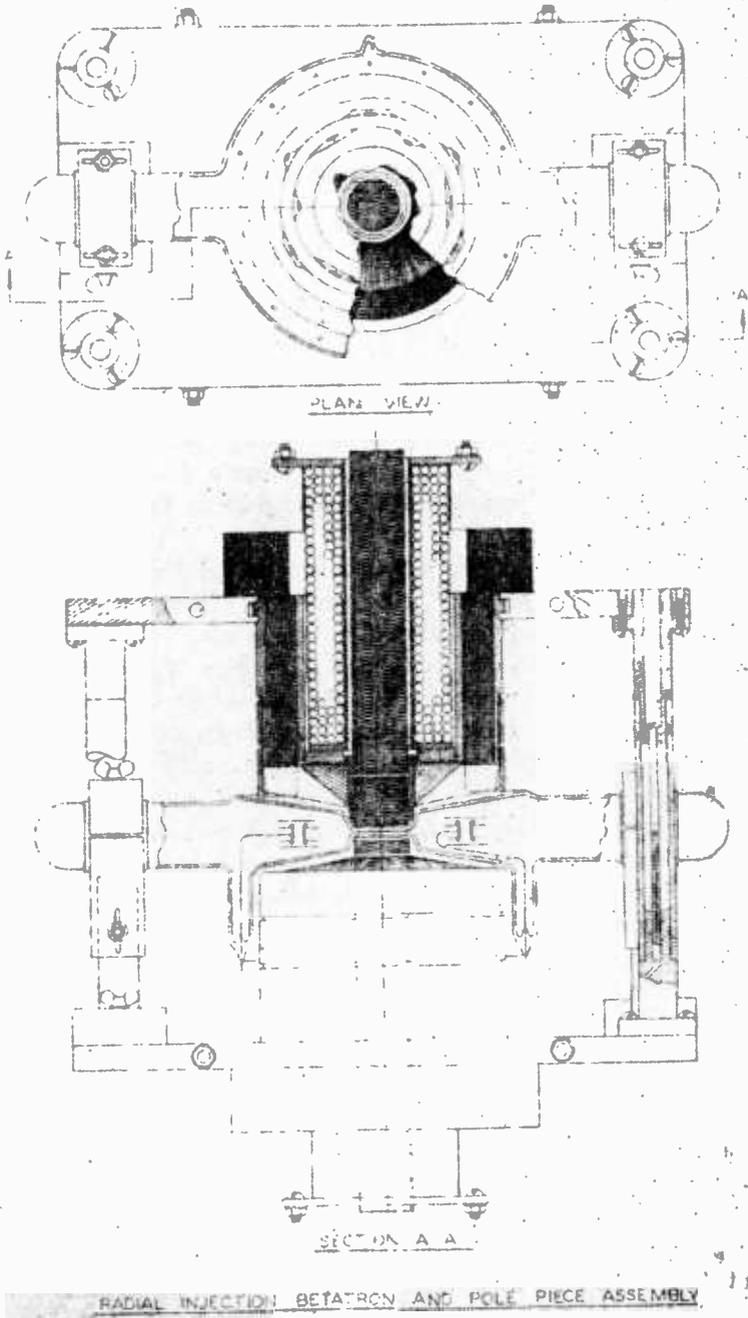
PLAN VIEW

SECTION A A

RADIAL INJECTION BETATRON AND POLE PIECE ASSEMBLY

Fig. 6—Cross-section of pole pieces and tube for radial injection test.

high, $3\frac{1}{2}''$ I.D., $4\frac{3}{8}''$ O.D. This wound section equalized the flux angularly to better than 1 per cent in spite of dissymmetries in the air gaps between the pole pieces and the yoke. The inner pole pieces were laminated cylinders, 1″ in diameter.

Measurements of the field distribution showed that when the angular dissymmetry exceeded 4 per cent no tubes were operative, but improvement down to 1 per cent had no further beneficial effect in conventional, purely magnetic betatrons. Instead of computing the main potential function from the radial distribution of the magnetic field obtained with a search coil, it was found less tedious to use a special system of 23 concentric coils whose numbers of turns were made inversely proportional to their respective radii. The position of the equilibrium orbit could be determined easily by switching to the coil with minimum signal on a peak voltmeter. The position of the equilibrium orbit at injection was obtained from a direct measurement while that at ejection was obtained by first passing the signal through an integrating circuit.

The glass envelope of the purely magnetic betatron had the shape of a doughnut $4\frac{1}{2}''$ O.D., $1\frac{3}{8}''$ I.D., and was blown by conventional techniques with the help of special carbon molds. The internal glass surface was platinized to insure that all the region of electron acceleration is an electrostatic equipotential. To prevent excessive heating by eddy currents, this metallization was very thin and was interrupted along a radius. The electron injector sealed in a side tube was of conventional type including an oxide coated cathode (or tungsten filament), a focusing negative element and an accelerating anode, and was made to inject electrons in both directions so that some charge was captured at each half cycle. The target was a tungsten rod mounted on the side diametrically opposed to the gun and on the inside of the equilibrium orbit.

An attempt was made at first to make the gun as small as possible (distance d from electron emission point to edge of gun .016″) in order to minimize its obstruction effects. It was soon observed, however, that the output increased with the voltage far beyond the limit given by the elementary notion by which missing of the gun occurs and the practical limit was the voltage breakdown of the gun. Larger guns ($d = .055''$) supported more voltage and provided more output. An auxiliary experiment was made in this connection. As an obstructing vane protruded into the orbital region for a distance varying up to about d beyond one of these large guns, the output at constant voltage diminished gradually to zero. The exact measurement of the dependence of the output on the voltage of injection was complicated by the

fact that the output is a critical function of the heater current and the grid bias. In spite of these, it was observed: that there is a threshold injection voltage of a few hundred volts below which no output could be detected, there is a general trend of increase of the output with the voltage, and finally that there are large fluctuations superimposed on that trend. Although no definite explanation of these observations is offered, it is probable that the threshold voltage is related to the scattering of slow electrons in their long, low velocity paths and that the general increase of output is due to the deeper potential troughs at higher voltages. An increase of the beam current at a given voltage of injection, resulting from the control of the grid bias or heater current, resulted in an initial increase of the output up to a flattening point, and for low injection voltages, to a maximum beyond which the output seemed actually to decrease. These effects are due probably to the role the space charge is playing in the mechanisms of missing the gun.

The output of the betatron was measured by the radiated X-rays rather than by the difficult direct measure of the small, high speed electron target current. The most sensitive detector was a Geiger-Mueller counter used either with an electronic counter or an oscilloscope. A slightly less sensitive but far more convenient instrument was obtained in irradiating a fluorescent screen with the X-rays and measuring the visible light thus produced with a multiplier phototube. The electrostatic focusing of the multiplier (RCA 931A or 1P22) being spoiled by the strong stray magnetic field near the betatron tube itself, it was necessary to remove the phototube to a distance of about 40 cm. and use an optical system to gather as much light as possible from the screen. A zinc sulfide screen (RCA Code 33-z-20A) with a time constant of about 100 microseconds was found particularly sensitive. Phase measurements of the X-ray bursts were made possible by displaying the multiplier output on an oscilloscope, while an accurate quantitative measurement of the intensity was obtained by a direct measure of the D.C. output current of the multiplier. The multiplier X-ray meter was calibrated with respect to a standard Victoreen Roentgen meter. The Victoreen meter itself was also used for the measurement of the stronger radiations of the betatron at the higher output level.

Absorption measurements of the hardness of the betatron radiation were consistent with the electron energies computed from flux measurements. Energies of 800 KV with an equilibrium orbit at 1.3" to 1.6" in a field of about 1,000 gauss were typical of our small instrument. The highest observed average output current was estimated at .02 microampere on the basis of extrapolation from the results of Van Atta and Northrup giving the yields of X-rays for high speed electrons.

This was obtained with an injection at every half cycle (840 times a second) with a pulsed injection voltage of 3,000 volts from a gun located at $r = 1.9''$ and an equilibrium orbit $r_e = 1.3''$. The limit of capturable space charge in the "normal" potential trough corresponding to those parameters is of the order of 4×10^{-9} coulomb and therefore about 160 times more than the observed amount. This large discrepancy may be due to the fact that most electrons which succeed in missing the gun are actually captured in a much shallower potential trough, as was explained in the theory.

An interesting phenomenon (40) observed with the multiplier phototube detector is the occurrence of multiple—most frequently double—peaks of X-rays in each quarter cycle when ejection was by means of the slow shift of the equilibrium orbit resulting from the saturation of a central core. The relative strength and phase of those peaks depended on the intensity of excitation of the magnet and all parameters of the gun injection. This effect is consistent with the view that the electrons are grouped in definite families according to their emission time and that they oscillate all at different amplitudes which do not diminish by a large proportion in a device with a relatively small final energy so that they can be "skimmed off" by the target at definite phases in the slow shift of their orbits.

An experimental test of a betatron with an additional electric field yielded a negative result. It required, unfortunately, rather elaborate techniques. The glass envelope had to have a re-entrant shape to permit the introduction of the ring shaped cathode-anode structure, shown in Figure 7 which also illustrates the detail of the pole piece construction. The necessary variations of the magnetic flux and field and of the electric field required special electronic circuits. The linearity of the variations of the electric field was obtained by charging and discharging a condenser with a constant current from current regulating tubes while the linear variations of the flux were obtained by applying suddenly, constant voltages derived from a thyratron tube to the almost entirely inductive coils of the control circuit. The variations of the parameters were within tolerances of the best checking instruments which themselves were just about at the limit of the theoretically computed tolerances. It is probable that by varying the parameters near their nominal values, the tube was actually tested under operative conditions. The lack of output resulted most likely from the defects in angular symmetry, which in spite of every effort— machining of the electrodes, careful tube centering and improvements in pole piece construction—remained sufficient for the electrons to ex-

perience some obstruction in their multiple initial swings to and from the cathode.

ACKNOWLEDGMENT

The authors wish to express their appreciation to Dr. V. K. Zworykin for his guiding interest and continued encouragement in their work.

REFERENCES

Articles

1. BREIT, G., TUVE, M. A., AND DAHL, O., Carnegie Institution Year Book (1927-28), No. 27, p. 209.
2. WIDEROE, R., "Über ein neues Prinzip zur Herstellung hoher Spannungen," *Archiv f. Elektrotechnik*, 21, 387-406 (1928).
3. WALTON, E. T. S., "The Production of High-Speed Electrons by Indirect Means," *Proc. Camb. Phil. Soc.*, 25, Part IV, 469-481 (October 1929).
4. JASSINSKY, W. W., *Journal of Experimental and Theoretical Physics, U.S.S.R.*, 5, 983 (1935).
5. JASSINSKY, W. W., *Electricity, U.S.S.R.*, 3, 37 (1936).
6. JASSINSKY, W. W., "Beschleunigung der Elektronen im Elektromagnetischen Weschselstromfeld," *Archiv f. Elektrotechnik*, 30, 590-603 (1936).
7. VAN ATTA, L. C. AND NORTHRUP, "Measurement of X-ray Production," *Am. J. Roentgen. Rad. Ther.*, 41, 633 (1939).
8. KERST, D. W., Letter to the Editor: "Acceleration of Electrons by Magnetic Induction," *Phys. Rev.*, 58, 841 (1940).
9. KERST, D. W., "The Acceleration of Electrons by Magnetic Induction," *Phys. Rev.*, 60, 47-53 (July 1941).
10. KERST, D. W., AND SERBER, R., "Electronic Orbits in the Induction Accelerator," *Phys. Rev.*, 60, 53-58 (July 1941).
11. KERST, D. W., Abstract: "Induction Electron Accelerator," *Phys. Rev.*, 59, 110 (1941).
12. TERLETZKY, J. P., *Journal of Experimental and Theoretical Physics, U.S.S.R.*, 11, 96 (1941).
13. KERST, D. W., "A 20-Million Electron-Volt Betatron or Induction Accelerator," *Rev. Sci. Inst.*, 13, 387-394 (1942).
14. KERST, D. W., "The Betatron," *American Journal of Physics*, 10, No. 5, 219-224 (October 1942).
15. KERST, D. W., "The Betatron," *Radiology*, 40, 115-119 (1943).
16. KOCH, H. W., KERST, D. W., AND MORRISON, P., "Experimental Depth Dose for 5, 10, 15 and 20 Million Volt X-rays," *Radiology*, 40, 120-127 (1943).
17. STEENBECK, M., "Beschleunigung von Elektronen durch elektrische Wirbelfelder," *Naturwiss.*, 31, 234-235 (May 7, 1943).
18. IWANENKO, D., AND POMERANCHUK, I., *C. R. Acad. Sci., U.S.S.R.*, 44, 315 (1944).
19. IWANENKO, D., AND POMERANCHUK, I., *Phys. Rev.*, 65, 343 (1944).
20. WANG, T. J., "The Betatron," *Electronics* (June 1945).
21. VEKSLER, V., "A New Method of Acceleration of Relativistic Particles," *Journal of Physics, U.S.S.R.*, IX, No. 3, 153 (1945).
22. TERLETZKY, J., "On the Stability of the Electron Motion in the Induction Accelerator of the Betatron Type," *Journal of Physics, U.S.S.R.*, IX, No. 3, 159 (1945).
23. WESTENDORP, W. F., AND CHARLTON, E. E., "A 100-Million Volt Induction Accelerator," *J. Appl. Physics*, 16, 591 (1945).
24. WESTENDORP, W. F., *J. Appl. Physics*, 16, 657 (1945).
25. McMILLAN, EDWIN M., "The Synchrotron, a Proposed High Energy Particle Accelerator," *Phys. Rev.*, 68, 143 (Sept. 1945).

26. McMILLAN, EDWIN M., "Radiation from a Group of Electrons Moving in a Circular Orbit," *Phys. Rev.*, 68, 144 (Sept. 1945).
27. BLEWETT, JOHN P., "Radiation Losses in the Electron Induction Accelerator," *Phys. Rev.*, 69, 87 (1946).
28. ARZIMOVICH, L., AND POMERANCHUK, I., "The Radiation of Fast Electrons in the Magnetic Field," *Jour. of Phys. of the U.S.S.R.*, IX, 267 (1945).
29. KERST, D. W., "Historical Development of the Betatron," *Nature*, 157, 90 (1946).
30. SCHIFF, L. I., "Production of Particle Energies Beyond 200 Mev," *Rev. of Sci. Inst.*, 17, 6 (1946).

U. S. Patents

31. SLEPIAN, J., 1,645,304 (Oct. 11, 1927), X-ray Tube.
32. STEENBECK, M., 2,103.303 (Dec. 28, 1937), Device for Producing Electron Rays of High Energy.
33. SMITH, C. G., 2,143,459 (Jan. 10, 1939), Electron Discharge Device.
34. PENNY, G. W., 2,193.602 (March 12, 1940), Device for Accelerating Electrons to very High Speeds.
35. KERST, D. W., 2,297,305 (Sept. 29, 1942), Magnetic Induction Accelerator.
36. BALDWIN, G. C., 2,331,768 (Oct. 12, 1943), Magnetic Induction Accelerator.
37. KERST, D. W., 2,335,014 (Nov. 23, 1943), Magnetic Induction Accelerator.

Notes added in proof reading

38. The successful ejection of a well defined electron beam has recently been announced. See: Skaggs, L. S., Almy, S. M., Kerst, D. W., and Lanze, L. H., Letter to the Editor, "Removal of the Electron Beam from the Betatron," *Phys. Rev.* 70, 95 (July 1946).
39. SCHWINGER, J., Abstract, "Electron Radiation in High Energy Accelerators," *Phys. Rev.* 70, 798 (Nov. 1946).
40. Similar observations made on the G.E. 70 Mev Synchrotron were announced by H. C. Pollock, R. V. Langmuir and F. R. Elder, at the American Physical Soiety Meeting on Feb. 1, 1947, in New York City.

THE MOTION OF ELECTRONS SUBJECT TO FORCES TRANSVERSE TO A UNIFORM MAGNETIC FIELD*†

By

PAUL K. WEIMER AND ALBERT ROSE

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—The paths of electrons in a uniform magnetic field, under the influence of forces transverse to the magnetic field, are of interest in a variety of vacuum tubes. In general, the force experienced by the electron varies with time. The time variation may arise from motion of the electron through an electrostatic field, from motion of the electron along the lines of a curved magnetic field (inertial forces), or from the deliberate application of a time-varying electric field. A graphical method for obtaining the electron paths is described as follows; the given transverse-field versus time curve is approximated by tangent sections; the complete analytic solution is obtained for any tangent section; the analytic solution is interpreted graphically; and a method for joining graphical solutions of neighboring sections is developed. Emphasis is placed on the resultant velocity components and displacement after the field has ceased to act. The graphical method is used to analyze the motion of electrons in the orthicon and image orthicon, television pickup tubes in which the velocity components of the scanning beam critically affect performance. Problems considered are: magnetic and electrostatic deflection in a magnetic field, an electrostatic lens immersed in a magnetic field, the effect of "tapering" the applied forces, and the possibility of canceling unwanted velocity components introduced in one part of the tube by equal and opposite components introduced in another part.

I. INTRODUCTION

A UNIFORM magnetic field has been used for controlling the paths of electrons in several types of television pickup tubes.[1][3] In these tubes the motion of the electron is sufficiently constrained by the field that the resulting path lies approximately along a magnetic line. The uniform field is particularly useful in the orthicon[2] and image orthicon[3] where the electron beam approaches the target with an energy of several hundred volts and must be decelerated to

---

[1] P. T. Farnsworth, "Television by electron image scanning," *Jour. Frank. Inst.*, Vol. 218, pp. 411-444; October, 1934.
[2] A. Rose and H. Iams, "Television pickup tubes using low-velocity electron beam scanning," *Proc. I.R.E.*, Vol. 27, p. 547; September, 1939.
[3] A. Rose, P. K. Weimer, and H. B. Law, "The image orthicon, a sensitive television pickup tube," *Proc. I.R.E.*, Vol. 34, pp. 424-432; July, 1946.

strike the target with less than a volt energy. Care must be taken, however, in the design of the tube that the beam does not acquire velocity components transverse to the magnetic field at the expense of its velocity parallel to the field. If such a transfer of energy does take place the beam will not possess sufficient longitudinal energy to reach the target. The loss of as little as a few tenths of a volt longitudinal energy in the scanning beam of the image orthicon is objectionable.

The necessity for reducing the transverse motion of electrons in the image orthicon has prompted a more general study of helical motion in a uniform magnetic field. The following problems were considered:

1. Motion of an electron along a uniform magnetic field and subject to perturbing transverse electrostatic fields.

2. Motion of an electron along a curved magnetic field.

3. Two-dimensional motion of an electron in a uniform magnetic field and subject to a time-varying transverse electric field.

The solution of the two-dimensional problem specified in 3 may be shown to be an approximate solution of the apparently diverse problems listed in 1 and 2. The approximation involved in this procedure is that the transverse velocity in problems 1 and 2 be sufficiently small compared to the longitudinal velocity that the variations induced in the transverse velocity may be considered to have negligible effect on the longitudinal velocity. Accordingly, a time may be assigned at the outset to each point along the prescribed path of the electron. Thus, in problem 1, the spatial variation of the perturbing electrostatic field along the path may be converted into a time variation of the field, and the solution of 3 applied.

In problem 2 the electron moves along the lines of a curved magnetic field and experiences an inertial force given by mv^2/r where v is the velocity of the electron and r is the radius of curvature of the field lines at each point. The inertial force, to the approximation considered here, is the equivalent of an electric field transverse to a uniform magnetic field. Owing to the motion of the electron and the spatial variation in curvature of the lines, the transverse field experienced by the electron varies with time. By first calculating the transverse field as a function of time the solution of 3 may be applied directly. This method of solving the motion of an electron in a curved magnetic field has been found to give results equal in accuracy to the mathematically more direct but physically less revealing method of obtaining an approximate solution to the three-dimensional equation of motion.[4]

---

[4] Albert Rose, "Electron optics of cylindrical electric and magnetic fields," *Proc. I.R.E.*, Vol. 28, pp. 30-40; January, 1940.
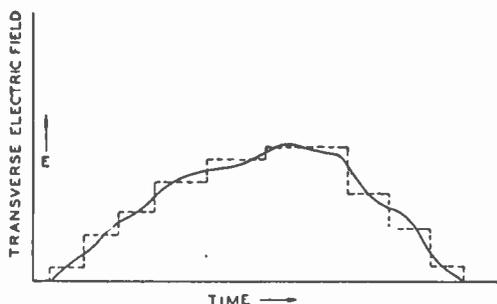
Fig. 1—A stepped approximation to a transverse-field versus time curve.

In general, the transverse electric field may vary with time in any arbitrary manner, as shown by the solid curve of Figure 1. A convenient graphical procedure for solving these problems was developed and will be described below. Application of the graphical method will be made to problems of the types 1 and 2 as they occur in the orthicon and the image orthicon.

II. TWO-DIMENSIONAL MOTION OF AN ELECTRON IN A MAGNETIC FIELD SUBJECT TO A TIME-VARYING TRANSVERSE ELECTRIC FIELD

Starting with the well-known cycloidal motion of an electron in crossed electric and magnetic fields, the curve representing the transverse field as a function of time might well be represented by a series of steps, as shown by the dashed lines of Figure 1. The solution for the first step would provide the initial conditions for the second step, and so on. The objection to this procedure is the number of steps required to attain a prescribed degree of approximation. Greater accuracy can be obtained in fewer steps if the curve of Figure 1 is approximated by straight-line tangents, as shown in Figure 2 (solid lines). Furthermore, the procedure can be greatly simplified if the analytical solutions for the successive linear sections are replaced by a single graphical construction. It has been found that by using
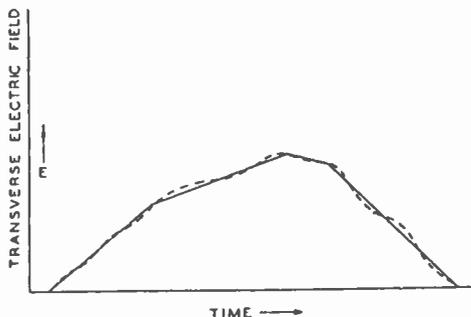
Fig. 2—A tangential approximation to a transverse-field versus time curve.

compasses and a protractor one may quickly find the final phase and magnitude of the circular motion acquired by an electron subjected to a time-varying electric field consisting of many linear sections. (See Figures 5 and 7.) The same construction gives information about the path of the electron *while* the field is acting. By supplementing the construction with a simple formula the complete paths may be plotted if required.

In the following description, the analytic solution of the motion for a single linear section will be derived first to form a basis for the graphical construction.

A. Analytical Solution for a Linear Section

The solution of the equations of motion for a single tangent section (Figure 2) is the solution for an electron in a uniform magnetic field subject to a transverse electric field that varies linearly with time. The equations of motion of an electron for any one section in electromagnetic units are

$$m\ddot{x} = -eH\dot{y}$$
$$m\ddot{y} = -e(E_0 + \dot{E}t) + eH\dot{x} \tag{1}$$

where E and H are directed positively along the Y and Z axes, respectively, in a right-handed co-ordinate system, and all derivatives are with respect to t. Taking the initial conditions for the section considered as $t = 0$, $E = E_0$, $x = x_0$, $y = y_0$, $\dot{x} = \dot{x}_0$, $\dot{y} = \dot{y}_0$, the solution of (1) is

$$x - x_0 = \frac{T}{2\pi} \left[\left(\dot{y}_0 + \frac{m\dot{E}}{eH^2} \right) \cos 2\pi \frac{t}{T} \right.$$
$$\left. + \left(\dot{x}_0 - \frac{E_0}{H} \right) \sin 2\pi \frac{t}{T} - \dot{y}_0 - \frac{m\dot{E}}{eH^2} \right] + \frac{E_0}{H} t + \frac{\dot{E}t^2}{2H} \tag{2}$$

$$y - y_0 = \frac{T}{2\pi} \left[\left(\dot{y}_0 + \frac{m\dot{E}}{eH^2} \right) \sin 2\pi \frac{T}{t} \right.$$
$$\left. - \left(\dot{x}_0 - \frac{E_0}{H} \right) \cos 2\pi \frac{t}{T} + \dot{x}_0 \right] - \frac{m}{eH^2} (E_0 + \dot{E}t) \tag{3}$$

where

$$T = 2\pi \left(\frac{m}{eH} \right). \tag{4}$$

For convenience in applying and demonstrating the graphical construction, these equations will be modified as follows:

1. The independent variable is changed[5] from t to τ where τ is the nondimensional measure of time in units of the electron period T.

2. The origin is chosen so that $x_0 = \dot{y}_0/2\pi$ and $y_0 = -(\dot{x}_0/2\pi)$, where the velocities $\dot{x}_0$ and $\dot{y}_0$ are in centimeters per electron period. This places the origin at the initial center of rotation of the electron in the magnetic field alone.

The complete solution with all derivatives taken with respect to τ, E expressed in volts per centimeter, H in gausses, and with the constants evaluated is

$$x = \frac{1}{2\pi}\left[\left(\dot{y}_0 + \frac{5.69\dot{E}}{H^2}\right)\cos 2\pi\tau + \left(\dot{x}_0 - \frac{35.7E_0}{H^2}\right)\sin 2\pi\tau\right]$$
$$- \frac{5.69\dot{E}}{2\pi H^2} + \frac{35.7}{H^2}\left(E_0\tau + \frac{\dot{E}\tau^2}{2}\right) \qquad (5)$$

$$y = \frac{1}{2\pi}\left[\left(\dot{y}_0 + \frac{5.69\dot{E}}{H^2}\right)\sin 2\pi\tau - \left(\dot{x}_0 - \frac{35.7E_0}{H^2}\right)\cos 2\pi\tau\right]$$
$$- \frac{5.69}{H^2}(E_0 + \dot{E}\tau) \qquad (6)$$

$$\dot{x}(cm/T) = -\left(\dot{y}_0 + \frac{5.69\dot{E}}{H^2}\right)\sin 2\pi\tau +$$

---

[5] The principal advantage of this change of variable is the generalization of the graphical solution. One construction may then apply for more than one value of magnetic field and of beam voltage. Also it provides a convenient measure of electron velocity and period, two quantities which are of inconvenient magnitude when expressed in centimeters per second and seconds, respectively. For example, the velocity of an electron in centimeters per second is

$$v = 5.93 \times 10^7 \sqrt{V}$$

where V is the energy of the beam in volts. When T is taken as the unit of time, the velocity in centimeters per electron period is

$$v = \frac{21.2\sqrt{V}}{H}$$

where H is the magnetic field in gausses. Thus, in the new units the velocity of the beam is numerically equal to the distance between successive nodes in the beam. This distance in the image orthicon is about 4 centimeters for the scanning beam. Another advantage of the change of variable is the simple relation between transverse velocity and the radius of the motion in the magnetic field alone. This is

$$v = 2\pi R.$$

$$+ \left(\dot{x}_0 - \frac{35.7E_0}{H^2} \right) \cos 2\pi\tau + \frac{35.7}{H^2} (E_0 + \dot{E}\tau) \qquad (7)$$

$$\dot{y}(cm/T) = \left(\dot{y}_0 + \frac{5.69\dot{E}}{H^2} \right) \cos 2\pi\tau + \left(\dot{x}_0 - \frac{35.7E_0}{H^2} \right) \sin 2\pi\tau - \frac{5.69\dot{E}}{H^2}. \quad (8)$$

Inspection of these equations shows that the actual motion of the electron is the resultant of two simpler motions. The harmonic terms indicate that one motion consists of a uniform velocity in a circle of constant radius R where

$$R = \frac{1}{2\pi} \left[\left(\dot{x}_0 - \frac{35.7E_0}{H^2} \right)^2 + \left(\dot{y}_0 + \frac{5.69\dot{E}}{H^2} \right)^2 \right]^{1/2}. \qquad (9)$$

The other motion is that of the center of this circle. It follows a parabolic path in space resulting from a constant velocity in the y direction given by $\dot{y} = -(5.69\dot{E}/H^2)$, and a uniformly accelerated velocity in the x direction given by $\dot{x} = (35.7/H^2)(E_0 + \dot{E}\tau)$. The breakdown of the actual motion into circular and translational components greatly simplifies the problem.

B. Graphical Method for Calculating the Circular Motion

The aim of this section is to derive from the above equations a graphical construction which will yield the magnitude and direction of circular motion that an electron has after being acted upon by an electric field whose variation with time has been approximated by linear sections. In the constructions which follow, the instantaneous circular motion is represented by a radius vector drawn outward from the center of the circle to the electron. The circumferential velocity is perpendicular to this vector and has a magnitude proportional to the length of the vector. In (5) to (9) the velocity is expressed in centimeters per electron period and is very simply related to the length of the corresponding radius vector by the proportionality factor 2π. The magnetic field is always assumed to be directed out of the paper and the radius vectors rotate counterclockwise through 360 degrees in the time interval $\tau = 1$.

The final construction arrived at will become clear if a complete picture of the electron path is first considered. By way of illustration, the simple transverse-field versus time curve shown in Figure 3 will be treated. The initial conditions are taken to be $\dot{x}_0 = \dot{y}_0 = E_0 = 0$. The origin of the co-ordinate system for (5) and (6) is at the center of the circle in which the electron moves at $\tau = 0$ by virtue of its initial
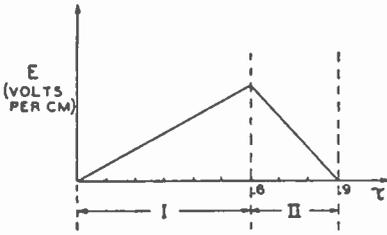
E
(VOLTS
PER CM)

I — II —

τ

velocity. In this instance, the initial velocity is zero and the origin and the initial electron position coincide. From (5) and (6) the complete path is made up of two parts, motion of the electron on a circle and motion of the center of the circle. It is convenient for purposes of the final method of construction to consider the constant term $5.69\dot{E}/2\pi H^2$ of (5) independently of the other terms representing the motion of the center of the circle. The remaining non-periodic terms are called the "translational terms" and are plotted in Figure 4 as the dotted curve C C' C''. The constant term of (5), defines a point X_1 displaced by the constant amount $-(5.69\dot{E}_1/2\pi H^2)$ from the dotted curve. ($\dot{E}_1$ is the slope of the first section of the force field when E is plotted against τ.) The center of the circle while the field is changing at the rate $\dot{E}_1$ is at X_1 and moves along a parabolic path displaced at the constant distance X_1C from the dotted curve. The electron itself is on the periphery of the circle and rotates around the center of the circle as the center slides along its parabolic path. The circular motion of the electron is represented by the rotation of the radius vector X_1C about its center X_1. The complete path of the electron is shown as the solid curve. Three positions, initial, intermediate, and final, of the rotating vector in the first section are shown. At R_1'' the rotating vector, by Figure 3, has completed 0.6 of a revolution. At R_1'', also, the electron has a total velocity, given by (7) and (8), which constitutes the initial velocity for the second section. It may be shown from (7) and (8) that the radius vector giving the total velocity of the electron at R_1'' is $R_1''O_{11}$. With O_{11} as the new origin,
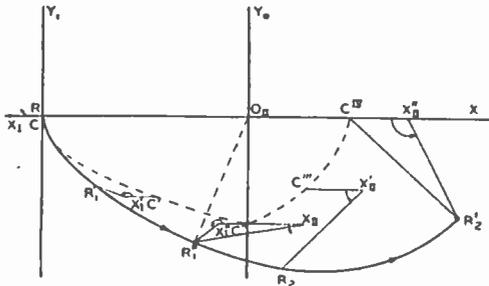
the translational terms of (5) and (6) furnish the parabolic curve $C^{II}C^{III}C^{IV}$. The center X_{II} of the new circle is displaced from the curve by the constant term $-(5.69\dot{E}_2/2\pi H^2)$. In this section $\dot{E}_2$ is negative and the displacement is to the right. The radius vector of the new circle must have a magnitude and direction such as to make the electron position (as well as total velocity) continuous across the boundary of the two sections. This means that it must extend from its center at X_{II} to the final position of the electron at the end of the first section, namely, $R_1{''}$. Again, the radius vector $X_{II}R_1{''}$ rotates about its center X_{II}, as its center slides along a parabolic curve displaced from the dotted curve. The total rotation, according to Figure 3,
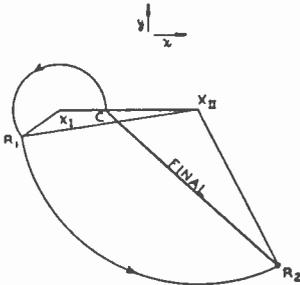
Fig. 5—The graphical construction giving the circular motion produced by the transverse field of Fig. 3. The radius vector CR_2 indicates the phase and magnitude of the motion at the instant the electric field ceases to act.

is 0.3 of a revolution, and the final position is $X_{II}{''}R_2{'}$. At $R_2{'}$ the force field has ceased to act and the electron continues to rotate about the center C^{IV}, its motion being described by the radius vector $C^{IV}R_2{'}$.

Figure 5 shows the simple graphical construction required to obtain the same information about the final circular motion as found in Figure 4. The translational terms have been dropped since they may be conveniently treated independently. The point C represents the center of rotation of the electron before and after the application of the transverse force. During the application of the force this point follows the parabolic paths indicated by the dotted curve of Figure 4, and its total translation may be readily calculated as described in the following section.

Figure 7 is the graphical construction applied to the more complex force field of Figure 6. For generality it is assumed that the electron has an initial velocity represented by the radius vector CR_0. The fulcrum of the rotating vector in the first section is displaced from C a distance $-(5.69\dot{E}_1/2\pi H^2)$ and CX_I is drawn proportional to this distance. The initial position of the rotating vector is then X_1R_0. During the time interval of section I, given by $\tau = 0.2$, X_1R_0 rotates through 0.2 of a revolution or 72 degrees. At the end of the interval the vector has the position X_1R_1.
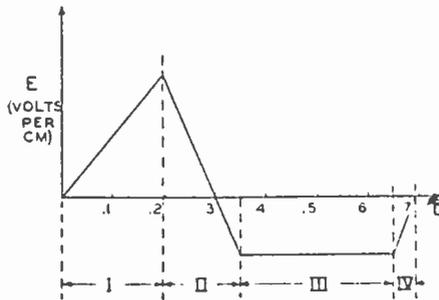
Fig. 6—A transverse-field versus time curve consisting of
four linear sections.

CX_{II} is then drawn proportional to $5.69\dot{E}_2/2\pi H^2$, giving $X_{II}R_1$ as
the initial value of the rotating vector in the second section. $X_{II}R_1$
rotates into $X_{II}R_2$.

In the third section $\dot{E}_3 = 0$ and the center of the rotating vector
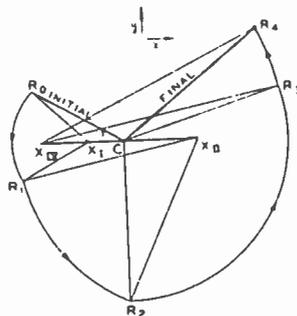is C. The rotating vector is CR_2 and rotates into CR_3.

Fig. 7—The graphical construction for the transverse field of Fig. 6. The
radius vector representing the initial circular motion is CR_o, its magnitude
being drawn proportional to the actual radius in centimeters, determined by

$$R = \frac{3.3\sqrt{V}}{H}$$

where V is the transverse energy of the circular motion in volts. The vectors
CX_I, CX_{II}, and CX_{IV} are drawn in the same proportion to the distance —
$(5.69\dot{E}/2\pi H^2)$ calculated for each section. The final vector CR_t gives the
phase and magnitude of the final circular motion and may be converted to
volts by the relation

$$V = \frac{H^2R^2}{3.3^2}.$$

CX_{IV} is drawn proportional to $-(5.69\dot{E}_4/2\pi H^2)$, locating the fulcrum of the vector in the fourth section. $X_{IV}R_3$ rotates into $X_{IV}R_4$. CR_4 represents the radius vector corresponding to the total motion of the electron at the instant the field ceases to act.

C. Complete Electron Paths in Transverse Fields

Reference to (5) and (6) shows that a transverse force field produces a net translation of the center of rotation along the X axis by an
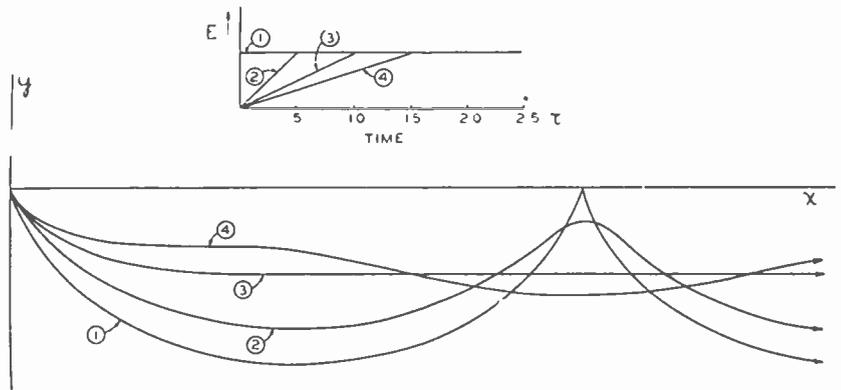
Fig. 8—Electron paths for different rates of application of a transverse electric field. The magnetic field is directed toward the reader.

amount proportional to the area under the curve of E plotted against τ. Thus

$$x = \frac{35.7}{H^2} \int_0^\tau E d\tau. \tag{10}$$

The translation in the y direction parallel to the electric field occurs only while the field is acting, and is proportional to the instantaneous value of E. Thus, after the field has dropped to zero we have

$$y = 0. \tag{11}$$

The complete electron paths while the field is acting may be conveniently plotted by using only the translational terms of (5) and (6) combined with a graphical construction of the type illustrated in Figures 5 and 7.

In Figure 8 the paths between two plates are plotted for several

cases in which the field is applied at different rates. The maximum circular motion occurs in case 1 where sudden application of the field gives the familiar cycloidal paths. As the field is applied more and more gradually the resulting circular motion is reduced. For those particular cases where the time for the field to reach its final value is an integral number of periods, the resultant circular motion between the plates is zero. The fact that an electric field introduces less circular motion when applied gradually is utilized in the orthicon, as will be described in the next section.

It is interesting to note from (5) to (8) that if an electron is introduced to an electric field E_0 in a time of τ electron periods, the amplitude of circular motion during the rise time of E_0 is reduced by the factor $1/2\pi\tau$ and the corresponding energy of circular motion by the factor $1/4\pi^2\tau^2$ relative to the amplitude and energy it would have if introduced suddenly to the same field. This would suggest that the starting electrons in a diode magnetron describe smooth paths concentric with the cathode. The anode field of the magnetron is applied in a time of many electron periods. Pulse rise times of one-half microsecond, for example, in combination with magnetic fields of a thousand gausses attenuate the energy of circular motion by a factor of the order of 10^{-8}.

III. ELECTROSTATIC DEFLECTION PLATES IMMERSED IN A MAGNETIC FIELD

In the 1840-type orthicon, electrostatic plates within the magnetic focusing field are used for the horizontal scanning of the target. The net displacement is parallel to the plane of the plates and can be calculated from (10). The plates are curved so that they are closely spaced at the center and flared out at the ends where the electron enters and leaves the deflecting field. The purpose of this shape is to reduce the helical motion produced in the beam by the transverse field. It is of interest to apply the graphical method to determine the conditions for which the flaring of the plates is advantageous.

The curvature of the plates is such that an electron passing between them at constant velocity experiences a transverse field whose time variation[6] is similar to that shown in Figure 3, except that now the maximum occurs at the midpoint in time. If the plates had been flat and so closely spaced that the fringe field were of negligible extent, the transverse field would, of course, have been constant with an abrupt

---

[6] The periodic variation of the electric field owing to the scanning process is so slow compared to the transit time of the electron that the field may be considered stationary for any one electron.

rise and fall. The helical motion acquired by the electron in both of these cases has been calculated by the graphical method and plotted in Figure 9 for various lengths of plates. It is noted that the curved plates actually introduce more helical motion for the same deflection than the flat plates if the plates are shorter than 1.4λ. (λ is the distance apart of successive nodal planes in the beam.) With longer
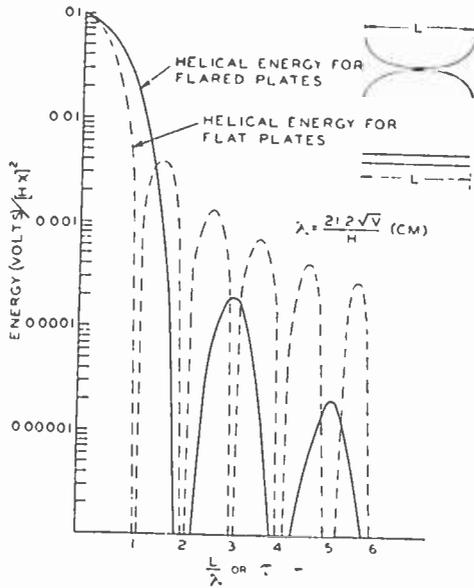

Fig. 9—Helical energy introduced by electrostatic deflection plates plotted as a function of the length of the plates. The abscissa represents the number of orders of focus within the plates and is also equal to the transit time in units of τ. The translation x is perpendicular to E and H.

lengths, both types of plates introduce less helical motion and the advantage in flaring the plates is apparent.

IV.   ELECTROSTATIC LENS IMMERSED IN A MAGNETIC FIELD

The deceleration field in front of the target in the image orthicon forms an electrostatic lens immersed in a uniform magnetic field (see the dotted lines of Figure 10). An electron path of special interest is the one approaching the edge of the target along the magnetic line indicated. Along this path the transverse electric field is larger than for paths near the axis of the tube. From the equipotential plot the transverse-field versus time curve has been calculated for the portion of the path from F to G. This curve is shown in Figure 11.

By use of (10) and Figure 11, the translation of the electron in the deceleration lens may be computed. This translation, at right angles to the electric and magnetic fields, appears as a slight rotation of the scanning pattern as a whole on the target.[7]

The helical motion acquired by the electron has been computed by a graphical construction similar to Figure 7. For the relatively small transverse motions involved here the electric field may be assumed to be uniform across the path. It was found that several volts of helical energy are acquired by an electron deflected toward the edge of the target.
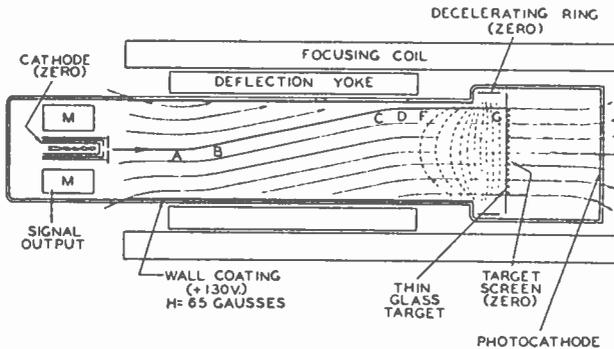
Fig. 10—Cross-sectional diagram of the image orthicon. The dotted lines in front of the target represent the equipotential surfaces of the electrostatic field. The curved lines within the deflection coil indicate the direction of the magnetic field resulting from the combined effects of the deflection and focusing coils. An electron path of particular interest is shown by the heavy line.

V. CURVED MAGNETIC FIELD

A deflection coil immersed in a uniform magnetic field causes the resultant field lines to bend as shown in Figure 10. An electron whose principal motion is along the magnetic lines experiences at each bend an mv^2/r force which may be expressed as an equivalent transverse electrostatic field. The value of this field in volts per centimeter is

$$E = \frac{2V_p}{r} \tag{12}$$

---

[7] This translation is also the chief reason why the electron on its return from the target does not retrace its initial path. The lack of retrace gives rise to a scanning pattern on the first stage of the electron multiplier. This pattern is oriented approximately 90 degrees to the scanning pattern on the target.

where V_p is the longitudinal energy of the beam in volts, and r is the radius of curvature of the magnetic lines at each point. By use of this equation, transverse-field versus time curves may be plotted for the bends AB and CD.

The effect of each bend is to produce a translation perpendicular to the plane of the curve as well as to introduce helical motion. The translation, whose magnitude is given by (10), is usually not significant in the image orthicon since that occurring at the first bend is equal and opposite to that produced by the second. The helical motions, however, introduced by the two bends usually do not cancel. Their magnitudes are not in general equal and their vectorial summation may vary depending on their relative phase as determined by the transit
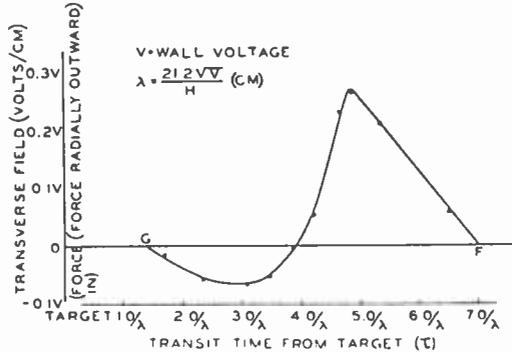
Fig. 11—Plot of transverse field versus time for an electron approaching the target along the path indicated from F to G in Fig. 10.

time between bends. The net helical energy impressed on the electron in passing through the deflection coil varies from zero for zero deflection to as much as several volts for maximum deflection.

Higher wall voltages and decreased magnetic field strengths tend to increase the helical energy acquired by the electron in passing through either the deflection coil or the deceleration lens. In the image orthicon, where zero helical motion at the target is desired, it has been found advantageous to balance the helical motion introduced by the deceleration lens against that produced by the deflection coil. The proper phase for cancellation can readily be obtained by sliding the deflection coil along the axis of the tube.

QUANTUM EFFECTS IN THE INTERACTION OF ELECTRONS WITH HIGH FREQUENCY FIELDS AND THE TRANSITION TO CLASSICAL THEORY*†

BY

LLOYD P. SMITH

Research Department, RCA Laboratories Division, Princeton. N. J.

Summary

The interaction of electrons with an electromagnetic field within a conducting enclosure is treated from the point of view of quantum mechanics. The various quantum effects to be expected with regard to energy exchanges are pointed out and it is shown how the various probabilities of energy change combine to give results in accord with classical theory for conditions where classical theory becomes appropriate. The details in the transition from quantum behavior to classical behavior are traced in a very illuminating way for two particular examples. The conditions under which quantum effects manifest themselves are given.

(15 pages, 12 figures)

CARBIDE STRUCTURES IN CARBURIZED THORIATED-TUNGSTEN FILAMENTS*†

BY

C. W. HORSTING

Tube Department, RCA Victor Division, Harrison. N. J.

Summary

A wide variety of carbide structures is described, as found in the surface layer of carburized thoriated-tungsten filaments. Their origin is traced back to carburizing conditions and subsequent processing during tube making. A frequently-occurring laminated carbide structure is found to contain less carbon than W_2C. The thyratron relay method of carburizing control is critically reviewed. Abnormal filament current in manufactured tubes is explained as due to surface conditions which cause changes in thermal emissivity.

(7 pages, 11 figures, 2 tables)

DETERMINATION OF CURRENT AND DISSIPATION VALUES FOR HIGH-VACUUM RECTIFIER TUBES*†

BY

A. P. KAUZMANN

Tube Department, RCA Victor Division.
Harrison, N. J.

Summary

Rectifier data are shown graphically with generalized parameters from which it is possible to determine the peak steady-state current, the maximum possible hot-switching current, and the dissipations in the diode and in any added series resistors. The paper covers capacitive-input filters with large capacitors and includes half-wave, full-wave, and voltage-doubler circuits. A table of operating conditions and efficiency for a group of typical rectifiers is included.

(16 pages; 5 figures)

GROUNDED-GRID RADIO-FREQUENCY VOLTAGE AMPLIFIERS*†

By

M. C. JONES

RCA Victor Division,
Camden, N. J.

Summary—Triode radio-frequency amplifiers have come into extensive use for medium-high-frequency applications. The use of triodes results from the reduced noise-equivalent resistance of a triode amplifier as compared to a multigrid-type amplifier tube. It is not possible with a triode to use conventional circuits with the input into the grid circuit and the output from the plate circuit because this connection results in excessive output to input feedback which produces regeneration and even oscillation. The grounded-grid amplifier[1] circuit alleviates these difficulties by utilizing the grid as a shield between the input or cathode circuit and the output or plate circuit. Such a circuit exhibits certain peculiarities, particularly when several such stages are operated in tandem. Following is an analysis of the performance of several types of grounded-grid radio-frequency amplifiers.

I. SINGLE STAGE AMPLIFIER

A SINGLE-STAGE amplifier consists of a high-impedance tuned circuit connected to the cathode of the amplifier tube and the load resistance connected in the plate. Such a circuit is shown in Fig. 1.

If we make the assumption that the impedance of the tuned circuit is high compared to the input resistance of the tube, it becomes evident

Fig. 1—Single stage grounded-grid amplifier.

that the signal current I_1 flowing to the cathode of the amplifier is the same as the plate current I_1 flowing into the plate of the amplifier. The tube then acts simply as a medium for transposing the input current, which may be from a low-impedance source, to the output circuit which may be a high-impedance source. From this we see that

---
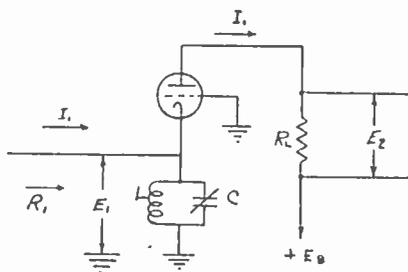
* Decimal Classification: R363.1.
† Reprinted from *Proc. I.R.E.*, July, 1944.
[1] C. E. Strong, "The inverted amplifier," *Electronics*, p. 87; July, 1940.

121

the gain of such a stage is the ratio of the output resistance (R_L) to the input resistance. The value of this input resistance will be determined later.

A marked similarity exists between this type of circuit and the so-called "cathode-follower" type of circuit. In the cathode follower the input and output voltages are identical to a first approximation, and in this circuit the input and output currents are identical. We might then coin the term "voltage-follower circuit" to apply to the cathode-follower and "current-follower circuit" to apply to the grounded-grid amplifier.

By inspection of Fig. 1 it is evident that

$$E_2 = I_1 R_L. \tag{1}$$

In any triode vacuum tube the fundamental frequency component of the signal current may be expressed by

$$I_1 = G_m (E_g + E_p/\mu). \tag{2}$$

Substituting the appropriate values for E_g and E_p in terms of E_1, I_1, and R_L we have,

$$I_1 = G_m [-E_1 + (-I_1 R_L - E_1)/\mu]. \tag{3}$$

Solving for E_1, (3) becomes

$$E_1 = I_1 [(R_p + R_L)/(\mu + 1)]. \tag{4}$$

We may now evaluate the input resistance to the amplifier tube as

$$R_1 = E_1/I_1 = (R_p + R_L)/(\mu + 1). \tag{5}$$

A curve showing R_1 as a function R_L for a typical amplifier tube is shown in Fig. 2. This equation may be simplified, for the condition where $R_L \ll R_p$ and $\mu \gg 1$, to

$$R_1 = 1/G_m. \tag{6}$$

We may now evaluate the gain of the single amplifier by dividing (1) by (4).

$$\text{Gain} = E_2/E_1 = [R_L (\mu + 1)]/(R_p + R_L). \tag{7}$$

$$R_1 = 1/G_m.$$

In the chart:

$$R_i = \frac{R_p + R_L}{\mu + 1}$$

$R_p = 15000 \,\Omega$

$\mu = 30$

$G_m = 2000 \,\mu Mhos.$

y-axis: Input Resistance R_i —Ohms

x-axis: Load Resistance R_L - Ohms.
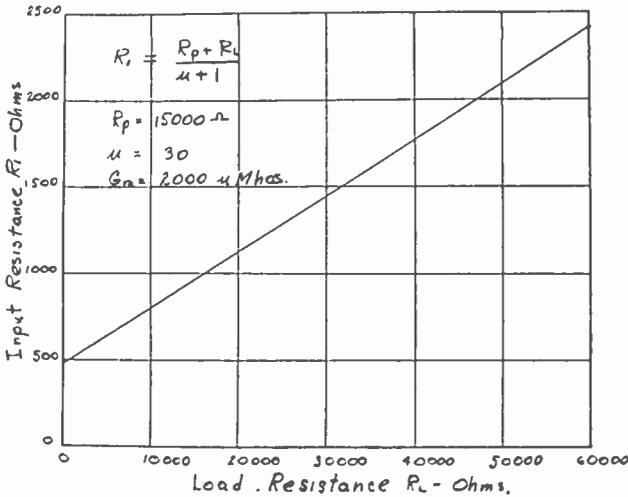
Fig. 2 — Curve showing input resistance as a function of load resistance.

This may be compared with the formula for the gain of an ordinary triode amplifier,

$$\text{gain} = R_L \mu / (R_p + R_L). \tag{8}$$

The slight increase in gain results from the fact that the output voltage is measured to ground which in the case of the grounded-grid amplifier is the grid circuit and in the case of the conventional amplifier is the cathode circuit.

If now we can make the assumption that $R_L \ll R_p$ and $\mu \gg 1$, equation (7) may be simplified to $\text{gain} = R_L G_m.$

II. Tandem Circuits

It is customary to operate two or more grounded-grid amplifiers in tandem. In this case the load resistance (R_L) is some function of the

Fig. 3—Three grounded-grid amplifier circuits in tandem.
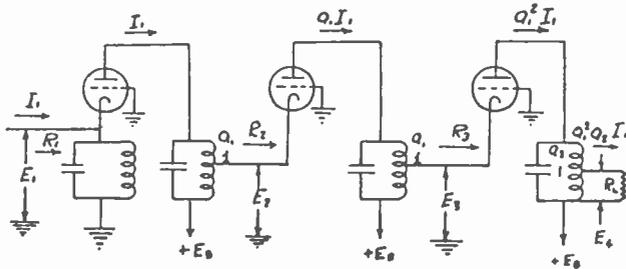
input resistance of the following stage. A circuit showing three such stages terminated finally in the load resistance R_L is shown in Fig. 3.

Before analyzing this circuit we must consider the coupling transformer which reflects the input resistance of each stage back to the load resistance of the preceding stage. For simplicity it is convenient to assume that this is a perfect transformer and that its output-to-input current ratio is the same as its input-to-output voltage ratio. If we define this ratio as a it is then only necessary to multiply the input current by the appropriate values of a through the various stages to determine the output current. In analyzing this circuit the most convenient method of approach is that of determining the input resistance of each stage as a function of the input resistance of the following stage. By reference to (5) we may write equations for the input resistances of the three stages as follows:

$$R_3 = (R_p + a_2{}^2R_L)/(\mu + 1) \tag{9}$$

$$R_2 = (R_p + a_1{}^2R_3)/(\mu + 1) \tag{10}$$

$$R_1 = (R_p + a_1{}^2R_2)/(\mu + 1). \tag{11}$$

We may now solve for the input resistance R_1 as a function of the output resistance, the tube characteristics, and the transformer characteristics:

$$R_1 = R_p/(\mu + 1) + a_1{}^2R_p/(\mu + 1)^2$$
$$+ a_1{}^4(R_p + a_2{}^2R_L)/(\mu + 1)^3. \tag{12}$$

This equation may now be generalized for any number (n) of stages:

$$R_1 = \frac{R_p}{\mu + 1} + \frac{a_1{}^2R_p}{(\mu + 1)^2} + \frac{a_1{}^4R_p}{(\mu + 1)^3} + \cdots$$
$$+ \frac{a_1{}^{2(n-2)} R_p}{(\mu + 1)^{n-1}} + \frac{a_1{}^{2(n-1)} (R_p + a_2{}^2R_L)}{(\mu + 1)^n}. \tag{13}$$

If the amplifier consists of a large number of stages such that the input resistance R_1 of the first stage is relatively independent of the output resistance R_L, or if the value of R_L is selected such that the input resistance of each stage is identical, we then may evaluate what might be termed the "characteristic impedance" for this amplifier by setting $R_1 = R_2$ in (11).

$$R_1 = R_p/(\mu + 1 - a_1^2). \tag{14}$$

From this it may be seen that the characteristic input impedance of a grounded-grid amplifier is determined by the tube characteristics and by the coupling transformer. A curve showing the variation of input resistance with current ratio of the coupling transformer for a typical amplifier tube is shown in Figure 4. From this curve it may be seen that the input resistance rises rapidly with values of a_1 approaching $\sqrt{\mu + 1}$. We may now determine the over-all gain of n grounded-grid amplifier stages as follows:
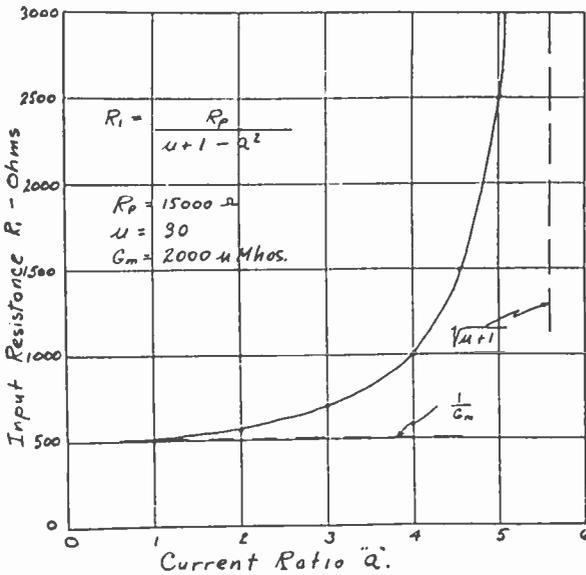
Fig. 4 — Curve showing input resistance of an amplifier terminated in its "characteristic impedance" as a function of a.

$$\text{Output current } (I_n) = a_1^{n-1} a_2 I_1 \tag{15}$$

$$\text{Output voltage } (E_n) = I_n R_L = I_1 R_L a_1^{n-1} a_2. \tag{16}$$

The input voltage E_1 is given by

$$E_1 = I_1 R_1. \tag{17}$$

The over-all gain may be determined by dividing (16) by (17).

$$\text{Over-all gain} = (R_L a_1^{n-1} a_2)/R_1 \tag{18}$$

or substituting the value of R_1 for n stages given in (13) we have,

$$\text{over-all gain} = \frac{R_L a_1{}^{n-1} a_2}{R_p/(\mu+1) + a_1{}^2 R_p/(\mu+1)^2 + \cdots}$$

$$\frac{}{+ a_1{}^{2(n-2)} R_p/(\mu+1)^{n-1} + a_1{}^{2(n-1)} (R_p + a_2{}^2 R_L)/(\mu+1)^n} . \quad (19)$$

If the stages are designed so that they exhibit the "characteristic input impedance" or if a large number of stages is used such that the input resistance is relatively independent of the output resistance R_L, equation (19) may be simplified to

$$\text{over-all gain} = [R_L a_1{}^{(n-2)} a_2 (\mu + 1 - a_1{}^2)]/R_p. \quad (20)$$

If $\mu \gg (1 - a_1{}^2)$, equation (20) may be simplified to

$$\text{over-all gain} = G_m R_L a_1{}^{n-1} a_2. \quad (21)$$

From this equation it is evident that the gain of several grounded-grid amplifier stages in tandem is a function only of the current ratio of the coupling transformers, the G_m of any one tube (assuming all tubes have the same value of G_m), and the load resistance R_L. It will be shown later an optimum value of a exists if we consider the input loading due to transit-time effect.

Fig. 5—π-section coupling trans-formers in two-stage grounded-grid amplifier.
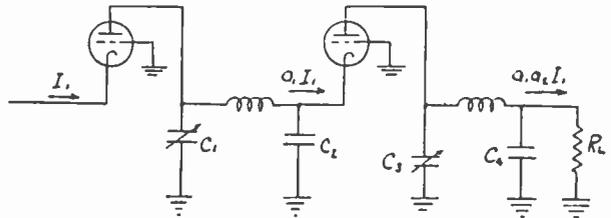
III. π-SECTION TANDEM CIRCUITS

It is sometimes more convenient practically to design the coupling transformers as a π-section filter in the manner shown in Figure 5.

It can be shown that the current ratio of these coupling trans-formers to a fair degree of approximation is given by

$$a_1 = C_2/C_1 \quad (22)$$

$$a_2 = C_4/C_3. \quad (23)$$

From this it may be seen that for large gains C_2 should be greater

than C_1, and C_4 should be greater than C_3. To obtain the maximum range from a given capacitance change in the tuning capacitor it is desirable then to make C_1 and C_3 the variable elements and to use a fixed capacitor for C_2 and C_4.

With the above definitions of a_1 the previous equations for over-all gain (18), (19), (20), (21), apply with equal accuracy to this circuit.

IV. ANTENNA INPUT CIRCUITS

Most amplifiers of this type operate from a low-impedance antenna circuit and feed into a constant impedance, usually the input impedance of a converter. A typical antenna circuit feeding directly into the input resistance of a grounded-grid amplifier is shown in Fig. 6.

If we assume that the loss in the input transformer is negligible and that the input circuit matches the antenna resistance it can be shown that the gain of the antenna circuit is given by
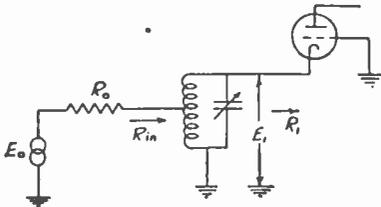
Fig. 6—Antenna input circuit for grounded-grid amplifier.

$$gain = G_1 = E_1/E_0 = 1/2 \ \sqrt{R_1/R_0}. \qquad (24)$$

The over-all gain for n stages plus the above antenna circuit is then given by

$$\text{over-all gain} = E_n/E_0 = ((R_L a_1{}^{n-1} a_2)/R_1)/2 \ \sqrt{R_1/R_0}. \qquad (25)$$

Simplifying the above equation

$$\text{over-all gain} = 1/2 \ (R_L a_1{}^{n-1} a_2)/2 \ \sqrt{R_1 R_0}. \qquad (26)$$

In this equation we may use values of R_1 as determined previously in (13) or (14). If the following approximation is valid,

$$R_1 = 1/G_m. \qquad (27)$$

The formula for over-all gain simplifies to

$$\text{over-all gain} = 1/2 R_L a_1{}^{n-1} a_2 \ \sqrt{G_m/R_0}. \qquad (28)$$

From this it may be seen that the over-all voltage gain depends entirely on the transformer design, the input and output impedances,

and the G_m of the amplifier tubes. An increase of the G_m of all the tubes produces an increase in over-all gain of only the square root of the G_m change of any one stage regardless of the number of stages. Consider a typical two-stage grounded-grid amplifier feeding from an antenna to the input of a converter with the following characteristics:

$G_m = 2000$ micromhos $a_1 = 3$

$R_0 = 50$ ohms $a_2 = 2$

$$R_L = 1000 \text{ ohms}$$

The over-all gain as calculated by (28) then becomes over-all gain $= 1/2 \times 1000 \times 3 \times 2 \sqrt{(2000 \times 10^{-6})/50} = 19$.

V. NOISE CONSIDERATIONS

The performance of any receiver is limited by noise voltages.[2,3] This noise is of two distinct types: Thermal-agitation noise, or noise developed within the circuits, and shot-effect noise; or noise developed in the plate of the amplier tubes. The thermal-agitation noise over any band of frequency can be predicated accurately if the value of circuit resistance is known. The shot-effect noise is more difficult to determine. In general this noise is referred to the grid circuit and expressed as a "noise-equivalent resistance" for the vacuum tube used, and calculations may then be made in the same manner as with thermal-agitation noise.

The noise voltage generated in series with any resistor R as measured by an amplifier having a bandwidth $B.W.$ is given by the expression:

$$E_n^2 = 4KTR \, B.W. \tag{29}$$

where

$K =$ Boltzmann's constant $= 1.37 \times 10^{-23}$ watt second per degree

$T =$ temperature in degrees Kelvin

$R =$ resistance in ohms

$B.W. =$ bandwidth of the amplifier in cycles per second

Since most of our calculations are based on normal or room tem-

[2] F. B. Llewellyn, "A rapid method of estimating the signal-to-noise ratio of a high gain receiver," *Proc. I.R.E.*, Vol. 19, pp. 416-420; March, 1931.
[3] E. W. Herold, "An analysis of the signal-to-noise ratio of ultra-high-frequency receivers," *RCA Review*, Vol. 6; January, 1942.

perature, it is possible to simplify (29) as follows:

$$E_n{}^2 = 1.66 \times 10^{-20} \times B.W. \times R \tag{30}$$

where E_n = the root-mean-square noise voltage in series with the resistance R at room temperature, 30 degrees centigrade, or 303 degrees Kelvin.

Measurement of Noise Factor (NF) by the 4KT Method

The noise factor of a receiver expresses its ability to receive and detect weak signals in the presence of noise generated both in the antenna resistance and the receiver itself. The noise factor may be defined as the ratio of the signal-to-noise ratio of a perfect receiver, to the signal-to-noise ratio of the receiver under test (both measured under the same conditions). A perfect receiver is defined as one with infinite input impedance, and no internal noise. Expressing this in an equation we have

$$NF = \frac{(E_s/E_n)}{(E_s/E_n{}^1)} = \frac{E_n{}^1}{E_n} \tag{31}$$

where,

E_s = signal voltage used to measure both receivers

E_n = equivalent noise voltage of the perfect receiver referred to the input circuit

$E_n{}^1$ = equivalent-noise voltage of the receiver under test referred to the input circuit.

From this it is apparent that the noise factor is the ratio of two noise voltages, one of which may be measured and the other calculated.

In order to measure the equivalent-noise voltage of a receiver, one would proceed as follows:

1. Connect a signal generator, of negligible internal impedance, to the receiver input terminals in series with a resistance equal to the input resistance of the receiver. It should be noted here that the $4KT$ method applies only when the internal impedance of the signal generator is small.

2. Connect a root-mean-square voltmeter to the output of the intermediate-frequency amplifier of the receiver. A direct-current voltmeter connected across the load resistor of second detector may be used with only slight error.

3. With the signal generator output reduced to zero, adjust the gain control until the noise voltage as read on the root-mean-square meter is some convenient value, say 1 volt.

4. Adjust the input voltage from the signal generator until the noise voltage has increased to $\sqrt{2}$ times its original value.

The voltage output from the signal generator is now equal to the equivalent-noise voltage of the receiver referred to the input circuit; or the voltage $E_n{}^1$ in (31).

The equivalent-noise voltage (E_n) of a perfect receiver, with characteristics similar to the receiver under test, is the noise voltage generated by the antenna resistance. This value of resistance is usually equal to the input resistance of the receiver under test. From (30) this voltage is

$$E_n = 1.29 \times 10^{-10} \sqrt{B.W. \times R_0} \tag{32}$$

where

$B.W. =$ bandwidth of receiver under test in cycles per second

$R_0 =$ antenna series resistance in ohms

$E_n =$ equivalent-noise voltage of the perfect receiver.

The noise factor may now be determined as the ratio $E_n{}^1/E_n$.

Since it is not possible to obtain a perfect receiver with infinite input impedance, it is useful to consider what might be called a "practically perfect" receiver with a finite input resistance, but which contains no internal noise sources other than this input resistance. The input circuit of such a receiver connected to a source of signal voltage E_0 is shown in Fig. 7(a). The noise sources and their equivalent noise voltage are shown in Fig. 7(b), and the previous perfect receiver is shown in Fig. 7(c).

Solving for the receiver input voltage E_s'' from Fig. 7(a) we have,

$$E_s'' = E_0 \left[R_{in}/(R_0 + R_{in}) \right]. \tag{33}$$

The noise voltage E_n'' from Fig. 7(b) is given by the expression

$$E_n'' = 1.29 \times 10^{-10} \sqrt{B.W. \left(R_0 R_{in}/(R_0 + R_{in}) \right)}. \tag{34}$$

Performing these same operations on the perfect receiver shown in Fig. 7(c) we have,

$$E_s = E_0 \tag{35}$$

$$E_n = 1.29 \times 10^{-10} \sqrt{B.W. \times R_0.} \tag{36}$$

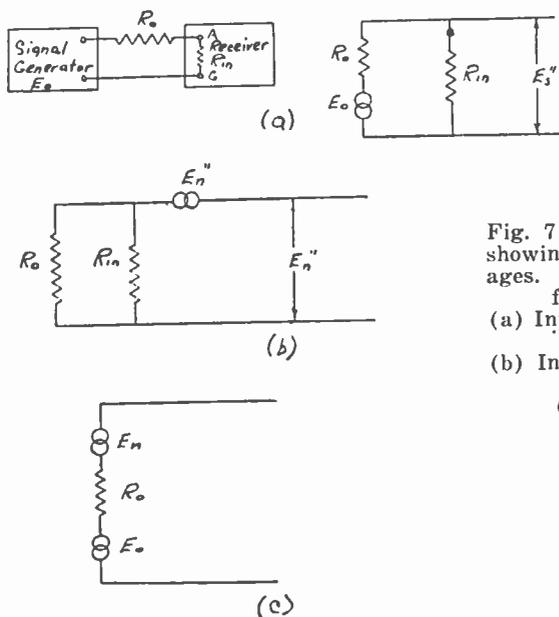If we now consider the practically perfect receiver as one under test we have, from (31),

Fig. 7 — Receiver input circuit showing noise and signal voltages. 4KT method of noise-factor measurement.
(a) Input circuit showing signal voltage
(b) Input circuit showing noise voltage
(c) Perfect receiver

$$NF = \frac{(E_s/E_n)}{(E_s''/L_n'')} = \frac{(E_0/(1.29 \times 10^{-10} \sqrt{B.W. \times R_0}))}{\dfrac{E_0 (R_{in}/(R_0 + R_{in}))}{1.29 \times 10^{-10} \sqrt{B.W. \times R_0} R_{in}/(R_0 + R_{in})}}$$

$$= \sqrt{\frac{R_0 + R_{in}}{R_{in}}} = \sqrt{\frac{1 + R_{in}/R_0}{R_{in}/R_0}}. \tag{37}$$

The curve in Fig. 8 shows the noise factor as a function of the ratio R_{in}/R_0. From this it may be seen that the noise factor of a matched receiver cannot exceed $\sqrt{2}$, or 3 decibels from thermal noise, even though there are no noise sources within the receiver except the input resistance. A slight advantage may be gained by mismatching the receiver in the direction of making R_{in} greater than R_0. This advantage

is more theoretical than practical, however, since mismatch always introduces additional losses due to reflections on the transmission line connecting the receiver with the antenna. It also tends to make the performance of the receiver a function of the transmission-line length since the transmission line is then a resonant circuit. At lower frequencies where the transmission line is short compared with a quarter wavelength, some advantage may be gained by mismatching the antenna circuit.

Measurement of Noise Factor by the 1KT Method

In the preceding analysis the signal-generator voltage (E_0) was connected in series with the antenna resistance (R_0), and as a result,
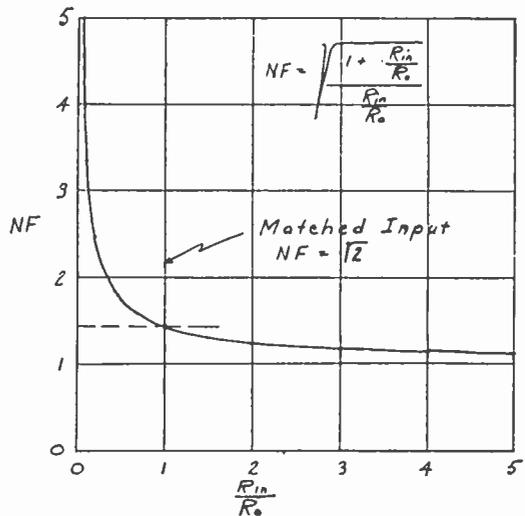
Fig. 8—Curve showing the noise factor as a function of receiver input resistance, for a receiver with no internal noise (4KT basis).

$$NF = \sqrt{\frac{1 + \frac{R_{in}}{R_0}}{\frac{R_{in}}{R_0}}}$$

$$\text{Matched Input} \quad NF = \sqrt{2}$$

the thermal-agitation noise was calculated on the so-called $4KT$ basis. In some signal generators, however, the internal resistance of the generator itself is equal to the antenna resistance, and the voltage as read by the signal generator is then equal to the voltage impressed on the receiver input terminals. A schematic diagram of such an arrangement, together with its equivalent circuits, is shown in Fig. 9.

With this arrangement the noise voltage must be re-evaluated so that it compares with the readings of voltage from the signal generator. The signal-generator voltage reading is now half its value in the preceding connection shown in Fig. 7(a) and the noise voltage must be reduced by a factor of 2 for a direct comparison with it. This gives rise to an expression for noise voltage which is not strictly accurate, but which gives proper results when compared with the equivalent-

noise voltage of the receiver shown in Fig. 9 measured by the method previously outlined. This expression for noise voltage is

$$(E_n'')^2 = 1KT \, R_0 \, B.W. \tag{38}$$

At $T = 303$ degrees Kelvin (30 degrees centigrade) this becomes

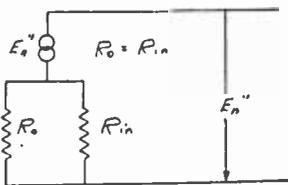$$(E_n'')^2 = 0.415 \times 10^{-20} \times B.W. \times R_0 \tag{39}$$

or

$$E_n'' = 0.645 \times 10^{-10} \sqrt{B.W. \times R_0} \tag{40}$$

This method of measuring noise factor is based on the so-called $1KT$ noise level. For any given receiver the noise factor as determined
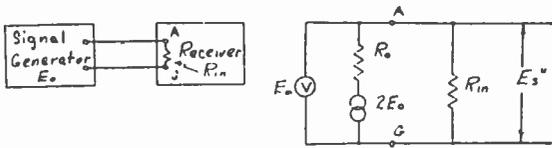
(a)

(b)

Fig. 9 — Receiver input circuit with matched impedance signal generator. $1KT$ method of noise-factor measurements.
(a) Input circuit showing signal voltage
(b) Input circuit showing noise voltage

by either of these methods should be identical. The method used depends on the internal resistance of the signal generator. In actual practice, the method used may quite often be a compromise between the $1KT$ and the $4KT$ methods. This condition occurs when the internal resistance of the signal generator is less than the input resistance of the receiver, but is still not negligible. By any method of measurement, however, a matched receiver with no internal noise has a theoretical maximum performance of 3 decibels from thermal noise.

Calculation of Receiver Noise and Noise Factor

The noise generated in the plate of a vacuum tube may be conveniently expressed as the noise-equivalent resistance of the tube

referred to the grid circuit. This is considered to be the resistance connected in series with the grid of a perfect amplifier tube having the same gain as the tube under consideration and generating the noise normally generated in the plate of the tube.

If one makes the assumption that the circuit noise in a radio-frequency amplifier is negligible compared with the noise generated in the antenna resistance, the input resistance of the receiver, and the shot-effect noise of the radio-frequency amplifier tubes, one may calculate the noise factor of the radio-frequency amplifier from a knowledge of its characteristics.

The noise-equivalent resistance of the first stage may be referred to the input circuit by dividing its value by the square of the gain of the antenna circuit. If additional stages, other than the first stage, contribute to the noise-equivalent voltage of the receiver, they may be taken into account by dividing the noise-equivalent resistance of each stage by the square of the total gain from the input circuit to the grid of the stage. Fig. 10 represents the input circuit to a receiver

Fig. 10—Noise sources within a receiver.
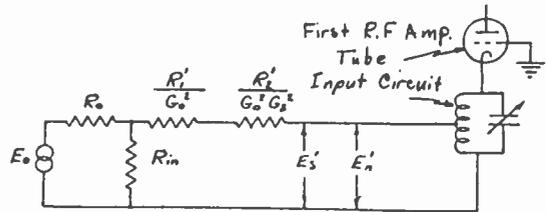
and the noise-equivalent resistances of the first and second radio-frequency amplifier tubes.

Since the noise-equivalent resistances and the parallel equivalent of the input and antenna resistance may be added numerically, the total noise voltage referred to the input of the amplifier is given by

$$E_n' =$$

$$1.29 \times 10^{-10}\sqrt{B.W. \; [R_0 R_{in}/(R_0 + R_{in}) + R_1'/G_0^2 + R_2'/G_0^2 G_2^2]} \quad (41)$$

where

$R_1' =$ noise-equivalent resistance of first radio-frequency amplifier tube

$R_2' =$ noise-equivalent resistance of second radio-frequency amplifier tube

$G_0 =$ antenna circuit gain measured from receiver input terminals to input of the first radio-frequency amplifier tube

$G_2 =$ gain of first radio-frequency amplifier stage.

The signal voltage with which this noise voltage should be compared is given by (33)

$$E_s' = E_0 \left[R_{in}/(R_0 + R_{in}) \right]. \tag{33}$$

The signal-to-noise ratio of this receiver is then,

$$\frac{E_s'}{E_n{}^1} =$$

$$\frac{E_0 \left[R_{in}/(R_0 + R_{in}) \right]}{1.29 \times 10^{-10} \sqrt{B.W.} \left[(R_0 R_{in}/(R_0 + R_{in})) + R_1' G_0{}^2 + R_2'/G_0{}^2 G_2{}^2 \right]}. \tag{42}$$

The signal-to-noise ratio of a perfect receiver, as shown in Figure 7(c) is,

$$\frac{E_s}{E_n} = \frac{E_0}{1.29 \times 10^{-10} \sqrt{B.W. \, R_0}}. \tag{43}$$

The noise factor of the receiver is obtained by dividing (43) by (42).

$$NF = \frac{(E_s/E_n)}{(E_s'/E_n')} = \frac{1/\sqrt{R_0}}{\dfrac{(R_{in}/(R_0 + R_{in}))}{\sqrt{R_0 R_{in}/(R_0 + R_{in}) + R_1'/G_0{}^2 + R_2'/G_0{}^2 G_2{}^2}}}$$

$$= \frac{R_0 + R_{in}}{R_{in}} \sqrt{\frac{R_{in}}{R_0 + R_{in}} + \frac{R_1'}{R_0 G_0{}^2} + \frac{R_2'}{R_0 G_0{}^2 G_2{}^2}}. \tag{44}$$

This equation simplifies to (37) when R_1', R_2', etc., $= 0$.

It should again be noted that the value for G_0, the antenna circuit gain, is now different from the value G_1 previously calculated in (24). Since our reference point for signal-to-noise calculations is the input circuit of the receiver, the gain must be measured from the input terminals to the grid or cathode of the first tube as shown in Figure 11.

This gain (G_0) is greater than the gain (G_1) calculated by (24) and is given by

$$G_0 = \text{Gain} = E_1/E_0' = \sqrt{R_1/R_0}. \tag{45}$$

Since the antenna-gain equations (24) and (45) assume matched-input conditions,

$$G_0 = 2G_1. \tag{46}$$

Simplifying (44) for matched-input conditions and substituting the generally accepted value for antenna gain G_1, as given by (24),

$$NF = \sqrt{2 + R_1'/Z_0G_1{}^2 + R_2'/Z_0G_1{}^2G_2{}^2 + \cdots}. \tag{47}$$

Consider now a grounded-grid amplifier circuit with several stages having the following characteristics:

$$R_0 = 50 \text{ ohms}$$

$$R_{in} = 50 \text{ ohms}.$$

Input resistance of first tube $= 500$ ohms

$G_1 = 1/2 \sqrt{500/50} = 1.58$

$G_2, G_3, G_4, \cdots = 3 \ (a = 3.3)$

$R_1', R_2', R_3', \cdots = 300$ ohms.

Calculating the noise factor by (45) we have,
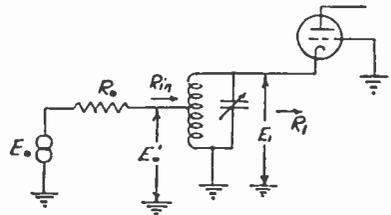
Fig. 11—Antenna input circuit.

$$NF = \sqrt{2 + 2.4 + 0.27 + 0.03 + \cdots} \tag{48}$$

$= 7$ decibels from thermal noise.

From this it may be seen that the principal source of noise in this amplifier is the shot-effect noise of the first radio-frequency amplifier tube. The following stages contribute a progressively smaller amount of noise and the noise contributed by the third stage is insignificant. The converter might take the place of the third stage. If we assume the noise-equivalent resistance of the converter to be 1000 ohms and recalculate the noise factor for a two-stage radio-frequency amplifier followed by the converter, we obtain the following results:

$$NF = \sqrt{2 + 2.4 + 0.27 + 0.1}$$

$= 7$ decibels from thermal noise. \tag{49}

VI. COUPLING TRANSFORMER DESIGNS AND OPTIMUM VALUES FOR R_L

From the previous analysis it appears that the gain increases without limit as the value of a for the coupling transformer increases. In practice, certain physical limitations produce an optimum value of a which should be used in any grounded-grid amplifier to obtain maximum gain. The most serious limitation is that imposed by a component of input resistance resulting from the transit-time effect.[4] Figure 12 shows a single-stage amplifier with this component of input resistance (R_x) in parallel with the normal input resistance R_1.

It is convenient to analyze this amplifier on the basis of power rather than voltage gain since the cathode and plate current are common. This power gain then becomes the ratio of output to total input resistance. Solving for power gain for the circuit shown in Fig 12 we have
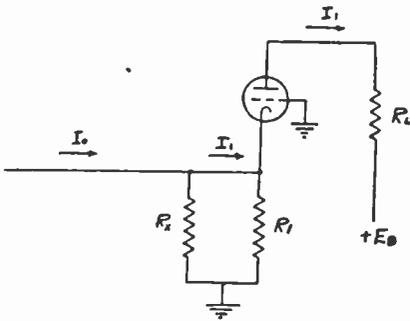
$$W_{in} = I_1^2 R_1 (R_1 + R_x)/R_x \quad (50)$$

$$W_{out} = I_1^2 R_L. \quad (51)$$

$$\text{power gain} = R_1 R_x/(R_1 + R_1 R_x). \quad (52)$$

Fig. 12 — Single-stage grounded-grid amplifier considering transit-time effect.

For any single-stage amplifier,

$$R_1 = (R_p + R_L)/(\mu + 1) \quad (53)$$

Substituting this value of R_1 in (52) and simplifying we have

power gain $=$

$$\frac{R_1 R_x (\mu + 1)^2}{R_L^2 + R_L [2R_p^2 + R_x (\mu + 1)] + R_p^2 + R_x R_p (\mu + 1)}. \quad (54)$$

By taking the derivative of this expression and setting it equal to zero we may determine the value of R_L which produces the maximum power gain. Since we may design a transformer with fairly high efficiency to match the output of this stage into input of another stage this maximum power gain then corresponds to a maximum voltage gain into any constant impedance. This optimum value for R_L is,

---

[4] W. R. Ferris, "Input resistance of vacuum tubes as ultra-high-frequency amplifiers," *Proc. I.R.E.*, Vol. 24, pp. 82-105; January, 1936.

$$R_L = R_p \sqrt{1 + [R_x (\mu + 1)/R_p]}. \tag{55}$$

For conditions where μ is considerably greater than 1, (55) simplifies to

$$R_L = R_p \sqrt{1 + R_x G_m}. \tag{56}$$

If we now consider the coupling transformer between this stage and the next stage whose input resistance is R_2 we have the expression

$$R_L = a^2 R_2. \tag{57}$$

From this we may determine an optimum value for a as follows:

$$a^2 = R_p/R_2 \sqrt{1 + R_x G_m}. \tag{58}$$

If the input resistance of the next stage is the "characteristic impedance" as given by (14) the value for a may be written in the form

$$a^2 = (\mu + 1) \sqrt{1 + G_m G_x}/(1 + \sqrt{1 + G_m R_x}). \tag{59}$$

At extremely high frequencies the value of R_x decreases rapidly. If the value of $G_m R_x$ is considerably less than unity, (59) simplifies to

$$a = \sqrt{\mu/2}.$$

For the amplifier considered in Fig. 2, the value for a under this condition is approximately 3.9. In practice other limitations are present such that a value slightly less than this is often used. For normal operating conditions the value for a of approximately 3 represents the practical optimum.

EXCESS NOISE IN CAVITY MAGNETRONS*†‡

By

ROBERT L. SPROULL

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—The magnetron oscillator of some frequency-modulation radars also furnishes the "local oscillator" excitation for the crystal mixer of the receiver. Excess noise generation by the magnetron was observed to reduce greatly the receiver sensitivity. This noise exhibited a strong dependence upon anode voltage and current, and changed with time in a perplexing manner. After many experiments, a hypothesis of the cause of excess noise was developed, and further experiments confirmed this hypothesis. The noise is thought to be caused by ionization of atoms of the cathode oxide coating, which atoms are removed from the cathode by electron bombardment. In order to reduce the generation of excess noise, and still preserve the advantages of an oxide-coated cathode, a special shape of cathode has been developed. The coated regions of this cathode are sheltered from electron bombardment, and the noise is much reduced. Use of cathodes of this general type should also produce magnetrons with longer operating lives than present tubes.

THE magnetron oscillator in some[1] frequency-modulation, continuous-wave radars also furnishes the "local oscillator" excitation of the crystal mixer of the receiver. It was observed that the magnetron oscillations were frequently modulated with large amounts of noise, thereby impairing the sensitivity of the receiver. The cause of this excessive noise modulation (many times ordinary shot noise) was made the subject of a special investigation, the results of which are reported here. Excess noise is not noticed in the usual use of a magnetron as a transmitter tube, since the signal-to-noise ratio is at least 60 db. However, the noise phenomena may increase our understanding of magnetron operation, are important in special applications of magnetrons, and are of interest in any development of receiving tubes utilizing magnetic fields.

I. DESCRIPTION OF THE NOISE PHENOMENA

Figure 1 shows the experimental arrangement used to detect

---

[1] Such as the AN/APQ-19, developed in this laboratory.

noise modulation and to estimate its magnitude. Since the magnitude of the oscillating field in the crystal cavity was kept constant by varying the attenuation between the magnetron and this cavity, the receiver detector current "I_n" was roughly inversely proportional to the signal-to-noise ratio of the magnetron oscillations.

Figure 2 shows the "noise current" I_n as a function of magnetron d.c. anode voltage (V_a) for several anode currents (I_a) in a 4000 megacycle, continuous-wave magnetron[2] developed in this laboratory. To obtain such a plot, the magnetic field was varied to produce different values of V_a at constant I_a. It is apparent that at low voltages the total noise current is substantially constant, and in this region the con-
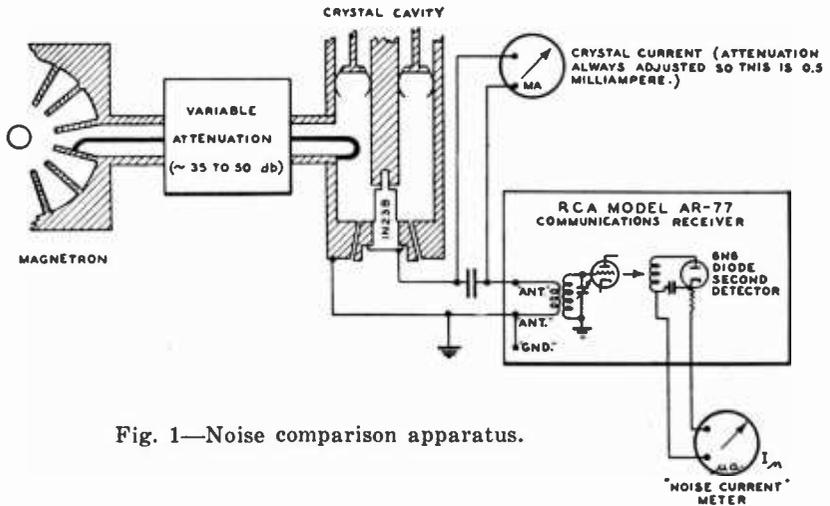
Fig. 1—Noise comparison apparatus.

tribution of the magnetron noise to the total receiver noise is small. At some voltage, which we shall call V_{ao}, a relatively sharp break in the curve appears, and above this voltage the magnetron noise dominates noise from other sources and becomes hundreds or thousands of times as large as at low voltages. This region ($V_a > V_{ao}$) of excess noise contained the ordinary operating point of the oscillator. Phenomena very similar to those portrayed in Fig. 2 were also observed in a magnetron[3] of quite different design.

Curves such as those of Fig. 2 are not reproducible in detail. The "peaks" and "valleys" of the curves cannot be repeated even a few

[2] G. R. Kilgore, C. Shulman, and J. Kurshan, "A Frequency-Modulated Magnetron for Super-High-Frequencies," *Proc. I.R.E.*, Vol. 35, No. 7, July, 1947.

[3] J. S. Donal, Jr., R. R. Bush, C. L. Cuccia, and H. R. Hegbar, "A One-Kilowatt Frequency-Modulated Magnetron for 900 Megacycles," *Proc. I.R.E.*, Vol. 35, No. 7, July, 1947.

minutes after they are observed. The most useful feature for comparing noise phenomena in different tubes is the voltage V_{ao}. Even this parameter changes with time, however, and may have quite different values in two tubes constructed (as nearly as possible) identically. Several tubes must, therefore, be constructed in order to be certain of the dependence or lack of dependence of noise upon some one feature of the design. Figure 2 is representative of the early 4000 megacycle tubes and exhibits the increase in noise with decrease in anode current which was observed with all tubes.

After vigorous oscillation in these magnetrons has begun at a sufficiently high anode voltage, the cathode heater power may be removed.
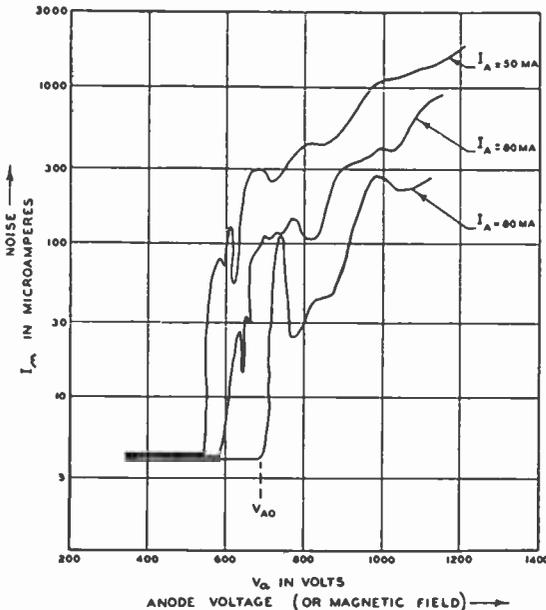
Fig. 2—Noise in early 4000-megacycle magnetrons. At constant anode current, anode voltage and magnetic field are approximately proportional. Therefore the abscissa may be considered (except for units) as either voltage or magnetic field strength.

Back-bombardment of the cathode by electrons which have received energy from the oscillating field produces secondary electrons and warms the cathode sufficiently to produce some thermionic electrons. The excess noise was invariably less when the ordinary heater power was removed. All of the measurements and correlations reported here refer to operation with no applied heater power.

The receiver (Fig. 1) was ordinarily tuned to pass modulation components in the frequency range 1496 to 1504 kilocycles, but other frequency regions were also explored. In general the noise was less the higher the receiver frequency, and the excess noise was quite small at a receiver frequency of 30 megacycles.

II. Correlations of Noise With Magnetron Properties

Many tubes were constructed with slightly different features from the ordinary magnetron in order to test the dependence of noise upon different aspects of the design. In each case correlation of noise properties with the changes made was attempted. Correlations were also attempted between noise properties and operating conditions of the tube.

The high-frequency load presented to the magnetron by the antenna and transmission system changed the details of the excess noise, but any standing wave ratios and phases which produced reasonably stable operation with moderate efficiency produced about the same values of V_{ao}. The presence or absence of discontinuities in the straps of the magnetron resonator, the presence or absence of "hats" on the ends of the cathode produced no consistent changes in noise. The geometry and potentials of electrodes near the ends of the cathode affected the noise only very slightly. Small changes in anode diameter produced no effect. A variation of almost a factor of two in cathode size produced marked changes in efficiency and stability; but the noise was little changed except with very small or very large diameters, and in these cases the noise was greater than with intermediate diameters.

These and other experiments were used to show that several more or less obvious hypotheses of the causes of excess noise were not tenable. For example, any axial oscillations (such as observed in magnetron structures many years ago by Megaw[1]) should be strongly affected by end plate potentials, and if the noise were related to such oscillations it should be a function of these potentials. Other hypotheses were concerned with the resonant modes of the resonator and with the space-charge and electric field distributions in the cathode-anode space. The results of the studies with various cathode and anode diameters, the fact that the excess noise was observed in a magnetron of quite different construction, and the occurrence of noise over a wide range of anode voltages and currents, made such hypotheses of the cause of noise very unlikely.

Three successful correlations emerged from this work, but none gave immediately a clear indication of the cause of the noise. These correlations will be briefly described here. The first was a correlation of noise with low efficiency accompanied by high back-bombardment of the magnetron cathode. Tubes which exhibited considerably lower efficiency than others of similar construction were usually very noisy.

---

[4] E. C. S. Megaw, "Fundamental Characteristics of Electronic Oscillations," *Nature* **137**, 189 (1936).

Experimental tubes were constructed with windows which allowed observation of the cathode during operation, and this correlation was verified in such tubes. This correlation is probably related to the reduction in noise as the normal cathode heater power is removed; in both cases high noise accompanies high cathode temperature.

The second correlation was between excess noise and the length of the cathode sleeve which was coated with oxides. It was found that if the emitting length were confined to less than two-thirds of the length of the anode, the voltage (V_{ao}) at which noise began could be materially increased. It was not certain that the ends of the cathode did not emit electrons, but the carbonate spray was carefully cleaned from all except the central region of the cathode sleeve in the hope that no emission would occur in the end regions. The suggestion was that the r-f and d.c. fields were distorted there, and it might be desirable to prevent the interaction of electrons with such fields.

The third correlation was between noise and the thermionic emission properties of the cathodes. The critical voltage (V_{ao}) was used as a measure of the former, and the heater current required to permit a thermionic current of 50 milliamperes at 100 volts (no magnetic field) was used as a measure of the latter. The correlation coefficient[5] of V_{ao} with this current was 0.79 for thirteen experiments with ten tubes. This shows a definite relation, with low noise accompanying low thermionic activity. This correlation was strengthened by observations on a single tube during the early part of its life, when both the cathode activity and the noise were changing. Magnetrons were constructed with strontium oxide coatings on the cathodes, since the emission from SrO is known[6] to be considerably less than that from the SrO-BaO mixtures ordinarily used. These tubes showed high values of heater current for 50 milliamperes thermionic current, but the noise properties did not differ appreciably from other tubes. This experiment indicated that noise was not directly related to the cathode's thermionic activity, but that for cathodes of the same composition, high noise accompanied high activity.

III. IONIZATION NOISE

One of the first hypotheses of the cause of the excess noise was that it originated in the ionization of gas atoms in the interaction space of the magnetron. To test this, several tubes were continuously pumped

---

[5] H. L. Rietz, *Mathematical Statistics*, Open Court (Chicago, 1927), p. 82.
[6] M. Benjamin and H. P. Rooksby, "Emission from Oxide-coated Cathodes," *Phil. Mag.* 15, 810 (1933).

while noise data were taken, and neon gas was admitted to vary the pressure. No variation of noise with pressure was observed for pressures less than about 5×10^{-3} mm of mercury. At pressures greater than this, the noise increased and the efficiency decreased rapidly; this behavior was not surprising since at 10^{-2} mm a glow discharge was beginning. The pressures could not be measured inside the magnetron itself, but they were measured in an ionization gauge connected to the magnetron.

Several magnetrons were constructed with tungsten helices in place of the usual oxide-coated cathodes. Two of these were continuously pumped, and noise was observed as a function of pressure. Figure
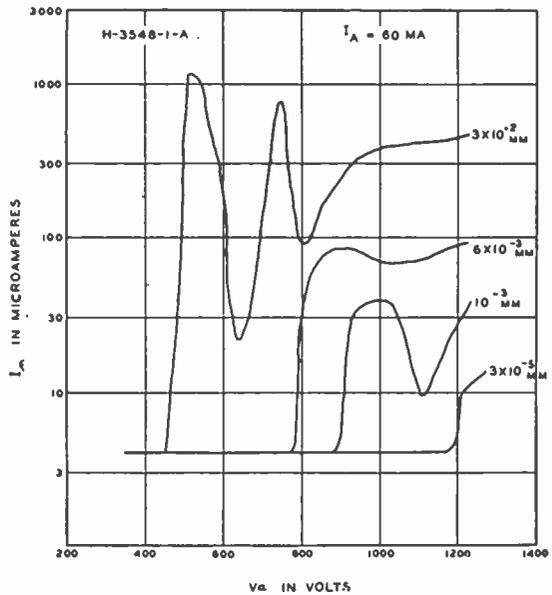
Fig. 3—Noise as a function of pressure. This tube had a helical tungsten filament in place of the usual oxide-coated cathode.

3 represents the behavior of these tubes. Unlike the results with oxide-cathodes, the data here show a continuous increase in V_{ao} (that is, decrease in noise) as the pressure decreases. For pressures less than 10^{-6} mm, V_{ao} was greater than 1200 volts for tungsten filament tubes. But lowering the pressure below 10^{-3} mm did not change the noise in oxide-coated cathode tubes, and even at 10^{-6} mm V_{ao} was no larger than 800 or 900 volts. If the cause of the noise is the same in the tubes with these two types of cathodes, it must be that there is a condensable gas in the oxide-coated cathode magnetrons which is present in the anode-cathode space but which is not measured by external pressure gauges.

Evidence to support this interpretation came from quite different

experiments. Observations upon tubes with windows revealed a luminescence in the anode-cathode space or upon the cathode surface. Spectrographic investigation showed that the line spectra of barium and strontium (the metallic constituents of the cathode coating) were represented by their most persistent lines. Since the lines were quite sharp, it was apparent that the luminescence was not fluorescence of the solid (such fluorescence is also observed on activated cathodes, but its spectrum is broad and diffuse). It appeared therefore that there were substantial quantities of Ba and Sr atoms in the interaction space. At least some of these were ionized, since some of the spectral lines observed were those of the positive, singly-charged ion.

In order to obtain further information about the atoms and ions in the anode-cathode space, "double-triode" tubes were built. In these tubes electrons from the central cathode bombarded both a nickel anode and an oxide-coated anode, which had been activated as if it were a cathode and which could be heated by a tungsten heater. A negative grid was placed in each half of the tube, and the positive ion currents to these grids were measured. Both anodes were maintained at approximately the same potential, and the electron currents in the two parts of the tube were the same. The grid current in the nickel anode space was about 0.01 microampere when 10 milliamperes electron current was flowing; this corresponds to roughly 10^{-6} mm pressure of residual gas in the tube. The grid current in the other side of the tube was always many times this value, and when the oxide-coated anode was heated to 700-800°C, grid currents of 10-40 microamperes were observed. It was apparent, therefore, that atoms or ions were released from the bombarded surface, and that these formed a condensable gas phase which could not be observed outside the immediate neighborhood of the bombarded surface. Noise measurements were also made on these tubes; excess noise was observed, and it was substantially proportional to the ion current collected by the grid.

From these experiments it was possible to derive the following interpretation of the phenomena of magnetron noise: (1) Some of the electrons (of the order of 20 percent) emitted from the cathode return to it after gaining considerable energy from the r-f field. If the cathode is oxide-coated, these back-bombarding electrons remove atoms or ions of the cathode material. Any ions emitted quickly return to the cathode, but they alter the space charge in the region of the potential minimum and permit a momentary excess of electrons to be emitted. Atoms emitted have between a 1 percent and a 10 percent chance of being ionized before they strike the anode or other parts of the tube. Ions thus produced alter the space charge and may release

other atoms or ions when they bombard the cathode. (2) Noise in excess of ordinary shot noise is produced by these processes. The noise occurring in space-charge limited electron currents when appreciable numbers of positive ions are present has been investigated by Thompson and North[7]. Most of the noise effects observed in the present tests of special triodes are consistent with their work. In general, however, the excess noise in magnetrons is greater than that obtained in any other vacuum tubes, and it would be necessary to assume pressures of the order of 0.1 to 0.01 mm of mercury in a magnetron in order to explain the observed excess noise by simple extrapolation of the work of Thompson and North. (3) The very great noise in magnetrons and the irregular dependence of noise upon time, current, and voltage, may be the result of cumulative or chain effects. The equivalent pressure of barium and strontium atoms near a hot magnetron cathode is of the order of 10^{-2} mm; this figure is derived from measurements on the "double-triode" tubes and is checked by the fact that at neon pressures less than 5×10^{-3} mm the noise in an oxide-coated cathode magnetron was independent of pressure. At such a large pressure the contribution of positive ions to the space-charge distribution is considerable, and increasing the pressure by a factor of ten would cause oscillation to cease and a gas discharge at low voltage to occur. Under operating conditions the magnetron may therefore be on the verge of a Townsend breakdown. The ionization of a single atom may result in a burst of charge reaching the anode that is many times the electronic charge, and a correspondingly large increase in noise may occur. Among the processes which could participate in such a breakdown are ionization by electron impact, removal of atoms from the cathode by electrons, removal of atoms from the anode surface (which is soon contaminated with cathode material) by electrons, secondary electron emission at the cathode, and removal of atoms and electrons from the cathode surface by positive ion bombardment. It is not clear why such a breakdown does not occur, however, instead of merely threatening to occur.

Our interpretation is therefore that the dominating cause of the excess noise is associated with the presence of gas atoms in the interaction space. In a tungsten filament tube, these atoms are of residual gases and by sufficiently extensive pumping and gettering the excess noise may be substantially removed. In an oxide-coated cathode tube which is well out-gassed and gettered, the atoms are at least pre-

---

[7] B. J. Thompson and D. O. North, "Fluctuations in Space-Charge-Limited Currents at Moderately High Frequencies; Part IV—Fluctuations Caused by Collision Ionization," *RCA Review*, Vol. V, No. 3, January, 1941.

dominantly atoms removed from the cathode by electron back-bombardment.

The correlations reported in the preceding section give some support to these conclusions. The first correlation (noise *vs.* degree of back-bombardment) is evidently related to the dependence of the rate of removal of cathode material upon the number of bombarding electrons and the cathode temperature, both of which are higher in magnetrons with excessive back-bombardment. Experiments showed that the equivalent pressure of cathode material was an exponentially increasing function of the cathode temperature. Therefore, under conditions of severe back-bombardment, it would be expected that noise would be greater than with moderate bombardment. This is just what was observed.

The reduction in excess noise by reducing the sprayed length of the cathode is presumed to be related to the first correlation. In several tubes with short-spray-length cathodes, the cathodes operated at temperatures less than 700°C. Most full-length-sprayed cathodes operated at considerably higher temperatures. This is evidence of smaller back-bombardment when the electrons do not interact with electric fields near the ends of the anode.

The third correlation is still obscure. It may be that a very active thermionic cathode, which may have a surface layer of barium or strontium atoms, yields more atoms or ions when bombarded with electrons than a less active cathode.

A number of apparently unconnected relations are consistent with the gas-ionization hypothesis of the cause of excess noise: The decrease of noise with increasing I.F. frequency, the decrease of noise when the cathode heater power is removed, the negative results of various experiments with end-plate potentials and geometries, observations of the appearance of anode and cathode surfaces after many hours of operation, the occurrence of excess noise in magnetrons of quite different construction, the lack of dependence of noise phenomena upon cathode and anode diameters, and the irregular changes of noise with time, are all susceptible to explanation by this hypothesis.

There are three outstanding difficulties: (1) The physical process whereby atoms or ions are released from the cathode coating by bombarding electrons is not known, though several processes seem to be capable of producing this result (2) The exact way in which ions and ionization of emitted atoms combine with the electron flow and the electric and magnetic fields to produce such large amounts of noise is not known. (3) The decrease of noise with increasing anode current does not seem to be explained.

Despite these difficulties, the hypothesis of the cause of the noise is felt to be reasonably well established. More detailed information is required about the mechanism of release of cathode material in order to advance our understanding of noise. Experiments with a mass spectrometer are planned to determine what molecules, atoms, or ions are released. The results of these experiments may make possible a quantitative theory of noise generation. Preliminary results with "double-triode" tubes indicate that at least some of the cathode material is removed as atoms and that the exponential increase of removed atoms (per unit electron curent) has an exponent which corresponds to an "activation energy" of 40-50 kilocalories per mole. The heats of sublimation of barium and strontium are 44 and 38 kilocalories. This may be a coincidence, or it may be that electron bombardment produces small patches of barium and strontium, from which atoms readily volatilize at the cathode operating temperature.

IV. Progress in the Reduction of Excess Noise

The increase in V_{ao} by reducing the sprayed length of the cathode has already been reported. This change and the improvement of the efficiency and reduction of back-bombardment resulting from further development of the frequency-modulated magnetrons[2] resulted in a large percentage of magnetrons which were not "noisy" in the ordinary operating region, near $V_a = 850$ volts.

The hypothesis of the cause of noise which was advanced in the preceding section suggested that the cathode be modified in order to reduce the excess noise and to raise V_{ao} still further. No hope was entertained of eliminating the back-bombarding electrons, so a means was sought for preventing the removal of cathode material by such electrons. There is considerable evidence that electrons returning to the cathode do not strike at normal incidence, but an angles varying between 45° and tangential incidence. Furthermore, for a fixed magnetic field, the sense of rotation of all electrons about the cathode is the same. These facts permit the construction of a cathode such that returning electrons strike a nickel surface instead of striking the cathode coating. One such cathode is shown in Fig. 4. The active length of the cathode is in the form of a 13-pointed star; only one side of each star point is coated with oxides. To accomplish this, the entire cathode was coated, and the spray was carefully scraped from areas where it was not wanted. This process was tedious, since the diameter of the cathode was only 0.070 inches, but was quite practical for experimental tubes.
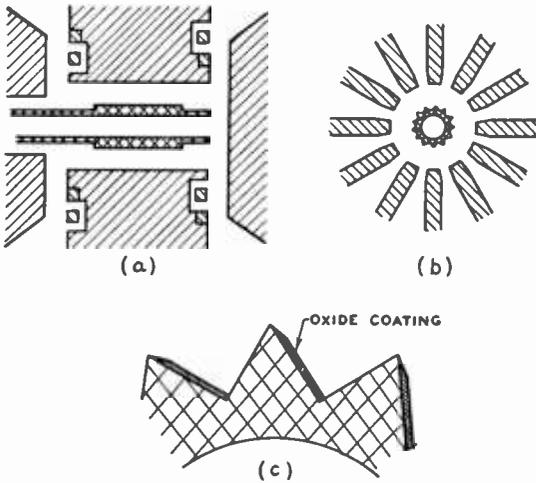
Fig.4—
Grooved cathode.

Figure 5 shows the noise behavior of one of the four tubes con-structed with the cathode of Fig. 4. By "correct" direction of the magnetic field we mean that direction which allows back-bombarding electrons to strike only the nickel surface (this is the clockwise sense of rotation in Fig. 4). It is evident that the noise is much less if the oxide-coating is sheltered from bombardment. This provides increased confidence in the hypothesis of the cause of noise which was advanced in the preceding section.
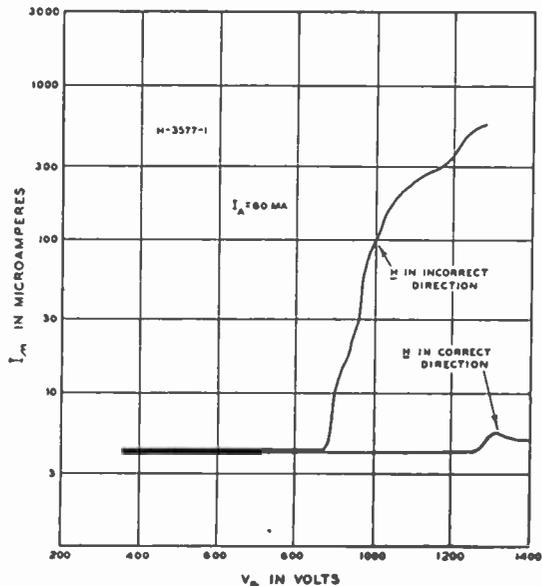
Fig. 5 — Noise in magnetron with grooved cathode. When the magnetic field is in the "correct" direction, the oxide-coated sur-faces are sheltered from back-bombarding electrons.

Three other tubes with cathodes as shown in Fig. 4 were built. One of these showed somewhat less noise for the "incorrect" direction of magnetic field than for the "correct" direction. This tube exhibited very high back-bombardment, more noise than other tubes of this type, and low efficiency. When it was dissected this tube was found to have a cathode which was not parallel to the axis of the anode. These facts make the information from this test of questionable value, but it is still not understood why this tube exhibited more noise when the electrons were rotating in the proper sense.

The remaining two tubes showed substantially the same behavior as the first (Fig. 5) and therefore confirmed the conclusions from that tube. The net result of all such experiments was that V_{ao} is 200-300 volts larger when the magnetic field is applied in the "correct" sense.

An interesting possibility of cathodes of the general type of Fig. 4 is that they may permit substantially longer life than has previously been experienced with magnetrons. There is much evidence (some of it contained in the present work) that cathode life in a magnetron is limited more by the destruction of the cathode by back-bombardment than by the processes of cathode decay observed in ordinary diodes and negative grid tubes. If long life, not low excess noise, were the chief requirement, the oxide-coating could initially cover the entire cathode. Removal of part of the surface by back-bombardment would not cause failure, because at least part of the surface would be sheltered from back-bombarding electrons.

THE MAXIMUM EFFICIENCY
OF REFLEX-KLYSTRON OSCILLATORS*†

By

ERNEST G. LINDER AND ROBERT L. SPROULL

Research Department, RCA Laboratories Division.
Princeton, N. J.

Summary—The theory of reflex-klystron oscillators is given in detail. It includes a discussion of relations in a loaded oscillator. It is shown that maximum efficiency for small amplitudes is given by

$$\eta_0 = 0.169\, M^2 i_1/G_c V_0,$$

where M is the coefficient of modulation of the gap, i_1 is the effective current, G_c is the shunt conductance of the unloaded resonator, and V_0 is the beam voltage. Possibilities of increasing efficiency are considered, including effects of grid transmission on effective current and on space charge, and effects of multiple electron transits.

THE THEORY presented in this paper is based mainly on the previous investigation of D. L. Webster,[1] but it embodies considerable modification and extension of his original treatment. The work was done several years ago but could not previously be made public because of wartime secrecy restrictions. The theory is a small-amplitude one, that is, the oscillatory voltage is assumed small in comparison with the steady voltages. This assumption is most applicable in the case of short wavelengths, such as below 10 centimeters, since there the efficiencies are seldom greater than a few per cent.

I. ENERGY RELATIONS

The essential phenomenon in bunching in klystrons is that an electron beam, initially uniform with respect to velocity and charge distribution, is velocity modulated; so that, after an interval during which the beam travels in a drift space, the charge distribution in the beam is no longer uniform, and hence the beam contains components of alternating current.

The expression for this current is

---

[1] D. L. Webster, "Cathode-ray bunching," *Jour. Appl. Phys.*, vol. 10, pp. 501-508; July, 1939.

$$i_2 = -i_1 \left[1 + 2 \sum_0^\infty J_n(nr) \cos n \left(\omega t_2 - \theta - \frac{\pi}{2} \right) \right] . \qquad (1)$$

The term containing the fundamental frequency ω is

$$i = -2i_1 J_1(r) \cos \left(\omega t_2 - \theta - \frac{\pi}{2} \right) . \qquad (2)$$

Here i_1 is the value of the current during the first passage of the beam through the gap at time t_1, and i_2 is its value during the second passage at time t_2. $J_n(nr)$ is the Bessell function of order n. The quantity r is called the bunching parameter and is given by $r = \theta M V_m/2V_0$, where θ is the transit angle of the bunch centers between the first and second transits, M is the coefficient of modulation of the gap, V_m is the amplitude of the oscillatory voltage, and V_0 is the beam voltage.

These expressions were originally derived by D. L. Webster[1] for the case of the two-cavity, single-transit klystron. It can be shown[2] that they are unaltered for the reflex case, except for a change of sign due to the reversal of the current.

In the present treatment the phase angle $\pi/2$ also is introduced to represent a phase shift inherent in the bunching process. This occurs since bunches form about those electrons traversing the gap when the voltage is zero and changing from accelerating to decelerating.[3] Thus the current maxima (bunch centers) are formed out of phase by angle of $\pi/2$ with respect to the voltage maxima. The gap voltage corresponding to the current given by (2) is

$$V = V_m \cos \omega t_2,$$

and the total phase angle between current and voltage is therefore $\theta + (\pi/2)$.

From (1) it is seen that the current resulting from the bunching process contains components of direct current, and alternating currents of the fundamental frequency and all harmonics. To obtain an expression for the power derivable from such a beam, consider its passage across a gap or between grids across which there is the voltage

[2] E. L. Gintzon, and A. E. Harrison, "Reflex-klystron oscillators," *Proc. I.R.E.*, vol. 34, pp. 97-113; March, 1946.

[3] For discussions of the bunching process see also D. M. Tombs, "Velocity-modulated beams," *Wireless Eng.*, vol. 17, pp. 55-60; February, 1940; and L. J. Black and P. L. Morton, "Current and power in velocity-modulated tubes," *Proc. I.R.E.*, vol. 32, pp. 477-482; August, 1944.

$V_m \cos \omega t_2$. The energy absorbed from the beam during one cycle will then be given by the integral of the product of the instantaneous current and the effective voltage which acts upon the beam during its passage. Thus the energy is

$$W' = - \int_0^{2\pi/\omega} i_2 M V_m \cos \omega t_2 dt. \tag{3}$$

In this integration, all terms will vanish except the one containing the frequency ω.

Hence,

$$W' = 2i_1 M V_m J_1(r) \int_0^{2\pi/\omega} \cos \left(\omega t_2 - \theta - \frac{\pi}{2} \right) \cos \omega t_2 dt$$

$$= \frac{2\pi}{\omega} i_1 M V_m J_1(r) \cos \left(\theta + \frac{\pi}{2} \right) .$$

Thus the energy absorbed per second, or the power, will be

$$W = \frac{\omega}{2\pi} W' = i_1 M V_m J_1(r) \cos \left(\theta + \frac{\pi}{2} \right) . \tag{4}$$

This is a maximum for

$$\theta + \frac{\pi}{2} = 2m\pi$$

or

$$\theta = 2\pi \left(m - \tfrac{1}{4} \right)$$

where m is any positive integer. Each integer corresponds to a mode of oscillation.

These values of θ giving maximum power correspond to the passage of the bunch center through the grids when the field has its maximum retarding effect. It is convenient to use the angle.

$$\gamma = \theta - 2\pi \left(m - \tfrac{1}{4} \right), \tag{5}$$

which indicates the departure from the power maximum or from a mode center. In terms of this

$$i = -2i_1 J_1(r) \cos(\omega t_2 - \gamma), \tag{6}$$

and

$$W = i_1 M V_m J_1(r) \cos \gamma. \tag{7}$$

The voltage between the grids may now be determined by considering that, if the shunt resistance between them is R, then

$$V_m^2 = 2RP,$$

or from (7), putting $\cos \gamma = 1$ to get the amplitude,

$$V_m^2 = 2Ri_1 M V_m J_1(r), \quad \text{or}$$

$$V_m = 2Ri_1 M J_1(r). \tag{8}$$

The factor M which occurs above, and which will be frequently used also in what follows, may be defined by the expression

$$M = \frac{V_m \text{ (effective)}}{V_m}.$$

An equivalent definition is

$$M = \frac{P}{eV_m},$$

where P is the actual maximum energy lost by an electron in traversing the gap. It would be equal to unity only in the case of an ideal gridded gap and zero transit time. It is a function of the variation of the field across the gap, but here it will be computed only for the case of parallel plane grids of sufficiently fine mesh so that the field may be considered uniform in intensity and normal to the grids at all points. The voltage across the gap is then given by $V_m \cos(\omega t_2 + \gamma)$.

The energy transferred to an electron traversing the gap is

$$P = \int_0^\tau F ds,$$

where τ is the transit time, F the force, and s the position co-ordinate.

Therefore,

$$P = e \int_0^\tau E \frac{ds}{dt_2} dt_2$$

$$= e \int_0^\tau \frac{V_m}{d} \cos(\omega t_2 + \gamma) \frac{ds}{dt_2} dt_2,$$

where E is the field and d the gap spacing.

Because of the assumption of small field, i.e., $V_m \ll V_0$, the velocity ds/dt_2 may be taken as constant and equal to v_0, the entrance velocity. Thus,

$$P = \frac{ev_0 V_m}{d} = \int_0^\tau \cos(\omega t_2 + \gamma) dt_2$$

$$\frac{P}{eV_m} = \frac{1}{\tau} \int_0^\tau \cos(\omega t_2 + \gamma) dt_2$$

$$= \frac{1}{\omega\tau} [\sin(\omega\tau + \gamma) - \sin\gamma]$$

$$= \frac{2}{\omega\tau} \cos\left(\frac{\omega\tau + 2\gamma}{2}\right) \sin\frac{\omega\tau}{2}.$$

Therefore,

$$\frac{P}{eV_m} = \frac{2}{\omega\tau} \cos\left(\frac{\omega\tau + 2\gamma}{2}\right) \sin\frac{\omega\tau}{2}.$$

The maximum value of this gives M, and evidently occurs when

$$\frac{\omega\tau + 2\gamma}{2} = 2n\pi, \qquad \text{or} \qquad \gamma = 2n\pi - \frac{\omega\tau}{2}.$$

Hence

$$M = \frac{2}{\omega\tau} \sin\frac{\omega\tau}{2}. \tag{9}$$

II. Efficiency of a Loaded Oscillator

The power given by (7) represents the total amount available from the beam. However, this cannot all be exploited in a useful load, ex-

cept for the hypothetical case in which the oscillator circuit, e.g., cavity resonator, is totally loss-free, so that it has infinite shunt resistance. In general, the power in the load may be determined as follows. The loaded resonator circuit is represented by a lumped-constant circuit in which the capacitance and inductance of the cavity, denoted by C and L, are in parallel. These are shunted by the unloaded shunt resistance[4] R_C, and also by the resistance of the load as reflected across the resonator gap, denoted by R_L.

Hence the total shunt resistance R of the circuit or the gap due to both the load and the cavity losses is given by

$$\frac{1}{R} = \frac{1}{R_C} + \frac{1}{R_L},$$

or in terms of the respective conductances

$$G = G_C + G_L. \tag{10}$$

In the following, resistances and conductances will be used interchangeably to facilitate simplification of equation writing.

The power dissipated in the load is

$$W_L = \tfrac{1}{2} G_L V_m^2, \tag{11}$$

and that dissipated in the cavity is

$$W_C = \tfrac{1}{2} G_C V_m^2. \tag{12}$$

By use of (10), (11) may be written

$$W_L = \frac{G_L V_m^2}{2} = \frac{V_m^2}{2}(G - G_C), \tag{13}$$

and by use of (8)

$$\eta = \frac{W_L}{V_0 i_1} = \frac{1}{V_0}\left[M V_m J_1\,(\theta\, M V_m/2V_0) - \frac{G_C V_m^2}{2 i_1} \right]. \tag{14}$$

In this expression the first term represents the contribution due to

---

[4] For methods of measuring R_C see R. L. Sproull and E. G. Linder, "Resonant-cavity measurements," *Proc. I.R.E.*, vol. 34, pp. 305-312; May, 1946.

the total energy delivered by the beam to the oscillating circuit. The second term represents the effect of the losses in the cavity.

From (14) the condition for the start of oscillation may be simply derived, since it corresponds to $\eta = 0$. Thus

$$MV_m J_1 \ (\theta \ MV_m/2V_0) = \frac{G_C V_m^2}{2i_1} \ .$$

But since the amplitude of oscillation is infinitesimal at the inception of oscillation, the Bessell function may be replaced by an approximation valid for small arguments. Thus

$$J_1 \ (\theta \ MV_m/2V_0) = \frac{\theta \ MV_m}{4V_0}$$

and

$$\frac{\theta \ M^2 V_m^2}{4V_0} = \frac{G_C V_m^2}{2i_1}$$

or

$$\beta \equiv \frac{\theta \ M^2 i_1}{2G_C V_0} = 1. \tag{15}$$

This defines β and gives the lowest value for which oscillations may occur. The value $\beta = 1$ corresponds to no loading of the cavity, i.e., $G_L = 0$. This is evident from (14), since the two right-hand terms represent the total power generated and the power dissipated in the cavity, respectively. If these are equal, then no power can be supplied to a useful load. The starting condition for a loaded cavity will require $\beta > 1$, its actual value depending upon the value of G_L.

To derive the conditions for maximum efficiency, consider that with a given oscillator for which V_0, ω, M, G_C, and i_1 are fixed, the efficiency may be varied by changing θ and V_m. The drift angle θ may be varied by altering the reflector voltage V_R. The quantity V_m is affected by varying the load resistance. Hence, to find the maximum efficiency obtainable from a given tube, it is necessary to maximize with respect to both of these two independent variables θ and V_m. Take first,

$$\frac{\partial \eta}{\partial \theta} = \frac{MV_m}{V_0} \ J_1' \left(\frac{\theta MV_m}{2V_0} \right) \frac{MV_m}{2V_0} = 0,$$

i.e.,

$$J_1'\left(\frac{\theta M V_m}{2V_0}\right) = J_1'(r) = 0, \tag{16}$$

or

$$r_0 = 1.84, \quad \text{and} \quad J_1(r_0) = 0.582. \tag{17}$$

This condition makes $J_1(r)$ a maximum, and hence means that the bunching is such that the alternating component of frequency ω of the beam has its greatest value.

Now maximize with respect to V_m, i.e.,

$$\frac{\partial \eta}{\partial V_m} = \frac{1}{V_0}\left[M V_m J_1'\left(\frac{\theta M V_m}{2V_0}\right)\frac{\theta M}{2V_0} + M J_1\left(\frac{\theta M V_m}{2V_0}\right) - \frac{G_C V_m}{i_1}\right] = 0$$

or

$$rJ_1'(r) + J_1(r) = \frac{G_C V_m}{M i_1},$$

or, since

$$J_1'(r_0) = 0, \qquad J_1(r_0) = \frac{G_C V_m}{M i_1}. \tag{18}$$

Since, from the theory of Bessell functions,

$$rJ_1'(r) + J_1(r) = rJ_0(r), \tag{19}$$

this condition may be written,

$$J_0(r) = \frac{2 G_C V_0}{\theta M^2 i_1} \equiv \frac{1}{\beta}. \tag{20}$$

Since, with a given tube, V_m is a function of the loading, (20) is the condition for optimum loading. That this condition represents a maximum, and not a minimum or a point of inflection, may be shown by investigating the usual conditions involving second-order derivatives. It should be noted that (18) holds only when η is maximized with respect to both θ and V_m, whereas (20) requires only maximization with respect to V_m and hence is valid for all values of r.

The meaning in terms of load conductance of the maximization

with respect to V_m may be seen from the following. The general expression for the voltage across the cavity is given by (8). Apply to this the condition for maximum efficiency (18). This gives

$$V_m = \frac{2i_1 M G_C V_m}{GM i_1},$$

from which

$$G = 2G_C, \quad \text{or} \quad G_L = G_C. \tag{21}$$

This indicates that equal amounts of power are dissipated in the oscillator resonator and in the load.

It is now possible to derive an expression for the maximum efficiency with respect to both θ and V_m. Let the two conditions for maxima, (16) and (20), be solved simultaneously, giving

$$J_0(1.84) = \frac{1}{\beta_0}, \tag{22}$$

or

$$\beta_0 = 3.17.$$

Then the maximum efficiency for a given tube is obtained by rewriting (14),

$$\eta = \frac{M^2 i_1}{G_C V_0} \left[\frac{r}{\beta} J_1(r) - \frac{r^2}{2\beta^2} \right], \tag{23}$$

and introducing the above optimum values of r and β. This yields

$$\eta_0 = \frac{M^2 i_1}{G_C V_0} \left[\frac{r_0}{\beta_0} J_1(r_0) - \frac{r_0^2}{2\beta_0^2} \right],$$

or

$$\eta_0 = 0.169 \frac{M^2 i_1}{G_C V_0}. \tag{24}$$

This is valid only if $MV_m \ll \frac{1}{2} V_0$, since this was assumed in the derivation of (1); also, if $MV_m > \frac{1}{2} V_0$, some electrons will be reversed between the grids. Such reversal introduces conditions not covered by

the present theory and which moreover clearly would result in a decrease in efficiency.

It is instructive to apply the above small-signal theory to the region where MV_m approaches the limiting value, $\frac{1}{2} V_0$, since, although it is inaccurate there, it nevertheless gives some idea of the ultimate values to be expected for the efficiency.

If it is assumed that the conditions are always such as to give maximum bunching, i.e., (16) is always satisfied then (14) may be written

$$\eta = \frac{1}{V_0} \left[0.58\, MV_m - \frac{G_C V_m^2}{2i_1} \right], \tag{25}$$

which expresses efficiency in terms of the amplitude V_m.

A set of curves of efficiency versus MV_m, for the following conditions

$$V_0 = 300 \text{ volts}$$
$$M = 1/\sqrt{2}$$
$$i_1 = 0.01 \text{ ampere}$$
$$MV_{m\ max} = 150 \text{ volts}$$

is given in Fig. 1 for values of R_C from 5000 to ∞ ohms. These curves are all terminated at the line $MV_m = 150$, since beyond that point electrons will be reversed by the alternating field.

The efficiencies given by (24) correspond to the maxima of curves such as a, b, c, and d, and also to e, the maximum of which occurs just at $MV_m = 150$. However, for curves such as f, (24) does not apply since its maximum occurs in the forbidden region $MV_m > 150$. The correct maximum efficiency in this case is given by the intersection of the curve with the line $MV_m = 150$, and may be computed from (25). Hence, on this basis, the maximum attainable efficiency for reflex tubes is 29 per cent, corresponding to curve g for $MV_m = 150$. It is of interest to compare this with the case of single-transit tubes. Here the restriction on MV_m is $MV_m \leqq V_0$. In this case it is seen from (25) that an ultimate efficiency of 58 per cent might be attained.

However, in most actual cases where the efficiency is of the order of a few per cent, MV_m is always much less than its limiting value of $V_0/2$, and the typical curve resembles that of case a, Fig. 1. In all such cases (24) gives the correct maximum efficiency.

To get an expression for V_m, under conditions for maximum η, rewrite (14) and substitute from (18). This yields

$$\eta_0 = \frac{MV_m}{V_0}\left[J_1(r_0) - \frac{G_c V_m}{2Mi_1}\right] = \frac{MV_m}{V_0}\frac{J_1(r_0)}{2} = 0.29\frac{MV_m}{V_0}. \qquad (26)$$

Since this expression and (23) must be equivalent they may be equated and solved for V_m. Thus,

$$V_{m0} = 0.58\,MR_c i_1. \qquad (27)$$

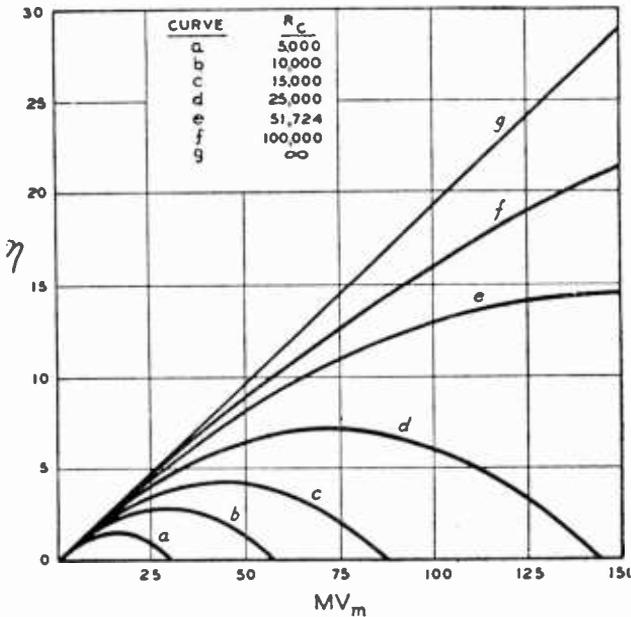This formula, like those from which it is derived, is valid only for $MV_m \ll \frac{1}{2}V_0$.

Fig. 1—Efficiency in per cent η versus effective oscillatory voltage MV_m for various values of unloaded cavity shunt resistance R_c.

III. Possibility of Increasing the Efficiency

In the expression for maximum efficiency,

$$\eta_0 = 0.169\frac{M^2 i_1}{G_C V_0},$$

there does not seem to be much of a possibility of greatly increasing η_0 by attaining more favorable values of M, G_c, or V_0. In general,

these factors are not independent, and changes which improve one frequently result in a compensating change in the others. Consider the case where the resonator is a simple cylindrical cavity, of height d and radius a, in which $d \ll a$. Then

$$G_C = \frac{a\delta}{189d^2}$$

where δ is the skin thickness. M may be expressed in terms of V_0 and d by making use of (9) in conjunction with the formula for transit angle.

$$\omega\tau = 1.07 \cdot 10^{-7} \frac{fd}{\sqrt{V_0}},$$

and that for wavelength,

$$\lambda = 2.61a.$$

These relations give

$$\frac{M^2}{G_C V_0} = \frac{5.77 \cdot 10^6}{f\delta} \sin^2 \frac{\omega\tau}{2}.$$

Therefore, for a fixed frequency, it is evident that the above quantity depends upon the sine term only, and is greatest when its argument is $\pi/2$, i.e., when the transit angle is π radians. Hence, so longs as $\omega\tau = \pi$, variations of M, G_C, and V_0 do not affect the efficiency, and no further increase of efficiency in this particular case is obtained by varying these quantities.

Since the efficiency is directly proportional to i_1, and the power to $i_1{}^2$, possibilities of increasing the current are of importance. It is of significance that i_1 is the effective current and not the cathode emission current which will be denoted by i_e. The former is obtained from the latter by subtracting lcsses principally to the grids and aperture edges.

The losses to the grids may be expressed in terms of the grid transmission coefficient α, which is taken as the ratio of open area to total area over the grid surface. This is a somewhat idealized definition since it assumes no component of lateral velocity to be introduced by the grid wires, and assumes that the shadows cast by previous grids do not affect the transmission coefficient. The area ratio may be shown to be

$$\alpha = \left(\frac{l-b}{l}\right)^2$$

where l is the distance between wire centers, and b is the wire diameter.

If the beam passes through one grid, the transmitted current will be

$$i_1 = \alpha i_e.$$

For two grids it is

$$i_1 = \alpha \cdot \alpha i_e.$$

and for u grids

$$i_1 = \alpha^u i_e. \tag{28}$$

For passage through u grids of transmission α_1, and u' grids of transmission α_2, the resulting current is

$$i_1 = \alpha_1{}^u \alpha_2{}^{u'} i_e.$$

It is obvious that i_1 can be increased by increasing α and decreasing u. The transmission α can be increased by using finer wire and larger spacings. But the fineness of the wire is limited by its temperature rise, and the largeness of the spacing is limited by its effect on M. Large spacings permit fringing electric fields and decrease M. The number u of grids depends upon the design of the beam electron-optical system. It is usually either two or three in reflex tubes, but the beam passes through some of the grids more than once.

Another important consideration that determines i_1 is that of multiple transits of electrons through the resonator. In the above theory only two transits are considered, first the modulating transit, and second the driving transit which occurs after the electrons have been bunched in the reflection space. Obviously there must be additional transits, since the cathode will act as a reflector. However, because of debunching and improper phasing these additional transits contribute little one way or the other to the oscillations of the resonator except in special cases. Nevertheless, they may contribute substantially to the space charge and thus decrease the cathode current i_e so that in some cases the beam current may be limited by space charge rather than by grid temperature.

The magnitude of this effect may be evaluated by considering a simple structure consisting of a plane cathode, grid, and reflector,

with the cathode and reflector at zero potential and the grid positive.
Let the grid have a coefficient of transmission α. Then for every N
electrons emitted from the cathode, a fraction $N\alpha^2$ will return after
one round trip. At the start of the second trip, the fraction $N\alpha^2$ will
be joined by a new group of N electrons, making a total of $(1 + \alpha^2) N$.
After q round trips the number will be

$$(1 + \alpha^2 + \alpha^4 + \cdots + \alpha^{2q}) N.$$

After equilibrium has been reached, i.e., $q = \infty$, this becomes $N/(1 - \alpha^2)$, which is the sum of this infinite series. Hence, if electrons are
emitted from the cathode at a rate of N per second, the number moving
through the space will be $N/(1 - \alpha^2)$. Or, if the emission current is
i_e, the space current i_s will be

$$i_s = \frac{i_e}{1 - \alpha^2}. \tag{29}$$

Thus space-charge limitation would occur with a cathode emission
current $1 - \alpha^2$ times that given by the Child-Langmuir law. If there
are u grids, the factor becomes $1 - \alpha^{2u}$.

Thus it is evident that there are two distinct grid transmission
effects: (1) losses to the grids which decrease the effective current,
and (2) transmission by the grids which allows multiple transits,
resulting in greater space charge and consequent decrease in cathode
emission. These effects are opposed to each other; large α increases
(2) while small α increases (1). Hence there is an optimum α.

This is clarified by reference to Fig. 2. A structure having three
grids is shown: a is an accelerating grid, whereas b and c are radio-
frequency grids. The effective current driving the oscillator is i_1,
which is the current remaining after four grid transits. The remain-
ing transits are assumed to contribute only to space charge. The sum
of currents due to all transits, regardless of direction of flow, gives
the space current i_s.

For the usual reflex-tube case with u grids, i_1 is given by

$$i_1 = \alpha^{u+1} i_e.$$

Then, from (29),

$$i_1 = \alpha^{u+1} (1 - \alpha^{2u}) i_s.$$
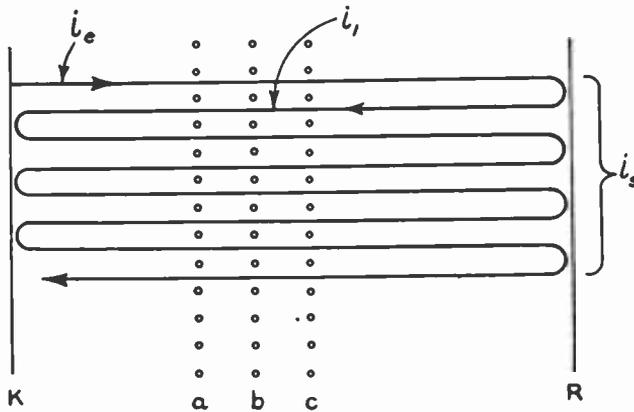
Fig. 2—Schematic electron paths in a triple-grid reflex oscillator, showing the effective current i_1 and the built-up space current i_s due to multiple transits, between cathode K and reflector R.

Here i_s is a constant determined by the Child-Langmuir law. By differentiating with respect to α, and equating to zero, the maximum for α is found to be

$$\alpha_0 = \left(\frac{u+1}{3u+1} \right) \frac{1}{2u}. \tag{30}$$

Hence, if

$$u = 2, \quad \alpha_0 = 0.810, \quad i_1 = 0.30 i_s,$$

and if

$$u = 3, \quad \alpha_0 = 0.858, \quad i_1 = 0.29 i_s.$$

These results represent a case in which the beam electrons continue to oscillate until all are captured by the grids. If there is any divergence of the beam so that electrons are captured elsewhere, corrections to the above are required. In some tubes all electrons are collected after two transits by an electrode which does not permit them again to approach the cathode after having once left it. Then space-charge effects are reduced to a minimum, and the grids may be designed for higher transmission limited only by considerations of heat dissipation and effect on M. Also it should be noted that i_1 can be increased by increasing i_s, as for example by diminishing the spacing from the cathode to the first grid. Here again heat dissipation is a limiting factor. The above maximization of α would usually be most useful in making best use of an i_s already at its highest value.

Another method of increasing efficiency lies in the possibility of utilizing electron transits after the second one. As has just been pointed out, multiple transits usually occur, but they are usually ineffective because of debunching and improper phasing. However, it is possible to design a tube so that the third transit will cause a considerable increase in efficiency. This is especially true in the case of
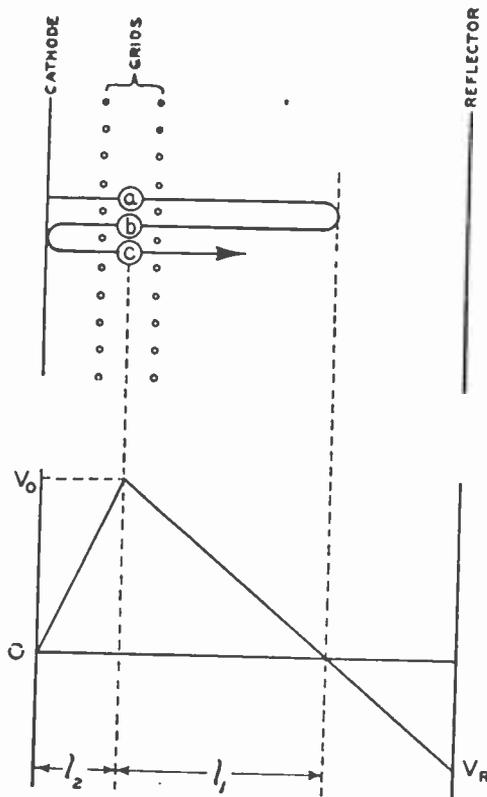
Fig. 3—Schematic diagrams of reflex klystron utilizing three electron transits.

low-amplitude (i.e., $V_m \ll V_0$) oscillators, since the velocity spread in the beam is small, and when once bunched the electrons stay sufficiently bunched long enough to permit a third transit with considerable effectiveness.

Referring to Fig. 3, we see that the power developed by the beam during the second transit, i.e., transit b, will be, according to (7),

$$W_1 = i_1 M V_m J_1(r_1) \cos \gamma_1.$$

During the third transit, i.e., transit c, it will be

$$W_1 = i_2 M V_m J_1 (r_2) \cos \gamma_2.$$

This assumes that the bunch shape is affected by the second transit only to the extent that r changes between the second and third transits. The fact that various electrons in the bunch are subjected to different action of the radio-frequency field and the bunch thus distorted is ignored. The deceleration of the bunch as a whole during transit b also is neglected, and the bunch therefore is assumed to be reflected at the cathode surface between transits two and three. Both of the above assumptions are justified on the basis of a small-signal theory.

Thus the total will be

$$W = W_1 + W_2 = MV_m (i_1 J_1 (r_1) \cos \gamma_1 + i_2 J_1 (r_2) \cos \gamma_2) = \tfrac{1}{2} G V_m^2.$$

Introducing this in (13) yields

$$W_L = \frac{V_m^2}{2} (G - G_C)$$

$$= MV_m [i_1 J_1 (r_1) \cos \gamma_1 + i_2 J_1 (r_2) \cos \gamma_2] - \frac{G_C V_m^2}{2}.$$

Therefore,

$$\eta = \frac{1}{V_0} \left[MV_m J_1 \left(\frac{MV_m \theta_1}{2V_0} \right) - \frac{G_C V_m^2}{2i_1} + \alpha^2 MV_m J_1 \left(\frac{MV_m \theta_2}{2V_0} \right) \right], \quad (31)$$

where the cosine factors have been given their maximum value of unity. This maximum value may be attained by varying the two voltages V_0 and V_R by small amounts. It is of interest that this exhausts the available parameters which can be adjusted for maximizing the phase angles γ_1 and γ_2. In attempting to utilize still more transits this maximization would not be possible.

These expressions are now similar to those for the double-transit case except for the addition of an extra term which represents the contribution of the third transit. The factor θ_2 which occurs in this added term is similar to θ_1 which has been discussed above and represents the transit angle of the bunch center from the first to the third transit, that is, it is equal to θ_1 plus the added angle caused by l_2.

As before, to maximize (31) with respect to the loading, the derivative of η with respect to V_m is equated to zero. This yields

$$J_0\left(\frac{MV_m\theta_1}{2V_0}\right) + \alpha^2\frac{\theta_2}{\theta_1} J_0\left(\frac{MV_m\theta_2}{2V_0}\right) = \frac{1}{\beta}. \qquad (32)$$

From the tube geometry and operating voltages, the quantities θ_1, θ_2, α_1 and β are known. Thus V_m may be determined from (32) and then η from (31).

As a specific example a case has been computed[5] for the following numerical values:

> $\omega = 5.89 \cdot 10^{10}$ cycles per second ($\lambda = 3.2$ centimeters)
>
> $M = 0.64$
>
> $\alpha = 0.81$
>
> $V_0 = 300$ volts
>
> $i_1 = 0.010$ ampere
>
> $R_C = 5000$ ohms.

The results are plotted in Fig. 4 where contour lines for η are plotted against l_1 and l_2. For a complete picture the factors $\cos \gamma_1$ and $\cos \gamma_2$, which were dropped to obtain (31), must now be considered. These are seen to be rapidly varying functions, compared to η, when l_1 and l_2 are varied, since it may be shown that γ changes by 2π radians for a change in l of only 0.274 millimeter. Thus the surface of Fig. 4, may be considered as an envelope of an array of individual maxima spaced 0.274 millimeter along each axial direction. Any value there plotted hence can be approached closely, if not actually realized.

For the above numerical case, but with only the usual two transits through the resonator, the maximum efficiency as computed from (24) is 1.16 per cent, whereas for three transits it is 3.16 per cent, an increase of 2.7 times. Somewhat greater values than this might theoretically be realized, since the figures for R_C and i_1 used above are conservatively low.

However, in practice, it is doubtful that increases as great as this could be obtained for several reasons. In the first place, the problem of the electron optics has been disregarded. It has been assumed that

---

[5] The writers are indebted to Mrs. L. A. Hartig for valuable assistance in making these computations.

the beam was ideally collimated. Actually, this condition would be difficult to realize. With the increased path length, control of the beam becomes more difficult. There would likely be more divergence and consequent loss of current to the aperture edges.
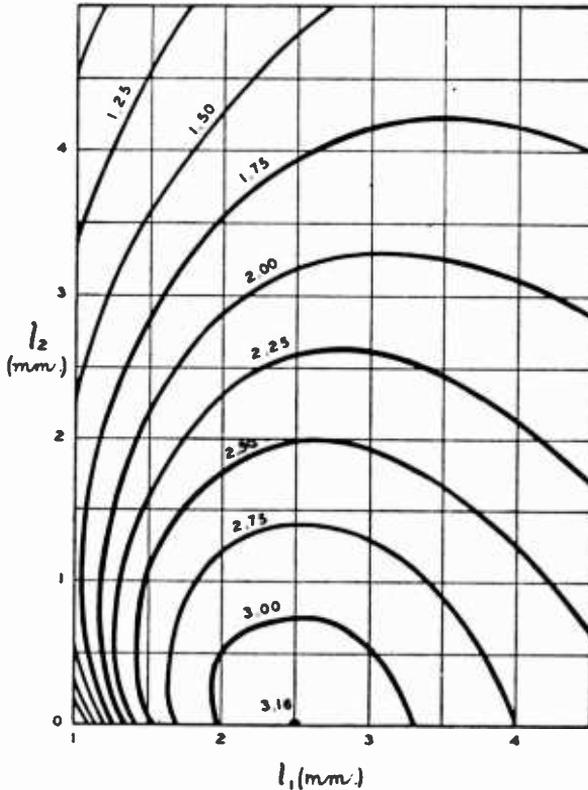
Fig. 4—Constant-efficiency contours (in per cent) versus drift distance in the reflector space l_1, and in the cathode space l_2, for a triple-transit reflex-klystron oscillator.

Closely associated with the beam control is the shape of the grids, since this is a primary factor in determining the shape of the electron wave fronts which form in the beam due to bunching. For a homogeneous, perfectly collimated beam, and with flat grids, flat wave fronts will form. The reflecting field should then be such as to return flat wave fronts to the grids. However, with curved grids, curved wave fronts are formed and the reflecting field must then be such as to return wave fronts having the same curvature as that of the grids. Otherwise all parts of the wave front or bunch will not pass through

the grids at the optimum phase. With two driving transits this problem becomes considerably more difficult, since the wave front shape must conform to the grid shape on both transits. On the first driving transit the shape of the reflector can be varied to produce the correct wave front shape, but on the second driving transit the cathode acts as the reflector, and it probably would be difficult to shape it so as to fulfill simultaneously its function both as cathode and reflector. A possible solution is to use parallel plane grids with a closely spaced parallel plane cathode. The construction of grids which will remain flat when raised to the usual high operating temperatures is a considerable problem that has not yet been solved, but this does not seem to be a difficulty which could not be overcome in a practical fashion.

The question of space charge must also be considered. It is highly desirable to operate with space-charge-limited current for the sake of stability. However, with a cathode closely spaced to the first grid, as it must be to act efficiently as a reflector, space-charge limitation may not always be obtained. In designing a multiple-transit tube this factor should be considered, and the cathode-grid spacing should be chosen to give space-charge limitation, if possible.

A DEVELOPMENTAL PULSE TRIODE FOR
200 KILOWATTS OUTPUT AT 600 MEGACYCLES*†

By

L. S. Nergaard, D. G. Burnside and R. P. Stone

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—The pulse triode A-2212 is a cylindrical triode which gives a peak power output of 200 kw. at 600 Mc. with a tunable external circuit. The tube and its pulse and c.w. performance are described. One of the single-tube circuits developed to test the tube is also described.

INTRODUCTION

EARLY IN 1942, the development of the pulse triode now known as the A-2212 was undertaken under a Navy contract.[1] The tube was intended for use in search radar and was to meet the following electrical requirements: (1) it must have a peak pulse-power output in excess of 100 kilowatts at 600 Mc. with a duty of 0.1 per cent and a 5-microsecond pulse length; (2) it must be operable with external circuits capable of a wide tuning range; (3) it must be air-cooled; and (4) it must operate with no applied voltage in excess of 15 kilovolts. In addition, it was considered essential that the tube be compact, have few and simple parts, and be easy to manufacture on a mass-production basis.

All these requirements were met in the developmental tube H-2614 and the subsequent modification known as the A-2212. After improved circuits for the tube had been developed, it was found that the peak power output per tube at 600 Mc. with 0.1 per cent duty and a 5-microsecond pulse was about 200 kilowatts. This provides a comfortable safety factor over the original requirement.

THE TUBE

The first problem in the design of the tube was the choice of basic geometry. After both cylindrical and planar geometries had been considered, the cylindrical structure was chosen for the following reasons: (1) In the planar electrode structure with cylindrical symmetry, the

---

voltage distribution across the cathode is a Bessel distribution. Because the area of the cathode can be increased only by increasing the radius of the cathode, the cathode area which can be usefully employed at a given frequency is definitely limited. With a cylindrical structure, the voltage distribution is sinusoidal axially and uniform angularly so that the cathode area can be increased indefinitely, in principle, by increasing the radius as long as the axial length is held constant. This argument was given considerable weight because it seemed likely that the H-2614, if successful, might be used as the basis for future tubes of higher power outputs. (2) The cylindrical structure is more likely to be mechanically stable under varying temperature conditions. (3) The cylindrical structure leads to a tube of smaller radius, which is of some consequence in the design of compact circuits for the tube. (4) The cylindrical structure is more economical of cathode-heating power than the planar structure.

When the basic geometry had been decided upon, the cathode was designed. Previous experience had led to two rough empirical relations for the design of oxide-coated cathodes for pulsed triodes: (1) The power output of pulsed triodes at about 600 Mc. is 1 kw. per ampere of emission. At first glance this relation seems a little absurd in that the operating voltage does not appear. However, the starting time and the peak power output of a pulsed triode both increase as the shunt load resistance is increased. When a reasonable compromise is made between the peak power output and the additional power dissipated at the anode because of the starting time, the load resistance turns out to be such that a peak power output of 1 kw. per ampere is obtained when the tube is operated up to its emission limit. (2) The peak-pulsed emission of an oxide-coated cathode is about 12 amperes per square centimeter with a 5-microsecond pulse.

On the basis of these two relations, a cathode area of 13.5 square centimeters was decided upon. This gives about 160 amperes of emission and allows a reasonable safety factor. From this point on, the electrical design proceeded along the usual lines for the design of high-frequency triodes.

Fig. 1 shows the tube H-2614 in section. The cathode is a thimble, oxide-coated on the cylindrical surface. The cathode is supported by tabs on a copper-plated Kovar cylinder about ¾ inch in diameter. This large-diameter cathode lead makes possible the use of a reasonably smooth transmission-line circuit between the cathode and grid. The lower end of the cathode lead is pierced by three eyelets. The axial eyelet carries the exhaust tubulation. The other eyelets carry the heater lead and the getter lead. The heater and getter currents are

returned through the cathode lead. After the tube is exhausted, the cathode lead is extended by a copper thimble which serves to protect the tubulation and leads. The heater lead is brought out axially through this thimble.

The cathode is heated by a small tungsten helix on the axis of the cathode. An upper and a lower heat shield serve to reduce the end losses so that the cathode may be brought to operating temperature with about 40 watts of heater power.
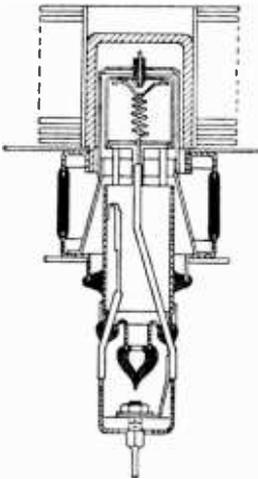

Fig. 1—Cross section of the H-2614 pulse triode.

The grid is a squirrel cage consisting of 90 platinum-clad molybdenum wires, 0.007 inch in diameter, supported by a cone welded to the grid flange. The upper end of the grid is held in alignment by a quartz bead on an axial pin on the cathode. The cathode-grid spacing is 0.019 inch.

The anode is a copper thimble with a U-shaped Kovar annulus silver-soldered to the lip. The internal diameter is such as to give a grid-to-anode spacing of 0.070 inch. The anode is cooled by a "horiozntal-type" radiator having an external diameter of 2 inches.

The subsequent tube A-2212 differs from the H-2614 in that the getter is dispensed with, the exhaust tubulation is of copper and is brought out through the end of the anode, the heater lead is brought through the cathode lead axially, and a "vertical" radiator is used.

Fig. 2 is a photograph of the individual tube parts of the H-2614, some of the subassemblies, and of the assembled tube. An A-2212 is also shown. The low-voltage plate characteristics of a typical tube are shown in Fig. 3. Other data of interest are given in Table I.

Table I

Heater voltage	5 volts
Heater current	8 amperes
Grid-to-plate capacitance	15 $\mu\mu$fd.
Grid-to-cathode capacitance	39 $\mu\mu$fd.
Plate-to-cathode capacitance	0.82 $\mu\mu$fd.
Amplification factor	30
Mutual conductance	0.024 mhos at $I_B = 0.25$ amperes
Plate resistance	1250 ohms at $I_B = 0.25$ amperes
Maximum anode dissipation	300 watts with 10 cubic feet of air per minute
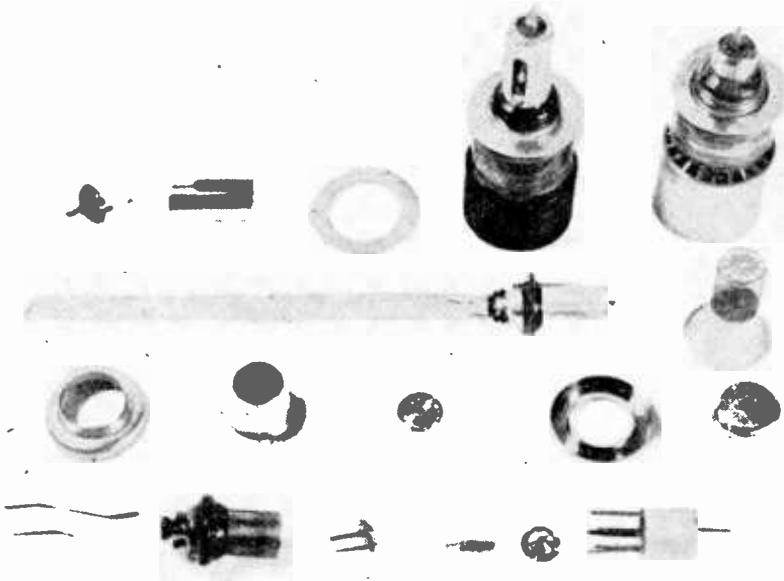
Fig. 2—Individual parts, subassemblies, and a completely assembled H-2614. An A-2212 is also shown.

CIRCUITS

While the principal object of the work described in this paper was to design a tube to meet certain specifications, a considerable amount of work was devoted to circuits for the tube. This attention to circuits is natural in the case of high-frequency tubes because a considerable
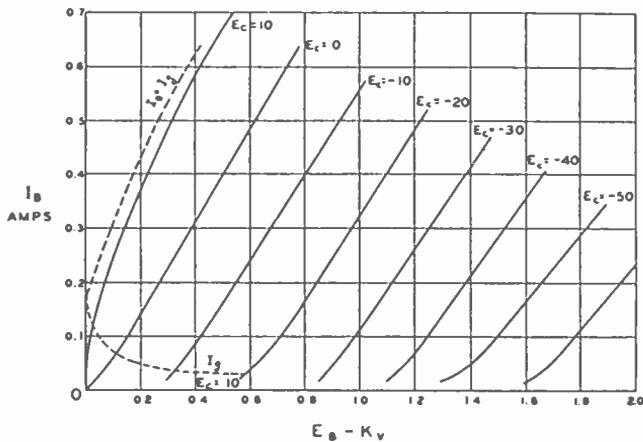
Fig. 3—Low-voltage plate characteristics of the H-2614.

portion of the circuit reactances lie within the envelope of the tube. Hence, it is impossible to design the mechanical features of a tube without a circuit in mind. The H-2614 was designed with two circuits in mind, one with a cavity between grid and anode, the second with a half-wave transmission line between grid and anode. Both of these circuits were built and used to test tubes.

A triode with flange "leads" fits most naturally into a grounded-grid circuit with a tuned circuit between the grid and anode and a second circuit between the cathode and grid. When such a circuit arrangement is used as an oscillator, the cathode-grid circuit is adjusted to have a capacitive reactance and the oscillator operates as a Colpitts oscillator. If the cathode-to-anode capacitance were adequate to support oscillation under loaded conditions, the shunt inductance in the cathode-grid circuit would serve only as a choke for the filament leads. However, in a tube with flange leads the internal shielding is usually good enough so that the cathode-anode capacitance is too small to support oscillation under loaded conditions. The cathode-anode capacitance can be increased by providing additional direct capacitance between the cathode and anode within the tube. This method of increasing feedback has several objections. First, the feedback can be adjusted only by tuning the cathode circuit. This method of adjustment gives only one degree of freedom so that the magnitude and phase of the feedback cannot both be adjusted to obtain optimum operation, a severe limitation when electron-transit times are large enough to produce appreciable phase shifts in the tube currents. Hence, while the use of a properly chosen feedback capacitance within the tube is quite satisfactory for a narrow frequency range, it is not too satisfactory when a very wide frequency range must be covered. Second, a tube with enhanced feedback is not well suited to both oscillator and amplifier use. The alternative is to use external feedback. Then it is relatively easy to obtain two degrees of freedom for the adjustment of feedback, and in addition the oscillator tube is not totally unsuited to amplifier use. These considerations led to the adoption of external feedback systems for the oscillators built to test tubes.

The "half-wave oscillator" circuit is shown in Fig. 4. The grid-anode circuit is a coaxial transmission line, effectively a half-wave-length in length. One quarter-wave of this circuit may be considered as the plate tank and the other quarter-wave as a blocking capacitance which presents a very low reactance to r.f. currents and a very high reactance to low-frequency currents. This feature is particularly important in pulse applications where the pulse shape may be badly distorted by reactance across the output of the pulser. Because this

circuit operates in its fundamental mode, the oscillator is free from mode switching. The part of the line external to the tube consists of an inner conductor comprising the radiator of the tube and a cylindrical extension of the radiator, and an outer conductor in the form of a cylinder 4 inches in diameter. The outer conductor extends beyond the inner conductor so that the line is terminated in a cutoff waveguide. This makes possible the admission of cooling air to a vertical-anode radiator through the end of the line without radiation losses. In order

that the circuit may withstand as high voltages as possible with the given external diameter of 4 inches, the external line is proportioned to effect a compromise between the voltage gradient across the line, which varies approximately inversely as the spacing between conductors, and the step-up in voltage between the low-surge-impedance interelectrode line within the tube and the relatively high-surge-impedance line outside the tube, which varies approximately in proportion to the spacing between conductors. The plate voltage is fed through a lead lying in the nodal surface of the circuit.
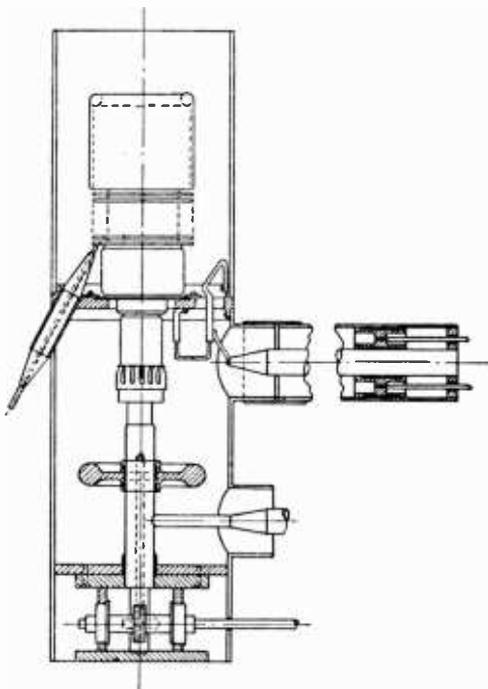
Fig. 4—Cross section of the "half-wave" oscillator circuit.

The cathode-grid circuit consists of a relatively high-surge-impedance coaxial transmission line operating in the three-quarter-wave mode, with a coaxial blocking capacitor adjacent to the tube. The line is tuned by a torus-shaped capacitor which slides on the inner conductor of the line. The oscillator output is obtained from the cathode circuit by a tap directly on the line. The output is taken from the cathode circuit in order to keep the anode circuit clear of objects which increase the voltage gradients and hence induce spark-over, and to take advantage of the low transformation ratio necessary to match a 50-ohm load into the cathode circuit.

The external feedback system consists of two loops in series, one in the grid-anode circuit and the second in the cathode-grid circuit. These loops are tuned by a stub tapped on the loop in the cathode-grid circuit. The stub and tuning capacitor provide the two degrees of freedom necessary for the proper adjustment of the feedback.

The "half-wave" circuit described above and the "cavity" circuit, which is not described for lack of space, were used to test tubes while they were being made in the laboratory. Other circuits were subsequently built. One of these circuits was a modification of the half-wave oscillator which could be operated over the frequency range 400 to 1200 Mc. Push-pull circuits which will be described elsewhere were also built.

PERFORMANCE

The performance data presented in this section were obtained in the half-wave circuit. In each case, the tube was biased by a cathode resistor. There were two reason for the use of cathode bias. With a grounded-grid circuit, the oscillator shell and output system can be operated at ground potential when cathode bias is used. Secondly, with cathode bias, operation is stable even when grid emission is large enough to make the plate current exceed the cathode current slightly.

As is customary with pulse tubes having oxide-coated cathodes, the tube was anode-pulsed. What is called the anode voltage in the data is actually the grid-to-anode voltage, i.e., the sum of the anode voltage and the grid-bias voltage. Similarly, the quoted efficiency is the over-all efficiency, not the anode efficiency. In the pulse tests, the duty was 0.1 per cent and the pulse was substantially square, so the average power output and average anode current may be computed by dividing the pulse power and current, respectively, by 1000.

Fig. 5 shows typical pulse operating data on a laboratory-made tube at 600 Mc. The peak power output, the over-all efficiency, and the anode current are plotted against the pulse anode voltage for two values of cathode-bias resistor. It will be noted that in each case the peak power output varies as the 5/2 power of the anode voltage, and the anode current varies as the 3/2 power of the anode voltage until grid emission becomes appreciable, at which point the current begins to rise more rapidly. The efficiency increases slowly with voltage for the lower voltages, and then drops as grid emission sets in. With the 60-ohm bias resistor, an output of 160 kw. was obtained at 15 kv. In this case, the output was limited by the pulser, but the indications of grid emission suggest that the useful operating point has been passed.

With the 10-ohm bias resistor, an output of 265 kw. was obtained at 12 kv. In this case, the output was limited by circuit flashover. In fact, carbontetrachloride vapor was used to coax the circuit up to this power level. However, again the evidence of grid emission suggests that the useful operating point has been passed.

Fig. 5 — Typical pulse operating data at 600 Mc.

Inspection of the efficiency curves shows that a given output may be obtained with a fixed input for a wide range of anode current. From the standpoint of voltage breakdown, it is advantageous to operate with a low voltage and a high current. At first, it was felt that high-current operation might seriously impair the cathode life. Life tests on tubes operating under conditions comparable to those with the 10-ohm cathode resistor and a power output of 150 kw. have shown that a life in excess of 1000 hours may be expected under these conditions.

While the H-2614 was designed for pulse operation and has a cathode much larger than is required for a continuous-wave tube of comparable average power rating, it was quite natural to make some tests of its performance under c.w. conditions. The results of such a test at 600 Mc. are shown in Fig. 6, in which the power output, plate current, efficiency, and filament voltage are plotted against the anode voltage. For each point on the curves, the cathode bias, feedback, and load were adjusted for maximum power output. The value of the cathode resistor for each point is shown on the power curve. The effects of the cathode-to-grid electron-transit time are evident in the reduction in heater power necessary to keep the tube stable and in the relatively large plate

Fig. 6—Typical continuous-wave performance at 600 Mc.

current required for best performance at low voltages. The effects of the grid-to-anode transit time are evident in the very rapid rise of power output with plate voltage. The maximum power output was limited by the tube stability. At the maximum power point, the heater power had been reduced to the point where any perturbation caused the emission to fail and the tube to drop out of oscillation, or caused the cathode temperature, and consequently the grid temperature, to rise so rapidly that the tube "ran away."

To get an idea of the "high-frequency limit" of the tube, it was operated as an oscillator with 250 volts on the anode. It oscillated at all frequencies up to 1100 Mc., at which frequency the efficiency dropped to zero. At 1000 Mc., a power output of 2 watts was obtained.

While the tube was designed specifically for pulse operation and is in some respects poorly designed for c.w. operation, its c.w. performance is such that it has had some application as a c.w. amplifier and oscillator.

CONCLUSION

In conclusion, it may be said that the pulse triode described above meets all the initial specifications with regard to power output, tunability, cooling, and maximum applied voltages. It meets the minimum pulse-power requirement of 100 kw. at 600 Mc., with a 5-microsecond pulse and 0.1 per cent duty, with a comfortable margin of about 100 per cent. It also gives a c.w. power output of 100 watts at 600 Mc., and has been operated as a c.w. oscillator at frequencies up to 1100 Mc.

ACKNOWLEDGMENT

A NEW 100-WATT TRIODE FOR 1000 MEGACYCLES*†

BY

W. P. BENNETT, E. A. ESHBACH, C. E. HALLER AND W. R. KEYE

Tube Department, RCA Victor Division,
Lancaster, Pa.

Summary—The design and development of a 100-watt, grounded-grid triode for operation at full ratings up to 1200 Mc is described. Unusual mechanical design features have been utilized to achieve a tube which not only is capable of excellent performance at ultra-high frequencies, but which can also be manufactured by production-line methods.

The outstanding design features making this performance possible include close spacing of a coaxial structure to give high perveance; precision cold working of metals at high unit pressures to fabricate all electrodes; and assembly of the tube using localized rf heating methods and precision jigs to maintain accurate spacing of electrodes.

Circuit and performance data of this new tube are given as a power oscillator, as well as a uniquely neutralized ultra-high-frequency power amplifier.

INTRODUCTION

THE design and development of a compact, forced-air-cooled uhf power triode capable of delivering 100 watts at an efficiency of 40 per cent in a cw amplifier operating at 1000 Mc presented stringent electronic and physical requirements which were met by novel methods of fabrication and assembly. The resultant type, the 5588, has a maximum of operational stability together with good circuit adaptability, achieved through the use of a coaxial electronic structure[1,2] with supporting elements likewise coaxially aligned.

DESIGN AND CONSTRUCTION

General Construction Features

The electrode structure of the 5588 consists of three closely spaced coaxial cylindrical elements: the unipotential oxide-coated cathode, the grid, and the anode. A cross-sectional view of the tube is given

---

[1] R. R. Law, D. G. Burnside, R. P. Stone, and W. B. Walley, "Development of pulse triodes and circuit to give one megawatt at 600 megacycles," *RCA Review*, Vol. 7, pp. 253-265; June, 1946.

[2] L. S. Nergaard, D. G. Burnside, and R. P. Stone, "A developmental pulse triode for 200-kilowatt output at 600 megacycles," *Proc. I.R.E.*, Vol. 36, pp. 412-416; March, 1947.

in Fig. 1. Because the axial length of the electrode structure is about ¼ inch, which is short in comparison with a quarter wavelength at 1000 Mc, essentially equal rf voltages exist at all points over this area. The grid, which consists of a cylindrical array of short wires aligned parallel to the tube axis, has low inductance. The short grid wires provide a maximum of end or conduction cooling, allowing greater grid dissipation before instability occurs due to grid emission. The short coaxial structure provides considerable freedom from instability due to shifts in electrode spacing, buckling, or warping of the electrodes. The members which support the electrodes and provide for the transition between the external circuit and the electrodes are continuous,
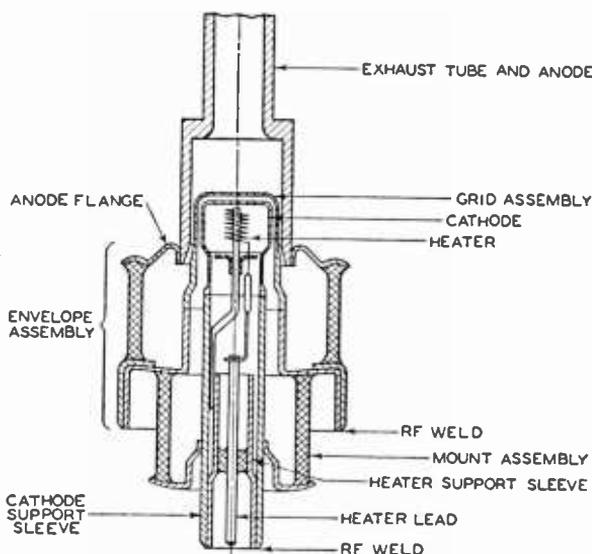
Fig. 1—Cross-sectional view of the 5588.

low-inductance cylinders. These members serve the further function of either providing or preventing thermal isolation of the various electrodes.

General Assembly Methods

The achievement of these construction features requires a well-integrated mechanical design based upon precision assembly methods and parts-manufacturing techniques. The mechanical design of the tube is such that the main supporting unit contains a base reference surface, the inside cylindrical surface of the cathode support sleeve. Each critically spaced element likewise has an appropriate mechanical reference surface. During assembly, as each element is fastened to the main unit its reference surface is aligned by the use of jigs with

the base reference surface. In this way, the proper cathode-grid-anode alignment is obtained.

The jigging tools consist of accurately ground V-blocks and cylindrical alignment mandrels as shown in Fig. 2. If mandrels of exactly equal diameters are placed in V-ways whose surfaces are true intersecting planes, the center lines of the mandrels will coincide. Hence the alignment of parts held by these mandrels will be limited only by the accuracy of the jigging surfaces. The critical intersecting plane surfaces of the V-ways can be ground to a high degree of accuracy in a surface grinder, and the pairs of cylinders or mandrels can be ground to virtually equal diameters in a cylindrical grinder between dead centers. The use of V-blocks eliminates the need for the bearing tolerances of a coaxial jig consisting of mandrel and coaxial bearings, or dowel pins and fitted holes.
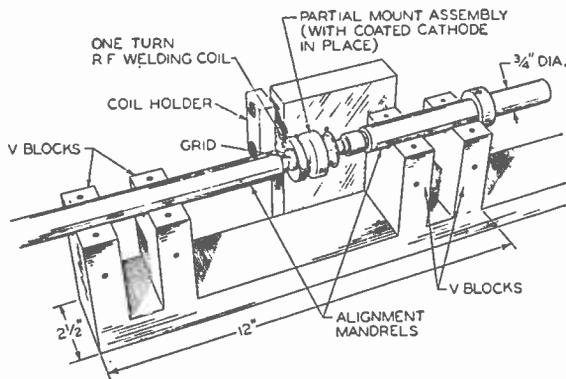
Fig. 2—Jigging equipment consisting of V-blocks and cylindrical mandrels.

Radio-Frequency Induction Welding

The reference-surface-assembly principle with V-blocks and mandrels is particularly useful when rf induction-welding methods are used for making the metal-to-metal joints. Figs. 2 and 3 show the arrangement of the parts of rf-welding the copper grid directly to its Kovar support. The grid is jigged by its reference surface, and the mount assembly is jigged by the inside bore of the cathode support cylinder. The welding is accomplished by sending a heavy surge of rf current through an appropriately shaped induction coil, the contours of which mate the members to be welded. The induced current is intensely concentrated in the weld area. The heat produced by the induced current brings the weld area up to the welding temperature in a fraction of a second. Because the heating is concentrated and requires only a fraction of a second, the welding is accomplished with

a minimum of total heat, and the delicate tube electrodes and glass-to-metal seals remain at low temperature and are unharmed.

In addition, because part reference surfaces and jigs remain cool, they retain their accuracy. It should be noted that the welding coil makes no direct contact with the work or jigs which might cause mis-alignment due to temporary deformations. The welds produced are symmetrical, continuous, and uniform, providing great mechanical strength and high thermal and electrical conductivity. The use of rf welding for assembly does not adversely deform the parts nor set up uneven stresses which pull the parts out of alignment when the jigs are removed. In addition, the process can be performed in any non-conducting medium. Neutral or reducing atmospheres can be used if it is desirable to prevent traces of oxidation.
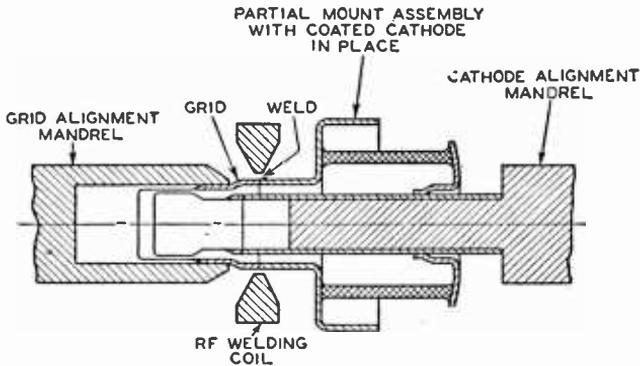
Fig. 3—Arrangement of parts for rf-welding the grid to Kovar support; cross-sectional view.

Assembly of Components

The tube is made up of three separate assemblies utilizing glass-to-metal seals and identified in Fig. 4 as mount assembly, envelope assembly, and heater assembly. As shown in Fig. 5, the mount assembly comprises the mount support assembly, the cathode assembly, and the grid. The separate parts (Fig. 6) of the mount support assembly include the cathode support, which is coaxially fastened by seals to the grid support and insulated from it with an intermediate length of glass tubing. The envelope assembly consists of the grid flange (subsequently welded to the grid support), which is separated by a length of glass tubing from the copper anode brazed to a sealing alloy flange. These three assemblies must be made so that the surfaces involved in the final welds to the critical electrode members are closely aligned. Glass seals are made through use of rf induction heating of the metal sealing
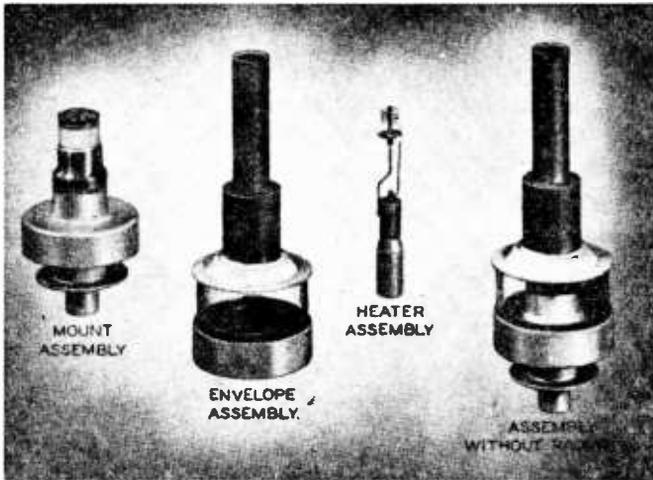
Fig. 4—Three main components and final assembly.

surfaces to bring the metal-glass boundaries to the temperature re-
quired for glass sealing. Because this technique obviates the use of
fires in the sealing operation, the jigging problem is immensely simpli-
fied. In addition, it is simpler to control the physical chemistry of the
glass-to-metal sealing phenomena by appropriate atmospheres and the
accurate localized heating. The use of suitable jigs and limit stops
permits placing cold parts on the jigs and making glass-to-metal seals
of high merit at a rapid rate.
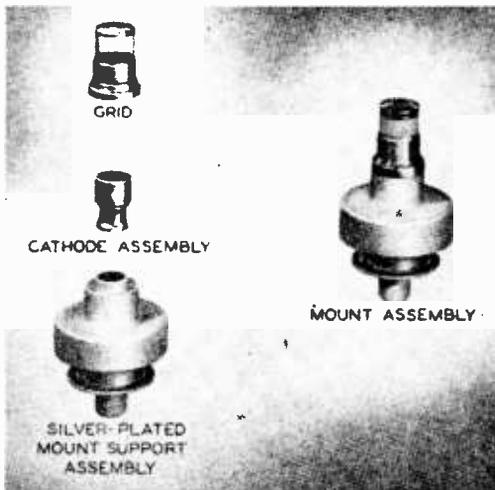
Fig. 5 — Mount assembly
and components.

Parts Features

The supporting members of the tube which make up the assemblies are produced by conventional drawing and forming methods. However, these members are specially shaped from a structural standpoint to yield sufficient "strain isolation," so that a deformation or "strain" in one section of the tube will not adversely affect some other critical area. For example, if the glass sealing surface of the sealing-alloy anode flange in Fig. 1 is to match the thermal expansion of the envelope assembly glass, it must be properly shaped so that the high thermal expansion of the copper anode does not affect this glass seal. Strain isolation is achieved by appropriate positioning of sharp changes in contour which hinder spreading of a strain or deformation over a great area.
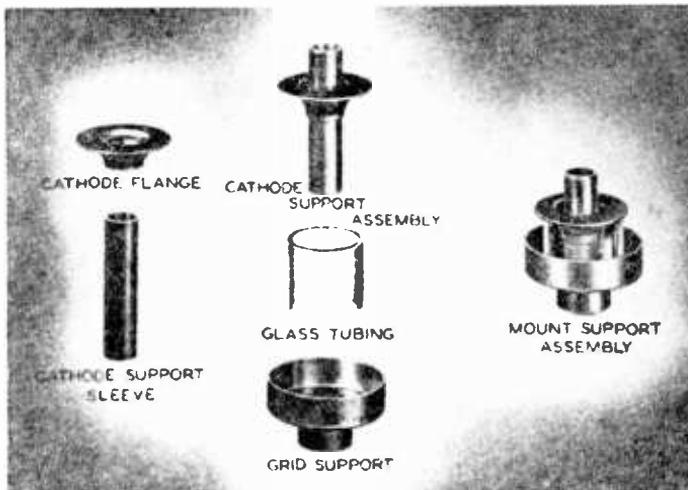
Fig. 6—Mount support assembly and components.

Chief among the parts that require special attention are the electrodes. For example, to reduce interelectrode capacitance to a minimum the cathode (Fig. 7) is formed into a re-entrant shape at high unit pressures in a precision die by an expanding punch. The intermediate cathode-supporting member thermally isolates the hot cathode, yet provides a mechanically strong support and a continuous low-inductance rf path. It achieves these qualities by virtue of its extremely thin wall section, which is fabricated by a special forming process from the parent thick-wall tubing. The cathode and the thin-wall support are rf-welded together to form the cathode assembly. Another special construction is the one-piece conduction-cooled grid

(Fig. 8). This construction was chosen because of the thermal grid dissipation necessary for the relatively high current and power densities required for uhf operation. A radiation-cooled grid would have to operate at a high temperature in order to dissipate the necessary energy. In a tube using an oxide-coated cathode there is usually a certain amount of barium deposited on the grid wires during tube processing, or possibly during operation. This barium lowers the temperature at which grid emission occurs and thus limits the working temperature of the grid to a value much lower than is permissible in a tube utilizing, for example, pure tungsten as a source of emission. Hence, because a radiation-cooled grid is inherently a high-temperature device, it has relatively low dissipation capabilities in a tube havng an oxide-coated cathode. It appears, therefore, that the most desirable

RE-ENTRANT CATHODE

WELD

THIN SUPPORTING MEMBER   A

WALL THICKNESS AT B IS 60 TIMES THAT AT A   B
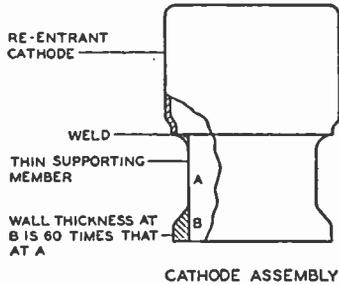
CATHODE ASSEMBLY

Fig. 7—Cathode assembly, cutaway view.

GRID BLANK

PARTIALLY PROCESSED GRID

COMPLETED ONE-PIECE GRID

Fig. 8—Fabrication of one-piece grid.

way to handle the grid-dissipation problem is to make the entire grid out of one piece of a material having high thermal conductivity, and to provide for removal of the heat to the outside of the tube by thermal conduction. A short vertical grid affords an effective thermal path for conducting the heat to a thermal "sink" outside the tube and makes it possible to hold the temperature of the grid wires to the desired value. Limiting the length of the grid verticals does not affect rf performance because the intended use of the tubes also dictates short electrode structures. The complete grid in the 5588 is fabricated from a single piece of copper by application of a special cold-forming process. By means of this construction, thermal barriers from indeterminate welded joints, or uneven thermal-mechanical strains that must be "stretched out," are not introduced. The outside cylinder of the supporting ring is automatically made into a jigging surface concentric to close tolerances with the grid-wire cylinder. In this way, induction welding to the grid support assembly is expedited. This joint com-

pletes the heat path to the outside of the tube. The thermal "sink" necessary for grid cooling is supplied by the anode-cooling air passing over the grid flange and by the contacting circuit fingers.

The anode, Fig. 9, is also a unique part in that it, together with a metal exhaust tube, is cold-extruded from a single cylindrical piece of copper cut from stock. This technique eliminates the possibilities of leaks due to overworked walls of drawn cups or oxygen-bearing copper incurred during intermediate annealings. In addition, an extremely fine finish is achieved and inherently accurate jigging surfaces are available for speedy assembly operations. One stroke of the press produces a finished anode complete with metal exhaust tube.

Assembly of Complete Tube

The mount support assembly (Fig. 5) contains the base reference surface and is the main supporting unit. To this, the cathode and the grid are radio-frequency welded in succession, while each is held very accurately concentric with the inside surface of the cathode support by means of the mandrels and V-blocks. In the final assembly of the tube, the three separate assemblies (Fig. 4) are brought together and the final metal-to-metal closures are then made utilizing rf welding. The envelope assembly containing the anode, and the external grid terminal separated by a section of glass, is placed over the mount

ANODE BLANK   EXTRUDED ANODE

Fig. 9—Fabrication of anode.

assembly and welded to it to make the main seal closure. The anode is held concentric with the grid-cathode structure in the V-blocks and mandrels. The final closure is made by placing the self-aligning heater assembly into the cathode-support tubing and rf-welding it in place. The tube is then evacuated followed by a cold-metal pinch-off at the exhaust tubulation above the anode. After the pinch-off, the radiator is soldered to the anode.
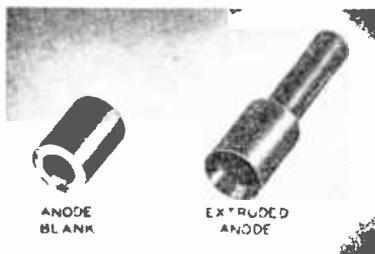
Because a uhf circuit is a precise mechanical device of which the tube must become a part during operation, the contact surfaces of the tube must also be precisely defined and maintained. The same coaxial alignment methods which produce the precise cathode-grid-anode alignment also serve to provide accurate alignment of the contact terminals. The external contact terminals consist of two coaxial cylindrical surfaces (the cathode and grid terminals) having progressively larger diameters, and a plane surface in the form of an annular ring (the

top surface of the anode ring) which is perpendicular to the center lines of the above cylinders (see Fig. 10). Since the outside diameter of the anode ring is not a contact surface, clearance may be allowed in the circuit adaptor around this ring so that motion in directions · perpendicular to the centerline is not limited by the anode ring. This clearance allows the tube to seek its position freely in the grid and cathode contacts, because it is plugged into the circuit before the clamp which anchors the tube in place is placed over the anode ring.

OPERATION AND APPLICATION

Operation Considerations

Electron-transit-time effects were evidenced by overheating of the cathode and resultant short cathode life at the higher frequencies.

Fig. 10—Uhf power triode, 5588

This cathode overheating is caused by back bombardment of the cathode by a portion of the space-current electrons at the end of each current pulse which do not have enough energy to pass through the grid plane as the grid swings negative at the end of the voltage pulse. Compensation for back bombardment is accomplished by a reduction in cathode heater input to maintain the cathode temperature at the normal value. Thus, when long life in continuous operation is desired, the tube should first be put into operation with full rated heater voltage and then, as back bombardment progresses, it should have its heater voltage reduced so as to bring the cathode temperature back to normal. The magnitude of the heating due to back bombardment is a function of the operating conditions and the frequency. The recommended heater operating voltage for oscillator service, as shown in the curves of Fig. 11, was determined on an empirical basis by operating tubes under varying conditions of plate current, grid current, plate voltage, and frequency, but with the grid bias adjusted at all times for essentially class-B operation. At each operating condition, the heater voltage was determined at which the tube became unstable, due to lack of emission as judged by the dropping of power output. The interpretation of these data was simplified by utilizing specially prepared graph paper which had equally spaced divisions on the horizontal axis, but scale divisions

on the vertical axis which were exponentially larger according to the exponent n in the relation $P = KE^n$. In this equation[3], which relates power input to a tungsten heater as a function of heater voltage,

$P =$ input to heater in watts

$K =$ constant

$E =$ heater volts

$n = 1.61$ for a tungsten heater.

If power input to the heater is plotted on the horizontal axis and the heater voltage on the vertical axis of this special paper, the resulting curve is a straight line passing through the origin.
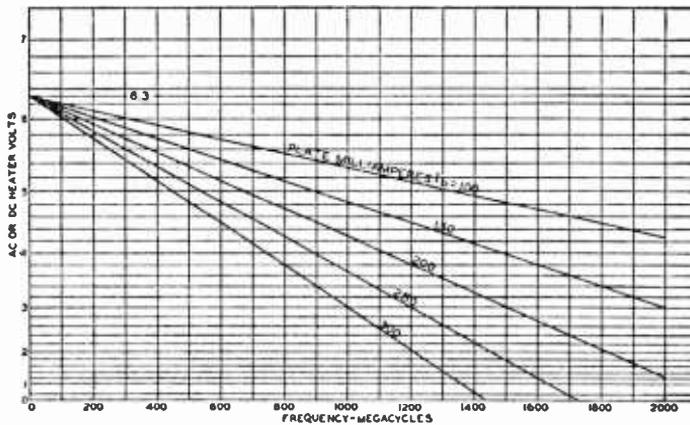
Fig. 11—Heater voltage versus frequency of the 5588 in oscillator service.

Instability curves were obtained by plotting on this paper the points at which the tube became unstable. The data indicated that, in class-B oscillator service, the back-bombardment heating power is essentially proportional to frequency and plate current, while in class-B amplifier service the back heating power is proportional to frequency and grid current. The recommended heater operating voltages as shown in Fig. 11 were arrived at by drawing a line passing through the vertical axis at 6.3 volts parallel to the instability curve. These values, then, include the normal safety factor for operation of oxide cathodes and, as a result, bring the cathode to normal operating temperature and thus insure long life. In oscillator service at an input of 250 watts at a plate current of 300 milliamperes and the heater

[3] Cecil E. Haller, "Filament and heater characteristics," *Electronics*, Vol. 17, pp. 126-131,; July, 1944.

voltage at the recommended 3 volts, the 5588 has operated at an efficiency of 30 per cent with a power output of 75 watts for well over 1000 hours. Tube data of a general nature is shown in Table I. The characteristics of a typical tube are shown in Fig. 12.

Table I—GENERAL DATA

Amplification Factor	16 ($I_b = 250$ ma)
Transconductance	15,000 micromhos ($I_b = 250$ ma)
Direct Interelectrode Capacitances	
Grid to plate	6 $\mu\mu$f
Grid to cathode	13 $\mu\mu$f
Plate to cathode	0.32 max. $\mu\mu$f
Plate Dissipation	200 watts
Over-all Length	$3\frac{5}{16}$ inches
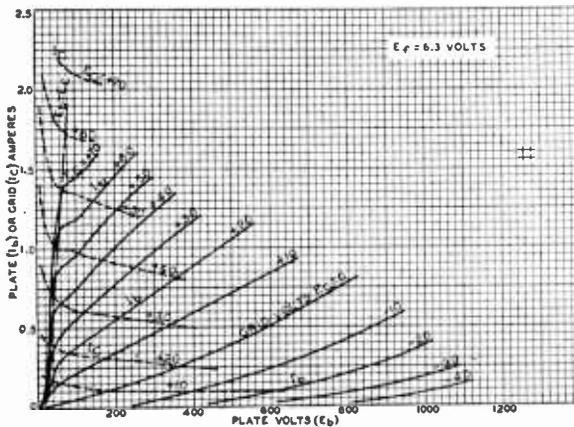
Fig. 12—Characteristic curves.

Application as Oscillator and Amplifier

The 5588 was initially designed for use in cw oscillator service up to 1000 Mc and for a power output of 50 watts with an efficiency of at least 20 per cent. The present tube, however, operated in circuits described in this paper has an average output of 75 watts with an efficiency of 30 per cent. The 5588 has also been operated as a stable grounded-grid amplifier at 1000 Mc with a power output of 100 watts. At 220 Mc, two tubes operated in a push-pull oscillator circuit gave an efficiency of 55 to 60 per cent.

At 1000 Mc the 5588 has been operated only in coaxial grounded-grid circuits because of the self-shielding and low-loss properties of such circuits. Both extended and folded-back types of coaxial circuits have been used. The advantage of the folded-back type of circuit is that the tube may be plugged directly into the circuit.

A diagrammatic sketch of a 1000-Mc oscillator of the extended co-axial type is shown in Fig. 13. The anode and cathode circuits are shorted concentric transmission lines operating in the $\frac{3}{4}$-wavelength mode. Feedback is provided by means of a tuned feedback loop extending into both the anode and cathode cavities. The power output is measured by means of a calibrated water-cooled load. This load is coupled into the anode circuit through a double stub tuner by means of a small probe. Cylindrical mica capacitors are used in the inner conductors of both the anode and cathode lines to isolate the dc plate and cathode-bias voltages. The tube is cooled by means of air blown down the inner conductor of the anode line. A typical operating condition for this circuit is shown in Table II.
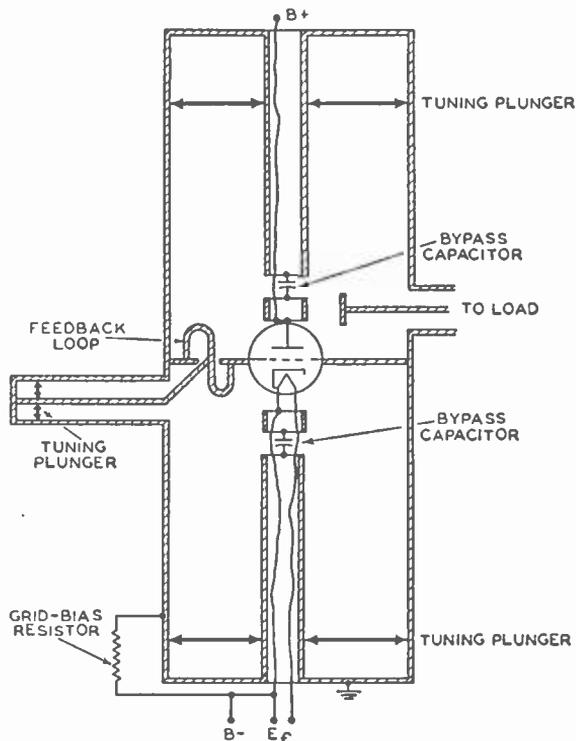
Fig. 13 — Diagrammatic sketch of 1000-Mc oscillator of extended-coaxial type.

Table II—TYPICAL OPERATION* OF 5588 AS GROUNDED-GRID OSCILLATOR AT 1000 Mc

Heater Voltage	3 volts
DC Plate Voltage	835 volts
DC Grid Voltage	—70 volts
From cathode-bias resistor of	205 ohms
DC Plate Current	300 ma
DC Grid Current (Approx.)	40 ma
Power Output (Approx.)	75 watts

* Continuous Commercial Service.

When the 5588 is used as a 1000-Mc amplifier in a folded-back type circuit such as that given in Fig. 14, neutralization is necessary to prevent oscillation. For the purpose, a feedback probe approximately ¼-wavelength long is used to couple the anode and cathode circuits. A probe such as is shown in Fig. 14 will prevent the amplifier from self-oscillation over a range of 100 Mc without requiring adjustment in the length of the probe. Although this type of neutralization does not completely isolate the anode and cathode circuits, it does permit stable amplifier operation. The amplifier circuit shown in Fig. 14 has the anode circuit folded back over the cathode circuit. This type of con-
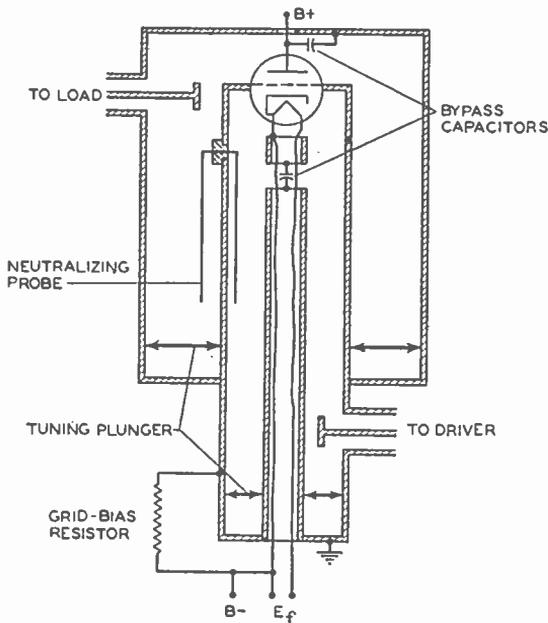
Fig. 14—Diagrammatic sketch of 1000-Mc amplifier of the folded-back type.

struction permits the insertion of the tube into the open end of the circuit. Connections are made to the cathode and grid of the tube by means of contact fingers, while the anode ring of the tube is clamped firmly in the circuit. The anode circuit is a coaxial line operating in the ¾-wavelength mode and the cathode line operated in the ⁵⁄₄-wavelength mode. Radio-frequency power is fed into the amplifier by means of a capacitive probe inserted into the portion of the cathode line which extends below the anode line. Power is taken from the amplifier by means of a probe inserted into the anode line near the tube. For measurement purposes, the power output was fed into a water-cooled load through a double-stub tuner.
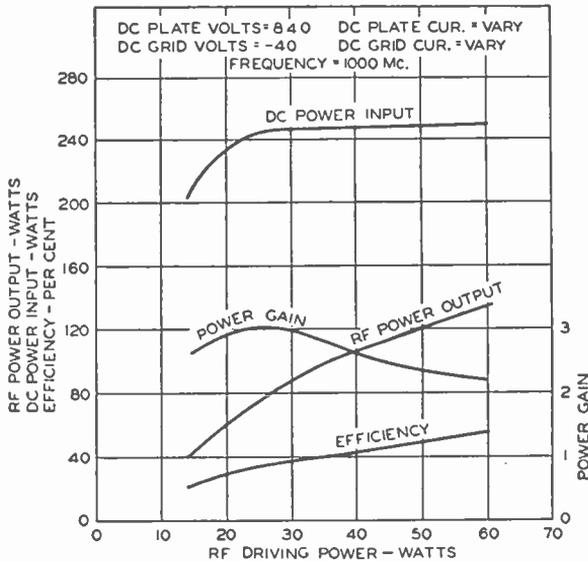
Fig. 15—Performance curve of 1000-Mc amplifier of
the folded-back type.

A performance curve of the amplifier is shown in Fig. 15. This curve shows dc power input, rf power output, apparent plate efficiency, and power gain of the amplifier as a function of rf driving power. The driving power was measured by means of a standing-wave-detector type of wattmeter inserted between the amplifier and driver.

ACKNOWLEDGMENT

The authors wish to acknowledge the numerous contributions to this development from other RCA members at the Lancaster and Camden plants. The coaxial structure employed is a further extension of a design originating at the RCA Research Laboratories at Princeton, N. J.

DUPLEX TETRODE UHF POWER TUBES*†

By

Philip T. Smith and Howard R. Hegbar

Research Department. RCA Laboratories Division,
Princeton. N. J.

Summary—Major factors affecting the design and development of wide-band uhf power tubes are considered and emphasis is given to the television application. A qualitative discussion of methods for obtaining the required performance is presented, and a 5-kw 300-Mc liquid-cooled, internally neutralized duplex tetrode is described.

INTRODUCTION

IN CONSIDERING the design and development of electron tubes suitable for use as grid-modulated television power amplifiers, there are certain performance characteristics that must be attained, and others that are highly desirable. The fixed tube-performance characteristics, such as bandwidth, power output, and carrier frequency, are determined by the standards adopted for television broadcasting, and must be accepted as minimum values in the design. When the other tube characteristics, grid currents, power gain, impedance presented to the modulator, efficiency, and feedback are considered, the desired values are not all attainable in a given design, and the best compromise is sought. Attempts to satisfy requirements of bandwidth, power gain, linearity, ease of modulation, and power output at the higher carrier frequencies can be resolved into a search for means of obtaining large cathode emission-current densities, large average anode-current densities, electrodes capable of handling large power dissipation per unit area of bombarded surface, small inter-electrode spacings, small electron currents to the grids, and a tube geometry and circuit-wise arrangement of tube elements that will provide adequate utilization of the aforementioned.

Realizing that these requirements could not be adequately met if the limitations imposed by conventional tube design were assumed, Zworykin organized a laboratory group for research in high-power electron tubes in 1937. It was his continued interest and good counsel which made possible the development of a 50-kw tube which served as the background for the development of the smaller. higher-frequency power tube herein described.

---

* Decimal classification: R339.2.
† Reprinted from *Proc. I.R.E.*, November, 1948.

Early in 1938 the senior author introduced a duplex tetrode with an electron-beam-forming electrode configuration and high-dissipation anodes. The initial tests were made in continuously pumped demountable metal envelopes. Sealed-off demountable tubes were made possible when fitted with a copper-gasket demountable seal introduced by Garner, a member of the group. Many laboratory tubes were built and tested before a tube capable of a 5-kw output at 300 Mc with a total output bandwidth of 10 Mc could be properly designed. This paper describes such a tube.

DESIGN CONSIDERATIONS

In arriving at a tube design, it is difficult to formulate a mathematical expression containing all of the factors affecting the design and to obtain the unique or the best solution. Tube design represents a compromise between conflicting factors which are individually studied to advantage, and which must be combined with care and ingenuity and with a view toward the circuit and application problems. A detailed analysis of the individual factors will not be attempted here, but only qualitative indications of the trends required for providing an improvement in the factors pertinent to the design of grid-controlled power tubes will be discussed.

The frequencies for the present commercial and experimental television channels are sufficiently high to make the electron-transit time between tube elements of importance. Many of the effects of long transit times are known, and have been observed in cathode back-bombardment, control-grid loading, and loss in efficiency and power output. Therefore, one of the present considerations is that of extending the usable frequency range of grid-controlled tubes by reducing the electron-transit time. Such a reduction in transit time is obtained by decreasing the interelectrode spacings and increasing the electron acceleration. Under conditions of space-charge-limited emission from the cathode, the increased electron acceleration is obtained only when increased cathode-current densities may be drawn. For example, the electron-transit time in a region between parallel-plane electrodes of large extent is proportional to the one-third power of the ratio of the electrode spacing to the current density when operated with space-charge-limited emission and assuming zero emission velocity. In grid-controlled high-frequency tubes, effective electrode spacings are obviously to be made as small as is practical without sacrificing mechanical rigidity of the electrode structure, and the longitudinal thermal conduction required for cooling. Large cathode-current densities are available from such surfaces as the thoriaon-tantalum cathode under steady-

state conditions with a reasonable life. Since we are concerned with continuous operation, the large pulse emission-current densities from barium-strontium-oxide cathodes cannot be utilized.

An attempt to utilize large current densities increases the difficulty of providing an adequate control-electrode configuration, particularly for the high-frequency applications where close interelectrode spacings are of importance. Since grid control of the large-density electron emission is sought, it is evident that a beam-forming electrode arrangement can be used to advantage. A focusing electric field is provided in the region adjacent to the cathode surface by means of focusing elements electrically connected to the cathode, and projecting slightly beyond the cathode into the cathode-grid space. Many such focusing arrangements are possible and the details must be chosen to fit the method of construction and other tube parameters. A typical cathode and focusing-element arrangement is shown in Fig. 1.
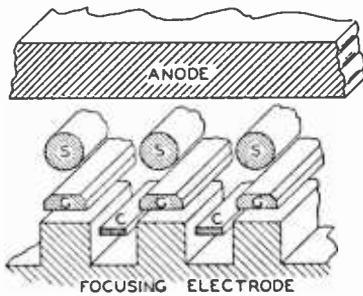
Fig. 1—A beam-forming electrode configuration. $S =$ screen-grid element, $G =$ control-grid element, $C =$ electron-emitting filament.

Since the instantaneous voltages of the control grids and the anodes vary with respect to time, the electron beam focusing is not constant over the operating cycle. In a practical design, the electrode configuration and electric fields are so arranged that at the instant of maximum beam spreading the portions of the beam intercepted by the No. 1 and No. 2 grid elements are small enough to prevent excessive power dissipation at these elements. Also, the portion intercepted by the No. 1 grid should be sufficiently small to cause a negligible variation in the impedance presented to the rf driving stage and to the modulator over the modulation cycle. It is just this beam spreading and the possible formation of potential minima in the interelectrode regions that determines the values of applied voltages and the amplitude of the control-grid and anode rf voltages for a given structure. The electric fields must be sufficiently great at the maximum of the grid-voltage swing to support the large space-current densities and to maintain sufficiently narrow beams. Considerations of transit time and beam spreading

determine the minimum gradients to be provided in the regions between the cathode and No. 1 grid, between the No. 1 grid and No. 2 grid, and between the No. 2 grid and the anode. Because of the gradients required in the latter region of a high-frequency power tetrode, the minimum of the instantaneous anode volage of a high-frequency tetrode should be substantially above the screen-grid voltage to utilize in the best manner the current available in the plane of the No. 2 grid.

It is well recognized that the wide-band high-frequency power tube requires large power dissipation per unit area at the anode. In order that the greatest power and bandwidth be obtained from a tube with a given anode current, the ratio of the tube output capacitance to the anode current should be small and the anode power dissipation per unit area should be large. When sufficient anode dissipation is available, the maximum tube output load impedance is determined by the required bandwidth and the tube output capacitance plus whatever effective capacitance is added by the output circuit. It is assumed that the anode voltage can be increased to the required value without failure of the tube or circuit. This assumption is reasonable for total bandwidths of 10 Mc or more, but leads to excessively high voltages for bandwidths less than 1 Mc. When the allowable anode dissipation is too low, in a given tube, the tube must be operated with a reduced anode voltage, a reduced output load impedance, and consequently a reduced efficiency and power output. At this expense, an output circuit is obtained whose bandwidth is increased beyond that required.

The requirement of large anode dissipation per unit area of bombarded surface can be met by the use of high-velocity water in cooling channels formed in a copper or silver anode body. This construction increases the area of metal in contact with the water and reduces the tendency for diversion of the cooling water by steam bubbles. Such an anode structure is shown in Fig. 6. This structure permits anode dissipations of from 500 to 1000 watts per square cm averaged over the anode face. This is to be compared with an allowable dissipation of 50 to 100 watts per square cm in conventional structures.

With an electron-beam system such that but a small portion of the beam current is collected by the No. 1 grid and such that the electron-transit-time effects are small, a large power gain can be achieved if the circuit losses are low and if there is no feedback. Small amounts of feedback from the output circuit to the input circuit give rise to assymmetric distortion of the sidebands, while larger values of feedback will produce instability and oscillation. It is common practice to add a load to the input grid circuit to minimize the effects of the feed-

back, even though such loading decreases the power gain. The feedback can be made small by designing the tube such that the anode and output circuit are shielded from the input by special grids, as is done in the tetrode, and somewhat similarly in the grounded-grid triode; or a neutralizing circuit may be applied. In the design of grids, the requirements for obtaining good shielding conflict with those for obtaining the desired electronic performance, and a compromise is usually made. As a result, most tetrodes and grounded-grid triodes require some neutralization for wide-band amplification at a high frequency. The selectivity of the neutralizing circuits for such application introduces an added difficulty in obtaining wide-band amplification. The selectivity of the neutralizing circuits can be decreased by making their elements very short compared to a quarter wavelength at the operating frequency. In a duplex tetrode arranged for push-pull operation, such short neutralizing elements can be located internally. This is done in the described developmental tube.

In order that the circuit losses be kept to low values to increase the power gain and to prevent mechanical failure as a result of heating, it is desirable that metal-to-glass seals be used that have low I^2R losses in the metal member. Among the seals suitable for power-tube construction, two such types are well known. These are the feather-edge copper-to-glass Housekeeper seal and the silver-plated-chrome iron-to-glass seal. The Housekeeper seal has good electrical conductivity but lacks the mechanical strength and rigidity of the kovar seal, and because of its construction is difficult to apply in some designs. The silver-plated-chrome iron-to-glass seals are made with rf induction heating and require carefully controlled heating conditions. These seals use a glass with a lower softening temperature than that of the kovar sealing glasses.

The kovar-to-glass seal is widely accepted and is satisfactory, except for its high electrical resistivity, which may become troublesome in some high-frequency tubes. During the work on the described tetrodes, methods have been developed for coating the kovar with a high-conductivity film and for sealing kovar matching glasses to this film. A coating of either copper, silver, gold, or chromium is electroplated to a thickness of several mils and is bonded to the kovar. Seals made to kovar coating in this manner have an electrical conductivity 10 to 20 times that of uncoated kovar at frequencies in excess of 50 Mc, and are particularly adaptable to the structures and requirements herein described.

In the design of a tube for application as a grid-modulated amplifier, a linear modulation characteristic is usually sought. While some

curvature of the modulation characteristic is acceptable, it is important that the shape and slope of the characteristic be independent of the frequency and amplitude of the modulating voltage. Large variations in the required rf driving power with modulating voltage applied will make the modulation characteristic substantially dependent on the modulating frequency, unless the input grid circuit has a bandwidth equal to, or greater than, the tube output circuit. Such variations in the required grid driving power can be reduced by reducing the feedback, the electron current to the No. 1 grids, and the electron-transit times. These reductions will also act to make the impedance presented to the modulator more nearly constant over the modulating cycle.

The duplex tetrode tube arrangement was selected by the authors as lending itself to internal neutralization and being capable of providing an effective utilization of beam-forming electrode configurations, thoria-coated filaments, and large-dissipation anodes in a tube
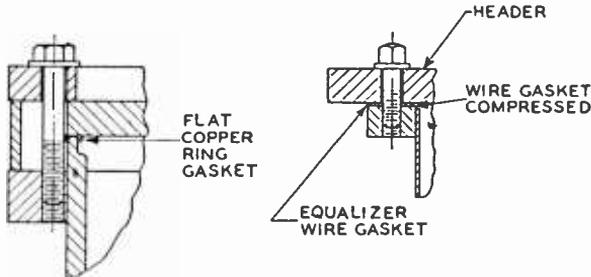
Fig. 2—Details of the copper-gasket vacuum seals.

to be operated as a wide-band grid-modulated power amplifier in the present commercial television channels. In the duplex tetrode the two tetrode units can be placed in very close proximity and can use, in effect, common screen-grid and cathode structures. The two screen grids are directly connected and by-passed to the cathode structure by low-impedance members, as in Fig. 4 and Fig. 5. The cathodes of the two tetrode units can be provided as the opposite legs of a "U"-shaped filament. These features permit the design of a high-gain wide-band grid-modulated high-frequency power tube that has exceptionally low feedback and good modulation characteristics.

Since a maximum of shielding is required between the input and output circuits, a metal envelope can be used to advantage to provide a convenient connection to external shielding in such a way as to make the envelope act as part of the shield. The metal envelope is also convenient when a demountable structure is desired. Two types of copper-gasketed compression joints are shown in Fig. 2. One of these uses a

flat copper-ring gasket compressed between a flat and a curved surface, while the other uses two round wire rings clamped between two flat surfaces, the outer ring acting only as a support ring to prevent distortion of the clamping plates. Ground and polished clamping surfaces are used with annealed OFHC copper gaskets. These demountable joints may be baked to 450°C in the processing of the tube, and are regularly used in sealed-off tubes. The tube shown in Fig. 7 is constructed with a compression joint and metal envelope, and is operated as a sealed-off tube.

THE DUPLEX TETRODE

The tube described is a developmental liquid-cooled duplex tetrode arranged for push-pull operation as a grid-modulated television power
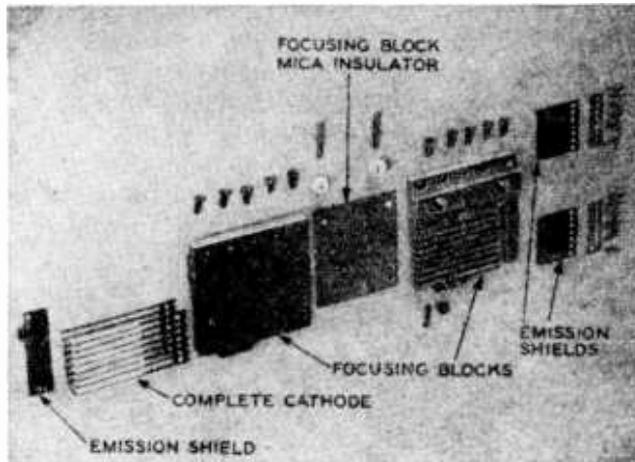
Fig. 3—The filament, filament shields, and focusing blocks.

amplifier, and is designed to give a power output, at the maximum of the synchronizing pulse, of 5 kw. at 300 Mc, and a total output bandwidth of 16 Mc. Neutralization is provided by elements attached to the No. 1 grids and included within the vacuum envelope.

An electron-beam-forming electrode configuration is used with a thoria-coated "U"-shaped filament in the arrangement shown in Fig. 1. The opposite sides of the filament act as the cathodes for the two tetrode units. The construction of the cathode and focusing-electrode assembly is shown in Fig. 3. Each of the two focusing-electrode blocks is connected to the filament, and functions as the connection to the filament-heating supply. A mica sheet is used to insulate the two focusing-electrode blocks, which are supported and cooled by water-

cooled blocks fixed to the header plate. Tantalum heat and emission shields are used, and are shown as part of the cathode assembly in Fig. 3. The electron-emitting surface of each of the two halves of the 8-strand filament is approximately 1.6 square cm in area. The filament strands are formed from tantalum sheet. A layer of thoria powder is sintered to the surface of the tantalum after the filament is formed. The filament operates at approximately 2000°K.

The No. 1 grids consist of molybdenum bars silver-soldered to water-cooled tubes which also carry the neutralizing elements. These grids are supported by the glass of the metal-to-glass vacuum seal
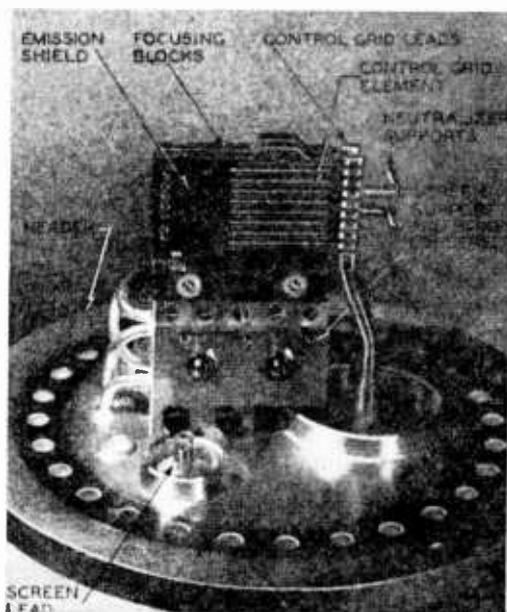
Fig. 4—The header with filament, filament shields, No. 2 mounting blocks, and No. 1 grids mounted.

mounted in the header plates as shown in Fig. 4. This figure shows the header plate with the cathode structure, and the No. 1 grids mounted and prepared for the addition of the No. 2 grids.

The two No. 2 grids are combined into a single structure which is rigidly clamped to the cathode assembly, and is insulated therefrom by mica sheets which provide a by-pass capacitance to the cathode. The No. 2 grid elements are molybdenum rods and are sliver-soldered to water-cooled copper plates. This structure with attached shields is shown in Fig. 5. The cooling of the electrode elements is obtained by thermal conduction along the length of the elements to the cooled supporting copper plates.
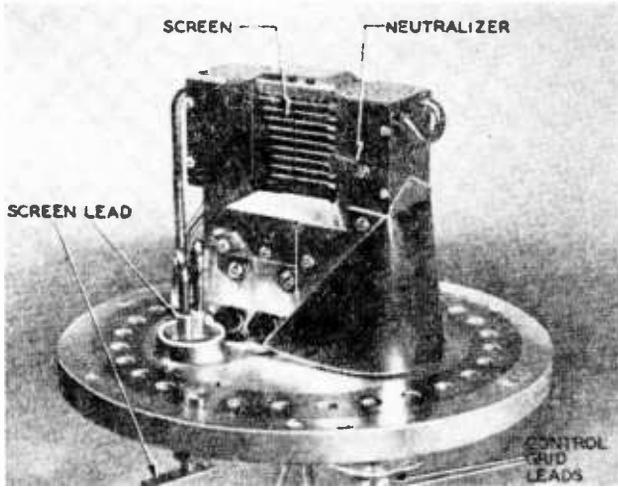
Fig. 5—A completed header assembly showing No. 2 grids and
neutralizers mounted.

The complete header assembly containing all of the electrodes excepting the anodes is shown in Fig. 5. The neutralizing tabs are shown mounted on elements fixed to the No. 1 grids and projecting through apertures in the No. 2 grid blocks. Thus use of a header for mounting the close-spaced electrodes permits accurate electrode alignment and inspection before the anode dome is mounted.

The anode dome and the details of the anode construction are shown in Fig. 6. The anode face has an area of 6.5 square cm., and is
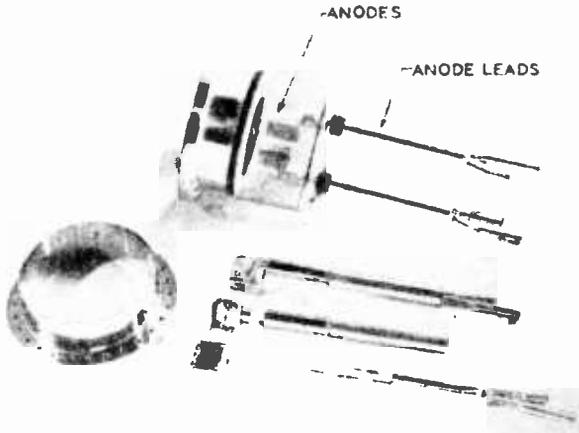
Fig. 6—The anode dome assembly, anode parts, and leads.
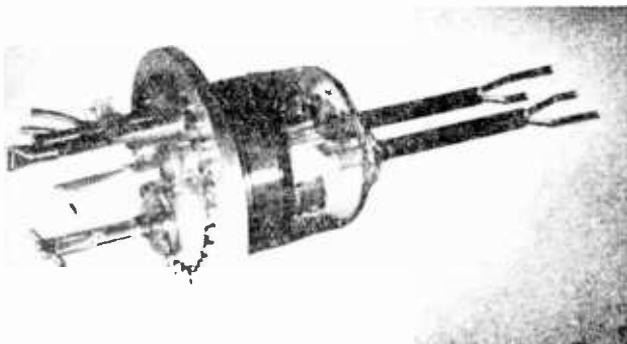
Fig. 7—The duplex tetrode.

ccoled by water circulated in the channels shown. Since the anode cooling water is carried by the anode supporting tube which functions as an element of the output tank circuit, the anode glass seals are water cooled. This cooling is adequate to permit the use of kovar at 300 Mc, if a high-conductivity coating is applied to kovar as previously described. With cooling water at a pressure of 60 pounds per square inch, these anodes have been operated without failure to a dissipation of 6 kw per anode.

TUBE DATA[1]

Direct interelectrode capacitance (each unit).

Grid No. 1 to anode	0.12 $\mu\mu$f
Input	24 $\mu\mu$f
Output	5 $\mu\mu$f
Grid No. 2 by-pass (approximate)	200 $\mu\mu$f
Grid No. 1 to anode feedback with neutralization	0.01 $\mu\mu$f

Filament—3.9 volts—105 amperes single phase

Electrode dissipation (maximum operating values):

Anodes	8 kilowatts total
No. 2 grids	400 watts total
No. 1 grids	50 watts total

Typical operating voltages:

Anode	5000–6000 volts dc
No. 2 grid	700 volts dc
No. 1 grid (cutoff bias)	—180 volts dc

Cooling water flow at 60 pounds per square inch pressure:

Anode of each unit	0.6 gallons per minute
Filament blocks in series	0.25 gallons per minute
No. 2 grid	0.3 gallons per minute
No. 1 grids in series	0.25 gallons per minute

---

[1] In the 8D21, or commercial model, some changes in these data were found necessary because of manufacturing techniques.
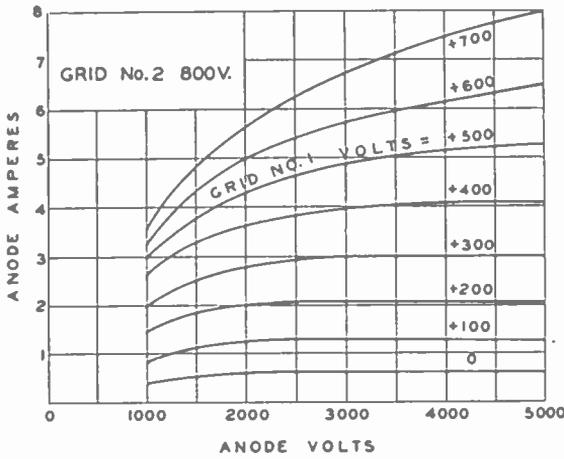
Fig. 8—Anode characteristics for each unit.

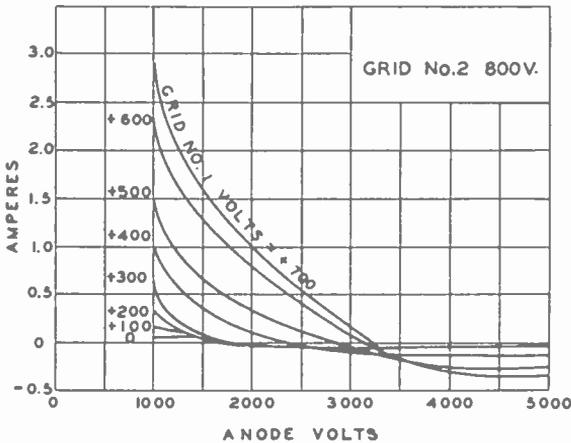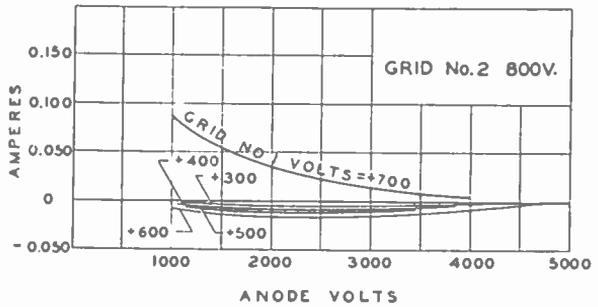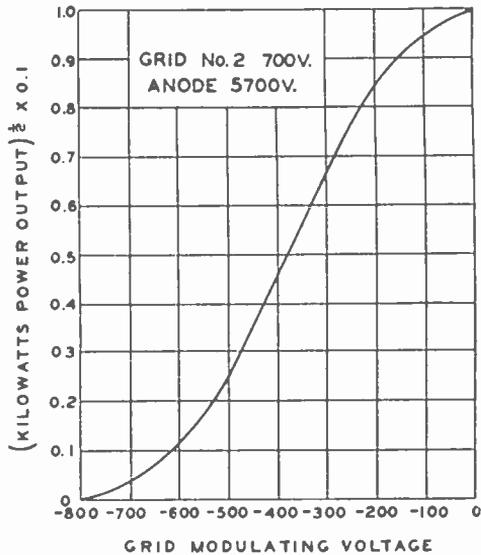Fig. 9—No. 1 grid characteristics for each unit.

Fig. 10—No. 2 grid characteristics for each unit.

Fig. 11—
Modulation characteristics.

The final step in assembly of this tube is the bolting of the anode dome to the header; this compresses the two copper-ring gaskets and provides the final vacuum closure of the envelope. The anodes, anode leads, and anode dome ring can be electroplated, cleaned, and washed before this operation, and are not subjected to any glass-working fires or other heating in making the final vacuum closure. The copper-ring gaskets are made from wire. This compression gasket seal, shown in Fig. 2, has proved to be an entirely satisfactory vacuum-tight joint which can be used both in demountable and in sealed-off tubes.

A completely processed tube is shown in the operating position in Fig. 7. This position is determined by the filament, which must hang vertically to prevent distortion and sagging. The tube is supported by the header plate rim, which acts as a flange, and which when installed is bolted to the shield wall separating the input and output circuits.

The static characteristics are given in Figs. 8, 9, and 10. A modulation characteristic, shown in Fig. 11, was obtained at 288 Mc with a total output bandwidth of 16 Mc. A maximum power output of 10 kw was obtained at an efficiency of 60 per cent. The instantaneous bias at this point was 100 volts positive with respect to the cathode. In the television application the bias at the maximum of the synchronizing pulse may be reduced to zero or made positive to reduce the amplitude of the rf grid driving voltage. This does not substantially change the required modulating voltage.

Under power output and bandwidth conditions as given above, a power gain at maximum power output of 25 to 30 is obtained. The driving power in these tests was approximately 350 watts. When the tube is used as a 5-kw amplifier with a bandwidth less than 1 Mc, power gains in excess of 100 are obtainable. These power gains for either application are not obtainable at 300 Mc, unless low-loss grid seals such as described in this paper are used.

ACKNOWLEDGMENT

THE DESIGN AND DEVELOPMENT
OF THREE NEW ULTRA-HIGH-FREQUENCY
TRANSMITTING TUBES*†

BY

CECIL E. HALLER

Research and Engineering Department, RCA Manufacturing Company, Inc.,
Harrison, N. J.

Summary

*A discussion and review are given of the service and design require-
ments of transmitting tubes intended for application in the ultra-high-
frequency spectrum. These requirements fall in two classes: (1) those
imposed by service conditions and (2) those imposed by the frequency at
which the tube is operated. The fulfilling of these requirements has led to
certain design and manufacturing problems, such as the reduction of grid
emission, choice of anode material, choice of a suitable mechanical structure,
etc.*

*A description of the novel features of construction and the operation
of three new ultra-high-frequency transmitting tubes is also given. Two
of these tubes are the RCA-815 and RCA-829 which are push-pull beam
tetrodes while the third tube, the RCA-826, is a triode. Some precautions
necessary for obtaining satisfactory operation with these tubes are given.*

(7 pages, 8 figures, 1 table)

DEVELOPMENT OF PULSE TRIODES AND CIRCUIT
TO GIVE
ONE MEGAWATT AT 600 MEGACYCLES*†

BY

R. R. LAW, D. G. BURNSIDE, R. P. STONE, W. B. WHALLEY

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary

*The work here described was done for the Army and Navy to develop
high-power air-cooled pulse triode transmitting tubes and circuits. The
A-2231 tube developed during the course of this project readily provides
500 kilowatts peak power at 600 megacycles with a duty cycle of 0.1 per
cent. In the push-pull oscillator described herein, two of the tubes easily
give a peak power of one megawatt. At this power level the operation is
stable and the frequency may be varied throughout the 560-640 megacycle*

tuning range with no tendency toward sparkover. By ganging together the cathode and anode tuning controls, it has been possible to accomplish tuning with a single electric motor which may be remote controlled. These same principles of tuning might be employed to cover a much wider range.
(12 pages; 7 figures)

POWER MEASUREMENTS OF CLASS B AUDIO AMPLIFIER TUBES*†

BY

DAVID P. HEACOCK

Tube Department, RCA Victor Division,
Harrison, N. J.

Summary

An accurate method of determining the performance of two triodes, or two triode units of the same tube, operating as a class B push-pull amplifier by a simple measurement of the plate current of a single triode unit is described. This method is particularly adaptable as a production test on electron tubes designed primarily for operation as class B amplifiers. The errors attendant upon the use of conventional measuring technique for class A power output when applied to the non-sinusoidal output of a single triode unit operated class B are discussed.
(11 pages; 8 figures)

COAXIAL TANTALUM CYLINDER CATHODE FOR CONTINUOUS-WAVE MAGNETRONS*†

BY

R. L. JEPSEN

Tube Department, RCA Victor Division,
Lancaster, Pa.

Summary

Major factors affecting the design of cathodes for continuous-wave magnetrons are considered. A particular type of structure, the coaxial tantalum cylinder cathode, is analyzed in some detail and its merits and shortcomings discussed. An approximate method is developed for computing optimum geometry and maximum life. Application to a particular magnetron is described and other applications suggested.
(11 pages; 5 figures)

STABILIZED MAGNETRON FOR BEACON SERVICE*†

Part I

DEVELOPMENT OF UNSTABILIZED TUBE

BY

J. S. DONAL, JR[a]., C. L. CUCCIA[a], AND B. B. BROWN[b]

Summary

The frequency of a magnetron varies rather widely with temperature, current, and load impedance. This paper describes one of the first successful attempts to reduce this frequency variation by means of a method of stabilization. Part I treats particularly the mechanical and electrical design of the unstabilized tube.

The mechanical design of the tube is unconventional in that all of the parts are supported on a header to which the envelope is welded. The necessary inserts in the magnetic circuit are at cathode potential, serving both as cathode supports and as improved cathode end-shields. The tube is designed for a pulsed input power of 2500 watts. The unstabilized peak power output is approximately 1 kilowatt at an anode potential of 2500 volts. The developmental cavity stabilizer served to demonstrate the principles of stabilization and influenced the design of the tube. For production purposes, the early stabilizer was superseded by a design differing in mechanical details.

(9 pages, 8 figures)

Part II

ENGINEERING OF TUBE AND STABILIZER

BY

C. P. VOGEL[b] AND W. J. DODDS[b]

Summary

To satisfy the frequency stability requirements of a beacon system, a 9310-megacycle magnetron was fitted with a frequency stabilization device that reduced the inherent frequency changes of a magnetron by a factor of approximately 10. The device included a turntable resonant cavity for the storage of additional electromagnetic energy, and a method of coupling the cavity of the tube.

The cavity is made of Invar steel to reduce the changes of dimension of the cavity with temperature. It is furthermore temperature-compensated by use of higher-expansion steel for the spindle which supports the cavity tuning plunger. The tube frequency is adjusted within the specified tuning range by changing the cavity frequency. This is done by means of tuning

mechanism comprising a movable plunger actuated by rotation of an accurately ground nut which rotates in an accurately ground bearing where it is held in place by spring loading.

The cavity and stabilizer system provides for hermetic sealing so that the system may be filled with dry nitrogen at atmospheric pressure. The wave-guide system which forms the circuit that couples the tube and stabilizer to the load contains adjustable screw tuners to permit compensating for the variations of internal impedance from tube to tube and to compensate for differences in line length due to the location of the tube in the system. Stabilization procedure consists of the proper adjustment of these screw tuners.

A basic, simplified mathematical theory of stabilization is contained in the Appendix.

(11 pages, 3 figures, 1 table, 1 appendix)

A FREQUENCY-MODULATED MAGNETRON
FOR SUPER-HIGH FREQUENCIES*†

By

G. R. Kilgore, Carl I. Shulman and J. Kurshan

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary

This paper is based on a wartime requirement for a 25-watt, 4000-megacycle continuous-wave oscillator capable of electronic frequency modulation with a deviation of at least 2.5 megacycles. A satisfactory solution was found in the addition of frequency control to a continuous-wave magnetron by the introduction of electron beams into the magnetron cavities in a manner described by Smith and Shulman. This method is referred to as "spiral-beam" control.

(8 pages, 10 figures)

A ONE-KILOWATT FREQUENCY-MODULATED
MAGNETRON FOR 900 MEGACYCLES*†

By

J. S. Donal, Jr., R. R. Bush, C. L. Cuccia, and H. R. Hegbar

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary

The method of Smith and Shulman has been used for the frequency modulation of a 1-kilowatt continuous-wave magnetron. This tube is of the "vane" type, having twelve resonant cavities, and it is mechanically tunable over a range from about 720 to 900 megacycles by a cylindrical element which varies the interstrap capacitance. At the applied magnetic field required for frequency modulation without change in amplitude, 1-kilowatt output at 900 megacycles is obtained with an anode voltage of 2.5 kilovolts and an efficiency of about 55 per cent; the efficiency rises with decreasing frequency or with increasing magnetic field.

At 900 megacycles, electron beams in nine of the magnetron resonant cavities give a frequency deviation of 3.5 megacycles (a total frequency swing of 7 megacycles) at an output of 1 kilowatt, rising to 4 megacycles at an output of 750 watts. The frequency deviation is reduced when the tube is tuned to lower frequencies. The modulator power required would be very low, since the grid-cathode capacitance of the frequency-modulation guns is small and the grids draw no current.

It would be practicable to increase the frequency deviation of this tube by about 15 per cent through an increase in beam current, and by an additional 20 per cent through the use of eleven beams. A change in the type of beam cathode would effect an even greater deviation.

(6 pages, 6 figures, 1 table)

THE OPERATION OF FREQUENCY CONVERTERS AND MIXERS FOR SUPERHETERODYNE RECEPTION*†

BY

E. W. HEROLD

RCA Manufacturing Company, Inc.,
Harrison, N. J.

Summary—This paper presents a general picture of superheterodyne frequency conversion followed by a detailed discussion of the behavior of tubes used with different types of oscillator injection. The general picture shows that the different methods of frequency conversion are basically similar for small signals. A strong local-oscillator voltage (which may be a pure sine wave) causes a periodic variation (which is usually nonsinusoidal) of the signal-electrode transconductance. The coefficient of each Fourier component of the transconductance-versus-time relationship is just twice the conversion transconductance at the corresponding harmonic of the local-oscillator frequency. For most tubes the conversion transconductance g_c at the oscillator fundamental is approximately 28 per cent of the maximum transconductance. At the second harmonic, g_c is about 14 per cent, and at the third harmonic it is about 9 per cent of the maximum transconductance. Fluctuation noise and input resistance at high frequencies of the different methods of conversion may be found from the time average over the oscillator cycle.

Using these general concepts, we discuss the detailed behavior of three conversion methods. In the first method, signal and local-oscillator voltages are impressed on the same electrodes. This method gives best signal-to-noise ratio, but has the disadvantage of bad interaction between signal and local-oscillator circuits. In the second method, the local-oscillator voltage is impressed on an electrode which precedes the signal electrode along the direction of electron flow. In this case, interaction of signal and oscillator circuits is somewhat reduced but is still bad at the higher frequencies because of space-charge coupling. The third method is that in which the local-oscillator electrode follows the signal electrode along the direction of electron flow. Most of the disadvantages of the third method may be overcome by special tube constructions, some of which are described.

I. INTRODUCTION

THE BETTER modern radio receivers are almost universally designed to use the superheterodyne circuit. In such a circuit, the received signal frequency is heterodyned with the frequency of a local oscillator to produce a difference frequency known as the intermediate frequency. The resultant signal is amplified by a selec-

212

tive, fixed-tuned amplifier before detection. Since the heterodyne action is usually accomplished by means of a suitable vacuum tube, it is the purpose of this paper to discuss the chief similarities and differences among the tubes which might be used, as well as to explain their behavior.

The combination of signal and local-oscillator frequencies to produce an intermediate frequency is a process of modulation in which one of the applied frequencies causes the amplitude of the other to vary. Although this process was originally called heterodyne detection and, later, first detection, it is now called frequency conversion. The portion of the radio receiver which produces conversion may, therefore, be identified as the converter. If conversion is accomplished in a single vacuum tube which combines the functions of oscillator and modulator, this tube may logically be termed a converter tube. When separate tubes are used for the oscillator and the modulator portions of the converter, respectively, the tube for the latter purpose is conveniently called a modulator or mixer tube. This terminology will be used in this paper.

Although in some of the earliest superheterodynes, frequency conversion was accomplished by a triode oscillator and a triode modulator,[1] other circuits used a single triode which served as both modulator and oscillator.[2] A triode used in the latter way could, therefore, be called a converter tube. The introduction of two-grid tubes (i.e., tetrodes) permitted a wide variety of modulator and converter arrangements which frequently gave superior performance to that possible with triodes.[3-7]

When indirectly heated cathodes became more common, conversion circuits in which the oscillator voltage was injected in the cathode circuit were used. These circuits reduced considerably the interaction between oscillator and signal circuits which would otherwise be

[1] E. H. Armstrong, "A new system of short-wave amplification," *Proc. I.R.E.*, Vol. 9, pp. 3-27; February, 1921.

[2] German Patent No. 324,514, 1918.

[3] J. Scott-Taggart, German Patent No. 383,449, 1919.

[4] J. deMare, R. Barthelemy, H. deBellescize, and L. Levy, "Use of double-grid valves in frequency-changing circuits," *L'Onde Elec.*, Vol. 5, pp. 150-180; 1926.

[5] "A four-electrode valve supersonic circuit," *Exp. Wireless*, Vol. 3, p. 650; October, 1926.

[6] R. Barthelemy, "Valve frequency changers," *Gen. Elec. Rev.*, Vol. 19, pp. 663-670; 1926.

[7] See also: M. Gausner, French Patent No. 639,028; G. Thebault, French Patent No. 655,738; and H. J. J. M. deRegnauld de Bellescize, United States Patent No. 1,872,634.

present.[8] When tetrodes and pentodes became available, the use of the triode was dropped except as the local oscillator. It was not long, however, before the desirability of more complete separation of oscillator and signal circuits became evident. Multigrid converter tubes were, therefore, devised to permit this separation in a satisfactory manner, at least for the frequencies then in common use.[9-11] In some of these it was also possible to control the conversion gain by an automatic-volume-control voltage, a decided advantage. The most satisfactory of the earlier multigrid tubes was known as the pentagrid converter, a type still widely used. A similar tube having an additional suppressor grid is used in Europe and is known as the octode.

When it became desirable to add high-frequency bands to super-heterodyne receivers which also had to cover the low broadcast frequencies, the converter problem became more difficult. The highest practicable intermediate frequency appeared to be about 450 to 460 kilocycles, a value which was only about 2 per cent of the highest frequency to be received. Its use meant that the oscillator frequency was separated from the signal frequency by only 2 per cent and the signal circuit, therefore, offered appreciable impedance at the oscillator frequency. A phenomenon known as "space-charge coupling," found in the pentagrid converter, indicated that signal and oscillator circuits were not separated as completely as would be desirable.[15] In addition, the permissible frequency variations of the oscillator had to be held to less than the intermediate-frequency bandwidth, namely, 5 to 10 kilocycles; at the highest frequency to be received, the oscillator frequency was required therefore to remain stable within 0.05 per cent. In the pentagrid converter, the most serious change in oscillator frequency occurred when the automatic-volume-control voltage was changed, and was sometimes as much as 50 kilocycles. Economic con-

[8] V. E. Whitman, United States Patent No. 1,893,813; H. A. Wheeler, United States Patent No. 1,931,338.

[9] F. B. Llewellyn, United States Patent No. 1,896,780.

[10] H. A. Wheeler, "The hexode vacuum tube," *Radio Eng.*, Vol. 13, pp. 12-14; April, 1933.

[11] W. Hasenberg, "The hexode," *Funk. Tech. Monatshefte*, pp. 165-172; May, 1933.

[12] Application Note No. 3, RCA Radiotron Co., Inc.

[13] E. Y. Robinson, British Patent No. 408,256.

[14] J. C. Smith, Discussion on H. A. Wheeler paper, "Image suppression and oscillator-modulators in superheterodyne receivers," *Proc. I.R.E.*, Vol. 23, pp. 576-577; June, 1935.

[15] W. A. Harris, "The application of superheterodyne frequency conversion systems to multirange receivers," *Proc. I.R.E.*, Vol. 23, pp. 279-294; April, 1935.

siderations have led to the use of at least a three-to-one frequency coverage for each band in the receiver. With capacitance tuning, the circuit impedance is very low at the low-frequency end of the high-frequency band so that failure to oscillate was occasionally observed in the pentagrid converter.

In Europe, where converter problems were similar, a tube known as the triode-hexode[16] was developed to overcome some of the disadvantages of the pentagrid converter. In the pentagrid tube, the oscillator voltage is generated by, and therefore applied to, the electrodes of the assembly closest to the cathode (i.e., the *inner* electrodes). In the European form of triode-hexode, the oscillator voltage is generated by a separate small triode section mounted on a cathode common to a hexode-modulator section. The triode grid is connected internally to the third grid of the hexode section. In this way, by the application of the oscillator voltage to an *outer* grid and the signal to the inner grid of the modulator, space-charge coupling was greatly reduced and automatic-volume-control voltage could be applied to the modulator section of the tube without seriously changing the oscillator frequency. In some European types, a suppressor grid has been added so that such tubes should be called triode-heptodes.

The first American commercial development to provide improved performance over that of the pentagrid converter also utilized oscillator voltage injection on an outer grid but required a separate tube for oscillator.[17] This development, therefore, resulted in a modulator or mixer tube rather than a converter. There were many advantages accompanying the use of a separate oscillator tube so that such a solution of the problem appeared to be reasonably satisfactory.

The demand arose shortly, however, for a one-tube converter system with better performance than the original pentagrid type for use in the standard all-wave receiver. A tube, the 6K8, in which one side of a rectangular cathode was used for the oscillator and the other side was used for the mixer section, was developed and made available.[18] This tube used inner-grid oscillator injection, as with the pentagrid converter, but had greatly improved oscillator stability. Another solution, also introduced in the United States, was a triode-heptode which

---

[16] E. E. Shelton, "A new frequency changer," *Wireless World*, Vol. 35, pp. 283-284; October 5, 1934.

[17] C. F. Nesslage, E. W. Herold, and W. A. Harris, "A new tube for use in superheterodyne frequency conversion systems," *Proc. I.R.E.*, Vol. 24, pp. 207-218; February, 1936.

[18] E. W. Herold, W. A. Harris, and T. J. Henry, "A new converter tube for all-wave receivers," *RCA Review*, Vol. 3, pp. 67-77; July, 1938.

is an adaptation of the European triode-hexode. This type used outer-grid injection of the oscillator voltage generated by a small auxiliary triode oscillator section. A recent converter (the SA7 type) for broadcast use is designed to operate with oscillator voltage on both cathode and first-grid electrodes.[19] This tube, in addition to having excellent performance, requires one less connecting terminal than previous converter tubes.

This paper will present an integrated picture of the operation of converter and modulator tubes. It will be shown that the general principles of modulating or mixing by placing the signal on one grid and the oscillator voltage on another, or by placing both voltages on the same grid, are the same for all types of tubes. The differences in performance among the various types particularly at high frequencies are due to a number of important secondary effects. In this paper, some of the effects such as signal-grid current at high frequencies, input impedance, space-charge coupling, feedback through interelectrode capacitances, and oscillator-frequency shift will be discussed.

II. General Analysis of Operation Common to All Types

A. *Conversion Transconductance of Modulator or Mixer Tubes*

The basic characteristic of the converter stage is its conversion transconductance, i.e., the quotient of the intermediate-frequency output current to the signal input voltage. The conversion transconductance is easily obtained by considering the modulation of the local-oscillator frequency by the signal in the tube and, as shown in another paper,[17] is determined by the transconductance of the signal electrode to the output electrode. The general analysis of a modulator, or mixer tube, is applicable to all mixers no matter how or on what electrodes the oscillator and signal voltages are introduced.

Under the assumption that the signal voltage is very small and the local oscillator voltage large, the signal-electrode transconductance may be considered as a function of the oscillator voltage only. The signal-electrode-to-plate transconductance g_m may, therefore, be considered as periodically varying at the oscillator frequency. Such a periodic variation may be written as a Fourier series

$$g_m = a_0 + a_1 \cos \omega_0 t + a_2 \cos 2\omega_0 t + \cdots$$

where ω_0 is the angular frequency of the local oscillator. Use of the cosine series implies that the transconductance is a single-valued function of the oscillator electrode voltage which varies as $\cos \omega_0 t$. When a small signal, $e_x \sin \omega_x t$, is applied to the tube, the resulting alternating plate current to the first order in e_x may be written

$$i_p = g_m e_s \sin \omega_x t$$

$$= a_0 e_s \sin \omega_x t + e_x \sum_{n=1}^{\infty} a_n \sin \omega_x t \cos n\omega_0 t$$

$$= a_0 e_x \sin \omega_x t + \tfrac{1}{2} e_x \sum_{n=1}^{\infty} a_n \sin (\omega_s + n\omega_0) t$$

$$+ \tfrac{1}{2} e_x \sum_{n=1}^{\infty} a_n \sin (\omega_x - n\omega_0) t.$$

If a circuit tuned to the frequency $(\omega_x - n\omega_0)$ is inserted in the plate, the modulator tube converts the incoming signal frequency ω_s to a useful output at an angular frequency $(\omega_x - n\omega_0)$ which is called the intermediate frequency. Since n is an integer, it is evident that the intermediate frequency, in general, may be chosen to be the difference between the signal frequency and any integral multiple of the local-oscillator frequency; this is true even though a pure sine-wave local oscillation is applied to the tube. The harmonics of the local-oscillator frequency need only be present in the time variation of the signal-electrode transconductance. The ordinary conversion transconductance is simply a special case when $n = 1$. The conversion transconductance at the nth harmonic of the local oscillator is given by

$$g_{cn} = \frac{i_{\omega x - n\omega_0}}{e_s} = \frac{a_n}{2}.$$

Substituting the value of the Fourier coefficient a_n it is found that

$$g_{cn} = \frac{1}{2\pi} \int_0^{2\pi} g_m \cos n\omega_0 t\, d(\omega_0 t).$$

When n is set equal to unity, this expression becomes identical with the one previously derived.[17]

Thus, the conversion transconductance is obtained by a simple Fourier analysis of the signal-to-electrode-to-output-electrode trans-

conductance as a function of time. Such an analysis is readily made from the tube characteristics directly by examination of the curve of signal-electrode transconductance versus oscillator-electrode voltage. The calculation of the conversion transconductance at the nth harmonic of the oscillator is made from this curve by assuming an applied oscillator voltage and making a Fourier analysis of the resulting curve of transconductance versus time for its nth harmonic component. The analysis is exactly similar to the one made of power output tubes, except that, in the latter case, the plate-*current*-versus-control-electrode-voltage curve is used. Figure 1 shows a curve of signal-grid transconductance versus oscillator-electrode voltage for a typical modulator or mixer tube. In the usual case, the oscillator voltage is applied
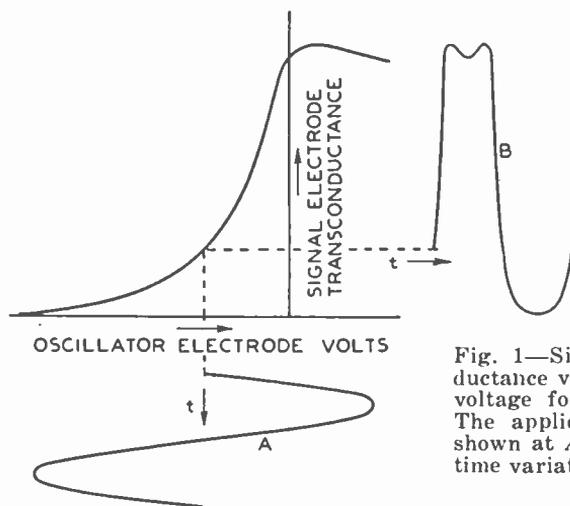

Fig. 1—Signal-electrode transconductance versus oscillator-electrode voltage for a typical mixer tube. The applied oscillator voltage as shown at A and B is the resulting time variation of transconductance.

from a tuned circuit and so is closely sinusoidal in shape as at A in the figure. The resulting curve of transconductance versus time is shown at B. Any of the usual Fourier analysis methods may be used to determine the desired component of curve B. Half of this value is the conversion transconductance at the harmonic considered. Convenient formulas of sufficient accuracy for many purposes follow. Referring to Figure 2a, a sine-wave oscillator voltage is assumed and a seven-point analysis is made (i.e., 30-degree intervals). The conversion transconductances g_{cn} are

$$g_{c1} = \tfrac{1}{12} \left[(g_7 - g_1) + (g_5 - g_3) + 1.73(g_6 - g_2) \right]$$

$$g_{c2} = \tfrac{1}{12} \left[2g_4 + \tfrac{3}{4}(g_3 + g_5 - g_6 - g_2) - (g_7 + g_1) \right]$$

$$g_{c3} = \tfrac{1}{12} \left[(g_7 - g_1) - 2(g_5 - g_3) \right].$$

The values g_1, g_2, etc., are chosen from the transconductance characteristic as indicated in Figure 2a. The values computed from the above formulas are, of course, most accurate for g_{c1} and of less accuracy for g_{c2} while a value computed from the formula for g_{c3} is a very rough approximation.
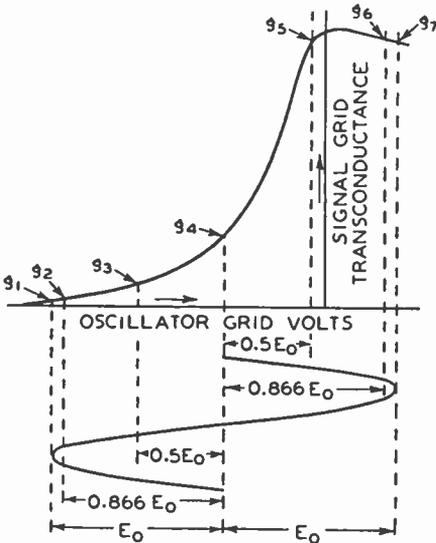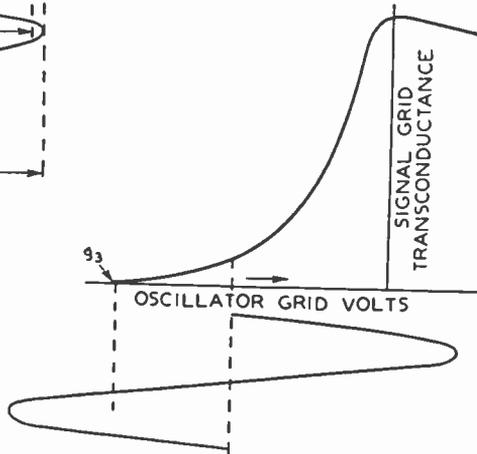

Fig. 2—a. Points used for 30-degree analysis of conversion transconductance.

b. Oscillator amplitude and bias adjusted for high conversion transconductance at oscillator fundamental, i.e., $g_1 = g_2 = g_3 = 0$.

Simple inspection of the formula for g_{c1}, the conversion transconductance used for conversion at the fundamental, is somewhat instructive. It is evident that highest conversion transconductance, barring negative values, as given by this formula, occurs when g_1, g_2, and g_3 are all equal to zero, and g_5, g_6, and g_7 are high. These requirements mean that sufficient oscillator voltage should be applied at the proper point to cut off the transconductance over slightly less than half the cycle as pictured in Figure 2b. For small oscillator voltages optimum

operation requires the differences $(g_7 - g_1)$, $(g_5 - g_3)$, and $(g_6 - g_2)$ to be as large as possible; this is equivalent to operation at the point of maximum slope. It should be noted that the minimum peak oscillator voltage required for good operation is approximately equal to one half the difference between the oscillator-electrode voltage needed for maximum signal-grid transconductance and that needed to cut off this transconductance. Thus, inspection of the curve of transconductance versus oscillator-electrode voltage gives both a measure of the fundamental conversion transconductance which will be obtained and the amount of oscillator excitation required. Conversion at a harmonic, in general, requires considerably greater oscillator exicitation for maximum conversion transconductance.

In practical cases using grid-controlled tubes of the usual kind, the maximum fundamental conversion transconductance which a given tube will give can quickly be determined within 10 per cent or so by simply taking 28 per cent of the maximum signal-grid-to-plate transconductance which can be attained. For conversion at second harmonic, optimum oscillator excitation gives a conversion transconductance of half this value, while for third-harmonic conversion the value is divided by three.

Although the same characteristic of all modulator or mixer tubes is used to determine the conversion transconductance, the shape of this characteristic varies between different types of mixers. This variation will be more clearly brought out in the later sections of the paper.

B.   *Conversion Transconductance of Converter Tubes*

In converter tubes with oscillator sections of the usual kind, the oscillator voltage is usually present on more than one electrode. Furthermore, the phase of the oscillator-control-grid voltage is opposite to that of the oscillator-anode alternating voltage, so that the two would be expected partially to demodulate each other. The transconductance curve which should be used in this case is the one in which the oscillator electrode voltages are simultaneously varied in opposite directions.

Fortunately, with most of the commonly used converter tubes such. as the pentagrid, octode, triode-hexode, etc., the effect of small variations of oscillator-anode voltage on the electrode currents is so small that usually it may be neglected. Thus, the conversion transconductance of these converter tubes may be found exactly as if the tube were a modulator or mixer, only.

With the circuit of Figure 3,[19][21] a Hartley oscillator arrangement is used and oscillator-frequency voltage is present on the cathode. The effect of such a voltage is also to demodulate the electron stream through the action of the alternating cathode potential on the screen-to-cathode and signal-grid-to-cathode voltages. When a relatively high-transconductance signal grid is present, as in the figure, this demodulation is considerably greater than in the normal cathode-at-ground circuit. In order to determine the conversion transconductance of a tube to be used in this circuit, a signal-grid transconductance curve is needed. Such a curve, however, must be taken with cathode and oscillator-grid potential varied simultaneously and in their correct ratio as determined by the ratio of cathode turns to total turns of the coil which is to be used. However, because the conversion transconductance is approximately proportional to the peak value of signal-grid transconductance, it is often sufficiently accurate to disregard the

Fig. 3—Converter circuit with oscillator voltage on both grid No. 1 and cathode.
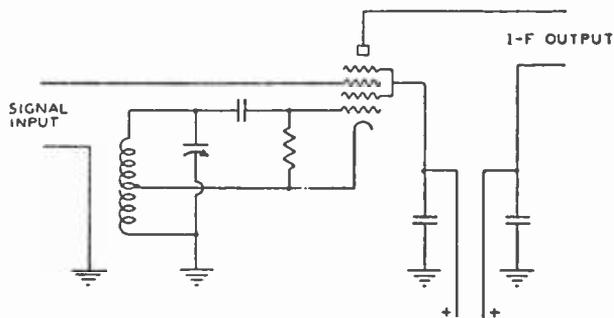
alternating-current variation of cathode potential and simply shift the signal-grid bias in the negative direction by the peak value of the alternating cathode voltage. If the resulting signal-grid-transconductance versus oscillator-grid-voltage curve is used for an analysis of conversion transconductance, the data obtained will not be far different from the actual values obtained in the circuit of Figure 3 where normal (unshifted) signal-grid bias values are used.

C. *Fluctuation Noise*

The fluctuation noise of a converter stage is frequently of considerable importance in determining the over-all noise. The magnitude of the fluctuation noise in the output of a converter or mixer tube may be found either by direct measurement using a known substitution noise source such as a saturated diode or by making use of the noise

---

[20] P. W. Klipsch, "Suppression of interlocking in first detector circuits," *Proc. I.R.E.*, Vol. 22, pp. 699-708; June, 1934.

[21] *Radio World*, p. 13; December 24, 1932.

of the same tube used as an amplifier and finding the average mean-squared noise over an oscillator cycle.[22,23] Since these methods give values which are substantially in accord and, since the noise of many of the usual tube types under amplifier conditions is readily derived from theory,[24] the latter procedure is convenient. Thus, if $\overline{i_{pn}^2}$ is the mean-squared noise current in the output of the converter or mixer tube considered as an amplifier (i.e., steady direct voltages applied) the mean-squared intermediate-frequency noise is

$$\overline{i_{i-f}^2} = \frac{1}{2\pi} \int_0^{2\pi} \overline{i_{pn}^2} \, d(\omega t)$$

or the average of $\overline{i_{pn}^2}$ over an oscillator cycle. The values of $\overline{i_{pn}^2}$ obtained from theory require a knowledge of the currents and transconductance of the tube and are usually proportional to these quantities. Thus, the converter-stage output noise, which is the average of $\overline{i_{pn}^2}$ over the oscillator cycle is usually proportional to the average electrode currents and average transconductance when the oscillator is applied. Specific examples will be given in following sections of this paper treating typical modes of converter operation.

Tube noise is conveniently treated by use of an equivalent grid-noise-resistance concept whereby the tube noise is referred to the signal grid. The equivalent noise resistance of a converter or mixer tube is

$$R_{eq} = \frac{\overline{i_{i-f}^2}}{(4kT_R\Delta f)\,g_{cn}^2}$$

where $k = 1.37 \times 10^{-23}$, T_R is room temperature in degrees Kelvin, and Δf is the effective over-all bandwidth for noise purposes. Since Δf is invariably associated with $\overline{i_{i-f}^2}$, the bandwidth cancels in the determination of R_{eq} which is one of the advantages of the equivalent-

[22] E. Lukacs, F. Preisach, and Z. Szepcsi, "Noise in frequency changer valves," (Letter to Editor), *Wireless Eng.*, Vol. 15, pp. 611-612; November, 1938.

[23] E. W. Herold, "Superheterodyne converter system considerations in television receivers," *RCA Review*, Vol. 4, pp. 324-337; January, 1940.

[24] B. J. Thompson, D. O. North, and W. A. Harris, "Fluctuations in space-charge-limited currents at moderately high frequencies," *RCA Review*, Vol. 4, pp. 269-285; January, 1940; Vol. 4, pp. 441-472; April, 1940; Vol. 5, pp. 106-124; July, 1940; Vol. 5, pp. 244-260; October, 1940; Vol. 5, pp. 371-388; January, 1941; Vol. 5, pp. 505-524; April, 1941; Vol. 6, pp. 114-124; July, 1941.

resistance concept. For $T_R = 20$ degrees centigrade,

$$R_{eq} = 0.625 \times 10^{20} \frac{1}{g_{cn}^2} \frac{\overline{i_{i-f}^2}}{\Delta f}.$$

A summary of values of R_{eq} for common types of converter will be found in a preceding paper.[23]

The equivalent noise resistance R_{eq} alone does not tell the entire story as regards signal-to-noise ratio, particularly at high frequencies. For example, if the converter stage is the first stage of a receiver, and bandwidth is not a consideration, the signal energy which must be supplied by the antenna to drive it will be inversely proportional to the converter-stage input resistance. On the other hand, the noise energy of the converter or mixer tube is proportional to its equivalent noise resistance. The signal-to-noise ratio therefore, will vary with the ratio of input resistance to equivalent noise resistance, and this quantity should be as high as possible. When bandwidth is important, the input resistance should be replaced by the reciprocal of the input capacitance if it is desired to compare various converter systems for signal-to-noise ratio.

III. THE OSCILLATOR SECTION OF CONVERTER TUBES

The oscillator section of converters is often required to maintain oscillation over frequency ranges greater than three to one for circuits using capacitance tuning. Although this requirement is easily net at the lower broadcast frequencies, the effect of lower circuit impedances, transit-time phenomena in the tube, and high lead reactances combine to make the short-wave band a difficult oscillator problem. Ability to oscillate has, in the past, been measured by the oscillator transconductance at normal oscillator-anode voltage and zero bias on the oscillator grid. Recent data have shown that, in the case of pentagrid and some octode converters, an additional factor which must be considered is the phase shift of oscillator transconductance (i.e., transadmittance) due to transit-time effects.[25,26]

The ability of a converter to operate satisfactorily at high frequencies depends largely on the undesirable oscillator frequency variations produced when electrode voltages are altered. The frequency changes

---

[25] M. J. O. Strutt[26] has published data on this phase shift in octodes. It was measured to be as high as 60 degrees at 33 megacycles.

[26] M. J. O. Strutt, "Electron transit-time effects in multigrid valves," *Wireless Eng.*, Vol. 15, pp. 315-321; June, 1938.

are mainly caused by the dependence on electrode voltages of oscillator-electrode capacitances, oscillator transconductance, and transit-time effects. There are many other causes of somewhat lesser importance. Because of the complex nature of the problem no satisfactory quantitative analysis is possible. In the case of the pentagrid and the earlier forms of octode converters there are indications that the larger part of the observed frequency shift is due to a transit-time effect. It is found that the phase of the oscillator transadmittance and, therefore, the magnitude of the susceptive part of this transadmittance varies markedly with screen and signal-grid-bias voltages. Since the susceptive part of the transadmittance contributes to the total susceptance, the oscillation frequency is directly affected by any changes.

IV. THE DETAILED OPERATION OF THE MODULATOR OR MIXER SECTION OF THE CONVERTER STAGE

This section will be devoted to a consideration of the modulator or mixer portion of the converter stage. This portion may be either a separate mixer tube or the modulator portion of a converter tube. Since with most of the widely used converter tubes in the more conventional circuits the alternating oscillator-anode voltage has a negligible effect on the operation of the modulator portion, only the effect of oscillator control grid need be considered. Thus the analysis of the operation of most converter tubes is substantially the same as the analysis of the same tubes used as a mixer or modulator only, just as in the treatment of conversion transconductance.

There are three methods of operation of mixer or modulator tubes. The oscillator voltage may be put on the same grid as the signal voltage, it may be put on the inner grid (the signal applied to an outer grid), or it may be impressed on an outer grid (with the signal on the inner grid). Each of these modes of operation has characteristics which depend on the mode rather than on the tube used in it. Tubes which may be used in any one mode differ from one another mainly in the degree in which they affect these characteristics. The treatment to follow, therefore, will not necessarily deal with specific tube types: instead, the phenomenon encountered will be illustrated by the use of data taken on one or more typical tubes for each of the modes of operation.

A. *Tubes with Oscillator and Signal Voltages Applied to Same Grid*

Typical tubes used for this type of operation are triodes and pentodes. The oscillator voltage may be introduced in series with the signal

voltage, coupled to the signal input circuit inductively, capacitively, and/or conductively, or it may be coupled into the cathode circuit. In all but the last case, by operating below the grid-current point, the oscillator circuit is not loaded directly by the mixer tube. When cathode injection is used, however, an effective load equal to the mean cathode conductance (slightly greater than the mean transconductance) is imposed on the oscillator circuit. The cathode injection circuit has the advantage that oscillator-frequency voltage between the signal input circuit and ground is minimized, thus reducing radiation when the converter stage is also the first stage of the receiver.

A typical transconductance versus bias curve for a variable-μ radio-frequency pentode is shown in Figure 4. The use of the Fourier analysis for conversion transconductance at oscillator fundamental indicates that a value of approximately a quarter of the peak transconductance can be attained. Because of the tailing off of the lower end of the curve, highest conversion transconductance requires a large oscillator swing. Very nearly the maximum value is obtained, however, at an operating bias shown by the dotted line, with an oscillator peak amplitude approximately equal to the bias. With lower oscillator amplitudes, and the same fixed bias, the fundamental conversion transconductance drops in approximate proportion to the oscillator amplitude.
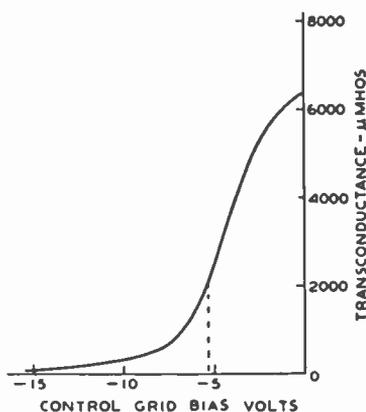

Fig. 4—Transconductance characteristic of a typical variable-μ, radio-frequency pentode.

Strictly speaking, when the cathode injection type of operation is used the effect of the oscillator voltage which is impressed between screen and cathode, and plate and cathode should be considered. Practically, however, there is little difference over the simpler circuit in which the oscillator voltage is impressed on the signal grid only. It is for this reason that the cathode-injection circuit is placed in the same category as those in which the oscillator voltage is actually impressed on the same electrode as the signal.

In a practical circuit the effective oscillator voltage is, of course, the oscillator voltage actually existent between grid and cathode of the tube. When the oscillator voltage is impressed in series with the signal circuit or on the cathode, this effective voltage is different from

the applied oscillator voltage by the drop across the signal circuit. In the usual case, with the oscillator frequency higher than the signal frequency, the signal circuit appears capacitive at oscillator frequency. This capacitance and the grid-to-cathode capacitance, being in series, form a capacitance divider and reduce the effective oscillator voltage. The reduction would not be a serious matter if it remained a constant quantity; but in receivers which must be tuned over an appreciable frequency range this is not the case. The result is a variation in conversion gain over the band. A number of neutralizing circuits have been described in the patent literature which are designed to reduce the oscillator-frequency voltage across the signal circuit and thus minimize the variations.[27,28]

Coupling of the oscillator voltage, into or across, the signal circuit is also accompanied by changes in effective oscillator voltage when the tuning is varied. These changes are not so great with pure inductive coupling as with pure capacitance coupling. In many practical cases, both couplings are present.

A method of reducing the variation of conversion gain with effective oscillator voltage in tubes in which oscillator voltage and signal are placed on the same grid, employs automatic bias. Automatic bias may be obtained either by a cathode self-bias resistor (by-passed to radio frequency) or by a high-resistance grid leak, or both. An illustration of the improvement which may be obtained in this way is shown in Figure 5. Three curves of conversion transconductance, at oscillator fundamental, against effective peak oscillator volts are shown for the typical variable-μ pentode of Figure 4 used as a mixer. For the curve a, a fixed bias was used at approximately an optimum point. The curve is stopped at the grid-current point because operation beyond this point is not practicable in a receiver. Curve b shows the same tube operated with a cathode self-bias resistor. This curve is also stopped at the grid-current point. Curve c shows operation with a high-resistance grid leak. It is evident that, above an oscillator voltage of about 3, curve b is somewhat flatter, and c is considerably flatter than the fixed-bias curve a. The high-resistance grid leak used for c may be made a part of the automatic-volume-control filter but care must be taken that its value is considerably higher than the resistance in the automatic-volume-control circuit which is common to other tubes in the receiver. If this is not done, all the tubes will be biased down with large oscillator swings. When a high-resistance leak is used, the

[27] H. J. J. M. deRegnauld de Bellescize, United States Patent No. 1,872,634; M. Gausner, French Patent No. 639,028.

[28] V. E. Whitman, United States Patent No. 1,893,813.

automatic-volume-control action does not begin in the mixer tube until the automatic-volume-control bias has exceeded the peak oscillator voltage. Because of the high resistance of the leak, the signal circuit is not loaded appreciably by the mixer tube. In a practical case, precautions must be taken that a pentode in the converter stage is not operated at excessive currents when accidental failure of the oscillator reduces the bias. A series dropping resistor in the screen-grid supply will prevent such overload. When a series screen resistor is used, the curve of conversion transconductance versus oscillator voltage is even flatter than the best of the curves shown in Figure 5. Series screen operation, therefore, is highly desirable.[23]
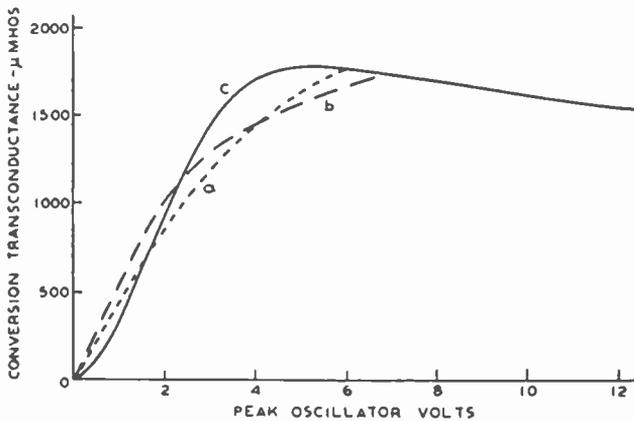

Fig. 5—Conversion transconductance of a typical variable-μ, radio-frequency pentode. Oscillator and signal voltages both applied to grid No. 1. *a*, fixed-bias operation; *b*, cathode resistor used to obtain bias; *c*, bias obtained by means of a high-resistance grid leak.

One of the effects of feedback through interelectrode capacitance in vacuum tubes is a severe loading of the input circuit when an inductance is present in the cathode circuit. Thus, in mixers using cathode injection, the signal circuit is frequently heavily damped since the oscillator circuit is inductive at signal frequency in the usual case. The feedback occurs through the grid-to-cathode capacitance and can be neutralized to some extent by a split cathode coil with a neutralizing capacitance.[28] Such neutralization also minimizes the voltage drop of oscillator frequency across the signal circuit.

Loading of the signal circuit by feedback from the plate circuit of modulators or mixers may also be serious when the signal-grid-to-plate capacitance is appreciable. This is especially true when a low-capacitance intermediate-frequency circuit, which presents a comparatively

high capacitive reactance at signal frequency, is used, as a wide-band intermediate-frequency circuits. The grid-plate capacitance of radio-frequency pentodes is usually small enough so that the effect is negligible in these tubes. In triodes, however, feedback from the intermediate-frequency circuit may be serious and the grid-plate capacitance should be minimized in tube and circuit design. Although neutralization is a possible solution to the plate feedback, a more promising solution is the use of a specially designed intermediate-frequency circuit which offers a low impedance at signal frequency by the equivalent of series tuning and yet causes little or no sacrifice in intermediate-frequency performance.

At high frequencies, the converter stage exhibits phenomena not usually observable at low frequencies. One group of phenomena is caused not by the high operating frequency, per se, but rather by a high ratio of operating frequency to intermediate frequency (i.e., a small separation between signal and oscillator frequencies). Among these phenomena may be listed pull-in and interlocking between oscillator and signal circuits and poor image response. In mixers in which oscillator and signal are impressed on the same grid. the first of these effects is usually pronounced because of the close coupling between the oscillator and signal circuits. It can be reduced by special coupling from the local oscillator at an increase in the complexity of the circuit.

Other phenomena, which are due to the high operating frequency, occur in mixers irrespective of the intermediate-frequency. The most important of these are those caused by transit-time effects in the tube and by finite inductances and mutual inductances in the leads to the tube. When the oscillator and signal are impressed on the same grid of a mixer, the effects are not dissimilar to those in the same tube used as an amplifier. So far as the signal is concerned, the operation is similar to that of an amplifier whose plate current and transconductance are periodically varied at another frequency (that of the oscillator). The effects at signal frequency must, therefore, be integrated or averaged over the oscillator cycle. The input conductance at 60 megacycles of the typical radio-frequency pentode used for Figures 4 and 5 as a function of control-grid bias is shown in Figure 6. The integrated or net loading as a function of oscillator amplitude, when the tube is used as a mixer at this frequency, is given in Figure 7, both with fixed-bias operation and with the bias obtained by a grid leak and condenser. The conductance for all other frequencies may be calculated by remembering that the input conductance increases with the square of the frequency. The data given do not hold for cathode

injection because of the loading added by feedback, as previously discussed.

When automatic volume control is used on the modulator tube, an important effect in some circuits is the change in input capacitance and input loading with bias. This is especially true when low-capacitance circuits are in use, as with a wide-band amplifier. With tubes having oscillator and signal voltages on the same grid, because of the integrating action of the oscillator voltage, the changes are not so pronounced as with the same tube used as amplifier. A small, un-by-passed cathode resistor may be used with an amplifier tube[29],[30] to reduce the variations; it should give a similar improvement with the modulator.
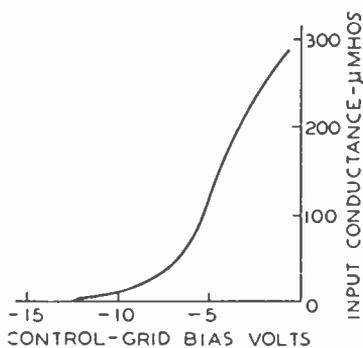
Fig. 6—Input conductance of a typical variable-μ, radio-frequency pentode, at 60 megacycles.
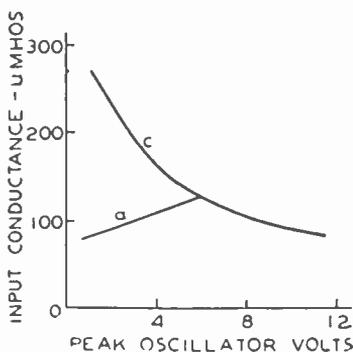
Fig. 7—Input conductance of a typical pentode when used as a mixer at 60 megacycles. *a*, fixed-bias operation; *c*, bias obtained by means of a high-resistance grid leak.

The question of tube noise (i.e., shot-effect fluctuations) is important in a mixer, or modulator, especially when this tube is the first tube in a receiver. There is little doubt that triode or pentode mixers, in which signal and oscillator voltages are impressed on the control grid, give the highest signal-to-noise ratio of any of the commonly used types of mixers. The reason for this has been made clear by recent studies of tube noise.[24] It is now well established that tube noise is the combined result of shot noise in the cathode current which is damped by space charge to a low value and additional fluctuations

[29] M. J. O. Strutt and A. van der Ziel, "Simple circuit means for improving short-wave performance of amplifier tubes," *Elek. Nach. Tech.*, Vol. 13, pp. 260-268; August, 1936.

[30] R. L. Freeman, "Use of feedback to compensate for vacuum-tube input-capacitance variations with grid bias," *Proc. I.R.E.*, Vol. 26, pp. 1360-1366; November, 1938.

in the plate current caused by random variations in primary current distribution between the various positive electrodes. Thus, in general, tubes with the smallest current to positive electrodes other than the plate have the lowest noise. It is seen that the tetrode or pentode modulator, with a primary screen current of 25 per cent or less of the total current, is inherently lower in noise than the more complex modulators in which the current to positive electrodes other than the plate usually exceeds 60 per cent of the total current. The triode, of course, has the lowest noise assuming an equivalent tube structure. The conversion transconductance of triode, tetrode, or pentode mixers is usually higher than that of multielectrode tubes using a similar cathode and first-grid structure. That this is so is again largely due to the lower value of wasted current to other electrodes.

The noise of triodes and pentodes used as mixers in the converter stage is conveniently expressed in terms of an equivalent noise resistance R_{cq} as mentioned in Section II C. The noise as a mixer, of both the triode and the pentode, may be expressed in one formula based on the now well-understood amplifier noise relations.[24] The equivalent noise resistance of the triode is obtained simply by equating the screen current to zero. An approximate formula for equivalent noise resistance of oxide-coated-cathode tubes is

$$R_{cq} \text{ (of triode and pentode mixers)} = \frac{2.2 \, \overline{g_m} + 20 \, \overline{I_{c2}}}{g_c^2} \frac{1}{1 + \alpha}$$

where $\overline{g_m}$ is the average control-grid-to-plate transconductance (averaged over an oscillator cycle), $\overline{I_{c2}}$ is the average screen current, g_c is the conversion transconductance, and α is the ratio of the screen current to plate current. Valuable additions to the above relation are given by formulas which enable a simple calculation of noise resistance from amplier data found in any tube handbook. These additional relations are approximations derived from typical curve shapes and are based on the maximum peak cathode current I_0 and the maximum peak cathode transconductance g_0. The data are given in Table I. It has been assumed that oscillator excitation is approximately optimum. In this table, E_{c0} is the control-grid voltage needed to cut off the plate current of the tube with the plate and screen voltages applied, and α is the ratio of screen to plate current.

As an example of the use of the table, suppose it is desired to find the equivalent noise resistance of a particular triode operated as a converter at the oscillator second harmonic. The local oscillator can be

Table I—Mixer Noise of Triodes and Pentodes
(Oscillator and Signal both Applied to Control Grid)

Operation	Approximate Oscillator Peak Volts	Average Transconductance $\overline{g_m}$	Average Cathode Current $\overline{I_k}$	Conversion Transconductance g_o	Equivalent Noise Resistance R_{eq}
At Oscillator Fundamental	0.7 E_{co}	$\dfrac{0.47}{1+\alpha}\,g_o$	0.35 I_o	$\dfrac{0.28}{1+\alpha}\,g_o$	$\dfrac{13}{g_o} + 90\,\dfrac{I_o}{g_o^2}\,\alpha$
At Oscillator 2nd Harmonic	1.5 E_{co}	$\dfrac{0.25}{1+\alpha}\,g_o$	0.20 I_o	$\dfrac{0.13}{1+\alpha}\,g_o$	$\dfrac{31}{g_o} + 220\,\dfrac{I_o}{g_o^2}\,\alpha$
At Oscillator 3rd Harmonic	4.3 E_{co}	$\dfrac{0.15}{1+\alpha}\,g_o$	0.11 I_o	$\dfrac{0.09}{1+\alpha}\,g_o$	$\dfrac{38}{g_o} + 260\,\dfrac{I_o}{g_o^2}\,\alpha$

permitted to swing the triode mixer grid to zero bias. With a plate voltage of 180 volts and zero bias, the tube data sheet shows a transconductance, $g_0 = 2.6 \times 10^{-3}$ *mho*. Thus the equivalent noise resistance is $31/g_0$ or 12,000 ohms and the conversion transconductance at second harmonic is 0.13 g_0, or 340 micromhos. Since, with this plate voltage the tube cuts off at about 8 volts, a peak oscillator voltage of around 12 volts will be required.

The above table may also be used to obtain a rough estimate of the input loading of pentode or triode mixers, since the high-frequency input conductance is roughly proportional to the average transconductance $\overline{g_m}$ and to the square of the frequency. Thus, if the loading at any transconductance and frequency is known, the loading as a mixer under the conditions of the table may quickly be computed.

B. Tubes with Oscillator Voltage on an Inner Grid, Signal Voltage on an Outer Grid.

When the oscillator voltage is impressed on the grid nearest the cathode of a mixer or converter, the cathode current is varied at oscillator frequency. The signal grid, on the other hand, may be placed later in the electron stream to serve only to change the distribution of the current between the output anode and the other positive electrodes. When the two control grids are separated by a screen grid, the undesirable coupling between oscillator and signal circuits is reduced much below the value which otherwise would be found.

The signal-grid-to-plate transconductance of the inner-grid injection mixer is a function of the total current reaching the signal

grid; this current, and hence the signal-grid transconductance, will vary at oscillator frequency so that mixing becomes possible. The signal-grid transconductance as a function of oscillator-grid potential of a typical modulator of this kind is shown in Figure 8. It will be observed that this characteristic is different in shape from the corresponding curve of Figure 4 for the tube with oscillator and signal voltages on the same grid. The chief point of difference is that a definite peak in transconductance is found. The plate current of the tube shows a saturation at approximately the same bias as that at which the peak in transconductance occurs, indicating the formation of a partial virtual cathode. The signal grid, over the whole of these curves, is biased negatively and so draws no current. The oscillator inner grid (No. 1 grid), however, draws current at positive values of bias. This separation of signal and oscillator grids is advantageous, inasmuch as the signal circuit is not loaded even though the oscillator amplitude is sufficient to draw grid current. In fact, in the usual circuit, the oscillator grid is self-biased with a low-resistance leak and condenser and swings sufficiently far positive to attain the peak signal-grid transconductance.

Fig. 8 — Signal-grid-(grid No. 4) to-plate transconductance versus oscillator-grid (grid No. 1) voltage curve of a typical mixer designed for inner-grid injection. Signal-grid bias = — 3 volts.

The conversion transconductance of such a tube has a maximum with an oscillator swing which exceeds the point of maximum signal-grid transconductance in the one direction and which cuts off this transconductance over slightly less than half the cycle, in the other. Curves of conversion transconductance against peak oscillator voltage are shown in Figure 9. Curve a is for fixed-bias operation of the oscillator grid, curve b is with a high-resistance (i.e., several megohms) grid leak and condenser for bias, and curve c is with the recommended value of grid leak (50,000 ohms) for this type of tube. It is seen that best operation is obtained with the lower resistance value of grid leak. With this value, the negative bias produced by rectification in the grid circuit is reduced enough to allow the oscillator grid to swing appreciably positive over part of the cycle. An incidental advantage

to the use of the low-resistance leak when the tube is self-oscillating (i.e., a converter) is that undesirable relaxation oscillations are minimized.

In mixers or converters in which the oscillator voltage is present on both the cathode and the oscillator grid in the same phase (e.g., Figure 3) it is usually necessary to utilize a relatively sharp cutoff in the design of the oscillator grid so as to cut off the cathode current when the signal grid is positive.[19] By this means the signal grid is prevented from drawing current. At the same time, however, the high currents needed for a high peak value of signal-grid transconductance cannot be obtained without a greater positive swing of the oscillator grid than with a more open oscillator grid structure. Thus, it is clear that it is desirable to have a negative bias on the oscillator electrode which is considerably smaller than the peak oscillator voltage. For this reason, optimum results are obtained on these tubes with very low values of oscillator grid leak (e.g., 10,000 to 20,000 ohms).

The effects of feedback through the interelectrode capacitance are small in well-designed multigrid mixers and converters of the kind covered in this section. The signal-grid-to-plate capacitance is usually small enough to play no part in the operation; even with a high *L*-to-*C* ratio in the intermediate-frequency transformer, the capacitive reactance of the intermediate-frequency circuit at signal frequency is only a very small fraction of the feedback reactance.

Fig. 9—Conversion transconductance of a typical mixer designed for inner-grid injection of oscillator. Signal-grid bias $= -3$ volts. *a*, fixed-bias operation of oscillator grid; *b*, oscillator-grid bias obtained through high-resistance grid leak; *c*, oscillator-grid bias obtained through a 50,000-ohm grid leak.

The other interelectrode capacitance which plays some part in determining circuit performance (excluding, of course, the input and output capacitances) is the capacitance from the oscillator electrode or electrodes to the signal grid. This capacitance is a source of coupling between these two circuits. In well-designed converter or modulator tubes of the type discussed in this section, however, the coupling through the capacitance may be made small compared with another form of internal coupling known as "space-charge coupling," which will be treated later in this discussion.
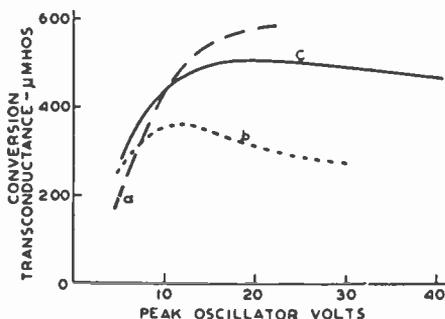
Coupling between oscillator and signal circuits is of no great consequence except when an appreciable voltage of oscillator frequency is built up across the signal-grid circuit. This is not usually possible unless the signal circuit is nearly in tune with the oscillator as it is when a low ratio of intermediate frequency to signal frequency is used. The effect of oscillator-frequency voltage induced across the signal circuit depends on its phase; the effect is usually either to increase or to decrease the relative modulation of the plate current at oscillator frequency and so to change the conversion transconductance. This action is a disadvantage, particularly when the amount of induced voltage changes when the tuning is varied, as usually occurs. In some cases, another effect is a flow of grid current to the signal grid; this may happen when the oscillator-frequency voltage across the signal-grid circuit exceeds the bias. Grid current caused by this effect can usually be distinguished from grid current due to other causes. By-passing or short-circuiting the signal-grid circuit reduces the oscillator-frequency voltage across the signal-grid circuit to zero. Any remaining grid current must, therefore, be due to other causes.

Current to a negative signal grid of a tube operated with inner-grid oscillator injection is sometimes observed at high frequencies (e.g., over 20 megacycles) even when no impedance is present in the signal-grid circuit. This current is caused by electrons whose effective initial velocity has been increased by their finite transit time in the high-frequency alternating field around the oscillator grid. These electrons are then able to strike a signal grid which is several volts negative. The magnitude of the signal-grid current is not usually as great as with tubes applying the oscillator voltage to an outer grid[31] although it may prevent the use of an automatic-volume-control voltage on the tube.

An investigation of coupling effects in the pentagrid converter showed that the coupling was much larger than could be explained by interelectrode capacitance. It was furthermore discovered that the apparent coupling induced a voltage on the signal circuit in opposite phase to that induced by a capacitance from oscillator to signal grid.[15] The coupling which occurred was due to variations in space charge in front of the signal grid at oscillator frequency. A qualitative explanation for the observed behavior is that, when the oscillator-grid voltage is increased, the electron charge density adjacent to the signal grid is increased and electrons are repelled from the

---

[31] The next part of this section contains a more detailed discussion of signal-grid current in outer-grid oscillator injection tubes.

signal grid. A capacitance between the oscillator grid and the signal grid would have the opposite effect. The coupling, therefore, may be said to be approximately equivalent to a negative capacitance from the oscillator grid to the signal grid. The effect is not reversible because an increase of potential on the signal grid does not increase the electron charge density around the oscillator grid. If anything, it decreases the charge density. The equivalence to a negative capacitance must be restricted to a one-way negative capacitance and, as will be shown later, is restricted also to low-frequency operation.

In general, the use of an equivalent impedance from oscillator grid to signal grid to explain the behavior of "space-charge coupling" is somewhat artificial. A better point of view is simply that a current is induced in the signal grid which depends on the oscillator-grid voltage. Thus, a transadmittance exists between the two electrodes analogous to the transconductance of an ordinary amplifier tube. Indeed, the effect has been used for amplification in a very similar manner to the use of the transconductance of the conventional tube.[32,33]

It is found that the transadmittance from the oscillator to the signal electrode Y_{m0-s} is of the form

$$Y_{m0-s} = k_1\omega^2 + jk_2\omega.$$

At low frequencies (i.e., $k_1\omega^2 << k_2\omega$) the transadmittance is mainly a transsusceptance but, as the frequency rises, the transconductance component $k_1\omega^2$ becomes of more and more importance, eventually exceeding the transsusceptance in magnitude. The early work on "space-charge coupling" indicated that the effect was opposite to that of a capacitance connected from oscillator to signal grid and could be canceled by the connection of such a capacitance of the correct value.[15,34] The effect of such cancellation could be only partial, however, since only the transsusceptance was balanced out by this arrangement. For complete cancellation it is also necessary to connect a conductance, the required value of which increases as the square of the frequency, between the oscillator grid and the signal grid so that the transconductance term is also balanced out.[35,36]

---

[32] C. J. Bakker and G. de Vries, "Amplication of small alternating tensions by an inductive action of the electrons in a radio valve," *Physica*, Vol. 1, pp. 1045-1054, October-November, 1934.

[33] A. M. Nicolson, United States Patent No. 1,255,211 (applied for in 1915).

[34] M. J. O. Strutt, "Frequency changers in all-wave receivers," *Wireless Eng.*, Vol. 14, pp. 184-192; April, 1937.

The cancellation of "space-charge coupling" may be viewed in another way. A well-known method of measuring the transadmittance of a vacuum tube is to connect an admittance from control grid to output electrode and to vary this admittance until no alternating-current output is found with a signal applied to the control grid.[37] The external admittance is then equal to the transadmittance. In exactly the same way, the transadmittance which results from the space-charge coupling may be measured. As a step further, if an admittance can be found which substantially equals the transadmittance at all frequencies or over the band of frequencies to be used, this admittance may be permanently connected so as to cancel the effects of space-charge coupling. As has been previously stated, the admittance which is required is a capacitance and a conductance whose value varies as the square of the frequency. Such an admittance is given to a first approximation by the series connection of a capacitance C and a resistance R. Up to an angular frequency $\omega = 0.3/CR$ the admittance of this combination is substantially as desired. At higher values of frequency, the conductance and susceptance fail to rise rapidly enough and the cancellation is less complete. Other circuits are a better approximation to the desired admittance. For example, the connection of a small inductance, having the value $L = 1/2CR^2$, gives a good approximation up to an angular frequency $\omega = 0.6/CR$. The latter circuit is, therefore, effective to a frequency twice as high as the simple series arrangement of capacitance and resistance. Inasmuch as in some cases the value of inductance needed is only a fraction of a microhenry, the inductance may conveniently be derived from proper proportioning and configuration of the circuit leads.

It is of interest to note the order of magnitude of the transadmittance which is measured in the usual converter and mixer tubes.[26,35,36,38] In the formula for Y_{mo-s} given above, k_1 is in the neighborhood of 10^{-21} and k_2 is around 10^{-12}. Cancellation is effected by a capacitance of the order of one or two micromicrofarads and a series resistance of 500 to 1000 ohms.

The correct value of the canceling admittance may be found ex-

[35] E. W. Herold, "Frequency changers in all-wave receivers," (Letter to Editor), *Wireless Eng.*, Vol. 14, pp. 488-489; September, 1937.

[36] E. W. Herold, United States Patent No. 2,141,750.

[37] F. B. Llewellyn, "Phase angle of vacuum tube transconductance at very high frequencies," *Proc. I.R.E.*, Vol. 22, pp. 947-956; August, 1934.

[38] M. J. O. Strutt, "Frequency changers in all-wave receivers," (Letter to Editor), *Wireless Eng.*, Vol. 14, p. 606; November, 1937.

perimentally by adjustment so that no oscillator voltage is present across the signal-grid circuit when the latter is tuned to the oscillator frequency. Another method which may be used is to observe either the mixer or converter plate current or the oscillator grid current as the tuning of the signal circuit is varied through the oscillator frequency. With proper adjustment of the canceling admittance there will be no reaction of the signal-circuit tuning on either of these currents.

There are two disadvantages which accompany the cancellation of space-charge coupling as outlined. In the first place, the signal-grid input admittance is increased by the canceling admittance. This point will be brought up again after discussing the input admittance. The second disadvantage is that the oscillator frequency shift with voltage changes in converter tubes may be somewhat increased by the use of this cancellating admittance. When separate oscillator and mixer tubes are used, the latter effect may be made less serious.

The next point to be considered is the input admittance of the signal grid. Signal-grid admittance curves of a typical modulator designed for use with the oscillator voltage impressed on the first grid are shown under direct-current conditions (i.e., as a function of oscillator-grid bias for several values of signal-grid bias) in Fig. 10. The admittance is separated into conductive and susceptive components, the latter being plotted in terms of equivalent capacitance. The admittance components of the "cold" tube (no electrons present) have been subtracted from the measured value so that the plotted results represent the admittance due to the presence of electrons only. The data shown were taken at 31.5 megacycles with a measuring signal which did not exceed 1.0 volt peak at any time. A modified Boonton Q meter was used to take the data. It should be noted that the presence of a marked conductive component of admittance is to be expected at frequencies as high as those used.

The most striking feature of the data of Fig. 10 is that both susceptive and conductive components are negative over a large portion of the characteristic. The Appendix discusses this feature in somewhat more detail. The measurements show that the susceptive component is analogous to a capacitance. The capacitance curves given are independent of frequency up to the highest frequency used (approximately 50 megacycles). The conductive component, on the other hand, increases as the square of the frequency also up to this frequency. The conductance is, therefore, negative even at very low frequencies although its magnitude is then very small. Thus, the conductance curves of Fig. 11 are valid for any frequency by multiplication

of the conductance axis by the square of the ratio of the frequency considered, to the frequency used for the data (i.e., 31.5 megacycles). Data taken at various frequencies for two particular values of grid bias voltage E_{c4} are plotted in Fig. 11. The square-law relation is shown to check very closely.
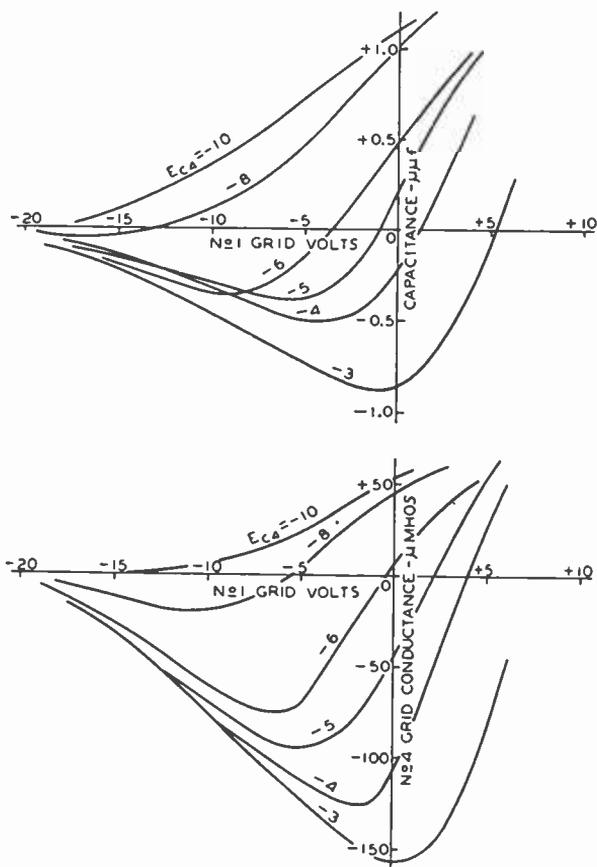

Fig. 10—Signal-grid (grid No. 4) admittance of a typical mixer designed for inner-grid injection of oscillator at 31.5 megacycles. Curves taken with no oscillator voltage applied. Data represents electronic admittance only (i.e., "cold" values were subtracted from measured values before plotting.

Fig. 10 should be considered remembering that the oscillator voltage is applied along the axis of abscissas. Considering an applied oscillator voltage, the admittance curves must be integrated over the oscillator cycle to find the admittance to the signal frequency. The operation is just as if the tube were an amplifier whose input admittance is periodically varied over the curve of Fig. 10 which corresponds to the signal-grid bias which is used. Curves of the modulator input conductance at 31.5 megacycles for various applied oscillator voltages are shown in Fig. 12. The oscillator-grid bias is obtained by means
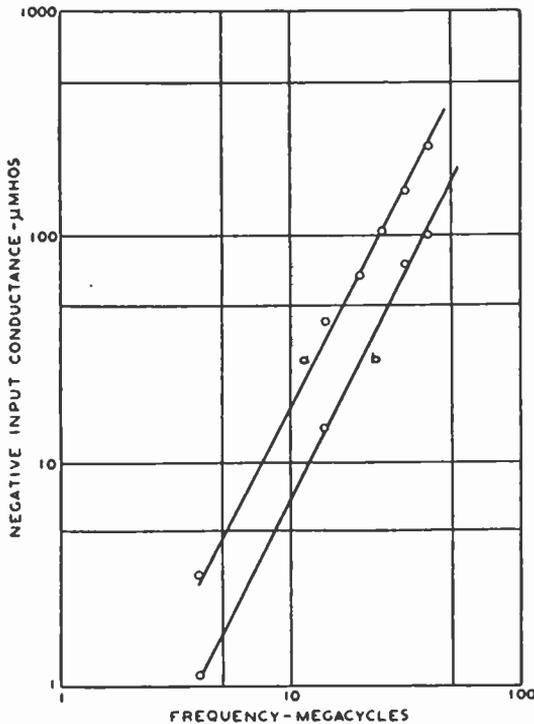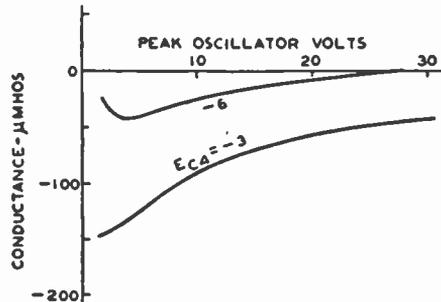
Fig. 11—Signal-grid (grid No. 4) conductance of a typical mixer designed for inner-grid injection of oscillator. Lines are drawn with slope of 2. Curve a taken with $E_{c1} = 0$, $E_{c4} = -3$ volts. Curve b taken with $E_{c1} = -6$, $E_{c4} = -6$ volts.

of the recommended value of grid leak for the tube (50,000 ohms). Curves are shown for two values of signal-grid bias voltage E_{c4}. As before, data for other frequencies are obtained by multiplying the conductance by the square of the frequency ratio.

The practical effect of the negative input admittance in a circuit is due to the conductive portion only, inasmuch as the total input capacitance remains positive in general.[39] An improved image ratio, and somewhat greater gain to the converter signal grid over other types of modulator is to be expected when this type of oscillator injection is used.

Fig. 12 — Signal-grid (grid No. 4) conductance of a typical mixer designed for inner-grid injection of oscillator, at 31.5 megacycles. Oscillator voltage applied. Oscillator-grid bias obtained through 50,000-ohm grid leak. Electronic portion of conductance, only, plotted.

[39] It should not be forgotten that the data given do not include the "cold" susceptance and conductance of the tube. The latter is a relatively small quantity, however.

At high frequencies, when a comparatively low intermediate frequency is used, it is usually desirable to cancel the space-charge coupling of the tube in the manner previously discussed. When this cancellation is made reasonably complete by the use of a condenser and resistor combination connected from the oscillator grid to the signal grid, the losses in this admittance at signal frequency are usually sufficient to wipe out the negative input admittance. The net positive input conductance however is often less than that found with other types of mixer.

The change in signal-grid input capacitance with automatic-volume-control voltage is small in this type of modulator, particularly with the larger values of oscillator swing because of the integrating action of the oscillator voltage.

The fluctuation noise which is found in the output of inner-grid oscillator-injection mixers and converters is not readily evaluated quantitatively. The fluctuation noise is primarily due to current-distribution fluctuations but is complicated by the possibility of a virtual cathode ahead of the signal grid. Data have been taken, however, which indicate some degree of proportionality between the mean-squared noise current and the plate current. The signal-to-noise ratio for this type of modulator is, therefore, approximately proportional to the ratio of conversion transconductance to the square root of the plate current. It is considerably less than for the pentode modulator with both signal and oscillator voltages on the control grid.

The noise of the converter or mixer with oscillator on an inner grid may be expressed in terms of an equivalent grid resistance as

$$R_{eq} = \frac{20\,\overline{I}_b}{g_c^2} F^2$$

where $\overline{I}_b$ is the operating plate current, g_c is the conversion transconductance, and F^2 is a factor which is about 0.5 for tubes with suppressor grids and at full gain. For tubes without suppressor or for tubes whose gain is reduced by signal-grid bias, F^2 is somewhat larger and approaches unity as a maximum. With this mode of operation there is not so much value in expressions for R_{eq} based on maximum transconductance and maximum plate current because these quantities are neither available nor are they easily measured. For operation at second or third harmonics of the oscillator (assuming optimum oscillator excitation) the plate current $\overline{I}_b$ and the conversion transconductance g_c are roughly ½ or ⅓, respectively, of their values with

fundamental operation so that the equivalent noise resistance for second-harmonic and third-harmonic operation is around two and three times, respectively, of its value for fundamental operation.

C. Mixers with Oscillator Voltage on an Outer Grid, Signal Voltage on Inner Grid

With this type of mixer, the cathode current is modulated by the relatively small signal voltage which is impressed on the control grid adjacent to the cathode. The oscillator voltage, on the other hand, is impressed on a later control grid so that it periodically alters the current distribution between anode and screen grid. The connections of signal and oscillator voltages to this type of modulator are just the reverse, therefore, of the mixer treated in the preceding section. The behaviors of the two types are also quite different although they both include internal separation of signal and oscillator electrodes through a shielding screen grid.

The signal-grid transconductance curve as a function of oscillator-grid voltage of a typical mixer designed for use with the oscillator on an outer grid is shown in Fig. 13. It differs in shape from similar curves for the other two classes of modulator in that an approximate saturation is reached around zero bias on the oscillator grid. The
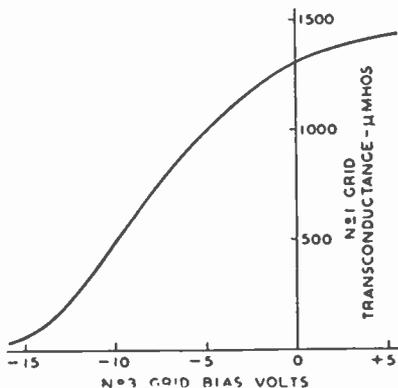

Fig. 13—Signal-grid (grid No. 1) transconductance versus oscillator-grid (grid No. 3) voltage of a typical mixer designed for use with outer-grid injection of oscillator. Signal-grid bias = — 3 volts.

conversion transconductance for such a tube is, therefore, more accurately predicted from normal amplifier transconductance. In fact, in the manufacture of this type of mixer, a test of signal-grid transconductance at somewhere near the saturation point (e.g., zero bias) on the oscillator grid has been found to correlate almost exactly with the conversion transconductance. The cutoff point of the curve must remain approximately fixed, of course, since this point affects the oscillator amplitude which is necessary.

The conversion transconductance of the typical outer-grid injection mixer tube which was used for Fig. 13 is shown in Fig. 14. Curve *a* which is for fixed bias on the oscillator grid is seen to be higher than curve *b* for which bias is obtained by a 50,000-ohm grid leak and

condenser. The latter connection is most widely used, however, because of its convenience. A compromise using fixed bias together with a grid leak is most satisfactory of all.[40] When this combination is used, the curve of conversion transconductance follows curve a of Fig. 14 to the intersection with curve b and then follows along the flat top of curve b.

In a well-designed mixer with the signal voltage on the grid adjacent to the cathode and the oscillator voltage on an outer grid, effects due to feedback through the interelectrode capacitance may usually be neglected. The only effect which might be of importance in some cases is coupling of the oscillator to the signal circuit through the signal-grid-to-oscillator-grid capacitance. In many tubes a small amount of space-charge coupling between these grids is also present and adds to the capacitance coupling (contrary to the space-charge coupling discussed in Section B which op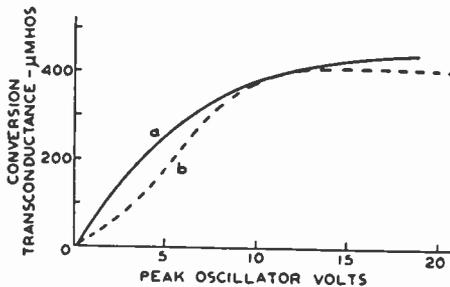poses the capacitance coupling in that case). Measurements of the magnitude of the space-charge coupling for this type of modulator show that it is of the order of 1/5 to 1/10 of that present in inner-grid-injection modulators. Coupling between oscillator and signal circuits causes a voltage of oscillator frequency to be built up across the signal input circuit. This oscillator-frequency voltage, d e p e n d i n g on its phase, aids or opposes the effect of the normal oscillator-grid alternating voltage. The action is additive when the signal circuit has capacitative reactance to the oscillator frequency, as in the usual case. When the oscillator-frequency voltage across the signal input circuit exceeds the bias, grid current is drawn to the signal grid, an undesirable occurrence. This grid current may be distinguished from signal-grid current due to other causes by short-circuiting the signal-input circuit and noting the change in grid current. With the majority of tubes, another cause of signal-grid current far exceeds this one in importance. This other cause will now be discussed.

Fig. 14—Conversion transconductance of a typical mixer designed for outer-grid injection of oscillator. Signal-grid bias, $E_{c1} = -3$ volts. Curve a corresponds to fixed No. 3 grid bias, $E_{c2} = -8$ volts. Curve b corresponds to bias obtained through a 50,000-ohm grid leak.

---

[40] E. W. Herold, United States Patent No. 2,066,038.

The most prominent high-frequency effect which was observed in mixers of the kind under discussion, was a direct current to the negative signal grid even when no impedance was present in this grid circuit. This effect was investigated and found to be due to the finite time of transit of the electrons which pass through the signal grid and are repelled at the oscillator grid, returning to pass near the signal grid again.[17,41,42] When the oscillator frequency is high, the oscillator-grid potential varies an appreciable amount during the time that such electrons are in the space between screen grid and oscillator grid. These electrons may, therefore, be accelerated in their return path more than they were decelerated in their forward path. Thus, they may arrive at the signal grid with an additional velocity sufficient to allow them to strike a slightly negative electrode. Some electrons may make many such trips before being collected; moreover, in each trip their velocity is increased so that they may receive a total increase in velocity equivalent to several volts. A rough estimate of the grid current to be expected from a given tube is given by the semiempirical equation

$$ I_{c1} = A \, I_k E_{osc} \omega \tau_{2-3} e^{BE_{c1}}. $$

Where A and B depend on electrode voltages and configuration, I_{c1} is the signal-grid current, E_{c1} is the signal-grid bias, I_k is the cathode current, E_{osc} is the impressed oscillator voltage on the oscillator grid, ω is the angular frequency of the oscillator, and τ_{2-3} is the electron transit time in the space between screen grid and oscillator grid.

Data on the signal-grid current of a typical mixer at 20 megacycles are shown in Fig. 15 where a semi-logarithmic plot is used to indicate the origin of the above equation.

The reduction of signal-grid current by operation at more negative signal-grid bias values is an obvious remedy. When this is done, in order to prevent a reduction in conversion transconductance, the screen voltage must be raised. A better method of reducing the undesired grid current lies in a change of tube design. It will be shown in a later part of this discussion that the constant A and/or the transit time τ_{2-3} of the above formula may be reduced considerably by proper electrode configuration.

---

[41] K. Steimel, "The influence of inertia and transit time of electrons in broadcast receiving tubes," *Telefunken-Röhre*, No. 5, pp. 213-218; November, 1935.

[42] K. S. Knol, M. J. O. Strutt, and A. van der Ziel, "On the motion of electrons in an alternating electric field," *Physica*, Vol. 5, pp. 325-334; May, 1938.

Another high-frequency phenomenon which is particularly noticed in outer-grid-injection mixers is the high input conductance due to transit-time effects. The cause for this was first made evident when the change of signal-grid admittance with oscillator-grid potential was observed. Fig. 16 gives data on the susceptive and conductive components of the signal-grid admittance of this type of modulator as a function of oscillator-grid bias (no oscillator voltage applied). The data were taken at 31.5 megacycles and, as in the other input admittance curves, show the admittance components due to the presence of electrons only. It is seen that when the No. 3 grid is made sufficiently negative the input admittance is greatly increased. This behavior coincides, of course, with plate-current cutoff. It seems clear that the electrons which are turned back at the No. 3 grid and which again reach the signal-grid are the cause of the increased admittance. Calculations based on this explanation have been published by M. J. O. Strutt[43] and show reasonable quantitative agreement with experiment. As in the other cases above, the upper curve of Fig. 16 is approximately independent of frequency while the lower one may be converted to any other frequency by multiplying the ordinates by the square of the frequency ratio.

Fig. 15—Signal-grid (grid No. 1) current in a typical mixer with a 20-megacycle oscillator voltage applied to grid No. 3. $E_{c3} = -10$ volts. $E_{c2 \text{ and } 4} = 100$ volts, $E_b = 250$ volts.

When an oscillator voltage is applied, the No. 3 grid bias is periodically varied at oscillator frequency. The net input admittance is then the average value over the oscillator cycle. Such net values of the conductance component are shown in Fig. 17. The frequency for these curves is 31.5 megacycles. Values for other frequencies are obtained by multiplying the ordinates by the square of the frequency

43 M. J. O. Strutt, and A. van der Ziel, "Dynamic measurements of electron motion in multigrid tubes," Elek. Nach. Tech., Vol. 15, pp. 277-283; September, 1938.

ratio. Curve *a* coincides with the fixed bias condition of curve *a* of Fig. 14 while curve *b* corresponds to the grid-leak and condenser bias as in *b* of Fig. 14. The conductance is approximately twice as high when the tube is used as a mixer as when it is used as an amplifier. This is a serious disadvantage, particularly at very high frequencies.

It is thus seen that two serious disadvantages of the outer-grid-injection mixer are both due to the electrons returned by the oscillator grid which pass again to the signal-grid region. It was found possible to prevent this in a practical tube structure by causing the returning electrons to traverse a different path from the one which they traveled in the forward direction.[44,45] The progressive steps towards an im-

provement of this kind are illustrated in Fig. 18 where cross-sectional views of the portion inside the oscillator grid of various developmental modulators are shown. The drawing (a) shows the original design, data on which have been given in Figs. 15, 16, and 17. Drawing (b) of Fig. 18 shows a tube in which two side electrodes operated at a high positive potential were added. In a tube of this kind many of the electrons returned by the No. 3 grid (oscillator grid) travel paths similar to the dotted one shown; they are then collected by the auxiliary electrodes and thus do not re-enter the signal
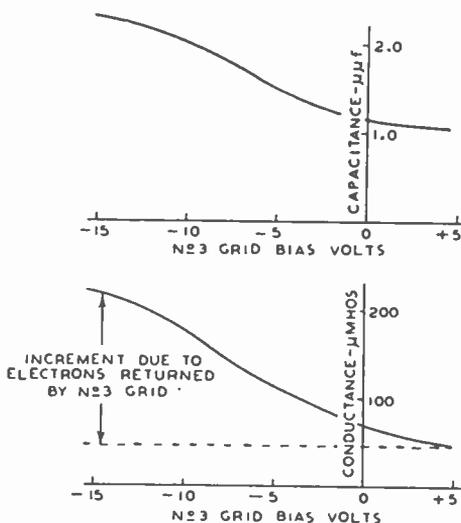

Fig. 16—Signal-grid (grid No. 1) admittance of typical mixer designed for outer-grid injection of oscillator. Data taken at 31.5 megacycles with no oscillator voltage applied. $E_{c1} = -3$ volts, $E_{c2 \text{ and } 4} = 100$ volts, $E_b = 250$ volts.

grid space. Tubes constructed similarly to (b) showed a considerable improvement in the signal-grid admittance increment due to returned electrons. Construction (c) shows the next step in which the side electrodes are increased in size and operated at somewhat lower potential. Because of the undesirability of an additional electrode and lead in the tube, the construction shown at (d) was tried. In this case the auxiliary

[44] The same principles have now been applied to inner-grid-injection mixers and converters. See references 19 and 45.

[45] A. J. W. M. van Overbeck and J. L. H. Jonker, "A new converter valve," *Wireless Eng.*, Vol. 15, pp. 423-431; August, 1938.
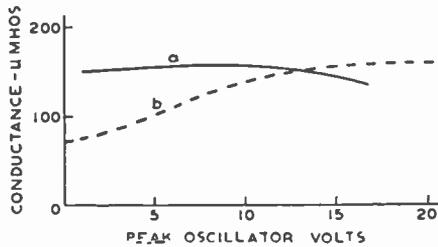
Fig. 17—Signal-grid (grid No. 1) conductance of typical mixer designed for outer-grid injection of oscillator. Frequency, 31.5 megacycles, signal-grid bias, $E_{c1} = -3$ volts. Curve a corresponds to fixed No. 3 grid bias, $E_{c3} = -8$ volts. Curve b corresponds to bias through a 50,000-ohm grid leak.

electrodes are bent over and connected electrically and mechanically to the screen grid. Curves showing the progressive reduction in the signal-grid conductance increment due to returned electrons are shown in Fig. 19. The curves are labeled to correspond with the drawings of Fig. 18. It should be noted that the use of the oscillator-grid support rods in the center of the electron streams as shown in Fig. 18(d) was found to improve the performance. No change in signal-grid conductance with oscillator-grid potential could be observed with this construction.[46] The conductance of the tube as a modulator, therefore, was reduced to less than half of that of construction (a). At the same time, a check of signal-grid current with a high-frequency oscillator applied to the No. 3 grid showed that this current was reduced to 1/20 of that of the original construction (a). The change in construction may be looked upon as dividing the constant A in the grid-current formula previously given, by a factor of more than 20.
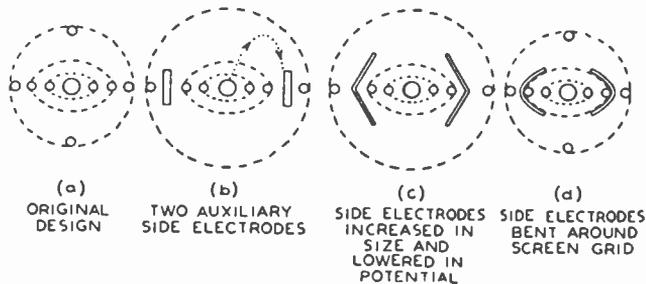

Fig. 18—Cross-sectional views of mixer designed for outer-grid injection of oscillator. The views show only the portions of the tube inside of and including the oscillator injection grid.

[46] It should be mentioned that it is also possible to construct tubes in which the signal-grid conductance decreases somewhat with increasingly negative No. 3 grid bias. This effect is caused by the inductance of the inner screen-grid lead which causes a negative conductance in the input circuit when the inner screen current is high, as at negative No. 3 grid bias values. This negative conductance cancels part of the positive conductance of the signal grid.

Another method of reducing the effect of electrons returned by the oscillator grid is to reduce the effect of electron transit time in the tube. This may be done by reducing the spacings, particularly the screen-grid-to-oscillator-grid spacing. This method of improving modulator performance has two disadvantages compared with the one discussed in connection with Fig. 18. The reduction in spacing is accompanied by a more sloping (i.e., less steep) signal-grid transconductance versus oscillator-grid voltage curve. This change in construction requires an increase in applied oscillator voltage to attain the same conversion transconductance. The second disadvantage is that such a method reduces the transit time and hence, the undesirable high-frequency effects only by an amount bearing some relation to the reduction in spacing. Since this reduction is limited in a given size of tube, the method whereby electron paths are changed is much more effective.

The method of reducing spacing, on the other hand, is extremely simple to adopt. A combination of both methods may be most desirable from the point of view of best performance with least complexity in the tube structure.

In a mixer which must operate at high frequencies, it is not usually sufficient to eliminate the effects of returned electrons in order to assure adequate performance. For this reason the development of the
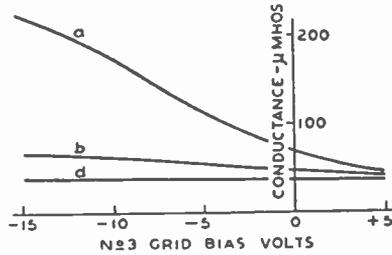

Fig. 19—Signal-grid (grid No. 1) conductance of the outer-grid-injection mixers shown in Fig. 18. Data taken at 31.5 megacycles with no oscillator voltage applied.

principles shown in Fig. 18 was carried on simultaneously with a general program of improving the tube. To this end, tubes were made with somewhat reduced spacings and with a rectangular cathode and a beam-forming signal grid (i.e., one with comparatively large supports). A number of developmental constructions are shown in Fig. 20. Construction (g), it will be noted, has finlike projections on the screen-grid channel members.[47] In construction (h) a reduction of spacing between screen and oscillator grids was combined with the channel construction. The relative performance of these constructions, so far as signal-grid current is concerned is shown in Fig. 21. The frequency used was 20 megacycles. The curve for the original design (taken from Fig. 15) is included and is drawn as *a*. All four of the constructions

---

[47] This construction was devised by Miss Ruth J. Erichsen who was associated with the writer during part of the development work herein described.

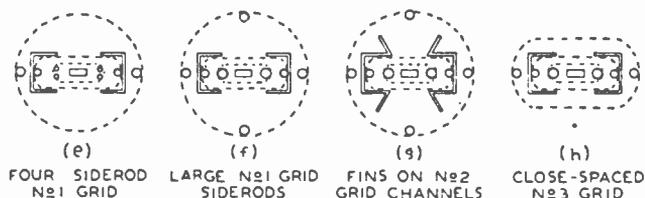| (e)	(f)	(g)	(h)
FOUR SIDEROD No1 GRID	LARGE No1 GRID SIDERODS	FINS ON No2 GRID CHANNELS	CLOSE-SPACED No3 GRID

Fig. 20—Cross-sectional views of improved mixers designed for outer-grid injection of oscillator. The views show only the portions of the tube inside of and including the oscillator injection grid.

of Fig. 20 were satisfactory as regards signal-grid conductance; in every case the change in conductance as the oscillator grid was made negative was a negligible factor. Construction (h) required approximately 20 per cent more oscillator voltage than (e), (f), or (g) because of the reduction in slope of the transconductance versus No. 3 grid voltage curve which accompanied the reduced spacing between the screen and the No. 3 grid.

Outer-grid-injection mixers have the same or slightly greater signal-grid capacitance changes with automatic volume control as are found in amplifier tubes. In this respect they are inferior to inner-grid-injection converters or mixers. The use of a small un-by-passed cathode resistance[29,30] is a help, however.

In closing this section, the subject of fluctuation noise will be considered. Experimental evidence indicates that the major portion of the noise in mixers with oscillator voltage on an outer grid is due to current-distribution fluctuations.[21] The oscillator voltage changes the current distribution from plate to screen so that the mixer noise is
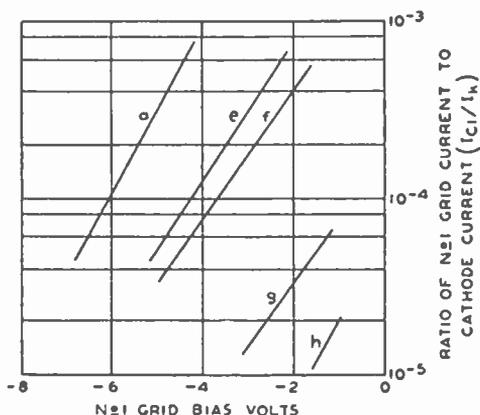

Fig. 21—Signal-grid (grid No. 1) current of the outer-grid injection mixers shown in Fig. 20. Curve a corresponds to the original design (a) of Fig. 18 and is shown for comparison. Data taken with a 20-megacycle oscillator voltage of 12 volts peak amplitude applied to the oscillator-grid. $E_{c3} = -10$ volts, E_{c2} and $_4 = 100$ volts, $E_b = 250$ volts.

given by the average of the distribution fluctuations over the oscillator cycle. In terms of the equivalent noise resistance the average has been found to be[23]

$$R_{eq} = \frac{20 \left[\overline{I_b} - \dfrac{\overline{I_b^2}}{I_a} \right]}{g_c^2}$$

where $\overline{I_b}$ is the average (i.e., the operating) plate current and $\overline{I_b^2}$ is the average of the square of the plate current over an oscillator cycle. I_a is the cathode current of the mixer section and is substantially constant over the oscillator cycle. This relation is not very useful in the form given. It is usually sufficiently accurate for most purposes to use an expression identical with that which applies to tubes with inner-grid oscillator injection, namely,

$$R_{eq} = \frac{20 \overline{I_b}}{g_c^2} F^2,$$

where F^2 is about 0.5 for tubes with suppressor grids and somewhat higher for others. By assuming a typical tube characteristic, the noise resistance may be expressed in terms of the cathode current I_a of the mixer section and the maximum signal-grid-to-plate transconductance g_{max} as

$$R_{eq} = 120 \frac{I_a}{(g_{max})^2}$$

for operation at oscillator fundamental. For operation at second or third harmonic of the oscillator, the noise resistance will be approximately doubled, or tripled, respectively.

V. Conclusion

It has been shown that the principle of frequency conversion in all types of tubes and with all methods of operation may be considered as the same (i.e., as a small-percentage amplitude modulation). The differences in other characteristics between various tubes and methods of operation are so marked, however, that each application must be considered as a separate problem. The type of tube and method of operation must be intelligently chosen to meet the most important

needs of the application. In making such a choice, it is frequently of assistance to prepare a table comparing types of tubes and methods of operation on the basis of performance data. An attempt has been made to draw such a comparison in a qualitative way for general cases and for a few of the important characteristics. Table II is the result. It must be understood, of course, that the appraisals are largely a matter of opinion based on experience and the present state of knowledge. Furthermore, in particular circuits and with particular tubes, the relative standings may sometimes be quite different. A study of the fundamentals brought out in the previous sections of this paper should help in evaluating such exceptions.

Table II—Approximate Comparative Appraisals of Methods of Frequency Conversion

Desirable Characteristic	Oscillator and Signal Voltages on No. 1 Grid		Oscillator Voltage on No. 3 Grid, Signal on No. 1 Grid		Oscillator Voltage on No. 1 Grid, Signal on No. 3 Grid
	Triode	Pentode	Pentode	Hexode or Heptode	Hexode or Heptode
High conversion transconductance	Good	Good	Fair	Fair	Fair
High plate resistance	Poor	Good	Poor	Good	Good
High signal-to-noise ratio	Good	Good	Poor	Poor	Poor
Low oscillator-signal circuit interaction and radiation	Poor	Poor	Good	Good	Fair
Low input conductance at high frequencies	Poor[1]	Fair	Poor	Poor[2]	Good
Low signal-grid current at high frequencies	Good	Good	Poor	Poor[2]	Fair
Low cost of complete converter system	Good	Fair	Fair	Poor	Good

[1] Due to feedback; may be increased to Fair by proper circuit design.

[2] May be increased to Fair by special constructions as described in text.

APPENDIX

Discussion of Negative Admittance of Current-Limited Grids

In Figs. 10, 11, and 12 it was seen that the electronic signal-grid (i.e., input) admittance components (i.e., the admittance due to the presence of electrons) of a mixer designed for No. 1 grid injection of the oscillator are negative over a considerable portion of the normal operating range. Figures 10 and 11, however, were taken with static voltages applied and so indicate that the phenomenon is not caused by an alternating oscillator voltage but is associated with the characteristics of the tube itself.

The input admittance of negative grids in vacuum tubes is the sum of three factors: (1) the "cold" admittance, or the admittance of the tube with the electron current cut off; (2) the admittance due to feedback from other electrodes through tube and external capacitance, etc.; and (3) the admittance due to the presence of the electrons in the tube. The first two factors have been well known for many years although certain aspects of the second have only recently received attention.[29,30,48] The third factor, however, is not so well understood although the excellent work done during the last ten years has paved the way for a complete understanding of the subject.[49] The present discussion is concerned only with this last point, namely the admittance of negative grids due to the presence of electrons in the tube.

Early work on transit-time effects in diodes and negative-grid triodes had indicated that, at very high frequencies, the conductance became negative in certain discrete bands (i.e., at large transit angles). It was not, at first, appreciated that conditions were possible with negative-grid triodes in which the input conductance could become negative even at low frequencies (i.e., at small transit angles). Data taken on the input (No. 4 grid) conductance of pentagrid converters by W. R. Ferris of this laboratory during 1934 showed that these tubes had a negative input conductance which varied as the square of the frequency and which remained negative at low frequencies. The conductance appeared, therefore, to behave in the same way as the positive input conductance of ordinary negative-grid tubes, except for a reversal in sign. The data on the pentagrid were taken with an external oscillator voltage applied to the No. 1 grid. The work of Bakkar and de

---

[48] M. J. O. Strutt and A. van der Ziel, "The causes for the increase of admittances of modern high-frequency amplifier tubes on short waves," *Proc. I.R.E.*, Vol. 26, pp. 1011-1032; August, 1938.

[49] An excellent historical summary of this work is found in W. E. Benham, "A contribution to tube and amplifier theory," *Proc. I.R.E.*, Vol. 26, pp. 1093-1170; September, 1938.

Vries[50] disclosed the possibility of a negative input conductance at small transit angles in a triode operated under current-limited conditions. They gave an experimental confirmation for a triode operated at reduced filament temperature. Data taken by the writer during 1936 on a pentagrid converter showed that the negative conductance was present in this tube even when direct voltages, only, were applied and that it was accompanied by a reduction in capacitance. A fairly complete theory of the effect was developed in unpublished work by Bernard Salzberg, formerly of this laboratory, who extended the theory of Bakker and de Vries to the more general case of multigrid tubes with negative controls in a current-limited region. Other experimental work was done on the effect during 1936 by J. M. Miller and during the first half of 1937 by the writer. In the meantime, the papers of H. Rothe,[51] I. Runge,[52,53] and L. C. Peterson[54] showed that independent experimental and theoretical work had been done on the negative-admittance effect in other laboratories.

In a rough way, the negative admittance found under current-limited conditions may be explained as follows: The electron current in a tube is equal to the product of the charge density and the electron velocity. If this current is held constant, a rise in effective potential of the control electrode raises the velocity and so lowers the charge density. A reduction in charge density with increase in potential, however, results in a reduction in capacitance, provided no electrons are caught by the grid. Thus, the susceptive component of the part of the admittance due to the current through the grid, is negative. Because of the time lag due to the finite time of transit of the electrons, there is an additional component of admittance lagging the negative susceptance by 90 degrees, i.e., a negative conductance. The value of the negative conductance will be proportional to both the transit angle and to the value of the susceptance. Since both of these quantities are proportional to frequency, the negative conductance is proportional to the square of the frequency.

[50] C. H. Bakker and C. de Vries, "On vacuum tube electronics," *Physica*, Vol. 2, pp. 683-697; July, 1935.

[51] H. Rothe, "The operation of electron tubes at high frequencies," *Telefunken-Röhre*, No. 9, pp. 33-65; April, 1937; *Proc. I.R.E.*, Vol. 28, pp. 325-332; July, 1940.

[52] I. Runge, "Transit-time effects in electron tubes," *Zeit. für Tech. Phys.*, Vol. 18, pp. 438-441; 1937.

[53] I. Runge, "Multigrid tubes at high frequencies," *Telefunken-Röhre*, No. 10, pp. 128-142; August, 1937.

[54] L. C. Peterson, "Impedance properties of electron streams," *Bell. Sys. Tech. Jour.*, Vol. 18, pp. 465-481; July, 1939.

The general shape of the curves of Fig. 10 may be explained as follows: At a No. 1 grid bias of about —20 volts, the cathode current is cut off and the electronic admittance is zero. At slightly less negative values of No. 1 grid bias, the electron current is too small to build up an appreciable space charge ahead of the signal grid (No. 4 grid). The latter grid, although it exhibits some control of the plate current does not control the major portion of the current reaching it and is thus in a substantially current-limited region. Its susceptance and conductance are, therefore, negative. Higher currents increase the negative admittance until at some value of No. 1 grid bias, the electron current is increased to the point at which a virtual cathode is formed in front of some parts of the signal grid. At these parts, the current which reaches the grid is no longer independent of this grid potential and, as a result, a positive susceptance and conductance begin to counteract the negative admittance of other portions of the grid. The admittance curves reach a minimum and for still higher currents approach and attain a positive value. The current necessary to attain the minimum admittance point is less when the signal-grid bias is made more negative so that the minima for increasingly negative No. 4 grid-bias values occur at increasingly negative No. 1 grid-bias values.

It may be noted that the signal-grid-to-plate transconductance is at a maximum in the region just to the right of the admittance minima of Fig. 10 (compare Fig. 8). The admittance of such a tube used as an amplifier remains negative, therefore, at the maximum amplification point.

BEAM-DEFLECTION CONTROL FOR
AMPLIFIER TUBES*†

By

G. R. Kilgore

formerly with
Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—It is the purpose of this paper: (1) to discuss the basic principles involved in obtaining high transconductance and high ratios of transconductance to current by means of deflection control; (2) to derive expressions for the ultimate transconductance at both low and high frequencies; (3) to discuss by means of elementary electron optics the design of a simple, beam-deflection gun for obtaining the desired results; and (4) to describe some of the early experimental results on amplifier tubes combining beam-deflection control and a multi-stage secondary-emission multiplier.

It is shown that deflection control offers a possibility of obtaining substantial transconductance with low capacitance and low beam currents and with a very high ratio of transconductance to plate current. It is found experimentally that useful values of transconductance with low capacitance and low current can be obtained with a simple deflection gun combining focusing and deflection. This type of control is ideally suited for use with a high gain secondary emission multiplier to obtain very high transconductance, without excessive capacitance, thus making possible a tube with a bandwidth figure of merit many times greater than for conventional tubes.

Experimental confirmation of some of the properties of deflection control in agreement with the analysis has already been obtained in experimental tubes.

Introduction

AFTER the successful demonstration of the secondary-emission multiplier in 1935[1] for amplifying photoelectric currents, it was natural that attempts should be made to obtain a high-transconductance, voltage-controlled amplifier with low input current. Efforts to produce such a tube with conventional grid control on the input soon indicated a basic limitation set by the ratio of transconductance to plate current. The maximum possible value of this ratio for grid control has been shown[2] to be e/kT. Thus, for ordinary

---

[1] V. K. Zworykin, G. A. Morton and L. Malter, "The Secondary Emission Multiplier—A New Electronic Device," *Proc. I.R.E.*, Vol. 24, pp. 351-376, March, 1936.
[2] F. Below, "Theory of Space-Charge-Grid Tubes," *Zeit. fur Fernmeldtechnik*, Vol. 9, pp. 113-118 and 136-142, 1928.

cathodes ($T \cong 1000$ degrees Kelvin) the maximum theoretical ratio cannot exceed about 10 to 12 micromhos per microampere and, in practice, is seldom more than one or two micromhos per microampere. This meant that high transconductance could be obtained only with high current on the output stage of the multiplier and this current limit was set by dissipation and space charge, as in a conventional tube. It was clear that a new means of control was needed to give higher ratios of transconductance to plate current.[3] Deflection control appeared as one of the possibilities and a decision to explore this method was somewhat influenced by a privately communicated report on the work of Nagashima.[4] He described a complex deflection tube in which he obtained a transconductance to current ratio of 100. Although his beam current was on the order of 10^{-3} amperes and the transconductance was only 0.5 micromhos, his experiments did indicate that deflection control offered a solution to the voltage-controlled multiplier for high transconductance. Other laboratories also appreciated the advantages of deflection control for multiplier tubes[5,6,7] but the absence of further publication makes the ultimate results obtained somewhat in doubt.

Well before the war, and at the suggestion of B. J. Thompson, an investigation was begun of beam-deflection control in combination with secondary-emission multiplication with the objective of obtaining a useful device for amplification and detection of ultra-high frequencies. A particular objective of the early work was to produce a high-transconductance tube with low capacitance for wide-band amplification at ultra-high frequencies. Work on deflection control, subsequent to that covered in this paper, was continued throughout the war in these laboratories and in other divisions of RCA. Some of this later work was sponsored by the Armed Services.

The present paper covers only the early work, namely: (1) The basic principles involved in obtaining high transconductance and high ratios of transconductance to current by means of deflection control;

[3] B. J. Thompson, "Voltage-Controlled Electron Multipliers," *Proc. I.R.E.*, Vol. 29, pp. 583-587, November, 1941.

[4] M. Nagashima, "Voltage Multipliers," a paper presented at combined Engineering Convention at Tokyo, April, 1938.

[5] W. Flechsig and M. Sandhagen, "Electronic Amplifier Valves with Secondary Emission Multiplication," *Fernseh Hausmitt*, Vol. 2, pp. 16-25, May, 1940.

[6] F. M. Colebrook, "Ultra-Short and Decimeter-Wave Valves—Deflection of a Focused Beam as a Possible Basis for Construction," *Wireless Eng.*, Vol. 15, pp. 198-201, April, 1938.

[7] J. H. O. Harries, "Discussion of Colebrook Paper," *Wireless Eng.*, Vol. 15, pp. 324-325, June, 1938.

(2) expressions for the ultimate transconductance at both low and high frequencies; (3) elementary electron optics and the design of a simple beam-deflection gun for obtaining the desired results; and (4) some of the early experimental results on amplifier tubes combining beam-deflection control and a multi-stage secondary-emission multiplier.

BASIC PRINCIPLES

Beam-deflection devices for use as detectors, amplifiers, and oscillators have been described in the literature[8] by many workers dating back at least to 1906, but none of these apparently had any remarkable success. Most of the earlier workers failed because they attempted to use high-current beams with consequent space-charge difficulties and because they apparently did not appreciate the basic current-density limitation that determines transconductance.

At the beginning of the present work it was decided to use very small currents in the deflection gun and to build up the current by secondary-emission multiplication. It was recognized that high current density and not high beam current was essential in obtaining high transconductance. A very important concept for electron beams was worked out by D. B. Langmuir,[9] who pointed out that the ultimate current density that could be brought to a focus was limited by thermal velocity of the electrons and was determined by the beam voltage, the cathode current density and temperature, and by the angle of convergence of the electron beam. While in his published work Langmuir treated only the point focus case, he had extended his work to the line focus case and derived an expression very similar to that later published by J. R. Pierce.[10] Combining this maximum current density with maximum deflection sensitivity, an expression for maximum low-frequency transconductance was found which gave a basis of comparison for the later work in evolving a high transconductance gun for high frequencies. This maximum low-frequency transconductance is worked out in the next section and extended to high frequencies in the section following that.

The concept of the current density limit is so basic to the beam-deflection work that it is worth while to review the concepts in terms

[8] The earliest work is usually credited to R. Von Lieben, German Patent 179,807 (1906), and M. Dieckmann and G. Glage, German Patent 184,710 (1906).

[9] D. B. Langmuir, "Theoretical Limitations of Cathode-Ray Tubes," *Proc. I.R.E.*, Vol. 25, pp. 977-992, August, 1937.

[10] J. R. Pierce, "Limiting Current Densities in Electron Beams," *Jour. Appl. Phys.*, Vol. 10, pp. 715-723, October, 1939.

of an elementary physical picture even though Langmuir and Pierce have covered the subject quite thoroughly. Their work shows that the maximum current density that can be obtained in a line focus is approximately

$$j_{max} \cong j_0 \frac{2}{\pi^{1/2}} \left(\frac{Ve}{kT} \right)^{1/2} \sin \theta \qquad (1)$$

where

$j_0 =$ cathode current density

$V =$ beam potential in volts

$$e/kT = \frac{\text{electronic charge}}{\text{Boltzmann constant} \times T} = \frac{11,600}{T} \text{ volts}^{-1}$$

$T =$ temperature in degrees Kelvin

$\theta =$ angle of convergence of beam
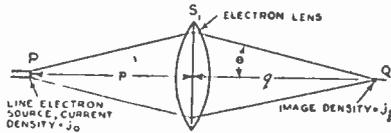
Fig. 1—Simplified picture of an electron-optical system.

Pierce's analysis shows that this theoretical maximum density requires infinite demagnification. He also shows that this expression is in error less than 10 per cent for values of $V > 0.5$ volt.

A good physical picture of the relation can be had by the use of elementary electron optics referring to Figure 1. Consider a source of electrons, P, such as a line cathode, perpendicular to plane of figure, emitting electrons which are focused at an image Q by a cylindrical lens S. If the current density of the object, P, is j_0 then in a perfect optical system, in which all the electrons leaving P reach Q, the current density at Q will be

$$j_q = \frac{j_0}{M} \qquad (2)$$

where M is the magnification ratio which is roughly equal to $\dfrac{q}{p}$. It might seem from this elementary picture that the current density at Q could be made as large as desired by making the magnification smaller and smaller by increasing p. The electrons, however, due to

their thermal velocities will spread out in every direction from the cathode and in any finite system the current reaching the image will be limited by the lens, by some aperture stop, or, in the case of beam-deflection tubes, possibly by the deflection plates themselves. Thus, as the object distance p is increased the beam will spread out more and more and a larger fraction of the electrons will be caught at S_1, and fewer electrons will get through to the image Q, thus compensating for the increased demagnification. As a result, after increasing p beyond a certain point, the current density j_0 might be expected to remain constant. A fairly good quantitative picture can be had by approximating the true emission condition by assuming that electrons are emitted from the cathode uniformly in all directions with a kinetic energy corresponding to a voltage $V_0 = \dfrac{kT}{e}$.
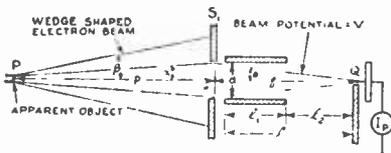
Fig. 2—Cross-sectional view of a beam-deflection tube showing reasons for loss of current density due to a lens stop and finite deflection plate spacing.
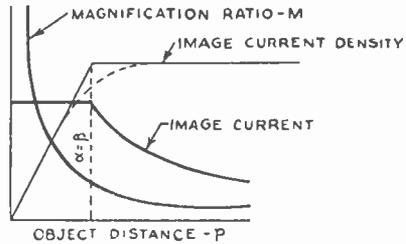

Fig. 3—Effect of varying object distance for an idealized tube.

This means the electrons will spread out in a wedge-shaped homogeneous beam, and the half-angle of spread β will be nearly equal to the ratio of initial velocity to the beam velocity or

$$\beta = \left(\frac{V_0}{V} \right)^{1/2} = \left(\frac{kT}{Ve} \right)^{1/2} \tag{3}$$

Referring to a simple beam-deflection tube as shown in Figure 2, the beam spread half-angle, β, may be greater or less than the entrance angle, α, which is determined by the object distance, p, and the aperture width, d. When β is less than α, all of the current gets through to Q and

$$j_q = \frac{j_0}{M} = \frac{j_0 \, p}{q} \tag{4}$$

If p is increased to the point where β just equals α, the current density will reach the maximum value and at this point $\dfrac{p}{q} = \dfrac{\theta}{\beta}$, so

$$j_{\max} = \frac{j_0 \, \theta}{\beta} = j_0 \left(\frac{Ve}{kT} \right)^{1/2} \theta \tag{5}$$

This expression differs from the more exact approximation, (1), only by a factor $2/\pi^{1/2}$ and that θ replaces $\sin \theta$. Following the elementary picture a little farther it is seen that, as the object distance is increased indefinitely, the fraction of current intercepted will just compensate for the magnification and the image current density will remain at a constant level while the beam current passing thru the deflection plates grows smaller and smaller. This is illustrated in Figure 3 which shows current, current density, and magnification ratio as a function of object distance.

To understand better the role of magnification factor it is well to look at the more general expression for current density in a line focus given by J. R. Pierce[10].

$$j = \frac{j_0}{M} \, \mathrm{erf} \, \left[\frac{M^2 \left(\dfrac{Ve}{kT} \right) \sin^2\theta}{1 - M^2\sin^2\theta} \right]^{1/2} \tag{6}$$

which assumes only that $Ve/kT \gg 1$, as in most practical cases. This expression, when expanded for very small values of $M \sin \theta$, gives

$$j = j_0 \frac{2}{\pi^{1/2}} \left(\frac{Ve}{kT} \right)^{1/2} \sin \theta \left[1 - \frac{(M \sin \theta)^2 \left(\dfrac{Ve}{kT} \right)}{3} \right.$$

$$\left. + \frac{(M \sin \theta)^4 \left(\dfrac{Ve}{kT} \right)^2}{10} \cdots \cdots \right] \tag{7}$$

which reduces to Equation (1) for the case of $M = 0$. Inspection of Equation (7) shows that it is not necessary to go to extremely small values of M to approach the maximum when θ is small, as is usually the case. For example, about 75 per cent of the ultimate current

density will be obtained if $(M \sin \theta)^2 \left(\dfrac{Ve}{kT} \right) = 1$

or $$M \lesseqgtr \left(\frac{kT}{Ve} \right)^{1/2} \frac{1}{\sin \theta} \qquad (8)$$

For a practical beam-deflection tube as an example $V = 300$ volts, $\dfrac{e}{kT} = 10$, $\sin \theta \sim \theta = \dfrac{1}{55}$, $M = 1$, or a one-to-one magnification ratio is sufficiently small to give 75 per cent of the maximum current density.

Although decreasing the magnification ratio beyond the above value does not materially increase the current density, it does reduce the total current in the image as was shown in the elementary picture of Figure 3. Except for aberrations, the current can be made as small as we please without affecting current density by making the magnification smaller and smaller. This indicates the possibility of using very low beam currents without loss of transconductance, which is an important consideration in connection with input loading and noise.

The total current reaching the image, Q, will depend on the total cathode current $I_c = j_o W_c h_c$ as defined below, and the amount intercepted at the stop, S_1 which determines the entrance angle α. An expression for this current can be derived from the density formula, recalling that $M \cong \dfrac{\alpha}{\theta}$.

$$I_{\text{image}} = j_o \, W_c \, h_c \, \frac{2}{\pi^{1/2}} \, \alpha \left(\frac{Ve}{kT} \right)^{1/2} \left[1 - \frac{\alpha^2 \left(\dfrac{Ve}{kT} \right)^{1/2}}{3} + \cdots \right] \qquad (9)$$

$$= I_c \frac{2}{\pi^{1/2}} \, \alpha \left(\frac{Ve}{kT} \right)^{1/2} \left[1 - \frac{\alpha^2 \left(\dfrac{Ve}{kT} \right)^{1/2}}{3} + \cdots \right] \qquad (10)$$

where W_c and h_c are width and height of cathode.

MAXIMUM LOW-FREQUENCY TRANSCONDUCTANCE

The maximum transconductance obtainable from an electrostatically focused beam-deflection gun at low frequencies, where transit

time is short compared to one period of the deflecting potential, can be derived by combining the above expression for maximum current density with the expression for deflection sensitivity. Consider a rectangular focused beam which is deflected by a pair of deflection plates past an intercepting edge, as illustrated in Figures 2 and 4. If the beam has a current density j_1 at the intercepting edge, then when the beam is deflected a distance $\triangle x$, the change in output current will be

$$\triangle I_p = j_1 W \triangle x$$

where W is the beam width as shown in Figure 4. If the deflection $\triangle x$ is a result of an applied incremental voltage $\triangle E$ to the deflection plates, then, dividing the equation by $\triangle E$ gives $\dfrac{\triangle I_b}{\triangle E} = j_1 W \dfrac{\triangle x}{\triangle E}$.
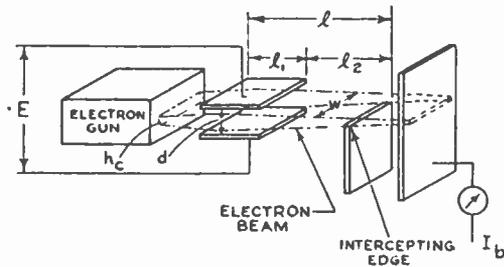
Fig. 4—Illustrative beam-deflection tube.

In the limit, where the changes are small, $\triangle I/\triangle E$ is the transconductance, g_m, and $\triangle x/\triangle E$ is the deflection sensitivity, S.

Thus, the transconductance is

$$g_m = j_1 W S \tag{11}$$

As was seen before, the maximum value of current density, j_1, in the line focus will be,

$$j_{max} = j_o \frac{2}{\pi^{1/2}} \left(\frac{Ve}{kT} \right)^{1/2} \sin \theta \tag{12}$$

Now, in a beam-deflection tube, the maximum angle of convergence, θ, is determined by the deflection plate spacing and the distance, l, from the entrance to the deflection plates to the intercepting edge.

For small values of θ, $\sin \theta \simeq \dfrac{d}{2l}$ so,

$$j_{max} = j_o \frac{1}{\pi^{1/2}} \left(\frac{Ve}{kT} \right)^{1/2} \frac{d}{l} \tag{13}$$

(In this analysis, it is assumed that the potential along the beam is constant and that there is no focusing after deflection.)

The deflection sensitivity at low frequencies is given by the standard formula

$$S = \frac{1}{4d\ V} (l_1{}^2 + 2l_1 l_2) \tag{14}$$

where $l_1 =$ length of deflection plates

$l_2 =$ distance from end of deflection plates to the intercepting edge

For a given beam length, it is evident that the maximum low-frequency deflection sensitivity will be obtained when the deflection plates extend the whole length of the focused beam in which case,

$$S_{max} = \frac{l^2}{4d\ V} \tag{15}$$

Combining Equations (13) and (15), the upper limit of low-frequency transconductance is,

$$g_{max} = j_o \frac{1}{4\pi^{1/2}} \left(\frac{e}{kT} \right)^{1/2} \frac{lW}{V^{1/2}} \text{ mhos} \tag{16}$$

Several interesting facts appear from this expression.

1. The maximum obtainable transconductance is independent of the deflection plate spacing. This suggests that a deflection gun may be built with large spacing and low capacity.

2. The maximum transconductance is proportional to cathode current density and is *independent* of *the total beam current*. This suggests again the possibility of high transconductance to current ratios.

3. Transconductance is proportional to the transit time through the deflection-plate space. Theoretically, therefore, there is no limit to transconductance as the transit angle is made longer. Practically, of course, there are several limits such as mechanical alignment, stray magnetic fields and surface effects.

An example of the ultimate theoretical transconductance for a tube of reasonable dimensions and potentials is as follows:

$$j_o = 0.100 \text{ amperes per square centimeter;}$$

$$T = 1160 \text{ degrees Kelvin,} \qquad \frac{e}{kT} = 10;$$

$$l = 3 \text{ centimeters,} \qquad W = 1 \text{ centimeter;}$$

$$V = 100 \text{ volts;} \qquad g_m = 14,000 \text{ micromhos.}$$

All the above was done without the consideration of space-charge limitations. This is justifiable since space-charge effects depend on the *total beam current* which theoretically can be made as small as we please without affecting the transconductance. Practically, the beam current will depend on physical size of the cathode, or the size of the first aperture and on the maximum demagnification. All of the analysis and experimental work included in this report was directed toward the low-current case where space-charge effects are small.

J. R. Pierce[11] has given an expression for the limiting transconductance of a deflection tube expressed in terms of the capacitance of the deflection plates. An examination of his expression shows that it is equivalent to Equation (16), although in a somewhat less convenient form for usual design purposes. The introduction of capacitance in the low-frequency case is not so important as in the high-frequency case where bandwidth considerations are important. The maximum ratio of transconductance to capacitance is derived in the next section for the high-frequency case.

The expression for maximum transconductance derived above was for a cylindrical-lens focusing system. It might appear possible to obtain still higher transconductance by the use of a spherical lens system where the beam can be compressed in width as well as thickness. This case is worked out in Appendix I and the value found to be only one-half the value given by Equation (16) above.

MAXIMUM HIGH-FREQUENCY TRANSCONDUCTANCE

At frequencies where the transit angle through the deflection plates reaches an appreciable fraction of the period, the force on the

---

[11] J. R. Pierce, "Theoretical Limitation to Transconductance in Certain Types of Vacuum Tubes," *Proc. I.R.E.*, Vol. 31, pp. 657-663, December, 1943.

electron as it moves through the deflecting plates is not constant and the deflection sensitivity will be less than the low-frequency case.[12] As an example, the ratio of high frequency to direct-current sensitivity is shown in Figure 5 as a function of transit angle, θ_1, through the deflection plates for two special cases. In the general case, the deflection sensitivity is made up of two components, one due to displacement in the deflection plate field and the other due to the transverse drift after the electron leaves the deflection plates. When transit angles are small, these components are in phase but, at high frequencies, these components vary in both magnitude and in phase as the transit angles are changed. General expressions for the magnitude of the deflection sensitivity have been given by a number of writers.[13] A convenient form for the present analysis is,

TRANSIT ANGLE THROUGH DEFLECTION PLATES- θ_1
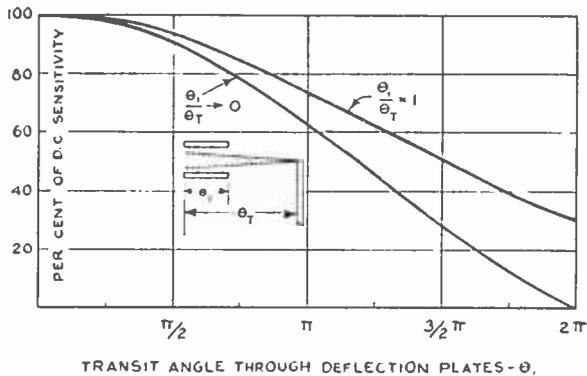
Fig. 5—Loss of deflection sensitivity with increasing frequency. (Curves plotted in terms of transit angle, $\omega\tau$.)

$$S = \frac{1}{\omega^2 d}\left(\frac{e}{m}\right)\sqrt{[(1-\cos\theta_1)+\theta_2\sin\theta_1]^2 + [(\theta_1-\sin\theta_1)+\theta_2(1-\cos\theta_1)]^2}$$

centimeters per volt (17)

where $\theta_1 = \omega\tau_1 =$ transit angle through the deflection field,

$\theta_2 = \omega\tau_2 =$ transit angle after deflection,

[12] J. T. MacGregor-Morris, "Measurements in Electrical Engineering by Means of Cathode Rays," *Jour. Inst. Elec. Eng. (Brit.)*, Vol. 63, p. 1098, 1925.

[13] L. Malter, "Deflection and Impedance of Electron Beams at High Frequencies in the Presence of Magnetic Fields," *RCA REVIEW*, Vol. V, No. 4, pp. 439-454, April, 1941. Also M. R. Gavin and G. W. Warren, "Deflected-Beam Valves for Ultra-High-Frequencies," *G.E.C. Jour.* Vol. 14, pp. 97-103, August, 1946.

$\omega = 2\pi f$ radians per second,

$d =$ deflection plate separation in centimeters,

$\dfrac{e}{m} =$ ratio of electronic charge to mass.

If the transit angle, θ_1, is varied while the total transit angle, θ_T, is held constant, the deflection sensitivity will, in general, go through one or more maxima as shown in Figure 6 where a deflection sensitivity factor is plotted for several values of total transit angle.* It is clear that the maxima are at $\theta_1 = \pi$, 3π, etc. as is found by maxi-
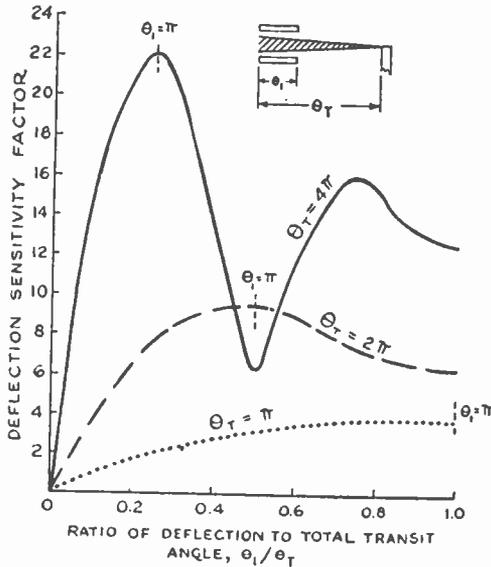
Fig. 6—Curves showing change in deflection sensitivity with length of deflection plates, holding overall beam length (and current density) a constant. (This figure shows maximum transconductance always occurs at a transit angle of π.)

mizing Equation (17). Further inspection shows that in each case the absolute maxima occur for $\theta_1 = \pi$.

Substituting $\theta_1 = \pi$ in Equation (17), the maximum deflection sensitivity is found to be,

$$S_{\max} = \frac{1}{\omega^2 d}\left(\frac{e}{m}\right)\sqrt{4 + (\pi + 2\,\theta_2)^2} \qquad \text{centimeters per volt} \qquad (18)$$

* This presentation is due to D. O. North.

It should be noted that even though the high-frequency sensitivity falls off compared to the direct-current sensitivity, that the actual high-frequency sensitivity increases indefinitely as θ_2 is increased. For values of $\theta_2 \gg \pi$, the sensitivity varies nearly in proportion to θ_2.

The high-frequency transconductance can now be had by combining Equation (18) with the expression for maximum current density given by Equation (1). Maximum transconductance for a finite transit angle θ_2 is very nearly equal to

$$g_m = \frac{j_0 W}{\pi^{1/2}} \left(\frac{e}{kT} \right)^{1/2} \frac{V^{1/2}}{\omega^2 l} \left(\frac{e}{m} \right) \sqrt{4 + (\pi + 2\,\theta_2)^2} \qquad (19)$$

where $\qquad\qquad W =$ the beam width, $\qquad\qquad V =$ beam voltage,

$l =$ total beam length from entrance to deflection plates to intercepting edge.

This can be put in better form by recognizing that,

$$\frac{\omega l}{\sqrt{\dfrac{2e}{m} V}} = \omega\tau_1 + \omega\tau_2 = \pi + \theta_2$$

then,

$$g_m = \frac{j_0 W}{2\omega\pi^{1/2}} \left(\frac{2e}{m} \right)^{1/2} \left(\frac{e}{kT} \right)^{1/2} \left[\frac{\sqrt{4 + (\pi + 2\,\theta_2)^2}}{\theta_2 + \pi} \right] \qquad (20)$$

As θ_2 is increased, the bracketed expression in Equation (20) rapidly approaches the value 2, as shown in Figure 7. For example, for $\theta = 4\pi$ the g_m is up to 90 per cent of the maximum value. The ultimate limit for high-frequency transconductance is therefore given by,

$$g_m \Big]_{max} \cong \frac{1}{\pi^{1/2}} \left(\frac{e}{kT} \right)^{1/2} \left(\frac{2e}{m} \right)^{1/2} \frac{j_0 W}{\omega}$$

$$\cong 5.4\, j_0 \left(\frac{e}{kT} \right)^{1/2} \frac{10^6}{f} W \qquad \text{mhos} \qquad (21)$$

For the usual oxide-coated cathode where T is roughly 1160 degrees Kelvin, the expression for maximum high-frequency transconductance reduces to the simple relation

$$g_m \Big]_{max} = \frac{18}{f_{(in\ megacycles)}} j_o W \qquad \text{mhos} \qquad (22)$$

RATIO OF HIGH-FREQUENCY TRANSCONDUCTANCE
TO INPUT CAPACITANCE

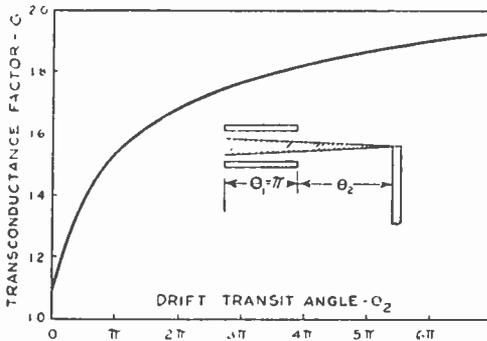One of the important figures of merit for wide-band, high-frequency tubes is the ratio of transconductance to capacitance.[14]

Fig. 7—Curve showing overall transconductance factor as a function of drift distance, assuming optimum deflection plate length.

This ratio is, in fact, proportional to the product of the voltage gain and bandwidth. Having derived an expression for maximum high-frequency transconductance it is interesting to relate this to the capacitance to obtain the bandwidth figure of merit. The case of the rectangular section beam with parallel-plane deflection plates will be considered as it is the one of most practical value. The capacitance between deflection plates will be considered as the only significant one since it is related to the factors affecting transconductance and is likely to be large compared to the output capacitance.

The capacitance between parallel-plane deflection plates is

$$C = a \frac{wl}{d} \times 10^{-12} \text{ farads} \qquad (23)$$

---

[14] A. P. Kauzmann, "New Television Amplifier Receiving Tubes," *RCA REVIEW*, Vol. III, No. 3, pp. 271-289, January, 1939.

where l = length of plates, d = separation of plates,

w = width of plates $>> d$,

a = a factor depending on the ratio of $\dfrac{l_1}{d}$

which is 0.0885 for $\dfrac{l_1}{d} >> 1$.

We have seen in the two preceding sections that the transconductance is apparently independent of the deflection plate separation, and thus it might appear that the capacitance could be made as small as we please by increasing the separation. This cannot be realized because of the effect of fringe fields which were neglected in the earlier work where it was tacitly assumed that $\dfrac{l_1}{d} >> 1$. In a practical deflection tube at high frequencies, where it is necessary to keep the deflection plates short, the fringe field is a limiting factor. A practical evaluation of this effect leads to a choice of .the upper limit of the ratio of spacing to length of about one-half. Using this ratio the capacitance is

$$C_{\text{min}} = 0.28\ w\ 10^{-12} \text{ farads} \tag{24}$$

It should be noted that this minimum capacitance is *independent of frequency*. The length of deflection plates, of course, will depend upon the frequency and beam voltage in order to maintain the optimum transconductance condition $\theta_1 = \pi$. In going to higher frequencies the length of deflection plates and spacing will decrease together, the capacitance remaining at a constant value given by Equation (24).

The ultimate transconductance-to-capacitance ratio for a high-frequency deflection tube using a rectangular beam is given by combining Equation (24) with Equation (21) assuming the width of beam W equal to the width of the plates, w.

$$\frac{g_m}{C} = 19 \times 10^{18} \left(\frac{e}{kT} \right)^{1/2} \frac{j_o}{f} \text{ mhos per farad} \tag{25}$$

An upper limit for the voltage gain times bandwidth can now be computed neglecting all capacitances except the deflection-plate capacitance,

$$B_{\text{max}} = \frac{g_m}{2\pi C} \cong 3 \times 10^{18} \left(\frac{e}{kT} \right)^{1/2} \frac{j_o}{f} \tag{26}$$

It is interesting to work out a practical case to compare with conventional tubes,

$$f = 300 \text{ Mc} = 3 \times 10^8 \text{ cycles/second,}$$

$$j_o = 0.100 \text{ amperes per square centimeter,}$$

$$\frac{e}{kT} = 10 \ (T = 1160 \text{ degrees Kelvin}),$$

$$B = 3150 \text{ megacycles.}$$

Although this bandwidth merit is an order of magnitude greater than for conventional tubes, it is only fair to point out that this figure could never be reached in practice because of the added capacitance due to the output electrode and leads and because of practical factors limiting the attainment of the maximum transconductance such as will be pointed out in a later section. It should also be noted that this figure of merit decreases inversely with frequency. It is probable that in actual practice that the bandwidth merit of the deflection gun itself may not be much different from that for conventional control. However, because of high transconductance-to-current possibility the deflection gun can be combined with secondary-emission multiplication to give a bandwidth merit much greater than for conventional tubes. Examples of this will be given in next to last section.

DEVELOPMENT OF A HIGH TRANSCONDUCTANCE DEFLECTION GUN FOR HIGH FREQUENCIES

In order that deflection control could be successfully used in high-frequency tubes it was necessary to evolve a simple deflection gun with the following qualities: (1) high transconductance; (2) high ratio of transconductance to capacitance; (3) high ratio of transconductance to plate current and, in some cases, low beam current.

The approach to these objectives was suggested by the analysis presented in the foregoing sections on the limiting transconductance and transconductance-to-capacitance ratio. From such relations it was possible to determine readily some of the design factors such as length of deflection plates, separation of deflection plates and general proportions of the tube without much regard for the type of focusing system used. Computation of these factors will be illustrated before considering the various focusing systems and the choice of a particular scheme. It is not the purpose of this discussion to set up an exact design procedure but rather to indicate in an approximate

quantitative fashion the factors in evolving the gun structure which formed the prototype of later guns.

The length of the deflection plates is a good starting point in the design since the transit angle through the plates should be approximately one-half period as shown in section C. The length of plates for a given frequency and voltage is,

$$l_1 = \frac{1}{2f} \left(\frac{2Ve}{m} \right)^{1/2} = \frac{3 \times 10^7}{f} V^{1/2} \tag{27}$$

where
$f =$ frequency in cycles per second,
$V =$ beam potential in volts,
$l_1 =$ length of deflection plates in centimeters.

The deflection plate separation, as pointed out in the preceding section, should be increased to about one-half the plate length in order to obtain the best transconductance-to-capacitance ratio (neglecting all other capacitances). Since the transconductance and capacitance are both independent of beam potential V, it is of advantage to use higher potentials to give a larger plate separation thus minimizing surface effects and mechanical difficulties. In receiving tubes it is desirable to keep potentials on the order of 100 volts. Assuming $V = 100$, the dimensions for $f = 3000$ megacycles would be $l_1 = 0.10$ centimeter, $d = 0.05$ centimeter.

It has been shown in Figure 7 that the length of beam from the end of the deflection plates to the intercepting edge need not be more than about three times the length of the plates. Thus, the total length of the deflection system required is about $4l_1$ and the maximum angle of convergence of the beam as shown in Figure 2 is

$$\theta_{max} = \frac{d}{2l} \cong \frac{1}{16} \tag{28}$$

Since it is desirable to have the beam fill approximately one-half the deflection space, θ reduces to a value of $1/32$.

In order to realize an appreciable fraction of the ultimate current density it has been seen that the magnification ratio must be made sufficiently small. The magnification ratio may be determined by substituting in Equation (8):

$$M \leqq \left(\frac{kT}{Ve} \right)^{1/2} \frac{1}{\theta} \cong 1 \tag{29}$$

using values: $V = 100$, $\dfrac{kT}{e} = \dfrac{1}{10}$, and $\theta = \dfrac{1}{32}$.

This means that in order to approach the maximum transconductance the length of the input end of the gun (object distance) should be approximately equal to the length of the deflection end (image distance). Thus, most of the dimensions of the tube are easily determined for a given frequency and given beam voltage. It should be noted that, if very low beam current is also required as well as high transconductance, and if the current cannot be limited sufficiently by the first aperture (one nearest cathode) it may be necessary to increase the object distance beyond that indicated above.

With the approximate dimensions chosen for a simple deflection gun, the type of focusing lens can be considered. A converging lens is desired that with reasonable potential ratios will give the focal length required for proper magnification ratio (roughly unity as shown in Equation (29)). Such a lens is conveniently constructed from a series of aperture lenses with the deflection plates in some cases acting as an aperture lens. Approximate figures for focal length can be computed from the slit aperture lens formula.[15]

$$F = \frac{2V}{E_2 - E_1} \tag{30}$$

where V = aperture potential. E_2 and E_1 are the electric fields on either side of the aperture. Note that $E_2 - E_1$ must be positive to give a converging lens.

There are an endless variety of focusing systems that could be used; some of the more likely forms are shown in Figure 8. In type A, for example, the main focusing lens is formed by the field at the entrance to the deflection plates. This can be treated approximately as an aperture lens where the field inside the deflection plates $E_2 = 0$. Applying Equation (30) it is seen that E_1 must be negative to form a converging lens. This means that V_2 must be less than V_1. It will be noted that the deflection plates are shown much closer to the entrance aperture than to the intercepting edge. This is done to minimize the focusing after deflection which reduces the deflection sensitivity, but this requirement makes it difficult to obtain the desired magnification ratio. A further disadvantage of this simple system is

---

[15] V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, A. W. Vance, ELECTRON OPTICS AND THE ELECTRON MICROSCOPE. John Wiley and Sons, New York, N. Y.. 1945.

that no stopping aperture is provided for limiting the current in the lens region.

Type B also combines focusing and deflection but the magnification ratio in this case is readily controlled by adjusting the distance between the first aperture (nearest to cathode) and the deflection plates. The current can be controlled by the stop S_3 or by an additional stop S_2. It will be shown later that with the double-aperture system formed by aperture S_3 and the deflection plates, a positive lens can be formed with either $V_1 < V_2$ or $V_1 > V_2$. The latter condition is
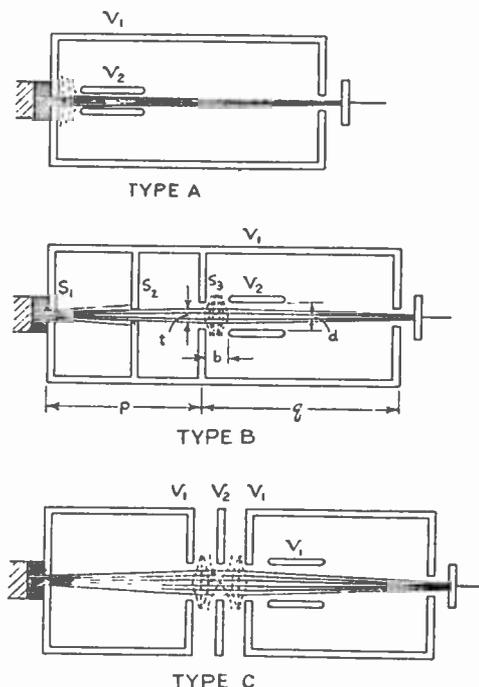
Fig. 8—Three types of electrostatic lens and deflection arrangements for beam-deflection tubes.

usually preferred because it suppresses secondaries from the aperture ahead of the deflection plates. It is this system which has been used in much of the experimental work and it forms the prototype of the deflection guns used in later tubes.

Type C allows the focusing to be controlled independently of the deflection plate potential, a desirable feature when attempting to maximize the transconductance by adjusting transit angle. This type of gun was useful in super-high-frequency tubes using cavity deflec-

tion systems. Generally the added complexity of this system has not been found worth while.

Considering the type B lens more closely it is seen that the focal length can be computed approximately by the combination of two aperture lenses to give

$$F \cong \frac{2b}{V_1/V_2 + V_2/V_1 - 2} \tag{31}$$

where $b =$ the effective distance from the aperture to the deflection plates.

A more exact expression can be had by referring to the work of L. H. Bedford.[16]

To obtain the one-to-one magnification ratio indicated for current density considerations, the focal length, F, should be one-half of the image distance, or roughly $2l_1$, in the tentative design discussed above. Assuming further that the separation, b, is equal to the plate separation, $d = \dfrac{l_1}{2}$, the resulting voltage ratio is $\dfrac{V_1}{V_2} = 2$ or 1/2. This was, in fact, the voltage ratio used in much of the early work. In these computations two minor effects were neglected; space charge and the focusing after deflection. The effect of space charge in these tubes is small but it is in the direction to require a stronger lens or a higher ratio of V_1/V_2. In a system such as type B where the potential in the space after the deflection plates is higher than the deflection plates there is a focusing action which tends to decrease the deflection sensitivity. However, in the design considered the field at the exit end of the plates is small compared to that at the entrance end and the focusing effect is generally negligible.

Now with a focusing system determined it is well to consider the electron emission system and to work forward from it to obtain the few remaining factors yet to be determined. It was seen that high cathode current density and low temperature are basic factors in obtaining high transconductance. In fact, as shown in Equation (21) increasing cathode current density is the only clear-cut means for increasing transconductance with a given cathode temperature. In the present work oxide-coated cathodes were used and current densities of from 0.100 to 0.200 amperes per square centimeter were considered reasonable values. To utilize the high current density of the cathode it is necessary to have a sufficient accelerating field in front of the

[16] L. H. Bedford, "Electron Lens Formulas," *Proc. Phys. Soc.*, Vol. 46, pp. 882-888, 1934.

cathode to overcome space charge. This accelerating field must be provided for by an accelerating electrode which may act either as a current limiting aperture or as the first lens in the optical system. In the first case as shown in Figure 9(a) the cathode is large and a slit aperture determines the magnitude of the beam current and forms the object which is imaged on the intercepting edge. In the second case shown in Figure 9(b), the current is determined by a line cathode which is imaged on the intercepting edge. In this case the first aperture acts as a diverging lens which should be sufficiently large to prevent serious aberrations. It should be noted that due to the accelerating field the apparent cathode plane will be at a distance
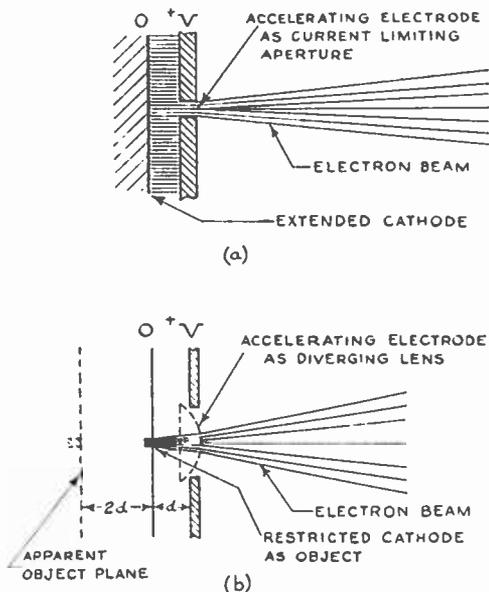
Fig. 9—Two methods of limiting beam current: (a) shows a narrow accelerating aperture; while (b) shows the use of a narrow emitting area on the cathode.

from the accelerator equal to twice the cathode aperture spacing.[15] In the case of full space charge this factor is increased to three. This effect is in most cases small and in any event is in the desired direction of decreasing the magnification ratio.

Both of the cathode arrangements shown in the Figure 9 are workable but it is found that imaging a slit aperture will usually give a sharper-edged image than for most practical cathodes and with the slit aperture it is also easier to limit the beam current to smaller values. In this connection it is well to stress again the point that the transconductance does not depend upon total current and this current

can theoretically be made as small as we please. One method of limiting the current is by making the first aperture very small, the other method is to increase the object distance, spreading the beam so that a large fraction of the current is absorbed at a later stop such as S_1. An approximate quantitative expression for the current has already been given in Equation (10). The actual amount of beam spread will be increased beyond that computed for initial velocities because of space charge. This can be readily computed for rectangular beams,[17] and is usually small and again has the effect of decreasing the magnification ratio.

We come now to a consideration of the lens system from the standpoint of aberrations which will reduce the transconductance below the theoretical value and which will have an even greater effect on the maximum ratio of transconductance to plate current than can be attained by working on the very edge of the beam. One important way of reducing aberrations is to restrict the fraction of the focusing lens occupied by the beam, as illustrated in Figure 8(b). The actual transconductance obtainable will thus be reduced from the theoretical transconductance of Equations (16) and (21), by at least the fraction $p = t/d$. In the early tubes to be described this factor was about one-half to one-third. Further reduction might be expected to enhance the transconductance to plate current ratio with some loss of transconductance.

EARLY EXPERIMENTAL RESULTS

In the early experimental work an effort was made to obtain the highest possible transconductance to current ratio in an electrostatically focused deflection gun. One of the high values which were measured was a ratio of 250 micromhos per microampere, or 20 times more than possible for conventional control methods. Expressed in another way, this corresponds to about a three-to-one change in plate current for input voltage increment of only four millivolts. This result was obtained with a gun similar in design to type B shown in Figure 8 using a 0.002-inch first aperture and 3-centimeter long deflection plates. The accelerating potential was 45 volts and the deflection plate potential 21 volts. The plate current was about 0.01 microampere and the transconductance 2.5 micromhos. Other tubes designed for higher frequencies generally gave ratios of 20 to 30.

Another early deflection gun having a transconductance to current ratio of 20 and an input transconductance of 40 micromhos was used

---

[17] B. J. Thompson, L. Headrick, "Space-Charge Limitations on the Focus of Electron Beams," *Proc. I.R.E.*, Vol. 28, pp. 318-324, July, 1940.

in combination with a 5-stage secondary-emission multiplier to obtain 100,000 micromhos transconductance at 5 milliamperes plate current. This tube was of the general design shown in Figure 10 except for the number of multiplier stages. The deflection plates were one centimeter long and with accelerator and deflection plate potentials of $+300$ volts and $+150$ volts. This tube had an input capacitance of 1.5×10^{-12} farads and an output capacitance of 3.5×10^{-12} farads which gives a bandwidth merit at low frequencies of

$$B = \frac{g_m}{2\pi \ (C_{in} + C_{out})} = 3000 \text{ megacycles} \tag{33}$$

or on the order of 50 times that of conventional tubes. One of these early tubes used as an amplifier gave a gain of 35 decibels at 450
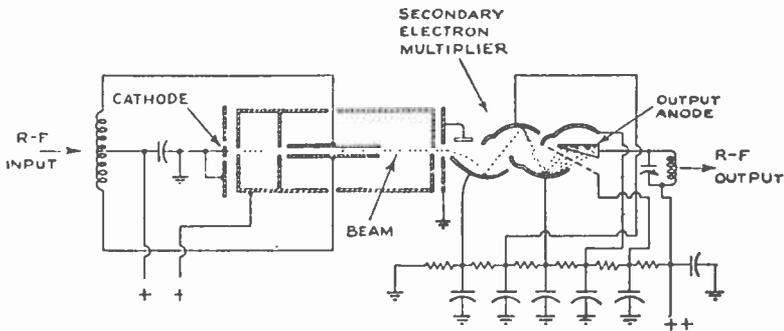
Fig. 10—Cross-sectional view of an early beam-deflection amplifier with 3-stage multipler.

megacycles, despite about 10-to-1 reduction in transconductance due to the combined effects of transit time through the multiplier and reduced deflection sensitivity. In these early tubes the width of beam was usually about one millimeter and the best gun transconductance was about 100 micromhos or 1000 micromhos per centimeter of beam width. In later tubes the beam width was increased to about tenfold without sacrifice of transconductance per unit width. In these earlier tubes the actual transconductance was never much higher than 10 per cent of the theoretical maximum given by Equation (21). One reason for this is that only a fraction of the deflection plate space was utilized. Aberrations and mechanical misalignment probably account for the balance of the reduction. Some idea of the sharpness of the beam can be had from the fact that the total beam thickness at the intercepting edge was 0.001 inch to 0.002 inch wide.

Conclusions

It has been shown that deflection control offers a possibility of obtaining substantial transconductance with low capacitance and low beam currents and with a very high ratio of transconductance to plate current. It has been shown experimentally that useful values of transconductance with low capacitance and low current can be obtained with a simple deflection gun combining focusing and deflection. This type of control is ideally suited for use with a high-gain secondary-emission multiplier to obtain very high transconductance, without excessive capacitance, thus making possible a tube with a bandwidth figure of merit many times greater than for conventional tubes.

The analysis and experimental confirmation of some of the properties of deflection control has already been useful in the development of practical tubes in these laboratories. It may be anticipated that this control method will be useful in other applications, a number of which have already been suggested in the literature. Among these
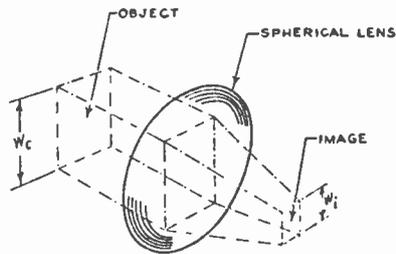
Fig. 11—Representation of electron optics using a spherical lens system to focus both width and thickness of a rectangular electron beam.

are wide-band amplifiers extending into the ultra-high or even the super-high frequency region, detectors and frequency multipliers, and low-frequency or direct-current control applications using the high-voltage sensitivity of deflection control.

Appendix I

The maximum low-frequency transconductance computed in the third section assumed a rectangular beam infinite in extent with focusing by means of a cylindrical lens. If one considers a rectangular beam of finite width as shown in Figure 11 it might appear possible to obtain still higher transconductance by the use of a spherical lens which would allow the beam to be compressed in width as well as in

thickness. It will be shown that actually less transconductance per unit width of cathode can be achieved by this method of focusing.

To evaluate the transconductance in this case it is necessary to use the current density expression for the point focus case which Pierce[14] has shown to be

$$j_1 = \frac{j_o}{M^2} \left[1 - (1 - \beta^2)\epsilon - \beta^2\phi/(1 - \beta^2) \right] \tag{A-1}$$

where
$$\beta = M \sin \theta,$$

$\theta =$ angle of convergence of beam,

$$\phi = \frac{Ve}{kT},$$

$j_o =$ cathode current density,

$j_1 =$ current density at image or intercepting edge.

In a practical deflection tube $\beta << 1$ in which case equation (A-1) reduces to

$$j_1 \cong j_o \phi \, \theta^2 \left(1 - \frac{M^2\theta^2\phi}{2} \right) \tag{A-2}$$

The transconductance is equal to

$$g_m = S \, j_1 \, W_1$$

where $W_1 =$ width of the beam at intercepting edge and S is the maximum deflection sensitivty

$$S = \frac{l}{4dV} \tag{A-3}$$

Also the maximum angle of convergence is

$$\theta_{max} = \frac{d}{2l} \tag{A-4}$$

Combining Equations (A-2), (A-3), and (A-4) and recalling that $W_1 = MW_c$ we obtain an expression for transconductance

$$\frac{g_m}{W_c} = j_o \frac{\phi}{16} \left(1 - \frac{M^2\theta^2\phi}{2} \right) \frac{dM}{V} \qquad (A\text{-}5)$$

Now in the cylindrical lens case it was seen that the transconductance approached a maximum as M approached zero. In the spherical lens case there is obviously an optimum value of M. By maximizing equation (A-5) with respect to M the optimum value is found to be

$$M = \left(\frac{2}{3} \right)^{1/2} \frac{1}{\theta \phi^{1/2}} \qquad (A\text{-}6)$$

(For example if $\theta = 1/32$, $\phi = 1000$, $V = 100$, and $\frac{e}{kT} = 10$ M $= 0.8$ or not much different from the practical value of unity assumed in the previous case.) Substituting the maximum value in Equation (A-5) the maximum transconductance per unit width of cathode is

$$\frac{g_m}{W_c} = \frac{(2/3)^{3/2}}{16} \phi^{1/2} \frac{d}{V\,\theta} \qquad (A\text{-}7)$$

but, since $\theta_{max} = \dfrac{d}{2l}$ and $\phi = \dfrac{Ve}{kT}$,

$$\left(\frac{g_m}{W_c} \right)_{max} = \frac{(2/3)^{3/2}}{8} \left(\frac{e}{kT} \right)^{1/2} \frac{l}{V^{1/2}} \qquad = 0.068 \left(\frac{e}{kT} \right)^{1/2} \frac{l}{V^{1/2}} \qquad (A\text{-}8)$$

mhos/centimeter

When compared with the maximum expression derived in the third section for the cylindrical lens case it is seen that the maximum transconductance per unit width for the spherical lens case is only about one-half.

SOME NOTES ON NOISE THEORY AND
ITS APPLICATION TO INPUT CIRCUIT DESIGN*†

By

WM. A. HARRIS

Tube Department, RCA Victor Division,
Harrison, N. J.

Summary—The mechanism by which noise is produced in an electron tube and the relation between induced grid noise and plate noise are discussed. An equivalent circuit with noise generators supplying voltages and currents to simulate noise derived from the plate current of a tube, from the grid by passage of this current, and from the input circuit is then analyzed to determine the optimum noise factor obtainable under various conditions. The frequency for which the quantity $R_{eq} g_m$ is unity is seen to be an appropriate figure of merit for the noise produced by an electron tube. The frequencies corresponding to chosen values for the noise factor are presented for several receiving tube types. The paper concludes with a discussion of the circuit requirements which must be satisfied in order to obtain noise factors approximating the theoretical values.

INTRODUCTION

NOISE generated in the first tube of a receiving system is frequently the factor controlling the over-all sensitivity of the system. An understanding of the mechanism by which such noise is produced is helpful in the design of receiving equipment, particularly with respect to the choice of tube types. If the electrons in a tube were to leave the cathode at a perfectly uniform rate, there would be no noise, or at least, none in the frequency range in which a tube is useful. The rate of emission of electrons, however, is not uniform. In any given interval of time there are probably a few more or a few less electrons leaving the cathode than the average number for that amount of time. The classical shot-effect derivations predict the magnitudes of fluctuations of this sort. Furthermore, because theory and experimental data have revealed the extent to which space-charge effects can reduce these fluctuations in electron tubes, the noise components of the plate current of a tube can be computed in many instances.[1]

At high frequencies, the fluctuation current induced in the grid of

---

* Decimal Classification: R138 × R161.6 × R361.211.

† Reprinted from *RCA Review*, September, 1948.

[1] B. J. Thompson, D. O. North, and W. A. Harris, "Fluctuations in Space-Charge-Limited Currents at Moderately High Frequencies", *RCA Review*, Vol. IV: No. 3, pp. 69-285, January 1940; No. 4, pp. 443-473, April, 1940; Vol. V: No. 1, pp. 106-124, July, 1940; No. 2, pp. 244-260, October, 1940; No. 3, pp. 371-388, January, 1941; No. 4, pp. 505-524, April, 1941; Vol. VI, No. 1, pp. 114-124, July, 1941.

a tube by the passage of the fluctuating plate current through the grid is another noise source which must be considered. The magnitude of the mean square of this current is proportional to the component of input conductance due to transit time.[2]

Since the grid noise and plate noise are derived in part from the same current fluctuations, they cannot be treated as entirely independent noise sources. Nevertheless, valuable working formulas and principles have been derived for conditions under which coherence between grid noise and plate noise can be ignored.[3] It is possible, moreover, that the improvements obtainable by taking coherence into account are not very important for the majority of tubes and circuits in current use. Theoretical considerations, however, indicate that a substantial improvement in noise factor may be obtained by taking advantage of the coherence between grid noise and plate noise if the conditions assumed for the theory can be realized in actual tubes. Experimental evidence shows that at least part of this improvement can be obtained in a practical system.[4]

In this paper, the relation between induced grid noise and plate noise is illustrated by an examination of the result of the passage of a single electron through a tube. Then, the conditions giving optimum noise factors are derived, using the methods employed by Herold and others. Herold[3] showed that the noise factor for a tube is a function of the product $R_{eq} g_{in}$, where R_{eq} is the equivalent noise resistance and g_{in} is the input conductance. In this paper, the frequency for which the product $R_{eq} g_{in}$ is unity is recommended as an appropriate noise "figure of merit" for a tube. The ratios of the operating frequencies to this reference frequency therefore can be used as the abscissas for curves of optimum noise factor.

CURRENT IMPULSES FROM ONE ELECTRON

Figure 1 shows the distribution of potential in a parallel-plate triode. The potential curve is based on the assumption of a Maxwellian distribution of initial velocities, with a cathode temperature of approximately 1000 degrees Kelvin. The dotted curve represents the velocity of an electron with just enough initial velocity to allow it to pass the point of minimum potential and continue to the anode. The time of

[2] D. O. North and W. R. Ferris, "Fluctuations Induced in Vacuum-Tube Grids at High Freqencies", *Proc. I.R.E.*, Vol. 29, pp. 49-50, February 1941.

[3] E. W. Herold, "An Analysis of the Signal-to-Noise Ratio of Ultra-High-Frequency Receivers", *RCA Review*, Vol. VI, No. 3, pp. 302-331, January, 1942.

[4] M. J. O. Strutt and A. vanderZeil, "Signal-Noise Ratio at VHF", *Wireless Engineer*, Vol. 23, pp. 241-249, September, 1946.
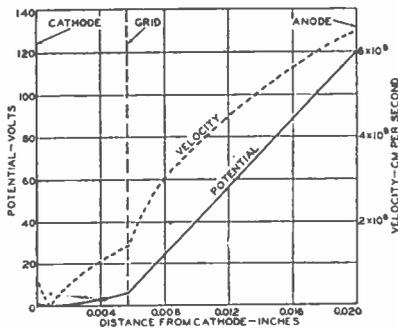
transit for such an electron can be computed by a graphical integration process.

When a charge is in motion between two electrodes in a tube, current flows between the electrodes bounding the region containing the charge. For a parallel-plane structure this current is proportional to the velocity of the charge and inversely proportional to the distance between the two plane boundaries; it does not depend on the position of the charge relative to the boundaries.[5] The velocity curve of Figure 1 is applicable for a charge of the indicated initial velocity. If the velocities for various points in the cathode-grid space are divided by the distance between cathode and grid, and the velocities for various points in the grid-anode space are divided by the distance between grid and anode, quantities proportional to the current due to the motion of the charge for these various positions are obtained. Then, the relation between position and time obtained by integration may be used to obtain a current-time curve.

The curves of current versus time for the tube structure of Figure 1 are shown in Figure 2. The solid curve shows the current to the grid and the dotted curve the current to the plate which would result from the passage of one electron. The choice of the velocity of the slowest electrons which can reach the anode leads to a computation difficulty; the time required for such electrons to pass the potential-minimum region is theoretically infinite. Consequently, transit times are computed from the cathode to a point near the potential minimum on the cathode side, and from the grid back to a point near the potential minimum on the grid side. The three rectangles between the ends of the two curves show the times and currents for charges passing between the terminal points of these curves with velocities exceeded by 90, 50, or 10 per cent of the electrons reaching the anode. The use of one of these velocities would cause some change in the remainder of the curve, both in the current and the time scale, but the shape of the curve would be about as shown. The effect of a change in initial velocity on the current between grid and anode would be almost negligible. The cathode-to-grid transit time for an electron with an initial velocity

---

[5] S. Ramo, "Currents Induced by Electron Motion", *Proc. I.R.E.*, Vol. 27, pp. 584-585, September, 1939.

corresponding to one of the rectangles in Figure 2 is, therefore, approximately the sum of the transit times represented by the two curves and the appropriate rectangle. The indicated range is from 6 to 8×10^{-10} seconds. The transit time from grid to anode for the conditions of Figure 2 is about 1×10^{-10} seconds.

The curve of Figure 2 does not show the compensating effect which takes place when an extra charge passes through a tube. The potential minimum is depressed by an amount depending on the position of the added charge, during the whole time this charge is between cathode and grid. The result is a reduction of the current, which can be considered equivalent to the passage of a series of charges of opposite sign between potential minimum and grid, and the passage of charges of the same sign but of opposite direction between potential minimum and cathode. These effects account for the shot-effect reduction factor computed by North[1]. It appears, however, that between the initiating pulse and the compensating current there is some time delay which may be important in the determination of the noise at very high frequencies.

As soon as the extra noise-producing charge leaves the cathode a small effect on the minimum potential will be noted; some electrons which are reaching the potential minimum at this instant turn back instead of continuing toward the plate. The effect of the extra charge persists until it reaches the grid. The compensating effect cannot be completed until the time at which an electron, turned back because of the depression of the potential minimum when the extra charge was near the grid, would have reached the grid had it not been turned back.
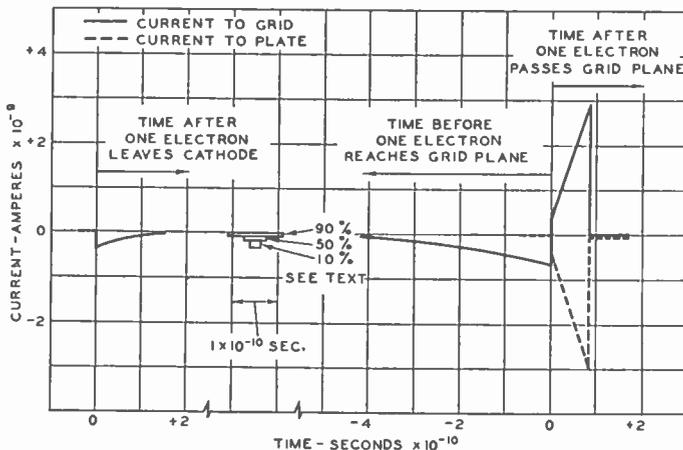

Fig. 2—Grid and plate currents due to passage of a single electron.

The compensating charges in motion after the passage of the causing charge, however, must themselves have an effect on the minimum potential. The complete result, consequently, is probably of the nature of a damped oscillation with a period related to the transit time of an electron from cathode to grid. Thus, the curves of Figure 2 do not present a complete picture of the generation of noise in a parallel-plane triode, but they do show some of the characteristics of the basic noise impulses.

It is pertinent at this point to discuss the extent to which grid and plate noise currents can be made to cancel each other. The pulse shapes are quite different, so it is evident that complete cancellation cannot be expected. Partial cancellation may be obtained if a voltage is developed at the grid by allowing the grid current to flow into a capacitor. This voltage is proportional to the integral of the grid current and has the effect of momentarily reducing the plate current. A suitable choice of capacitor value can give a plate-current pulse of zero net area for electrons of a particular initial velocity. However, there will always be some conductance in the grid circuit which will result in a grid-voltage component tending to increase the noise output. Moreover, electrons leaving the cathode with velocities too low to allow them to pass the point of minimum potential produce pulses of grid current without producing corresponding plate-current pulses.

DETERMINATION OF FREQUENCY SPECTRA

The curves of Figure 3 illustrate the method by which the frequency spectra corresponding to the grid-current and plate-current pulses may be obtained. The current which could be measured in the small frequency range represented by $d\omega$ at a frequency $\omega/2\pi$ is obtained from the Fourier integral:

$$A d\omega = \frac{2}{\pi} \int F_{(\lambda)} \cos \omega\lambda \, d\lambda \cos \omega t \, d\omega + \frac{2}{\pi} \int F_{(\lambda)} \sin \omega\lambda \, d\lambda \sin \omega t \, d\omega.$$

$$(1)$$

The function $F_{(\lambda)}$ represents the pulse.

When the frequency is low in comparison with the reciprocal of the transit time, the value of the term $\cos \omega\lambda$ in Equation (1) is nearly constant over the region in which $F_{(\lambda)}$ has a value other than zero. In addition, the value of the term $\sin \omega\lambda$ in Equation (1) can be represented as a straight line with a slope directly proportional to the frequency over the same region.

Because the grid pulse has equal positive and negative areas, the

Fig. 3—Development of frequency spectra.
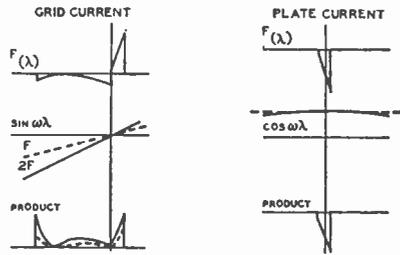
integral containing the cosine terms is zero for any frequency low enough so that cos ωλ can be considered constant. The integral containing the sine terms will have a value which is repre-
sented graphically by the area under the curve labeled "product" in the grid-current curves of Figure 3. This area will be directly proportional to the frequency because the slope of the sin ωλ line is proportional to frequency. It will also be proportional to transit time because, if the areas and shapes of the parts of the $F_{(\lambda)}$ curve are maintained constant and the base line is extended, the area under the "product" curve will increase in proportion to the increase in base-line length.

The plate pulse will give a zero value for the integral containing sine terms if a suitable point of origin is chosen. The integral containing cosine terms, then, gives the current, which is independent of the frequency when the frequency is low. Because the area of the plate-current pulse represents the amount of charge producing the pulse, the current in a small frequency band resulting from a given amount of charge is also independent of the transit time.

The mean-square noise current measurable in any frequency band results from large numbers of pulses distributed at random with respect to time. For the plate current, consequently, the mean-square current $di^{\overline{2}}$ in a frequency band of width df can be represented by the

equation
$$di^{\overline{2}} = k_2 \, df \tag{2}$$

and for the grid current, by
$$di^{\overline{2}} = k_1 \, \omega^2 \, \tau^2 \, df \tag{3}$$

where τ is the transit time, or
$$di^{\overline{2}} = k_1 \, \theta^2 \, df \tag{3a}$$

where θ is the transit angle.

The electronic component of input conductance is proportional to the square of the transit angle[6], so a proportionality between the mean-square noise current and the input conductance is indicated, thus:

$$di^{\overline{2}} = k_3 \, g_1 \, df. \tag{4}$$

North and Ferris[2] found that the complete relation for grid-current noise is
$$di^{\overline{2}} = \theta_1 \, g_1 \cdot 4k \, T_o \, df \tag{5}$$

[6] D. O. North, "Analysis of the Effects of Space Charge on Grid Impedance", *Proc. I.R.E.*, Vol. 24, pp. 108-136, January, 1936.

where θ_1 has a numerical value of approximately 5 when the cathode temperature is 1000 degrees Kelvin and the reference temperature T_o is approximately 300 degrees Kelvin. The dependence on temperature is discovered only when the analysis is extended to include the compensating currents.

The plate current noise from a tube may be represented as if it were derived from a noise voltage at the grid sufficient to produce the noise current. The appropriate equations[1] are $\quad d\bar{e^2} = 4kT\,R_{cq}\,df \quad$ (6)

where, for oxide-coated cathode tubes, theory indicates approximately that, for triodes, $\qquad\qquad R_{cq} = 2.5/g_m \qquad\qquad$ (7)

and for pentodes $\qquad R_{cq} = \dfrac{I_b}{I_b + I_{c2}}\left(\dfrac{2.5}{g_m} + \dfrac{20\,I_{c2}}{g_m^2}\right) \qquad$ (8)
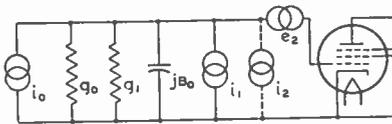
Fig. 4—Equivalent circuit.

CIRCUIT ANALYSIS

The circuit of Figure 4 represents the replacement of a real tube by a fictitious noise-free tube with zero input admittance and suitable noise generators and external circuit elements. The plate noise is introduced by a constant-voltage generator delivering a voltage e_2 in series with the grid. The noise current to the grid, i_1, is represented by a constant-current generator across the grid circuit. The noise from the input system, i_o, is represented by a second constant-current generator. The plate-noise generator can be replaced by another constant-current generator; the voltage output of the plate-noise generator is multiplied by the total admittance of the input circuit to give the required current i_2. The relations between the noise currents and the tube and circuit parameters are given by the equations:

$$e_2 = K\sqrt{R_{eq}} \qquad (9) \qquad\qquad\qquad i_1 = (-j)\,K\sqrt{\theta_1 g_1} \qquad (10)$$

$$i_2 = e_2(g_o + g_1 + j\,B_o) = K\sqrt{R_{eq}}(g_o + g_1 + j\,B_o) \qquad (11)$$

$$i_o = K\sqrt{\theta_o g_o} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (12)$$

where $\quad K = \sqrt{4kT\Delta f}$

$\qquad g_1$ is the electronic component of input conductance;

R_{eq} is the resistance equivalent for the plate noise, referred to the grid;

θ_1 is a multiplier relating grid noise to input conductance; its value is approximately 5 for tubes with oxide-coated cathodes;

θ_o is a multiplier representing the ratio of antenna noise to the noise in a resistor at room temperature;

g_o is the antenna conductance, referred to the grid;

B_o is the net susceptance of the circuit at the operating frequency.

It is assumed that the conductance g_o can be varied arbitrarily by some such means as a variable-ratio transformer between antenna and grid. Also, it is assumed that means such as a tuning capacitor are provided so that B_o can be varied arbitrarily. Ohmic losses in the input circuit are neglected.

The quantity $(-j)$ in parenthesis in the expression for i_1 indicates that i_1 may be in quadrature with e_2 over a specified frequency range. The preceding discussion suggests that this assumption is legitimate in the case of a triode, when the frequency is not too high and the frequency band is not too wide. The assumption is not valid, however, for a pentode because in that case the larger part of the plate noise results from the division of current between plate and screen grid[1], and consequently it cannot be correlated with the grid noise.

The total mean-square current from the three generators of Figure 4 can be found as follows: Add i_1 and i_2, taking coherence, if assumed, into account. Then, determine the sum of the squares of i_o, the real part of $(i_1 + i_2)$, and the imaginary part of $(i_1 + i_2)\cdot$ When coherence is not assumed, simply add the mean-square values of i_o, i_1, and i_2 The results follow:

When a quadrature relation between grid and plate noise is assumed, the mean-square current is

$$\overline{i^2} = K^2\{g_o\theta_o + R_{eq}(g_1 + g_o)^2 + (B_o\sqrt{R_{eq}} - \sqrt{\theta_1 g_1})^2\} \qquad (13)$$

When no coherence is assumed

$$\overline{i^2} = K^2\{g_o\theta_o + R_{eq}(g_1 + g_o)^2 + R_{eq}B_o^2 + \theta_1 g_1\} \qquad (14)$$

OPTIMUM NOISE FACTORS

Optimum performance with respect to noise is obtained when the term $g_o\theta_o$ is as large as possible in comparison with the other terms.

and, in fact, the noise factor as defined by North[7], Friis[8], and others is obtained by dividing Equation (13) or (14) by $K^2 g_o \theta_o$ and assuming $\theta_o = 1$. The first step in finding conditions for minimum noise is the adjustment of B_o to eliminate the term in which it appears in either equation.

Then, either equation can be differentiated with respect to the ratio g_1/g_o and an optimum value of noise factor can be obtained. The noise factors after adjustment of B_o are given by the equation

$$NF = 1 + R_{eq} g_1 \left(\frac{g_1}{g_o} + \frac{g_o}{g_1} + 2 \right) \tag{15}$$

when coherence is assumed, and the equation

$$NF = 1 + R_{eq} g_1 \left(\frac{g_1}{g_o} + \frac{g_o}{g_1} + 2 \right) + \theta_1 \frac{g_1}{g_o} \tag{16}$$

when coherence is not assumed.

The minimum noise factors, with the conditions for obtaining them are

$$NF = 1 + 4 R_{eq} g_1 \quad (17) \qquad \frac{g_1}{g_o} = 1 \quad (18) \qquad B_o^2 = \theta_1 g_1{}^2 / R_{eq} g_1 \quad (19)$$

when coherence is assumed; and, when coherence not assumed

$$NF = 1 + 2 R_{eq} g_1 + 2 \sqrt{\theta_1 R_{eq} g_1 + (R_{eq} g_1)^2} \tag{20}$$

$$\frac{g_1}{g_o} = \sqrt{R_{eq} g_1 / (\theta_1 + R_{eq} g_1)} \qquad (21) \qquad\qquad B_o = 0 \qquad (22)$$

The quantities R_{eq} and g_1 are both tube parameters Since they appear as the product $R_{eq} g_1$ in Equations (17) and (20), the magnitude of this product indicates the noise performance obtainable from a tube. The quantity g_1, however, varies with the square of the frequency. For purposes of computation, it is preferable to use as a reference parameter the square root of the product $R_{eq} g_1$, which

[7] D. O. North, "The Absolute Sensitivity of Radio Receivers", *RCA Review*, Vol. VI, No. 3, pp. 332-343, January, 1942.

[8] H. T. Friis, "Noise Figures of Radio Receivers", *Proc. I.R.E.*, Vol. 32, pp. 419-422, July, 1944.

Fig. 5—Minimum noise factor.

varies with the first power of
the frequency. The curves of
Figure 5 show the optimum noise
factors for the two cases con-
sidered, plotted against the quan-
tity $\sqrt{R_{eq}\,g_1}$ for the condition

$\theta_1 = 5$. If the frequency for which $\sqrt{R_{eq}\,g_1}$ is unity is designated as
f_n, the quantity $\sqrt{R_{eq}\,g_1}$ for any frequency f is equal to the ratio f/f_n.

The method of analysis described above is essentially the same
as that used by Herold[3]. The curve for the case of no coherence
(Figure 5) can be identified with one of the curves (Figure 5) of
Reference 3 when differences in the coordinates used are taken into
account. Equation (21), giving the required ratio of tube input
conductance to circuit conductance, is equivalent to Equation (7) of
Reference 3.

The susceptance required for the case of quadrature, as found
from Equation (19), is obtained by the same amount of capacitance
at any frequency. Equation (19) can be rewritten

$$B_o{}^2 = \theta_1 g_1/R_{eq}. \tag{19a}$$

Because g_1 is proportional to the square of the frequency and R_{eq}
and θ_1 are independent of frequency, it is evident that the susceptance
B_o is directly proportional to the frequency and, consequently, can be
produced by a fixed capacitance.

COMPARISON OF TUBES

The data for Tables I and II were obtained by calculating values
for the equivalent noise resistance and using measured values for input
conductance for the tube types listed. Table I gives the reference
frequency for noise, f_n, and the frequencies for which noise factors of
1, 3, and 10 decibels are calculated for a number of pentode types. No
coherence is assumed between plate noise and grid noise for this case.
Table II gives similar data for two triodes and for several pentodes,
connected as triodes, under the alternate assumptions of no coherence
between plate and grid noise, and a quadrature relation between plate
and grid noise. The 10-decibel column for the quadrature case is omitted
because the indicated frequencies are too high to make the assumption
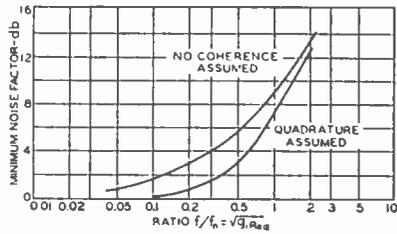appear reasonable.

Table I—Pentodes

Type	R_{eq} ohms	g_{in} (100 mcs.) micromhos	f_n mcs.	Frequency for Noise Factor (1 db.) mcs.	(3 db.) mcs.	(10 db.) mcs.
6SK7	11,600	440	45	2.5	8.9	56
6AC7	650	1,730	94	.5.3	19	119
6BA6	3,800	580	67	3.8	13.4	85
6AG5	1,900	300	133	7.5	27	169
6AK5	1,900	125	208	11.6	41	262
6BH6	2,360	340	122	6.8	24	154
6BJ6	3,800	275	98	5.5	19.3	124

The input conductance values used in Tables I and II were meas-
ured by the susceptance-variation method[9] and include the effects of
lead inductance. For pentodes, the predominant lead effect is that
of the cathode-lead inductance, which tends to increase the input
conductance. For triodes, inductance in the plate lead tends to reduce
the input conductance and this effect may be equal or greater than
the effect of cathode-lead inductance. For the triode-connected
pentodes, the input-conductance data obtained with the tubes con-
nected as pentodes are used.

Triode "A" in Table II is a developmental triode, designed pri-
marily for use as a high-frequency oscillator. The low input con-
ductance and the consequent high "f_n" value recorded for this type

[9] "Input Admittance of Receiving Tubes", RCA Application Note
AN-118, RCA Tube Department, Harrison, N. J., April, 1947.

Table II—Triodes and Triode-Connected Pentodes

Type	R_{eq} (Triode) ohms	g_{in} (100 mcs.) micromhos	f_n (Triode) mcs.	No Coherence Assumed (1 db.) mcs.	(3 db.) mcs.	(10 db.) mcs.	Quadrature Assumed (1 db.) mcs.	(3 db.) mcs.
6SK7.....	970	440	72	4.0	14	91	18	36
6AC7.....	214	1,730	164	9.2	33	207	41	82
6BA6.....	410	580	204	11.5	41	258	51	102
6AG5.....	380	300	294	17	59	374	74	148
6AK5.....	380	125	476	26	92	580	116	230
6BH6.....	390	340	274	15.4	54	345	68	137
6BJ6.....	485	275	274	15.4	54	345	68	137
6J6......	470	195	320	18.0	63	410	80	160
"A"*.....	360	50	747	42	147	940	186	373

* Developmental triode.

is probably accounted for by close spacing, high current density, and a symmetrical cylindrical structure which contributes to uniformity in the cathode-to-grid and grid-to-plate transit times.

RELATION OF REFERENCE FREQUENCY TO TRANSIT TIME

The reference frequency for noise for a triode depends primarily on the electron transit time between cathode and grid. The noise equivalent resistance for a triode is approximately $R_{eq} = 2.5/g_m$ (7)

and the electronic component of the input conductance[6] is approximately

$$g_1 = g_m (\omega \tau_1)^2/20. \tag{23}$$

The product, therefore, is

$$R_{eq} g_1 = (\omega \tau_1)^2/8. \tag{24}$$

This product is equal to unity when

$$\omega \tau_1 = 2.83 \tag{25}$$

so

$$f_n = 0.45/\tau_1. \tag{26}$$

The values of f_n obtained from Equation (26) are even higher than the values given in Table II. The cathode-to-grid transit time for a tube such as Type 6AK5 is of the order of 7×10^{-10} seconds, so the value of f_n from the above equation is

$$f_n = 0.064 \times 10^{-10} \text{ cycles} = 640 \text{ megacycles}.$$

The value obtained for Type 6AK5 from input conductance data (Table II) is 476 megacycles.

It appears that the only way to increase the frequency for a given noise factor with electron tubes of conventional design is to reduce the transit time. Triode types such as the 6J6, 6J4, and 2C43 are designed with close enough spacings and, consequently, short enough transit times to give promise of good results in equipment designed for minimum noise.

EFFECT OF CIRCUIT LOSSES

An important question with reference to the application of the curves and tables presented is the attainability of the circuit conditions assumed. The conditions are not hard to realize in practice, as the following examples illustrate:

1. Consider the use of Type 6AK5 as a pentode amplifier at 40 megacycles. The reference frequency f_n is 208 megacycles, so the ratio f/f_n is 0.192; the product $R_{eq} g_1$ is 0.037. The calculated noise factor is 3 decibels. The required ratio g_1/g_o is 0.046. Because the tube input conductance for 40 megacycles is 19.7 micromhos, the required antenna loading is 230 micromhos. For a tube input capacitance of 6 micromicrofarads, the quantity ωC is 1500 micromhos; because

the total conductance at the grid is 250 micromhos, the minimum value of Q is 6. Higher Q values may be obtained by adding more capacitance with appropriate inductance values. It is evident that there will be no serious increase in the noise factor until the conductance of the added elements becomes appreciable in comparison with 250 micromhos. If the Q is improved to 50 by addition of a resonant circuit with a Q of 200, the added conductance is approximately 60 micromhos. The noise factor would be increased from 3 to 3.5 decibels by the added circuit losses.

2. Consider the 6AK5 or an equivalent tube connected as a triode used at 200 megacycles. Neutralization may be used to avoid feedback, but feedback generally does not have an important effect on the question of obtainable noise factors. The reference frequency f_n is 476 megacycles; the ratio f/f_n is 0.42; the expected noise factors, from the two curves of Figure 5, are 5.1 decibels for no coherence, 2.3 decibels if the quadrature relation holds. In the first case, the required antenna loading is 2600 micromhos and the resulting Q for the input circuit is only 2. Adjustment of Q to any moderate desired value can be made by the addition of circuit elements as before without materially affecting the noise factor. In the second case, the antenna loading would be adjusted to equality with the tube conductance, which is 500 micromhos for this frequency. Then, the susceptance which must be added is 5700 micromhos, corresponding to a capacitance of 4.6 micromicrofarads.

Conclusions

The conclusions which may be drawn from this discussion may be summarized as a set of principles to be followed in the design of amplifiers for low noise.

1. Choose an input tube with low transit time. For frequencies above 30 megacycles, use a triode or a triode-connected pentode.

2. Adjust the input circuit with signal-to-noise ratio as the criterion. This adjustment is most readily made by using a noise generator, such as a diode, as a signal source.

3. Try the effect of detuning the input from resonance and the effect of increasing the coupling to the antenna beyond the value for maximum gain.

When theoretical considerations indicate a very low noise factor, it may be necessary to pay considerable attention to the design of the load circuit of the first tube and the input circuit for the second tube to obtain optimum results.

SUPERHETERODYNE FREQUENCY CONVERSION USING PHASE-REVERSAL MODULATION*†

BY

E. W. HEROLD

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary

In radio reception using the superheterodyne principle the incoming signal is changed in frequency by the converter stage of the receiver to a new and lower frequency known as the intermediate frequency. The electron tubes used in the converter stage have been characterized in the past by poor performance as compared with that of tubes used for amplification. This paper describes a new principle whereby frequency conversion may be accomplished with substantially improved performance over that available from conversion methods heretofore used.

The principle of conversion herein described is to reverse the phase of the signal output periodically at a rate which differs from the signal frequency by the intermediate frequency. This may be done either by continuous variation of phase or by continuous variation of tube transconductance from positive to negative. The result is a conversion transconductance which is twice as high as had heretofore been believed ideal. Furthermore, if the phase-reversal rate is made by an integral multiple of an applied local-oscillator frequency, equally good conversion is obtained at a harmonic of the local oscillator without spurious responses at any other harmonic than the one chosen. An electron tube with a multihumped characteristic has been devised as a means to this end since the transconductance characteristic will then vary from positive to negative as the control voltage is varied. An analysis of such a tube is carried out in detail, including the effect of fluctuation noise.

The analysis shows that the new conversion method doubles the conversion gain possible in a tube with a given maximum transconductance. In an ideal case with no second-stage noise, the signal-to-noise ratio is as good as with the same tube used as amplifier; even in practical cases, the mixer is only 10 per cent to 20 per cent poorer than the amplifier. This is in contrast with conventional mixer methods in which the signal-to-noise ratio is from two to three times poorer than when the same tube is used as an amplifier.

Conversion at a harmonic may also be achieved with high gain but it is found that the signal-to-noise ratio is not as favorable as with fundamental operation.

(15 pages, 17 figures)

RADIO-FREQUENCY PERFORMANCE OF SOME RECEIVING TUBES IN TELEVISION CIRCUITS*†

By

ROBERT M. COHEN

Tube Department, RCA Victor Division,
Harrison, N. J.

Summary

Several types of receiving tubes may be used to advantage in television receivers designed to tune all thirteen channels. This paper discusses the performance of these tube types in radio-frequency amplifier, mixer, and local oscillator applications. Both push-pull "balanced" circuits and single-ended "unbalanced" circuits are discussed. Data are presented for over-all gain, noise, image rejection, and, to a lesser extent, on oscillator frequency stability. These data are taken at two representative channels in the television band: Channel No. 4 (66 to 72 megacycles) and Channel No. 11 (198 to 204 megacycles).

(13 pages, 4 figures, 5 tables)

THE TRANSITROL, AN EXPERIMENTAL AUTOMATIC-FREQUENCY-CONTROL TUBE*†

By

JEROME KURSHAN

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary

A tetrode may be operated with its #2 grid as the anode of an oscillator and its plate as an electron reflector. The reflector potential controls the electron transit time and hence the oscillation frequency. This and other forms of transit time control are investigated theoretically and experimentally. An approximate expression for the control sensitivity is

$$s = df/V = - kY_m0 \ (\cos \Theta)/4\pi C \ (V_o - V)$$

where k is the feedback fraction which is of the order of 0.2, Y_m the magnitude of the transadmittance, C the circuit capacitance, V_a the anode voltage, V the reflector voltage and Θ the total transit angle. This predicts a maximum in s near $\Theta = \pi/4$ and a second and greater maximum near $\Theta = 2\pi$.

In a circuit suitable for local oscillator use in the frequency-modulation

(FM) broadcast band (88-108 megacycles), $s = 100$ kilocycles/volt was readily obtained with a special tube. In a particular application to an FM receiver, warm-up frequency drift at the high-frequency end of the band was reduced by a factor of 4.5. A receiver with the automatic-frequency-control (AFC) circuit requires 2 resistors and 1 condenser more than a conventional receiver, but saves a radio-frequency choke. Another successful application has been as a one-tube FM transmitter with the reflector elec-trode modulated directly by a microphone output.

Standard pentodes also allow transit time control, but do not have optimum spacing. Optimizing the design for one frequency will result in less control sensitivity at lower frequencies, but there will still be a residual improvement over an uncontrolled oscillator.

(17 pages; 13 figures)

A PHOTOTUBE FOR DYE IMAGE SOUND TRACK*†

BY

ALAN M. GLOVER AND ARNOLD R. MOORE

Tube Department, RCA Victor Division,
Lancaster, Pa.

Summary.—In view of the use of dye image sound track on new-type color film, a phototube has been developed with characteristics suitable for sound reproduction from these films as well as from ordinary silver tracks. It is a gas-filled, high-sensitivity phototube for use in standard reproducing equipment. The maximum spectral response occurs in the blue blue-green region of the spectrum. Details of construction, sensitivity, characteristic curves, spectral response, frequency response, and life are presented.

Color is a subject of considerable interest to the Society of Motion Picture Engineers. Color complicates the problem of the sound track engineer. As an introduction to the discussion of a new photosensitive surface which will be of considerable value for use with dye image sound tracks, a brief review of the characteristics of the phototubes now in general use is in order.

The advent of sound with motion pictures in the late 1920's was in large measure made possible by the introduction of the caesium-silver-oxygen photosurface. Considerable effort had been expended in the previous decade in an attempt to raise the hitherto limiting sensitivity of the photosurfaces then available and, as is so frequently the case in technical progress, it is difficult to assay the part which the photosurface itself played in the new sound industry. Concomitant with the new photosurface were improved amplifying tubes and steady progress in the quality of the film, but the industry would have progressed slowly without the new photosurface.

The most advantageous characteristic of the caesium-silver-oxygen photosurface was its high sensitivity in the near infrared, a region of the spectrum in which the major portion of the energy of the incandescent light source is concentrated. About three-quarters of the total sensitivity of this photosurface to light from an incandescent source lies in the infrared. This sensitivity to infrared was of great advantage as long as the sound track consisted of a developed silver image which

---

296

may be exposed to densities as high as three in this region. With the advent of dye image tracks a possibility, it immediately became apparent that the marked transparency of such tracks in the near infrared to which the *S-1* surface is so sensitive would seriously limit the modulation obtainable. This is true whether the sound track is of the variable-area or variable-density type. Other authors have already touched on this subject in a previous issue of the JOURNAL.[1]
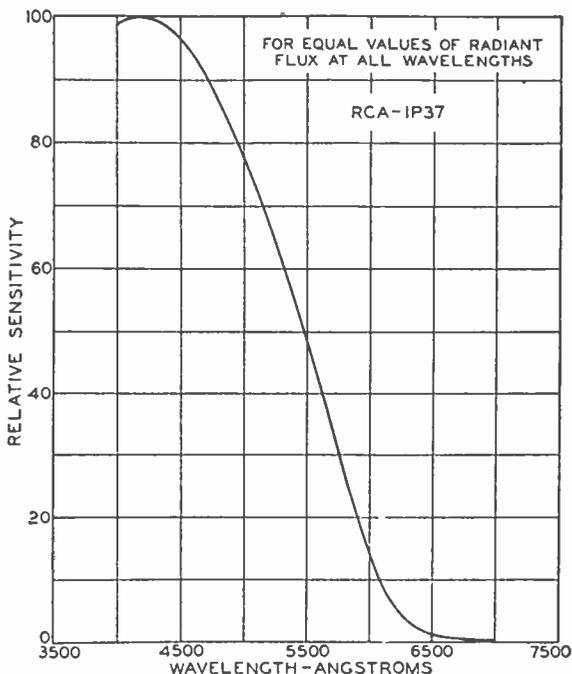

FIG. 1. Spectral sensitivity of phototube having *S-4* response.

In 1940 there appeared information on a new photosurface, the sensitivity of which is largely concentrated at the short wavelength portion of the visible spectrum.[2] Most of the phototubes which have employed this surface are of the high-vacuum type and this is probably the reason for the slow acceptance of such tubes by the motion picture industry. The gas-filled phototube has been popular in the industry for 3 reasons: high sensitivity with resultant high signal-to-noise ratio, adaptability of the voltage sensitivity of the gas tube as a

volume control, and the lower impedance level of the gas tube. Although the cause is not yet completely known, some difficulty has been encountered in introducing inert gas into phototubes containing the *S-4* surface. Such tubes have suffered from short life. However, this paper will outline the characteristics of a new tube, the RCA *1P37*, a gas-filled phototube of good life whose properties are such as to indicate the broad possibilities for its use with dye image sound tracks.
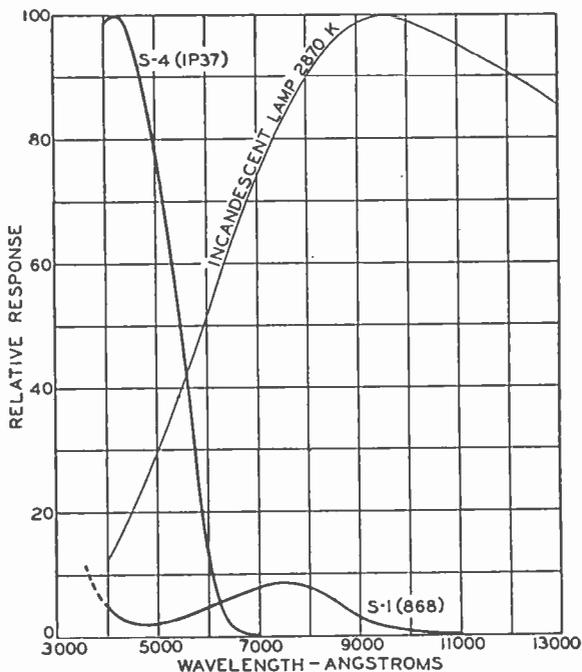
FIG. 2.   Spectral response of *S-1* and *S-4* phototubes.

The important characteristic of the new photosurface, its spectral sensitivity, is shown in Fig. 1. By contrast the same characteristic for the caesium-silver-oxygen *S-1* surface is also shown in Fig. 2. It should be borne in mind that these characteristics are typical of an average tube and that considerable variation from tube to tube may be expected. It is believed that the variations in the *S-4* spectral response are proportionately much less than those encountered in the *S-1* surface. In addition to our own data, the data taken by a number of investigators have recently been analyzed by us, and the curve

shown is suggested as one which might be adopted as a typical standard. Variations in the position of the spectral maximum from 4000 to 4500 A are commonly found with occasionally a peak at as short a wavelength as 3700 A being encountered. Based on data in greater quantity than previously available a curve giving the position of the maximum at 4200 A is believed typical. Little variation in the longer
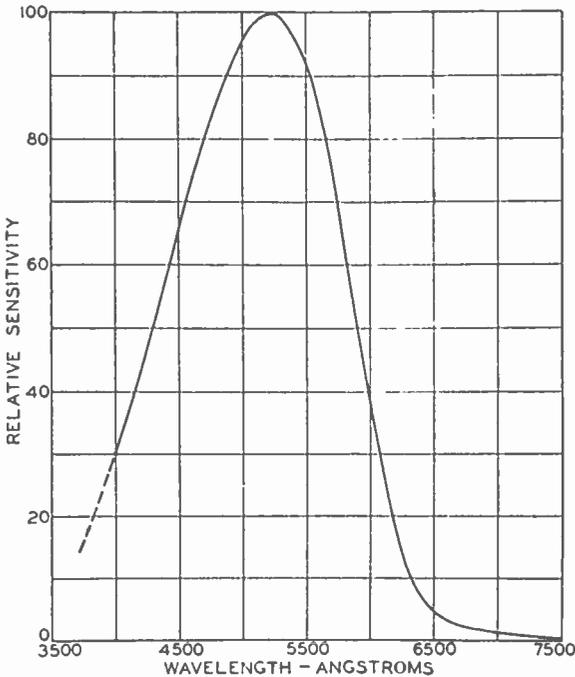
Fig. 3. Spectral response of *S-4* surface with tungsten light at 2870 K.

wavelength portion of the curve is encountered. Since the phototube will be commonly used in conjunction with an incandescent light source, the product of the spectral response curve of the tube and of a light source operating at 2870 K is shown in Fig. 3. This represents the effective spectral sensitivity of the tube as used in motion picture equipment.

The sensitivity of a gas-filled phototube containing the *S-4* photo-surface to a tungsten light source operating at a standard color temperature, 2870 K, may be made to vary considerably. For the pur-

pose of introducing a replaceable phototube for the *868* into the motion picture industry a relatively high sensitivity is not required. The amount of argon gas added can be varied to make the resultant over-all sensitivity comparable to that of the *868* even though the energy output of the light source is not well adapted to the *S-4* surface. The frequency response of the phototube is also a function of the gas pressure employed as is the breakdown voltage of the tube. With these factors in mind, an average gas amplification factor of three has been chosen. The gas amplification is less than that of the
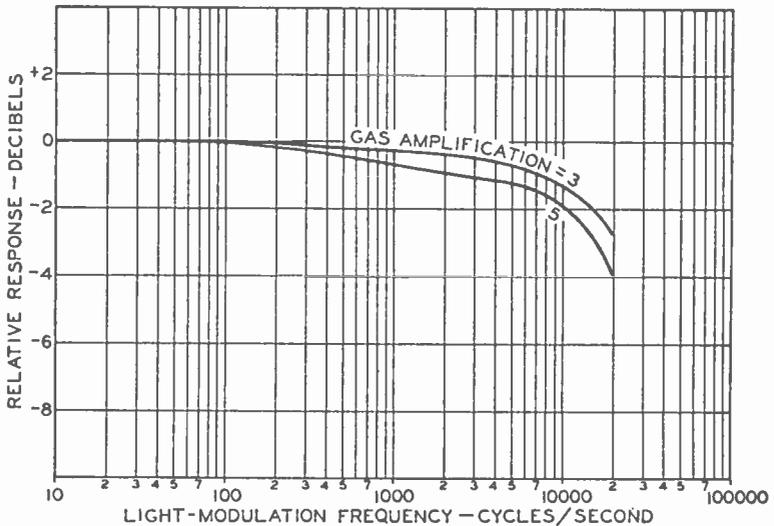
Fig. 4. Frequency response of *1P37*.

868 and therefore the frequency response of the *1P37* is slightly better. The frequency response for 2 different gas amplification factors is shown in Fig. 4. These curves were obtained using a glow lamp modulated light source,[3] the data being corroborated by measurements with standard frequency sound track.

It may be readily seen from the spectral characteristics that the blue-sensitive photosurface is more sensitive to variations in the temperature of the light source. This characteristic is believed to be its only point of inferiority when compared with the *S-1* surface. Data are shown in Fig. 5. Regulation of the light source voltage should be designed accordingly. It should be emphasized that the sensitivity

figures quoted for any photosurface when expressed in microamperes per lumen vary with color temperature of the light source. Care should be taken to state on what basis data are quoted.

Data giving the anode or output characteristic curves of the RCA *1P37* phototube are shown in Fig. 6. These curves are comparable to those for the *868*, thus permitting replacement of the latter tube by the *1P37* without modification of the circuit.
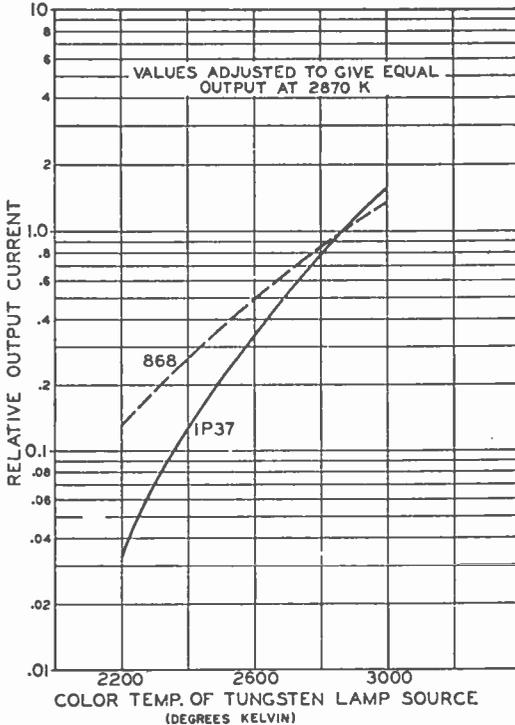
FIG. 5. Relative output of *868* and *1P37* as function of light source color temperature.

Prior to the war, one of the authors began work on the problem of making gas-filled tubes using the *S-4* surface. Considerable loss in sensitivity on use was encountered, and this conclusion was later stated in foreign articles on the subject.[4] Renewed efforts to eliminate or reduce this loss in sensitivity have met with considerable success. Life data for 2500 hr of continuous service are shown in Fig. 7. It is expected that life equal to that obtained with the *868* phototube will be obtained with the new tube. However, the life characteristic
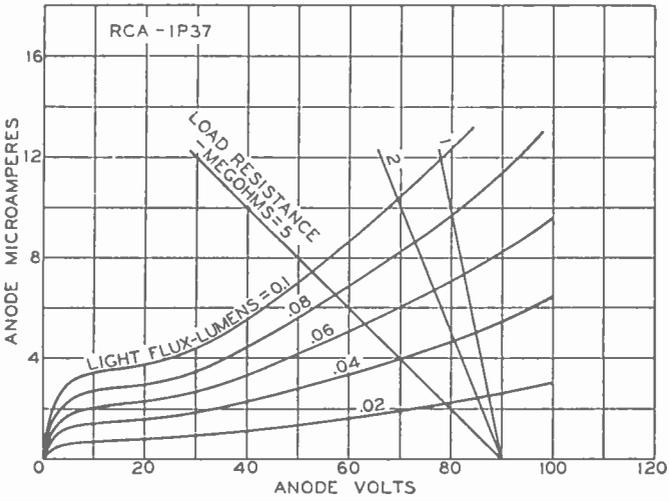
Fig. 6.   Anode characteristics of *1P37*.
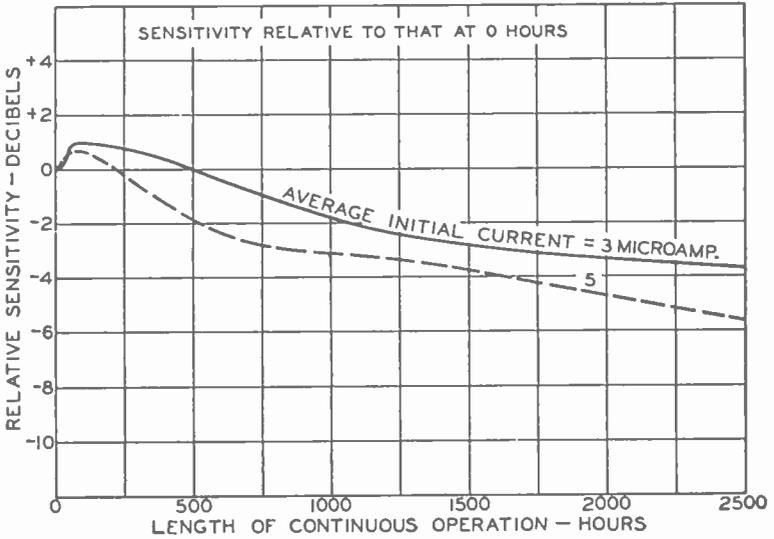
Fig. 7.   Life data on *1P37*.

is somewhat different from that of the *S-1* in that the extremes of variation encountered each time the tube is operated are not so marked. As in other phototubes the life varies with the level of current drawn and the rated maximum average current is set with this in mind. The currents shown are about those drawn with tubes of average sensitivity in a 35-mm optical system with the film in place. Greater currents may be drawn for brief intervals such as the time required for changing reels but limiting series resistance should always be included with a gas-filled phototube to prevent damage from extreme current in the case of voltage breakdown of the gas.

Data taken with the *1P37* phototube, a photograph of which is shown as Fig. 8, with dye image tracks, are the subject for other papers presented before the Society.[5, 6] The structure of the *1P37* has been dictated by the replacement problem of the *868*. The use of the photomultiplier type RCA *931A* in which the *S-4* surface is included may also be of interest for developmental study.

Fig. 8.   Photograph of
            1P37.

REFERENCES

[1] Görisch, R., and Görlich, P.: "Reproduction of Color Film Sound Records," *J. Soc. Mot. Pict. Eng.*, **43**, 3 (Sept., 1944), p. 206.

[2] Glover, A. M., and Janes, R. B., "A New High-Sensitivity Photosurface," *Electronics*, **13**, (Aug., 1940), p. 26.

[3] Kruithof, A. A.: "Time Lag Phenomena in Gas-Filled Photoelectric Cells," *Philips Tech. Rev.*, **4**, (Feb., 1939), p. 48.

[4] Sommer, A.: "The Influence of Illumination on the Fatigue of Photoelectric Cells," *Elec. Eng.*, **17**, (May, 1945), p. 504.

[5] Drew, R. O., and Johnson, S. W.: "Preliminary Sound Recording Tests with Variable-Area Dye Tracks," *J. Soc. Mot. Pict. Eng.*, **46**, 5 (May, 1946), p. 387.

[6] Phyfe, J. D.: "Behavior of a New Blue-Sensitive Phototube in Theater Sound Equipment," *J. Soc. Mot. Pict. Eng.*, **46**, 5 (May, 1946), p. 405.

BEHAVIOR OF A NEW BLUE-SENSITIVE PHOTO-TUBE IN THEATER SOUND EQUIPMENT*†

BY

J. D. PHYFE

RCA Victor Division, Indianapolis, Ind.

Summary.—A new phototube designed to provide optimum performance when used in reproducers with films having standard black-and-white silver sound tracks, or with color films having either dye or edge-treated sound tracks, has been developed in the RCA laboratories.

Some results of laboratory tests and field observations when the new phototube is substituted for the standard red-sensitive 868-type phototube are discussed.

Need has recently arisen for a phototube that could be interchanged with the type commonly used in theater sound heads. The necessity is the result of an apparent trend toward an increasing use of color in motion picture film productions. Also, there is the possibility that a larger percentage of these color films will have dye sound tracks instead of the usual silver tracks. The desired tube should be interchangeable with the red-sensitive type, therefore, and must perform as well when used with films that have the regular silver sound tracks.

Such a phototube has been developed in anticipation of these needs, and is known commercially as the type *1P37*. To date the observed performance of this tube indicates that it accommodates this change-over very well. Mechanically and electrically the *1P37* is interchangeable with the type *868* phototube which has been used in RCA theater sound equipment for more than a decade.

In order to evaluate the merit of this new blue-sensitive phototube for theater use, comparisons of performance have been made in the laboratory with the type *868* red-sensitive tube. In addition a number of the blue-sensitive phototubes have recently been distributed to various theaters throughout the country with instructions to use them in place of the *868* tube, and report their relative behavior.

Too short a time has elapsed since these phototubes were placed in the field for observation, so reports on performance are not complete at present.

Laboratory tests have been made, however, to determine how well they would function in theater sound heads with standard release

---

304

prints, and if there were any adverse operating characteristics which might preclude their being substituted for the *868* tube. Some of the characteristics investigated were:

(1) Relative gain as compared to the *868* tube when used with films having the usual silver sound tracks.

(2) Relative hiss level and microphonics.

(3) Ionization or glow point.

(4) Relative distortion.

(5) Relative hum level when used with raw (unfiltered) a-c on the exciter lamp.

(6) Gain variation with changes in exciter lamp brilliancy owing to line voltage fluctuations.

(7) Ease of balancing output between sound heads by varying the anode voltage.

(8) Frequency response relative to the *868* tube.

(9) Refocusing of the optical system because of different spectral sensitivities of blue-sensitive and red-sensitive phototubes when interchanged.

Results of Laboratory Tests.—The results of these tests are reported in the order in which they were enumerated above:

(1) *Gain Check.*—This was made by running a 1000-cycle loop of standard silver sound track through a regular theater sound head and amplifier channel, and noting the levels. Both types of phototubes were used. A sufficient number of both types were checked to represent an average cross section of sensitivity variation. Exciter lamp and phototube anode voltages were held constant at $8^1/_2$ and 75 v, respectively.

Results of this check showed that the sensitivity of the 2 types of phototubes is practically the same with a slight superiority of the *1P37* tube. It was noted that the *1P37* tube showed somewhat less variation in output from one tube to another than tube *868*.

(2) *Hiss Level and Microphonics.*—This test was made by checking the relative outputs of a group of red-sensitive and blue-sensitive phototubes using a 1000-cycle film loop as a signal source. The film was then removed, the amplifier gain was raised to approximately 130 db and the hiss level was measured. The hiss level was checked both on an output meter and a sound-level meter used in conjunction with a standard 2-way theater loudspeaker system.

Microphonics was then checked by starting the projector motor and noting the output level, as was done in measuring hiss.

The results of these tests revealed no apparent difference in hiss or microphonics when the blue-sensitive phototube was used.

(3) *Ionization or Glow Point.*—No changes in voltage supply are

necessary when changing from the type *868* phototube to the *1P37* type. The same maximum supply voltage limitations exist for both types.

(*4*) *Distortion.*—A constant-frequency film having 80 per cent modulation was run through a standard theater sound head and theater amplifier channel using both types of phototubes, and the rms harmonic distortion measured by means of a distortion factor meter. The measured distortion was found identical for the *1P37* tube and the *868*.

EXCITER LAMP VOLTS A.C.	OUTPUT WITH 1000 CYCLE FILM		HUM OUTPUT		HUM OUTPUT BELOW 1000 CYCLE FILM OUTPUT	
	IP37	868	IP37	868	IP37	868
7.5	17.7	19.1	-0.4	-3.4	18.1	22.5
8.0	20.0	20.4	+2.0	-1.6	18.0	22.0
8.5	21.8	21.8	+3.8	-0.2	18.0	22.0
9.0	23.3	22.7	+5.7	+1.1	17.6	21.6
9.5	24.8	23.8	+7.4	+2.6	17.4	21.2
10.	26.5	24.9	+10.0	+4.0	16.5	20.9

FIG. 1.   Relative film and hum output variations in db—*1P37* versus *868*.   Film removed for hum output measurements.

(5) *Relative Hum Level.*—Using a-c on the filament of a standard 10-v, 7.5-amp exciter lamp, the *1P37* tube showed 4 db more hum when the exciter lamp voltage was adjusted to a normal operating value of 8.5 v. This was based on equal signal outputs for both types of phototubes using a 1000-cycle film loop. Figs. 1 and 2 show relative signal output and hum levels between the *1P37* tube and the *868* when the exciter lamp voltage was varied between 7.5 and 10 v Hum level was measured by removing the film after output measurements were taken.
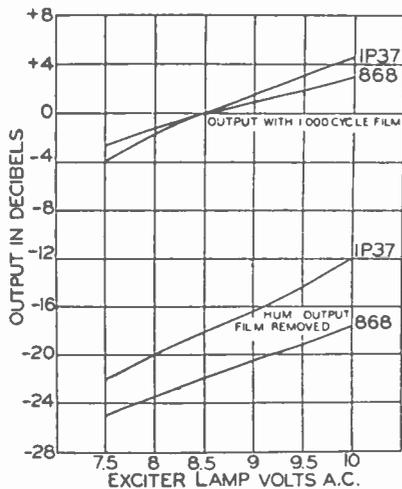
FIG. 2.   Relative signal-to-hum response.

(6) *Gain Variation with Changes in Exciter Lamp Voltage.*—Referring to Figs. 1 and 2, a gain change of 5.8 db is observed for the *868* phototube and 8.8 db for the *1P37* tube, using a 1000-cycle film. This shows a 3-db increase in gain variation for the *1P37*

tube when the exciter lamp voltage is varied between 7.5 and 10 v.

(7) *Balancing Sound Head Outputs by Adjusting Anode Potential of Phototube.*—The adjustments used for balance of the sound head outputs are the same for both the *1P37* tube and the *868*. These are made by control of anode supply voltage.

(8) *Optical System Focus When Using Both Types of Phototube.*— A standard $1^1/_4$-mil slit image optical system was focused for maximum output using a 7000-cycle loop of film and an *868* red-sensitive phototube. Output readings were then taken for both types of phototubes. No observable increase in output was obtainable with the *1P37* tube by refocusing the optical system. This test was then repeated using a 9000-cycle loop of film. The results were identical.

This observation seems to substantiate the assumption that there is no need to refocus a standard $1^1/_4$-mil slit image optical system when the *1P37* tube is used in place of the *868*.

(9) *Relative Frequency Response.*—For this check a standard theater reproducing channel having an optical system with a $1^1/_4$-mil slit image was adjusted for maximum focus using a 9000-cycle loop. No low-pass filter was used since this might have had the effect of making relative output at the higher frequencies. Frequency runs were then made using a calibrated test film which included 31 different frequencies between 30 and 9000 cycles. Response measurements revealed no difference in relative frequency response when the *1P37* tube was substituted for the *868*.

Conclusion.—Reports from the field are awaited in order to better judge the seriousness of the increased hum, and greater variation in signal level with changes in exciter lamp voltage obtained with the *1P37* tube. No attempt will be made at this time to evaluate the seriousness of these 2 points.

It is felt, however, that the increased hum might be satisfactorily compensated by modification of the 120-cycle hum filter in those installations which operate with raw a-c on the exciter lamp.

For those installations which operate with d-c exciter lamps and which have some form of regulation of the exciter lamp voltage, it does not appear that the latter point would become a problem.

In the meantime, however, the tubes that have been substituted are working very well, indicating that no differences in operating characteristics have been observed, or else are not of sufficient magnitude to justify an immediate report.

AN INFRARED IMAGE TUBE AND ITS MILITARY APPLICATIONS*†#

BY

G. A. MORTON AND L. E. FLORY

Research Department, RCA Laboratories Division,
Princeton, N. J.

Summary—The military value of the security obtained by the use of infrared for nocturnal vision was recognized even before the entry of the United States into World War II. A program for the development of infrared viewing devices employing electron image tubes was consequently set up by the National Defense Research Committee. Before the close of the war a number of types of infrared telescopes had been manufactured in quantity and had seen service in fairly large numbers.

The 1P25 image tube is the essential element in the infrared electron telescope, and serves to convert the invisible infrared image into a visible image. The tube contains a semi-transparent photocathode which is processed to be sensitive to infrared radiation, and an electron lens for imaging the electrons from the photocathode onto a fluorescent screen which becomes luminous upon bombardment by electrons. When an infrared image is focused on the photocathode, a visible reproduction of this image is formed on the fluorescent screen.

Basically, the infrared telescope consists of the image tube, an objective for forming the infrared image on the photocathode and an ocular for viewing the reproduced image. Associated with the telescope is a battery operated vibrator power supply which furnishes the 4000 to 5000 volts and the several intermediate voltages required by the image tube.

A variety of types of telescopes was developed and produced for a number of different applications. These included a signalling telescope employing a large aperture reflective optical system as objective, the Sniperscope which is a carbine-mounted telescope and infrared source permitting aiming and shooting in complete darkness and the Snooperscope composed of the same infrared units mounted on a handle for short range reconnaissance work. Binocular telescopes, helmet-mounted driving and flying instruments, long-range reconnaissance units and other special night-seeing devices were also developed in the course of this project.

EVEN before the entry of the United States into World War II, it was recognized that many military operations would require the secrecy afforded by complete visual darkness. Therefore, the National Defense Research Committee, under Army and Navy directives, undertook the development of infrared viewing devices employing electron image tubes and an investigation of the applications

---

* Decimal Classification: R138.31 × R800 (621.375).

† This paper is based in whole or in part on work done for the National Defense Research Committee under Contracts OEMsr-169 and OEMsr-440 with Radio Corporation of America.

Reprinted from *RCA Review*, September, 1946.

of such devices. This work was carried out by these laboratories.

A variety of infrared telescopes was developed suitable for different tasks and a number of types saw considerable service during the war. Among the most widely used were the Navy infrared signalling equipment and the Sniperscope and Snooperscope procured by the Army. Figure 1 illustrates an infrared telescope, while Figure 2 shows the laboratory prototype of the Sniperscope.

Basically, all of these telescopes consist of an objective for forming an infrared image of the scene being viewed upon the photosensitive cathode of the image tube, the image tube itself, and an ocular for viewing the reproduced image. The general form of the electron tele-
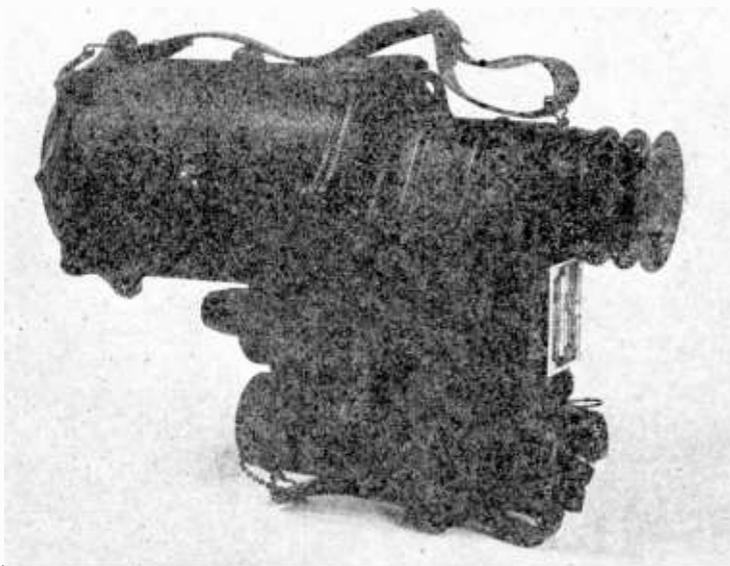
Fig. 1—Infrared telescope.

scopes using refractive and reflective optics is shown schematically in Figure 3.

The fundamental component of these infrared telescopes is the electron image tube. This tube consists of a semi-transparent photocathode processed so that it has high sensitivity in the infrared portions of the spectrum, a fluorescent screen and an electron optical arrangement for focusing the electrons onto the screen.

In undertaking the design of these instruments and tubes, the requirements of mass production as well as those relating to the particular application, were taken into consideration. As a result, the U. S.
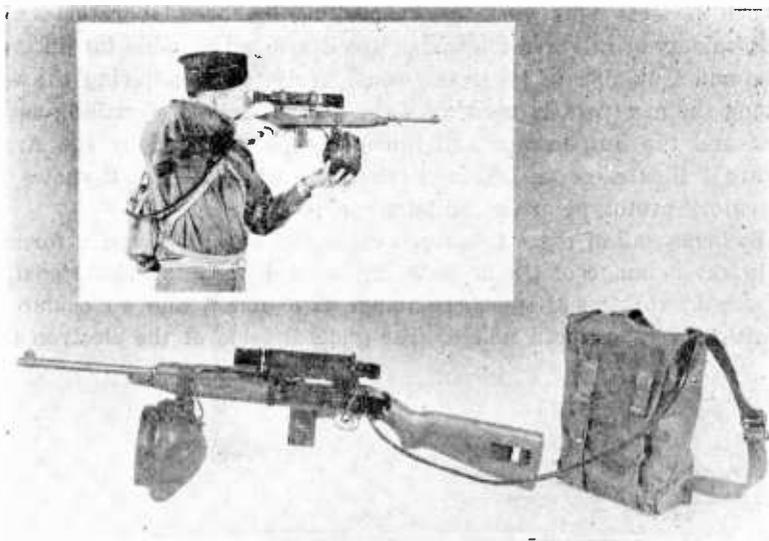
Fig. 2—Laboratory prototype sniperscope.

Armed Forces were able to obtain these instruments in far larger quantities than could either the Germans or the Japanese whose instruments were not suitable for quantity production.

INFRARED TELESCOPES
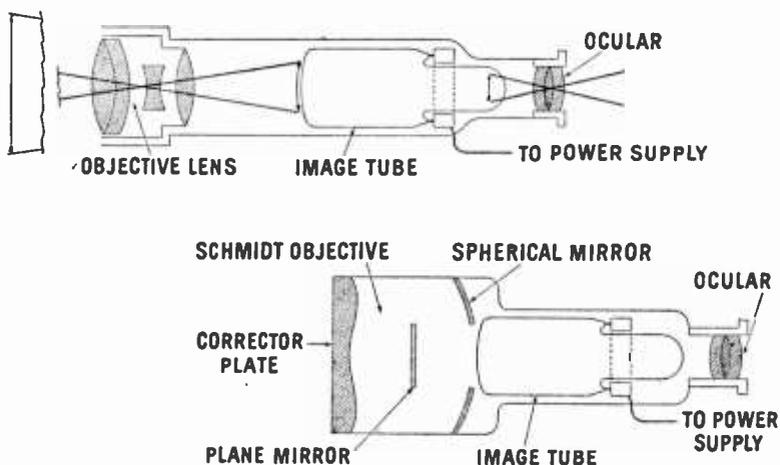
Fig. 3—Schematic diagram of two types of infrared telescopes.

The Image Tube

An investigation of electron imaging dating back to the early 1930's had resulted in successful image tubes.[1,2] However, the requirements placed on an image tube for military applications were so very different from any heretofore encountered that it was necessary to re-examine the entire subject again.

The most important considerations were, of course, sensitivity of the tube and perfection of the image. These are properties of the activation, phosphor efficiency and electron optics, and will be discussed in greater detail. A second very important consideration was that the tube be of such a form that it could be produced quickly in fairly large numbers. Finally, it should be so designed that a single type of tube could be used for all of the various applications envisaged.

Tube dimensions selected as being the best compromise between the very small size which would be desirable for portable instruments, and the larger tube suitable for fixed units, were 4½ inches length and 1⅞ inches maximum diameter. This size was found convenient from the production standpoint, and useful in a wide variety of instruments. Consideration of power supply design, cable insulation and tube stability dictated an overall voltage range of 4000 to 6000 volts.

The first decision which had to be made concerned the most practical way of imaging the electrons from the cathode onto the fluorescent screen. There are essentially three systems which may be used, namely:

    (1) uniform field between cathode and screen;

    (2) magnetic lens; and

    (3) electrostatic lens.

The first was rejected because of the close spacing between cathode and screen and high field strength necessary in the vicinity of the cathode. This makes the activation difficult and the tube prone to cold discharge. Also, the image produced in this way is erect where preferably it should be inverted. Magnetic focusing was also rejected from the standpoint of weight and complexity, and because of the difficulty of obtaining an inverted image.

An electrostatic lens system is capable of a sharp, clear image over a wide range of magnifications. The image is inverted making it unnecessary to use an inverting ocular for viewing the reproduced image. It is necessary to curve the photo cathode in order to produce an undistorted image over a large angular field. Where a reflective optical system is used as objective, the curvature of the cathode can be made

[1] V. K. Zworykin and G. A. Morton, "Applied Electron Optics", *Jour. Opt. Soc. Amer.*, Vol. 26, No. 4, pp. 181-189, April, 1936.

[2] G. A. Morton and E. G. Ramberg, "Electron Optics of an Image Tube", *Physics*, Vol. 7, No. 12, pp. 451-459, Dec., 1936.

to match that inherently present in the image surface of these optics. It is, however, sometimes necessary to use an optical field corrector lens when an ordinary refractive objective is employed, if the field of view is to be flat. However, since the electrostatic lens is also free from the objections mentioned in connection with the magnetic and uniform field systems, it was selected as the most satisfactory for the purpose.

The magnification of the image tube has an important bearing on its performance. This is because the brightness of the reproduced image varies inversely with the square of the magnification. Thus, if a telescope with a given overall magnification employing an image tube

IMAGE TUBE
(IP 25)
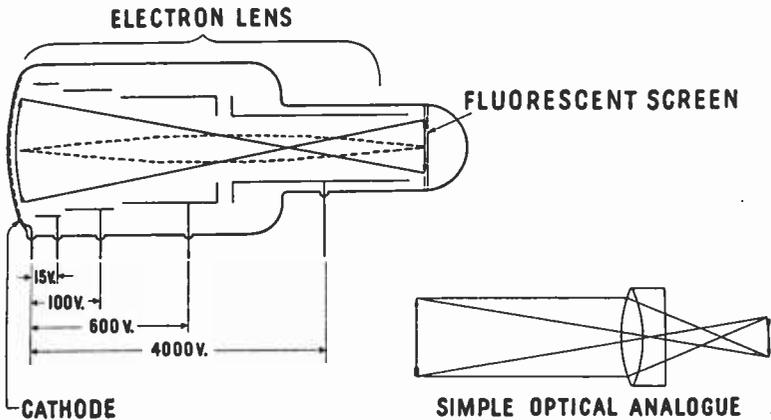
Fig. 4—Schematic diagram of 1P25 Image Tube.

with unity magnification and an X5 ocular is compared with one using an image tube with magnification ½ and an X10 ocular, the image in the latter will be four times brighter. However, for a given size of image tube and angular field of view, the magnification cannot be decreased indefinitely because as the power of the ocular increases the size of the exit pupil decreases until a point is reached where the pupil of the dark adapted eye is not filled. Beyond this point, the brightness of the retinal image does not increase with decreasing magnification of the image tube. For many applications, it is also essential that the exit pupil be much larger than the pupil of the eye, so that the observer's eye does not have to be located too exactly with respect to the

instrument. Experiment showed that for a tube the size of the image tube under discussion, the magnification should not be less than one-half. Figure 4 illustrates schematically the construction and action of the image tube adopted.

ELECTRON OPTICAL CONSIDERATIONS

The electron optical system of the image tube consists essentially of a strong main lens as the principal imaging means and a series of relatively weaker correcting lenses between the cathode and main lens. The potential distribution along the axis of the tube is shown in the upper portion of Figure 5. Two electron paths, one of an electron
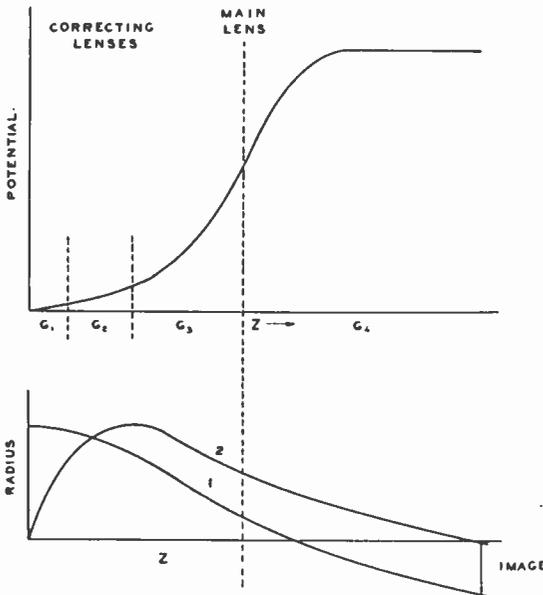
Fig. 5—Potential distribution and electron paths in the Image Tube.

originating from the cathode on the axis of symmetry with radial initial velocity, the other originating off the axis with no initial velocity, are illustrated in the lower part of the figure. These two paths are sufficient to determine the first order imaging properties of the system.

A detailed theoretical study of this type of system leads to the following conclusions:

(1) Curvature of the image field and pincushion distortion can only be eliminated by the use of a curved cathode (or a radial potential gradient on the cathode).

(2) Curvature of the image surface and astigmatism limit the off-axis definition.

(3) Chromatic aberration due to the spread of initial velocities of the photoelectrons establishes the limit of resolution at the center of the image.

(4) Spherical aberration and coma play a negligible role in limiting the definition.

The radius of curvature of 2.38 inches selected for the cathode was a compromise between that required for a flat electron image surface and optical considerations of the objective. With this curvature very little pincushion distortion remained, and a definition of 350 lines (television nomenclature) or better could be obtained near the margins of the picture.

At the center, the diameter ($\triangle$) of the circle of confusion due to chromatic aberration is given approximately by:

$$\triangle \simeq 2m \, V/E \qquad (1)$$

where E is the gradient near the cathode and V the initial electron energies in electron volts. Evaluating this from the gradient known to exist in the tube and from the initial velocities expected near the infrared threshold, the limiting definition at the center is 2000 or more lines. Definitions of 1000 lines were realized in laboratory tubes, and of 450 lines or better in production tubes. In general, the difference in definition between the theoretical estimated definition and that achieved in practical tubes is due to misalignment of the electrodes, inhomogenities in the photo-cathode and granularity of the fluorescent screen.

Photo-Electric Cathode

The photo-sensitization of the cathode is one of the critical steps in the preparation of the image tube. Research to date has lead to the conclusion that a complex surface involving caesium, oxygen and silver yields the highest infrared response of any of the surfaces yet studied. This surface is formed by evaporating a thin layer of silver on the cathode disk, oxidizing it completely, then adding alternately silver, caesium and silver while subjecting it to an appropriate thermal treatment. The completed surface is semi-transparent so that, when illuminated from the outside, electrons are emitted from the inner surface. The photoemission from a well activated surface of this type will be 30 to 50 microamperes per lumen for whole light (visible + infrared) from a tungsten source at a color temperature of 2870 degrees Kelvin.

Figure 6 illustrates the spectral response of this type of emitter.

FLUORESCENT SCREEN

The requirements of the fluorescent screen are the following:

(1) It must have a high efficiency of conversion of electron energy into visible light of a color suitable for scotopic vision.

(2) It should have a fine grain structure capable of giving high definition.

(3) Its time constant must be short so that moving images do not blur.

(4) It should be inert to the chemical action of caesium.

Synthetic willemite was found to satisfy these requirements fairly well, although its phosphorescent decay time is somewhat longer than
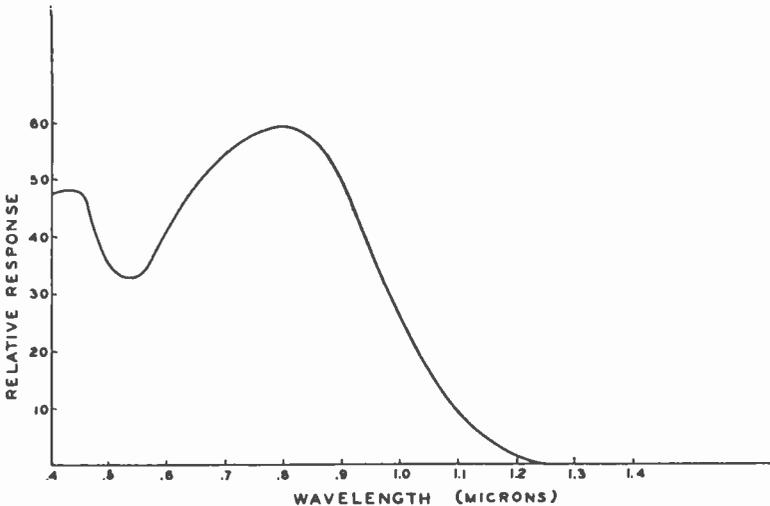
Fig. 6—Spectral response of 1P25 Image Tube.

might be desired. In spite of its shortcoming as far as persistence is concerned, it was selected as the phosphor most suitable for the 1P25 because of its availability and ease of handling together with its chemical stability.

The efficiency of this phosphor in the vicinity of 5000 volts is between 1 and 3 candles per watt. Its color is green or yellow-green which is quite satisfactory from the standpoint of scotopic vision. With a little care, the grain and aggregate size can be made small enough so that the screen does not limit the definition of the image. The decay characteristic of the material cannot be expressed by a single time constant. However, for the brightness involved in such applications of the tube

as night driving, reconnaissance, and sniperscopes, the relation:

$$B = B_o \varepsilon^{-60t} \qquad (2)$$

where t is the time in seconds after excitation ceases and B_o is the brightness at $t = o$ is entirely adequate. The expression for phosphorescent decay indicates that the image brightness falls to 10 per cent of its initial value in 0.04 secs. While rapid enough for most purposes, it causes some loss in definition for rapidly moving objects. At very low brightness levels, the decay becomes less rapid than is indicated by this expression. This long low-level afterglow is of consequence in the detection of an infrared marker and signal light near the visual threshold.
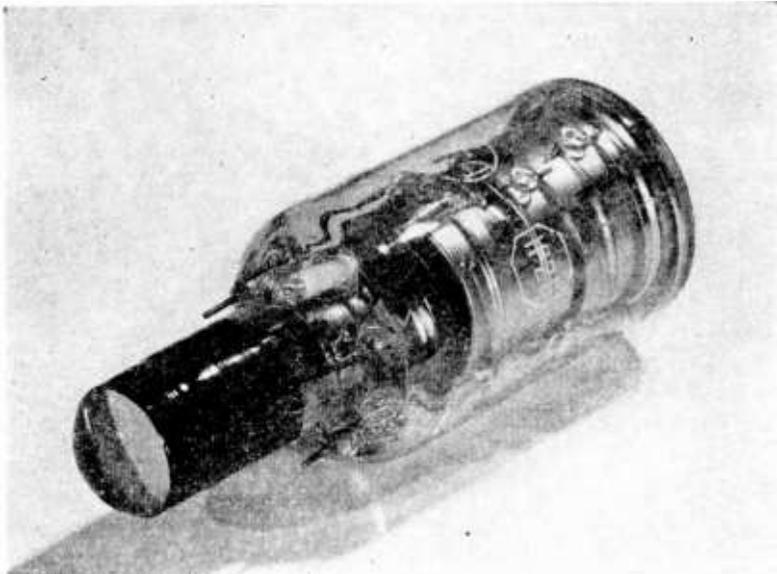
Fig. 7—1P25 Image Tube.

IMAGE TUBE PERFORMANCE

The performance of the 1P25 image tube may be summarized as follows: The light output per lumen of light incident on the photocathode, or conversion, is in the neighborhood of 0.5 to 1 lumen. In conjunction with an infrared filter, the conversion is reduced by the corresponding filter factor. It has been customary to express this filter factor in terms of the ratio of the image-tube response for whole light from a given incandescent source at color temperature 2870 degrees Kelvin to the response from the same source when filtered.

The central definition of the image is 450 lines or more and the

peripheral definition 300 lines. This definition is such that for most brightness levels encountered in practice, the eye rather than the instrument is the limiting element.

A photograph of the finished tube is shown in Figure 7.

During the later stages of the research program, a number of new types of tubes were developed to meet special problems. One of these was a single-voltage tube, contained in an envelope identical to that of the 1P25, but requiring no intermediate or focusing voltages. This tube is interesting in that it employed an electron optical system with an unconventional departure from circular symmetry. A second tube operated at an overall-voltage of 16 kilovolts employing a multiple lens anode.

INSTRUMENTS EMPLOYING THE 1P25

During the course of the investigation, many different types of infrared instruments were developed employing the image tube. The number is so large that only a small fraction of them can be described in this paper. Therefore, a few representative instruments have been selected which will be described and their performance indicated.

Signalling Telescopes

One of the widest and at the same time most exacting use of the electron telescope is for the observing of infrared signal and marker lights. Here, since the object observed is an unresolved luminous point, the considerations involved in determining the sensitivity of the instrument are quite different from the case of an extended image. The two primary optical factors are the area of the objective and the magnification, while for the image tube the conversion and background only are involved. It will be noticed that the f-number of the objective and magnification of the image tube do not affect the sensitivity. However, if a lens with a large f-number or an image tube with high magnification is used, the angular field of view will be small which is undesirable for a marine signalling or search instruments. For these reasons the signalling telescope was designed with an objective having a short focal length and large aperture. The only practical way of achieving such a system is by the use of reflective optics, as illustrated in the lower part of Figure 3. The corrector plate and spherical mirror were of transparent plastic, assembled as a unit in a plastic barrel. The system was arranged so that the image was folded back by means of a plane mirror onto the cathode of the image tube. The focal length of this objective was 2.4 inches and its effective f-number was about 0.9. The image on the fluorescent screen was viewed through an X11

ocular. A hemisphere, which was an integral part of the image tube bulb, was a component of the ocular system. The telescope and power supply were assembled in a light, weatherproof and hermetically sealed magnesium casting. A discussion of power supplies for this type of instrument will be postponed until a later section.

A much larger reflection-type telescope was also developed having a 7-inch focal length and approximately the same effective numerical aperture. This instrument was very much more sensitive as a signalling telescope but had a much smaller angle of view. Because this instrument was also designed for reconnaissance, its optical focus was made variable through an adjustment which moved the plane mirror in and out.

Reconnaissance Telescope

With the exception of the large reflective-type telescope mentioned in the preceding paragraph, the reconnaissance instruments were in general of the small portable variety. In order to give them greater depth of focus than could be obtained with a reflective optical system, these telescopes employed refractive optics with f-numbers down to about 2.0. With an f/2.0 lens and an image tube having a conversion of 1.0 the ratio of brightness of the image of an object illuminated with whole light as seen through the telescope to brightness as seen directly is about 0.10. This ratio is reduced by the appropriate filter factor when a filtered source is employed.

In their simplest form, these instruments consist of a barrel (usually of mu-metal or other high permeability alloy for a shield) containing the image tube, to which are affixed the objective and ocular, both in focusing mounts. Tests indicated that objective focal lengths in the range 2½ to 3½ inches and ocular magnification of X8 to X12 were most satisfactory. For example, when used as the basis of a driving telescope, as will be discussed below, an instrument with a 2½-inch objective and an X8 ocular giving an overall magnification of unity and a 24-degree field of view gave best results, while for devices such as the Sniperscope and Snooperscope a 3½-inch focal length objective and a X8 or X12 ocular were to be preferred.

Power for the image tube was supplied through a three wire, insulated cable providing ground, the overall voltage and the variable focusing voltage. A resistance voltage divider at the image tube socket provided the other voltage steps for the 1P25.

Other instruments were designed with the power supply an integral part of the telescope. One unit employs a 2½″ focal length f/2.0 plastic

objective and an X12 ocular. A model is illustrated in Figure 8. Its size and weight is only about one third that of the telescope shown in Figure 1.

For general purpose observation, these simple in-line monocular telescopes served as very useful tools. For example, this type of instrument was frequently carried by an observer during night driving. Also, it was used to supplement the large reflective-type telescopes as general orientation instruments, and for many other supporting operations. Under these circumstances, the illuminator providing the infrared radiation was a separate unit over which the user of the telescope had little or no direct control.
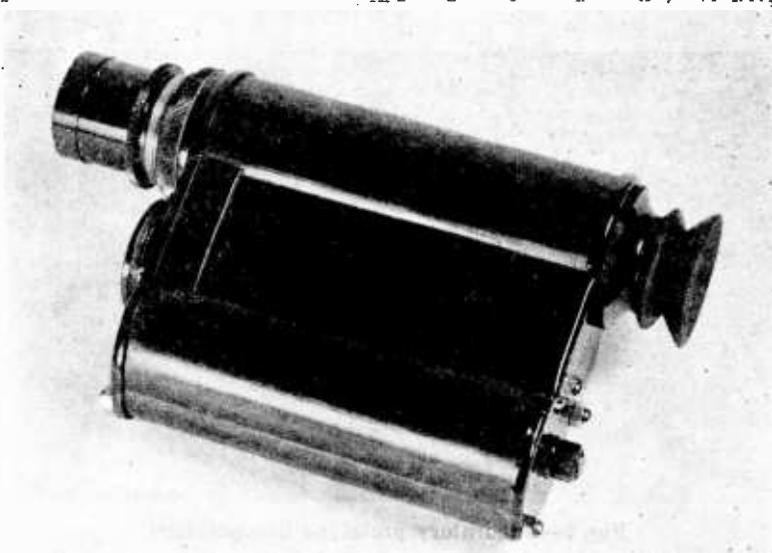
Fig. 8—Experimental telescope.

Frequently it is advantageous to have a portable light source to use in conjunction with the telescope. Therefore, a study was made of instruments involving source-telescope combinations. In particular, two instruments of this class were developed, namely, a monocular telescope and a light source mounted on a handle for relatively short range reconnaissance (see Figure 9) and a similar telescope and source mounted on a carbine in such a way that the telescope could be used for aiming in complete visual darkness. These instruments were christened Snooperscope and Sniperscope respectively and were later

named Molly and Milly by members of the Armed Forces.

The telescopes used in the laboratory prototypes of these units were essentially the same as the monocular telescope described in earlier paragraphs. The objectives were 3½-inch focal length f/2.0 lenses while an X8 ocular was used for viewing the screen. A chevron was placed on the surface of the field-corrector lens to serve as the aiming index for the Sniperscope. This chevron was accurately aligned with the direction of fire of the piece. By placing the aiming index at the objective, distortion or deflection of the electron image had no effect on the accuracy of aiming.

Tests were made to determine whether the telescope, including the image tube, was sufficiently rugged to withstand the rough usage involved in this application. No particular difficulties were encountered
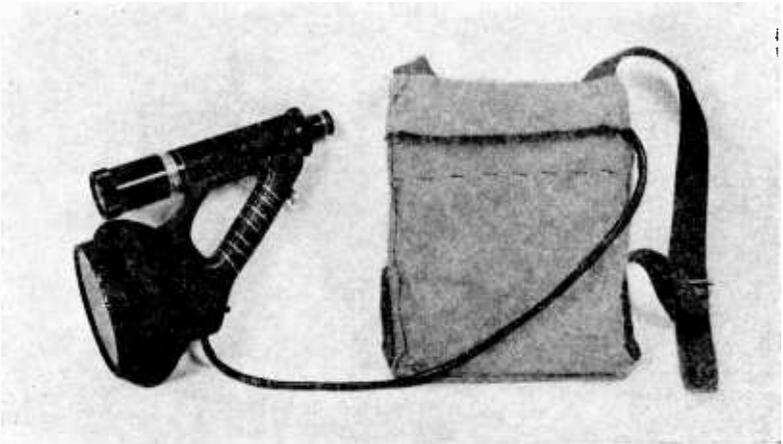
Fig. 9—Laboratory prototype Snooperscope.

in the case of the Snooperscope. In the Sniperscope, a certain percentage of tubes were found to fail as a result of the shock of firing, due to minute particles of the phosphor becoming dislodged from the fluorescent screen and settling on the lens electrodes and causing flashing in the tube because of cold discharge. This made it necessary to shock-test production tubes before employing them for this purpose.

The selection of the size and form of light source was a result of compromises in a number of directions. These included angular field, range, operating life for the allowable battery weight, and considerations of security. The source chosen for the laboratory models was a special General Electric sealed beam lamp with a 12 to 15 degree spread and a maximum beam candle power of 80,000. Various infrared filters

were used including Corning 2540 heat transmitting glass and the Polaroid XRX series.

The high voltage power supply and storage batteries were carried in a knapsack with a cable from it to the instrument carrying both the image tube voltages and the current for the infrared source. The design of the power supply is discussed in a later section.

The weight of the telescope and source was about 5 pounds while the power supply and batteries weighed approximately 13 pounds. The unit was capable of 3 to 4 hours continuous operation before it was necessary to recharge the batteries. The Snooperscope and Sniperscope were manufactured on a fairly large scale, the production design being carried out by the Engineer Board of the Army for the Corps of Engineers.

Driving Instruments

In some applications, particularly vehicle driving, it is advantageous to have binocular vision. It is interesting to note that while the observer feels a very definite need of being able to use both eyes it makes very little difference whether or not he has stereovision.

The first experimental driving instruments were in the form of a single large barrel carrying the objective and eyepiece, and enclosing an image tube which was much larger than the 1P25. The eyepiece was so arranged that the virtual image of the fluorescent screen was at infinity and the observer saw this image with both eyes.

This type of driving telescope was found to be generally quite satisfactory, but suffered from two serious drawbacks. It was quite large and occupied considerable space in the vehicle and it was difficult to use on short turns.

To overcome these difficulties, a small binocular instrument was developed using two 1P25's. This instrument, illustrated in Figure 10 gave the observer true stereovision. The binocular consisted of a pair of in-line telescopes mounted parallel to one another by means of hinges so that the interpupillary distance could be adjusted to fit the user. In order to obtain satisfactory register of the images seen by the two eyes, it was necessary to provide means for moving the two images relative to one another. This was accomplished by mounting the ocular lenses in such a way that their axes were slightly displaced with respect to the axes about which the ocular fittings could be rotated. When the eyepiece fittings were turned the two virtual images seen by the observer moved in circles about two different centers. The points of intersection of these circles are the points of register of the images.

Separate focus of the two objectives and two oculars was provided. The power supply was designed so that the electrical focus of the two image tubes was also independent.

These binoculars, with a suitable head rest, served as excellent driving telescopes, when supported on a pivoted arm in front of the driver. As designed the instrument not only gave the operator use of both eyes but also permitted true stereovision.

Some experiments were undertaken in helmet mounting the in-line binoculars. However, even when counterbalanced to neutralize the forward torque of the instrument, the moment of inertia was rather high which made it awkward to handle.

To overcome this, an investigation was made on a series of helmet-
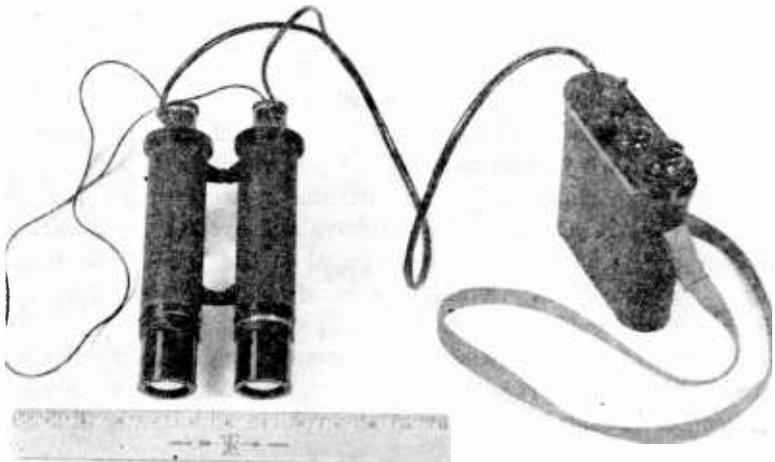
Fig. 10—Infrared In-line Binocular.

mounted instruments. Figure 11 illustrates an early right-angle peri-scopic unit. The results of these tests were so encouraging that the development of a light-weight, Z-shaped binocular suitable for night flying and driving, was undertaken in collaboration with the Johnson Foundation of the University of Pennsylvania. The completed instrument is illustrate in Figure 12. The unit was made of aluminum, and plastic optics were used throughout (except for the ocular lenses) to reduce the weight to a minimum. The folding was accomplished by means of plastic prisms. Again, independent optical and electrical focusing adjustments were provided for the two sides, the electrical focus being controlled by means of potentiometers mounted at the bases of the telescope barrels. Register of the image was effected by

means of a pair of rotatable ring magnets mounted on each barrel just below an inner mu-metal shield which surrounded the image tube from cathode to main lens. The resultant field of these ring magnets, which could be varied in intensity and direction by rotating the rings, made it possible to deflect the electron images into exact alignment.

These helmet telescopes appeared to be a very adequate solution to the problem of infrared night driving, and were on the verge of going into production when the war ended.
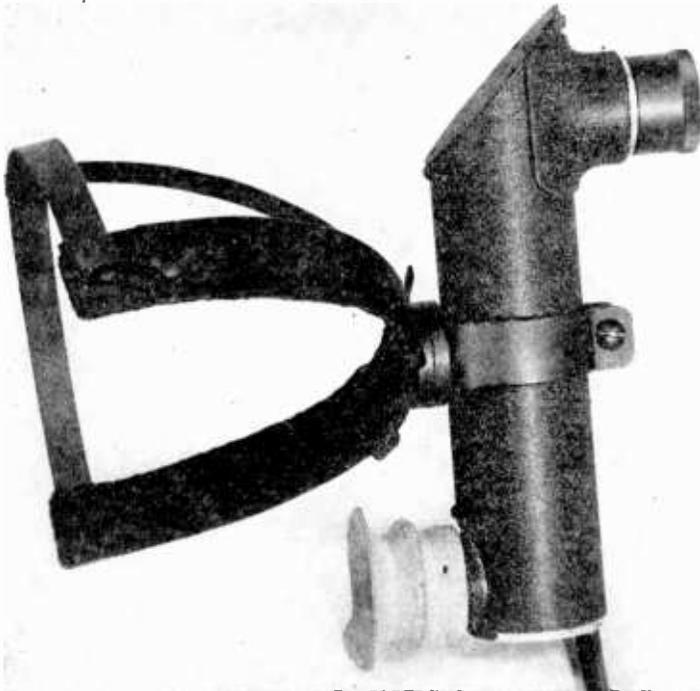
Fig. 11—Periscopic Helmet Monocular.

HIGH VOLTAGE POWER SUPPLIES

The 1P25 image tube requires a rather high voltage for its operation. Since portability was one of the aims of the development, it was essential that the power supplies be small and operate from a small primary battery source. At the same time, the battery life had to be above a certain minimum if the instrument was to be practical.

The design of a power supply meeting these requirements is possible only because of the fact that while a 4000 to 6000-volt output is

required, the actual power needed is very small. The tube itself requires only a fraction of a microampere of current even under bright light conditions. The total power output required to supply the tube and the voltage divider necessary for the various focusing electrode voltages of the 1P25 is on the order of a tenth of a watt.

The only practical available means of converting the low voltage
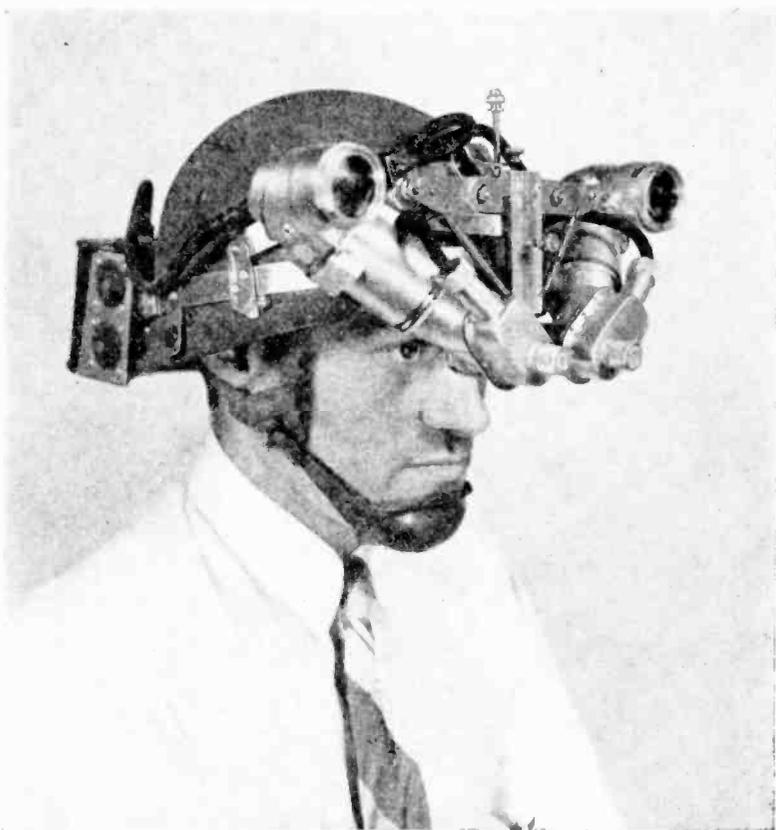
Fig. 12—Type Z Helmet Binocular.

from the batteries to the high potential necessary to actuate the image tube is a vibrator-transformer-rectifier combination. A typical vibrator power supply is shown in Figure 13. It differs from the conventional vibrator power units used in battery operated radios, in that, due to the low power requirement, use can be made of the relatively high voltage peaks appearing across the primary of the transformer when the magnetic field collapses as the primary circuit is broken by

the vibrator. In addition the primary is tuned to resonate with the natural period of the secondary to obtain maximum transfer of energy. By this method an effective primary voltage of ten to twenty times the battery voltage is realized. This makes possible a great reduction in the size of the transformer required.

A standard automobile type vibrator was used because of ease of procurement. The frequency was of the order of one hundred interruptions per second and the power consumed was about 0.2 watt. A conventional rectifier circuit and capacity filter was used employing the special rectifier described later.

The design of the transformer was necessarily a compromise between light weight and efficiency. A light-weight transformer with
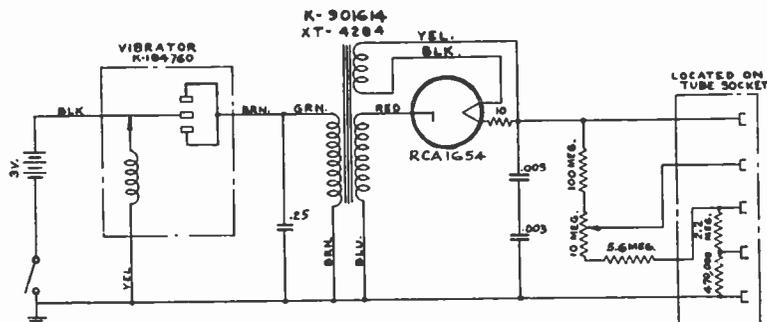
Fig. 13—Basic circuit of vibrator power supply.

somewhat lower efficiency means a larger battery or shorter battery life, while a heavier transformer will give a higher efficiency due to lower core losses. As a result, the optimum transformer is usually a design unique to the particular application. For an overall efficiency of 10 per cent, including vibrator power and rectifier filament, at an output of 4000 volts, the weight of the transformer is on the order of 20 ounces per watt of output. This means that in order to supply one-tenth watt at 4000 volts a two-ounce transformer would be required at an input of one watt. At somewhat higher power ouputs the efficiency may run as high as 20 per cent, since the power taken by the vibrator and rectifier will be constant.

A typical transformer design may be approximated by making the following assumptions:

Vibrator frequency—100 per second
Time of contact—.005 second
Battery voltage—3 volts
Maximum battery current (peak)—3 amperes
Peak output voltage—4000 volts.

Under the type of operation required, the secondary current is small and most of the power dissipated in losses. Therefore, as a first approximation the effect of the secondary circuit on the primary may be neglected except as it affects the resonant frequency.

If the time constant of the primary is made equal to the contact time, then

$$\frac{L}{R} = .005$$

Since the maximum current is to be 3 amperes and the voltage is 3 volts, the primary resistance is 1 ohm. Consequently,

$$L = .005 \text{ Henries.}$$

If the decay time of the primary current is now assumed to be one-tenth of the contact time as determined by the resonant frequency of the secondary to which frequency the primary is tuned, then

$$e = L \frac{\Delta i}{\Delta t}$$

$$= .005 \frac{3}{.0005} = 30 \text{ volts.} \tag{3}$$

Since the peak output is to be 4000 volts, the turns ratio of the transformer becomes

$$\frac{4000}{30} = 133$$

To obtain the necessary primary inductance requires about 100 turns. Thus 13,300 turns will be required for the secondary.

The direct current from the 3-volt battery under these conditions is about 0.5 amperes with the secondary delivering about 50 microamperes. A core cross section of 0.25 square inch ordinary silicon steel with a 0.010 air gap was found sufficient.

Since no rectifier of small size and low filament power consumption was available, a special tube was developed. This tube, shown in Figure

14, is now in production as the 1654 and in the special circuit shown will deliver 100 microamperes at 5000 volts.

A typical example of this form of power supply is shown in Figure 15. This supply delivers 0.15 watts at 4000 volts with an input of one watt. The total weight including the battery, which will operate the instrument for 2½ hours, is 2½ pounds.

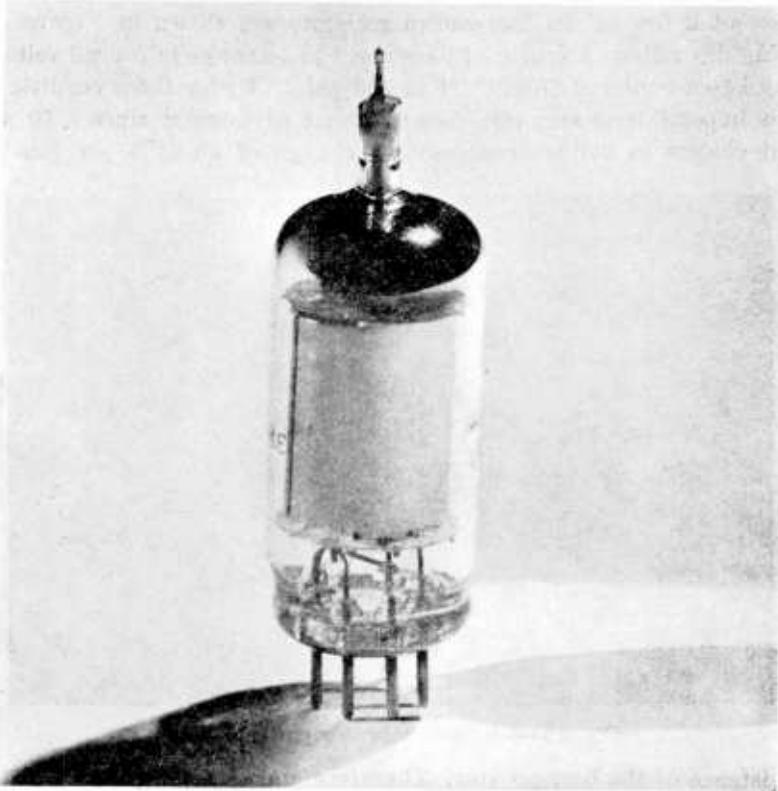A high degree of stability of the overall voltage is not essential but the ratio of voltages on the various electrodes must be maintained to

Fig. 14—1654 Rectifier Tube.

keep the image in good focus. The regulation of the power supply is not important since the load is essentially constant. As a matter of interest the equivalent resistance of the power supply shown in Figure 15 is approximately 40 megohms.

Since the overall voltage varies considerably as the batteries discharge and since the instruments may be subjected to wide ranges of temperature, behavior of the components of the voltage divider as

regards temperature and voltage was a matter of considerable concern. It would be highly desirable to be able to maintain the proper voltage ratios over the range of temperatures and voltage encountered in the field. However, this is not always possible and occasional refocusing may be necessary although the variations can be greatly reduced by proper choice of components in order to balance their characteristics.

All of the available high value resistors (50 megohms or more) show considerable change of resistance with voltage. The voltage characteristics of a few of the best-known resistors are shown in Figure 16. Using dry cells as a source of power, a 2 to 1 change in overall voltage may be encountered from start to end point. Under these conditions, it is impossible to maintain focus without adjustment since a 50 per cent change in voltage represents a change of about 5 per cent in
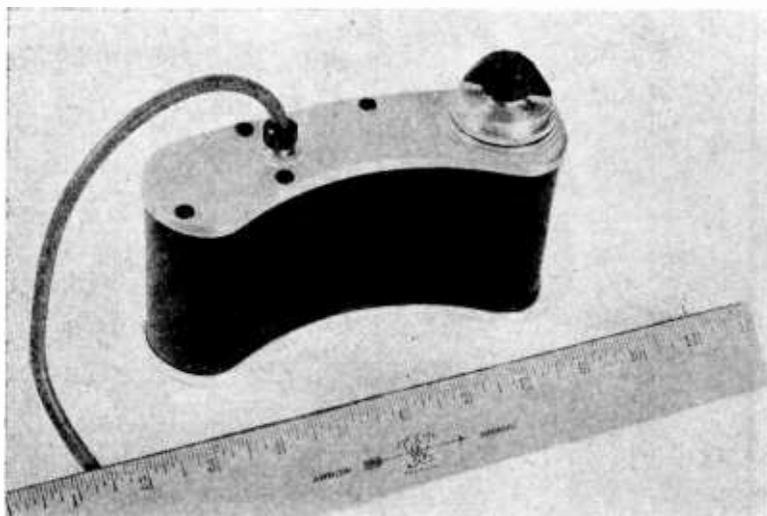
Fig. 15—5-kilovolt vibrator power supply.

resistance of the best resistor. Therefore, unless compensation can be provided, it is necessary to refocus as the batteries deteriorate. In the case of storage batteries, about 10 per cent change in voltage may be expected over the operating life. This produces a negligible change in resistance of the No. 5 resistor and no refocusing is necessary.

Most resistors have a high temperature coefficient and in order to design a voltage divider which will maintain the tube focus independently of temperature, it is necessary to select components which either have the same coefficient, so that the ratio remains the same over the temperature range, or which have coefficients which tend to compensate

for each other. Variation of resistance with temperature for a variety of resistors is shown in Figure 17.

In making up a divider, many combinations of resistors tending to compensate are possible. From the curves in Figure 17, two combinations were selected and the characteristics of the dividers plotted in Figure 18. In both cases, the G_3 voltage remained essentially constant over the entire temperature range, the small variations being in such a direction as to compensate for the variation in G_2. With divider No. 1 adjusted for focus at 20 degrees Centigrade, the voltage on G_2 remains in the region of good focus over the range from —10 degrees Centigrade to + 60 degrees Centigrade. Divider No. 2 remains in focus
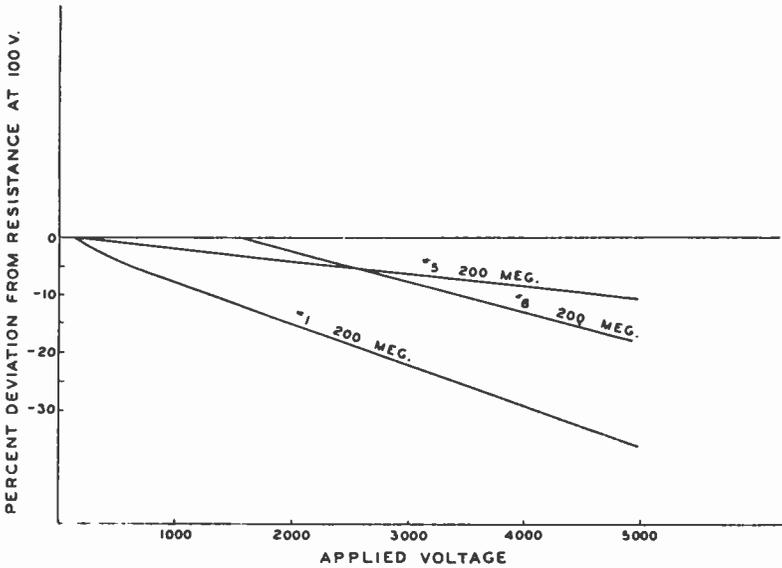
Fig. 16—Voltage characteristics of resistors.

from —40 degrees Centigrade to +75 degrees Centigrade. Therefore, using storage battery supply and selected components for the voltage divider, it is possible to build an instrument which will not require electrical focusing in the field under the range of conditions usually encountered.

The Type S_2 supply shown in Figure 19 is an interesting modification of the vibrator power supply. This arrangement is similar to the conventional voltage doubler circuit except that the two halves of the doubler are brought out separately. In this way, it is possible to place a voltage divider across one side without disturbing the other. In these vibrator supplies, the alternating-current wave is non-symmetri-
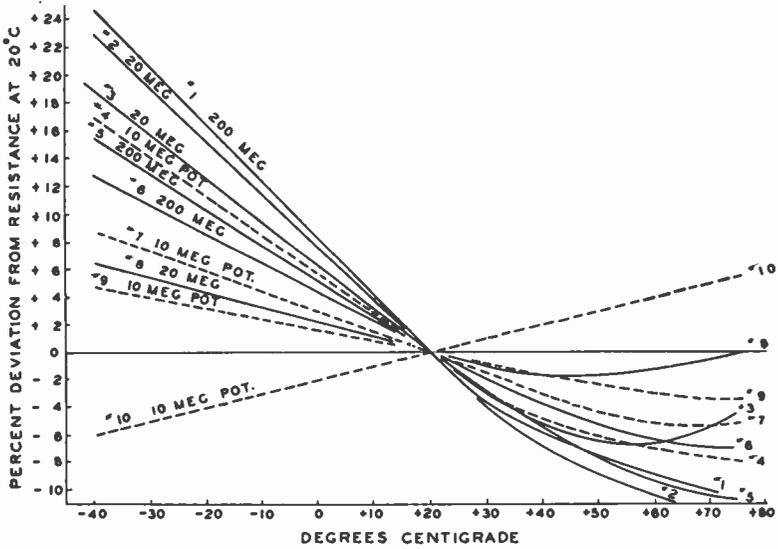
Fig. 17—Temperature characteristics of high-valued resistors.

cal, being in the nature of a damped oscillation, so that in the circuit shown, the voltage across the high voltage section, which is determined by the first loop of the wave, is about 4000 volts while the voltage in the opposite section, determined by the second or negative loop, is about 1000 volts. Therefore, by putting the voltage divider across
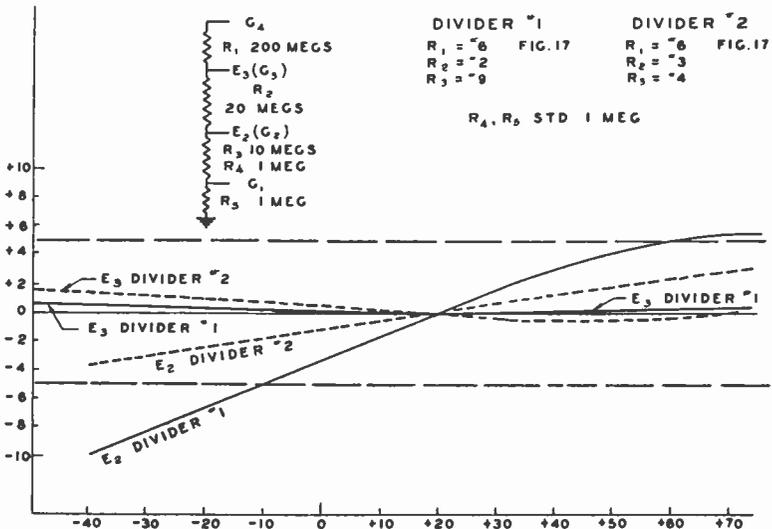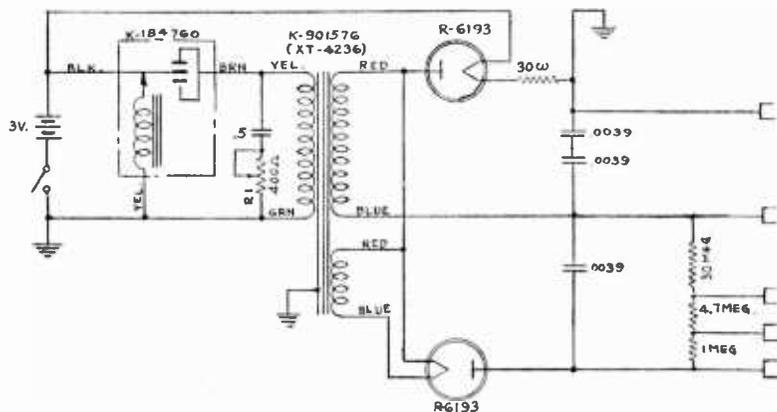

Fig. 18—Temperature characteristics of composite voltage divider.

Fig. 19—Voltage doubler power supply (Type S-2).

only the low voltage section, the desired low voltages may be obtained without loading down the high voltage section. Another very interesting feature of this circuit is the fact that by introducing resistance in the tuned primary circuit, the damping of the circuit is increased which tends to decrease the second or negative loops and thus the low voltage without appreciably affecting the high voltage. This action is shown in the curves on Figure 20. This affords a means of varying the
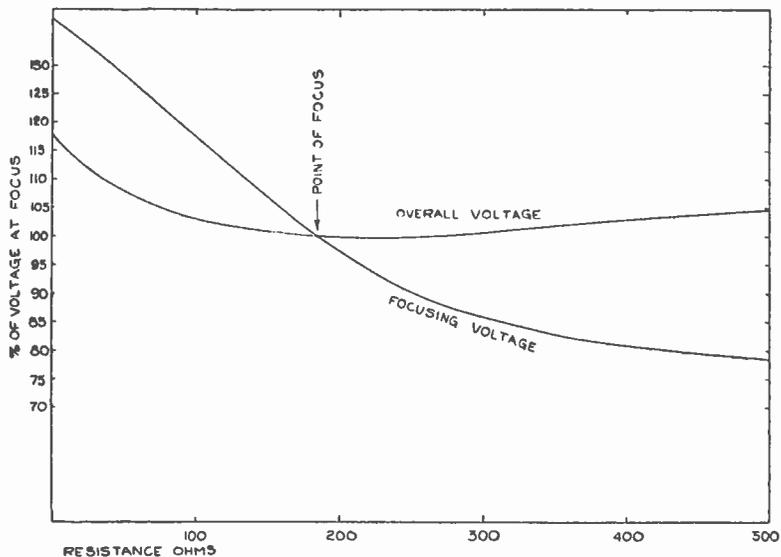
Fig. 20—Voltage control of type S-2 Power Supply.

focusing voltage by an element in the primary circuit which is a great advantage from the standpoint of electrical design.

Another power supply of interest was developed for use with the single voltage tube. With this tube the only load on the power supply is the actual photocurrent and leakage. By careful design, the entire load resistance can be made as high as 10^{10} ohms. Using a relatively large capacity in the output, the time constant of the circuit can be made to equal several seconds so that a quite infrequent charging of the circuit is required. For this purpose, an interrupter was designed consisting of an electrically-driven balance wheel having a period of
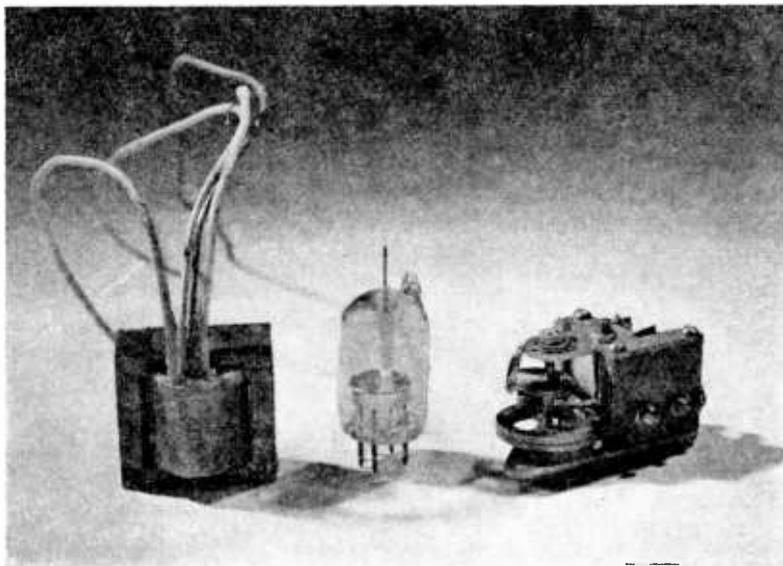

Fig. 21—Electrical components of pulsed power supply.

about $\frac{1}{4}$ second. The design was such that the transformer primary is open most of the time and is closed for a short time, to allow the current to build up, and immediately opened. In this way the drain on the battery is extremely small, the supply operating for as long as 50 hours on a single size D flashlight cell. A photograph of the interrupter, the 1-ounce transformer designed for the purpose and the special rectifier described below are shown in Figure 21.

Since the primary power required by this supply is so small, the power taken by the usual rectifier filament becomes very large in comparison. Consequently, a new type "filamentless" rectifier was developed, known experimentally as the KR-31.

This rectifier depends for its action upon a gas discharge in Helium,

Neon, or other inert gas at about 0.5 millimeters pressure. The cathode is an aluminum cup, mounted so that its closed side faces the anode. The anode is a nickel rod or tube, over which is fitted a woven fibreglass sleeve. The entire anode is covered with the fibreglass; the sleeve fits down over the glass seal at one end and is closed by fusing the glass at the free end.

The peak inverse voltage of the KR31 is 6000 volts and the forward breakdown voltage is 300-600 volts. The peak current may be several milliamperes but the allowable average current is low. In the applications for which the tube was designed the average current is under 10 microamperes. An average current of 50 microamperes may not be exceeded except for very short periods due to sputtering and clean-up of the gas.

This tube has not been put into production and is not available commercially.

The power pack for the Snooperscope and Sniperscope involved some special considerations. A 6-volt, 25-ampere-hour storage battery was used to operate the infrared source so the power required to operate the high-voltage power supply was a negligible drain on the battery. It was necessary to silence the vibrator to a surprisingly high degree since the most obvious uses of the instruments were under conditions of extreme quiet and where the utmost in secrecy was essential. The usual rubber-mounted automobile radio-type mounting is effective for damping out the high frequencies, but the fundamental vibrator frequency (100 cycles) is not sufficiently suppressed. One method used in experimental models was to suspend the mounted vibrator by two flat spiral springs of at least one turn, coiled around the vibrator can, the inner ends being fastened to the vibrator can and the outer ends to the power supply chassis or box. By proper choice of spring thickness, a period of only a few cycles per second can be obtained with sufficient stiffness to support the vibrator adequately. By this means it was possible to silence the vibrator so that the user himself could not detect the vibration. Another method used, with some increase in bulk, was the addition of one or more stages of sponge-rubber cushioning around the usual vibrator can.

The type MA4, high-voltage image tube raised some special problems in power supply design. The overall voltage required is on the order of 15 to 20 kilovolts and in addition, a number of intermediate voltages are required. These intermediate voltages, particularly those over 4000 volts are difficult to obtain efficiently by conventional means because of the relatively large power which would be wasted in a voltage divider of sufficiently low resistance to be stable. Also, it is

possible to obtain higher voltages from the previously described power supplies only by increasing the flux in the transformer. This in turn can be accomplished only by increasing the primary power, necessitating larger transformer and batteries. Lastly, if a conventional power supply is used, a rectifier tube capable of withstanding 20 to 30 kilovolts inverse voltage would be necessary. This type of rectifier is not available in small size and low filament power. Consequently, a cascade type (voltage-adding) power supply was designed which overcame most of the objections and automatically provided the necessary four steps of high voltage without a voltage divider.

A schematic diagram of the S-5 power supply making use of this circuit is shown in Figure 22. As can be seen, this supply is made up of four rectifiers which are essentially in parallel for alternating cur-
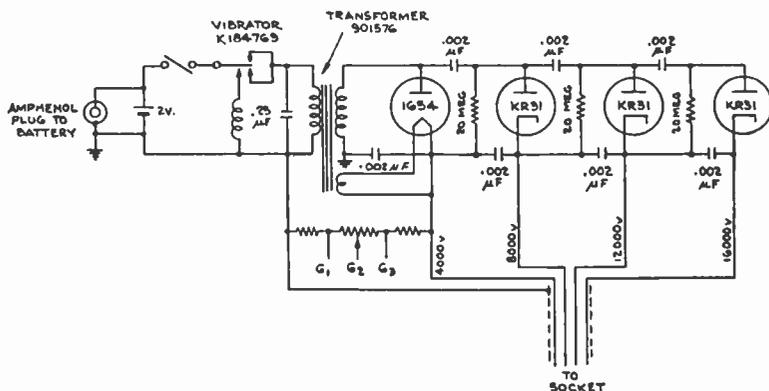

Fig. 22—Circuit of voltage quadrupler power supply (Type S-5).

rent. The direct-current voltages developed across the rectifiers, however, are added by means of the resistors which connect the anode of one rectifier to the plate of the next and thus place all the rectifiers in series for direct current. These resistors offer much higher impedance to the alternating current than do the capacitors so they do not affect the parallel alternating-current connection. Any number of stages may be cascaded in this manner, provided, of course, that the transformer will deliver the proper voltage to all the rectifiers in parallel. Four stages were chosen in this case because four steps of voltage are necessary for operation of the MA4 tube. The lower voltages required for the tube are obtained in the usual way by a voltage divider across the first section of the power supply. A thermionic rectifier (1654) is used in this stage in order to supply the divider current but the following stages make use of KR31 gas rectifiers, thus eliminating the need for

filament supply circuits, with a high degree of voltage insulation. The current drain at the high voltages is very low so that the voltages shown are obtained with a total battery current of only 0.4 ampere at 2 volts. A photograph of the power supply is shown in Figure 23.

The chief problems in connection with this supply are leakage and corona. These must both be kept to a minimum since the internal resistance of the power supply is quite high (approximately 10^9 ohms at the 16-kilovolt tap). Leakage can be minimized by use of high-quality insulation and protection from humidity. Hermetical sealing, or other provisions for drying, are essential with this type of voltage supply.
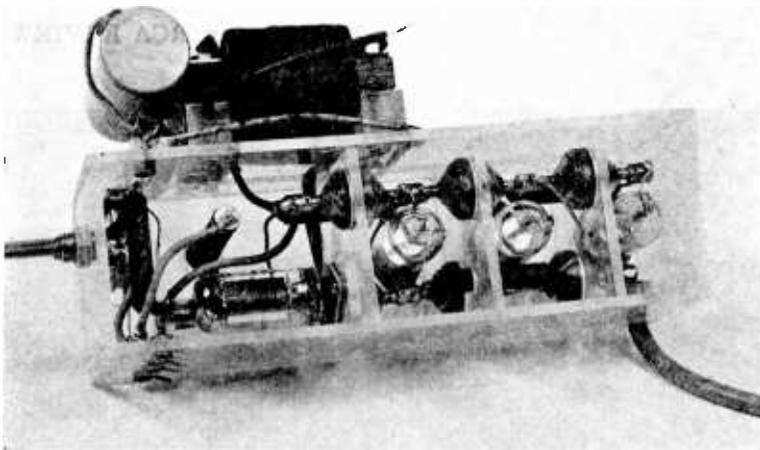
Fig. 23—16-kilovolt voltage quadrupler power supply.

Corona can be prevented by eliminating all sharp edges at the high voltage connections or by coating with a closely-adhering insulating material such as wax.

CONCLUSION

The above discussion stresses only the military application of infrared imaging equipment. There are, however, a number of peacetime uses for these instruments. Among these applications are their possible value in police work, their use in the field of medicine, the viewing of the usual types of photographic film during processing and production and for the inspection and control of a number of other industrial and scientific processes where visible light is undesirable.

In closing, the authors wish to express their appreciation to Dr.

V. K. Zworykin, Director of Electronic Research, for his advice and encouragement during the course of this development and also their recognition of the contribution made by Dr. J. E. Ruedy, G. L. Krieger and Dr. P. Rudnick to this project. Credit should go to Dr. L. B. Headrick, Miss H. C. Moodey and Dr. R. B. Janes of the Lancaster plant for work on the production design of the image tube.

As an interesting example of the effectiveness of the infrared devices described in this paper, two illustrations of a war street scene at night are included below. The upper picture shows the scene as viewed by the unaided eye. The lower picture shows, in the circle, the details of the scene when viewed by the infrared devices.

The Manager, RCA REVIEW
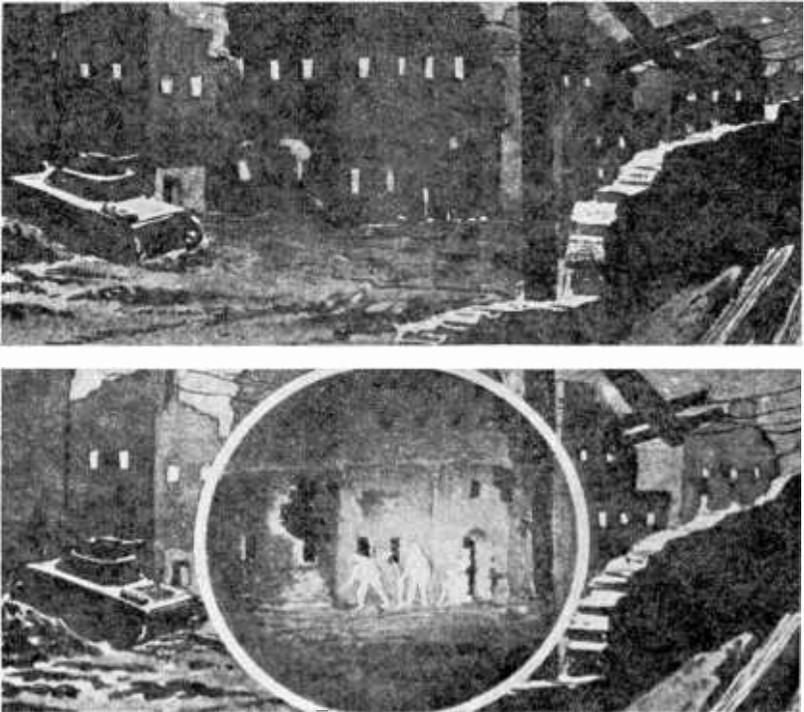
MULTIPLIER PHOTO-TUBE CHARACTERISTICS: APPLICATION TO LOW LIGHT LEVELS*†

By

RALPH W. ENGSTROM

Tube Department, RCA Victor Division,
Lancaster, Pa.

Summary—Measurements in such fields as photometry, spectrometry, and astronomy, dealing with very low light levels, may be assisted by the use of photo-multiplier tubes. Spectral characteristics of visible (S-8), blue (S-4), and ultraviolet (S-5) sensitive multiplier tubes are presented. Data are shown illustrating the wide range of linearity of photo-multiplier types 931-A, 1P21, 1P22, and 1P28. The tubes are linear at currents up to the point where space charge in the output stages reduces the ratio of anode current to cathode light flux. Dark currents and the limitations which they introduce in practice are discussed. Fatigue effects are discussed and data shown which indicate good stability at low current levels. The ultimate sensitivity of these tubes when refrigerated to various degrees has been investigated; with suitable precautions, pulses attributable to individual photoelectrons can be counted.

INTRODUCTION

IN RECENT years the availability of multiplier photo-tubes capable of measuring very weak light sources has resulted in a number of important applications in such fields as Raman spectra[1], spectrographic quantitative analyses[2-5], radioactive measurements[6], and astronomy[7,8]. The success of the multiplier photo-tube is principally the result of its tremendous amplification factor which, in addition to its great convenience, permits extension of measurement to lower light

---

[1] D. H. Rank, R. J. Pfister, and P. D. Coleman, *Jour. Opt. Soc. Amer.*, Vol. 32, 390 (1942).

[2] D. H. Rank and R. V. Wiegand, *Jour. Opt. Soc. Amer.*, Vol. 36, 325 (1946).

[3] E. A. Boettner and G. R. Brewington, *Jour. Opt. Soc. Amer.*, Vol. 34, 6 (1944).

[4] G. H. Dieke and H. M. Crosswhite, *Jour. Opt. Soc. Amer.*, Vol. 35, 471 (1945).

[5] J. L. Saunderson, V. J. Caldecourt, and E. W. Peterson, *Jour. Opt. Soc. Amer.*, Vol. 35, 681 (1945).

[6] M. Blau and B. Dreyfus, *Rev. Sci. Inst.*, Vol. 16, 245 (1945).

[7] G. E. Kron, *Astrophys. Jour.*, Vol. 103, 324 (1946).

[8] A. E. Whitford and G. E. Kron, *Rev. Sci. Inst.*, Vol. 8, 78 (1937).

fluxes than is possible with a simple photo-tube. This advantage is gained through the elimination of the coupling resistor and the associated Johnson noise[9-17].

The purpose of this paper is to make data on the properties of the RCA series of multiplier photo-tubes more generally available. The data include spectral response curves; representative data illustrating the magnitude of fatigue to be expected at various output levels; and a study of the linearity of the tubes with variations in light flux. Emphasis is placed on those properties affecting the measurement of very low light levels and the relative advantages of various methods of utilizing the tubes are presented.
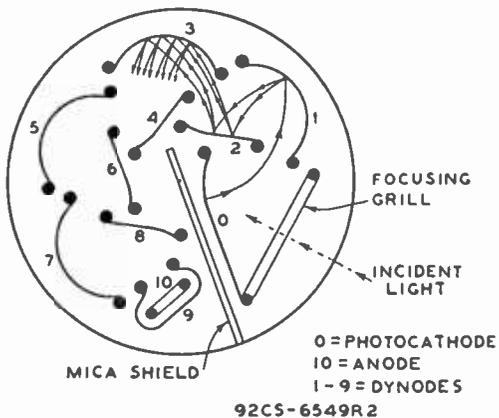

Fig. 1 — Schematic arrangement of type 931-A.

0 = PHOTOCATHODE
10 = ANODE
1 - 9 = DYNODES
92CS-6549R2

GENERAL CHARACTERISTICS

Gain and Voltage

Schematically illustrated in Fig. 1 is the RCA type of electrostatically focused multiplier photo-tube with the paths of the electrons and secondary electrons indicated as they cascade through the nine dynode stages to the collector anode. The result of this series of secondary-emission multiplications is an over-all gain of approximately one million at 100 volts per dynode stage. Variation of the amplifica-

[9] A. M. Glover, *Proc. I.R.E.*, Vol. 29, 413 (1941).

[10] J. A. Rajchman and R. L. Snyder, *Electronics*, Vol. 13, 20 (Dec. 1940).

[11] V. K. Zworykin and J. A. Rajchman, *Proc. I.R.E.*, Vol. 27, 558 (1939).

[12] R. B. Janes and A. M. Glover, *RCA Review*, Vol. 6, 43 (1941).

[13] A. Sommer, *Elec. Eng.*, Vol. 17, 164 (1944).

[14] R. C. Winans and J. R. Pierce, *Rev. Sci. Inst.*, Vol. 12, 269 (1941).

[15] F. Preisach, *Wireless Eng.*, Vol. 16, 169 (1939).

[16] C. C. Larson and H. Salinger, *Rev. Sci. Inst.*, Vol 11, 226 (1940).

[17] W. Shockley and J. R. Pierce, *Proc. I.R.E.*, Vol. 26, 321 (1938).

tion factor with the voltage per stage is shown in Fig. 2 for a typical 931-A tube. This rapid variation is often useful in that it allows a simple, wide-range gain-control which may be obtained either by varying the stage voltage of all the dynodes[18], or that of just one dynode. The first method controls gain by changing the secondary emission per stage; the second controls gain by defocusing the electron stream in one section of the tube. In any case the amplification factor is practically independent of the output stage voltage.

A disadvantage of a gain control sensitive to voltage changes is that it necessitates a rather well-regulated power supply to prevent the output current of the multiplier tube from being modulated by fluctua-
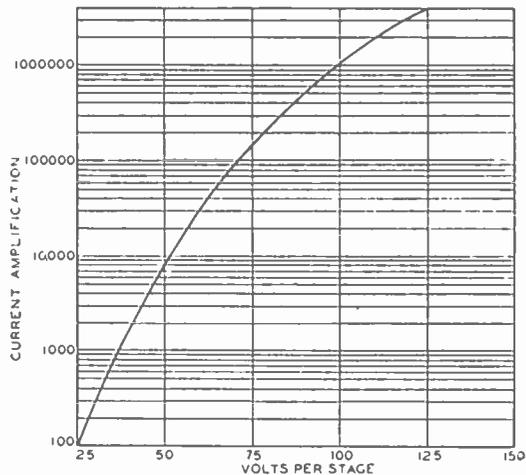
Fig. 2—Average amplification of type 931-A.

tions in supply voltage. For low output currents, a fairly compact and inexpensive battery pack can be constructed by use of miniature batteries. An analysis of fundamental electronic regulating circuits, with possible utility in multiplier photo-tube applications, exists in an article by Hunt and Hickman[19]. For the higher output currents a useful circuit may be found in a paper by W. R. Hill[20]. Much of the work that has been done on voltage supplies for the Geiger counter is directly applicable to multiplier tubes. In low current applications in this laboratory a modification of the amplification-bridge type of regulated supply[19] has been found very useful. This circuit has the advantage of providing a positive ground which is desirable in eliminating electrical pick-up. A wide range of regulated voltage may be obtained

---

[18] M. H. Sweet, *Jour. Opt. Soc. Amer.*, Vol. 36, 720 (1946).

[19] F. V. Hunt and R. W. Hickman, *Rev. Sci. Inst.*, Vol. 10, 6 (1939).

[20] W. R. Hill, *Proc. I.R.E.*, Vol. 33, 38 (1945).

without rebalancing the bridge by choosing a high μ-triode (75 or greater) with a flat amplification factor *versus* plate voltage characteristic.

Spectral Characteristics

At present three types of spectral response are available in the RCA line of multiplier photo-tubes. Data recently obtained for the S-4, S-5, and S-8 responses with a Bausch and Lomb quartz monochometer are presented in Fig. 3. The S-4 curve, characteristic of a caesium-antimony surface on a nickel base, may be compared with that published earlie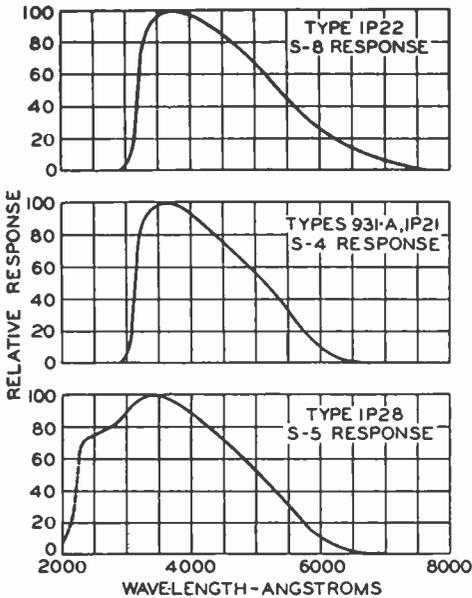r by this laboratory[9]. It is not known to what extent the minor differences are attributable to changed and v a r i a b l e manufacturing techniques, or to errors in measurement.

When a caesium-antimony surface is enclosed in an ultraviolet transmitting envelope (C o r n i n g 9741 glass), the useful range is extended as shown in Fig. 3 for type 1P28. The ultraviolet cut-off is critically dependent upon the composition and thickness of the glass, so that tube-to-tube variations may be expected.

A caesium-bismuth surface gives the additional red sensitivity of the S-8 response compared to the S-4 curve. This surface, in general, reduces the absolute sensitivity so that a tube with an S-8 response, such as the 1P22, should be chosen only when extension to the red is necessary. The range of response of the S-8 characteristic is approximately that of the eye but the shape of the curve is, unfortunately, quite different.

Fig. 3—Spectral response curves.

Color-response variations of these types are of a lesser degree than the rather large tube-to-tube variations of the caesium-oxygen-silver (S-1)types[21]. The S-4 spectral response is particularly stable. Produc-

---

[21] H. E. Ives and E. F. Kingsbury, *Jour. Opt. Soc. Amer.*, Vol. 21, 541 (1931).

tion variations in the response and cut-off characteristics of S-4 types are approximately one-tenth of those of the S-1 types.

Stability

Of great interest to the average user of multiplier tubes is the degree of stability that may be expected in the current output with constant light flux and voltage per stage. Each tube is individual in this respect but a fatigue, characterized by a decrease in sensitivity, is typical. Fatigue is also greater the higher the output current. Illustrating this trend are the data shown in Fig. 4 which were obtained from the average behavior of 10 representative 931-A tubes. In this graph the sensitivity loss in a given time is plotted as a function of the anode current. For each current value at which data were taken, the output current was held constant during the 40-minute test by
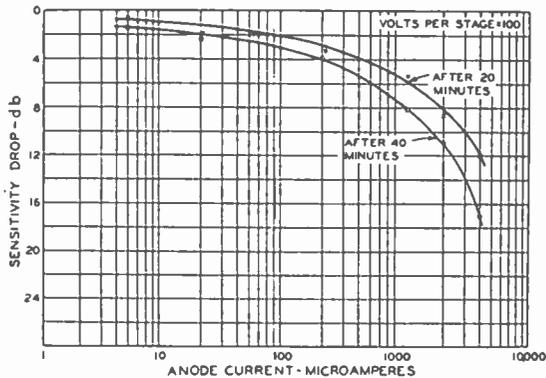

Fig. 4—Sensitivity loss in type 931-A.

adjusting the light flux. Sensitivities were recorded at the end of 20 and 40 minutes. The same tubes were used for all the points on the graph; after each 40-minute period the tubes were stored in darkness for at least one day before being operated at another current value. During this rest period, sensitivity generally recovers to approximately the initial value, except after currents in excess of 1 milliampere, in which case the recovery is incomplete. Loss in sensitivity results almost entirely from the loss in gain in the last few stages of secondary emission.

It is usually advantageous to operate the tube for at least a half-hour at the level of intended operation since the rate of fatigue is greatest during the first half-hour. Operation above 1 milliampere is not recommended because of the excessive fatigue. For most stable operation very low currents are recommended—certainly never over

100 microamperes. How stable the average tube is below 10 micro-amperes has not been determined, because errors in realigning the light spot and adjusting the voltage were larger with the particular test equipment used than the degree of instability being investigated. At currents less than 1 microampere, however, very little change in sensitivity may be expected.

Linearity

It is well known that photoelectric emission varies linearly with light flux over an extremely wide range. There is likewise no reason to expect secondary emission in the multiplier to be other than linear until space charge distorts the electrostatic fields. Figure 5 illustrates the linearity of output current in an RCA multiplier photo-tube as a function of light flux.
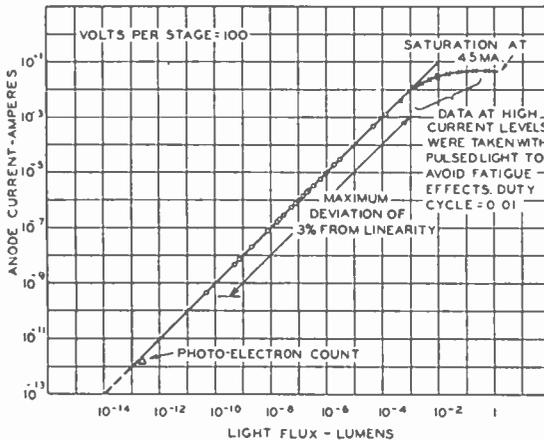

Fig. 5—Linearity characteristics of RCA photo-multipliers.

Within an experimental error of 3 percent, the anode current in the range from 10^{-9} to 10^{-3} ampere is proportional to light flux at the cathode. The low levels of light indicated on the graph were obtained by reflection of light consecutively from a pair of convex mirrors. A base of 20:1 in linearity was first established by varying the distance from the light source to the first diverging mirror. In this way any change in light spot or position on the multiplier cathode was made negligible. By means of this 20:1 range neutral filters were calibrated, and used to extend the measurements to the complete range shown in Fig. 5. Since the data for this curve were taken for several tubes, the value of the light flux plotted was arbitrarily adjusted to coincide with the sensitivity of an average tube (10 amperes per lumen), thus permitting a more simple comparison of results.

The last point on the lower end of the curve (Fig. 5) was obtained by counting and measuring the heights of the output pulses from the multiplier, each pulse being an amplified photo-electron. This point, representing a cathode photo-emission of only a few electrons per second, is included on the graph not so much to indicates the linearity of the multiplier at this level (of which there is very little doubt), but to illustrate the wide range of its applicability.

In order to obtain points in the range from 1 to 45 milliamperes, a technique using pulsed light was employed. Continuous operation at high levels of output current would fatigue the output stages of the multiplier so rapidly that reliable data could not have been obtained. The duty cycle, therefore, was reduced to 0.01 by means of a light-chopper disk. Measurements of the peak current output were then made with a cathode-ray oscillograph. The reduced duty cycle also reduced the bleeder current requirements by a factor of 100. Inter-stage capacitors maintained dynode voltages during the pulse time.

At the top end of the curve a saturation characteristic is observed. The chief limitation of current by space-charge saturation occurs not, as might be expected, at the ninth dynode, but at the surface of the eighth dynode (Fig. 1). This condition is the result of the much closer spacing between dynode 9 and the anode, than between the dynodes 8 and 9. Calculations from the geometry of the tube and the electro-static fields (with reference to a rubber dam model) checked quite well with the measured 45-milliampere current maximum at 100 volts per stage. Further confirmation of this point was obtained in the approximate ³⁄₂-power variation of the saturated current as a function of the voltage per stage.

While these data indicate anode currents as high as 45 milliamperes, it should be emphasized that stability cannot be expected at such levels. These levels are included chiefly for their academic interest and for possible application of the tube in light-pulse detection. The stability of the tube with pulsed light is approximately that of the integrated average current level.

APPLICATION TO LOW LIGHT LEVELS

Dark Current

In any application of the multiplier photo-tube to very low light levels, one of the most important considerations is the dark current of the tube. It is important to understand the source of this current so that operation with maximum sensitivity may be obtained.

Illustrated in Fig. 6 is the anode dark current in microamperes as

a function of the voltage per stage. The points were obtained from one tube chosen as typical of the behavior of RCA multiplier photo-tubes. There are three domains in each of which a different type of dark current dominates: (1) ohmic leakage; (2) amplified thermionic emission; (3) regenerative ionization.

At low voltages ohmic leakage is dominant. Leakage current is caused by traces of caesium and other conductive materials on the stem and insulating spacers inside of the tube. Base leakage may also be included in this category. Base leakage is particularly evident if the
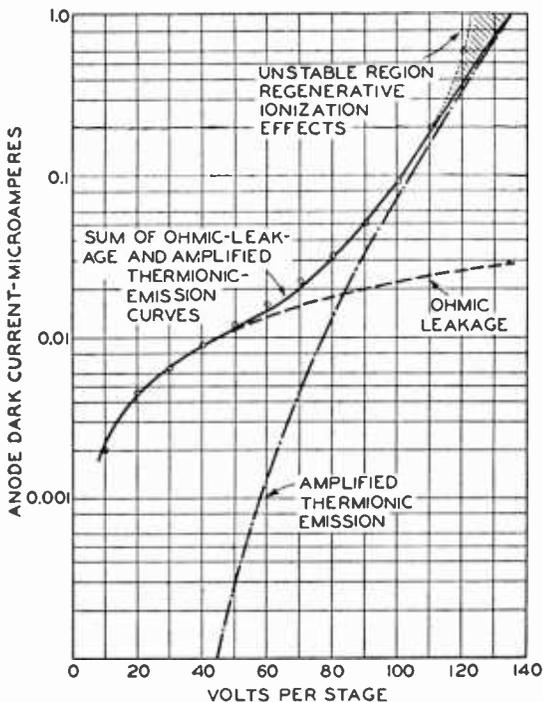

Fig. 6 — Dark current in photo-multiplier tubes.

tube is operated in a very humid atmosphere or in a dirty or dusty locale. As a general rule the leakage in the base itself is less than that inside of the tube. If high humidity causes difficulty it may be advantageous to apply a coating of ceresin wax to the base. In an atmosphere of reduced pressure an electric breakdown may also occur across adjacent pins at high voltage. This type of breakdown is important in airborne equipment.

In the characteristic plotted in Fig. 6 the ohmic leakage is dominant up to about 50 or 60 volts per stage. An extrapolated dashed line indicates the calculated continuation of the ohmic leakage. This leak-

age is directly proportional to the voltage per stage, but the shape of the curve is convex because a logarithmic scale is used for the dark current.

Above 60 volts per stage dark current caused by thermionic emission becomes important. In the sample tube the equivalent thermionic emission at the cathode was 0.6×10^{-14} ampere. This emission is amplified through the tube in the same manner as photo-current; hence, the output current caused by thermionic emission is proportional to the gain characteristic of the multiplier tube, as illustrated by the dot-dash line in Fig. 6. The sums of the values of the extrapolated ohmic-leakage curve, and the values of the amplified thermionic emission curve give the solid line which quite closely fits the observed points and, thus, gives credence to this interpretation.

In almost any method of using the multiplier photo-tube the ultimate limitation to signal detection is thermionic emission. Associated with this component of the dark current is a shot noise resulting from random thermionically emitted electrons, variably multiplied by the secondary-emission gains of nine stages.

Above about 110 volts per stage, Fig. 6, a third region of dark current begins. Deviation from the ohmic and thermionic dark-current curve is associated with regenerative ionization effects, which cause complete breakdown and an uncontrollable discharge at sufficiently high voltage. For most tubes this breakdown occurs between 140 and 180 volts per stage depending principally on the gain of the tube. In this unstable region, indicated by the shaded area on the graph, dark currents fluctuate violently and if the current is viewed on a cathode-ray oscilloscope, rather large bursts of pulsed noise will be observed in contrast to the fairly steady "rain" associated with thermionic-emission noise. The exact cause of this gaseous noise is not known, but when the gain of the tube has reached a value of 10^7 to 10^8, very little feedback from the anode to the cathode will obviously initiate a self-maintained discharge. This feedback may be caused either by secondary emission from positive ions (of residual gas or vapor), or by photo-emission resulting from light emitted from excited gas, or from fluorescence in the glass or other insulators in the tube. Although the tube has been designed to avoid feedback, it will occur in all tubes when the gain is sufficiently high.

When the tube is operated for optimum signal-to-noise ratio, the region where gas effects become noticeable should be avoided in any system of measurement. If a direct-current method of observation is used, it is best to operate at a stage voltage level where the ohmic