The Bell System Technical Journal

Vol. XXIX *October, 1950* *No. 4*

Theory of Relation between Hole Concentration and Characteristics of Germanium Point Contacts

By J. BARDEEN

(*Manuscript Received Apr. 7, 1950*)

The theory of the relation between the current-voltage characteristic of a metal-point contact to *n*-type germanium and the concentration of holes in the vicinity of the contact is discussed. It is supposed that the hole concentration has been changed from the value corresponding to thermal equilibrium by hole injection from a neighboring contact (as in the transistor), by absorption of light or by application of a magnetic field (Suhl effect). The method of calculation is based on treating separately the characteristics of the barrier layer of the contact and the flow of holes in the body of the germanium. A linear relation between the low-voltage conductance of the contact and the hole concentration is derived and compared with data of Pearson and Suhl. Under conditions of no current flow the contact floats at a potential which bears a simple relation, previously found empirically, with the conductance. When a large reverse voltage is applied the current flow is linearly related to the hole concentration, as has been shown empirically by Haynes. The intrinsic current multiplication factor, α, of the contact can be derived from a knowledge of this relation.

I. Introduction

IN DISCUSSIONS of the theory of rectification at metal-semiconductor contacts, it is usually assumed that only one type of current carrier is involved: conduction electrons in *n*-type material or holes in *p*-type material.[1] In the case of metal-point contacts to high-purity *n*-type germanium, such as is used in transistors and high-back-voltage varistors, it is necessary to consider flow by both electrons and holes. A large part of the current in the direction of easy flow (metal point positive) consists of holes which flow into the *n*-type germanium and increase the conductivity of the material in the vicinity of the contact.[2,3] The conductivity is increased not only by the presence of the added holes but also by the additional conduction electrons which flow in to balance the positive space charge of the holes. There is a small concentration of holes normally present in the germanium under equilibrium conditions with no

[1] For a discussion of the nature of current flow in semi-conductors see the "Editorial Note" in *Bell Sys. Tech. Jour.* 28, 335 (1949).

[2] J. Bardeen and W. H. Brattain, *Bell Sys. Tech. Jour.* 28, 239 (1949).

[3] W. Shockley, G. L. Pearson and J. R. Haynes, *Bell Sys. Tech. Jour.* 28, 344 (1949).

current flow. When the contact is biased in the reverse (negative) direction, these holes tend to flow toward the contact and contribute to the current. The hole current is increased if the concentration of holes in the germanium is enhanced by injection from a neighboring contact or by creation of electron-hole pairs by light absorption.

Much has been learned about the effect of an added hole concentration on the current voltage characteristics of contacts from studies with germanium filaments. Part of this work is summarized in a recent article of W. Shockley, G. L. Pearson and J. R. Haynes.[3] These authors have investigated the way the low-voltage conductance of a point contact to a filament of *n*-type germanium varies with the concentration of holes in the filament and have shown that there is a linear relation between conductance and hole concentration. They have shown that the current to a contact biased with a large voltage in the reverse direction varies linearly with hole concentration. Suhl and Shockley[4] have shown that by applying a large transverse magnetic field along with a large current flow holes may be swept to one side of the filament. Changes in hole concentration produced in this way are detected by measuring changes in the conductance of a point contact.

Shockley[5] has suggested that the floating potential measured by a contact made to a semiconductor in which the concentration of carriers is not in thermal equilibrium may depend on the nature of the contact and differ from the potential in the interior. Pearson[6] has investigated this effect for point contacts on germanium filaments, and has shown that the floating potential is related to the conductance of the contact. This effect provides an explanation for anomalous values of floating potentials measured by Shockley[6] and by W. H. Brattain.[6] They found that potentials measured on a germanium surface in the vicinity of an emitter point biased in the forward direction may be considerably higher than expected from the conductivity of the material.

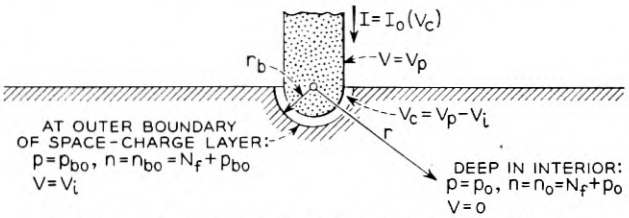
The purpose of the present paper is to develop the theory of these relations. We are particularly interested in effects produced by changes in hole concentration in *n*-type germanium resulting from hole injection or photoelectric effects. The equations developed also apply to injected electrons in *p*-type semiconductors with appropriate changes in signs of carriers and bias voltages. The methods of analysis used are similar to those which have been employed by Brattain and the author in a discussion of the forward current in germanium point contacts[2].

[4] H. Suhl and W. Shockley, *Phys. Rev.* 74, 232 (1948).
[5] W. Shockley, *Bell Sys. Tech. Jour.* 28, 435 (1949), p. 468.
[6] Unpublished.

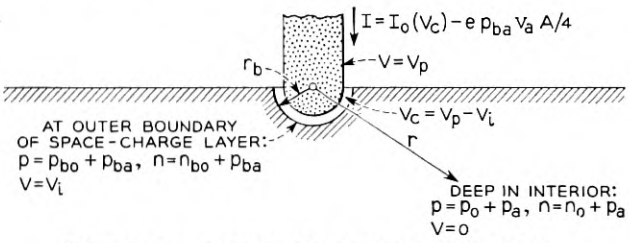
The problem may be divided into two parts, which can be treated separately:

(a) The first deals with the current-voltage characteristics of the space charge region of the rectifying contact. The current flowing across the contact is expressed as the sum of the current which would flow if the hole concentration in the interior were normal and the current which results from the added hole concentration.

(b) The second is concerned with the current flow in the semiconductor outside the space charge region. In general, both diffusion and conduction



(a) EQUILIBRIUM CONCENTRATION OF HOLES IN INTERIOR



(b) ADDED CONCENTRATION OF HOLES IN INTERIOR

Fig. 1.—Model and notation used for calculation of current flow in low-voltage case.

are important in determining the flow of carriers, although, depending on conditions, one may be much more important than the other. In case the applied voltage and current flow are small, holes in an *n*-type semiconductor move mainly by diffusion. This situation applies to the problems discussed in the first part of the memorandum. In Section IV we discuss the opposite limiting case of large voltages in which the electron current flowing is so large that the hole current is determined by the electric field and diffusion is unimportant.

The model which is used to investigate the low-voltage case is illustrated in Fig. 1. For purposes of mathematical convenience, the contact is represented as a hemisphere extending into the germanium. Recombination, both at the surface of the semiconductor and in the interior, is

assumed to be negligible so that the lines of current flow are radial. The spherical symmetry of the resulting problem simplifies the mathematics. A calculation is given in an Appendix for a model in which the contact is a circular disk and recombination takes place at the surface. The latter does not give results which are significantly different from the simplified model.

Figure 1(a) applies to the case in which the hole concentration deep in the interior has its normal or thermal equilibrium value, p_0. The subscript zero is used to denote values which pertain to this situation. Of a voltage V_P applied to the contact, a part V_c occurs across the space-charge barrier layer of the contact and a part V_i occurs in the body of the semiconductor. Thus V_P represents the voltage of the contact and V_i the voltage in the semiconductor just outside the barrier layer, both measured relative to a point deep in the interior. It should be noted that V_P does *not* include the normal potential drop which occurs across the barrier layer under equilibrium conditions with no voltage applied. In the examples with which we shall deal in the present memorandum, the spreading resistance is small compared with the contact resistance, so that V_i is small compared with V_P. Obviously,

$$V_P = V_c + V_i. \tag{1}$$

When a current is flowing to the contact the hole concentration, p_{b0}, measured just outside of the barrier layer, differs from the concentration deep in the interior, p_0. It is the concentration gradient resulting from the difference between p_{b0} and p_0 which produces a flow of holes from the interior to the contact. In the forward direction, p_{b0} is larger than p_0; in the reverse direction, p_{b0} is less than p_0.

The total current, $I_0(V_c)$, flowing across the contact includes both electron and hole currents. It will not be necessary to distinguish between these two contributions to the normal current flow across the barrier layer in the subsequent analysis.

Figure 1(b) applies to the case in which the hole concentration deep in the interior has been increased to $p_0 + p_a$ by adding a concentration p_a to the normal concentration, p_0. The concentration just outside the barrier layer is increased to $p_{b0} + p_{ba}$. In addition to the normal current, $I_0(V_c)$, flowing across the contact, there is an additional current of holes resulting from the added hole concentration, p_{ba}, at the barrier.

The magnitude of this added hole current is determined in the following way. It is assumed that all holes which enter the barrier region are drawn into the contact by the field existing there. The number of holes

entering the barrier region per second is given by the following expression from kinetic theory:

$$p_b v_a A/4, \tag{2}$$

where v_a is the average thermal velocity, $2(2kT/\pi m)^{1/2}$, of a hole and A is the contact area. This expression gives the average number of particles which cross an area A from one side per second in a gas with concentration p_b. It follows that the current due to the added holes is:

$$I_{pa} = -e p_{ba} v_a A/4. \tag{3}$$

Since, by convention, a current flowing into the semiconductor is positive, a current of holes flowing from the interior to the contact is negative.

The diffusion current resulting from the added holes depends on the difference between p_{ba} and p_a. We shall show in Section III that when p_a is small compared with the normal electron concentration,

$$I_{pa} = 2\pi r_b kT\mu_p(p_{ba} - p_a), \tag{4}$$

where r_b is the radial distance to the outer boundary of the barrier layer and μ_p is the hole mobility. The value of p_{ba} is found by equating (4) and (3), i.e., the added current flowing from the interior to the barrier layer and the current flowing across the barrier layer. This gives

$$p_{ba}/p_a = a/(1 + a), \tag{5}$$

where a, defined by

$$a = 4(kT/er_b)\mu_p/v_a, \tag{6}$$

is the ratio of the velocity acquired by a hole in a field $4kT/er_b$ to thermal velocity. This ratio is generally a small number so that the a in the denominator of (5) can be neglected in comparison with unity. Equation (3) then becomes:

$$I_{pa} = -ea p_a v_a A/4 = -p_a kT\mu_p A/r_b. \tag{7}$$

If p_a is not assumed small, a similar procedure may be used but the expressions for I_{pa} in terms of p_a are more complicated than (4) and (7)

It is possible that the added hole current, I_{pa}, will affect the contact in such a way as to change the normal current flowing. If there is such a change, one might expect it to be proportional to I_{pa} as long as I_{pa} is sufficiently small. The total current flow may then be expressed in terms of an "intrinsic α" for the contact as follows:

$$I = I_0(V_c) - \alpha I_{pa}(p_a). \tag{8}$$

There is no good theoretical reason to expect that α is different from unity for small current flow in normal contacts unless trapping is important.

Equation (8) is used as the basis for the analysis of the low-voltage data. One important consequence of the equation is that if p_a is different from zero, there is a voltage drop across the barrier layer even though no net current flows to the point. The presence of the added holes in the interior produces a floating potential on the point. The magnitude of this floating potential, V_{cf}, is obtained by setting $I = 0$ in Eq. (8) and finding the value of V_c which solves the equation. This potential can be observed on a voltmeter and is analogous to a photovoltage.

Associated with the floating potential is a change in conductance of the contact. The conductance near $I = 0$, given by

$$G = (dI/dV_c)_{V_c=V_{cf}} = (dI_0/dV_c)_{V_c=V_{cf}}, \tag{9}$$

is just the conductance for normal hole concentration in the interior at an applied voltage equal to V_{cf}. In setting the conductance equal to the derivative of I with respect to V_c, we have neglected the difference, V_i, between V_c, the voltage drop across the barrier, and V_P, the total drop from the contact to the interior. This corresponds to neglecting the spreading resistance in comparison with the barrier resistance.

Equation (8) may be used to relate the floating potential with change of conductance of the contact. The appropriate equations, together with applications to data of Pearson and of Brattain, are given in Section II. In Section III we derive Eq. (4) which relates the added hole current with the added hole concentration in the interior. This relation is used to show that the point conductance G varies linearly with the added hole concentration, p_a. The theoretical expression for conductance is compared with data of Pearson and of Suhl.

In section IV we discuss the dependence of the current-voltage characteristic at large reverse voltages on hole concentration. Under these conditions it is the electric field rather than diffusion which produces the hole current in the body of the germanium. The electron and hole currents are then in the ratio of the electron to hole conductivity. With introduction of an "intrinsic α" for the contact, a simple relation is derived for the dependence of current on hole concentration for fixed voltage on the point. This relation is used to determine α for several point contacts from some data of J. R. Haynes.

II. Floating Potential of Point Contact

In order to get analytic expressions for the floating potential and admittance, it is necessary to make some assumption about the normal cur-

rent-voltage characteristic, $I_0(V_c)$. It is found empirically[7] that as long as V_c is not too large (a few tenths of a volt for a point contact on *n*-type germanium), it is a good approximation to take:

$$I_0 (V_c) = I_c (\exp(\beta e V_c/kT) - 1), \tag{10}$$

where I_c is a constant for a given contact. Except for the factor β, this is of the form to be expected from the diode theory of rectification. The empirical value of β is usually less than the theoretical value of unity in actual contacts.

If (10) is inserted into (8), the following equation is obtained for the current when there is an added concentration of carriers, p_a, in the interior:

$$I = I_c (\exp(\beta e V_c/kT) - 1) - \alpha I_{pa}. \tag{11}$$

Setting $I = 0$ and solving the resulting equation for the floating potential, $V_c = V_{cf}$, we find:

$$V_{cf} = (kT/e\beta) \log [1 + \alpha(I_{pa}/I_c)]. \tag{12}$$

The floating potential may be simply related to the conductance corresponding to small current flow. Using Eqs. (9) and (11), we find:

$$G = (dI_0/dV_c)_{V_c=V_{cf}} = (\beta e I_c/kT) \exp (\beta e V_{cf}/kT). \tag{13}$$

Since the normal low-voltage conductance is just

$$G_0 = \beta e I_c/kT, \tag{14}$$

we have

$$G = G_0 \exp (\beta e V_{cf}/kT). \tag{15}$$

By using (12), G can be expressed in terms of p_a. This relation is given and compared with experiment in Section III. Equation (15) may be solved for the floating potential:

$$V_{cf} = (kT/e\beta) \log (G/G_0). \tag{16}$$

It should be noted that (16) does not involve p_a directly. Thus it is possible to determine V_{cf} from a measurement of the change in conductance without direct knowledge of the added hole concentration. It holds for large as well as small p_a.

The logarithmic relation (16) between floating potential and conductance has been demonstrated by an experiment of Pearson. The experi-

[7] See H. C. Torrey and C. A. Whitmer, "Crystal Rectifiers", McGraw-Hill Company, New York, N. Y., (1949), p. 372–377.

mental arrangement is illustrated in Fig. 2. Holes are injected into a germanium filament by an emitter point and the circuit is closed by allowing the current to flow to the large electrode at the left end. The right end of the filament is left floating. Some of the injected holes diffuse

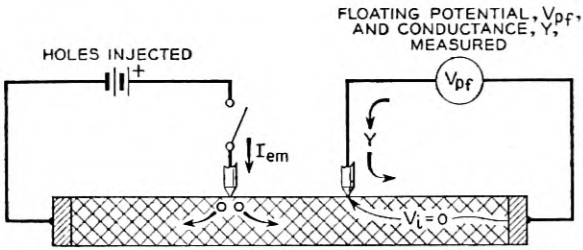


Fig. 2.—Schematic diagram of experiment of G. L. Pearson to investigate relation between floating potential and impedance of point contact.

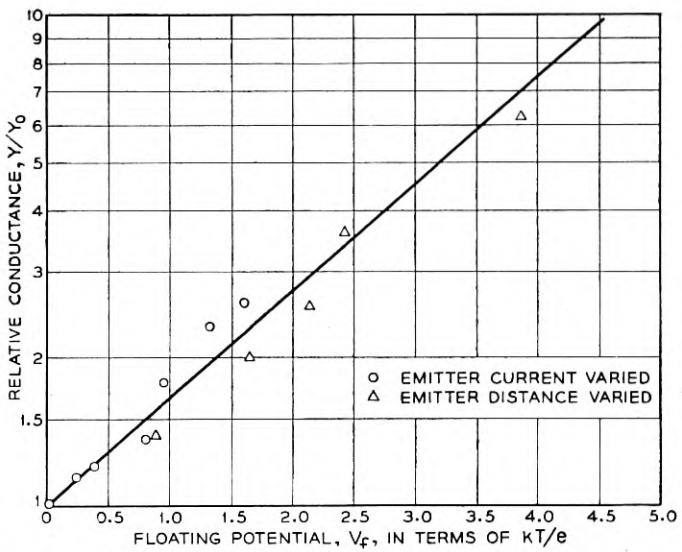


Fig. 3.—The relationship of admittance ratio to potential, measured at a point on a germanium filament into which holes are emitted, with no current flow, from G. L. Pearson's data of September 21, 1948.

down the filament and increase the local concentration in the neighborhood of the probe point. This concentration can be varied by changing the emitter current and also by changing the distance between emitter and probe. Both the floating potential and the conductance between the probe point and the large electrode on the right end were measured. Under the conditions of this experiment, the potential drop in the in-

terior of the floating end of the filament is small. The small drop which does exist results from the difference in mobility between electrons and holes. Almost all of the potential difference between the probe and the right end is the floating potential, V_{cf}, across the barrier layer of the probe point.

Pearson's data are plotted in Fig. 3. The data can be fitted by an equation of the form (16) with $\beta = 0.5$.

The difference in potential between a floating point contact and the interior which exists under non-equilibrium conditions explains anomalously high values of probe potential which were sometimes observed by Shockley and by Brattain in the vicinity of an emitter point operating in the forward direction. As an example of a case in which the effect is

TABLE I

Measurements of probe potential, V_{pf}, at a contact on an etched germanium surface .005 cm from a second contact carrying a current I. The conductance of the probe point is G_P. The voltage drop across the probe contact, $V_{pf} - V_i$, at zero current is calculated from $V_{pf} - V_i = 2.5(kT/e) \log (G_p/G_0)$. Data from W. H. Brattain.

I amps	V_{pf} volts	V_{pf}/I ohms	G_p mhos	G_p/G_0	$\log (G_p/G_0)$	$V_{pf} - V_i = .062 \log (V_p/V_0)$	V_i volts	V_i/I ohms
2.0×10^{-3}	0.189	94	8.3×10^{-4}	6.9	1.93	0.120	0.069	35
1.0	0.141	141	5.0	4.2	1.435	0.090	0.051	51
0.5	0.096	190	3.3	2.8	1.030	0.064	0.032	64
0.2	0.052	260	2.2	1.8	0.588	0.037	0.015	75
0.1	0.030	300	1.7	1.4	0.336	0.021	0.009	90
−0.1	−0.0096	96	1.2	1.0				
−0.2	−0.0186	93	1.2	1.0				
−0.5	−0.044	88	1.25	—				
−1.0	−0.10	100	1.35	—				

large, some data of Brattain are given in Table I for the experimental arrangement of Fig. 4. Two point contacts were placed about .005 cm apart on the upper face of a germanium block. The surface was ground and etched in the usual way. A large-area, low-resistance contact was placed on the base. The potential, V_P, of one point, used as a probe, was measured as a function of the current flowing in the second point. In this case, the potential on the probe point is produced in part by the V_{cf} term and in part by a potential, V_i, in the interior which comes from the IR drop of the current flowing from the emitter point to the base electrode. Reasonable values are obtained for V_i from measurements of V_P if a correction for V_{cf} is properly made.

The first column of Table I gives the current and the second column the probe potential, V_P, measured relative to the base. The third column gives values of V_P/I. In the reverse direction (negative currents) V_P/I

is approximately constant at a little less than 100. Values of V_P/I in the forward direction are much larger, starting at 300 for $I = 0.1$ ma and decreasing to 94 at $I = 2$ ma. If anything, one would expect a decrease rather than an increase in V_P/I in the forward direction as injection of holes lowers the resistivity of the germanium in the vicinity of the point. We shall show that V_i/I actually does decrease and that the anomalously high values of V_P/I in the forward direction result from the drop, V_{ef},

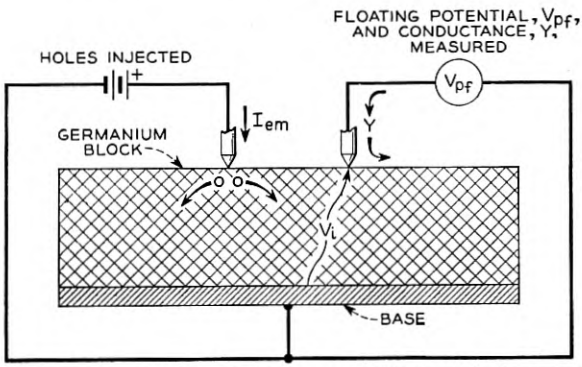


Fig. 4.—Schematic diagram of experiment of W. H. Brattain for measuring floating potential and admittance at point near emitter.

across the barrier layer between the contact point and the body of the germanium. Thus,

$$V_i = V_P - V_{ef}. \tag{17}$$

Values of V_{ef} can be estimated from the change in conductance corresponding to small currents in the probe point. The conductance increases with increasing forward emitter current. Values of V_{ef}, calculated from

$$V_{ef} = 2.5\ (kT/e)\ \log\ (G_p/G_0), \tag{18}$$

are given in column 6. The value 2.5, chosen empirically to give reasonable values of V_i, is not far from the value 2.0 required to fit Pearson's data in Fig. 1. Values of V_i obtained from Eq. (17) are given in column 7. The ratios V_i/I given in column 8 are reasonable. The decrease in V_i/I with increasing forward current is caused by a decrease in the resistivity of the germanium resulting from hole injection.

In another case, in which no such anomaly was observed in the forward direction, it was found that V_{ef}, calculated from the change in conductance, was small compared with V_p.

There have as yet been no measurements which permit a comparison of the values of β required to correlate probe potential and conductance

with values of β obtained directly from the current-voltage characteristic of the probe. Such a comparison would provide a valuable test of the theory.

III. Low Voltage Conductance of Point Contacts

In this section we calculate the hole current flowing in the body of the germanium from diffusion and find an expression relating change of conductance with added hole concentration. The results shall be applied to data of Pearson and of Suhl. We need to derive Eq. (4) which gives the hole current in terms of the added hole concentrations, p_{ba}, measured just outside the barrier layer, and p_a, measured deep in the interior.

The model which is used for the calculation is illustrated in Fig. 1. The diffusion equation for hole flow is to be solved subject to the boundary conditions that $p = p_b$ just outside the barrier layer and $p = p_i$ at large distances from the contact in the interior. It is assumed that the total current flow is zero or small.

We shall first derive the more general equations[8] which include flow by the electric field as well as by diffusion in order to show the conditions under which the electric field can be neglected. In the body of the semiconductor, conditions of electric neutrality require that the electron concentration, n, be given by:

$$n = N_f + p, \tag{19}$$

where N_f, the net concentration of fixed charge, is the difference between the concentrations of donor and acceptor ions. We shall assume that N_f is constant so that

$$\text{grad } n = \text{grad } p. \tag{20}$$

The general equations for electron and hole current densities, i_n and i_p, are:

$$i_n = \mu_n (enF + kT \text{ grad } n) \tag{21}$$

$$i_p = \mu_p (epF - kT \text{ grad } p), \tag{22}$$

where F is the electric field strength. By using (19) and (20), and setting $\mu_n = \beta\mu_p$, we can express i_n in the form:

$$i_n = b\mu_p (e(N_f + p) F + kT \text{ grad } p). \tag{23}$$

The magnitude of F for zero net current,

$$i = i_p + i_n = 0, \tag{24}$$

[8] A discussion of the equations of flow is given in the article by W. van Roosbroeck in this issue of the *Bell System Technical Journal*.

can be obtained by adding (22) and (23) and equating the result to zero. This gives:

$$\frac{eF}{kT} = -\frac{b-1}{N_f b + p(b+1)} \text{ grad } p. \tag{25}$$

The field vanishes for $b = 1$, corresponding to equal mobilities for holes and electrons. For b greater than unity and for equal concentration gradients of holes and electrons, the diffusion current of electrons is larger than that of holes. The field is such as to equate these currents by increasing the flow of holes and decreasing the flow of electrons.

If (25) is substituted into (22), the following equation is obtained for i_p:

$$i_p = -kT\mu_p \left[\frac{(b-1)p}{N_f b + p(b+1)} + 1\right] \text{grad } p. \tag{26}$$

If recombination is neglected, the hole current is conserved and

$$\text{div } i_p = 0. \tag{27}$$

Using this relation, an equation of the Laplace type can be obtained for p which may be integrated subject to the appropriate boundary conditions. This derivation is given in Appendix B. The results do not differ significantly from those obtained below for p assumed small.

Rather than continue with the general case, we shall at this point assume that $p \ll N_f$ so that the first term in the parenthesis of Eq. (26) is negligible in comparison with unity. This amounts to setting $F = 0$ in Eq. (3) and assuming that the holes move entirely by diffusion. This is a very good approximation in most cases of practical interest and is valid for small i as well as for $i = 0$. We then have

$$i_p = -kT\mu_p \text{ grad } p. \tag{28}$$

The condition div $i_p = 0$ gives Laplace's equation for p:

$$\nabla^2 p = 0. \tag{29}$$

Equation (29) is to be solved subject to the appropriate boundary conditions. For the model illustrated in Fig. 1 we can assume that p depends only on the radial distance r and that

$$p = p_b \text{ at } r = r_b, \tag{30}$$

$$p = p_i \text{ at } r = \infty. \tag{31}$$

The solution of (29) which satisfies (31) is:

$$p = p_i + (I_p/2\pi kT\mu_p r), \tag{32}$$

in which I_p is the total hole current. The boundary condition (30) gives the relation between I_p and p_b:

$$p_b = p_i + (I_p/2\pi kT\mu_p r_b). \tag{33}$$

Since the equations are linear, an equation of the form (33) applies to the hole current due to the added holes as well as to the entire hole current. For the former we have:

$$p_{ba} = p_a + (I_{pa}/2\pi kT\mu_p r_b), \tag{34}$$

which is equivalent to Eq. (4).

In the derivation of Eq. (34) we have neglected recombination at the surface as well as in the interior. In the Appendix we give a solution for a contact in the form of a circular disk and assume that recombination takes place at the surface. The hole concentration then satisfies Laplace's equation subject to more complicated boundary conditions at the surface. The results are not significantly different from those of the simplified model.[9]

Equation (34), or rather its equivalent, Eq. (4), was used in the derivation of Eq. (12) for the floating potential, V_{cf}. If this value for V_{cf} is inserted into Eq. (15), an equation relating the conductance directly with the added hole concentration is obtained:

$$G = G_0 + (\alpha e^2 a v_a A \beta p_a/4kT). \tag{35}$$

This expression may be simplified by substituting for a from Eq. (6):

$$G = G_0 + \alpha\beta\mu_p e A p_a/r_b. \tag{36}$$

By using the expression for the normal conductivity:

$$\sigma_0 = b\mu_p e n_0, \tag{37}$$

the conductance can be given in the form:

$$G = G_0 + (\alpha\beta\sigma_0 A/br_b)(p_a/n_0). \tag{38}$$

If σ_0 is in practical units (mhos/cm), G is in mhos.

We shall compare (38), which gives a linear variation between G and p_a, with experimental data of Pearson[10] and of Suhl. The arrangement used

[9] In the applications, these equations are applied to situations in which the contact is on a germanium filament and there is a flow of current along the length of the filament in addition to the flow to the contact. A question may arise as to whether it is justified to neglect the filament current when discussing flow to the contact. There is no difficulty as long as p_a/n_0 is small compared with unity because the equations are then linear and the solution giving the flow to the contact can be superimposed on the solution giving the flow along the length of the filament. The neglect of the filament current cannot be rigorously justified in case p_a/n_0 is large, as is assumed in the calculations of Appendix B. It is not believed, however, that the exact treatment would yield results which are significantly different.

[10] See reference 3, p. 356 and Fig. 6.

by Pearson is shown in Fig. 5. Two probe points were placed about .009 cm apart near one end of a germanium filament. The concentration

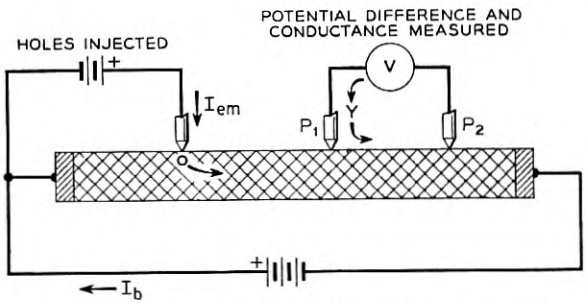


Fig. 5.—Experimental arrangement used by G. L. Pearson to investigate relation between admittance and hole concentration in germanium filament.

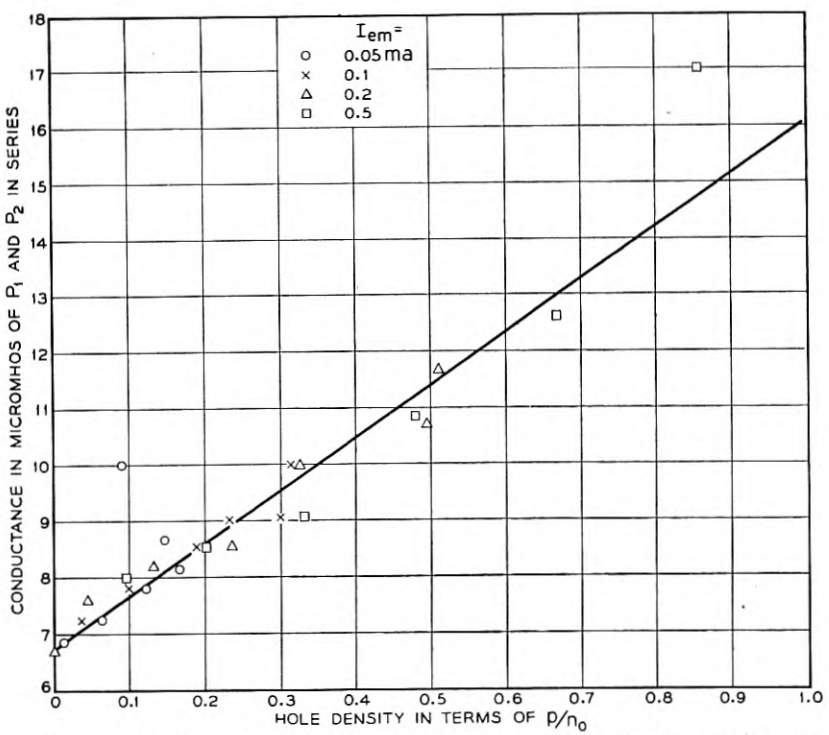


Fig. 6.—The relationship between point admittance and relative hole concentration, for a germanium filament from G. L. Pearson's data of September 28, 1948.

of holes was varied by current from an emitter point near the opposite end of the filament. There was an additional current flowing between

electrodes at the two ends so that the field pulling the holes along the filament could be varied. The concentration of holes was determined from the change in resistivity of that segment of the filament between the two probes. Measurements of admittance were made by passing a small current between the two probes connected in series. The area of the filament is about 1.6×10^{-4} cm^2 and the normal resistance between the probes about 1800 ohms. The normal conductivity is thus

$$\sigma_0 = .009/(1800 \times 1.6 \times 10^{-4}) = 0.03 \text{ (ohm cm)}^{-1}. \tag{39}$$

As shown in Fig. 6, Pearson finds a linear relation between G and p_a. The line best fitting Pearson's data is

$$G = G_0 + (8 \times 10^{-6})\ (p_a/n_0) \text{ (mhos)}. \tag{40}$$

The theoretical value of the coefficient may be obtained from Eq. (38). Taking

$$\alpha = 1, \quad \beta = 0.5, \qquad \sigma_0 = 0.03$$
$$b = 2.0, \ A = 10^{-6} \text{ cm}^2, \ r = 5 \times 10^{-4} \text{ cm}, \tag{41}$$

we get

$$\alpha\beta\sigma_0 A/b\, r_b = 15 \times 10^{-6} \text{ mhos}. \tag{42}$$

Pearson's data, represented by (40), apply to the conductance of two point contacts in series, and the conductance of each one may be about twice that given by (40). Thus the theoretical value is in good agreement with the observed. There is no indication that α differs from unity at low voltage.

Suhl varied the concentration of holes in the vicinity of probe points by application of a transverse magnetic field as well as by injection from an emitter point. The experiment is illustrated in Fig. 7. He used a filament with a cross-section of about .025 × .025 cm. Four probe points were placed along the length of the filament at intervals of about .04 cm. A total current of 4 ma flowed in the filament.

In one experiment, none of this current was injected, so that the concentration of holes was normal in the absence of the magnetic field. Measurements were made of the floating potentials and of the conductances of the probe points. Then a transverse magnetic field was applied and the conductances measured again. We are interested here only in the case of a large field (30,000 gauss) in such a direction as to sweep the holes to the opposite side of the filament. Suhl believes that under these conditions the concentration of holes near the probe points is practically zero. The difference between the conductances with and without the field

then gives the contribution to the conductance from the normal concentration of holes.

In a second experiment 1 ma of the current of 4 ma flowing in the filament was injected from an emitter point near one end of the filament. From the probe potentials, estimates have been made of the change in

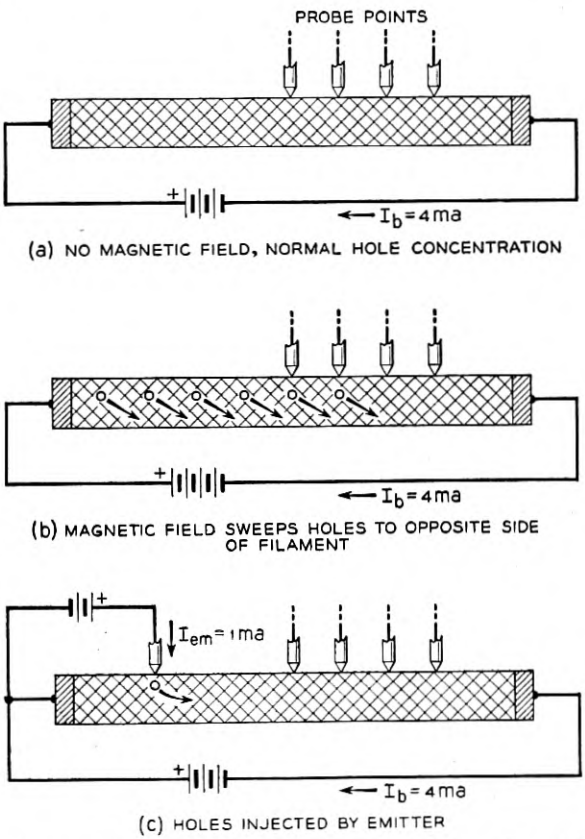


Fig. 7.—Schematic diagram of experiment of H. Suhl to investigate relation between hole concentration and impedance of point contacts.

resistivity and thus of the added hole concentration at the different probe points. Changes in hole concentration from injection have been correlated with changes in admittance of the probe points.

The filament with dimensions .025 × .025 × 0.4 cm has a resistance of 4,600 ohms. The normal resistivity, ρ_0, is then about 7.2 ohm cm. Since the concentration of electrons corresponding to 1.0 ohm cm is about

1.8×10^{15}, the concentration corresponding to a resistivity of 7.2 ohm cm is[11]:

$$n_0 = 1.8 \times 10^{15}/7.2 = 2.5 \times 10^{14}/\text{cm}^3. \tag{43}$$

The product of the equilibrium concentrations of electrons and holes is about 4×10^{26} in germanium at room temperature[12]. Thus, for this sample,

$$p_0 = 4 \times 10^{26}/2.5 \times 10^{14} = 1.5 \times 10^{12}/\text{cm}^3. \tag{44}$$

If there is an added concentration of holes, p_a, resulting from injection, the added conductivity is:

$$\sigma_a = (1 + b)\, e\mu_h p_a = 8.4 \times 10^{-16}\, p_a\,. \tag{45}$$

The resistivity is changed to:

$$\rho = \rho_0\sigma_0/(\sigma_a + \sigma_0) \simeq \rho_0\,(1 - \sigma_a\rho_0), \tag{46}$$

the approximate expression holding if the relative change is small. The resistance per unit length of filament is:

$$R = 1.15 \times 10^4\,(1 - \sigma_a\rho_0). \tag{47}$$

The change in voltage gradient, $dV/dx = RI$, resulting from hole injection is, for a current of 4×10^{-3} amps,

$$\Delta(dV/dx) = d(\Delta V)/dx = -46\rho_0\sigma_a\,. \tag{48}$$

Suhl measured the change in probe potential, ΔV, which resulted when 1 ma of the total current of 4 ma was injected from the emitter instead of having the entire 4 ma flowing between the ends of the filament. His values of ΔV for the four probe points are given in Table II. We have made a plot of these as a function of position and have estimated the gradients at each of the four probe positions. Using these values we have calculated σ_a from Eq. (48) and the corresponding injected hole concentration from Eq. (45). These are given in the last column of the table.

Suhl's measurements of conductances, G, of the probe points are given in Table III. Also given are differences, ΔG, from the normal values with no magnetic field and no injection and also these differences multiplied by n_0/p_a. Values of p_a for the case of hole injection were obtained from Table II. Values of $\Delta G(n_0/p_a)$ are to be compared with the theoretical value,

$$\Delta G(n_0/p_a) = \alpha\beta\sigma_0 A/cr_b\,, \tag{49}$$

[11] These values are based on taking $\mu_n = 3500$ cm²/volt sec and $\mu_p = 1700$ cm²/volt sec, as measured by J. R. Haynes. They correspond to room temperature (295°K).

[12] This value is obtained from an intrinsic resistivity of about 60 ohm cm for *Ge* at room temperature and the mobility values in reference 11.

from Eq. (38). Taking $\alpha = 1$, $\beta = 0.5$, $\sigma_0 = 0.14$, $b = 1.5$, $A = 10^{-6}$ and $r_b = 5 \times 10^{-4}$, we get

$$G(n_0/p_a) \sim 100 \text{ micromhos.} \tag{50}$$

This value is of the same order as the values obtained from Suhl's data listed in Table III. There is a large scatter in the latter and the values are

TABLE II

Calculation of hole concentrations from probe potential measurements. ΔV measures potential difference resulting from hole injection of 1 ma when total current is kept at 4 ma; data from H. Suhl.

Point No.	Relative Position (cm)	ΔV (volts)	$\frac{d\Delta V}{dx}$ (volts/cm)	$\rho_0\sigma_a$	σ_a (mhos)	p_a (cm^{-3})
#6	0	−.04	−0.6	.013	.0018	2.2×10^{12}
#5	.044	−.073	−1.10	.024	.0033	4.0
#4	.084	−.13	−1.8	.039	.0054	6.5
#3	.12	−.21	−2.5	.055	.0077	9.0

TABLE III

Changes in conductance resulting from application of magnetic field and from hole injection. Units are micromhos. Data from H. Suhl.

Point	No Field	With—30,000 gauss field			With hole injection		
	G	G	ΔG	$\Delta G \frac{n_0}{(-p_0)}$	G	ΔG	$\Delta G \frac{n_0}{p_a}$
#6	17.2	16.4	−0.8	130	22.5	7.8	880
#5	6.55	4.35	−2.2	365	7.0	0.45	28
#4	3.7	3.2	−0.5	80	5.1	1.4	54
#3	13.0	9.2	−3.8	630	19	6	165

not consistent. It has been suggested that the abnormal values may result from local sources of holes.

IV. Hole Flow for a Collector with Large Reverse Voltage

Haynes has shown that there is a linear relation between the current to a collector point operated in the reverse direction and the concentration of holes in the interior of a germanium filament. Under the conditions of his experiment, the current flowing to the collector point is small compared with the total current flowing down the filament, so that the collector current does not alter the concentrations very much. Holes are injected into the filament by an emitter point placed near one end, and the concentration is determined from the change in resistance of the filament in the neighborhood of the collector point.

Haynes' measurements may be fitted by an empirical equation of the following form:

$$I = I_0[1 + \gamma p_a/n_0)], \tag{51}$$

in which I_0 is the normal collector current flow for a given collector voltage, I is the collector current flowing for the same collector voltage when the hole concentration is increased by p_a, and n_0 is the normal electron concentration. Values of I_0 and γ for four different formed phosphor-bronze collector points are given in Table IV. The collector bias is -20 volts in each case. It can be seen that the variations in γ are much less than those in I_0. It will be shown below that γ is related to the intrinsic α of the point contact.

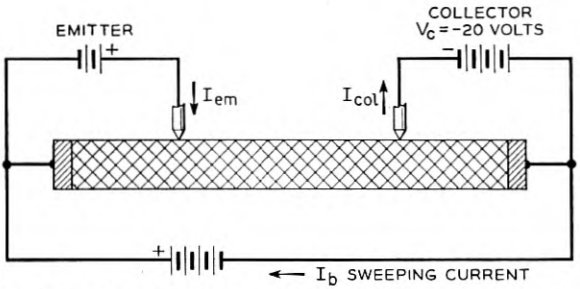


Fig. 8.—Experimental arrangement used by J. R. Haynes to determine relation between hole concentration and current to collector point biased with large voltage in reverse direction.

In Haynes' experiment, holes are attracted to the collector by the field produced by the electron current and diffusion plays a minor role. In contrast to the preceding examples, the terms involving the field F in Eqs. (21) and (22) are large and the diffusion terms represented by the concentration gradients are small. It follows from (21) and (22) that the ratio of electron to hole current density is then:

$$i_n/i_p = bn/p, \tag{52}$$

which is equal to the ratio of the electron and hole contributions to the conductivity. If n and p do not vary with position, the ratio is the same everywhere and equal to the ratio of total electron and hole currents, I_n and I_p:

$$I_n/I_p = i_n/i_p = bn/p. \tag{53}$$

The currents I_n and I_p can also be related to the intrinsic α for the contact by use of an equation of the form:

$$I = I_{n0} + \alpha I_p, \tag{54}$$

in which I_{n0} is the electron current for zero hole current. The electron current is:

$$I_n = I_{n0} + (\alpha - 1)I_p . \tag{55}$$

Thus we have

$$\frac{I_n}{I_p} = \frac{I_{n0} + (\alpha - 1)I_p}{I_p} = \frac{bn}{p} = \frac{b(N_f + p)}{p}. \tag{56}$$

This equation may be solved for I_p to give:

$$I_p = pI_{n0}/(bN_f + (\alpha - 1 - b)p). \tag{57}$$

The term $(\alpha - 1 - b)p$ is generally small compared with bN_f and may be neglected. We thus have approximately for p/N_f small and $N_f \simeq n_0$,

$$I = I_{n0} + \alpha I_p = I_{n0}[1 + (\alpha p/bn_0)]. \tag{58}$$

When expressed in terms of the normal current,

$$I_0 = I_{n0}[1 + (\alpha p_0/bn_0)], \tag{59}$$

the equation for I is of the form (51):

$$I = I_0 [1 + (\alpha p_a/bn_0)]. \tag{60}$$

From a comparison of (51) and (60) it can be seen that:

$$\gamma = \alpha/b \text{ or } \alpha = b\gamma. \tag{61}$$

Values of α determined from empirical values of γ for the four point contacts of Haynes are given in Table IV. The values are of a reasonable order of magnitude for formed collector points.

An estimate of the importance of diffusion can be obtained by comparing the hole current in Haynes' experiments with the hole current which would exist if the electron current were zero, so that holes move by diffusion alone. Equations (28) to (33) apply to the latter case. In addition to (33) we need an equation which expresses the hole current flowing into the contact in terms of the hole concentration, p_b, at the contact. If the reverse bias is large, no holes will flow out and the entire hole current is that from semiconductor to metal as given by an equation similar to (3):

$$I_p = -ep_b v_a A/4. \tag{62}$$

Substituting this value for I_p into equation (33) we get an equation which may be solved for p_b, to give:

$$p_b = ap_i(1 + a) \simeq ap_i, \tag{63}$$

with a given by Eq. (6). Using (63) for p_b, we get:

$$I_p = kT\mu_h p_i A/r_b = (kT\sigma_0 A/ebr_b)(p_i/n_0). \tag{64}$$

With $kT/e = .025$ volts, $\sigma_0 = bn_0e\mu_h = 0.2$ (ohm cm)$^{-1}$, $A = 10^{-6}$ cm^2 and $r_b = 5 \times 10^{-4}$ cm, we get for the diffusion current:

$$I_p = (5 \times 10^{-6})(p_i/n_0) \text{ amps.} \tag{65}$$

Comparing (65) with (57) we see that diffusion of holes will not be important if

$$I_{n0} \gg 5 \times 10^{-6} \text{ amps.} \tag{66}$$

This condition is satisfied in Haynes' experiments.

In the case of point contacts formed to have a high reverse resistance as diodes, I_0 may be of the order of 10^{-7} to 10^{-6} amps at room temperature. Diffusion of holes will then play a role, and the hole current will

TABLE IV

Relation between hole concentration and collector current from data of J. R. Haynes. Data represented by

$$I = I_0(1 + (\gamma p_n/n_0))$$

where I is current flowing to collector point biased at -20 volts and p_n/n_0 is ratio of added hole concentration to the normal electron concentration.

Probe Point	I_0	$\alpha = 2.1\gamma$
0	0.94	4.6
2	0.33	4.4
3	0.54	6.9
4	1.20	4.6

be larger than indicated by Eq. (53). As discussed in reference (4) there is still a question as to the importance of holes in the saturation current observed by Benzer in diodes with high reverse resistance. Experiments similar to those of Haynes would be valuable to determine the influence of hole concentration on reverse current.

ACKNOWLEDGMENT

The author is indebted to G. L. Pearson, J. R. Haynes, W. H. Brattain, and H. Suhl for use of the experimental data presented herein; to W. Shockley for a critical reading of the manuscript and a number of valuable suggestions, and to W. van Roosbroeck for aid with some of the analyses and for suggestions concerning the manuscript.

APPENDIX A

DIFFUSION OF HOLES WITH SURFACE RECOMBINATION

In the calculation of the diffusion of holes given in Section III of the text it was assumed that no recombination of electrons and holes oc-

curred. In the present calculation it is assumed that recombination occurs at the surface, but not in the volume. This is a good approximation for a point contact on germanium. It is further assumed that the hole concentration is sufficiently small so that Laplace's equation (29) may be used.

The model which we shall use is illustrated in Fig. 9. The contact is in the form of a circular disk of radius ρ on the surface of the semiconductor. Cylindrical coordinates, r, θ, z, are used, with the origin at the center of the disk and the positive direction of the z-axis running into the semiconductor. We calculate the flow due to the added holes, and shall use the symbol p without subscript to denote the added hole concentration.

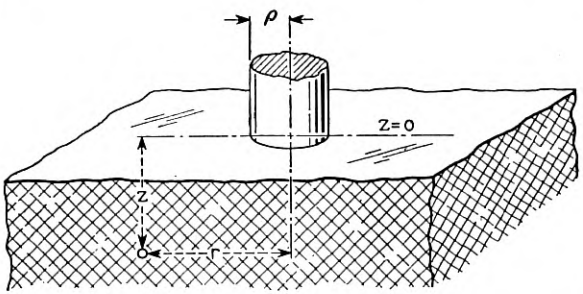


Fig. 9.—Coordinates used for calculation of hole flow to contact area in form of circular disk.

With recombination at the surface, it is necessary to have a gradient in the interior which brings the holes to the surface.

It is assumed that the rate of recombination at the surface is:

$$sp = \text{holes/cm}^2, \tag{1A}$$

where the factor s has the dimensions of a velocity and p is evaluated at the surface $z = 0$. According to measurements of Suhl and Shockley, s is about 1500 cm/sec for a germanium surface treated with the ordinary etch. The current flowing to the surface is:

$$(\mu_p kT/e)(\partial p/\partial z)_{z=0} \text{ holes/cm}^2. \tag{2A}$$

The boundary condition for p at the surface $z = 0$ outside of the contact area is obtained by equating (1A) and (2A). This gives:

$$\partial p/\partial z = \lambda p \text{ at } z = 0, r > \rho \tag{3A}$$

where

$$\lambda = se/\mu_p kT, \tag{4A}$$

has the dimensions of a length. For $s = 1500$ cm/sec and $\mu_p = 1700$ cm^2/volt sec, corresponding to germanium at room temperature, λ is about 35 cm^{-1}.

The boundary condition on the disk is similar to (3A) except that s is replaced by $v_a/4$ (cf. Eq. (3)). Thus for $r < \rho$,

$$\partial p/\partial z = \lambda_e p \qquad z = 0, r < \rho, \tag{5A}$$

where

$$\lambda_e = v_a e/4\mu_p kT. \tag{6A}$$

Evaluated for germanium at room temperature, λ_e is about 6×10^4.

In order to have a dependent variable which vanishes at infinity, we replace p by:

$$y = p_a - p + \lambda p_a z, \tag{7A}$$

so that $p \to p_a$ for $z = 0$ as $r \to \infty$. The variable y satisfies Laplace's equation subject to the boundary conditions:

$$\partial y/\partial z = \lambda y \qquad z = 0, r > \rho \tag{8A}$$

$$\partial y/\partial z = \lambda_e (y - p_a) \; z = 0, r < \rho \tag{9A}$$

$$y = 0 \qquad r, z \to \infty. \tag{10A}$$

An exact solution of the problem is difficult. We shall obtain an approximate solution which satisfies (8A) but not (9A) and which applies when

$$\lambda\rho \ll 1 \ll \lambda_e\rho. \tag{11A}$$

This approximation is valid for a germanium point contact, since, for $\rho \sim 10^{-3}$ cm,

$$\lambda\rho \sim .035, \lambda_e\rho \sim 60. \tag{12A}$$

We shall first discuss the limiting case for which $\lambda \to 0$ and $\lambda_e \to \infty$. The former implies neglect of surface recombination and the latter

$$y = p_a \text{ for } z = 0, r < \rho. \tag{13A}$$

The problem is the same as that of finding the potential due to a conducting circular disk. The solution of this problem, which is well known, is:

$$y = (2p_a/\pi) \int_0^\infty e^{-zt} J_0(rt) \frac{\sin \rho t}{t} \, dt. \tag{14A}$$

The current flowing to the disk is obtained from integrating:

$$i_z = kT\mu_p(\partial y/\partial z), \tag{15A}$$

over the area of the disk. This gives:

$$I_{pa} = -4\rho p_a kT\mu_p. \tag{16A}$$

The analogous expression for a hemispherical contact area of radius r_b, obtained from (7), is:

$$I_{pa} = -2\pi r_b p_a kT\mu_p. \tag{17A}$$

If a comparison is made on the basis of equal radii, (17A) is larger than (16A) by a factor of $\pi/2$. On the more reasonable basis of equal contact areas, (16A) is larger than (17A) by a factor of $4/\pi$.

An approximate solution which includes surface recombination can be obtained as follows. A solution of Laplace's equation which satisfies (8A) and (10A) is:

$$y = \frac{2y_0}{\pi}\int_0^\infty e^{-zt}J_0(rt)\frac{\sin \rho t}{t+\lambda}\,dt. \tag{18A}$$

That (18A) satisfies (8A) may be verified by direct substitution:

$$\left[-\frac{\partial y}{\partial z} + \lambda y\right]_{z=0} = \frac{2y_0}{\pi}\int_0^\infty J_0(rt)\sin \rho t\,dt = 0 \quad \text{for} \quad r > \rho. \tag{19A}$$

$$= (2y_0/\pi)(\rho^2 - r^2)^{-1/2} \quad \text{for} \quad r < \rho. \tag{20A}$$

Expression (18A) satisfies (9A) approximately if λ_c is large. Using (20A) and neglecting λ in comparison with λ_c, we have:

$$y = p_a - (2y_0/\pi\lambda_c)(\rho^2 - r^2)^{-1/2} \quad \text{for} \quad z = 0, r < \rho. \tag{21A}$$

Except for r almost equal to ρ, the second term on the right of (21A) is very small. It is not possible to obtain an explicit expression for y for $r < \rho$. For $z = 0, r = \rho$,

$$y = \frac{2y_0}{\pi}\int_0^\infty \frac{J_0(\rho t)\sin \rho t}{t+\lambda}\,dt = y_0 F(\lambda\rho). \tag{22A}$$

The integral, $F(\lambda\rho)$, can be evaluated from a more general integral in Watson's Bessel Functions, p. 433. We have:

$$F(k) = \frac{2}{\pi}\int_0^\infty \frac{J_0(x)\sin x\,dx}{x+k} = \cos k\,J_0(k) + \sin k\,Y_0(k). \tag{22B}$$

The factor multiplying y_0 is unity for $\lambda\rho = 0$, and decreases as $\lambda\rho$ increases. Since y is approximately equal to p_a, we have, approximately,

$$y_0 = p_a/F(\lambda\rho). \tag{23A}$$

The value of y can also be found for $r = 0$. For $z = 0$, $r = 0$, we have:

$$y = \frac{2y_0}{\pi} \int \frac{\sin \rho t \, dt}{t + \lambda} = \frac{2y_0}{\pi} G(\lambda\rho). \tag{24A}$$

The integral can be expressed in terms of integral sine and cosine functions:

$$G(k) = \frac{2}{\pi} \int_0^\infty \frac{\sin x \, dx}{x + \lambda} = \frac{2}{\pi} \left[-\cos k \left(\mathrm{Si}\ k - \frac{\pi}{2} \right) + \sin k \ \mathrm{Ci}\ k \right]. \tag{25A}$$

If k is not too large, $G(k)$ is nearly equal to $F(k)$, so that y is approximately constant over the area of the disk.

The total current flowing from the contact is found from integrating $kT\mu_p \, (\partial y/\partial z)$ over the disk:

$$I_{pa} = -kT\mu_p y_0 \int_0^\rho \int_0^\infty \frac{4rtJ_0(rt) \sin \rho t}{t + \lambda} \, dt \, dr \tag{26A}$$

$$= -4kT\mu_p y_0 \int_0^\infty \frac{\rho J_1(\rho t) \sin \rho t}{t + \lambda} \, dt. \tag{27A}$$

The integral can be evaluated with use of the general integral of Watson, to give:

$$I_{pa} = -4\rho kT\mu_p y_0 \, H(\lambda\rho), \tag{28A}$$

where

$$H(k) = \int_0^\infty \frac{J_1(x) \sin x \, dx}{x + k} = -\frac{\pi}{2} [\cos k \, J_1(k) + \sin k \, Y_1(k)]. \tag{29A}$$

Using (23A) for y_0, we have:

$$I_{pa} = -4\rho kT\mu_p p_a [H(\lambda\rho)/F(\lambda\rho)]. \tag{30A}$$

Except for the factor $H(\lambda\rho)/F(\lambda\rho)$, this expression for the current is identical with (16A). This factor, which gives the effect of recombination on the current, is plotted in Fig. 10. Recombination gives an increase in current flow, but the effect is small for the normal rate of surface recombination, which corresponds to $k = \lambda\rho \sim .035$.

APPENDIX B

Calculation of Hole Flow for Arbitrary Hole Concentration

In the text it was assumed that the concentration of holes was sufficiently small so that the first term in the brackets of Eq. (26) could be neglected in comparison with unity, yielding Eqs. (28) and (29). We give

here the general integration of Eqs. (26) and (27) for p arbitrarily large. Equation (26) may be written in the form:

$$i_p = -\text{grad } \psi, \tag{1B}$$

where

$$\psi = kT\mu_p \left[\frac{2bp}{b+1} - \frac{b(b-1)N_f}{(b+1)^2} \log \left(1 + \frac{(b+1)p}{bN_f} \right) \right]. \tag{2B}$$

Equation (27) then becomes:

$$\nabla^2 \psi = 0. \tag{3B}$$

The radial solution of this equation corresponding to a total current I_p is:

$$\psi = \psi_\infty + I_p/2\pi r. \tag{4B}$$

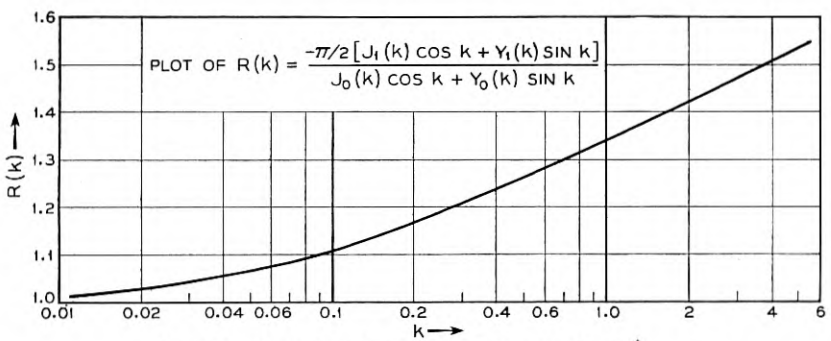


Fig. 10.—Correction factor for surface recombination.

The constants I_p and ψ_∞ are determined from the boundary conditions (30) and (31) of the text corresponding to $r = r_b$ and $r = \infty$. These conditions give:

$$\psi_\infty = kT\mu \left[\frac{2bp_i}{b+1} - \frac{2(b-1)N_f}{(b+1)^2} \log \left(1 + \frac{(b+1)p_i}{bN_f} \right) \right], \tag{5B}$$

$$I_p = 2\pi r_b \left(\psi(r_b) - \psi_\infty \right),$$

$$= 2\pi r_b \left[\frac{2b(p_b - p_i)}{b+1} - \frac{b(b-1)N_f}{(b+1)^2} \log \frac{bN_f + (b+1)p_b}{bN_f + (b+1)p_i} \right]. \tag{6B}$$

This equation is the appropriate generalization of Eq. (33) of the text. Since the equations are no longer linear, they do not apply strictly to the added hole concentration. However, if the normal hole concentration, p_0, is small, p_0 will be negligible in comparison with p_{ba} and p_a when the equa-

tions are not linear. Accordingly, to a close approximation, we may take for the added hole current:

$$I_{pa} = 2\pi r_b \left[\frac{2b(p_{ba} - p_a)}{b+1} - \frac{b(b-1)}{(b+1)^2} \log \frac{bN_f + (b+1)p_{ba}}{bN_f + (b+1)p_a} \right], \quad (7B)$$

which is the generalization of Eq. (34) of the text.

The value of p_{ba} and thus of I_{pa} may then be found by equating this expression with that of Eq. (3) for I_{pa}. This procedure yields the transcendental equation:

$$p_{ba} = -a \left[\frac{2b(p_{ba} - p_a)}{b+1} - \frac{b(b-1)N_f}{(b+1)^2} \log \frac{bN_f + (b+1)p_{ba}}{bN_f + (b+1)p_a} \right], \quad (8B)$$

where a is again defined by Eq. (6) of the text. This equation must be solved in general by numerical methods for a particular case. The equation simplifies for p_a either large or small compared with N_f. The latter case is treated in the text. The opposite limiting case of large hole concentrations is treated below.

For p_a large compared with N_f, the logarithm may be neglected, so that

$$p_{ba} = -2ab(p_{ba} - p_a)/(b+1). \quad (9B)$$

If, as in the text, it is assumed that a is small in comparison with unity, there results:

$$p_{ba} = 2abp_a/(b+1), \quad (10B)$$

and, using (3):

$$I_{pa} = -[2b/(b+1)]p_a kT\mu_p A/r_b. \quad (11B)$$

This differs from (7) by a factor $2b/(b+1)$. The equation corresponding to (8) will have this additional factor, and also the expression for the conductance, G, which, for large hole concentrations is:

$$G = G_0 + [2b/(b+1)](\alpha\beta\sigma_0 A/br_b)(p_a/n_0), \quad (12B)$$

in place of (38) of the text. Equation (16) which relates floating potential and conductance is general, and applies for arbitrary hole concentration.

Design Factors of the Bell Telephone Laboratories 1553 Triode

By J. A. MORTON and R. M. RYDER

(*Manuscript Received Aug. 3, 1950*)

IN DEVELOPING microwave relay systems for frequencies around 4000 megacycles, one of the major problems is to provide an amplifier tube which will meet the requirements on gain, power output, and distortion over very wide bands. As the number of repeaters is increased to extend the relay to greater distances, the requirements on individual amplifiers for the system become increasingly severe. A tube developed for this service is the microwave triode B.T.L. 1553, the physical and electrical characteristics of which were briefly described in a previous article.[1] In the development of such a tube, both theoretical and experimental factors are involved; illustration of these factors in some detail is the purpose of the present paper.

Given the application, a number of questions arise at the outset. What determines the tube type—why pick a triode for development, rather than a velocity variation tube, or perhaps a tetrode? What electrode spacings are necessary in such a tube, and what current must it draw? How is its performance rated, and how does it compare with other tubes? To what extent can the performance be estimated in advance? What experimental tests can give more precise information? Some answers to these questions were obtained by the use of figures of merit, which led up to the choice of a triode as most promising for development, and which also led to the subsequent method of optimizing the design for the particular system application of microwave amplifiers and modulators.

The design process may be said to proceed by the following series of steps:

1. Formulate the system requirements, frequently with the aid of one or more figures of merit. The purpose here is to concentrate attention upon the limitations inherent in the tube alone by eliminating considerations of circuitry or of other parts of the system. The figure of merit measures tube performance in an arbitrary environment, so chosen as to be simple, and also directly comparable to the actual system requirement.
2. Make tentative choices of tube type, and analyze further to find out

[1] J. A. Morton, "A Microwave Triode for Radio Relay," *Bell Laboratories Record* 27, 166–170 (May 1949).

how the figure of merit depends on the internal parameters of the tube, such as spacings, current density, and so on.

3. Optimize the internal parameters to make the figure of merit as good as possible, with due regard to practical limitations like cathode activity, life, cost, etc.

4. Use enough experimental checks to make sure the estimates are sound. Build then the type of tube which appears to fill the requirements best, including the practical as well as the technical limitations. The figures of merit serve now as quantitative checks both of how well the tube satisfies the application, and also of how accurate is the theory.

Given a good accurate design theory, the whole process could in principle be calculated in advance. Such a theory would permit great savings in effort, since spot checks of relatively few parameters are sufficient to insure accuracy even when the theory is used to predict a wide range of phenomena. The extent to which presently available microwave tube theory meets this need is considerable, as will appear from some of the results below.

The degree of accuracy required of a theory increases as the development process continues. For preliminary estimates, such as deciding what tube type to develop, the theory can be rather rough and still be satisfactory. For complete predictions of final performance, only experimental construction can suffice. By this means the theory can be checked, so that it can serve future designs with improved accuracy.

The method outlined here is not new, but rather follows standard practice fairly closely. It does, however, give more than usual quantitative emphasis to the figures of merit, using them to codify the procedure; and it incorporates a certain amount of quantitative calculation at microwave frequencies. It will be seen that the theory of Llewellyn and Peterson needs only some semi-empirical supplementation in the low-voltage input space, as has already been pointed out by Peterson.[2]

Preliminary Estimates—Choice of Tube Type

For the New York to Boston microwave relay, an output amplifier was developed using already available velocity-modulation tubes.[3] With four stagger-tuned stages, the amplifier proved satisfactory for this service, and in fact tests indicated that this system could be extended to considerably greater distances and still give good performance. It was apparent, however, that these amplifiers would not be satisfactory for a coast-to-coast system.

[2] L. C. Peterson, "Signal and Noise in Microwave Tetrodes," *I. R. E. Proc.* (Nov. 1947).

[3] H. T. Friis, "Microwave Repeater Research," *B. S. T. J.* 27, 183–246 (April 1948).

When this limitation became clear several years ago, a study was undertaken to determine which particular type of electron tube amplifier then known had the best possibilities of being pushed to greater gain-band products. The results of this study indicated that a very promising prospect was to build, for operation at 4000 megacycles, an improved planar triode, that is, one in which the active elements are on parallel planes.

In arriving at this conclusion, two general types of device were considered: velocity-modulated, as in a klystron, and current-modulated, as in a triode. (Nowadays, such a study would of course include traveling-wave tubes.) The conclusions were reached with the aid of the gain-band figures of merit, along the following lines:

Gain-Band Product

The system performance requirements demand amplifiers capable of reasonable gains and power outputs over prescribed bandwidths. However, it is known that bandwidth can be increased by complicating the circuits (double-tuning, stagger-tuning, etc.). Such factors, being common to whatever tube may be used, are extraneous to a discussion of tube performance, and accordingly the tubes are rated by their performance with simple, synchronous resonant circuits. Furthermore, even then the bandwidth can be increased at the expense of a corresponding reduction of gain, by simply depressing the impedance levels of the interstages. Since the product of gain and bandwidth remains constant, it is a suitable figure of merit, independent of the particular choice of bandwidth, provided the definition of gain is suited to the device.

Unfortunately there is more than one possible gain-band product, the appropriate form depending on how many simple resonant circuits shape the band of the amplifier stage. For example, a conventional pentode or a velocity-variation tube is usually used in conjunction with two high-Q resonant circuits, one each on input and output. If these are adjusted to give the same Q, then it is well known that, no matter what the bandwidth, the product of voltage gain and bandwidth is constant. (See Appendix 1)

$$|\Gamma_0| B = |Y_{21}| / 2\pi\sqrt{C_{in} C_{out}} \tag{1}$$

Here Γ_0 is the mid-band voltage gain, B the bandwidth 6 db down (3 for each circuit), Y_{21} the stage transadmittance, and C_{in} and C_{out} the total effective capacitances of the resonant circuits, including the contributions of the tube*. It is assumed that the stage is matched into transmission lines of some suitable constant admittance level G_0.

In amplifiers using triodes such as the B.T.L. 1553 (or tetrodes) in

* As shown in Appendix 1, all quantities in equations (1) and (2) are the values effective at the electrodes adjacent to the electron stream.

grounded-grid circuits, the situation is different because the Q of the input circuit is always very much smaller than that of the output. Here a figure of merit independent of bandwidth is obtained from the product of power gain and bandwidth:

$$|\Gamma_0|^2 B = |Y_{21}|^2/4\pi G_{in} C_{out} \qquad (2)$$

Here G_{in} is the total conductance of the input circuit, including tube contributions*. The gain is again measured with the tube matched at an arbitrary admittance level G_0. The band, being now limited by only one tuned circuit, is somewhat different in shape from the above, and is taken 3 db down.

While each figure of merit gives an unequivocal rating of tubes of appropriate type, the intercomparison of the two types still depends on the bandwidth. In particular, as the band is widened, the two-circuit type (klystron) loses gain at the rate of 6 db per octave of bandwidth, while the one-circuit type (triode) loses only 3 db per octave. Consequently, if the two devices start with equal gains at some narrow bandwidth, the triode rapidly pulls ahead in gain as the bandwidth is increased.

The figure of merit equation (1) states that improved klystron performance implies either an increase in transadmittance Y_{21} or a decrease in the band-limiting capacitances C_{in} or C_{out}. According to the simplest klystron bunching concept,[4] the transconductance of such a tube may be increased indefinitely simply by making the drift time longer. Unfortunately, this simple kinetic picture does not take account of the mutually repulsive space-charge effects which set an upper limit to the useful drift time by debunching the electrons.[5] For a 2000-volt beam in the 4000-megacycle range, this limit is approximately three micromhos per milliampere. The 402A tube used in the New York to Boston system has already approached this limit within a factor of two. Since the capacitances are also quite small, the prospect is quite dubious for any considerable improvement in gain-band merit if the simple klystron type of operation were to be used.

Improvements are possible in a klystron by changing the manner of operation so as to lower the drift voltage V_0, because the aforesaid transadmittance limit is proportional to V_0^{-n}.* This prospect is also relatively unattractive. To get transadmittance values anywhere near the triode would require low voltages and close spacings somewhat like the latter, and would encounter space-charge difficulties involved in handling a large current in a low-voltage drift space. Furthermore, the tube would be more

[4] D. L. Webster, *Jour. App. Phys.* 10, 501–508 (July 1939).
[5] S. Ramo, *Proc. I. R. E.* 27, 757–763 (December 1939).
* The value of n may vary between ¼ and ¾. See reference 5.

complex, having several grids instead of one. A number of modifications of klystron operation were considered, but all looked more complex mechanically and more speculative theoretically than a triode.

In a triode there is also an upper limit to the transconductance that can be achieved by spacing cathode and grid more closely. This limit would be reached if the spacing were so close that the velocity produced by the grid voltage were of the same order as the average thermal velocity of cathode emission. The triode limit of some 11,000 micromhos per milliampere is, however, many times greater than that for ordinary klystrons. What is still more important is the fact that previous microwave triodes were still a factor of twenty to twenty-five below this limit, leaving considerable room for improvement. Thus, if mechanical methods could be devised for decreasing the cathode-grid spacing and at the same time maintaining parallelism between cathode and grid, it seemed highly probable that great improvements would be available from a new triode.

The choice to develop a triode for this application was therefore taken not merely on the basis of simplicity, but also with the expectation that performance improvements would be not only larger but also more certainly obtainable than by use of a modified klystron. Moreover, the possibilities of using the triode over a wide frequency range in other ways—as a low noise amplifier, modulator and oscillator—lent additional weight to its choice. By translating the known requirements on gain, bandwidth and power output into triode dimensions as discussed below, it was found that the input spacings of existing commercial tubes would have to be reduced by a factor of about five. In addition, cathode emission current densities would have to be increased about three times. A design was evolved in which the required close spacings could be produced to close tolerances by methods consistent with quantity production requirements. The B.T.L. 1553 tube was the result (Fig. 1). Many of its design features were adopted for use in the Western Electric 416A tube, which is an outgrowth of this investigation.

Description of B.T.L. 1553 Triode*

The electrode spacings of this tube and of a 2C40 microwave triode are shown in Fig. 2. In the 1553, the cathode-oxide coating is .0005″ thick, the cathode grid spacing is .0006″, the grid wires are .0003″ in diameter, wound at 1000 turns per inch, and the plate-grid spacing is .012″. It is interesting to note that the whole input region of the 1553 including the grid is well within the coating thickness of the older triode.

The arrangement of the major active elements of the tube is shown in

* This section is repeated from reference 1 for completeness.

Fig. 3. This perspective sketch has been made much out of scale so that the very close spacings and small parts would be seen. The nickel core of the cathode is mounted in a ring of low-loss ceramic in such a manner that the nickel and ceramic surfaces may be precision ground flat and coplanar. A thin, smooth oxide coating is applied to the upper surface of

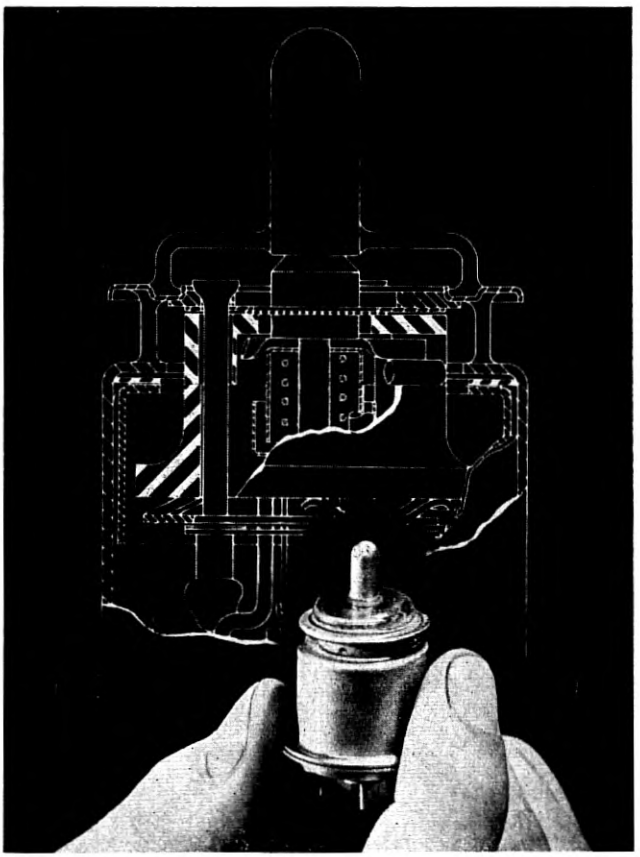


Fig. 1.—The B.T.L. 1553 microwave triode with a cross-section drawing of it in the background.

the cathode by an automatic spray machine developed especially for this tube. With this machine, a coating of 0.0005″ ± 0.00002″ may be put on under controlled and specifiable conditions. To insure long life with such a thin coating, it was necessary to develop coatings from two to four times as dense as those used in existing commercial practice.

The grid wires are wound around a flat, polished molybdenum frame

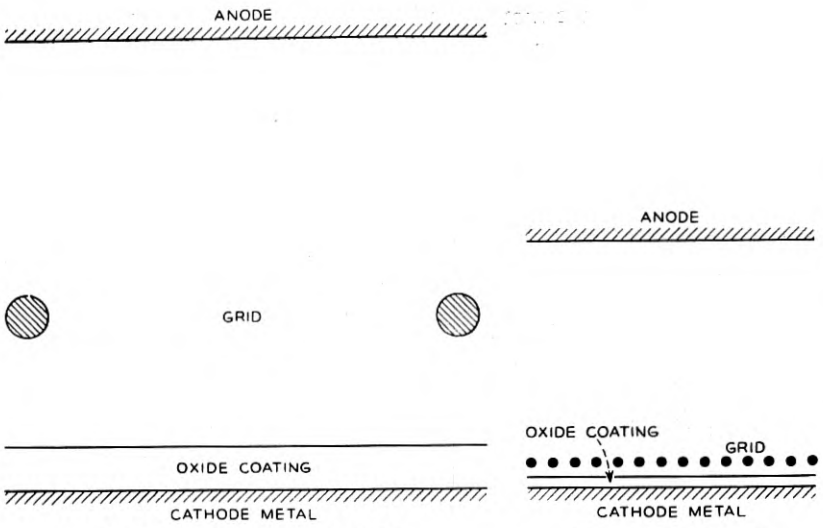


Fig. 2.—Comparison of the spacings of the 1553 triode at the right with a previously existing microwave triode at the left.

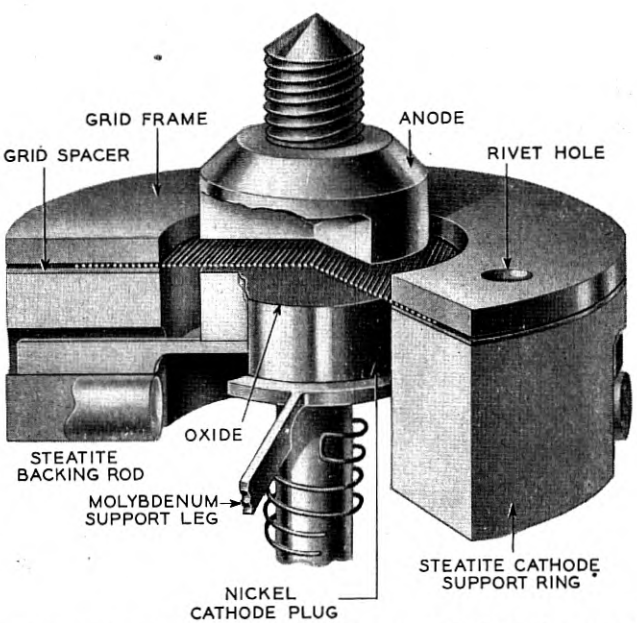


Fig. 3.—Perspective drawing of the active elements of the 1553 close-spaced triode.

that has been previously gold sputtered. The winding tension is held within ±1 gram weight to about 15 gram weight, which is about sixty per cent of the breaking strength of the wire. This is accomplished by means of a small drag-cup motor brake, a new method which was developed especially for these fine grids. The grid is then heated in hydrogen to about 1100°C, at which point the gold melts and brazes the wires to the frame. The mean deviation in wire spacing is less than about ten per

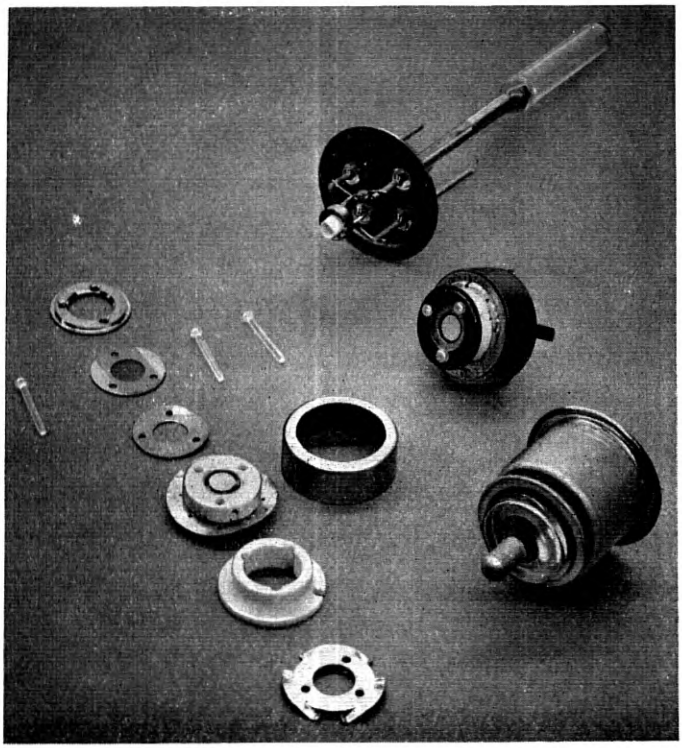


Fig. 4.—Physical appearance of the elements comprising the 1553 triode.

cent, and in fact these grids are fine enough and regular enough to be diffraction gratings as is shown in Fig. 5. In this figure, a fourth order spectrum diffracted by one of these grids can be seen. The third order, which should be absent because the wire size is about one-third of the pitch, is much less intense than the fourth. Proper spacing of the grid is then obtained by a thin copper shim placed between the cathode ceramic and the grid frame. Its thickness must be equal to the coating thickness, plus the thermal motion of the cathode, plus the desired hot spacing.

The cathode, spacer, and grid comprising the cathode-grid subassembly are riveted together under several pounds of force maintained by the molybdenum spring on the bottom of the assembly. The rivets are three synthetic sapphire rods fired on the ends with matching glass. In Fig. 4, the parts comprising this assembly are shown in appropriate pile-up sequence at the left, and the completed cathode-grid subassembly is shown at the right between the bulb and the press. The grid-anode spacing of .012″ is easily obtained by means of an adjustable anode plug the surface of which is gauged relative to the bulb grid disc.

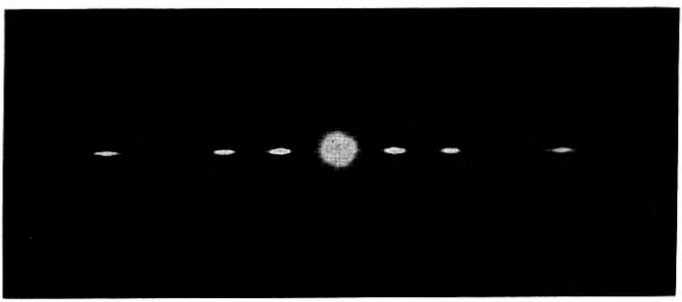


Fig. 5.—Spectrum formed by the grid of the 1553 microwave triode.

TABLE I

LOW-FREQUENCY CHARACTERISTICS

For $V_p = 250$ V, $I_p = 25$ ma, $V_g = -0.3$ V

$g_m = 50{,}000$ μmhos	$C_{kg} = 10$ μμf
$\mu = 350$	$C_{gp} = 1.05$ μμf
$r_p = 7000$ ohms	$C_{kp} = .005$ μμf

The higher current density of 180 milliamperes per square centimeter, the thin dense cathode coating, and the very close spacings, posed a problem in obtaining adequate emission and freedom from particle shorts, and had to be solved by quality control methods because of the large number of factors involved and the precision required. Tubes, subassemblies, and testers have been made in batches and studied by statistical methods. To achieve a state of statistical control on emission, and freedom from dust particles, it is necessary to process the parts and assemble the tubes in a rigorously controlled environment. Completely air-conditioned processing and assembly rooms operating under rigorous controls have been found necessary[6]. Under such controlled conditions, good production yields with satisfactory cathode activity have been obtained,

[6] R. L. Vance, *Bell Laboratories Record*, 27, 205–209 (June 1949).

whereas without such conditions not only was the yield low but it was difficult to ascertain just what factors were operating to inhibit emission and to cause cathode-grid shorts.

A summary of the pertinent low-frequency characteristics of the 1553 triode is given in Table I. It should be noticed that, at plate currents of 25 milliamperes, the transconductance per milliampere is about 2000, that is, about one-fifth of the theoretical upper limit. At lower currents this ratio is higher: at 10 milliamperes, for example, it is 3000 micromhos per milliampere. Diodes with the same spacings have about twice these values of transconductance per milliampere, showing that the grid is fine enough to obtain fifty per cent of the performance of an ideal grid.

Triode Design Requirements

Analysis of the figure of merit can well begin by devoting attention to the band-limiting capacitance C_{out} of the output circuit. First, some question may be raised as to the applicability of the concept of a simple *L-C* shunt resonant circuit at high frequencies, where the circuit parameters are actually distributed, not lumped. Suppose the actual circuit admittance is $Y_x = G_x + jB_x$. In order to represent it as a simple shunt resonant circuit of admittance $Y_p = G_p + j\omega C_p + 1/j\omega L_p$, we need only require that the two be equal and have equal derivatives with respect to frequency at the center frequency $f_0 = \omega_0/2\pi$. Accordingly the "effective values" of the actual admittance are given by the following equations:

$$\begin{aligned} G_p &= G_x\,(\omega_0) \\ C_p &= \tfrac{1}{2}(B_x' + B_x/\omega_0) \\ \frac{1}{L_p} &= \tfrac{1}{2}(\omega_0^2\, B_x' - \omega_0 B_x) \end{aligned} \tag{3}$$

From this development one sees that the representation neglects G_x', the first derivative of the conductance, but otherwise is correct to first order as a function of frequency.

There are important cases where this representation as a simple circuit does not hold. For example, double-tuned circuits having two local resonances have a fundamentally different band shape. However, such complication of the circuits has been excluded from the figure of merit on the ground that it is purely a circuit "broad-banding" problem: having determined the performance of the tube for simple circuits, any broad-banding (double-tuning, staggering, etc.) will give a calculable improvement which does not depend upon the tube. Accordingly, to compare tubes it is sufficient to consider standard simple circuit terminations, tuned to the same frequency.

The total capacitance C_{out} includes two contributions: from the active electrode area inside the tube (C_{22}) and from the passive resonating circuit (C_{p2}). It is convenient to consider these separately, writing the figure of merit as follows:

$$|\Gamma_0|^2 B = \frac{|Y_{21}^2|}{4\pi G_{11} C_{22}} \frac{1}{\left(1 + \frac{G_{p1}}{G_{11}}\right)\left(1 + \frac{C_{p2}}{C_{22}}\right)} \tag{4}$$

The first factor is the "intrinsic" electronic figure of merit of the active transducer alone, while the second factor expresses the deterioration caused by input passive circuit loss G_{p1} and output passive circuit capacitance C_{p2}, both of which should ideally be held as small as possible.

Consider the first factor, the intrinsic electronic gainband product which depends only upon the properties of the electron stream and the electrode dimensions in the regions occupied by the electron stream.

It is the responsibility of the tube design engineer to maximize this product consistent with any limitations which may be imposed by mechanical, emission, thermal or circuital considerations.

On the other hand, in maximizing this intrinsic gain-band product, the tube engineer must not proceed in ignorance of the effect of his actions on the possibility of obtaining a favorable value for the second factor. For example, he may attempt to make C_{22} so small (in order to maximize the first factor) that it becomes physically impossible to obtain an effective circuit capacitance C_{p2} which is not large compared to C_{22}. In such a case, the actual gain-band product would be much smaller than the intrinsic product of which the tube would be capable if circuit capacitance were negligible. Such a balancing of effects will become apparent from the subsequent discussion.

It is desired, therefore, to express the transadmittance, input conductance and output capacitance of the electronic transducer in terms of such parameters as cathode current density, electrode dimensions, frequency and potentials in such a way that it will become clear how a maximizing process may be carried out by adjusting these parameters.

As a first approximation let us use the results of Llewellyn and Peterson's analysis of plane-parallel flow[7], which makes the following assumptions:

1. All electrons are emitted with zero velocity.
2. All electrons in a given plane have the same velocity.

[7] F. B. Llewellyn and L. C. Peterson, "Vacuum Tube Networks," *Proc. I. R. E.*, 32, 144–166 (1944).

3. The dimensions of the grid are infinitesimal compared to the electrode spacings.
4. The electrode dimensions are small compared to the wavelength.

It can be shown that the intrinsic gain-band product may be expressed in the following two ways:

$$\begin{aligned} M_i &= K\left[\frac{1}{x_1}\right]\left[\frac{F_1^2(\theta_1)}{\theta_1 F_3(\theta_1)}\right][\theta_2 F_2^2(\theta_2)\sqrt{V_p}] \\ &= K'\left[\frac{1}{j}\right]\left[\frac{F_1^2(\theta_1)}{\theta_1^4 F_3(\theta_1)}\right][\theta_2 F_2^2(\theta)\sqrt{V_p}] \end{aligned} \tag{5}$$

where K, K' are parameters which are functions only of frequency.

x_1 is the cathode-grid spacing in cm

θ_1 is cathode-grid transit angle and $\theta_1 = \dfrac{126}{\lambda}\left(\dfrac{x_1}{j}\right)^{1/3}$

j = cathode current density in amp/cm^2

θ_2 = grid-anode transit angle and $\theta_2 = \dfrac{6300\ x_2}{\lambda\sqrt{V_p}}$

and $F_1(\theta_1)$, $F_2(\theta_2)$ and $F_3(\theta_1)$ are complicated functions of their respective transit angles.

Consider frequency to be given as part of the specifications on the tube.

Variation with Current Density, j

In the first formulation the current density is involved only in the second factor. This factor is a function only of $\theta_1 = \left(\dfrac{x_1}{j}\right)^{1/3}$ and is shown plotted in Fig. 6. If x_1 and λ are considered to be held fixed for the moment the first maximum at $\theta_1 \to 0$ requires j to be as large as possible consistent with emission limitations and life. For the 1553 the cathode current density is set at 180 ma/cm^2.

The other maxima at larger values of θ_1 (and smaller values of j), where $F_3(\theta_1)$ goes through zero, correspond to transit angles where $G_{11} \to 0$ in the single-valued velocity theory. These maxima cannot be taken at face value, however, to indicate maxima in the unequal-Q gain-band product since they violate the assumption that $Q_1 << Q_2$ for which the formula was developed. To make a study of gain-band variation in this region therefore entails a study of gain-band product as a function of bandwidth, as was pointed out previously in connection with comparison of the equal-Q and unequal-Q cases. Such maxima are of interest pri-

marily in narrow band cases so that for the present we shall concern ourselves only with the first maximum at $\theta_1 \rightarrow 0$ and j indefinitely large.

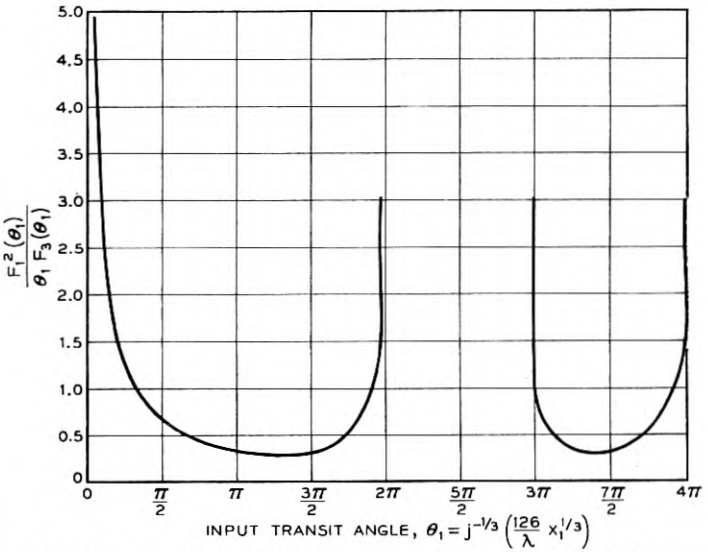


Fig. 6.—Gain-band product dependence on current density (j), with input spacing (x_1) fixed.

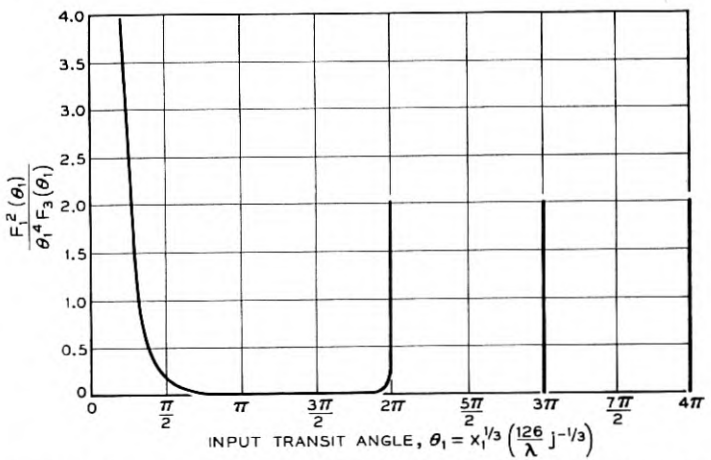


Fig. 7.—Gain-band product dependence on input spacing (x_1), with current density (j) fixed.

Variation with Cathode-Grid Spacing, x_1

Now consider that j has been fixed at the largest permissible value according to the previous section and consider the second formulation

for M_i. The spacing x_1 is involved only in the second factor which again is a function only of $\theta_1 = \left(\frac{x_1}{j}\right)^{1/3}$. We again have a strong first maximum at $\theta \rightarrow 0$ requiring x_1 to be as small as possible (Fig. 7). Other maxima are indicated at larger values of θ_1 (and larger values of x_1) again at points where $G_{11} \rightarrow 0$ and the same remarks apply here as were made in the previous section. For broad-band optima we are therefore interested in minimum values of x_1.

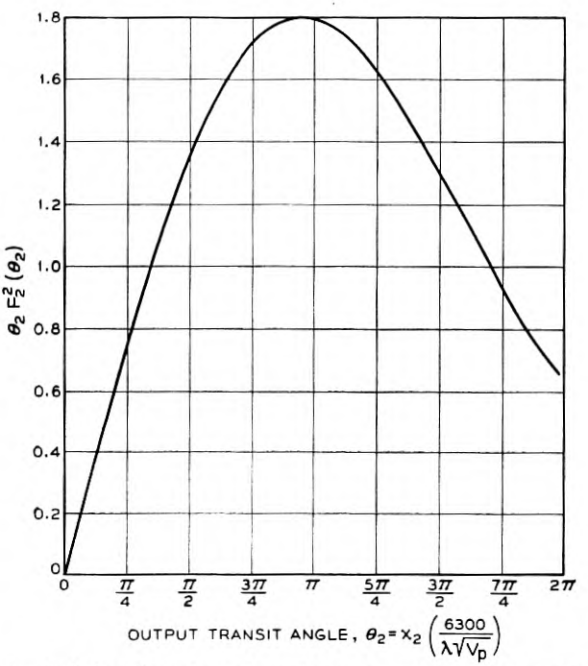


Fig. 8.—Gain-band product dependence on output spacing (x_2).

Variation with Anode-Grid Spacing, x_2

The anode-grid spacing x_2 is involved only in the third factor of either formulation. This factor is a function of output transit angle θ_2 and exhibits a maximum for $\theta = 2.9$ radians as shown in Fig. 8. This optimum at a fairly large value of θ_2 is due to the fact that the capacitance C_{22} varies as $1/x_2$ whereas the coupling coefficient of the stream to the gap decreases more slowly at first than the capacitance so that the ratio Y^2_{21}/C_{22} improves as the spacing becomes moderately wide. The optimum θ_2 corresponds to an optimum value of x_2 which of course depends upon the plate voltage and frequency of operation. For the 1553 at 250 volts and 4000 Mc/s, the optimum output spacing is .022″.

Limitations in Choosing Optimum Parameters

Generally, there are mechanical, thermal, emission and specification limits which prevent the realization of optimum values for all of the above parameters simultaneously. A good design is one in which a nice balance is effected between these various optima and their limitations.

Limitations on Emission Current Density, j

It is generally true that the life of a thermionic electron tube varies inversely as the average cathode current density in a complicated fashion. The maximum permissible value of j is therefore always a compromise between our desire for highest figure of merit and long life. In the present state of the cathode art as it has been evolved for the 1553 triode it is possible to operate at a current density of 180 ma/cm^2 and obtain an average life of several thousands of hours. It is perhaps of interest to note that it was necessary to develop much more dense and smooth oxide coatings in order to make possible such life in the thin coatings necessary for operation at such close spacings.

Limitations on Cathode-Grid Spacing, x_1

Consider the limitations in reaching the optimum in x_1. There is, of course, the obvious one that it is mechanically and electrically not possible at present to make x_1 equal to zero and still retain the essential features of unilateral controlled space charge flow. Granting then that the spacing cannot be zero, we must choose the smallest value of x_1 for which parallelism and reasonable tolerances can be maintained. To this end in the 1553 a value of $x_1 = .0006''$ is very near this limit with present structures.

There is, however, at present another limitation which is essentially mechanical in nature but makes itself felt electrically in a way not indicated in the above simplified theory. This theory has assumed that the grid dimensions are infinitesimally thin compared to the electrode spacings. However, if this is not the case then the grid has less control action than an ideal fine grid, and the intrinsic gain band product must be reduced by still another factor F_4 which is a function of the grid transmission factor a $= \frac{p-d}{p}$ and the ratio $\frac{x_1}{p}$ where p is the pitch distance between grid wire centers and d is the diameter of grid wires. This function has the form shown in Fig. 9.*

Thus if the grid pitch and wire diameter are mechanically limited to some finite though small values, the optimum in input spacing x_1 will

* Data transmitted informally from C. T. Goddard and G. T. Ford.

still be for $x_1 \rightarrow 0$ but will not increase so strongly as $x^{-4/3}$ as before but much more slowly, about as $x^{-1/3}$. The grid dimensions should consequently be made as small as possible while still maintaining a transmission fraction at no less than 0.5 and at the same time not allowing mean deviations in pitch more than about 15%.

In the 1553 our best grid techniques today have led to a stretched grid (which does not move appreciably during temperature cycling) having a transmission factor of approximately 0.7, a pitch distance of .001″ and a mean deviation in pitch of less than 15%. For such a grid further decreases in input spacing without refining the grid will not pay off very rapidly, since we are on the maximum slope portion of the function F_4.

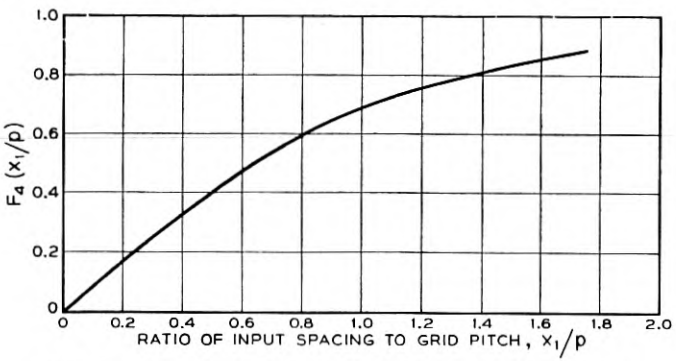


Fig. 9.—Dependence of gain-band product on grid pitch.

Limitations on Anode-Grid Spacing, x_2

In considering the choice of output spacing we must attain a balance among the following considerations:

a. The optimum transit angle $\theta_2 = 2.9$ radians requires a spacing which varies with plate voltage and with frequency. For 250 volts and 4000 Mc/s, this optimum is .022″.

b. The anode heat dissipation must be closely watched because the glass seal in this type of tube is very close to the anode. For the 1553, a maximum of 50 watts per square centimeter of anode active surface is safe. With a maximum cathode current density of 180 ma/cm², set by life considerations, heat dissipation limits the plate voltage to 275 volts unless the current is lowered.

c. If the anode is moved too far out, keeping its voltage constant, then in order to draw the desired current the grid must go positive, perhaps drawing excessive grid current. The grid shielding factor μ cannot be reduced without harming the transadmittance and feed-

back values; accordingly the cathode current would have to be reduced below the maximum permissible from life considerations.

d. The circuit degradation factor $(1 + C_{p2}/C_{22})^{-1}$ becomes more unfavorable as the active capacitance C_{22} is reduced by widening the output spacing. For discussion and calculation of this factor, see Appendix 2.

e. A wider output spacing, by virtue of the reduced capacitance, permits a higher maximum frequency limit on the tube.

The actual choice of output spacing in the 1553 is .012". This compromise between the foregoing factors appears to be suitable at 4000 Mc/s. The output transit angle of 1.6 radians gives 78% of the theoretical optimum intrinsic gain-band product. The anode dissipation is near the maximum safe value for the maximum allowable cathode current. The grid runs very close to cathode potential so that grid current is small. The circuit degradation factor has a value of about 0.8, while the upper frequency limit of the tube is satisfactory (about 5000 Mc/s).

The optimum design just described is an attempt to get the best possible gain-band product in the resulting tube, and is based on a particular electronic theory (that of Llewellyn and Peterson). Two points remain to be discussed. (1) What would be the result of optimizing for other merit figures such as power-band product or noise figure, and (2) how valid is the theory?

Power-Band Product

The radio relay amplifier requires not only gain, but perhaps even more, power output. In such a case, the design specification of greatest importance is the bandwidth over which a certain power output can be obtained with a specified maximum distortion, and is expressed by an analogous figure of merit, the power-band product.

Of the many methods of specifying distortion, one which is particularly useful in this connection is the "compression", that is, the amount by which the gain is reduced from the small-signal value. In an amplitude-modulated system, the compression would be a direct measure of nonlinear amplitude distortion in the amplifiers. In the actual relay, using FM, compression is an indication that the amplifier is approaching its maximum limit of power output.

The maximum power output depends not only on how much current the tube can carry, but also on the magnitude of the load impedance into which this current works, which in turn depends upon the bandwidth of the load. To compare tubes without need of specifying any bandwidth, one notes that the product of power output and band-

width is a constant, a figure of merit. The derivation is outlined in Appendix 1.

$$P_0 \cdot B = \frac{I_{20}^2 F^2(C) F_2^2(\theta_2)}{4\pi C_{\text{out}}} \tag{6}$$

The numerator here is just the square of the maximum ac current; that is, the dc current I_{20}, multiplied by a factor $F(C)$ depending on the allowable compression C, and by the gap coupling coefficient $F_2(\theta_2)$ of the electron stream to the output gap. The latter is of course a function of the output transit angle θ_2. It is assumed that the load is a matched simple resonant circuit and the band is taken 3 db down.

The power optimum must clearly be somewhat different from the gain optimum previously discussed. For example, the transadmittance does not appear here, nor does any property of the input circuit; while the magnitude of the direct electron current, which did not appear in the gain-band product, is now important. The capacitance of the output circuit appears in both figures of merit.

In terms of internal parameters of the tube, application of Llewellyn and Peterson's theory along the lines previously discussed leads to the following expression for power-band product:

$$M_i(P) = K[Aj^2 F^2(C)] [\theta_2 F_2^2(\theta_2) \sqrt{V_p}] \tag{7}$$

where A is the electrode area, $F^2(C)$ is a function of the allowable distortion limits, K is a constant which may depend upon frequency, and the other symbols are as before.

Considering first the dependence on output transit angle and plate voltage, one sees that this figure of merit has exactly the same form as the gain-band product. It is, however, not quite safe to assume therefore that exactly the same output configuration is still optimum, because the factors entering into the choice of output spacing have not exactly the same relative importance any longer; for example, a positive grid may be less objectionable, or a higher plate voltage may be permissible. Still, as a first approximation one may assume the output configuration to be already somewhere near optimum.

Other factors of the power-band figure of merit show considerable difference from the gain-band product. For instance, the electrode area enters the picture explicitly, suggesting that a larger area tube would give more power. The current density enters squared instead of only to the $\frac{2}{3}$ power; the explicit dependence on input spacing is missing. The compression function $F(C)$ depends mostly on the input conditions in a complicated way difficult to calculate. It can be approximated graphically from static characteristics.

A power tube similar to the 1553 might therefore be larger in electrode area, might have a coarser grid and wider input spacing, and perhaps would differ somewhat in output configuration, particularly if the plate voltage were raised. Any cathode development permitting a higher current density would improve the power output more than the gain, and might well lead to a drastic anode redesign to permit larger plate dissipation.

Similarly, a design to optimize noise figure would lead to still a third version of the tube, in which one might consider such things as critical relationships between input and output spacings.

For the 1553 at 4000 megacycles the following quantitative data may be quoted in order to check the gain-band product estimates.[8]

$$| Y_{21} | = 39.10^{-3} \text{ mhos}$$

$$G_{11} = 73 \cdot 10^{-3} \text{ mhos}$$

Note that the transadmittance is less than the dc value of $45 \cdot 10^{-3}$ mhos by only about 15%, while the input conductance, instead of being equal to the transadmittance as at low frequencies, is almost twice as large, on account of loading of the input gap by electrons returning to the cathode. Using the active capacitance C_{22} of .477 $\mu\mu f$, the intrinsic gain band product is:

$$\Gamma \cdot B = Y_{21}{}^2/4\pi G_{11} C_{22} = 3480 \text{ megacycles.}$$

With the somewhat optimistic capacitance degradation factor of .81 computed in Appendix 2, the gain band product would be reduced to 2820 megacycles.

The experimental average value is about 1100 megacycles.[9] The difference is probably due in part to resistive loss in the passive input circuit, which may be calculated as follows: Neglecting feedback, the input circuit may be represented as containing a resistance R_s in series with the short-circuit input admittance $g_{11} + jb_{11}$. Robertson gives the following values for these elements:

$$g_{11} = 73 \cdot 10^{-3} \text{ mhos}$$

$$b_{11} = 26 \cdot 10^{-3} \text{ mhos}$$

$$R_s = 7.6 \text{ ohms}$$

Accordingly, the input degradation factor $R_{11}/(R_{11} + R_s)$ should be $11.2/(11.2 + 7.6) = .60$, giving a computed overall gain-band product of 1690 megacycles. The best tubes sometimes exceed this figure. Tubes

[8] S. D. Robertson's measurements at 4000 megacycles, *B. S. T. J.*, 28, 619–655 (October 1949).

[9] A. E. Bowen and W. W. Mumford "Microwave Triode as Modulator and Amplifier," this issue of *B. S. T. J.*

with lower values may have excessive input circuit loss or may have narrower bandwidth on the input side than has been assumed. Further measurements, by elucidating this point, might lead to a better design of tube and circuit.

An entirely similar calculation can be made for the power-band product. The additional assumptions required are that the compression function $F^2(C)$ has the conservative value of $\frac{1}{2}$, and the output coupling coefficient $F_2(\theta_2)$ is taken as 0.9. The power-band product at 4000 megacycles is then computed to be 50 watt megacycles, which is quite close to the figures found by Bowen and Mumford.

Refinements of the Electronic Theory

In the electronic computations above, the single-valued theory was used because it is the simplest theory which describes the high frequency case at all accurately. The most important discrepancy between the rigorous theory and the actual situation is the first theoretical assumption listed above, that the electrons are emitted from the cathode with zero velocity. For actual cathodes the velocity of emission is not zero nor uniform but has a Maxwellian distribution such that the average energy away from the cathode is $\frac{1}{2}\ k\ T_k$, or about equivalent to the velocity imparted by a potential drop of 0.04 volt for an oxide cathode at 1000°K. There result several effects whose general nature is known but which have not yet been formulated into a rigorous quantitative theory valid at high frequencies.

(1) A potential minimum is formed at a distance on the order of .001″ in front of the cathode instead of at the cathode as in the simple theory. This distance is not negligible for close-spaced tubes; so that, for very close spacings, even perfect "physicists' grids" approach a finite trans-conductance limit. [van der Ziel, Philips Research Reports 1, 97–118 (1946); Fig. 2.]

(2) Because the potential minimum implies a retarding field near the cathode many electrons emerging from the cathode are forced to return to it. These returning electrons absorb energy from the signal and also induce excess noise in it, both effects becoming important at high frequencies.

The effects of initial velocities on the figures of merit can be measured experimentally. For example, the circuit and electronic impedances of diodes and triodes at 4000 Mc have been measured by Robertson.[8] Such measurements can determine the electronic loading and noise separately from the circuit degradation effects and are therefore a highly effective

[8] loc. cit.

method of circuit design as well. Robertson found that the input circuit structure of the 1553 produces a measurable impairment in its gain-band product, which redesign of both tube and circuit may be able to improve. Comparison of his results with the theory has given a better understanding of the limits of high-frequency performance, and has lent some support to the following set of rules of thumb which have been in use for some time:

1. The input loading arising from the returning electrons is considerable, the input conductance of these tubes at 4000 Mc being about double the theoretical value of Llewellyn and Peterson.[7]
2. The input noise of these close-spaced tubes checks well with what one would expect of a low-frequency diode with Maxwellian velocities, whose solution is known. In high-frequency noise calculations, therefore, one can use with some confidence Rack's suggestion that cathode noise can be regarded as an effective velocity fluctuation at the virtual cathode.[10]
3. Single velocity theory seems to hold well when velocities are much larger than Maxwellian, drift times are not more than a few cycles, electron beams are short compared to their diameter, and no exact cancellations of large effects are predicted. In particular it holds well for the 1553 output space and for calculations of the high-frequency trans-admittance.

Extensive calculations of signal and noise behavior in planar multigrid tubes have been made by L. C. Peterson, using the single-velocity theory except for an empirical value of input loading, and using Rack's suggestion for cathode noise.[11] The results so far checked have agreed well with experiment.

In short, the optimum design for the tube is still given fairly closely by the figures of merit based on the approximate theory, but the performance will fall somewhat short of the predictions of the simple theory; performance can be estimated with the aid of the experimental measurements and rules of thumb just described.

Summary

From the foregoing calculations we draw a number of conclusions:

1. The figures of merit can be validly analyzed into their dependence on more elementary properties like transadmittance, circuit capacitance, input loss resistance, and so on.

[7] loc. cit.

[10] A. J. Rack "Effects of Space Charge and Transit Time on the Shot Noise in Diodes," *B. S. T. J.*, 17, 592–619 (October 1938).

[11] L. C. Peterson "Space Charge and Noise in Microwave Tetrodes," *Proc. I. R. E.*, 35, 1262–1274 (November 1947).

2. Even rough calculations, such as the coaxial line approximations used in Appendix 2 are close enough to the facts to indicate whether the design is close to an optimum with respect to such parameters as output spacing, anode diameter, grid diameter, and the like. More accurate calculations and experiments can give more precise answers to these questions.
3. Some considerations such as cathode activity, tube life, heater power and so on have not yet been included in the analysis. However, systematic optimization for such parameters as are treated quantitatively is greatly facilitated. In general, each different figure of merit leads to a somewhat different optimum and hence a different version of the tube.

The design of tubes by the method of figure of merit has been outlined. The method is very general, but in essence has just three steps:

1. Formulate the system performance of the projected device with the aid of a figure of merit.
2. Find how the figure of merit depends upon the parameters of the tube, such as spacings, current, etc.
3. Adjust the tube parameters, subject to physical limitations, to optimize the figure of merit.

Acknowledgments

The development of this microwave triode has required not only the expert and highly cooperative services of a large team of electrical, mechanical, and chemical engineers but also the indispensable assistance of skilled technicians, all of whom worked smoothly together to develop these new materials and techniques to a point where they are specifiable and amenable to quantity production. It is not practical to mention all those who have made significant contributions to this development. The contributions of A. J. Chick, R. L. Vance, H. E. Kern and L. J. Speck, however, are of such outstanding nature that mention of them cannot be omitted.

APPENDIX 1

Derivation of the Figures of Merit

Gain-Band Figure of Merit

Let the problem be stated as the design of an amplifier tube to operate with as large gain over as wide a frequency band as practicable. As a standard environment, we use a single-stage amplifier working between equal resistive impedances. For three reasons this standard is suitable: it is simple; it corresponds closely to practicality in many cases especially

in the microwave field; and in most cases, it turns out that performance is limited by the same transadmittance to capacitance ratios as apply when the source and load impedances are not purely resistive. The terminology of high frequencies will be used but the analysis applies at all frequencies under the conditions stated.

Consider the over-all single-stage amplifier of Fig. A1-1 consisting of input resonator, tube and output resonator, to be a single transducer

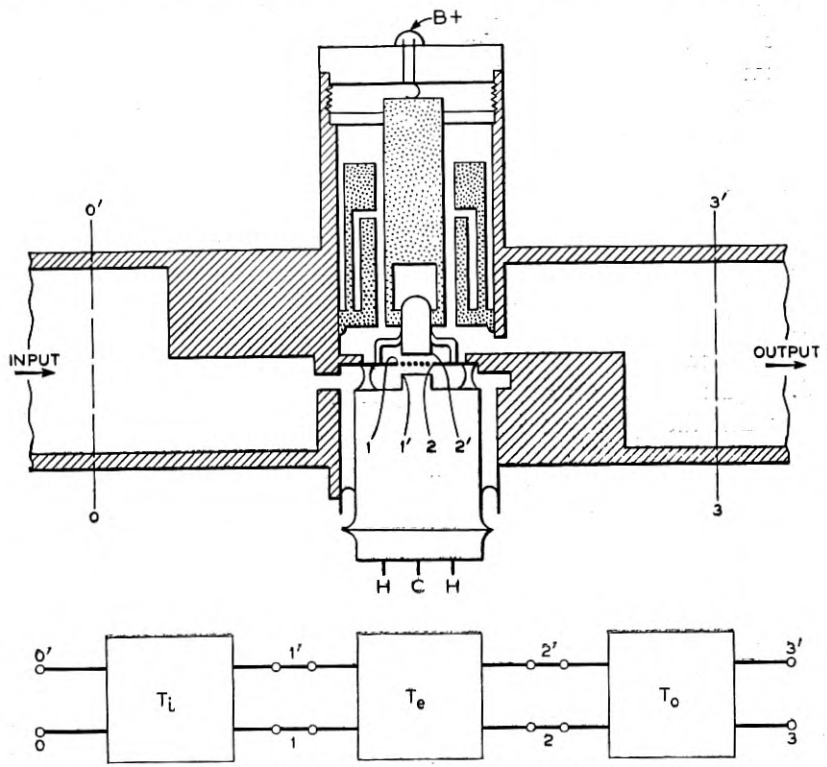


Fig. A1-1.—Microwave triode amplifier.

whose gain and bandwidth we wish to relate to the geometry and other pertinent characteristics of the circuits, bulb and electrode characteristics.

It is instructive to consider the whole transducer to be made up of three transducers in tandem as follows:

1. The input passive transducer, extending from the externally available input terminals (perhaps located somewhere in the driving wave guide or coaxial line) up to the internal input electrodes right at the boundary of the electron stream. Call this transducer T_i; in the

case of the grid-return triode of Fig. 1 it begins somewhere in the input wave guide at 0-0′ where only the dominant wave exists, includes the input external cavity and that portion of the tube interior right up to but not including the cathode-grid gap adjacent to the electron stream at 1-1′.

2. The output passive transducer, extending from the externally available output terminals located in the output wave guide through the output part of the bulb right up to the internal output electrodes at the boundary of the electron stream. Call this T_0; in the triode it

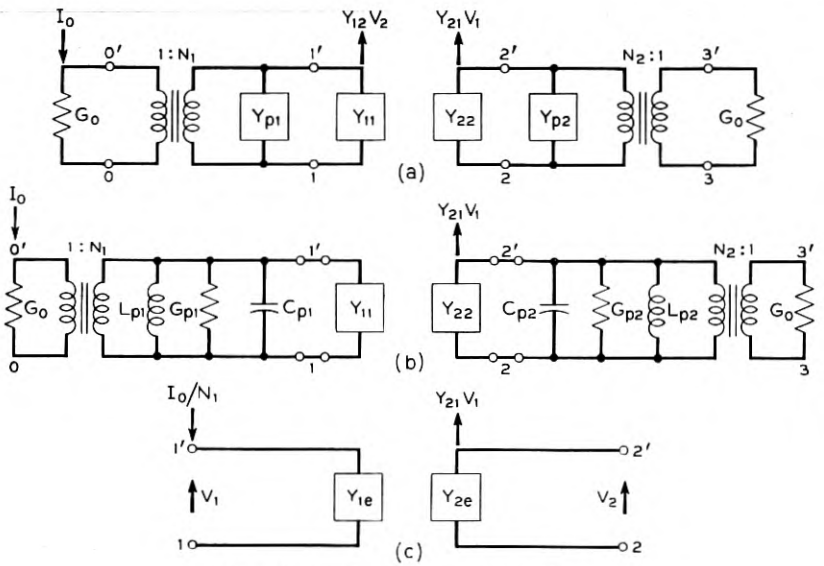


Fig. A1-2.—Amplifier representations.

extends from somewhere in the output wave guide at 3-3′ where only the dominant wave exists, includes the external coupling window, resonator cavity and output portion of the bulb, right up to the grid-anode gap adjacent to the electron stream at 2-2′.

3. The active electron transducer enclosing everything between the internal terminals of the above two passive coupling transducers—call this T_e—in the triode it extends from the cathode-grid gap adjacent to the electron stream at 1-1′ to the grid-anode gap adjacent to the electron stream at 2-2′. Geometrically it includes the stream and active portions of the electrodes. The term "active" will be applied to the electron stream and to those portions of the electrodes which interact directly with the stream.

We may represent these three transducers as in Fig. A1-2a, where the input and output transducers have each been replaced by an ideal transformer of turns ratio N and a shunt admittance Y_p. This representation is general enough for present purposes, provided that Y_p and N are allowed to be complex functions of frequency and provided that terminals 0-0′ and 3-3′ are chosen so that a potential minimum occurs at those points when points 1-1′ and 2-2′ are shorted.

The short-circuit admittances for the whole transducer as seen at terminals 0-0′ and 3-3′ are then

$$\begin{aligned} Y^*_{11} &= N_1^2\,(Y_{11} + Y_{p1}) \\ Y^*_{22} &= N_2^2\,(Y_{22} + Y_{p2}) \\ Y^*_{21} &= N_1 N_2 Y_{21} \\ Y^*_{12} &= N_1 N_2 Y_{12} \end{aligned} \tag{A1-1}$$

where the Y_{ij} are the short-circuit admittances of the electron transducer alone as seen at terminals 1-1′ and 2-2′.

If the feedback admittance Y_{12} is assumed negligible the insertion voltage gain may be written as

$$\Gamma(\omega) = \frac{2N_1 N_2\, Y_{21}}{G_0(1 + \sigma_1)(1 + \sigma_2)}$$

where the sigmas are admittance-matching factors:

$$\sigma_1 = \frac{N_1^2(Y_{11} + Y_{p1})}{G_0} = \frac{Y^*_{11}}{G_0}; \qquad \sigma_2 = \frac{N_2^2(Y_{22} + Y_{p2})}{G_0} = \frac{Y_{22}^*}{G_0} \tag{A1-2}$$

The gain is maximum when σ_1, σ_2 are minimum, i.e., when tube and circuits are resonant and losses are minimum.

We may rewrite this in terms of the total Y^*_{ij} as follows:

$$\Gamma(\omega) = \frac{2Y^*_{21}}{G_0(1 + \sigma_1)(1 + \sigma_2)} \tag{A1-3}$$

Many practical cases are well approximated by the more special representation of Fig. A1–2b, where the turns ratios of the ideal transformers are real and independent of frequency, and the shunt admittance consists of ordinary lumped constant circuit elements. The feedback admittance Y_{12} is neglected.

This representation as simple, lumped-constant elements holds very well for any admittance, even a distributed, cavity-type microwave circuit, or an electronic admittance, provided that the combined circuit has no series and only one shunt resonance near the frequency band in

question. The "effective values" of the actual admittance are given by equations (3) of the text, as follows:

$$\begin{aligned} G_p &= G_x\,(\omega_0) \\ C_p &= \tfrac{1}{2}\,(B_x' + B_x/\omega_0) \\ \frac{1}{L_p} &= \tfrac{1}{2}\,(\omega_0^2\,B_x' - \omega_0\,B_x) \end{aligned} \tag{A1-4}$$

Let the complete admittances across nodal pairs 1-1′ and 2-2′ be called Y_{1e} and Y_{2e} as in Fig. A1-2c, which is an abbreviation of Fig. A1-2b from the point of view of the active transducer.

$$\begin{aligned} Y_{1e} &= G_1 + G_{p1} + G_{11} + j\omega C_{p1} + j\omega C_{11} + \frac{1}{j\omega L_{p1}} \\ &= G_{1e} + j\omega C_{1e} + \frac{1}{j\omega L_{1e}} \end{aligned} \tag{A1-5}$$

$$\begin{aligned} Y_{2e} &= G_2 + G_{p2} + G_{22} + j\omega C_{p2} + j\omega C_{22} + \frac{1}{j\omega L_{p2}} \\ &= G_{2e} + j\omega C_{2e} + \frac{1}{j\omega L_{2e}} \end{aligned}$$

where G_1 and G_2 are the line admittances as seen from the active transducer:

$$G_1 = G_0/N_1^2;\; G_2 = G_0/N_2^2.$$

The Q's of the circuit are defined as

$$\begin{aligned} Q_{1e} &= \omega_0\,C_{1e}/G_{1e} \\ Q_{2e} &= \omega_0\,C_{2e}/G_{2e} \end{aligned} \tag{A1-6}$$

The insertion voltage gain (2) may be written as follows to emphasize the manner in which it depends upon frequency:

$$\Gamma = \frac{2Y_{21}}{Y_{1e}Y_{2e}}\sqrt{\frac{G_{1e}G_{2e}}{(1+\mu_1)(1+\mu_2)}} \tag{A1-7}$$

Here $\mu = \sigma(\omega_0)$ is the matching factor at band center. Frequently the circuits are matched ($\mu_1 = \mu_2 = 1$) to avoid standing waves in system applications, and we shall discuss this case; but in any case μ_1 and μ_2 are constants with respect to frequency. For our standard circuits, G_{1e} and G_{2e} are independent of frequency; also ordinarily the transadmittance Y_{21} may be considered constant for bandwidths commonly encountered. There results then the fact that the voltage gain (and phase) depends on frequency in the same way as $(Y_{1e}\,Y_{2e})^{-1}$.

Since the gain varies with frequency, the amplifier will give approximately constant response only within a certain range of frequencies. The band of the amplifier is defined as that frequency interval within which the magnitude of the gain is constant within some specified tolerance; the bandwidth is the size of this interval. We wish to express the gain of the amplifier in terms of its bandwidth, in the following way:

The voltage gain of this amplifier has a maximum, called Γ_0, at band center frequency f_0. Take the band of the amplifier $B_N(A)$ as that interval within which the voltage gain is within a factor of 1/N times the maximum.

$$\left| \frac{\Gamma(\omega)}{\Gamma(\omega_0)} \right| \geq \frac{1}{N} \text{ defines } B_N(A) \tag{A1-9}$$

We can analogously define the band of a simple circuit $B_n(C)$ by the relation

$$\left| \frac{Y_{2e}(\omega_0)}{Y_{2e}(\omega)} \right| \geq \frac{1}{n} \text{ defines } B_n(C). \tag{A1-9}$$

It follows directly that

$$B_n(C) = \frac{G_{2e}}{2\pi C_{2e}} \sqrt{n^2 - 1}\,. \tag{A1-10}$$

Since the amplifier gain is inversely proportional to the product of the circuit admittances, it follows that $n_1 n_2 = N$.

The intrinsic bandwidth resulting from the tube admittance may not be suitable for the intended application. In that case the band may be widened by increasing G_{1e} or G_{2e} with a corresponding decrease in gain. We have then the problem of adjusting G_{1e} and G_{2e} for greatest band efficiency, i.e., maximum gain for a given bandwidth, with synchronous tuning. It turns out that if the bandwidth is less than that needed, then the circuit of higher Q should be lowered until either (a) the band becomes wide enough, or (b) the Q's become equal. In case (b), both Q's should then be lowered, maintaining equality, until the band is wide enough.

Two important limiting cases are to be considered: (a) $Q_{1e} = Q_{2e}$, i.e. the band is shaped equally by the input and output circuits; and (b) $Q_{1e} << Q_{2e}$, i.e. the band is shaped by only the output circuit. In the equal-Q case we have

$$\begin{aligned} \frac{G_{1e}}{C_{1e}} &= \frac{G_{2e}}{C_{2e}} \\ n^2 &= N \end{aligned} \tag{A1-11}$$

$$B_N(A) = \frac{1}{2\pi} \sqrt{\frac{G_{1e}}{C_{1e}} \frac{G_{2e}}{C_{2e}}} \sqrt{N - 1}\,.$$

If only the output circuit is involved, then $N = n_2$ and the band of the amplifier, being shaped differently, is given by a different relation:

$$B_N(A) = \frac{G_{2e}}{2\pi C_{2e}} \sqrt{N^2 - 1}. \tag{A1-12}$$

In other words, a band shaped by only one circuit has the shape of (12), while a band shaped by two circuits has the shape (11). The maximum voltage gain is

$$| \Gamma_0 | = | \Gamma(\omega_0) | = \frac{2 | Y_{21} |}{\sqrt{G_{1e}G_{2e}(1 + \mu_1)(1 + \mu_2)}} \tag{A1-13}$$

Substituting for the G's in terms of the bandwidth, we have for the equal-Q case (from 11)

$$| \Gamma_0 | = \frac{| Y_{21} |}{2\pi\sqrt{C_{1e}C_{2e}}} \frac{2\sqrt{N - 1}}{\sqrt{(1 + \mu_1)(1 + \mu_2)}} \frac{1}{B_N} \tag{A1-14}$$

and for the unequal-Q case (from 12)

$$| \Gamma_0 | = \frac{| Y_{21} |}{\sqrt{G_{1e}}\sqrt{4\pi C_{2e}}} \frac{\sqrt{8}\sqrt[4]{N^2 - 1}}{\sqrt{(1 + \mu_1)(1 + \mu_2)}} \frac{1}{\sqrt{B_N}} \tag{A1-15}$$

These equations give the relationship between the gain and bandwidth of a transmission system shaped by two or one independent circuits, respectively. The comparison between these two cases is not quite straightforward. First, the band shapes (11) and (12) are different, although this difference is small enough to be ignored for $N < 2$ (6 db down). Second, the gain varies differently as the band is widened; the equal-Q case loses gain at 6 db per octave in bandwidth, the unequal-Q case only 3 db per octave. The comparison therefore depends on the bandwidth chosen. However, these formulas are still quite useful, especially in comparing two amplifiers of the same type or in optimizing an amplifier of one of the types.

From the equal-Q formula one notices that the product of insertion voltage vain and bandwidth does not depend on the bandwidth, but is a figure of merit by which two amplifiers of the same type (i.e. equal Q) but different gains and bandwidths can be compared. Since

$$C_{1e} = C_{11} + C_{p1}\ ;\ C_{2e} = C_{22} + C_{p2}$$

$$| \Gamma_0 | B_N = \left(\frac{Y_{21}}{2\pi\sqrt{C_{11}C_{22}}}\right)\left(\frac{1}{\sqrt{1 + \frac{C_{p1}}{C_{11}}}\sqrt{1 + \frac{C_{p2}}{C_{22}}}}\right) \cdot \left(\frac{2\sqrt{N - 1}}{\sqrt{(1 + \mu_1)(1 + \mu_2)}}\right) \tag{A1-16}$$

This expression for the gain-band figure of merit of a two-circuit, line-to-line amplifier is particularly useful for grounded-cathode pentodes and klystrons. It is the product of three factors. The first may be called the electronic figure of merit because it depends only upon electron stream parameters (ratio of transadmittance to mean capacitance of the electronic transducer T_e). The second is the degradation factor giving the effect of adding passive circuit capacitance both inside and outside the bulb to the active capacitance already present in the electronic transducer. The third factor, called the matching factor, depends only on the matching conditions and on the arbitrary definition of bandwidth. If the band is taken 6 db down (3 db for eacy circuit) and the tube input and output are matched, the third factor is unity.

In amplifiers using triodes and tetrodes in grid-return circuits, the Q of the input circuit is usually very much smaller than that of the output. Here it is appropriate to use the single-circuit limiting concept, with $Q_{1e} << Q_{2e}$. Here a figure of merit independent of bandwidth is obtained from the product of power gain and bandwidth:

$$|\Gamma_0|^2 B_N = \left(\frac{|Y_{21}|^2}{4\pi G_{11} C_{22}}\right)\left(\frac{1}{\left(1+\frac{G_{p1}}{G_{11}}\right)\left(1+\frac{C_{p2}}{C_{22}}\right)}\right)\left(\frac{8\mu_1\sqrt{N^2-1}}{(1+\mu_1)^2(1+\mu_2)}\right) \qquad \text{(A1-17)}$$

This expression for the gain-band figure of merit of a one-circuit, line-to-line amplifier is also the product of three factors. The first is again the intrinsic electronic figure of merit of the active transducer alone; the second is the degradation produced by the addition of passive circuit capacitance to the output and circuit loss to the input; the third is a band-definition matching factor which is unity when the band is taken 3 db down and the tube is matched.

In the application of the figures of merit, the third factors are usually omitted, since they depend only on the matching conditions and on the particular definitions of bandwidth used.

Power-Band Figure of Merit

In the problem of power output amplifier stages, the design specification of greatest importance is the bandwidth over which a certain power output can be obtained with a specified maximum of distortion. Of the many methods of specifying distortion, one which is particularly useful for microwave systems is known as the "compression". If the power gain is plotted in decibels as a function of the power output, as shown in Fig.

A1-3, it will normally be constant for low power levels (for which the device is essentially linear) and equal to the low level power gain $|\Gamma|^2$. However, at some higher power level non-linearities appear in some or all of the various short-circuit admittances, usually causing the power gain to decrease below the small-signal value by an amount called the compression, C. If P_0/P_i be power gain for any power output and $|\Gamma|^2$

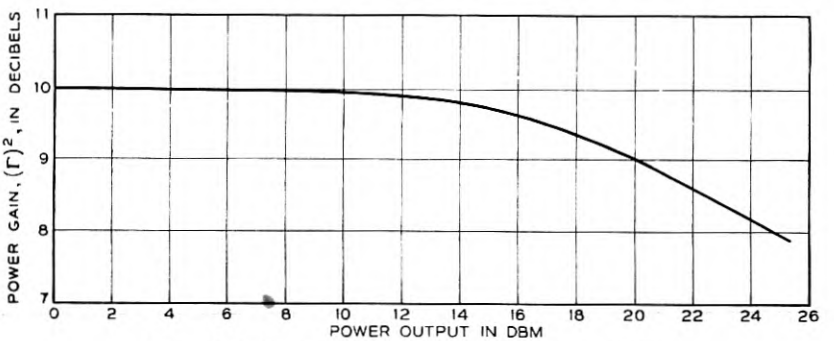


Fig. A1-3.—Typical gain variation with power output.

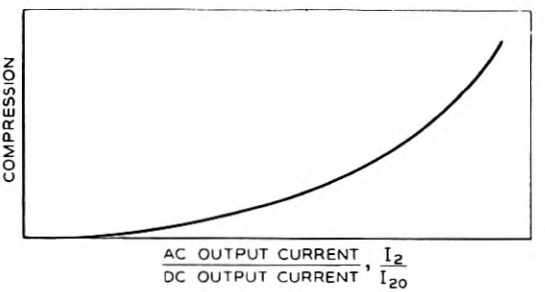


Fig. A1-4.—Compression vs. Alternating current in output / Direct current in output

the small-signal power gain, the compression C is defined in decibels as follows:

$$C = 10 \log_{10} |\Gamma|^2 - 10 \log_{10} P_0/P_i \qquad \text{(A1-18)}$$

$$= 10 \log_{10} \frac{|\Gamma|^2 P_i}{P_0}$$

Naturally, the compression depends upon how hard the tube is driven. It is therefore a function of the amount of drive, which may be conveniently expressed in terms of the ratio of the alternating output current to the operating direct current, as in Fig. A1-4.

The power output depends on operating parameters thus:

$$P_0 = I_2^2 \frac{G_2}{(G_{22} + G_2)^2} \tag{A1-19}$$

As the output power level is continually raised, more and more current is required to drive the load, until finally the non-linear distortion limit is reached. The maximum output current is therefore limited to a certain proportion of the direct current I_{20}, thus:

$$I_{2m} = I_{20} \cdot F(C)\ F_2(\theta_2) \tag{A1-20}$$

where $F(C)$ shows the dependence upon the compression C and will naturally be the larger, the more the allowable compression. $F_2(\theta_2)$ indicates a dependence upon output transit angle; it is the output gap coupling coefficient.

The power output depends also upon the output circuit conductance G_2 and can be greater if G_2 is smaller. However, a smaller G_2 implies a smaller bandwidth. It results that the power is inversely proportional to the bandwidth of the output circuit, or in other words, the product of power output by the bandwidth of the output circuit is a constant—a figure of merit of the tube. As in the case of the gain-band merit, this also can be broken up into factors:

$$P_0 \cdot B_N = \left(\frac{I_{20}^2 F^2(C) F_2^2(\theta_2)}{4\pi C_{22}}\right)\left(\frac{1}{1 + C_{p2}/C_{22}}\right)\left(\frac{2\sqrt{N^2 - 1}}{1 + \mu_2}\right) \tag{A1-21}$$

This expression for the power-band figure of merit is the product of three factors. The first is the intrinsic figure of merit of the active transducer alone; the second is the degradation caused by the addition of passive circuit capacitance to the output circuit; the third is a band definition—matching factor which is unity when the output is matched and the band of the output circuit is taken 3 db down.

The power-band computation does not depend upon the input circuit. Variations in the latter affect the gain of the amplifier, but not its overload point. Accordingly in the power band formula only properties of the tube and its output circuit appear. When feedback has to be considered, then the input circuit also affects the power, and the analysis becomes more complicated.

We have now three figures of merit: namely, two gain-band products applying to different kinds of amplifiers, and one power-band product. They relate the performance of an amplifier to certain internal parameters. For wide band service, the tube design should make the appropriate figure of merit as large as practicable.

It should be understood that many other factors may have a bearing on amplifier design, such as power consumption, noise performance or amount of feedback. Where such factors are important, they too must be considered, and frequently appropriate merit figures like plate efficiency or noise figure are useful.

APPENDIX 2

The Circuit Capacitance Degradation Factor

The capacitance degradation factor $C_{22}/(C_{22} + C_{p2})$ which applies to both gain-band and power-band products, can be calculated approximately

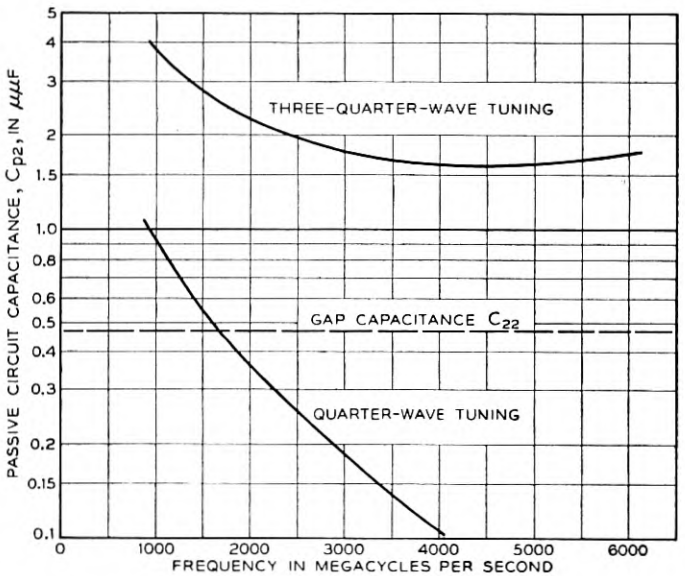


Fig. A2-1.—Passive circuit capacitance C_{p2}.

as shown below. As the frequency is varied, this factor changes by considerable amounts for the 1553 tube; accordingly, both figures of merit vary with frequency, and design control has been exercised to produce maximum merit around 4000 megacycles.

The capacitance degradation factor is just the proportion which the active tube capacitance bears to the total capacitance of tube and circuit, and would therefore have a maximum of unity if the circuit passive capacitance were made zero. For the 1553, we may begin by assuming that the plate circuit is to be tuned by a resonant coaxial line. As the frequency is lowered the effective capacitance will be increased, since the line must be lengthened; its variation is shown in Fig. A2-1.

The calculation is based on the following assumptions (Fig. A2-2):

1. The output cavity has inner diameter .180", outer .850", consequently a characteristic admittance G_0 :

$$G_0 = 7250/\log \frac{d_2}{d_1} = 10{,}710 \text{ micromhos} \qquad \text{(A2-1)}$$

2. The gap capacitance is that of a parallel plate condenser of .180" diameter and .012" spacing, namely

$$C_{22} = \epsilon_0 A/d = 0.477 \ \mu\mu f \qquad \text{(A2-2)}$$

3. The effect of the glass vacuum envelope is neglected for simplicity.

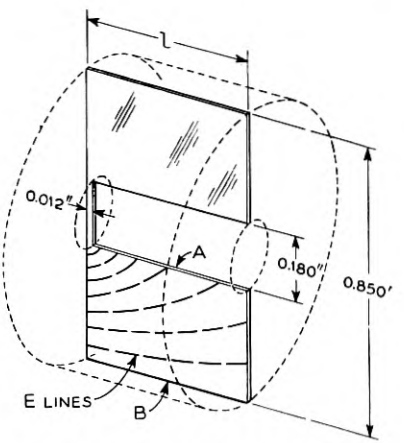


Fig. A2-2.—Output cavity dimensions. A, B are concentric cylindrical portions. Actual lines of electric force are partly dotted into sketch.

Consequently the length l of the line is given by the well-known tuning relation

$$\omega C_{22} = G_0 \cot \theta = G_0 \cot \frac{\omega l}{c} \qquad \text{(A2-3)}$$

The distributed capacitance of the line is determined from the formulas (3) of the text, which in this case reduces to the following:

$$\omega C_{p2} = \frac{G_0 \theta}{2}\left(1 + \frac{\omega^2 C_{22}^2}{G_0^2}\right) - \frac{\omega C_{22}}{2} \qquad \text{(A2-4)}$$

The cavity distributed capacitance is thus comparatively easy to calculate at high frequencies because of the simplicity of the geometry. At low frequencies the computation of the distributed capacity of a coil is no

different in principle, but would be harder to carry out in practice because of the helical geometry. The value can of course in any case be found by measurement of the tuning admittance as a function of frequency. From these equations the circuit degradation factor can be calculated, and is shown in Fig. A2-3 as a function of frequency.

The accuracy of the coaxial line assumptions decreases as the cavity becomes shorter. For 4000 and 6000 megacycles, since the length of the cavity is less than its diameter, it would be more nearly correct to regard it as a radial transmission line loaded by the inductive "nose" in the

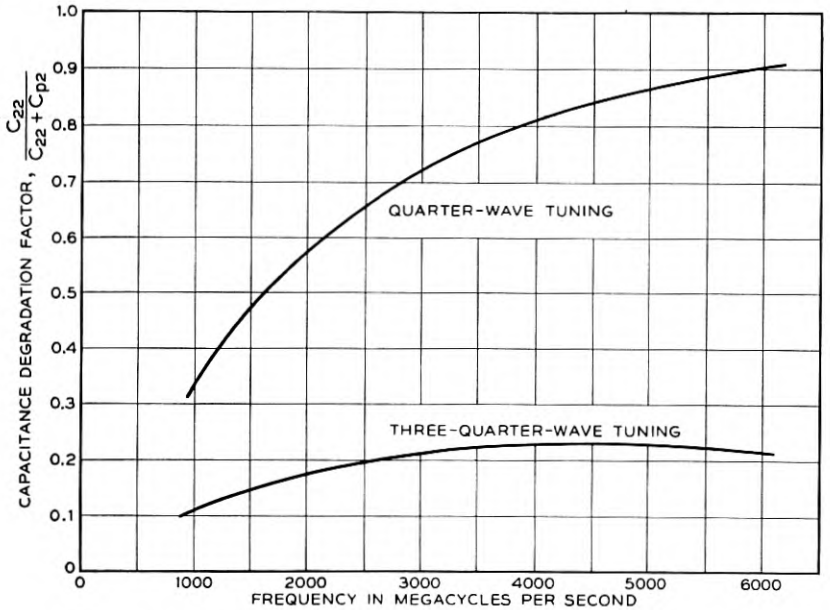


Fig. A2-3.—Capacitance degradation factor, $\frac{C_{22} + C_{p2}}{C_{22}}$.

center. The admittance of such a cavity can be calculated[12] or measured; but the additional precision hardly warrants the effort in the present case.

The capacitance degradation factor at 4000 megacycles is indicated from Fig. A2-3 as .81, or only 0.9 db less than the intrinsic limit of unity if the passive capacitance were entirely negligible compared to the active 0.5 $\mu\mu f$. This indication is somewhat optimistic, as appears from Fig. A2-2. The coaxial line formulas assume that the capacitance corresponds to a radial electric field between concentric cylinders A and B. This capacitance is found to be quite small (.11 $\mu\mu f$ at 4000 Mc.). The actual lines of

[12] S. Ramo and J. R. Whinnery, "Fields and Waves in Modern Radio," N. Y., Wiley, 1944.

force, dotted in the figure, clearly correspond to a somewhat larger capacitance, especially when the length of the cavity is smaller than its diameter; but this larger capacitance is probably still less than the active capacitance C_{22}.

In so far as the gain-band product depends on the circuit capacitance degradation factor (Fig. A2-3), the curve is probably fairly accurate up to 2000 megacycles and somewhat optimistic for higher frequencies where the coaxial line predictions are evidently too small.

Above 5000 megacycles the quarter-wave tuning cannot be used for the 1553 tube since the glass would interfere with the tuning plunger. A glance at Fig. A2-3 shows that moving the plunger back a half-wave to the next node involves a drastic loss in gain-band product—a factor of four at 6000 megacycles—because of the great increase in circuit passive capacitance. Redesign of the tube for good figure of merit at 6000 megacycles would therefore require the use of first-node tuning. A reduction in outer diameter would be necessary, and the use of an internal pre-tuned cavity might also be indicated.

A New Microwave Triode: Its Performance as a Modulator and as an Amplifier

By A. E. BOWEN* and W. W. MUMFORD

(*Manuscript Received Mar. 20, 1950*)

This paper describes a microwave circuit designed for use with the 1553-416A close-spaced triode at 4000 m.c. It presents data on tubes used as amplifiers and modulators and concludes with the results obtained in a multistage amplifier having 90 db gain.

Introduction

MICROWAVE repeaters are of two general types: those that provide amplification at the base-band or video frequency and those that amplify at some radio frequency. Of the latter there are two types: those that involve no change in frequency and those that do involve a change in frequency, that is, the radiated frequency is different from the received frequency. The Boston-New York link[1] is of this last type as is also the New York-Chicago link. This paper deals chiefly with a discussion of the application of the close-spaced triode[2] in a repeater of the type to be used between New York and Chicago.

A block diagram of this type of repeater appears in Fig. 1. The received signal comes in at a frequency of, say, 3970 mc. It is converted to some intermediate frequency, say 65 mc, in the first converter which is associated with a beating oscillator operating at a frequency of 3905 mc. After amplification at 65 mc it is converted in the modulator back to another microwave frequency 40 mc lower than the received signal and then it is amplified by the r.f. amplifier at 3930 mc and transmitted over the antenna pointed toward the next repeater station. Our attention will be focussed upon the performance of the close-spaced triode in the transmitting modulator and in the r.f. power amplifier in this type of repeater.

The close-spaced triode was assigned the code number 1553 during its experimental stage of development and, with subsequent mechanical improvements, it became the 416A. Some of the data reported herein were taken on one type, and some on the other; references to both the 1553 and 416A tubes will be noted throughout the text. The difference in electrical performance was not significant.

An early experimental circuit for the 1553 type tube will be described

* Deceased.

in detail and the performance as amplifier and modulator will be presented. Measurements of noise figure will be included with a discussion of the performance of multistage amplifiers.

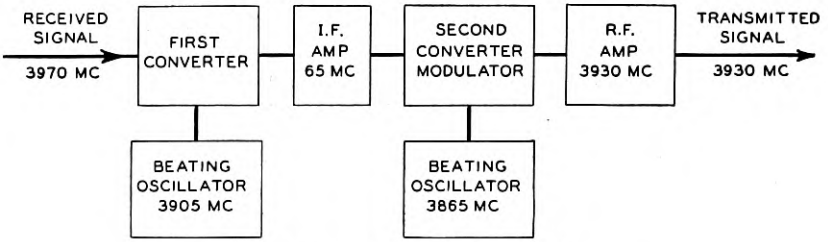


Fig. 1.—Typical microwave repeater.

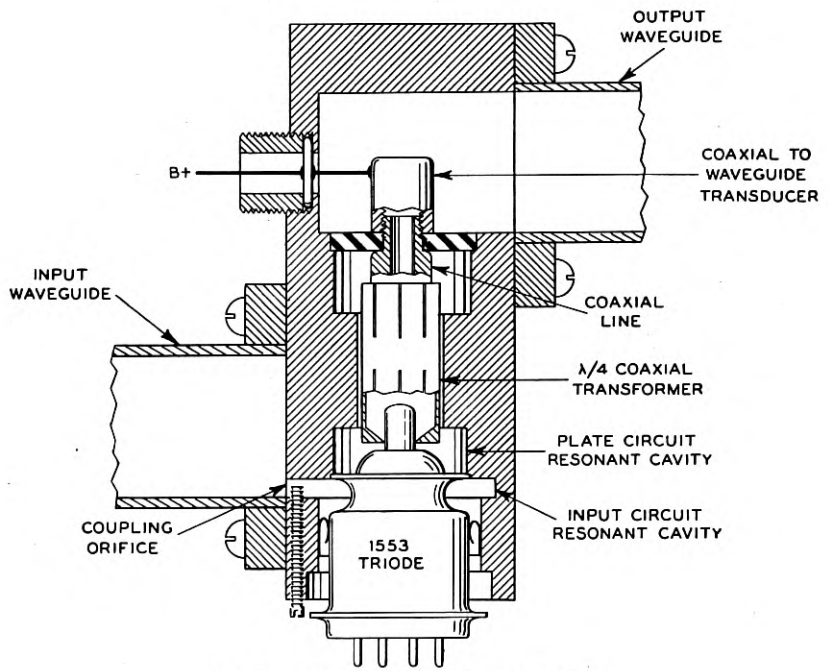


Fig. 2.—Microwave circuit for 1553 triode.

The Microwave Circuit

The experimental circuit which has to date met with greatest favor consists of cavities coupled to input and output waveguides, as shown in Fig. 2. The grid, of course, is grounded directly to the cavity walls and separates the input cavity from the output cavity. An iris with its orifice

couples the input waveguide to the input cavity and is tuned by a small trimming screw across its opening. The metal shell of the base of the tube makes contact to the input cavity through spring-contact fingers around its circumference and forms a part of the input cavity. The cathode and its by-pass condenser, located within the envelope, complete the input circuit cavity. The heater and cathode leads, brought out through eyelets in the base of the tube, are isolated from the microwave

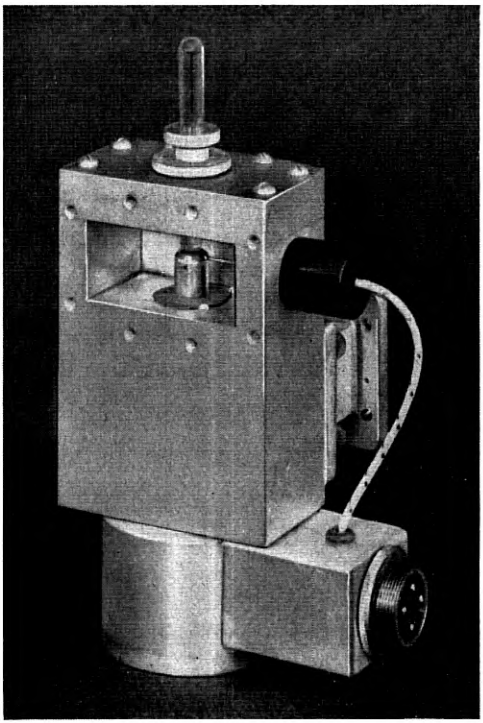


Fig. 3.—Model eleven microwave circuit for the close-spaced triode looking into the output waveguide.

energy in the input cavity by means of the internal by-pass condenser. When the tube is used as a modulator, this by-pass condenser acts as a portion of the network through which the intermediate frequency signal power is fed onto the cathode.

The output circuit cavity is coupled to the output waveguide through a coaxial transformer, a coaxial line and a wide-band coaxial-to-waveguide transducer. The output cavity is bounded by the grid, the coaxial line outer conductor, the radial face of the quarter-wave coaxial transformer

and the sealed-in plate lead of the tube. The plate impedance of the tube is transformed by the resonant cavity to a very low resistance (a fraction of an ohm) on the plate lead just outside the glass seal. The quarter-wave coaxial transformer serves to match this low impedance to the surge impedance of the coaxial line (45 ohms). Coarse tuning is accomplished by moving the slug of the outer conductor; fine tuning by moving the inner conductor. The coaxial line is supported at its end by a dielectric washer. Plate voltage is applied to the tube through a high impedance quarter-wave wire brought out to the low impedance probe through the side wall of the waveguide. Both the modulator and the amplifier used this type of circuit, which we call model eleven.

Fine tuning of the plate cavity is obtained by sliding the inner conductor of the coaxial transformer up and down on the plate lead. This movement is derived through a low-loss plastic screwdriver inserted through the hollow probe transducer; the driving mechanism is housed inside the inner conductor of the transformer, thus isolating the mechanical design problem from the electrical design problem effectively. The hollow stud at the top of the structure serves two purposes: screwing it into the waveguide introduces a variable capacitive discontinuity which serves to improve the match between the cavity and the waveguide. The length of the hollow plug provides a length of waveguide beyond cutoff which keeps the r.f. energy from leaking out through the plastic tuning screwdriver.

The heater and cathode leads from the tube are housed in a cylindrical metal can and are brought out through by-pass condensers to a standard connector. The photograph, Fig. 3, illustrates these features.

The input face of the circuit is illustrated in Fig. 4. The long narrow slot near the base of the rectangular block is the iris opening which couples the input waveguide to the cathode-grid cavity. The single tuning screw provided at the input iris is not adequate to match all of the tubes over the whole frequency band of 500 megacycles; an auxiliary tuner shown at the right of the circuit provides the necessary flexibility. This tuner, described by Mr. C. F. Edwards of the Bell Telephone Laboratories[3], is, in effect, two variable shunt tuned circuits about an eighth of a wavelength apart in the waveguide. Each variable tuned circuit is made up of a fixed inductive post (located off center in the waveguide) and a variable capacitive screw. It is capable of tuning out a mismatch corresponding to four db standing wave ratio of any phase.

As shown in Fig. 5, the tube slides into the bottom of the circuit and the grid flange is soldered to the wall of the cavity with low melting point

solder.† The shell of the tube is grasped by the springy contacts around the bottom of the input cavity. Above the tube the plate lead projects into the cylindrical space which can be adjusted to the desired size by the quarter wave slug seen to the right of the circuit. This makes contact to the walls of the outer cylinder by spring fingers on each end. Contact to the plate lead is then made through the movable slotted inner conductor, seen on the extreme right of Fig. 5.

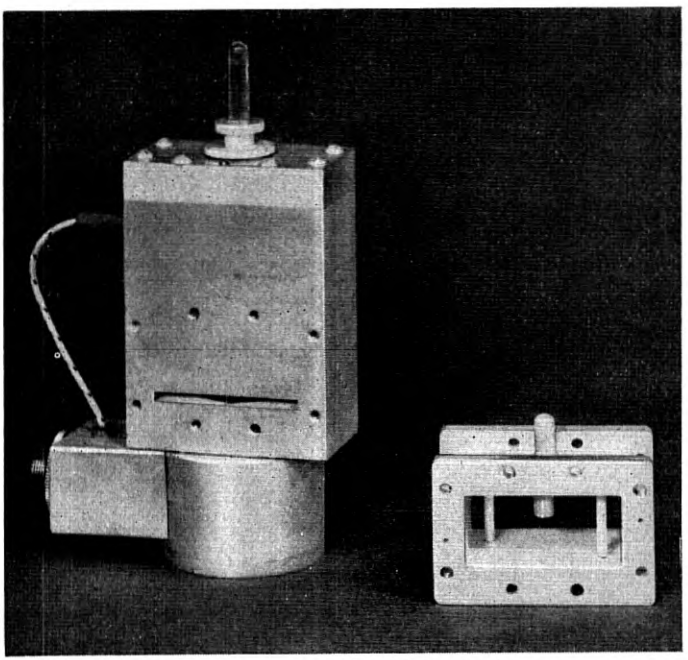


Fig. 4.—The input face of the circuit.

Figure 6 gives an exploded view of the details of the circuit, showing the simplicity of the construction which permits easy assembly. The guide pin which serves to keep the inner conductor of the transformer from rotating as it slides up and down on the plate lead during the tuning process can be seen on the third detail to the right of the main block. Also there is provision for external resistive loading to be introduced into the plate cavity through the small square holes in each side of the block. A screw mechanism adjusts the penetration of the loading resistive strip into the

† The early experimental tubes were soldered into the circuits. Chiefly through the efforts of Mr. C. Maggs and Mr. L. F. Moose, of B. T. L., who undertook the development of the tube for production by the Western Electric Company, the present 416A tubes come with a threaded grid flange to facilitate replacement.

plate cavity to provide for a limited adjustment of the bandwidth of the circuit. These are not always used, however, and most of the data to be presented here are for the condition of no external resistive loading.

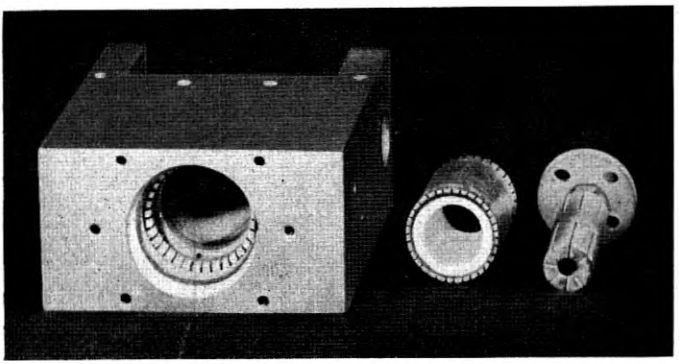


Fig. 5.—Bottom view of circuit.

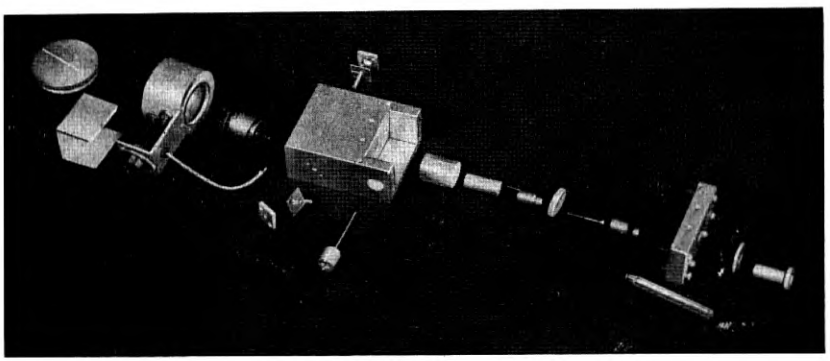


Fig. 6.—Exploded view of details.

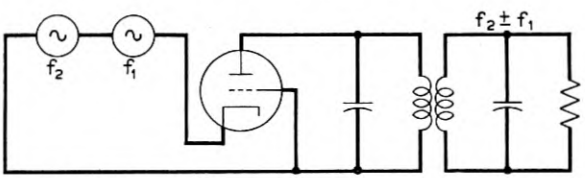


Fig. 7.—Elementary grounded grid converter schematic.

Modulator

The grounded grid transmitting converter shown schematically in Fig. 7 includes the two generators, a microwave beating oscillator, f_2, and an intermediate frequency signal, f_1, which impress voltages on the cathode,

the grid itself being grounded. The output circuit in the plate is tuned to the sum or difference frequency, $f_2 \pm f_1$.

By-pass condensers, traps and filters for other frequencies present in the modulator must be considered. Besides the beating oscillator and the signal, their sum and difference frequencies appear in both the input circuit and the output circuit and of course bias voltage on the cathode and plate voltage on the plate must be applied. Some of the traps and by-pass condensers which influence the converter performance are indicated in Fig. 8. It is obvious that microwave energy should be kept from flowing into the i.f. signal circuit and vice versa if the highest conversion gains are to be obtained. Both of these conditions are easily achieved. It is not so readily apparent that the components of the wanted and the unwanted sidebands present in the input circuit must be handled

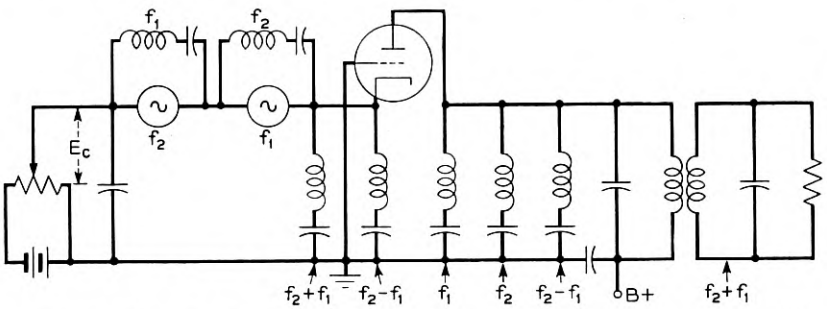


Fig. 8.—Schematic diagram of converter with traps and filters for fundamental frequencies of signal, f_1, beating oscillator, f_2, and sidebands, $f_2 \pm f_1$.

properly. Of these two, the more important is the wanted sideband and the next figure illustrates just how necessary it is to treat it properly.

The simplest way to keep the wanted sideband component of the input circuit from being absorbed by the beating oscillator branch is to reflect the energy back into the converter by means of a reflection filter. This reflected energy arrives back at the tube and may conspire to reduce the conversion gain of the modulator if the phase is wrong. The phase depends upon the spacing along the waveguide between the tube and the filter and Fig. 9 illustrates how badly the gain is affected when the wrong spacing is used. Data for two different tubes are given which indicate that the correct spacing for one tube may be incorrect for another. It should be pointed out, however, that these two tubes were early experimental models and that production tubes behave more consistently.

The i.f. impedance of the modulator is also affected by the filter spacing for the wanted sideband on the input. This effect can be utilized to

vary the i.f. impedance by small amounts to achieve a better i.f. match, since the proper spacing for best gain is not a critically exact dimension. That is to say, there is a fairly large range of spacings which give good performance as far as conversion gain is concerned so that, as long as the critical distance which gives poor gain is avoided, the i.f. impedance can be adjusted by varying the spacing of the input filter.*

It is important that the i.f. impedance of the modulator be adjusted to match the impedance of the i.f. amplifier which drives it, since any mismatch would cause a degradation of the system performance. In the design of the matching transformer the inductance of the leads, the capacity of the tube and by-pass condenser and the resistance of the elec-

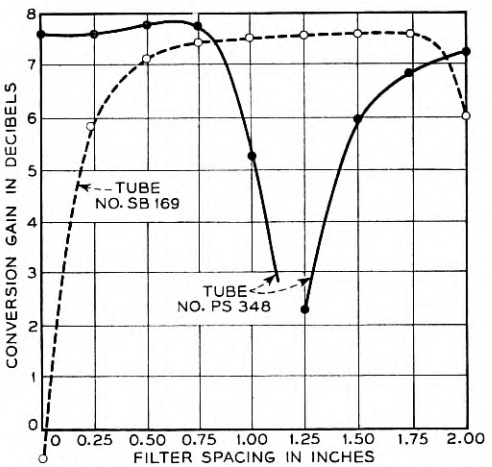


Fig. 9.—Data showing the effect of the spacing of a rejection filter for the wanted sideband in the input circuit.

tron stream were measured at the base of the tube. A broad-band transformer was designed and the inductances were thrown into an equivalent T network, thereby utilizing the lead inductance inside the tube as a part of the transformer, absorbing it in the L_2-M branch as indicated in Fig. 10. In several experimental tubes the lead inductance was .04μH. The impedance match obtained with such a transformer gave less than two db SWR over a band from 55 to 75 mc with the loop at the cusp on the reflection coefficient chart characteristic of slightly over-coupled tuned transformers as shown in Fig. 11.

The broadband matching of the output circuit of the modulator required a different technique. Not only is this filter called upon to provide a broad-band impedance match, but also it should provide dis-

* The spacing of the input filter also affects the plate impedance in a complicated way.

crimination against the other microwave-frequency components present in the modulator output circuit; consideration of the beating oscillator and

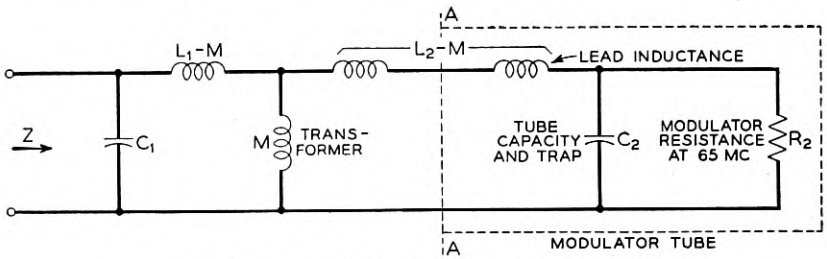


Fig. 10.—Equivalent circuit of modulator at I. F.

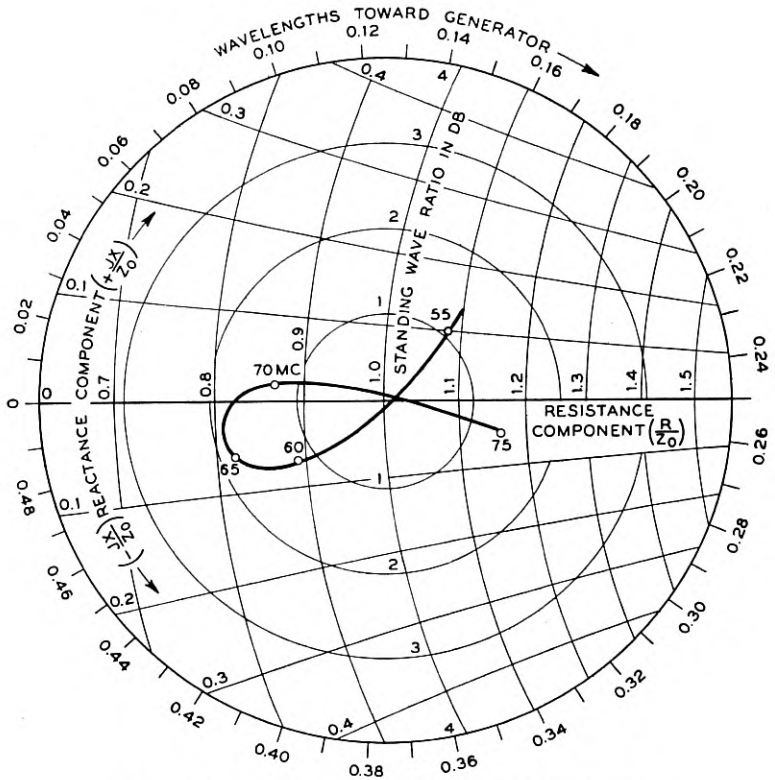


Fig. 11.—Modulator I. F. impedance with transformer.

both sidebands is necessary. The variables at our disposal are the bandwidth of the modulator output circuit and the number of cavity resonators which follow it. The desired quantities are the specified transmission

bandwidth and the attenuation required at the beating oscillator frequency. With two equations and two unknowns, the maximally-flat filter theory was applied to the circuit shown schematically in Fig. 12.[4] This indicated that an output circuit bandwidth of 84 mc (to the three db loss points), associated with two external resonant branches having bandwidths of 42 and 84 mc respectively, were needed to obtain a 20 mc flat band with 30 db suppression of the beating oscillator.

Such cavities were designed and attached to the output of a modulator whose bandwidth had been adjusted by means of small resistive strips.

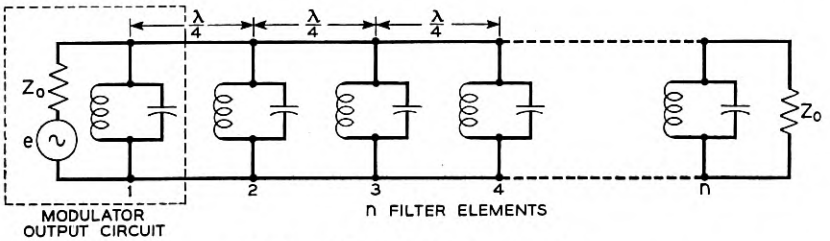


Fig. 12.—Sideband filter in waveguide.

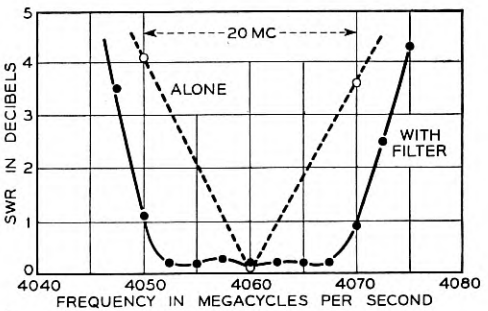


Fig. 13.—Output circuit impedance match.

The resulting impedance match gave a standing wave ratio of less than one db over a 20 mc band (the plate circuit alone without the filter was only about 5 mc wide to corresponding points) as shown in Fig. 13, and the beating oscillator power at the output of the filter was less than one tenth of a milliwatt, corresponding to 33 db discrimination.

The requirements and specifications for this particular experimental model do not necessarily reflect out present thoughts upon the requirements for any particular microwave radio relay system; they are presented here in some detail to indicate how certain specifications can be met, rather than to express what those specifications should be.

Other factors which influence the performance of the 416A modulator

are the plate voltage and the beating oscillator drive. The beating oscillator power affects the low level gain only slightly but has quite an effect on the gain at high power levels, that is, when the output power becomes comparable with the beating oscillator power. It is seen in Fig. 14 that compression becomes noticeable when the output power approaches within ten db of the beating oscillator driving power.

Varying the plate voltage on the modulator from 150V to 300V had little effect upon the conversion gain at low levels, but more power output

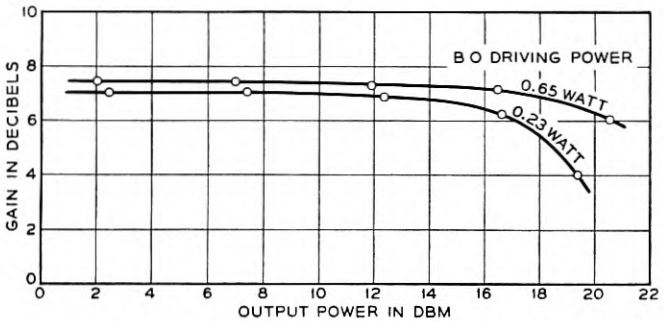


Fig. 14.—Modulator compression data for tube #PS62.

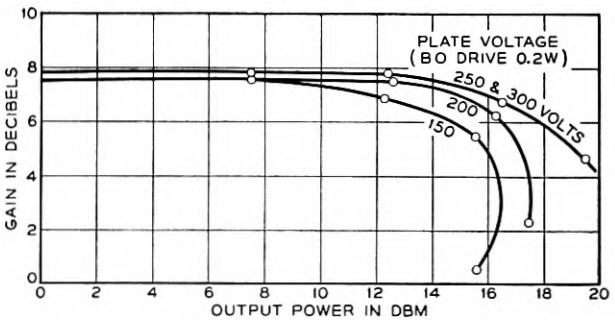


Fig. 15.—Modulator compression data for tube #PS348.

was obtained at the higher voltages. At 15 dbm power output, very little difference between 200V and 300V was observed, but at 150 V the gain was down two db, as shown in Fig. 15.

Fig. 16 shows the compression data for seven early experimental tubes, used as modulators with 200V on the plate, 14 ma cathode current, and 200 mw of beating oscillator drive. Half of these tubes had over seven db low level gain and only slight compression at power output levels of 13 dbm. The two poorest tubes would probably have been rejected before shipment, according to present standards of production. Each of the seven tubes was matched in impedance on the r.f. and i.f. inputs and also

on the r.f. output. The curves represent unloaded gain; no external loading was added to increase the bandwidth.

The performance of the close-spaced triode when used as a modulator appears to be superior in some respects to that of the silicon crystal modulators which are used in the New York-Boston microwave relay system.[1]

Single tubes had from 5 to 9 db gain compared with from 8 to 11 db loss for the crystals for corresponding power outputs. To get this performance the beating oscillator drive was only 200 milliwatts, compared with about 700 milliwatts for the crystal modulator. This reduction in r.f. power requirements means considerable simplification in a repeater.

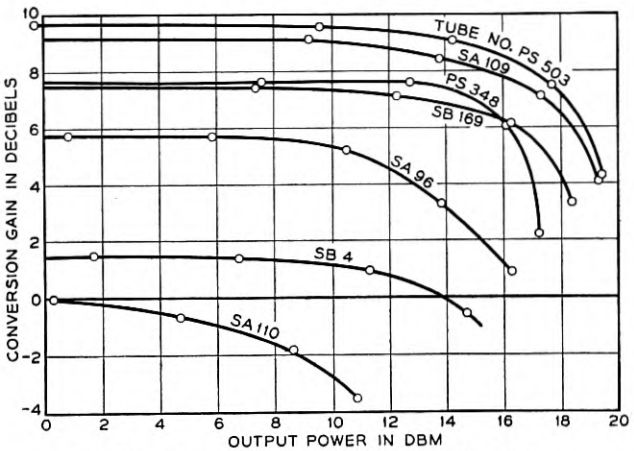


Fig. 16.—Compression data on seven 1553 triodes.
Plate voltage 200V
Plate current 14 Ma
B.O. Power 200 MW
Matched inputs and output

To offset this, the tube requires power supplies which are not necessary for the crystals, but low voltage power supplies should be cheap. The bandwidth of the tube modulator, 60 to 80 mc is less than the very wide (500 mc) band of the crystal modulator but it is comparable with the band width of the extra i.f. stages needed to drive the crystal modulator. The life of the tubes, although very little data are available as yet, will probably be less than the practically indefinite life of the silicon point contact modulators.

Amplifier

The performance of the close-spaced triode as an amplifier can best be described by referring to its impedance match, gain, transmission bandwidth and compression.

In some of the experimental tubes, bandwidths to the half power points of 21 mc to 250 mc have been measured. Typical of one of the better tubes, though not the best one, are the data contained in Fig. 17. The bandwidth of the input circuit is about twice that of the output circuit, and the SWR slumps outside the band on the low frequency side. The output impedance is more regular, exhibiting the familiar standing wave

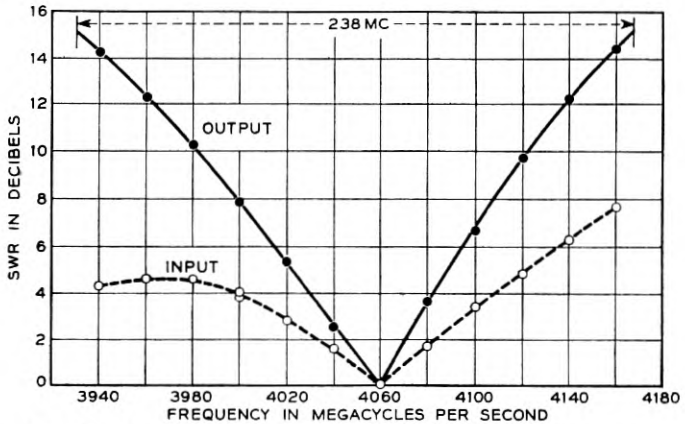


Fig. 17.—Input and output standing wave ratio versus frequency.

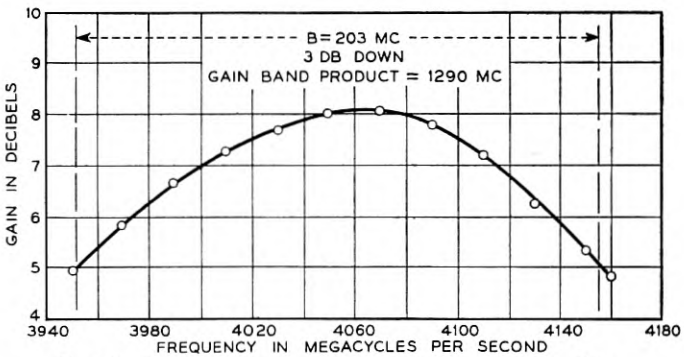


Fig. 18.—Transmission characteristic of a one-stage amplifier.

ratio of a simple single tuned resonant circuit. When the output impedance is plotted on the Smith reflection coefficient chart, the circle which results is also similar to that of a single tuned circuit. This is desirable since it then becomes a simple matter to incorporate the plate circuit in a maximally-flat filter of as many resonant branches as are needed, in the same way that the modulator output circuit was treated.

The transmission bandwidth for this single stage amplifier was 203 mc to the half power points, as shown in Fig. 18. This, with a gain of 8.05 db

at midband, gave a gain-band product of 1290 mc. The bandwidth of 203 mc was considerably greater than the average for these tubes. Similar results on 35 experimental tubes yielded the following averages: Low-level gain 10 db; Bandwidth 103 mc; Gain-band product 916 mc. The 416-A tubes produced by Western Electric Company exhibit comparable averages with much less spread; for example, a recent sample of 138 tubes had average values and standard deviations as follows:

TABLE I
GAIN AND BANDWIDTH OF 138 W. E. CO. 416-A TRIODES

	Average	St'd. Dev.
Low-level gain	9.9 db	1.1 db
Bandwidth	110 mc	9 mc
Gain-band product	1080 mc	350 mc

It is indeed gratifying to realize that such a remarkable tube can be produced with such uniformity.

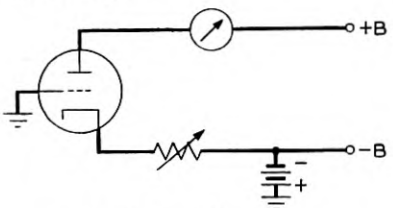


Fig. 19.—Stabilizer circuit.

In operating these tubes, it has been found that small variations in gain due to power line fluctuations and due to other disturbing influences can be minimized by using a stabilizing bias network which provides a large amount of negative feedback for the dc. path. This circuit is similar to one proposed by Mr. S. E. Miller of the Bell Telephone Laboratories for use in coxial repeaters which also use high transconductance tubes. In this circuit, shown schematically in Fig. 19, a few volts negative are applied to the cathode through a suitable dropping resistor. In the absence of plate voltage, the grid draws current, being positive with respect to the cathode. When plate voltage is applied, the drop in the cathode resistor tends to bring the cathode nearer ground potential until a stable voltage is reached. The resistor is set to a value which allows the desired cathode current to flow and subsequent variations in g_m or plate voltage then have little effect on the total cathode current.

Maintaining the cathode current constant does have an appreciable effect on the gain of the tube when operating at high output levels. This is characterized by a decrease in gain as the driving power is increased.

Fig. 20 illustrates this point. The low-level gain of this tube was 12.3 db but when the tube was driven so as to have an output power of 400 mw the gain was only about 3 db. At this point, retuning the circuit to rematch the tube at the high output level increased the gain to about 5 db. Now, returning to low level, the gain was only 10 db. Presumably in between these two points, 5 db at 500 mw output and 12.3 db at less than one milliwatt output, the performance could have been better than either of these two curves shows, i.e., the performance could have been improved by rematching at each intermediate power level.

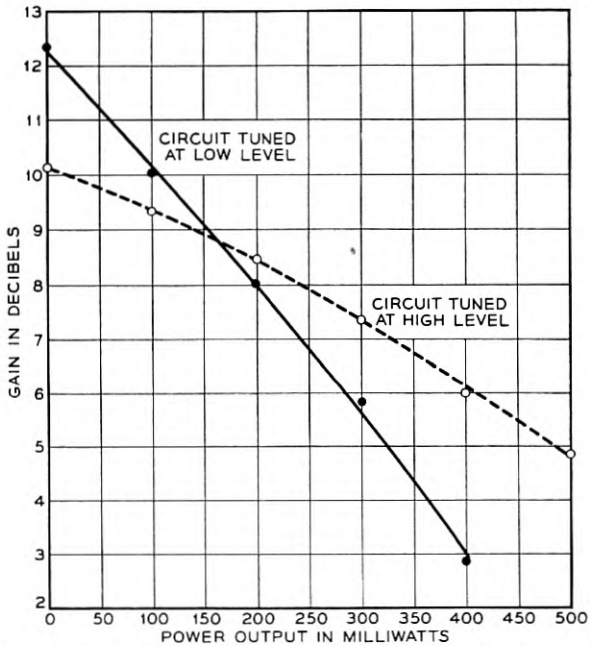


Fig. 20.—"Compression" in a one-stage microwave amplifier $I_p = 30$ ma.

This tube is not representative of all of the tubes tested. It is rather poorer in the spread of the two curves than most. It was picked merely to illustrate that besides a drop in gain also a detuning effect takes place when the driving power is changed. In the example given here the cathode current was held at or near 30 ma by the stabilizing bias circuit.

Without the stabilizing circuit, these so-called "compression" curves would be quite different. For instance, if the bias were held constant, we should expect that the gain would not drop as fast as indicated here, since the plate current would rise as the drive was increased.

At any rate, in an F.M. system, we are not concerned with how much

"static compression" exists, but rather with how much gain can be realized without exceeding the dissipation ratings of the tubes.

With this in mind data were taken on 25 of the experimental tubes. In each case they were matched to the input and output waveguides and the cathode current was stabilized at 30 ma. After driving the tube to a high level of output power, the circuits were rematched and the resulting "compression" curves revealed the capabilities tabulated.

TABLE II
SUMMARY OF DATA ON 25 EXPERIMENTAL CLOSE-SPACED TRIODES

	Highest	Lowest	Average
Low level gain	12.3 db	3.8 db	7.8 db
Gain (500 mw output)	7.0 db	−8.0 db	1.82 db
Power Output (3 db gain)	950 mw	50 mw	455 mw

It can be seen from the table that we might expect to obtain a gain of 20 or 25 db with three or four stages with a power output of about 500 mw and a flat band of over 20 mc.

THREE STAGE AMPLIFIER

A three-stage amplifier with 24 db gain has been assembled using an earlier type of circuit and loop tested at low levels on the equipment of Messrs. A. C. Beck, N. J. Pierce and D. H. Ring.[5] This amplifier had a bandwidth of about 30 mc to the 1 db points and while it does not represent the best that can be done with the 416A tube, the results of the loop test are interesting.

The recirculating pulse test, or loop test, is performed on a repeater component to determine its ability to reproduce a pulse faithfully after repeated transmissions. The output of the amplifier is connected to its input through a long delay line and an adjustable attenuator. The overall gain of the loop thus formed is adjusted to unity or zero db so that an injected pulse will recirculate through the loop without attenuation but accumulating distortion with each round trip. After allowing the pulse to recirculate long enough the amplifier is blanked out or quenched and the recirculating pulse amplitude dies out, thus preparing the loop for the next injected pulse, when the process is repeated. With a pulse length of one microsecond and an overall delay of two microseconds, one hundred round trips occur in 0.2 milliseconds, thus allowing the process to be repeated at the rate of two or three thousand times per second. A cathode ray oscilloscope is used to examine the pulse shapes, and its sweep is synchronized to the injected pulse so that successive corresponding pulses are superposed, enabling the operator to examine the pulse after any

number of round trips or select individually the nth round trip for inspection.

Fig. 21(a) shows the complete cycle between successive injected pulses, and the individual pulses that follow cannot be resolved at this slow sweep speed. Fig. 21(b) shows the first 26 round trips resolved so that they are distinguishable. Figure 21(c) shows the first and second round trips

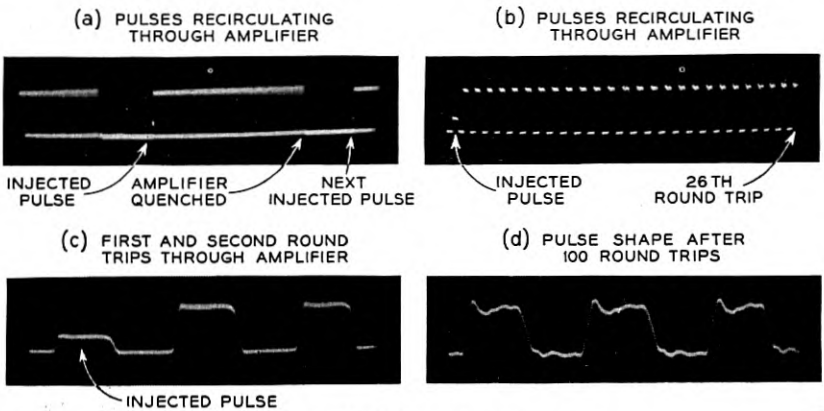


Fig. 21.—Recirculating pulse test patterns.

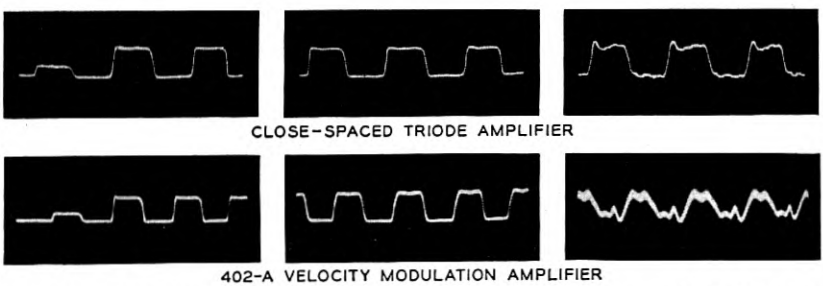


Fig. 22.—Recirculating pulse patterns showing 1st, 10th and 100th round trips for: Top: Close-spaced triode amplifier. Bottom: 402-A velocity modulation amplifier.

through the amplifier, with little or no distortion discernible. Fig. 21(d) gives, to the same scale as the preceding picture, the pulse shape after 100 round trips. A little overshoot and subsequent oscillation is now visible, although the whole pulse shape is still not too bad.

In Fig. 22, these results are compared with the results of a similar test performed on a four-stage, stagger tuned, stagger-damped amplifier using the 402 velocity variation amplifier tubes; the first, the tenth and the hundredth round trips are shown. Little or no distortion is seen at the

tenth round trip, but the superiority of the 416A amplifier is clearly shown in the hundredth round trip.

Both amplifiers were operating at low levels, and the pulse was an amplitude modulated one. Since these are not the conditions under which our microwave radio relay circuits operate, conclusions should not be drawn about how many repeater stations can now be put in tandem. The

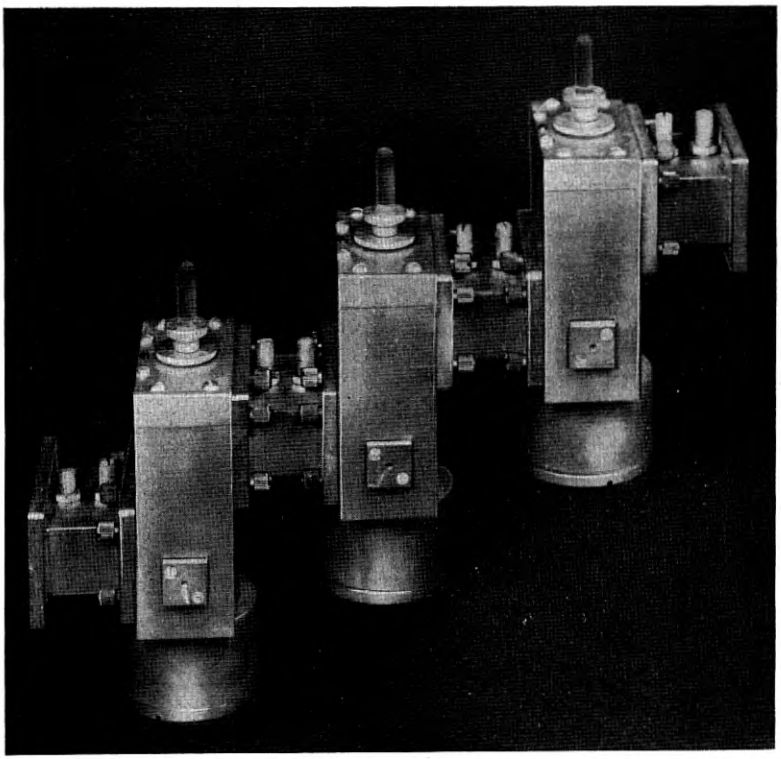


Fig. 23.—An assembled three-stage microwave amplifier.

test merely indicates that an improvement has been made, thus corroborating the evidence obtained by other tests.

Still further improvement has been made since loop testing the model ten amplifier. A three stage 416A amplifier (see Fig. 23) using model eleven circuits had comparable gain, 23 db, but a bandwidth of 50 mc to points 0.1 db down. These data again are for low level operation, but it is reasonable that half a watt might be expected from four such stages with comparable gain and slightly narrower bandwidth, surely 30 mc.

Noise Figure

In a forward looking program it is well to keep in mind other possibilities for this tube, such as use in a straight through type of repeater in which all of the amplification is obtained at microwave frequencies. In such an application the noise figure of the triode becomes one of its limitations, since the 416A must compete with the low noise figure of the silicon crystal converter which, for the New York-Boston circuit, is around 14 db. Data on thirty five early experimental and production 416A tubes gave an average value of 18.08 db at 4060 mc.* Each of the

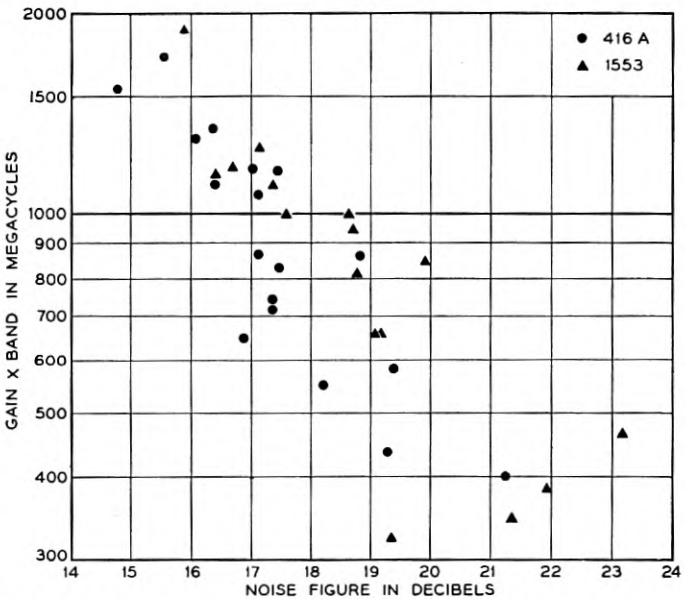


Fig. 24.—Noise figure vs gain-band product for close-spaced triode.

tubes was operated at 200 volts with 30 ma space current and was tuned so as to present matched impedances to the input and the output waveguides. The best of this batch had a noise figure of 14.79 db and the poorest 23.2 db. These measurements were made with a fluorescent light noise source.[6]

An interesting correlation between noise figure and gain-band product was uncovered during these tests, as can be seen in Fig. 24, which gives the noise figure in db on the abscissa and the gain-band product in mega-

* More recently, a sample of twelve production 416A tubes ranged from 13.5 to 16.2 db and averaged 15.06 db noise figure, with a standard deviation of 0.8 db.

cycles along the logarithmic ordinate. The points scatter between the extremes of 15 db noise figure for a gain-band product of 2000 megacycles to 23 db noise figure at 400 megacycles gain-band product. Extrapolating from these data, a noise figure of 10 db might be achieved if the gain-band product could be increased to 5500 mc. It is reasonable to expect that an improvement of this amount can be achieved if the resistance and return electron losses inside the tubes can be eliminated.[7]

We may use these data to determine the expected noise figure of a straight through amplifier, thus:

$$F = F_A + \frac{F_B - 1}{G_A} + \frac{F_C - 1}{G_A G_B} \cdots \tag{1}$$

If, for example, we assume that all stages are alike in noise figure and in gain, equation (1) approaches the expression, as the number of stages increases without limit:

$$F \underset{n\to\infty}{\overset{\lim}{=}} \frac{F_A G_A - 1}{G_A - 1} \tag{2}$$

Using an average value of 10 db gain per stage, the overall noise figure would be as follows:

(1) For $F_A = 30$ (best tube, 14.79 db)
$F = \frac{29.9}{9} = 33.2$ or 15.2 db
(2) For $F_a = 64$ (average tube, 18.08 db)
$F = \frac{63.9}{9} = 71$ or 18.5 db.

Straight-Through Amplifier

The actual performance of a ten-stage amplifier was about what should be expected from the considerations above. The best tube (10 log F = 14.79 db) was used in the first stage, and the next best tube in the second stage. The measured overall noise figure was 15.96 db. The overall gain was 90 db and the band was flat to 0.1 db for 44 mc. Such an amplifier with its associated power supply and individual control panels is shown in Fig. 25.

Conclusions

A circuit is described which lends itself readily to utilizing the 416A close-spaced triode as a modulator or a cascade amplifier for microwave repeaters operating at 4000 mc. Data are presented on early experimental models of the tube.

As modulators, single tubes had from 5 to 9 db gain with 10 to 20

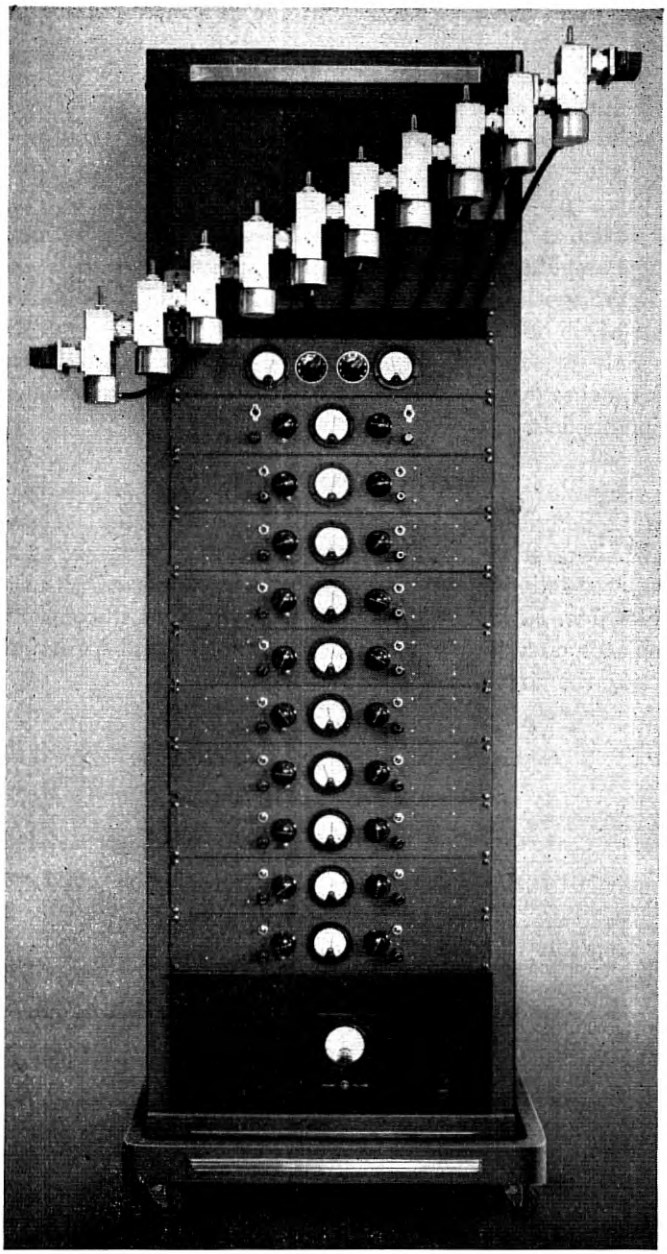


Fig. 25.—A ten-stage microwave amplifier operating at 4000 mc.

mw output when driven with 200 mw of beating oscillator power. A bandwidth of twenty megacycles was readily obtained.

As amplifiers at 4060 mc, the average gain of 60 tubes was 9 db, the average bandwidth of 34 tubes was 103 mc to the half power points, the average noise figure of 35 tubes was 18.08 db and the average power output (for 3 db gain) was 455 mw for 25 tubes. Operating the tubes in cascade produced an amplifier which had less distortion of pulse shape than an earlier amplifier which used the 402-A velocity variation tube. A ten-stage amplifier has been assembled and tested, yielding 90 db gain, a noise figure (with selected tubes) of 15.96 db and a bandwidth of 44 mc to the 0.1 db points.

These data are for early experimental models of the tube and it is likely that subsequent alterations may improve the performance in the production models.

Acknowledgments

The work described in this paper took place at the Holmdel Radio Research Laboratories. Mr. Bowen, with the able assistance of Mr. E.L. Chinnock, was active in pursuing the problems connected with the amplifier circuits. Mr. R. H. Brandt helped with the work in connection with the modulator circuits and many others were helpful in designing and constructing the circuits and facilities for testing.

References

1. "Microwave Repeater Research," H. T. Friis, *B. S. T. J.*, Vol. 27, pp. 183–246, April 1948.
2. "A Microwave Triode for Radio Relay," J. A. Morton, *Bell Labs. Record*, Vol. 27, #5, May 1949.
3. "Microwave Converters," C. F. Edwards, *Proc. I. R. E.*, Vol. 35, pp. 1181–1191, Nov. 1947.
4. "Maximally-Flat Filters in Wave Guide," W. W. Mumford, *B. S. T. J.*, Vol. 27, pp. 684–713, Oct. 1948.
5. "Testing Repeaters with Circulated Pulses," A. C. Beck and D. H. Ring, *Proc. I. R.E.*, Vol. 35, pp. 1226–1230, November 1947.
6. "A Broad-Band Microwave Noise Source," W. W. Mumford, *B. S. T. J.*, Vol. 28, pp. 608–618, October 1949.
7. "Electron Admittances of Parallel-Plane Electron Tubes at 4000 Megacycles," Sloan D. Robertson, *B. S. T. J.*, Vol. 28, pp. 619–646, October 1949.
8. "Design Factors of the Bell Telephone Laboratories 1553 Triode," J. A. Morton and R. M. Ryder, *B. S. T. J.*, Vol. 29, #4, pp. 496–530, Oct. 1950.

A Wide Range Microwave Sweeping Oscillator

By M. E. HINES

(*Manuscript Received July 24, 1950*)

1. Introduction

A SWEPT frequency oscillator is a useful laboratory tool for testing wide-band circuit components. It permits an oscillographic display of a frequency characteristic, avoiding much of the labor of point-by-point testing at discrete frequencies. There was a particular need in the Bell Telephone Laboratories for a sweeping oscillator to cover the communications band between 3700 and 4200 megacycles to facilitate the testing of components for radio relay repeaters.

This paper describes one type of oscillator designed to satisfy this need. It utilizes the BTL 1553 (or the Western Electric 416A) microwave triode. The tuning is accomplished mechanically so that the frequency varies continuously back and forth over the band at a low audio frequency rate. Continuous oscillations have been obtained over a 900 megacycle band from 3600 to 4500 megacycles.

2. Circuit Structure

Basically, the rf circuit consists of a tunable cavity for a grid-anode resonant circuit, a means for feedback to an untuned grid-cathode circuit, and a means for coupling the cavity to a waveguide output. The grid-anode cavity is the only sharply tuned circuit, and it was found that oscillations could be obtained over the entire band by changing the resonant frequency of that cavity alone. In this application, the electronic conductance between the grid and cathode is so high that this portion of the circuit has an inherent broad band such that separate tuning is unnecessary.

The necessity for continuous, rapid tuning virtually requires that there be no sliding contacts in the tuning mechanism. A type of cavity was chosen so that tuning could be accomplished by a simple variable capacitor of the non-contacting type. Reduced to its simplest elements, it consists of a short coaxial line, resonant in the half-wave mode. Actually the line is much shorter than a half wavelength because of excess capacitance at both ends. At one end is the capacitance of the grid-anode gap, and at the other end is the variable capacitor used for tuning.

The actual cavity is illustrated in Fig. 1. This is somewhat more complicated than a half wave line, but the mode of resonance is essentially the same. The variable capacitor utilizes a thin-walled copper cup which is movable vertically. This cup fits rather closely inside, and is coaxial with, a cylindrical hole in the main body of the cavity. It forms the center

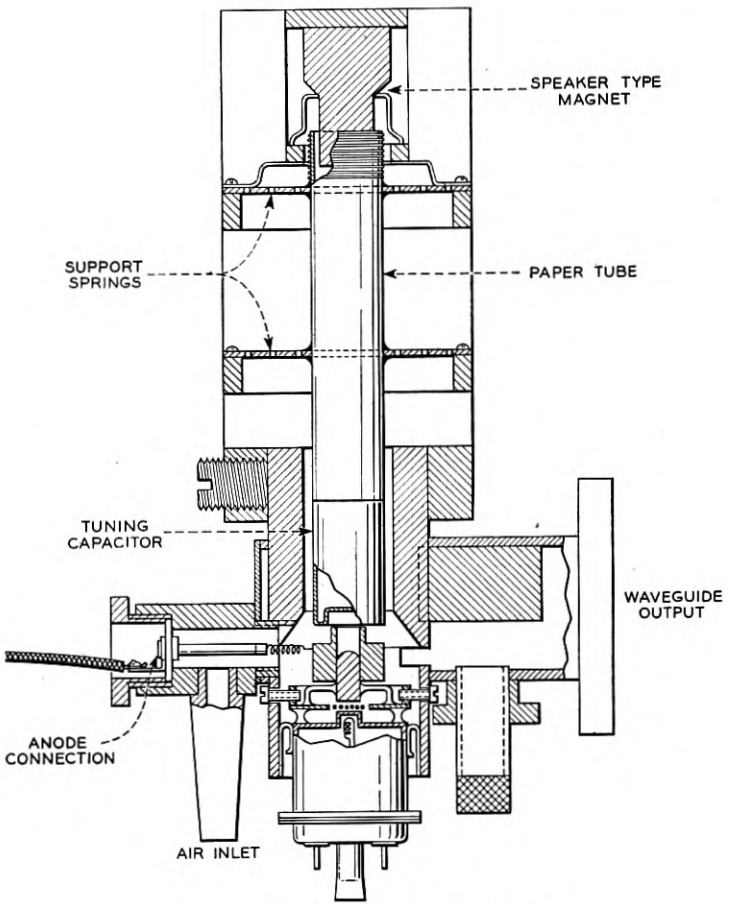


Fig. 1.—Construction of the oscillator.

conductor of a low impedance coaxial line approximately one-fourth wavelength long, so that in this frequency range it is effectively short-circuited to the cavity wall. Vertical motion of the cup is therefore roughly equivalent to moving the end wall, thereby changing the capacitance between the wall and the center conductor of the main cavity. The reso-

nant frequency is lowest when the two surfaces are nearly in contact, and highest when the cup is fully extracted.

The recessed end of the cup fits over a protuberance on the center conductor when they are nearly in contact. This special shape was designed to give a reasonably straight curve of frequency vs. displacement. With planar surfaces, the frequency would change more rapidly with displacement at the low than at the high frequency end of the band.

The grid disk of the tube is separated from the wall of the cavity by a narrow annular space, and contact is made across the gap by a number of small screws. These screws act as an inductive reactance in series with the circulating currents of the resonant cavity. The voltage developed across this reactance is applied between the grid and the main envelope of the tube, and in this way energy is fed into the grid-cathode space to provide feedback.

The mechanical tuning device was adapted from an inexpensive permanent magnet loudspeaker of the type used in small home radios. The construction is shown in Figs. 1 and 5. The speaker cone was removed and the voice coil was attached to a thin-walled paper cylinder which supports the tuning cup inside the cavity. Two sheet fiber springs support the paper cylinder and maintain the axial alignment in the magnet and cavity. These springs are cut with a number of incomplete circular slits to reduce the stiffness for axial motion. With the voice coil actuated from a small filament transformer, peak to peak motion $\frac{1}{4}$ of inch is obtainable.

The heater and cathode connections are made at the base of the tube which protrudes from the cavity. The grid is internally connected to the main body of the cavity. The anode lead is brought out through a quarter-wave choke and mica button condenser.

To prevent overheating of the anode of the tube, air must be blown through the cavity. This is done by connecting a low pressure air hose to the air inlet shown in Fig. 1. Excessive air flow must be avoided, as it will cause erratic vibrations of the tuning plunger.

3. Adjustment and Operation

The degree of feedback is adjustable by changing the number and relative positions of the feedback screws which connect the cavity to the grid ring of the tube. There are 16 possible screw positions, but only about 5 or 6 are needed to obtain optimum feedback. Reducing the number of screws increases the amount of feedback.

Care should be taken that the spring which contacts the anode for dc connection is not of such a length to have resonances within the band. When such resonances exist, "holes" or other irregularities will be found in the output spectrum. This spring can act as a helical line, and when it

is too long, resonances will occur which can absorb power and otherwise affect the cavity impedance.

When properly adjusted and sweeping, the output is continuous and the frequency varies approximately sinusoidally back and forth over the band of interest. The width of the sweeping band depends upon the ac current in the voice coil of the speaker drive, and the center frequency depends upon the mean position of the tuning plunger. The latter can be adjusted mechanically by loosening the clamping screw and raising or lowering the sweeping mechanism by hand. It is also possible to make small adjustments of the center frequency electrically by adding a dc component to the voice coil driving current.

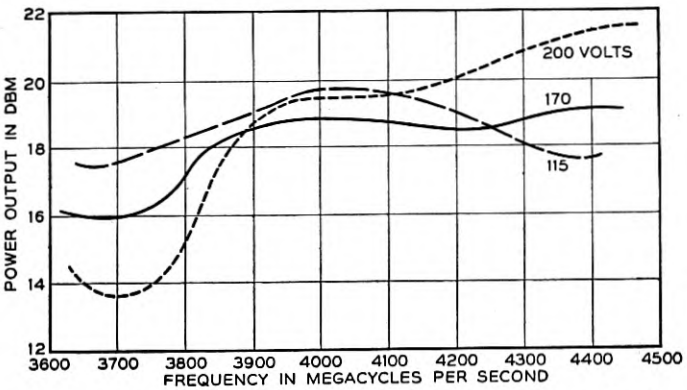


Fig. 2.—Power output curves at 115, 170, and 200 Volts on the anode, for a mean anode current of 25 ma.

Typical curves of power output vs. frequency, taken at different anode voltages, are shown in Fig. 2. The flattest curve requires a voltage considerably lower than the tube rating. The feedback phase is not optimum for best power output, a larger phase shift being desirable in this oscillator. The lowered voltage helps in this regard, increasing the electron transit time in the tube and thereby increasing the phase shift. Efficiency was sacrificed in this design to increase the tuning band. A longer feedback path would increase the power output, but would tend to narrow the band over which oscillation could be obtained by a single tuning adjustment.

The anode power supply should be variable between 100 and 250 volts, but need not be regulated because this voltage is not critical. A rheostat is used for cathode self-bias. The cathode heater and the sweeping mechanism are supplied from a single 6.3 volt filament transformer, with a potentiometer control to vary the sweep range.

A crystal detector and an oscilloscope are used to view the output. It is convenient to use a sinusoidal horizontal sweep on the oscilloscope, driven from the same 6.3 volt transformer as the mechanical sweeping mechanism. In this case, a phase shifter is needed to synchronize the oscilloscope sweep with the motion of the tuning plunger, because there is an appreciable mechanical phase shift in the loudspeaker mechanism.

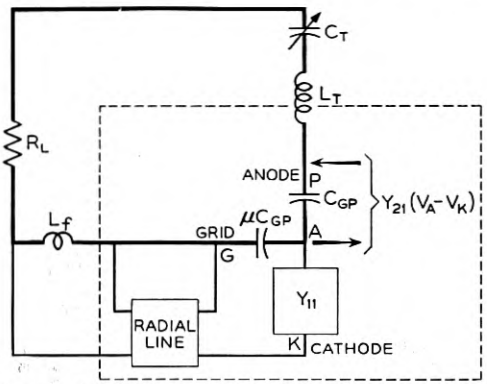


Fig. 3.—Simplified equivalent circuit of the oscillator.

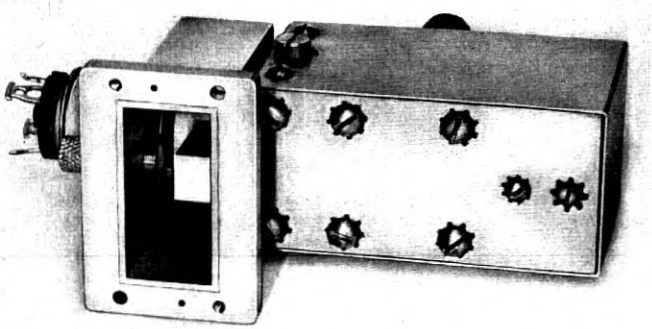


Fig. 4.—The complete oscillator, showing the output coupling window and the ridged waveguide coupling transformer.

When properly phased, the output spectrum will be displayed across the oscilloscope screen with the minimum and maximum frequencies at opposite ends of the trace. In addition, a vibrating relay (such as the Western Electric 275 B Mercury Relay) is used to short out the input to the oscilloscope during half of each cycle. This converts the return trace into a zero-signal reference line, so that the complete picture is a closed loop with a flat bottom. The separation of the active from the reference trace

is a direct indication of signal strength, displayed as a function of frequency.

The results reported here were obtained using the BTL 1553 tube, which is a laboratory model. Samples of the production model, Western Electric 416A, have also been used in this oscillator with quite similar results. To adapt the oscillator for the 416A, the grid ring should be threaded on the inside to fit the threads on the grid disk of that tube.

4. An Equivalent Circuit

The field configuration in the cavity of the oscillator is quite complex, and cannot be readily described in any quantitative fashion. The formulation of an equivalent circuit would require many approximations and

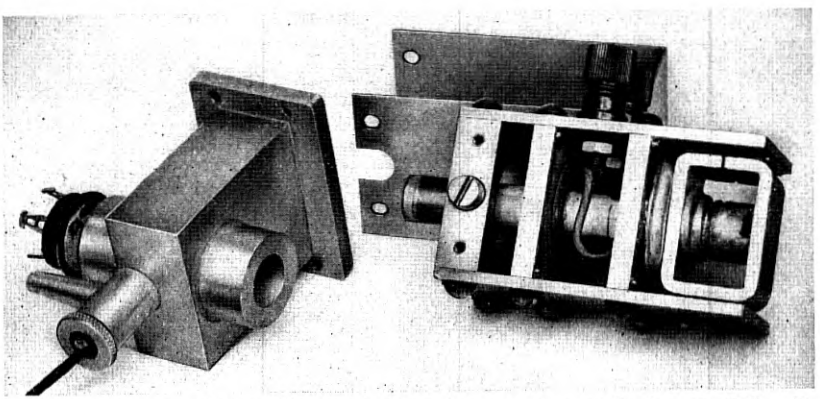


Fig. 5.—The complete oscillator, showing the sweeping mechanism partially dismantled.

judicious guesses if values are to be specified for the various circuit parameters. The circuit of Fig. 3 is believed to be equivalent in a qualitative sense.

A portion of the circuit is within the tube itself. This is the region enclosed by the dotted line in Fig. 3. The T of elements which include Y_{11}, C_{gp}, μC_{gp} and the injected currents, is the equivalent circuit of Llewellyn and Peterson[1] for the active region of a triode. Experimentally determined values for these quantities are reported by Robertson[2]. Y_{11} is the admittance of an equivalent diode between the grid and cathode, and the injected currents indicated by the arrows are the electronic transfer currents associated with the grid voltage and the transadmittance. At high

[1] Llewellyn and Peterson, "Vacuum Tube Networks," *I.R.E. Proc.*, Vol. 32, pp. 144–166 (March 1944).

[2] S. D. Robertson, "Electronic Admittances of Parallel-Plane Electron Tubes at 4000 Megacycles," *B. S. T. J.*, Vol. 28, p. 619, Oct. 1949.

frequencies, both the admittance Y_{11} and the transadmittance Y_{21}, are complex quantities which vary with frequency as shown by Llewellyn and Peterson. The 4-pole box shown represents the passive radial line between the glass seal at the edge of the tube and the cathode-grid gap. This line is heavily loaded with dielectrics and is believed to be electrically about a quarter wavelength long at 4000 Mc. The inductance connected to the anode is that of the anode pin itself and the coaxial center-conductor attached to it. A series resistor R_L is added to include the effects of cavity losses and loading by the output coupling window. C_t is the tuning capacitor which varies with tuning plunger position. The inductance L_f is the feedback reactance introduced by the screws connected to the grid disk.

5. Acknowledgments

I wish to acknowledge the assistance of Messrs. J. A. Morton, R. M. Ryder, and the late A. E. Bowen for many helpful suggestions in the design of this oscillator.

Theory of the Flow of Electrons and Holes in Germanium and Other Semiconductors

By W. VAN ROOSBROECK

(*Manuscript Received Mar. 30, 1950*)

A theoretical analysis of the flow of added current carriers in homogeneous semiconductors is given. The simplifying assumption is made at the outset that trapping effects may be neglected, and the subsequent treatment is intended particularly for application to germanium. In a general formulation, differential equations and boundary-condition relationships in suitable reduced variables and parameters are derived from fundamental equations which take into account the phenomena of drift, diffusion, and recombination. This formulation is specialized so as to apply to the steady state of constant total current in a single cartesian distance coordinate, and properties of solutions which give the electrostatic field and the concentrations and flow densities of the added carriers are discussed. The ratio of hole to electron concentration at thermal equilibrium occurs as parameter. General solutions are given analytically in closed form for the intrinsic semiconductor, for which the ratio is unity, and for some limiting cases as well. Families of numerically obtained solutions dependent on a parameter proportional to total current are given for *n*-type germanium for the ratio equal to zero. The solutions are utilized in a consideration of simple boundary-value problems concerning a single plane source in an infinite filament.

Table of Contents

1. Introduction 560
2. General Formulation 565
   2.1 Outline 565
   2.2 Fundamental equations for the flow of electrons and holes 566
   2.3 Reduction of the fundamental equations to dimensionless form 571
      2.31 The general case 571
      2.32 The intrinsic semiconductor 577
   2.4 Differential equations in one dimension for the steady state of constant current and properties of their solutions 578
3. Solutions for the Steady State 583
   3.1 The intrinsic semiconductor 584
   3.2 The extrinsic semiconductor: *n*-type germanium 586
   3.3 Detailed properties of the solutions 588
      3.31 The behavior for small concentrations 590
      3.32 The zero-current solutions and the behavior for large concentrations 593
4. Solutions of Simple Boundary-Value Problems for a Single Source 594
5. Appendix 599
   5.1 The concentrations of ionized donors and acceptors 599
   5.2 The carrier concentrations at thermal equilibrium 600
   5.3 Series solutions for the extrinsic semiconductor in the steady state 601
   5.4 Symbols for quantities 605

1. Introduction

IN A semiconductor there are current carriers of two types: electrons in the conduction band, and positive holes in the filled valence band; and the increase of their concentrations in the volume of the semicon-

ductor over the concentrations which obtain at thermal equilibrium is fundamental to a number of related phenomena, of which transistor action is a familiar instance. In an *n*-type semiconductor, for example, in which the carriers are predominantly electrons, the carrier concentrations are increased by the introduction of holes which, through a process of space-charge neutralization, produce additional electrons in the same numbers and concentrations. The bulk conductivity of the semiconductor is thereby so increased that power gain is obtainable.[1] Holes can be introduced by the local application of heat, or by irradiation with light, X-rays, or high-velocity electrons—in fact, by any agency which transfers electrons from the highest filled band to the conduction band. They can be introduced also through an emitter, which may be a positively biased point contact or a positively biased $p - n$ junction[3], as exemplified in the transistor. In this case the emitter introduces holes, which flow into the volume of the semiconductor[4], by the removal of electrons from the filled band.[2, 5] Entirely analogous considerations apply to the introduction of electrons into a *p*-type semiconductor.[6]

In their flow in a semiconductor, added electrons and holes are subject to drift under electrostatic fields and to diffusion in the presence of concentration gradients as a consequence of their random thermal motions. They are subject also to recombination, which results in concentration gradients in source-free regions even for the steady state in one dimension, or which augments those which may otherwise be associated with the time-dependence of the flow, or with its geometry in the steady state. From fundamental equations which take into account these phenomena of drift, diffusion, and recombination, for the existence of each of which there is experimental evidence[1], general differential equations and boundary-condition relationships in suitable reduced or dimensionless variables and parameters may be derived, and solutions which give the concentrations and flow densities of added carriers obtained for various cases of physical interest.

This paper presents results of a theoretical analysis, along these lines, of the flow of electrons and holes in semi-conductors. The treatment is intended particularly for application to germanium. An initial formulation,

[1] W. Shockley, G. L. Pearson and J. R. Haynes, *B. S. T. J.* 28, (3), 344–366 (1949).
[2] J. Bardeen and W. H. Brattain, *Phys. Rev.* 74 (2), 230–231 (1948); W. H. Brattain and J. Bardeen, *Phys. Rev.* 74 (2) 231–232 (1948).
[3] W. Shockley, G. L. Pearson and M. Sparks, *Phys. Rev.* 76 (1), 180 (1949); W. Shockley, *B. S. T. J.* 28 (3), 435–489 (1949).
[4] E. J. Ryder and W. Shockley, *Phys. Rev.* 75 (2), 310 (1949); J. N. Shive, *Phys. Rev.* 75 (4), 689–690 (1949); J. R. Haynes and W. Shockley, *Phys. Rev.* 75 (4), 691 (1949).
[5] J. Bardeen and W. H. Brattain, *Phys. Rev.* 75 (8), 1208–1225 (1949); *B. S. T. J.* 28 (2), 239–277 (1949).
[6] W. G. Pfann and J. H. Scaff, *Phys. Rev.* 76 (3), 459 (1949); R. Bray, *Phys. Rev.* 76 (3), 458 (1949).

which retains, wherever convenient, such generality as is instructive per se or of manifest utility, is specialized so as to apply to the steady state of constant current in a single cartesian distance coordinate. For the intrinsic semiconductor, general analytical solutions are obtainable in closed form, and such solutions are given, as well as general solutions obtained numerically for *n*-type germanium in which the hole concentration at thermal equilibrium may be neglected compared to the electron concentration. Solutions for these cases are given explicitly for each of two recombination laws: recombination according to a mass-action law, and recombination such that the mean lifetime of the added carriers is constant. Methods are described for the fitting of boundary conditions, and the following relatively simple boundary-value problems are considered: a source at the end of a semi-infinite semi-conductor filament; and a single source in a doubly-infinite filament.

To indicate the presumed scope and application of the results obtained, it may suffice to outline briefly the principal assumptions on which they are based and the approximations employed: The assumption is made at the outset that trapping effects may be neglected, which provides the important simplification that the recombination rates of holes and electrons are equal at all times. One justification for this is the circumstance that the fairly high hole mobilities found by G. L. Pearson from Hall-effect and conductivity measurements[7] are no larger than those found by J. R. Haynes from transit times under pulse conditions[1]. With hole trapping, holes injected in a pulse would initially fill traps; and if there were subsequent relatively slow release of the holes from the traps, an apparent reduction of mobility would be manifest. It is further assumed that substantially all donor and acceptor impurities are ionized. With the assumption that the semi-conductor is homogeneous in its bulk, and free from grain boundaries[8] or rectifying barriers, the assumption of the electrical neutrality of the semiconductor, or of the neglect of space charge, is in general an excellent approximation: Small departures from electrical neutrality in the volume would vanish rapidly, with time constant equal to that for the dielectric relaxation of charge, which for germanium equals $1.5 \cdot 10^{-12}$ sec per ohm cm of resistivity[9] and is in general small compared with the mean lifetime of added carriers. A uniform local departure from electrical neutrality in germanium of only one per cent in relative concentration would produce appreciable changes in field in a

[7] G. L. Pearson, *Phys. Rev.* 76 (1), 179–180 (1949).

[8] G. L. Pearson, *Phys. Rev.* 76 (3), 459 (1949); W. E. Taylor and H. Y. Fan, paper OA5, and N. H. Odell and H. Y. Fan, paper OA2 of the 1950 Annual Meeting of the American Physical Society, February 3, 1950.

[9] A value of 16.6 for the dielectric constant of germanium is obtained from optical data of H. B. Briggs: *Phys. Rev.* 77 (2), 287 (1950).

mean free path for the carriers, equal to $1.1 \cdot 10^{-5}$ cm at room temperature, which would even preclude the applicability of the fundamental equations employed. In qualitative terms, the conductivity of the semiconductor is sufficiently large that the currents which commonly occur are produced by moderate fields whose maximum gradients are relatively small. Space charge may persist in the steady state, but then only in surface regions whose thickness[10] in germanium is generally less than about 10^{-4} cm and whose effects may be dealt with through suitable boundary conditions.

The steady-state solutions, in their qualitative aspects, are illustrative of the phenomena taken into consideration. In an extrinsic semiconductor, if the concentrations of added carriers are not too large, the solutions for moderate and large fields are in general approximately ohmic in their local behavior. The effect of diffusion is then comparatively small, and the added carriers largely drift under a field which varies with distance through the increased conductivity which these recombining carriers themselves produce. Diffusion effects are incident in addition to this behavior, and become pronounced for large concentrations or small applied fields. For example, solutions which specify the concentrations of added holes as functions of distance, for different total currents or applied fields in a source-free region, all approach a common solution for large hole concentrations, regardless of applied field; those for the hole current and the electrostatic field behave similarly. This behavior results from diffusion in conjunction with the increase in conductivity. Another example is that of the solutions for zero total current: As the result of diffusion in conjunction with recombination, a flow of added holes can occur along a semi-conductor filament with no flow of current. It is, of course, accompanied by an equal electron flow, so that the hole and electron currents cancel, and occurs in any open-circuited semi-conductor filament which adjoins a region in which added holes flow. It can also be realized by suitable irradiation of an end of a filament, with no applied field. A closely related effect is illustrated in the flow of holes injected through a point-contact emitter into a semi-conductor filament along which a sweeping field is applied: Some of the holes will flow against the field, an appreciable proportion, unless the current in the filament is sufficiently large. As a further example, if the mobilities of holes and electrons were equal, the electrostatic field would be given by Ohm's law as the total current

[10] The (largest) distance over which the increment in electrostatic potential exceeds kT/e may be expressed in units of the length $L_d \equiv (kT\epsilon/8\pi n_i e^2)^{\frac{1}{2}}$, where n_i is the thermal-equilibrium concentration of electrons (or holes) in the intrinsic semiconductor; see the paper of reference 3, also W. Schottky and E. Spenke, *Wiss. Veröff. Siemens-Werken* 18 (3), 1–67 (1939). This distance increases with resistivity, never exceeding the value 1.4 L_d for the intrinsic semiconductor. In high back voltage n-type germanium, it exceeds about 0.5 L_d, and L_d for germanium is about $7.4 \cdot 10^{-5}$ cm at room temperature.

divided by the local increased conductivity. With electrons more mobile than holes, this ohmic field is modified by a contribution which is directed away from a hole source and proportional to the magnitude of the concentration gradient divided by the local conductivity. This contribution gives a non-vanishing electrostatic field for zero total current.

The intrinsic semiconductor has, as the result of a conductivity which is everywhere proportional to the concentration of carriers of either type, the property that the flow in it is as if the added carriers were actuated entirely by diffusion, with only the carriers normally present drifting under a field equal to the unmodulated applied field. The extrinsic semiconductor becomes in effect intrinsic if the concentrations of carriers are sufficiently increased, by whatever means, the ohmic contribution to the current density of either electrons or holes then becoming proportional to the total current density and, in this case, negligible compared with the contribution due to diffusion. It may, for example, be expected that the transport velocity of added carriers in an extrinsic semiconductor can be increased by an increase in the applied field only if the consequent joule heating does not unduly modify the semiconductor in the intrinsic direction.

General solutions for the steady state in one dimension are obtainable analytically in closed form for a number of important special cases. Aside from that for which diffusion is neglected, they include the general cases for no recombination, for the intrinsic semiconductor, and for zero total current, and the limiting cases of small and of large concentrations of added carriers. W. Shockley has made use of small-concentration theory in an analysis of $p - n$ junctions[3]. J. Bardeen and W. H. Brattain have given solutions for the steady-state hole flow in three dimensions, neglecting recombination, in the neighborhood of a point-contact emitter.[5, 11] Transient solutions are obtainable analytically for the intrinsic semiconductor for constant mean lifetime, and for the extrinsic semiconductor if the concentrations of added carriers are sufficiently small that the change in conductivity is negligible. For concentrations unrestricted in magnitude, Conyers Herring has described a general method for graphical or numerical construction of transient solutions in one dimension from a first-order partial differential equation appropriate to the case for which diffusion is neglected in the extrinsic semi-conductor, and has given some solutions so obtained, with estimates of the effect of diffusion. Reference might be made to his paper[12] also for discussion of various physical con-

[3] loc. cit.
[5] loc. cit.
[11] See the paper of J. Bardeen in this issue.
[12] Conyers Herring, *B. S. T. J.* 28 (3), 401–427 (1949).

siderations and of certain interesting transient effects. Steady-state alternating-current theory for relatively small total hole concentrations in the n-type semiconductor has been used to describe the action of the filamentary transistor[13] for which diffusion may in general be neglected.[1]

The steady-state solutions in one dimension apply to single-crystal semiconductor filaments, and for critical comparisons between theory and experiment, the ideal one-dimensional geometry should be simulated as closely as possible. Experimental estimates of hole concentrations and flows are frequently obtained from measurements of potentials and conductances of point contacts along a filament[1]. These estimates require a knowledge of the dependence of the current-voltage characteristics of point contacts on hole concentration. Theory for this dependence has been presented by J. Bardeen[11], and the determination of hole concentrations by means of the solutions here given should provide an essential adjunct to this point contact theory for its comparison with experiment.

2. General formulation

2.1 *Outline*

The formulation of the general problem is initiated by writing the fundamental equations for the time-dependent flow of holes and electrons in a source-free region of a homogeneous semiconductor under the assumption that there is no trapping. Conditions for their validity are discussed. Neglecting changes in the concentrations of ionized donors and acceptors, the fundamental equations are expressed in reduced or dimensionless form by suitable transformations of the dependent and independent variables. They are simplified so that the general problem is formulated by means of second-order partial differential equations in two dependent variables, one for concentration and the other for electrostatic potential; corresponding equations are derived for the intrinsic semiconductor. Various properties of the equations are adduced. For the flow in one dimension, a differential equation in the hole concentration is given for the n-type semiconductor, accompanied by expressions for the electrostatic field and hole flow density, as well as by some boundary-condition relationships involving specification of the latter. The equations for this case are found to depend on three parameters: the ratio of electron to hole mobility; a reduced concentration of holes at thermal equilibrium; and a parameter which fixes the total current density.

The recombination of holes and electrons is specified by means of a

[1] loc. cit.
[11] loc. cit.
[13] W. Shockley, G. L. Pearson, M. Sparks, and W. H. Brattain, *Phys. Rev.* 76 (3), 459 (1949).

suitable function of the concentration of the added carrier, whose form is specified for two recombination laws: recombination according to a mass-action law, and recombination characterized by constant mean lifetime. It is shown that essentially the same reduced equations apply to the case for which recombination is neglected.

Second-order differential equations in the hole concentration for the n-type semiconductor with the thermal-equilibrium value of the hole concentration assumed negligible compared to the electron concentration, and for the intrinsic semiconductor, are then written for the steady state of constant current in one dimension. These are converted into first-order equations which have, as dependent variable a reduced concentration gradient G, and as independent variable a reduced concentration of added holes, ΔP. Boundary conditions are expressed as relationships between these variables. Properties of the general solutions and of the boundary conditions are accordingly examined in the $(\Delta P, G)$-plane. It is found that there are two intersecting solutions through the $(\Delta P, G)$-origin, which is a saddle-point of the differential equation, and that these are the solutions for field directed respectively towards and away from sources in semi-infinite regions which have sources only to one side. They are called field-opposing and field-aiding solutions, and possess two degrees of freedom. Solutions which do not intersect at the origin are asymptotic to these, possess three degrees of freedom, and are called solutions of the composite type. This is the general type, and applies to a finite region in distance at both ends of which boundary conditions are specified. The region may, for example, be one between a source and either another source, a sink, a non-rectifying electrode, or a surface upon which recombination takes place. While the analysis of composite cases is straightforward, the present treatment is confined to the simpler cases of field opposing and field aiding, the latter being the one most generally applicable to experiments in hole injection. Also, where the differential equations involved are linear, solutions for composite cases can be written as linear combinations of field-aiding and field-opposing solutions.

From the properties of the curves in the $(\Delta P, G)$-plane is determined the qualitative behavior of the hole concentration at a hole source at the end of a semi-infinite filament as the total current is indefinitely increased.

2.2 *Fundamental equations for the flow of electrons and holes*

The equations for the flow in three dimensions of electrons and holes in a homogeneous semiconductor contain, as principal dependent variables, the hole and electron concentrations, p and n, the flow densities $\mathbf{J}_p$ and $\mathbf{J}_n$, and the electrostatic field, $\mathbf{E}$, or potential, V. With no trapping,

the equations may be written in a symmetrical form, so that they are applicable to either an n-type, a p-type, or an intrinsic semiconductor, as follows:

$$
(1)\quad
\begin{cases}
\dfrac{\partial p}{\partial t} = -\left[p/\tau_p - g_0\right] - \operatorname{div} \mathbf{J}_p \\[2ex]
\dfrac{\partial n}{\partial t} = -\left[n/\tau_n - g_0\right] - \operatorname{div} \mathbf{J}_n \\[2ex]
\mathbf{J}_p \equiv \dfrac{1}{e}\mathbf{I}_p = \mu_p\left[p\mathbf{E} - \dfrac{kT}{e}\operatorname{grad} p\right] = -\mu_p p \operatorname{grad}\left[V + \dfrac{kT}{e}\log p\right] \\[2ex]
\mathbf{J}_n \equiv -\dfrac{1}{e}\mathbf{I}_n = \mu_n\left[-n\mathbf{E} - \dfrac{kT}{e}\operatorname{grad} n\right] \\[2ex]
\qquad\qquad = -\mu_n n \operatorname{grad}\left[-V + \dfrac{kT}{e}\log n\right] \\[2ex]
\operatorname{div} \mathbf{E} = \dfrac{4\pi e}{\epsilon}\left[(p - p_0) - (n - n_0) + (D^+ - D_0^+) - (A^- - A_0^-)\right] \\[2ex]
\mathbf{E} = -\operatorname{grad} V.
\end{cases}
$$

In the first two equations, which are the continuity equations for holes and electrons written for a region free from external sources, g_0 is a constant which represents the thermal rate of generation of hole-electron pairs per unit volume; for cases in which hole-electron pairs are produced also by penetrating radiation, appropriate source terms in the form of identical functions of the space and time coordinates can be included on the right in the respective equations. The mean lifetimes of holes and electrons, τ_p and τ_n, are in general considered to be concentration-dependent and, since trapping is neglected, the quantities p/τ_p and n/τ_n are equal, being the rate at which holes and electrons recombine. Evaluated for the normal semiconductor, or the semiconductor at thermal equilibrium with no injected carriers, they equal g_0.

The equations for $\mathbf{J}_p$ and $\mathbf{J}_n$, which are vectors whose magnitudes equal, respectively, the numbers of holes and of electrons which traverse unit area in unit time, are diffusion equations of M. von Smoluchowski, written for hole flow and for electron flow[14]. Of the type frequently employed, after C. Wagner, in theories of rectification, each expresses the dependence of the flow density on the electrostatic field and on the concentration gradient, the diffusion constant for holes or electrons having been expressed in terms of the mobility, μ_p or μ_n, in accordance with the

[14] S. Chandrasekhar, *Rev. Mod. Phys.* 15, 1–89 (1943).

well-known relationship of A. Einstein[15]. In them, e denotes the magnitude of the electronic charge; T is temperature in degrees absolute; and k is Boltzmann's constant. With transport velocity defined as flow density divided by concentration, the product of the mobility and the quantity in square brackets in the expression for $\mathbf{J}_p$ or $\mathbf{J}_n$ on the extreme right gives the corresponding velocity potential, which is thus proportional to the sum of an electrostatic potential and a diffusion potential.

The next to last equation is Poisson's equation, which relates the divergence of the field to the net electrostatic charge. Here ϵ is the dielectric constant; p_0 and n_0 are the concentrations of holes and electrons at thermal equilibrium, in the normal semiconductor. The concentrations of ionized donor and acceptor impurities at thermal equilibrium are represented by D_0^+ and A_0^-, while D^+ and A^- are dependent variables which denote the respective concentrations in general of ionized donors and acceptors in the semiconductor with added carriers. As shown in the Appendix, variations in D^+ and A^- may be neglected if the impurity centers are substantially all ionized in the normal semiconductor, despite the effect large concentrations of added carriers may have on the equilibria[16].

The expression of the electrostatic field as the gradient of a potential according to the last equation is consistent with the circumstance that the effects of magnetic fields, with none applied, are in general quite negligible.

Subtracting the first continuity equation from the second, it is found that

$$\text{(2)} \qquad \operatorname{div} (\mathbf{J}_p - \mathbf{J}_n) = -\frac{\partial}{\partial t}(p - n),$$

since, with no trapping, p/τ_p equals n/τ_n. Neglecting changes in the concentrations of ionized donors and acceptors, this equation and Poisson's equation give

$$\text{(3)} \qquad \mathbf{J}_p - \mathbf{J}_n = \mathbf{J} - \frac{\epsilon}{4\pi e}\frac{\partial \mathbf{E}}{\partial t}; \qquad \mathbf{I}_p + \mathbf{I}_n = \mathbf{I} - \frac{\epsilon}{4\pi}\frac{\partial \mathbf{E}}{\partial t},$$

where $\mathbf{J}$ and $\mathbf{I}$ are solenoidal vector point functions, in general time-dependent. The latter is the total current density, and the term which follows it in (3) gives the displacement current density.

[15] A. Einstein, *Annalen der Physik* 17, 549–560 (1905); Müller-Pouillet, *Lehrbuch der Physik*, Braunschweig, 1933, IV (3), 316–319.

[16] It has been found from measurements of the temperature dependence of the conductivity and Hall coefficient that the energy of thermal ionization of the donors in n-type germanium of relatively high purity is only about $10^{-2}eV$, whence most of the donors are ionized at room temperature: G. L. Pearson and W. Shockley, *Phys. Rev.* 71 (2), 142 (1947).

It may be well to point out that the validity of the diffusion equations depends on two assumptions, which, while hardly restrictive in general for homogeneous semiconductors, indicate the nature of the generalizations which might otherwise be necessary[14]. The first assumption is that there are no appreciable time changes in the dependent variables in the relaxation time for the conductivity, or the time of the elementary fluctuations. This is tantamount to the requirement that the carriers undergo many collisions in the time intervals of interest. The second assumption is that the changes in the carriers' electrostatic potential energy over distances equal to the mean free path are small compared with the average thermal energy. In accordance with this assumption, very large fields in the electrically neutral semiconductor for which the carriers are not substantially in thermal equilibrium with the lattice are ruled out. The neglect of space charge then in general validates the two assumptions, if the resistivity is not too small, since the neglect of changes in the dependent variables which occur in the dielectric relaxation time obviates their change in the relaxation time for conductivity; and the neglect in the steady state of appreciable variations in electrostatic potential, and thus in the other dependent variables, in the distance[10] L_d, obviates their variation in a mean free path. The dielectric relaxation time for germanium, $1.5 \cdot 10^{-12}$ sec per ohm cm of resistivity, in high back voltage material exceeds the relaxation time for conductivity, which is about $1.0 \cdot 10^{-12}$ sec; and in semi-conductors in which the mobilities and the conductivity are smaller than the comparatively large values for germanium, the dielectric relaxation time may be appreciably larger than the relaxation time for conductivity. Similarly, L_d for germanium is about 7 times the mean free path, and this ratio, which is essentially inversely proportional to the square root of the product of mobility and intrinsic conductivity, may be appreciably larger for other semiconductors.

If, on the other hand, it should be desired to consider space-charge effects in germanium, the diffusion equations may be of rather marginal applicability, and the use of their appropriate generalization indicated, since with L_d equal to 7 mean free paths, appreciable space-charge variation of potential, corresponding to a field which is not small compared with the free-path thermal-energy equivalent of about 3500 volt cm^{-1}, may occur in at least one of the free paths. For example, diode theory, rather than diffusion theory, provides the better approximation for the characteristics of germanium point-contact rectifiers, and is particularly applicable to those from low resistivity material for which the potential variation is largely confined to one mean free path or less[17].

[10] loc. cit.
[14] loc. cit.
[17] H. C. Torrey and C. A. Whitmer, "Crystal Rectifiers," New York, 1948, Sec. 4.3.

Neglecting space charge, Poisson's equation becomes simply the condition of electrical neutrality:

$$(p - p_0) - (n - n_0) = 0, \tag{4}$$

assuming substantially complete ionization of donors and acceptors. Similarly, equations (3) become

$$\mathbf{J}_p - \mathbf{J}_n = \mathbf{J}; \qquad \mathbf{I}_p + \mathbf{I}_n = \mathbf{I}. \tag{5}$$

With electrical neutrality, the two continuity equations merge into one: Since derivatives of p equal the corresponding ones of n,

$$\begin{aligned} \operatorname{div} \mathbf{J}_p &= -[p/\tau_p - g_0] - \frac{\partial p}{\partial t} \\ &= \operatorname{div} \mathbf{J}_n = -[n/\tau_n - g_0] - \frac{\partial n}{\partial t}. \end{aligned} \tag{6}$$

The neutrality condition in conjunction with the two equations obtained by substituting for $\mathbf{J}_p$ and $\mathbf{J}_n$ from the diffusion equations in (6) thus provide three equations for the determination of p, n, and $\mathbf{E}$ or V.

It is instructive to rewrite equations (6) in accordance with

$$\begin{cases} \operatorname{div} \mathbf{J}_p = \mathbf{s} \cdot \operatorname{grad} p \\ = \operatorname{div} \mathbf{J}_n = \mathbf{s} \cdot \operatorname{grad} n, \\ \mathbf{s} \equiv \left[\dfrac{\partial \mathbf{J}_p \cdot \mathbf{i}}{\partial x} \Big/ \dfrac{\partial p}{\partial x}\right]\mathbf{i} + \left[\dfrac{\partial \mathbf{J}_p \cdot \mathbf{j}}{\partial y} \Big/ \dfrac{\partial p}{\partial y}\right]\mathbf{j} + \left[\dfrac{\partial \mathbf{J}_p \cdot \mathbf{k}}{\partial z} \Big/ \dfrac{\partial p}{\partial z}\right]\mathbf{k}, \end{cases} \tag{7}$$

where $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are unit vectors in the directions of the respective axes. The velocity $\mathbf{s}$, which is given as well by the expression for electrons analogous to that written for holes, may be defined alternatively as follows: Suppose, for definiteness, that the second-order system of equations (4) and (6) have been solved, so that the concentrations and flow densities are known in terms of the cartesian coordinates x, y, and z, and the time t. The x-component of $\mathbf{s}$ is then the partial derivative with respect to p of the x-component of $\mathbf{J}_p$ in which x has been replaced by the proper function of p, y, z, and t, and similarly for the other components. Thus, with $\mathbf{s}$ a known function, p or n may be considered to satisfy the first-order partial differential equation obtained by substituting from (7) in (6), from which it is evident that $\mathbf{s}$ is the velocity with which concentration transients are propagated[18]. This velocity, which is here called the differential

[18] The identification of $\mathbf{s}$ as this propagation velocity follows the example of C. Herring, in whose method for solving the transient constant-current problem in one dimension the velocity depends in a known manner on concentration only, through the neglect of diffusion, so that the general solution of the differential equation in which thus neither independent variable x nor t occurs explicitly may be obtained; cf. reference 12, pp. 412 ff.

transport velocity and loosely referred to as the transport velocity of added carriers, of course differs in general from the transport velocity proper, defined as the ratio of flow density to concentration; its general definition, which is applicable to the steady state, has been introduced to facilitate later interpretations.

2.3 *Reduction of the fundamental equations to dimensionless form*

2.31 *The general case*

In order to obtain solutions in forms which exhibit such generality as they may possess, the fundamental equations are to advantage written in terms of dimensionless dependent and independent variables which are the original variables measured in suitable units. Through formal consideration of the equations (1), in conjunction with (3) or with (4) and (6), these units can be so chosen that the system of reduced equations will exhibit independent parameters on which it may be considered to depend. The best choice of suitable units is by no means unique; those choices which have been made are natural ones, in that they have been found to result in greater formal simplicity and ease of interpretation in the theory than others which may be equally valid in principle.

The choice for an n-type semiconductor consists in definitions of dimensionless variables and parameters as follows:

$$
\left[\begin{array}{l}
X \equiv x/L_p,\ Y \equiv y/L_p,\ Z \equiv z/L_p\ ;\ L_p \equiv \left[\dfrac{kT\mu\tau}{e}\right]^{\frac{1}{2}} = [D_p\tau]^{\frac{1}{2}} \\
U \equiv t/\tau \\
P \equiv p/(n_0 - p_0);\ P_0 \equiv p_0/(n_0 - p_0) = g_0\tau/(n_0 - p_0) \\
N \equiv n/(n_0 - p_0);\ N_0 \equiv n_0/(n_0 - p_0) \\
\mathbf{C} \equiv \mathbf{I}/I_0 = \mathbf{E}_a/E_0 = \mu\mathbf{E}_a/[D_p/\tau]^{\frac{1}{2}};\ I_0 \equiv \sigma_0 E_0\ ;\ E_0 \equiv kT/eL_p \\
\mathbf{C}_p \equiv \mathbf{I}_p/I_0 \\
\mathbf{C}_n \equiv -\mathbf{I}_n/I_0 \\
\mathbf{F} \equiv \mathbf{E}/E_0 \\
W \equiv V/E_0L_p = eV/kT \\
Q \equiv \tau/\tau_p\ .
\end{array}\right. \tag{8}
$$

The rectangular cartesian space coordinates are x, y, and z. The quantity τ is the mean lifetime of holes for concentrations of added holes small compared with the thermal-equilibrium electron concentration, n_0 ; and σ_0 is the conductivity of the normal semiconductor. The hole mobility,

originally μ_p, is denoted by μ for simplicity. If b is the ratio of electron to hole mobility, σ_0 is given in general by

$$(9) \quad \sigma_0 = \mu e(bn_0 + p_0) = M_0\, b\mu e(n_0 - p_0), \qquad M_0 \equiv 1 + \frac{b+1}{b} P_0,$$

the symbol M_0 being introduced for brevity. If $p_0 << n_0$, M_0 is unity and σ_0 equals $b\mu e n_0$.

The independent dimensionless distance variables are X, Y and Z, where the distance unit, L_p, is a diffusion length for a hole for the mean lifetime, τ, the diffusion constant for holes being D_p. This mean lifetime is the unit for the independent dimensionless time variable, U. The hole and electron concentrations are measured in units of the excess in concentration of electrons over holes[19], $n_0 - p_0$, the reduced variables being P and N, respectively. The reduced total current $\mathbf{C}$ is total current density measured in units of the current density I_0 which flows in the semiconductor with no added carriers under the characteristic field E_0, which is a field such that a carrier would expend the energy kT in drifting with it through the distance L_p. A more illuminating alternative description is that $\mathbf{C}$ is the ratio of the average drift velocity of holes under the applied or asymptotic field, $\mathbf{E}_a$, to the hole diffusion velocity $(D_p/\tau)^{\frac{1}{2}}$. The field $\mathbf{E}_a$ is that which produces the current density $\mathbf{I}$ in the semiconductor with no added carriers. The corresponding reduced hole and electron flow densities are $\mathbf{C}_p$ and $\mathbf{C}_n$. The electrostatic field measured in units of E_0 is denoted by $\mathbf{F}$, and W is the corresponding reduced electrostatic potential. The lifetime ratio Q is a function of P which characterizes the recombination process. While it appears from experiment that the recombination rate for holes depends on both physical and chemical properties of the semiconductor, in a particular semiconductor at given temperature it may be considered to depend on hole concentration alone.

Representative values for germanium of units in terms of which the dimensionless quantities are defined are as follows: The mean lifetime τ may be of the order of 10^{-5} sec. With a mobility for holes[7] of 1700 cm^2 $volt^{-1}$ sec^{-1} in germanium single crystals at 300 deg abs, the length L_p is then about $2 \cdot 10^{-2}$ cm; the characteristic field, E_0, 1.2 volt cm^{-1}; and the current density I_0, 0.12 ampere cm^{-2} for a resistivity of 10 ohm cm.

With these definitions[20], the fundamental equations for a region free from external sources, neglecting changes in the concentrations of ionized

[7] loc. cit.

[19] The excess in concentration of electrons over holes is of course equal to that of ionized donors over ionized acceptors.

[20] The definitions given appear best if there is a region in which $P - P_0$ is small, with $P_0 \neq 0$. Modified definitions of the reduced flow densities, in which the conductivity σ_0 is replaced by the conductivity $b\mu e(n_0 - p_0)$ due to the excess electrons alone, result in equations obtainable formally by setting M_0 equal to unity.

donors and acceptors and neglecting space charge[21], are given in reduced form as follows:

$$
(10)\quad \begin{cases} \dfrac{\partial P}{\partial U} = -[bM_0 \operatorname{div} \mathbf{C}_p + PQ - P_0] \\ \dfrac{\partial N}{\partial U} = -[bM_0 \operatorname{div} \mathbf{C}_n + PQ - P_0] \\ \mathbf{C}_p = \dfrac{1}{bM_0}[\mathbf{F}P - \operatorname{grad} P] = -\dfrac{1}{bM_0} P \operatorname{grad} [W + \log P] \\ \mathbf{C}_n = \dfrac{1}{M_0}[-\mathbf{F}N - \operatorname{grad} N] = -\dfrac{1}{M_0} N \operatorname{grad} [-W + \log N] \\ (P - P_0) - (N - N_0) = P - N + 1 = 0 \\ \mathbf{F} = -\operatorname{grad} W, \end{cases}
$$

and the reduced form of equations (5) is

$$
(11)\qquad \mathbf{C}_p - \mathbf{C}_n = \mathbf{C}.
$$

These reduced equations may be simplified and two differential equations in the dependent variables P and W written as follows:

$$
(12)\quad \begin{cases} -b\, M_0 \operatorname{div} \mathbf{C}_p = \operatorname{div} P \operatorname{grad} [W + \log P] = [PQ - P_0] + \dfrac{\partial P}{\partial U} \\ \operatorname{div} \mathbf{C} = 0, \qquad \mathbf{C} = -\Sigma \operatorname{grad} \left[W - \dfrac{b-1}{b+1} \log \Sigma \right], \end{cases}
$$

where Σ is the conductivity σ in reduced form:

$$
(13)\qquad \Sigma \equiv \frac{\sigma}{\sigma_0} = \frac{bN + P}{bN_0 + P_0} = \frac{1}{M_0}\left[1 + \frac{b+1}{b} P\right].
$$

An alternative formulation, due to R. C. Prim, which is obtained by evaluating div $[\mathbf{C}_n \pm b\, \mathbf{C}_p]$, consists of the two equations,

$$
(14)\quad \begin{aligned} \frac{b}{b-1} \operatorname{div} (1 + 2P) \operatorname{grad} W &= -\frac{b}{b+1} \operatorname{div} \operatorname{grad} [W - (1 + 2P)] \\ &= [PQ - P_0] + \frac{\partial P}{\partial U}, \end{aligned}
$$

[21] It may be desirable to take space charge into account in cases involving high frequencies or high resistivities. Poisson's equation and equations (3) are in reduced form,

$$P - N + 1 = bM_0\Gamma \operatorname{div} \mathbf{F} \text{ and } \mathbf{C}_p - \mathbf{C}_n = \mathbf{C} - \Gamma \frac{\partial \mathbf{F}}{\partial U}, \text{ where } \Gamma \equiv \epsilon/4\pi\sigma_0\tau.$$

The term containing Γ may often be omitted from one of these equations, depending on the nature of the particular case considered.

in which the use of $(1 + 2P)$ as dependent variable may be desirable. This variable is equal to the concentration of carriers of both kinds divided by the excess of electron concentration over hole concentration, which is a constant.

The expression in the equations which specifies the recombination rate may be written more simply. Since the lifetime ratio Q is unity for $P = P_0$,

$$PQ - P_0 = (P - P_0)R, \tag{15}$$

where R, which will be called the recombination function, depends on P and also equals unity for $P = P_0$. The lifetime ratio and the recombination function which, of course, differ in general, both equal unity for the case of constant mean lifetime. Recombination of holes and electrons at a rate proportional to the product of their concentrations, called mass-action recombination, and recombination characterized by a constant mean lifetime for holes are frequently of interest. For a combination of independent mechanisms of both types, it is easily seen that

$$\begin{cases} Q \equiv \tau/\tau_p = 1 + a\,(p - p_0)/n_0 = 1 + a\,(P - P_0)/(1 + P_0), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad a \equiv \tau/\tau_v,\ 0 \leq a \leq 1 \\ R = 1 + ap/n_0 = 1 + aP/(1 + P_0), \end{cases} \tag{16}$$

where τ_v is the mean lifetime for small concentrations associated with mass-action recombination alone, so that $a = 0$ for constant mean lifetime, and $a = 1$ for mass-action recombination. If both recombination mechanisms are operative, that of mass-action recombination will, of course, determine the mean lifetime where the concentration of added carriers is sufficiently large.

Recent experiments have shown that the mean lifetime for holes in n-type germanium can be increased materially, to at least 100 microseconds, by minimizing surface recombination through decreases in surface-to-volume ratios.[1] On the other hand, comparatively short mean lifetimes, of the order of one microsecond, occur in p-type germanium produced, for example, from n-type by nucleon bombardment. It should be possible to determine in various cases which recombination law would provide the better approximation by use of the technique of H. Suhl and W. Shockley of hole injection in the presence of a magnetic field[22] or by the photoelectric technique of F. S. Goucher[23].

[1] loc. cit.

[22] H. Suhl and W. Shockley, *Phys. Rev.* 75 (10), 1617–1618; 76 (1), 180 (1949).

[23] F. S. Goucher, paper I 11 of the Oak Ridge Meeting of the American Physical Society, March 18, 1950; *Phys. Rev.* 78 (6), 816 (1950).

It appears that solutions neglecting recombination furnish useful approximations for some applications. If recombination is neglected, by assuming that the mean lifetime is infinite, the definitions (8) of the dimensionless quantities no longer have meaning, but essentially the same differential equations and corresponding boundary-condition equations can still be used. The reduced equations become essentially homogeneous in τ for τ large, and it suffices to suppress the recombination terms, $PQ - P_0$, retaining formally the definitions of the dimensionless quantities in which now τ, and thus L_p and E_0 or I_0 no longer have physical significance. One of these unitary quantities may be chosen arbitrarily. It might be noted that if Poisson's equation is retained the length unit is advantageously chosen as L_d, which gives a dielectric relaxation time for the time unit.[21]

In one cartesian dimension, with total current a function of time only, W may be eliminated by means of the equation for $\mathbf{C}$ in (12) and, upon substituting for it in any of the three remaining equations in (12) and (14), a differential equation for P results which depends on b, P_0, and $\mathbf{C}$ as parameters. Dropping vector notation, this equation is

$$\frac{\partial P}{\partial U} = \frac{\left[(1+2P)\left(1+\frac{b+1}{b}P\right)\right]\frac{\partial^2 P}{\partial X^2} + \frac{b-1}{b}\left[\frac{\partial P}{\partial X}\right]^2 - M_0 C\frac{\partial P}{\partial X}}{\left[1+\frac{b+1}{b}P\right]^2} - (P-P_0)R. \tag{17}$$

Similarly, from (10),

$$\begin{cases} C_p = \dfrac{M_0 CP - (1+2P)\dfrac{\partial P}{\partial X}}{bM_0\left[1+\dfrac{b+1}{b}P\right]} \\ F = \dfrac{M_0 C - \dfrac{b-1}{b}\dfrac{\partial P}{\partial X}}{1+\dfrac{b+1}{b}P}. \end{cases} \tag{18}$$

The expressions for F and C_p possess some interesting features. That for the reduced field, F, is composed of two terms, the first of which expresses Ohm's law, since C is reduced total current density and the denominator is proportional to the local conductivity. The second term is a contribution which is directed away from a hole source, since b is greater than

unity, or since electrons are more mobile than holes. If b were equal to unity, the field would be independent of the concentration gradient. The second term thus represents a departure from Ohm's law which is due to diffusion and which is associated with the presence of current carriers of differing mobilities. It gives a non-vanishing electrostatic field for the case of zero total current. The two terms in the expression for C_p are likewise ohmic and diffusion terms, but here the diffusion term would be present even if the hole and electron mobilities were equal.

Boundary-condition relationships might be illustrated by some examples for this one-dimensional case. If it be specified that for $U > 0$ a fraction f of the total current to the right of a source at the X-origin, say, be carried by holes, then, from (18),

$$\frac{\partial P}{\partial X} = -M_0 \frac{b + (b+1)P}{1 + 2P}\left[f - \frac{P}{b + (b+1)P}\right]C, \qquad X = +0, \quad U > 0. \tag{19}$$

The solution in an X-region to the right of the origin may be determined by this condition and an additional one. The simplest is that for the flow in the semi-infinite region, namely $P = P_0$ for $X = \infty$. This relationship holds for some finite X for an idealized non-rectifying electrode there. For the region between the source and a surface at $X = X_a$ on which there is recombination characterized by a hole transport velocity s, which is also the differential transport velocity for s constant, it is clear that $C = 0$, so that, for $X = X_a$,

$$C_p = -\frac{1}{M_0}\frac{(1+2P)}{b + (b+1)P}\frac{\partial P}{\partial X} = \frac{1}{bM_0}SP; \tag{20}$$

$$S \equiv s/[D_p/\tau]^{\frac{1}{2}}, \qquad s \equiv J_p/p.$$

Consistently with these examples, boundary conditions may in general be expressed as relationships between P, $\frac{\partial P}{\partial X}$, and the parameter C, for given values of X.

A simple transformation of dimensionless quantities serves to extend all of the analytical results which have been given for the n-type semiconductor to the p-type semiconductor: Consider the substitutional transformation which consists in replacing the original dimensional quantities for holes by the corresponding ones for electrons, and vice versa, and in replacing the electrostatic field by its negative. The original set of fundamental equations (1) is invariant under this substitution, which defines an equivalent transformation from the dimensionless quantities of

equations (8) to the desired new set, in which the ratio b of electron to hole mobility is replaced by its reciprocal.

2.32 *The intrinsic semiconductor.*

For the intrinsic semiconductor, in which $p_0 = n_0$, the reduced concentrations given in (8) are inapplicable. As p_0 approaches n_0, these reduced concentrations increase indefinitely, and the equations which those given for the n-type semiconductor approach in the limit are homogeneous in the concentration unit. These limiting equations therefore apply to the intrinsic semiconductor in terms of a concentration unit which may be chosen arbitrarily. The quantity n_0 will be chosen as this unit. Thus, redefining the reduced concentration variables as

$$P \equiv p/n_0, \qquad N \equiv n/n_0; \qquad P = N, \tag{21}$$

from equations (12) and (14) any two of the equations in the dependent variables P and W given by

$$\left\{\begin{aligned} -(b+1)\ \mathrm{div}\ \mathbf{C}_p &= \frac{2b}{b-1}\ \mathrm{div}\ P\ \mathrm{grad}\ W \\ &= \frac{2b}{b+1}\ \mathrm{div\ grad}\ P = [PQ - 1] + \frac{\partial P}{\partial U}, \\ \mathbf{C}_p &= -\frac{1}{b+1} P\ \mathrm{grad}\ [W + \log P]; \\ \mathrm{div}\ \mathbf{C} &= 0, \qquad \mathbf{C} = -P\ \mathrm{grad}\left[W - \frac{b-1}{b+1}\log P\right], \end{aligned}\right. \tag{22}$$

and including the right-hand member which is common at least once, characterize the intrinsic semiconductor[24].

It is noteworthy that one of these equations contains only P as dependent variable, W being absent; and this equation indicates that the spatial distribution of carrier concentration is not subject to drift under the field, but only to a diffusion mechanism with diffusion constant $2D_pD_n/(D_p + D_n)$, where $D_n = bD_p$ is the diffusion constant for electrons.[25] This result is readily accounted for as being due to a conductivity in the intrinsic case which is everywhere proportional to the concentration of carriers of either type, so that $\Sigma = P$. The expression for $\mathbf{C}$

[24] These equations for the intrinsic case were first derived quite unambiguously as those for the special case of the parameter p_o/n_o equal to unity in the general equations written in terms of the concentration unit n_o. This unit is, however, less advantageous than $(n_o - p_o)$ which, in obviating much of the formal dependence on p_o, makes for greater generality.

[25] The equations for the intrinsic case might be written in somewhat simpler form by redefining the length unit in terms of $2D_pD_n/(D_p + D_n)$ as a diffusion constant instead of D_p, but their relationship to those of the general case would then be less evident.

in (22) owes its special form simply to this circumstance, while that for $\mathbf{C}_p$ applies also to the general case, and the differential equation in P is a consequence of the equations in P and W from div $\mathbf{C}$ and div $\mathbf{C}_p$. Or, in more detailed terms, since the ohmic contribution to $\mathbf{C}_p$ must be proportional to $\mathbf{C}$, div $\mathbf{C}_p$ contains only the contribution due to diffusion. This is evident from the relationship obtained from (22),

$$\mathbf{C}_p = \frac{1}{b+1}\left[\mathbf{C} - \frac{2b}{b+1}\,\text{grad}\,P\right], \tag{23}$$

from which it follows also that, despite the dependence of the local field on concentration gradient, the ohmic contribution to the hole flow density is the flow density of holes normally present in the intrinsic semiconductor under the unmodulated applied field.

The equations which have been given for one-dimensional flow in the n-type semiconductor can readily be transformed, in the manner indicated, into the corresponding equations for the intrinsic semiconductor.

2.4 *Differential equations in one dimension for the steady state of constant current and properties of their solutions*

The steady state of constant current in one dimension will be considered explicitly for two limiting cases: the n-type semiconductor with $P_0 = 0$, and the intrinsic semiconductor. These serve to illustrate and delimit the qualitative features of the general case. Furthermore, the case $P_0 = 0$ frequently applies as a good approximation[26], as does the intrinsic case, which is of particular interest not only in itself but also because the extrinsic semiconductor exhibits intrinsic behavior for large concentrations, and because moderate increases in temperature above room temperature, such as joule heating may produce, suffice to bring high back voltage germanium into the intrinsic range of conductivity[27]. The temperature dependence of P_0 and of other reduced quantities is evaluated for germanium in the Appendix.

The ordinary differential equations in the reduced hole concentration, P, for the steady state in one dimension, which result from equations (17) and (22) by equating the time derivatives to zero are as follows:

$$\frac{d^2P}{dX^2} = \frac{C\dfrac{dP}{dX} - \dfrac{b-1}{b}\left[\dfrac{dP}{dX}\right]^2}{[1+2P]\left[1+\dfrac{b+1}{b}P\right]} + \frac{P\left[1+\dfrac{b+1}{b}P\right]}{1+2P}R \tag{24}$$

[26] In n-type germanium of resistivity about 5 ohm cm, for example, the electron concentration exceeds the equilibrium hole concentration by a factor of about 70.

[27] Germanium which is substantially intrinsic at room temperature has been produced: R. N. Hall, paper I5 of the Oak Ridge Meeting of the American Physical Society, March 18, 1950.

for the n-type semiconductor with $P_0 = 0$, and

$$\frac{d^2P}{dX^2} = \frac{b+1}{2b}(P-1)R \tag{25}$$

for the intrinsic semiconductor, with R given as $(1 + aP)$ by (16); P has the same meaning in both equations, the concentration unit being n_0 for each case. With time variations excluded in this way, the parameter C is a constant and the differential equations apply to the steady state of constant current.

Since the equations involve only the single independent variable X which does not appear explicitly, their orders may be reduced by one, in accordance with a well-known transformation, which consists in introducing P as a new independent variable, and

$$G \equiv \frac{dP}{dX} \tag{26}$$

as new dependent variable: Noting that $\frac{d}{dX}$ is equivalent to $G\frac{d}{dP}$, the differential equations become

$$\frac{dG}{dP} = \frac{C - \frac{b-1}{b}G}{[1+2P]\left[1+\frac{b+1}{b}P\right]} + \frac{P\left[1+\frac{b+1}{b}P\right]R}{[1+2P]G} \tag{27}$$

for the n-type semiconductor, and

$$\frac{dG}{dP} = \frac{b+1}{2b}\frac{(P-1)R}{G} \tag{28}$$

for the intrinsic semiconductor. These are differential equations of the first order.

The solutions sought in the semi-infinite region, $X > 0$, are those for which $G = 0$ for $\Delta P = 0$, that is, those which pass through the $(\Delta P, G)$ — origin, where ΔP, which denotes $P - P_0$, equals P for the n-type semiconductor and $P-1$ for the intrinsic semiconductor. This condition is that the concentration gradient vanish with the concentration of added holes, as it must for X infinite. It will be shown that the differential equations possess singular points at the $(\Delta P, G)$-origin, and the physical interpretation of the solutions through these singular points will be examined. For this purpose, consider equation (27) for the n-type semiconductor which, in the neighborhood of the origin, assumes the approximate form,

$$\frac{dG}{dP} = \frac{G}{P} = C + \frac{P}{G}, \tag{29}$$

since R is close to unity for P small, whence

$$\frac{dG}{dP} = \frac{G}{P} = \frac{1}{2}[C \pm \sqrt{C^2 + 4}]. \tag{30}$$

Similarly, for the intrinsic semiconductor, for $P-1$ small,

$$\frac{dG}{d(P-1)} = \frac{G}{P-1} = \pm\sqrt{\frac{b+1}{2b}}. \tag{31}$$

There are thus, in each case, two solutions through the $(\Delta P, G)$-origin, one with a positive derivative and the other with a negative derivative. Consider now the doubly-infinite region with a source at $X = 0$. Then, for $X > 0$, the negative derivatives apply, since the concentration gradient G is negative. Similarly, for $X < 0$, the positive derivatives apply. Now, the value of the current parameter C will be substantially the same in both regions, since it has been assumed that ΔP is small. For C positive, equation (30) for the n-type semiconductor indicates that the magnitude of dG/dP for $X < 0$ exceeds that for $X > 0$, and the situation is reversed if the sign of C is changed. That is, the magnitude of the concentration gradient increases more slowly with concentration for field directed away from a source than for field directed towards a source, which is otherwise plausible. For the intrinsic semiconductor, on the other hand, equation (31) shows that corresponding magnitudes of the concentration gradient are equal and entirely independent of C, a result which the differential equation (28) establishes in general.

It thus appears that a differential equation for the steady state possesses two solutions through the $(\Delta P, G)$-origin, and that one of the solutions corresponds to the case of field directed towards a source, the other to the case of field directed away from a source. Field directed towards a source is called field opposing, while field directed away from a source is called field aiding, the latter being the one commonly dealt with in hole-injection experiments. It should be noted that the cases of field opposing or field aiding can be realized in a given X-region only if it adjoins a semi-infinite region free from sources and sinks. In the region between two sources, neither of these cases applies. L. A. MacColl has shown, through a more detailed consideration of the singularity at the $(\Delta P, G)$-origin, that the two solutions through this point are the only ones through it. The origin is thus a saddle-point of the differential equation, and there exist families of nonintersecting solutions in the $(\Delta P, G)$-plane for which the solutions which intersect at the origin are asymptotes. A solution for an X-region between two sources, for example, is a member of such a family, as is in general any solution determined by boundary conditions at the ends of a finite region in X. Such a solution will be called a solution for a composite case; it approaches asymptotically both a field-opposing and a

field-aiding solution, which is consistent with the qualitative geometry associated with a saddle-point, and with the fact that, in the X-region, a

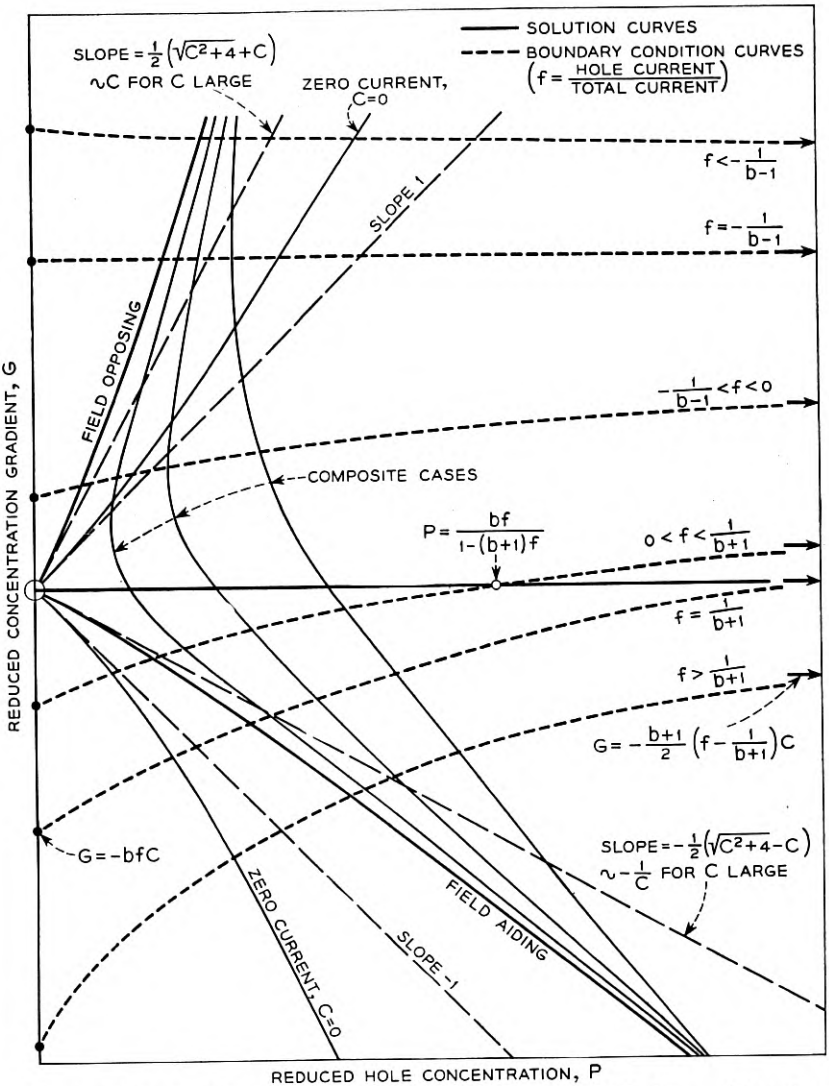


Fig. 1.—Diagrammatic representation in the (P, G)-plane of solutions and boundary conditions for the steady-state one-dimensional flow of holes in an n-type semiconductor.

total current directed away from one source is necessarily directed towards the other. This behavior is illustrated diagramatically for the n-type semiconductor in Fig. 1, which shows, in the (P, G)-plane, solution

curves as well as boundary-condition curves for a source, for a given positive value of C. Those for the intrinsic semiconductor differ only in that the solution curves in the $(P-1, G)$-plane do not depend on C, all being given by the ones for zero total current density, and the corresponding boundary-condition curves are straight lines.

Once a solution, $G(P)$, for field opposing, field aiding, or a composite case, specifying G as a function of P has been obtained, the dependence of P on X is determined by evaluating

$$X = \int_{P^0}^{P} \frac{dP}{G(P)} \tag{32}$$

in accordance with the definition of G, equation (26). For the general composite case, $G(P)$ is that one of the family of solutions for the given C such that the integral between values of P determined by the intersections with the boundary-condition relationships provides the correct interval in X. If P^0 is determined by the condition that for $X = 0$, a fraction f of the total current is carried by holes, then, from (19), P^0 is the point on the solution curve which satisfies either

$$G^0 = -\frac{b + (b+1)P^0}{1 + 2P^0}\left[f - \frac{P^0}{b + (b+1)P^0}\right]C \tag{33}$$

for the n-type semiconductor, or

$$G^0 = -\frac{(b+1)^2}{2b}\left[f - \frac{1}{b+1}\right]C \tag{34}$$

for the intrinsic semiconductor, G^0 being the corresponding value of G.

From the manner of derivation of the boundary conditions (33) and (34), it is evident that they are perfectly general, holding in particular for the cases of field opposing and field aiding, and whatever be the sign of C. The concentration gradient G^0 may be seen to have the correct sign for these cases if it is taken into account that f, defined as C_p/C or I_p/I, may assume any positive or negative value, being positive for field aiding, and negative for field opposing, for which the hole flow is opposite to the applied field. For f negative, the quantities in brackets in equations (33) and (34) are negative. The general principle that the sign of the concentration gradient G is such as to be consistent with the flow of holes from a source requires also that the quantities in brackets be positive for field aiding, or whenever f is positive. For the intrinsic semiconductor, this requires that f for field aiding never be less than $1/(b+1)$. This is clearly a consistent requirement which holds in all generality since, for zero concentration of added holes, or for the normal semiconductor, G^0 vanishes and the ratio of hole current to total current equals $1/(b+1)$.

In the case of the *n*-type semiconductor, f is not restricted in this way. Consider, for this case, hole injection into the end of a semi-infinite filament, to which the field-aiding solutions apply. As the total current is increased indefinitely, the tangent to the solution in the (P, G)-plane at the origin approaches the P-axis, as does the solution itself, and it is evident from the boundary-condition curves of Fig. 1 that if f is less than $1/(b + 1)$ the hole concentration P^0 at the source approaches as a limit the indicated abscissa of intersection of the appropriate boundary-condition curve with the P-axis, or the value for which the quantity in brackets vanishes. It is similarly evident that P^0 increases indefinitely with total current in either semiconductor if f is greater than or equal to $1/(b + 1)$. This is a result otherwise to be expected from the qualitative consideration that an extrinsic semiconductor becomes increasingly intrinsic in its behavior as the concentration of injected carriers is increased.

Figure 1 serves also to facilitate a count of the number of degrees of freedom which the steady-state solutions possess: Corresponding to values of the concentration and concentration gradient at a point in a semiconductor filament in which added carriers flow, there is a point (P, G) in the half-plane, $P > 0$, of the figure. If the total current density is specified in addition, the value of C and the solution through the point (P, G) are determined. This solution applies in general to a composite case, which therefore possesses three degrees of freedom. That is to say, at a point in a filament, any given magnitudes of both concentration and concentration gradient can be realized for a preassigned total current density by a suitable disposition of sources to the right and left. The cases of field opposing or field aiding, however, possess only two degrees of freedom, since the given concentration and gradient determine the total current density and the solution, which must pass through the origin; and which of the two cases applies depends on whether the point (P, G) lies to the left or to the right of the curves, shown in the figure, for the zero-current solution. Thus, in a filament with a single source of holes, for example, the concentration, concentration gradient, total current density, and any functions of these, such as hole flow density and electrostatic field, are all quantities the specification of any two of which at a point completely determines the solution for a source-free X-region which includes the point.

3. Solutions for the Steady State

For a given value of the current parameter C, solutions for the steady state of constant current in a single cartesian distance coordinate, specifying G in terms of the relative hole concentration P, and P, the reduced hole

flow density C_p, and the reduced electrostatic field F, in terms of reduced distance X are found in general by numerical means, which include numerical integration and the evaluation of appropriate series expansions.

General solutions which have been evaluated numerically for n-type germanium for a number of values of the current parameter are given in the figures. In the limiting cases of P small and P large, analytical approximations for the extrinsic semiconductor are readily obtained, that for P large being derived from an analytical solution for C equal to zero, or zero current. If the steady-state problem for the extrinsic semiconductor is simplified by neglecting either recombination or diffusion, solutions are obtainable which, like the zero-current one, are expressible in closed form.

For the intrinsic semiconductor, the general problem considered in this section is solved quite simply by analytical means. The solution provides, as physical considerations indicate it should, the same analytical approximation for large P as does the zero-current solution for the extrinsic case. It may be well to consider first the intrinsic semiconductor which, aside from the extrinsic semiconductor for the case of zero current, appears to constitute the only analytically solvable steady-state case in one dimension which has physical generality according to the present approach.

3.1 *The intrinsic semiconductor*

Integrating the differential equation (28), it is found that

$$G^2 = \frac{b+1}{b} \int (P-1) R dP, \tag{35}$$

with R given as $1 + aP$ by (16), for an arbitrary combination of the two recombination mechanisms, assumed independent. Thus

$$G^2 = \frac{b+1}{2b} (P-1)^2 \left[(1+a) + \frac{2}{3} a(P-1) \right] \tag{36}$$

for the cases of field opposing or field aiding, for which $G = 0$ for $P-1 = 0$; for a composite case, a suitable constant is included on the right-hand side. Excluding composite cases, the root may be taken in (36) and G replaced by its definition, which gives

$$\frac{d(P-1)}{dX} = \pm \left[\frac{b+1}{2b} \right]^{\frac{1}{2}} [P-1] \left[(1+a) + \frac{2}{3} a(P-1) \right]^{\frac{1}{2}} \tag{37}$$

and if the X-origin is selected more or less arbitrarily as the point at which P is infinite, then (37) gives

$$P - 1 = \frac{3(1+a)}{2a} \operatorname{csch}^2 \left[\frac{(1+a)(b+1)}{8b} \right]^{\frac{1}{2}} X \tag{38}$$

provided $a \pm 0$; for mass-action recombination $a = 1$. For $a = 0$ or for constant mean lifetime, (37) gives an exponential dependence of $P-1$ on X:

$$P - 1 = (P^0 - 1) \exp\left[\pm\left[\frac{b+1}{2b}\right]^{\frac{1}{2}} X\right], \tag{39}$$

where P_0 is the relative hole concentration for $X = 0$. Linear combinations of the two solutions in (39) give solutions for composite cases, since the differential equation from which (39) was derived is linear in P. A similar result does not hold if there is mass-action recombination present, and the more general procedure above referred to must then be followed.

A characteristic feature of these solutions for the intrinsic semiconductor is their independence of the current parameter C, this parameter occurring only through a boundary condition, such as the one given in equation (34) of Section 2.4. They are symmetrical in shape about a source, the dependence of the concentration on the magnitude of the distance from the source being the same for field opposing as for field aiding, which follows quite simply from the symmetrical forms of the solutions, and the condition that the concentration is everywhere continuous.

Equations (22) and (23) of Section 2.32 provide the hole flow density and the electrostatic field for this case. With G given for mass-action recombination or for constant mean lifetime by the appropriate special case of equation (36), and using the positive sign for an X-region to the left of sources and the negative sign for an X-region to the right,

$$\begin{cases} C_p = \dfrac{1}{b+1}\left[C - \dfrac{2b}{b+1} G\right] \\ F = \dfrac{1}{P}\left[C - \dfrac{b-1}{b+1} G\right]. \end{cases} \tag{40}$$

The electrostatic potential, V, is readily expressed in terms of P: From

$$F = -\frac{eL_p}{kT}\frac{dV}{dx} = -\frac{e}{kT}\frac{dV}{dX} = -\frac{e}{kT} G \frac{dV}{dP} \tag{41}$$

and (40), it is found that

$$\frac{e}{kT}\frac{dV}{dP} = \frac{b-1}{b+1}\frac{1}{P} - C\frac{1}{GP}, \tag{42}$$

whence

$$\frac{eV}{kT} = \frac{b-1}{b+1}\log P - C\int\frac{dP}{GP} = \frac{b-1}{b+1}\log P - C\int\frac{dX}{P}, \tag{43}$$

with the integral to be evaluated for the particular case it is desired to consider.

3.2 *The extrinsic semiconductor: n-type germanium*

The evaluation of steady-state solutions for the extrinsic semiconductor involves, as a first step, the determination of G as a function of P from the differential equation (27), which is accomplished by numerical integration and by the use of series expansions. These variables are subsequently found in terms of X in the manner described in Section 2.4. The series expansions, which are Maclaurin's series in P, and series in powers of the current parameter, C, with coefficients functions of P, are given explicitly for the n-type semiconductor in the Appendix; they readily furnish the corresponding series for the p-type semiconductor by means of the transformation discussed at the end of Section 2.31. The Maclaurin's series in P are useful for starting the solutions at the (P, G)-origin. As P increases, these series converge increasingly slowly, and it becomes necessary to extend the solutions by other means. For the larger values of C, however, the numerical integration for the important case of field aiding becomes increasingly difficult, and it is advantageous to use the appropriate series in the current parameter, which converges the more rapidly the larger is C. The first term alone in this series for field aiding gives in closed form the solution for the case in which diffusion is neglected; and the existence of the series itself was, in fact, originally suggested by the form of the solution for this case[28]. Series of this type are given also for field opposing, and it seems probable that such series are obtainable for composite cases as well, though this has not been investigated.

Solutions were evaluated numerically for n-type germanium, by the means described, using the value 1.5 for the mobility ratio[29], b. For the case of mass-action recombination, solutions for values of the current parameter, C, up to 50, specifying $| G |$ in terms of P, are given in Fig. 2, both for field opposing and field aiding. These solutions in the (P, G)-plane are given to permit the fitting of boundary conditions at a hole source, according to a method described in Section 4. Solutions specifying P in terms of X for field aiding are given in Fig. 3, with the X-origin chosen more or less arbitrarily at $P = 100$. The corresponding solutions for the reduced hole flow density, C_p, and the reduced field, F, are given

[28] The solution for this case was communicated by Conyers Herring and is given in his paper of reference 12.

[29] The hole mobility and the value 1.5 for the mobility ratio were determined by G. L. Pearson from the temperature dependence of the conductivity and Hall coefficient in p-type germanium. J. R. Haynes has recently obtained, from drift-velocity measurements, the same hole mobility, but the larger value 2.1 for the ratio of electron mobility in n-type germanium to hole mobility in p-type: Paper L2 of the Chicago Meeting of the American Physical Society, November 26, 1949.

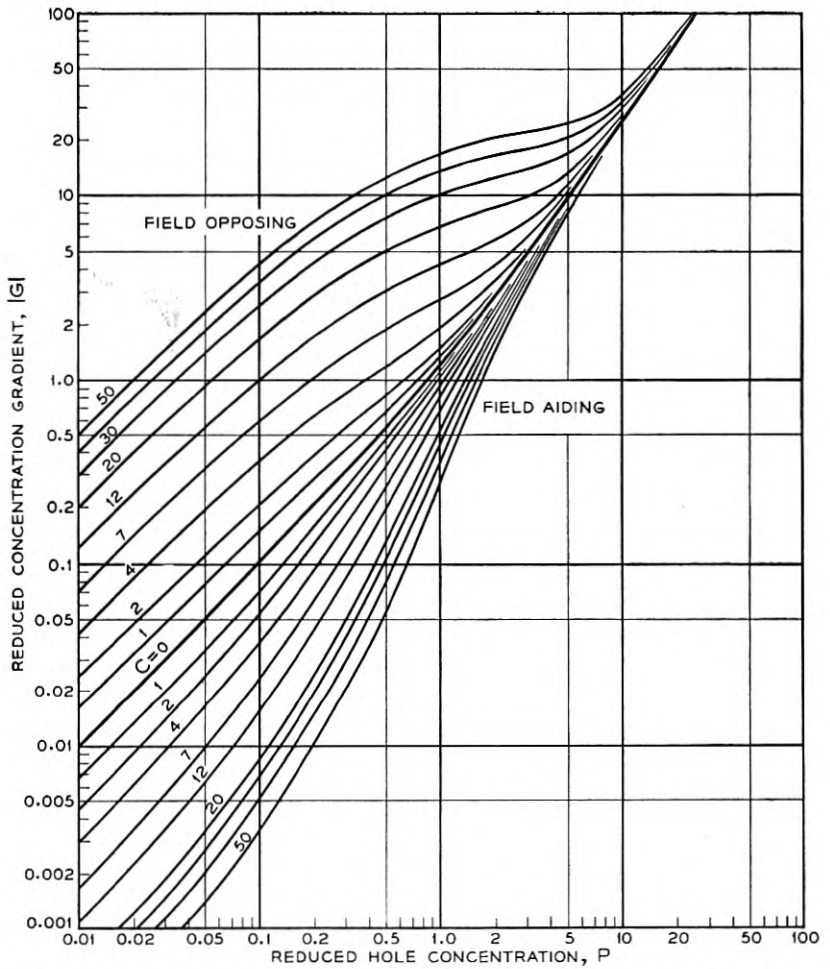


Fig. 2.—The dependence of the reduced concentration gradient on reduced concentration for the steady-state one-dimensional flow of holes with mass-action recombination in n-type germanium.

respectively in Fig. 4 and in Fig. 5. In accordance with equations (10), (18), and (26), the solutions for C_p and F are found from

$$C_p = \frac{1}{b}(FP - G) = \frac{CP - (1 + 2P)G}{b + (b + 1)P}, \tag{44}$$

and

$$F = \frac{C - \frac{b - 1}{b}G}{1 + \frac{b + 1}{b}P}. \tag{45}$$

The electrostatic potential may be evaluated from F in a manner similar to that followed in the preceding section.

3.3 *Detailed properties of the solutions*

The general solutions given in the figures illustrate certain properties which can be established through the analytical approximations obtainable for small and for large values of the relative concentration of added holes. The principal qualitative properties evident from the figures are:

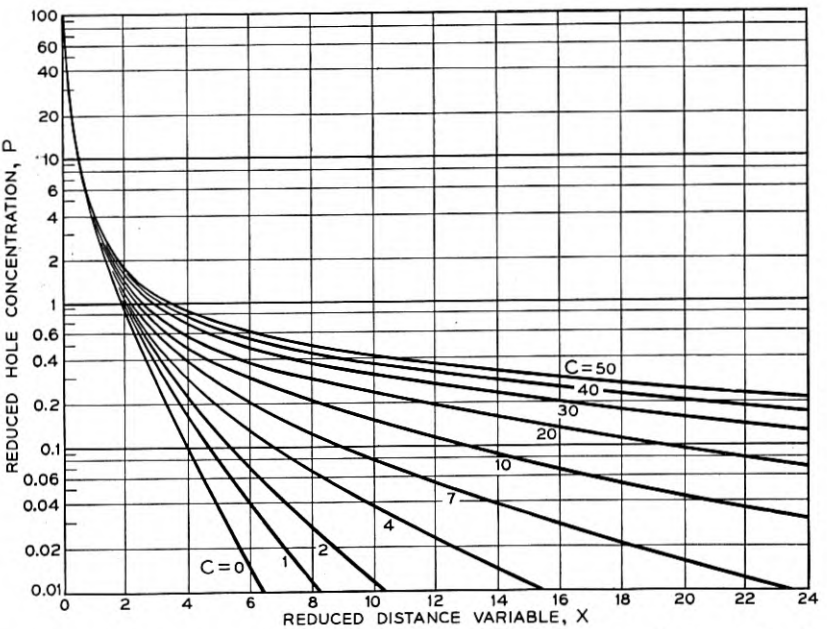


Fig. 3.—The dependence of the reduced concentration on reduced distance for the steady-state one-dimensional flow of holes with mass-action recombination in n-type germanium.

The relative hole concentration, P, and the reduced hole current, C_p, depend exponentially on distance for small concentrations; and for large concentrations all solutions for a given dependent variable run together, independently of the value of the current parameter, and give comparatively rapid variations of hole concentration and current with distance[30]. The property that a common solution independent of total current or

[30] These rapid variations would account for the observation of J. R. Haynes that estimates, for a given emitter current, of hole concentrations or currents in a filament at a point contact removed from the emitter, with no additional applied field, are largely independent of changes in f_e for the emitter.

applied field obtains for large P results from diffusion in conjunction with the increase of conductivity. As may be expected, the solutions for the case of constant mean lifetime also have this property, the recombination law merely affecting the form of the common solution.

In Fig. 6 are shown curves for P, C_p, and F for the case of constant mean lifetime in n-type germanium, evaluated for C equal to 16.3. These curves are intended to illustrate the qualitative differences between the solutions for this case and those for mass-action recombination, which are manifest

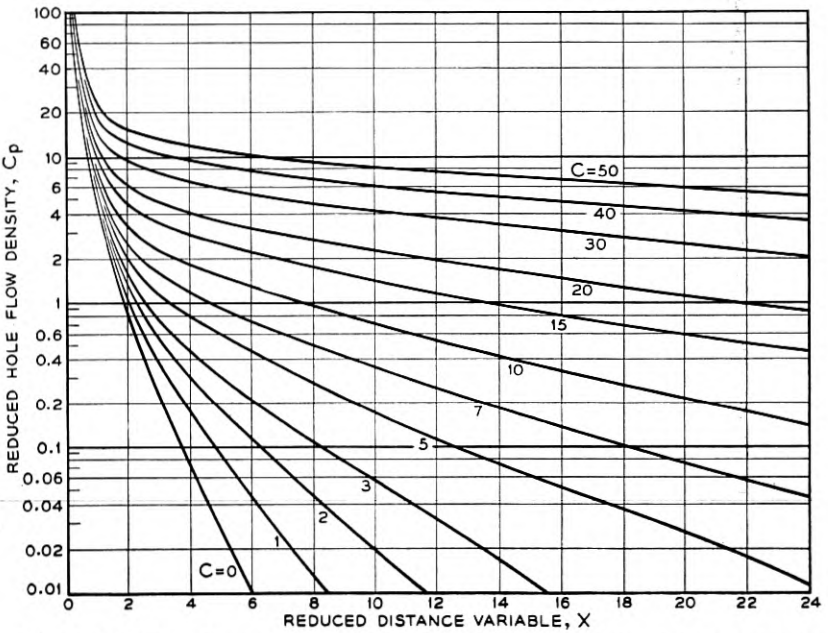


Fig. 4.—The dependence of the reduced hole flow density on reduced distance for the steady-state one-dimensional flow of holes with mass-action recombination in n-type germanium.

primarily at the larger concentrations. The dashed curves in the figure give the corresponding solutions for the case of mass-action recombination; and the X-origins for the two cases have been so chosen that corresponding curves, which exhibit essentially the same dependence on X for small P, coincide in the limit of small P. As the figure shows, constant mean lifetime gives an exponential dependence of P on X for large P, while mass-action recombination gives larger concentration gradients, with an increase of P to indefinitely large values in the neighborhood of a vertical asymptote.

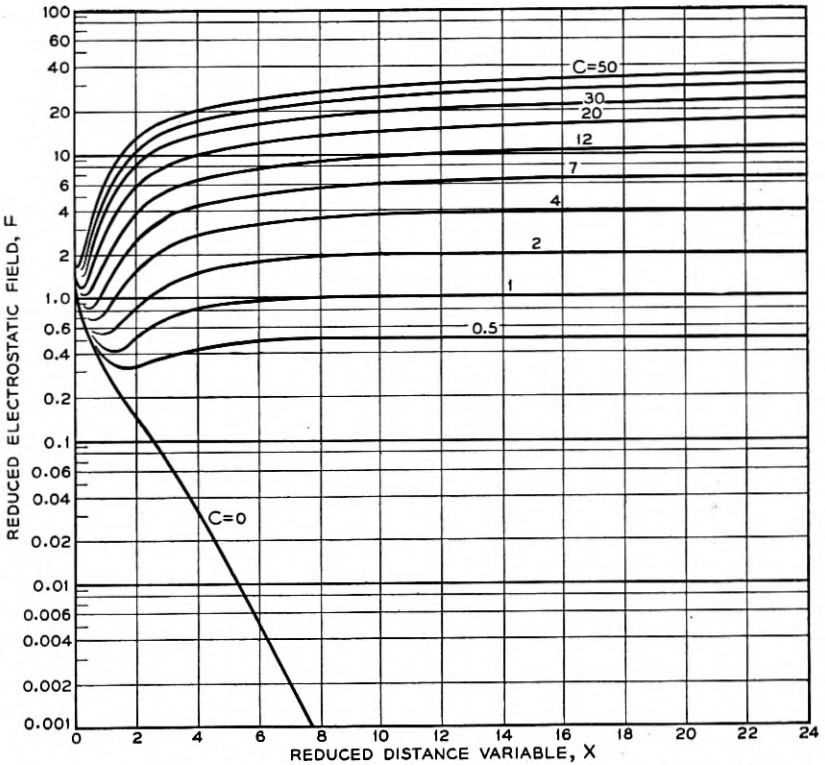


Fig. 5.—The dependence of the reduced electrostatic field on reduced distance for the steady-state one-dimensional flow of holes with mass-action recombination in *n*-type germanium.

3.31 *The behavior for small concentrations*

The exponential dependence of P and C_p on distance for P small is given for the *n*-type semiconductor by the analytical approximations,

$$\begin{cases} P = P_s \exp\left[-\tfrac{1}{2}[\pm\sqrt{C^2+4} - C]X\right] \\ C_p = \dfrac{1}{2b}[\pm\sqrt{C^2+4} + C]P, \end{cases} \tag{46}$$

where P_s is a suitable constant. If C is positive, the plus sign holds for field aiding[31] and the minus sign for field opposing. These approximate

[31] It is evident from the curves for C_p in Fig. 4 that the exponential extrapolation back to the emitter location of estimates of hole concentrations or currents at a point contact on a germanium filament lead to values of f_ϵ for the emitter which are too small. Using moderately large injected currents and no additional applied fields, J. R. Haynes once obtained in this way an apparent f_ϵ of about 0.2. From the figure, this is the apparent f_ϵ to be expected for moderate and large values of C for the true f_ϵ equal to unity.

solutions, which hold for any recombination law, are obtained quite simply, by integration, from G in terms of P to the first term of the Maclaurin's expansion, given in the Appendix. It might be noted that for this approximation the electrostatic field is equal to the applied field, so that F equals C.

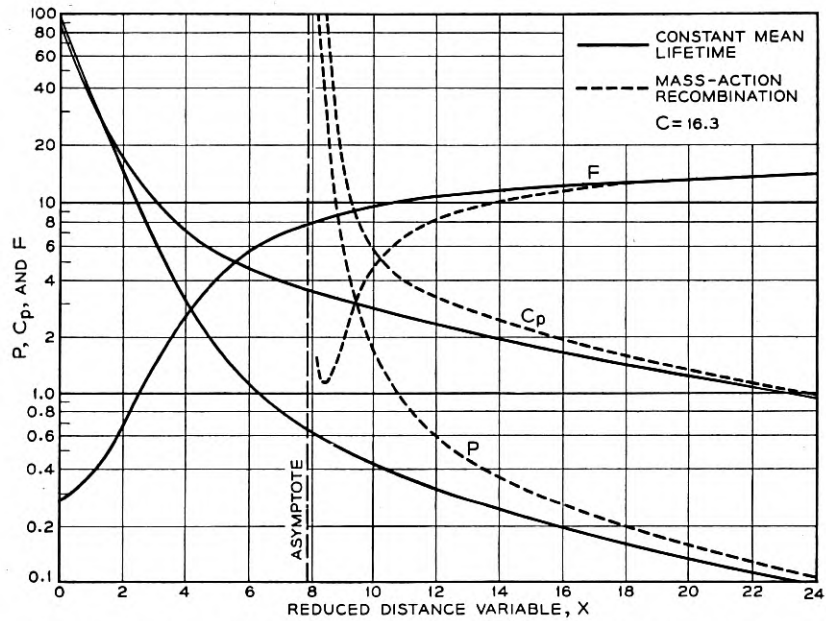


Fig. 6.—The dependence of the reduced hole concentration, hole flow density, and electrostatic field on reduced distance for steady-state one-dimensional hole flow in n-type germanium, for the cases of constant mean lifetime and mass-action recombination.

Since P is small, the transport velocity of holes is equal to their differential transport velocity[32]. Writing the equation for C_p in dimensional form, the transport velocity is found to equal

$$s = \tfrac{1}{2}\left[\pm\sqrt{(\mu E_a)^2 + 4D_p/\tau} + \mu E_a\right], \tag{47}$$

with the plus sign for field aiding and the minus sign for field opposing, if the applied field, E_a, is positive. This result is consistent with the

[32] In accordance with equations (7), (8), and (10), the differential transport velocity for the steady state in one dimension may be found from the general formula,

$$bM_0 dC_p/dP = -(P - P_0)R/G.$$

Its equalling the transport velocity proper for P small appears to result from the property of non-composite cases that the dependent variables, for a given C, are all functions of P which do not depend on any quantity determined by the boundary values, a property which composite cases, with their additional degree of freedom, do not possess.

equation for P, which may be written as

$$P = P_s \exp(-x/s\tau). \tag{48}$$

For a large aiding field, s reduces to the velocity of drift under this field while, for a large opposing field, the magnitude of s is approximately $D_p/\mu E_a\tau$. For zero field, s equals the diffusion velocity $(D_p/\tau)^{\frac{1}{2}}$, which is a diffusion distance for a mean lifetime divided by the mean lifetime. This diffusion velocity can be specified in terms of its field equivalent, or the field which gives an equal drift velocity, and for germanium it is found that the equivalent field is about 8 volt cm^{-1} for τ equal to one microsecond and about 2.5 volt cm^{-1} for τ equal to 10 microseconds.

For small concentrations of added holes in the intrinsic semiconductor, or $(P-1) << 1$, equations (38) and (40) give the approximate solutions,

$$\begin{cases} P - 1 = (P^0 - 1) \exp\left[\pm\left[\dfrac{(1+a)(b+1)}{2b}\right]^{\frac{1}{2}} X\right] \\ C_p = \dfrac{1}{b+1}\left[C - \dfrac{2b}{b+1} G\right] \sim \dfrac{1}{b+1} C, \end{cases} \tag{49}$$

the X-origin being selected arbitrarily at the point at which the relative concentration is P^0 according to the approximation. It is evident from the equation for C_p that, for $(P-1)$ small, the transport velocity is the drift velocity under the applied field, which is the velocity of the holes normally present in the semiconductor. The differential transport velocity, obtainable by differentiating the equation for C_p with respect to P and using the differential equation (28), or by writing the exponent in the equation for $(P-1)$ in the form given in (48), is, on the other hand, given by

$$s = \left[\frac{2b}{(1+a)(b+1)}\right]^{\frac{1}{2}}\left[\frac{D_p}{\tau}\right]^{\frac{1}{2}} = \left[\frac{1}{1+a}\frac{2D_pD_n}{D_p+D_n}\Big/\tau\right]^{\frac{1}{2}}, \tag{50}$$

and is a diffusion velocity. This holds for holes added in any concentration if $a = 0$, or for constant mean lifetime, since the first of equations (49) is then the general solution given in (39).

The nature of the flow for small concentrations of added carriers in the general case, which depends on the parameter P_0, is illustrated qualitatively by the n-type and intrinsic cases considered, for which P_0 is respectively zero and infinite. Solutions for the general case are easily evaluated analytically from the linear differential equation which results from (17) if $P - P_0 << \frac{1}{2} + P_0$. It can be shown from the field-aiding steady-state solution that the ratio of the differential transport velocity to the velocity, proportional to C, of drift under the applied field is for $C^2 >> (1 + 2P_0)M_0$ equal to the quantity $1/M_0$. This result is consistent

with those already derived: For large applied aiding fields, the differential transport velocity changes from the drift velocity, for P_0 equal to zero and M_0 unity, to the diffusion velocity given in (50) as P_0 and M_0 increase indefinitely.

3.32 *The zero-current solutions and the behavior for large concentrations*

The solutions for the intrinsic semiconductor for the current parameter equal to zero are, of course, the same as the general ones given in Section 3.1, since the current parameter does not occur in the differential equation. For the n-type semiconductor, the differential equation (27) becomes an equation of the Bernouilli type for C equal to zero, and may be solved by quadratures. It is then linear in G^2, and gives, for field aiding or field opposing,

$$G^2 = 2\left[\frac{1+\frac{b+1}{b}P}{1+2P}\right]^2 \int_0^P \frac{P(1+P)(1+aP)}{1+\frac{b+1}{b}P}\,dP, \tag{51}$$

expressing the recombination function R according to equation (16) for a combination of the two recombination mechanisms. Writing, for brevity,

$$\begin{cases} \beta \equiv \dfrac{b}{b+1} \\ M \equiv 1 + \dfrac{b+1}{b}P, \end{cases} \tag{52}$$

and evaluating the integral in (51), the following result is obtained:

$$\begin{aligned} G^2 = 2\beta^2\left[\frac{M}{1+2P}\right]^2 \Big[& [1-a] \\ & \cdot [\beta(M^2-1) + (1-4\beta)(M-1) - (1-2\beta)\log M] \\ & + a[\tfrac{2}{3}\beta^2(M^3-1) + \tfrac{3}{2}(1-2\beta)(M^2-1) \\ & + (1-6\beta+6\beta^2)(M-1) - (1-\beta)(1-2\beta)\log M]\Big]. \end{aligned} \tag{53}$$

For P large, this solution gives the approximations,

$$G = \pm\left[\frac{b+1}{2b}\right]^{\frac{1}{2}} P \tag{54}$$

for constant mean lifetime, with $a = 0$, and

$$G = \pm\left[\frac{a(b+1)}{3b}\right]^{\frac{1}{2}} P^{\frac{3}{2}} \tag{55}$$

if there is mass-action recombination present, so that $a \neq 0$. The dependence of P on X for these approximations is readily obtained by integrating the differential equations which result from writing in place of G, its definition, dP/dX; constant mean lifetime gives an exponential dependence. An examination of (54) and (55) in conjunction with the general differential equation (27) shows that, for P large, the dominant term in the differential equation is independent of C. It follows that solutions for all values of C approach a common solution for P large, which is given by (54) or (55). The solutions run together appreciably for P sufficiently large that P and M are substantially proportional, that is, for P large compared with $b/(b + 1)$, which is of order unity. It is to be expected that the approximations (54) and (55) should apply equally well to the intrinsic semiconductor, and this expectation is easily verified by evaluating the integral in equation (35) for the intrinsic semiconductor, for P large, for the two recombination cases here considered.

4. Solutions of Simple Boundary-Value Problems for a Single Source

Among the boundary-value problems whose solutions are useful in the interpretation of data from experiments in hole injection are the following: the semi-infinite filament for field aiding, with holes injected at the end, which constitutes a relatively simple case; and the doubly-infinite filament with a single plane source, with which this section will be primarily concerned.

Consider first the semi-infinite filament, and suppose that it starts at the X-origin and extends over positive X, so that the current parameter is positive for field aiding. If two quantities are specified, namely the current parameter and the fraction f_ϵ of the current carried by holes at the origin or injection point, then the solution of the boundary-value problem is completely determined. It is merely necessary to select the general field-aiding solution for P or C_p in terms of X, for the particular value of the current parameter, and then to determine the X-origin, corresponding to the source, which is simply the X at which the ratio f of C_p to C equals f_ϵ.

Use in the boundary-condition equations (33) and (34) of the approximate expressions given in (54) and (55) for G in terms of P, for large P, permits the complete analytical determination of the dependence of P^0 on total current as this current is indefinitely increased. It was shown in Section 2.4 that, if f_ϵ is less than $1/(b + 1)$ for the n-type semiconductor, P^0 approaches as a limit the value for which G^0 vanishes according to the boundary-condition equation (33); in all other cases for the n-type semiconductor, or if f_ϵ exceeds $1/(b + 1)$ for the intrinsic semiconductor, P^0 increases indefinitely with C. For $f_\epsilon > 1/(b + 1)$, it is readily seen that

P^0 is proportional in the limit to C for constant mean lifetime, and to $C^{\frac{1}{2}}$ for mass-action recombination; and, for $f_\epsilon = 1/(b+1)$ in the case of the n-type semiconductor, P^0 increases as $C^{\frac{1}{2}}$ for constant mean lifetime, and as $C^{\frac{1}{3}}$ for mass-action recombination.

Consider now the doubly-infinite semiconductor filament with a source at the origin, and suppose that the total injected current at the source is C_ϵ, in reduced form, with a fraction f_ϵ of this current carried by holes. Denote by C^- and by C^+ the reduced total currents for $X < 0$ and for $X > 0$, respectively. Since the injection of holes requires that C_ϵ be positive, at least one of C^- and C^+ must be positive, since total current is conserved. Let f^- and f^+ denote, respectively, the ratio of the hole current at the origin to the left, C_p^-, to the total current C^-, and the ratio of the hole current at the origin to the right, C_p^+, to the total current, C^+. It might be noted that, for a flow of holes to the left, say, against the field, C^- and C^+ are positive and f^- is negative, and that, if C^- is (plus) zero, f^- is (negatively) infinite, corresponding to the flow of holes under zero applied field. Now, general boundary-condition equations of the form of (33) or (34) hold with the sign conventions here employed, as indicated in Section 2.4. One may be written for the flow to the left, another for the flow to the right, making use of the condition that the relative concentration P is everywhere continuous; G exhibits a discontinuity of the first kind at the source, with a change in sign. Writing G^- for the limiting value of the reduced concentration gradient as the origin is approached from the left, and G^+ the limiting value as the origin is approached from the right, the boundary-condition equations are, for the n-type semiconductor,

$$(56)\qquad \begin{cases} G^- = -\dfrac{b+(b+1)P^0}{1+2P^0}\left[f^- - \dfrac{P^0}{b+(b+1)P^0}\right]C^- \\ G^+ = -\dfrac{b+(b+1)P^0}{1+2P^0}\left[f^+ - \dfrac{P^0}{b+(b+1)P^0}\right]C^+. \end{cases}$$

For the intrinsic semiconductor, they are

$$(57)\qquad \begin{cases} G^- = -\dfrac{(b+1)^2}{2b}\left[f^- - \dfrac{1}{b+1}\right]C^- \\ G^+ = -\dfrac{(b+1)^2}{2b}\left[f^+ - \dfrac{1}{b+1}\right]C^+. \end{cases}$$

There are, in addition, an equation which expresses the conservation of hole flow, and one which expresses the conservation of total current, as follows:

$$(58)\qquad \begin{cases} f^+C^+ - f^-C^- = f_\epsilon C_\epsilon \\ C^+ - C^+ = C_\epsilon \quad . \end{cases}$$

The solution of the problem is determined by f_ϵ and the three parameters which specify the total currents: With these four quantities known, then, from equations (56) or (57) in conjunction with (58) and the known general solutions in the $(\Delta P, G)$-plane which apply to the left and to the right of the origin, all of the quantities P^0, G^-, G^+, f^- and f^+ can be found and the problem completely solved.

The technique of obtaining the solution depends on a simple fundamental result which may be expressed as follows:

For fixed f_ϵ and C_ϵ, consider the sum of the magnitudes of the concentration gradients at a single common source from which holes flow into a number of similar filaments in parallel, for any consistent distribution among the filaments of total currents, some of which may be produced by opposing fields. This sum is equal to the magnitude of the concentration gradient at the source if the entire flow, under the appropriate aiding field, were confined to a single filament.

The total magnitude of the concentration gradient, in this sense, is an invariant for fixed f_ϵ and C_ϵ. Specifically, for the n-type semiconductor, it follows from equations (56) and (58) that

$$G^+ - G^- = -\frac{b + (b+1)P^0}{1 + 2P^0}\left[f_\epsilon - \frac{P^0}{b + (b+1)P^0}\right]C_\epsilon. \tag{59}$$

Similarly, for the intrinsic semiconductor,

$$G^+ - G^- = -\frac{(b+1)^2}{2b}\left[f_\epsilon - \frac{1}{b+1}\right]C_\epsilon. \tag{60}$$

The left-hand sides of these equations are the negative of the sum of the magnitudes of the reduced concentration gradients, since G^- is always positive and G^+ always negative, and their right-hand sides are similar in form to those of equations (56) and (57), with the quantities f_ϵ and C_ϵ, characteristic of the source, replacing f^- and C^-, or f^+ and C^+.

The particular utility of these equations arises from their independence of the unknowns f^- and f^+. By means of equation (59) for the n-type semiconductor the evaluation of the five unknown quantities can now be effected as follows: With the current parameters known, the solutions in the (P, G)-plane to the left and right of the X-origin are determined; either both solutions are for field aiding, or else one is for field aiding and the other for field opposing. From them, the sum of the magnitudes of the reduced concentration gradients can be found as a function of P. It is also given, for the origin, as a function of the unknown P^0, by equation (59). The values of the sum for the origin and of P^0 are accordingly found as those which satisfy both relationships. The value of P^0 thus found determines both G^- and G^+ from the respective solutions

in the (P, G)-plane, and f^- and f^+ may be obtained by solving for them in equations (56).

For the intrinsic semiconductor, this method can be applied analytically, and the solution so obtained serves at the same time as an approximation for large relative hole concentrations in the n-type semiconductor, for which the method is otherwise essentially graphical or numerical in the general case. Making use of the symmetry of the solutions for the intrinsic semiconductor about a source, it follows from (57) and (58) that

$$\begin{aligned} G^+ = -G^- &= -\frac{(b+1)^2}{4b}\left[f_\epsilon - \frac{1}{b+1}\right]C_\epsilon \\ &= -\frac{(b+1)^2}{2b}\left[f^+ - \frac{1}{b+1}\right]C^+ = \frac{(b+1)^2}{2b}\left[f^- - \frac{1}{b+1}\right]C^-, \end{aligned} \tag{61}$$

whence

$$\begin{cases} f^- = \dfrac{1}{b+1} - \dfrac{1}{2}\left[f_\epsilon - \dfrac{1}{b+1}\right]\dfrac{C_\epsilon}{C^-} \\ f^+ = \dfrac{1}{b+1} + \dfrac{1}{2}\left[f_\epsilon - \dfrac{1}{b+1}\right]\dfrac{C_\epsilon}{C^+}. \end{cases} \tag{62}$$

It is easily verified that this result holds approximately for large relative concentrations in the n-type semiconductor. Three simple special cases of (62) might be considered: The first is

$$\begin{cases} C^- = -C^+ = -\tfrac{1}{2}C_\epsilon \\ f^- = f^+ = f_\epsilon . \end{cases} \tag{63}$$

This is the rather trivial case of symmetrical flows from a source which supplies all currents. A second special case is that for which C^- and C^+ are both positive, say, and such that there is no hole flow to the left against the field. It is readily found that, for this case,

$$\begin{cases} C^- = \dfrac{b+1}{2}\left[f_\epsilon - \dfrac{1}{b+1}\right]C_\epsilon; \quad C^+ = \dfrac{b+1}{2}\left[f_\epsilon + \dfrac{1}{b+1}\right]C_\epsilon \\ f^- = 0; \quad f^+ = \dfrac{2}{b+1}\dfrac{f_\epsilon}{f_\epsilon + 1/(b+1)}. \end{cases} \tag{64}$$

Here, the drift from the left under the applied field of holes normally present in the intrinsic semiconductor just cancels the diffusion from the source to the left.[33] A third special case is that in which the total current

[33] Using the numerically obtained solutions, the validity of (64) as an approximation for large concentrations in n-type germanium may be seen as follows: For f_ϵ equal to unity and C_ϵ, C^- and C^+ equal to 2, 1.5, and 3.5 respectively, P^0 is about 0.6 and the fraction of injected holes which flows against the field is nearly one-half; doubling these current densities increases P^0 to 1.45 and decreases the fraction to about one-fourth, and the fraction is less than about one-tenth if the current densities are increased so that C^+ exceeds 15

to the left of the source is zero, the left-hand side of the filament being open-circuited. For this case, equations (62) are better written in the form obtained by multiplying through by C^- or C^+, and the special case in question is then found to be given by

$$(65)\quad \begin{cases} C^- = 0; & C^+ = C_\epsilon \\ C_p^- = -\frac{1}{2}\left[f_\epsilon - \frac{1}{b+1}\right] C_\epsilon; & C_p^+ = \frac{1}{2}\left[f_\epsilon + \frac{1}{b+1}\right] C_\epsilon, \end{cases}$$

according to which, if f_ϵ is equal to unity, the magnitude of the hole flow to the left into the open-circuit end is $b/(b+2)$ times that into the circuit end, to the right; or a fraction $b/2(b+1)$ of the holes flows to the left, and a fraction $(b+2)/2(b+1)$ to the right. Thus, for germanium, the hole flow into the open-circuit end is 0.43 as large as that into the circuit end, a fraction 0.30 flowing to the left, and 0.70 to the right. It might be observed that the fractions of the injected holes which flow to the left and right are, in this case, proportional to the total currents C^- and C^+ of the preceding case, for which there is zero hole flow to the left.

Another general limiting case for the n-type semiconductor is that for P_0 small, so that the exponential approximations of Section 3.31 apply. The restriction on the magnitude of P is $P << \frac{1}{2}$. This restriction obtains if C_ϵ is sufficiently small that C^- and C^+ do not differ appreciably. Equation (59) then gives

$$(66)\qquad G^+ - G^- = -bf_\epsilon C_\epsilon .$$

Writing C for C^- and C^+, equations (30) and (46) result in

$$(67)\quad \begin{cases} G^- = \frac{1}{2}[\sqrt{C^2+4} + C]\, P^0 = bC_p^+ \\ G^+ = -\frac{1}{2}[\sqrt{C^2+4} - C]\, P^0 = bC_p^-, \end{cases}$$

whence, solving for $G^+ - G^-$ and comparing with equation (66),

$$(68)\qquad P^0 = bf_\epsilon C_\epsilon/\sqrt{C^2+4}.$$

In accordance with (67), then,

$$(69)\quad \begin{cases} C_p^- = -\frac{1}{2}[1 - C/\sqrt{C^2+4}]\, f_\epsilon C_\epsilon \\ C_p^+ = \frac{1}{2}[1 + C/\sqrt{C^2+4}]\, f_\epsilon C_\epsilon. \end{cases}$$

These are the reduced hole flows to the left and right of the source. While it has been assumed that C_ϵ is small compared with C, no restriction has been placed on C itself. For C small compared with unity, the equations indicate that the hole flows to the left and right are the

same in magnitude, while for C large compared with unity,

$$\begin{cases} C_p^- \sim -\frac{1}{C^2} f_\epsilon C_\epsilon \\ C_p^+ \backsim f_\epsilon C_\epsilon . \end{cases} \tag{70}$$

Thus, according to this approximation, C should exceed about 10 if no more than one per cent of the holes are to flow against the field. From (75) in the Appendix, a value of 10 for C corresponds to a current density of about 1.2 amp cm^{-2} in germanium of 10 ohm cm resistivity, with τ equal to 10 μsec. This current density is moderately large among those which have been employed in experiments with germanium filaments.

Experimentally, the ideal one-dimensional geometry postulated in the present treatment of the problem of the single source in an infinite filament cannot easily be realized, hole injection generally being accomplished through a point contact or a side arm on one side of the actual filament. If suitable averages are employed, non-uniformity in P at the injection cross-section does not, however, vitiate the approximate results for ΔP large and ΔP small, since their applicability depends largely on the validity over the injection cross-section of the approximation assumed.

ACKNOWLEDGMENT

The author is indebted to a number of his colleagues for their stimulating interest and encouragement; to J. Bardeen and W. Shockley for a number of valuable and helpful comments, as well as to W. H. Brattain, J. R. Haynes, C. Herring, L. A. MacColl, G. L. Pearson, and R. C. Prim. J. Bardeen also suggested the numerical analysis for n-type germanium which constituted one of the initial points of attack, and aided materially in its inception. The rather difficult numerical integrations and associated problems were ably handled by R. W. Hamming, Mrs. G. V. Smith and J. W. Tukey.

5. APPENDIX

5.1 *The concentrations of ionized donors and acceptors*

While the donor and acceptor concentrations need not, of course, be considered for the intrinsic semiconductor, for the extrinsic semiconductor the fundamental equations, as they have been written, are in principle incomplete: Two additional equations in the variables D^+ and A^- are required. One of the required equations is trivial, since changes in the concentration of ionized centers which are compensated by those which determine the conductivity type of the extrinsic semiconductor

can certainly be neglected. For an n-type semiconductor, for example, the term $(A^- - A_0^-)$ in Poisson's equation may be suppressed. This procedure is strictly consistent with the neglect of p_0 and g_0, but undoubtedly holds to an even better approximation. If D is the total donor concentration in the n-type semiconductor, the concentration of ionized donors may be considered to satisfy the equation,

$$\frac{\partial D^+}{\partial t} = H(D - D^+) - KD^+n, \tag{71}$$

which applies to the homogeneous semiconductor, with H and K constants which characterize, respectively, the rate of ionization of unionized donors, and the rate of recombination of an ionized donor with an electron. If, as a result of a small thermal ionization energy, most of the donors are ionized, so that $KD/H << 1$, the change in ionized-donor concentration for the steady state is given by (71) as

$$D^+ - D_0^+ \sim -\frac{KD}{H}(n - n_0), \tag{72}$$

which is small compared with the corresponding change in electron concentration. In other cases, the use of the general expression obtainable from (71) for the steady-state concentration of ionized donors in terms of the electron concentration, or the expression for the other limiting case of relatively few ionized donors, might provide a more precise description provided the conditions under which solutions are sought do not involve unduly rapid changes with time.

5.2 *The carrier concentrations at thermal equilibrium*

The ratio of the thermal-equilibrium values of the hole and electron concentrations may be evaluated for n-type germanium from[5]

$$\begin{cases} np = 3\cdot 10^{32}T^3 \exp\left(-\dfrac{8700}{T}\right) \equiv n_i^2 \\ n - p = n_s \sim n_0 = 1/b\mu e\rho_0 = 2.40\cdot 10^{15}/\rho_0, \end{cases} \tag{73}$$

where the electron concentration excess n_s corresponds to complete ionization of the donors, and is approximately n_0 at the highest temperature at which P_0 is still negligible, which may be taken as room temperature[29]. The resistivity ρ_0 is that which determines n_0. Thus,

$$P_0 = \tfrac{1}{2}\left[\sqrt{1 + 4(n_i/n_0)^2} - 1\right], \tag{74}$$

[29] loc. cit.
[5] loc. cit.

with n_i, the concentration of holes or electrons in intrinsic germanium at T deg abs, given in (73). It may be estimated that temperature rises of less than 100 deg C will make 10 ohm cm n-type germanium substantially intrinsic in its behavior.

The range of values of the parameter C for which the numerical solutions are given corresponds, for example, to current densities up to the order of 10 amp cm^{-2} in germanium filaments of about 10 ohm cm resistivity, for the mean lifetime τ about 10 μsec; for this mean lifetime, the distance unit L_p is approximately $2 \cdot 10^{-2}$ cm. Current densities corresponding to the larger values of C will ordinarily produce appreciable joule heating in filaments some 10^{-3} cm^2 in area of cross-section, cemented to a backing, with temperature rises of the order of 100 deg C.

The effect of joule heating on L_p and C may be evaluated from

$$(75) \qquad \begin{cases} L_p = 6.6 \left[\dfrac{T}{300}\right]^{-\frac{1}{4}} \tau^{\frac{1}{2}} \\ C = 2.6 \cdot 10^2 \left[\dfrac{T}{300}\right]^{-\frac{3}{4}} \tau^{\frac{1}{2}} \rho I, \end{cases}$$

where τ is expressed in sec, I in amp cm^{-2}, and ρ is the normal resistivity in ohm cm of the germanium at T deg abs. These are obtained from the definitions (8), taking the hole mobility in the thermal scattering range to be proportional to $T^{-\frac{3}{2}}$, with the value 1700 cm^2 volt^{-1} sec^{-1} at 300 deg abs.[7]

5.3 *Series solutions for the extrinsic semiconductor in the steady state*

Maclaurin's series for G in the relative concentration P are of the form

$$(76) \qquad G = a_1 P + a_2 P^2 + a_3 P^3 + \ldots$$

for the cases of field opposing and field aiding, the solutions passing through the (P, G)-origin. Substituting the series (76) for G in the differential equation (27) for the n-type semiconductor in the steady state, it is found, in accordance with (30), that

$$(77) \qquad a_1 = \tfrac{1}{2} \left[C \pm \sqrt{C^2 + 4}\right],$$

the sign of C being taken before the radical for field opposing, the other sign for field aiding. The other coefficients are given in terms of a_1 and

[7] loc. cit.

also the b, C, and the constant, a, of the recombination function:

$$(78)\quad \begin{cases} a_2 = \dfrac{4a_1^2 - \left[2\dfrac{b+1}{b} + a\right]}{C - 3a_1} \\ a_3 = \dfrac{2a_2^2 + \dfrac{11b+1}{b} a_1 a_2 + 2\dfrac{b+1}{b} a_1^2 - \dfrac{b+1}{b}\left[\dfrac{b+1}{b} + 2a\right]}{C - 4a_1} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases}$$

The series in the current parameter are series in ascending powers of the reciprocal of C. Writing, for convenience,

$$(79)\quad \gamma \equiv 1/C,$$

the differential equation (27) may be put in the form,

$$(80)\quad \begin{aligned} &\gamma[1 + 2P]\left[1 + \frac{b+1}{b} P\right] GG' \\ &\qquad + \gamma \frac{b-1}{b} G^2 - G - \gamma P\left[1 + \frac{b+1}{b} P\right]^2 R = 0, \end{aligned}$$

using the prime to denote differentiation with respect to P. Consider expansions of the form,

$$(81)\quad G = \sum_{j=j_0}^{\infty} A_j \gamma^j,$$

in which the A's are functions of P to be determined. Substituting in the differential equation, there results

$$(82)\quad \begin{aligned} &\sum_{j=j_0}^{\infty} \sum_{m=j_0}^{\infty} \left[[1 + 2P]\left[1 + \frac{b+1}{b} P\right] A_j A_m' + \frac{b-1}{b} A_j A_m\right] \gamma^{j+m+1} \\ &\qquad - \sum_{j=j_0}^{\infty} A_j \gamma^j - P\left[1 + \frac{b+1}{b} P\right]^2 R\gamma = 0. \end{aligned}$$

Since the expansions are to hold for arbitrary values of γ, the A's must, for the cases of field opposing and field aiding, for which the solutions pass through the (P, G)-origin, vanish identically for P equal to zero, and be determined by equating to zero the coefficients of given powers of γ in (82). It can, without loss of generality, be assumed that the coefficient of the leading term in the expansion, A_{j_0}, is not identically zero. Then, from (82), it is found that there is no expansion for $j_0 = 0$, that is, no expansion starting with a term independent of γ. Formal expansions can be obtained, however, for $j_0 = -1$ and for $j_0 = +1$. These may be identified,

respectively, with the solutions for field opposing and field aiding, as will be seen.

For $j_0 = -1$, or field opposing, (82) leads to differential equations of the first order for the determination of the A's. The condition that these functions vanish identically for $P = 0$ suppresses all A's of even order. The first term of the expansion is found by solving

$$(83)\quad A'_{-1} + \frac{b-1}{b}\frac{A_{-1}}{[1+2P]\left[1+\frac{b+1}{b}P\right]} = \frac{1}{[1+2P]\left[1+\frac{b+1}{b}P\right]}$$

whence

$$(84)\quad A_{-1} = \frac{P}{1+2P}.$$

The second term is found from

$$(85)\quad A'_1 + \frac{b-1}{b}\frac{A_1}{[1+2P]\left[1+\frac{b+1}{b}P\right]} = \left[1+\frac{b+1}{b}P\right]R,$$

whence, with R equal to unity and $(1+P)$, respectively,

$$(86)\quad \begin{cases} A_1 = \dfrac{P[1+P]\left[1+\frac{b+1}{b}P\right]}{1+2P} & \text{for constant mean lifetime} \\ A_1 = \dfrac{P\left[1+\frac{3}{2}P+\frac{2}{3}P^2\right]\left[1+\frac{b+1}{b}P\right]}{1+2P} & \text{for mass-action recombination.} \end{cases}$$

For the third term, making use of (84), (85) and (86),

$$(87)\quad \begin{cases} A'_3 + \dfrac{b-1}{b}\dfrac{A_3}{[1+2P]\left[1+\frac{b+1}{b}P\right]} \\ \qquad = -[1+P]\left[1+\frac{b+1}{b}P\right]^2 & \text{for constant mean lifetime} \\ A'_3 + \dfrac{b-1}{b}\dfrac{A_3}{[1+2P]\left[1+\frac{b+1}{b}P\right]} \\ = -[1+P]\left[1+\frac{3}{2}P+\frac{2}{3}P^2\right]\left[1+\frac{b+1}{b}P\right]^2 & \text{for mass-action recombination} \end{cases}$$

whence

$$
(88)\quad \begin{cases}
A_3 = \dfrac{-P\left[1 + \dfrac{4b+1}{2b}P + \dfrac{5b+3}{3b}P^2 + \dfrac{b+1}{2}P^3\right]\left[1 + \dfrac{b+1}{b}P\right]}{1+2P} \\
\qquad\qquad \text{for constant mean lifetime} \\
A_3 = \dfrac{-P\left[1 + \dfrac{33b+6}{12b}P + \dfrac{70b+27}{18b}P^2 + \dfrac{73b+43}{24b}P^3 + \dfrac{38b+30}{30b}P^4 + \dfrac{2b+2}{9b}P^5\right]\left[1 + \dfrac{b+1}{b}P\right]}{1+2P} \\
\qquad\qquad \text{for mass-action recombination.}
\end{cases}
$$

For $j_0 = +1$, or field aiding, the A's are determined somewhat more simply, recursive relationships obtaining. The results are:

$$
(89)\qquad A_1 = -P\left[1 + \frac{b+1}{b}P\right]^2 R,
$$

and

$$
(90)\quad \begin{cases}
A_3 = [1+2P]\left[1 + \dfrac{b+1}{b}P\right]A_1A_1' + \dfrac{b-1}{b}A_1^2 \\
A_5 = [1+2P]\left[1 + \dfrac{b+1}{b}P\right][A_1A_3]' + 2\dfrac{b-1}{b}A_1A_3 \\
A_7 = [1+2P]\left[1 + \dfrac{b+1}{b}P\right][[A_1A_5]' + A_3A_3'] \\
\qquad\qquad + 2\dfrac{b-1}{b}\left[A_1A_5 + \dfrac{1}{2}A_3^2\right] \\
A_9 = [1+2P]\left[1 + \dfrac{b+1}{b}P\right][[A_1A_7]' + [A_3A_5]'] \\
\qquad\qquad + 2\dfrac{b-1}{b}[A_1A_7 + A_3A_5] \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{cases}
$$

The identification of the series in the parameter γ as series for field opposing and field aiding is accomplished by evaluating them for small P and then comparing them with the first terms of the corresponding Maclaurin's series in P, expanded in powers of γ. Further agreement is

obtained by comparing the first terms of the series in γ with the functions of P which result from evaluating the Maclaurin's series for γ small.

5.4 *Symbols for Quantities*

$a \equiv \tau/\tau_v$, constant in recombination function.
$a_j \equiv$ coefficients in the Maclaurin's expansion of G in powers of P; j an integer.
$A_j \equiv$ coefficients in the expansion of G in powers of γ; j an integer.
$A^- \equiv$ concentration of ionized acceptors.
$A_0^- \equiv$ thermal-equilibrium concentration of ionized acceptors.
$b \equiv$ ratio of electron mobility to hole mobility.
$C \equiv I/I_0$, reduced total current density.
$C_\epsilon \equiv$ reduced emitter current.
$C^- \equiv$ reduced total current to the origin from the left.
$C^+ \equiv$ reduced total current from the origin to the right.
$C_n \equiv -I_n/I_0$, reduced electron flow density.
$C_p \equiv I_p/I_0$, reduced hole flow density.
$\gamma \equiv I/C$.
$\Gamma \equiv \epsilon/4\pi\sigma_0\tau$, reduced time for the dielectric relaxation of charge.
$D \equiv$ total donor concentration.
$D^+ \equiv$ concentration of ionized donors.
$D_0^+ \equiv$ thermal-equilibrium concentration of ionized donors.
$D_n \equiv kT\mu_n/e$, diffusion constant for electrons.
$D_p \equiv kT\mu_p/e$, diffusion constant for holes.
$e \equiv$ magnitude of the electronic charge.
$E \equiv$ electrostatic field.
$E_a \equiv$ applied or asymptotic field.
$E_0 \equiv kT/eL_p$, characteristic field.
$\epsilon \equiv$ dielectric constant.
$f \equiv$ fraction of total current carried by holes.
$f_\epsilon \equiv$ fraction of total current carried by holes at an emitter.
$f^- \equiv$ fraction of total current carried by holes at a source, to the left.
$f^+ \equiv$ fraction of total current carried by holes at a source, to the right.
$F \equiv E/E_0$, reduced electrostatic field.
$g_0 \equiv$ thermal rate of generation of hole-electron pairs, per unit volume.
$G \equiv dP/dX$, reduced concentration gradient.
$G^0 \equiv$ value of G for $X = 0$.
$G^- \equiv$ limiting value of G at a source, approached from the left.
$G^+ \equiv$ limiting value of G at a source, approached from the right.
$H \equiv$ probability of thermal ionization of an unionized donor, per unit time.

$I \equiv$ total current density.
$I_n \equiv$ current density of electrons.
$I_0 \equiv \sigma E_0$, characteristic current.
$I_p \equiv$ current density of holes.
$J \equiv \frac{1}{e} I$, total carrier flow density.
$J_n \equiv -\frac{1}{e} I_n$, electron flow density.
$J_p \equiv \frac{1}{e} I_p$, hole flow density.
$k \equiv$ Boltzmann's constant.
$K \equiv$ probability per unit time of electron capture by an ionized donor, per unit electron concentration.
$L_d \equiv (kT\epsilon/8\pi n_i e^2)^{\frac{1}{2}}$, characteristic length associated with space charge in the steady state.
$L_p \equiv (kT\mu\tau/e)^{\frac{1}{2}}$, diffusion length for holes for time τ.
$M \equiv 1 + \frac{b+1}{b} P$.
$M_0 \equiv 1 + \frac{b+1}{b} P_0$.
$\mu \equiv \mu_p \equiv$ mobility for holes.
$\mu_n \equiv$ mobility for electrons.
$n \equiv$ concentration of electrons.
$n_i \equiv$ thermal-equilibrium concentration of electrons (or holes) in the intrinsic semiconductor.
$n_0 \equiv$ thermal-equilibrium concentration of electrons.
$n_s \equiv$ saturation concentration excess of electrons, corresponding to complete ionization of donors.
$N \equiv n/(n_0 - p_0)$, reduced electron concentration for an n-type semiconductor.
$p \equiv$ concentration of holes.
$p_0 \equiv$ thermal-equilibrium concentration of holes.
$P \equiv p/(n_0 - p_0)$, reduced hole concentration for an n-type semiconductor.
$\Delta P \equiv (p - p_0)/(n_0 - p_0)$, reduced concentration of added holes.
$P_0 \equiv p_0/(n_0 - p_0)$, reduced hole concentration at thermal equilibrium.
$P^0 \equiv$ value of P for $X = 0$.
$Q \equiv \tau/\tau_p$, lifetime ratio.
$R \equiv$ general recombination function, equal to $1 + aP/(1 + P_0)$ for mass-action and constant-mean-lifetime mechanisms combined.
$\rho \equiv$ volume resistivity in ohm cm.
$s \equiv$ differential transport velocity.
$S \equiv s/(D_p/\tau)^{\frac{1}{2}}$, reduced differential transport velocity.

σ ≡ conductivity of semiconductor.
σ_0 ≡ normal conductivity of semiconductor, with no added carriers.
Σ ≡ $\sigma/\sigma_0 = M/M_0$, reduced conductivity of semiconductor.
t ≡ time variable.
T ≡ temperature in degrees absolute.
τ ≡ mean lifetime for holes for small added concentrations, in an *n*-type or in an intrinsic semiconductor.
τ_n ≡ mean lifetime for electrons (concentration-dependent).
τ_p ≡ mean lifetime for holes (concentration-dependent).
τ_v ≡ mean lifetime for holes, for small added concentrations in an *n*-type semiconductor, due to mass-action recombination alone.
U ≡ t/τ ≡ reduced time variable.
W ≡ eV/kT, reduced electrostatic potential.
x ≡ distance variable.
X ≡ x/L_p, reduced distance variable.
V ≡ electrostatic potential.

Traveling-Wave Tubes

By J. R. PIERCE

[FOURTH INSTALLMENT]

CHAPTER XII

POWER OUTPUT

A THEORETICAL EVALUATION of the power output of a traveling-wave tube requires a theory of the non-linear behavior of the tube. In this book we have dealt with a linearized theory only. No attempt will be made to develop a non-linear theory. Some results of non-linear theory will be quoted, and some conclusions drawn from experimental work will be presented.

One thing appears clear both from theory and from experiment: the gain parameter C is very important in determining efficiency. This is perhaps demonstrated most clearly in some unpublished work of A. T. Nordsieck.

Nordsieck assumed:

(1) The same a-c field acts on all electrons.
(2) The only fields present are those associated with the circuit ("neglect of space charge").
(3) Field components of harmonic frequency are neglected.
(4) Backward-traveling energy in the circuit is neglected.
(5) A lossless circuit is assumed.
(6) C is small (it always is).

Nordsieck obtained numerical solutions for such cases for several electron velocities. He found the maximum efficiency to be proportonal to C by a factor we may call k. Thus, the power output P is

$$P = kCI_0V_0 \tag{12.1}$$

In Fig. 12.1, the factor k is plotted vs. the velocity parameter b. For an electron velocity equal to that of the unperturbed wave the fractional efficiency obtained is $3C$; for a faster electron velocity the efficiency rises to $7C$. For instance, if $C = .025$, $3C$ is 7.5% and $7C$ is 15%. For 1,600 volts 15 ma this means 1.8 or 3.6 watts. If, however, $C = 0.1$, which is attainable, the indicated efficiency is 30% to 70%.

Experimental efficiencies often fall very far below such figures, although some efficiencies which have been attained lie in this range. There are three apparent reasons for these lower efficiencies. First, small non-uniformities in wave propagation set up new wave components which abstract energy from the increasing wave, and which may subtract from the normal output. Second, when the a-c field varies across the electron flow, not all electrons

are acted on equally favorably. Third, most tubes have a central lossy section followed by a relatively short output section. Such tubes may overload so severely in the lossy section that a high level in the output section is never attained. There is not enough length of loss-free circuit to provide sufficient gain in the output circuit so that the signal can build up to maximum amplitude from a low level increasing wave. Other tubes with distributed loss suffer because the loss cuts down the efficiency.

Some power-series non-linear calculations made by L. R. Walker show that for fast velocities of injection the first non-linear effect should be an expansion, not a compression. Nordsieck's numerical solutions agree with this. A power series approach is inadequate in dealing with truly large-signal be-

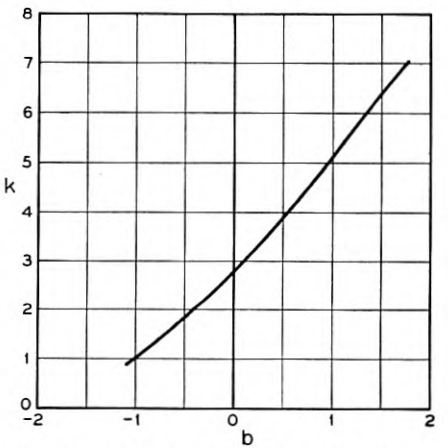


Fig. 12.1—The calculated efficiency is expressed as kC, where k is a function of the velocity parameter b. This curve shows k as given by Nordsieck's high-level calculations.

havior. In fact, Nordsieck's work shows that the power-series attack, if based on an assumption that there is no overtaking of electrons by electrons emitted later, must fail at levels much below the maximum output.

Further work by Nordsieck indicates that the output may be appreciably reduced by variation of the a-c field across the beam.

It is unfortunate that Nordsieck's calculations do not cover a wider range of conditions. Fortunately, unlikely as it might seem, the linear theory can tell us a little about what limitation of power we might expect. For instance, from (7.15) we have

$$\frac{v}{u_0} = -j\frac{\eta V}{u_0^2 \delta C}$$

$$\frac{v}{u_0} = -j\left(\frac{V}{2V_0}\right)\left(\frac{1}{\delta C}\right) \tag{12.2}$$

while from (7.16) we have

$$\frac{i}{I_0} = -\left(\frac{V}{2V_0}\right)\left(\frac{1}{\delta C}\right)^2 \tag{12.3}$$

We expect non-linear effects to become important when an a-c quantity is no longer small compared with a d-c quantity. We see that because $(1/\delta C)$ is large, $|\, i/I_0 \,|$ will be larger than $|\, v/u_0 \,|$.

The important non-linearity is a sort of over-bunching or limit to bunching. For instance, suppose we were successful in bunching the electron flow into very short pulses of electrons, as shown in Fig. 12.2 As the pulses approach zero length, the ratio of the peak value of the fundamental component of convection current to the average or d-c current I_0 approaches 2. We may, then, get some hint as to the variation of power output as various parameters are varied by letting $|\, i \,| = 2I_0$ and finding the variation of power in the circuit for an a-c convection current as we vary various parameters.

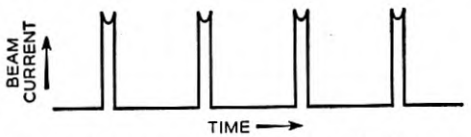


Fig. 12.2—If the electron beam were bunched into pulses short compared with a cycle, the peak value of the component of fundamental frequency would be twice the d-c current I_0.

Deductions made in this way cannot be more than educated guesses, but in the absence of non-linear calculations they are all we have.

From (7.1) we have for the *circuit field* associated with the *active mode* (neglecting the field due to space charge)

$$E = \frac{\Gamma^2 \Gamma_1 (E^2/\beta^2 P)}{2(\Gamma_1^2 - \Gamma^2)} \tag{12.4}$$

This relation is, of course, valid only for an electron convection current i which varies with distance as $\exp(-\Gamma z)$. For the power to be large for a given magnitude of current, E should be large. For a given value of i, E will be large if Γ is very nearly equal to Γ_1. This is natural. If Γ were equal to Γ_1, the natural propagation constant of the circuit, the contribution to the field by the current i in every elementary distance would have such phase as to add in phase with every other contribution.

Actually, Γ_1 and Γ cannot be quite equal. We have from (7.10) and (7.11)

$$-\Gamma_1 = \beta_e(-j - jCb - Cd) \tag{12.5}$$

$$-\Gamma = \beta_e(-j + jCy_1 + Cx_1) \tag{12.6}$$

For a physical circuit the attenuation parameter d must be positive while, for an increasing wave, x must be positive. We see that we may expect E to be greatest for a given current when d and x are small, and when y is nearly equal to the velocity parameter b.

Suppose we use (12.4) in expressing the power

$$P = \frac{E^2}{\beta^2(E^2/\beta^2 P)} = \left| \frac{\Gamma^4 \Gamma_1^2 (E^2/\beta^2 P)}{4\beta^2(\Gamma_1^2 - \Gamma^2)^2} i^2 \right|. \tag{12.7}$$

Here we identify β with $-j\Gamma_1$. Further, we use (2.43), (12.5) and (12.6), and assuming C to be small, neglect terms involving C compared with unity. We will further let i have a value

$$i = 2I_0 \tag{12.8}$$

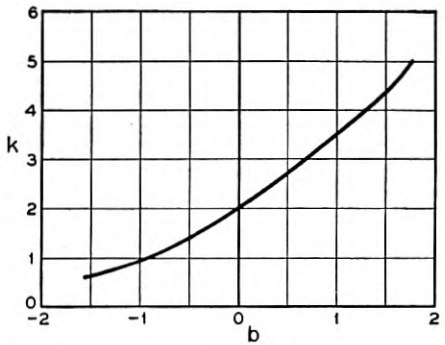


Fig. 12.3—An efficiency parameter k calculated by taking the power as that given by near theory for an r-f beam current with a peak value twice the d-c beam current.

We obtain

$$P = kCI_0V_0 \tag{12.9}$$

$$k = \frac{2}{(b + y)^2 + (x + d)^2} \tag{12.10}$$

We will now investigate several cases. Let us consider first the case of a lossless circuit ($d = 0$) and no space charge ($QC = 0$) and plot the efficiency factor k vs. b. The values of x and y are those of Fig. 8.1. Such a plot is shown in Fig. 12.3.

If we compare the curve of Fig. 12.3 with the correct curve of Nordsieck, we see that there is a striking qualitative agreement and, indeed, fair quantitative agreement. We might have expected on the one hand that the electron stream would never become completely bunched ($i = 2I_0$) and that, as it approached complete bunching, behavior would already be non-linear. This would tend to make (12.10) optimistic. On the other hand, even after i

attains its maximum value and starts to fall, power can still be transferred to the circuit, though the increase of field with distance will no longer be exponential. This makes it possible that the value of k given by (12.10) will be exceeded. Actually, the true k calculated by Nordsieck is a little higher than that given by (12.10).

Let us now consider the effect of loss. Figure 12.4 shows k from (12.10) vs. d for $b = QC = 0$. We see that, as might be expected, the efficiency falls as the loss is increased. C. C. Cutler has shown experimentally through unpublished work that the power actually falls off much more rapidly with d.

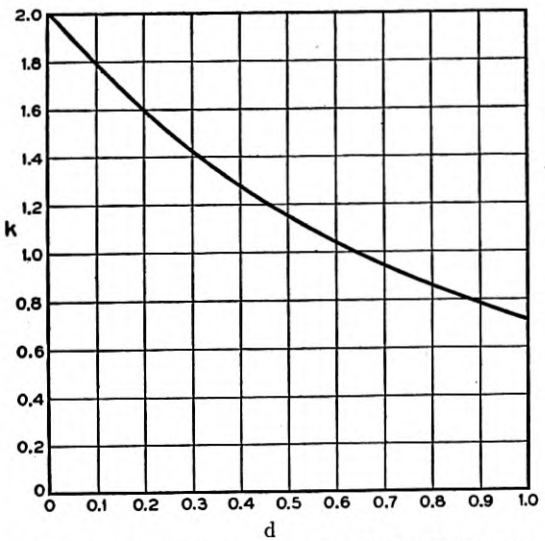


Fig. 12.4—The efficiency parameter k calculated as in Fig. 12.3 but for $b = 0$ (an electron velocity equal to the circuit phase velocity) and for various values of the attenuation parameter d. Experimentally, the efficiency falls off more rapidly as d is increased.

Finally, Fig. 12.5 shows k from (12.10) vs. QC, with $d = 0$ and b chosen to make x_1 a maximum. We see that there is a pronounced rise in efficiency as the space-charge parameter QC is increased.

J. C. Slater has suggested in Microwave Electronics a way of looking at energy production essentially based on observing the motions of electrons while traveling along with the speed of the wave. He suggests that the electrons might eventually be trapped and oscillate in the troughs of the sinusoidal field. If so, and if they initially have an average velocity Δv greater than that of the wave, they cannot emerge with a velocity lower than the velocity of the wave less Δv. Such considerations are complicated by the fact that the phase velocity of the wave in the large-signal region will not

be the same as its phase velocity in the small-signal region. It is interesting, however, to see what limiting efficiencies this leads to.

The initial electron velocity for the increasing wave is approximately

$$v_a = v_c(1 - y_1 C) \tag{12.11}$$

where v_c is the phase velocity of the wave in the absence of electrons. The quantity y_1 is negative. According to Slater's reckoning, the final electron velocity cannot be less than

$$v_b = v_c(1 + y_1 C) \tag{12.12}$$

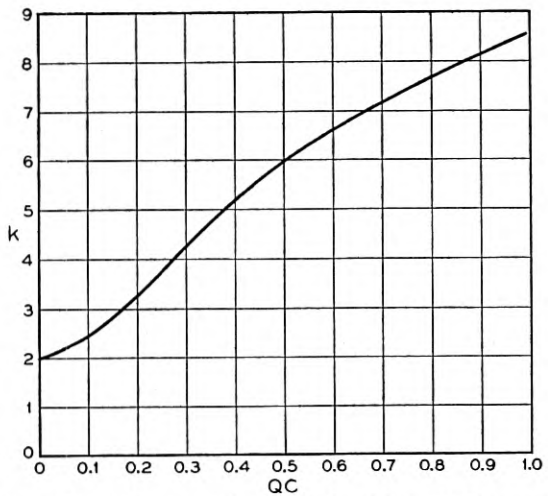


Fig. 12.5—The efficiency parameter k calculated as in Fig. 12.3, for zero loss and for an electron velocity which makes the gain of the increasing wave greatest, vs the space-charge parameter QC.

The limiting efficiency η accordingly will be, from considerations of kinetic energy

$$\eta = \frac{v_a^2 - v_b^2}{v_a^2}$$

$$\eta = \frac{4y_1 C}{(1 - y_1 C)^2}.$$

If $y_1 C \ll 1$, very nearly

$$\eta = 4\, y_1 C \tag{12.13}$$

We see that this also indicates an efficiency proportional to C. In Fig. 12.6 $4y_1$ is plotted vs. b for $QC = d = 0$. We see that this quantity ranges

from 2 for $b = 0$ up to 5 for larger values of b. It is surprising how well this agrees with corresponding values of 3 and 7 from Nordsieck's work. Moreover (12.13) predicts an increase in efficiency with increasing QC.

Thus, we may expect the efficiency to vary with C from several points of view.

It is interesting to consider what happens if at a given frequency we change the current. By changing the current while holding the voltage constant we increase both the input power and the efficiency, for C varies as $I_0^{1/3}$. Thus, in changing the current alone we would expect the power to vary as the 4/3 power of I_0

$$P \approx I_0^{4/3} \tag{12.14}$$

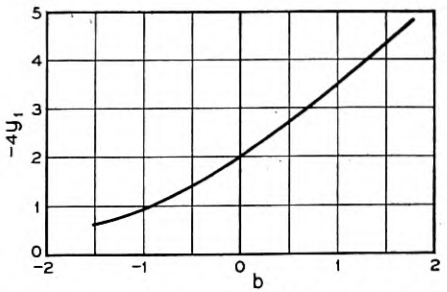


Fig. 12.6—According to a suggestion made by Slater, the velocity by which the electrons are slowed down cannot be greater than twice the difference between the electron velocity and the wave velocity. If we use the velocity difference given by the linear theory, for zero loss ($d = 0$) this would make the efficiency parameter k equal to $-4y_1$. Here $-4y_1$ is plotted vs b for $QC = 0$.

Here space charge has been neglected, and actually power may increase more rapidly with current than (12.14) indicates.

A variety of other cases can be considered. At a given voltage and current, C and the efficiency rise as the helix diameter is made smaller. However, as the helix diameter is made smaller it may be necessary to decrease the current, and the optimum gain will come at higher frequencies. For a given beam diameter, the magnetic focusing field required to overcome space-charge repulsion is constant if $I_0/V_0^{1/2}$ is held constant, and hence we might consider increasing the current as the 1/2 power of the voltage, and thus increasing the power input as the 3/2 power of the voltage. On the other hand, the magnetic focusing field required to correct initial angular deflections of electrons increases as the voltage is raised.

There is no theoretical reason why electrons should strike the circuit. Thus, it is theoretically possible to use a very high beam power in connection with a very fragile helix. Practically, an appreciable fraction of the beam current is intercepted by the helix, and this seems unavoidable for wave

lengths around a centimeter or shorter, for accurate focusing becomes more difficult as tubes are made physically smaller. Thus, in getting very high powers at ordinary wavelengths or even moderate powers at shorter wavelengths, filter type circuits which provide heat dissipation by thermal conduction may be necessary. We have seen that the impedance of such circuits is lower than that of a helix for the broadband condition (group velocity equal to phase velocity). However, high impedances and hence large values of C can be attained at the expense of bandwidth by lowering the group velocity. This tends to raise the efficiency, as do the high currents which are allowable because of good heat dissipation. However, lowering group velocity increases attenuation, and this will tend to reduce efficiency somewhat.

It has been suggested that the power can be increased by reducing the phase velocity of the circuit near the output end of the tube, so that the electrons which have lost energy do not fall behind the waves. This is a complicated but attractive possibility. It has also been suggested that the electrode which collects electrons be operated at a voltage lower than that of the helix.

The general picture of what governs and limits power output is fairly clear as long as C is very small. If attenuation near the output of the tube is kept small, and the circuit is constructed so as to approximate the requirement that nearly the same field acts on all electrons, efficiencies as large as 40% are indicated within the limitations of the present theory. With larger values of C it is not clear what the power limitation will be.

The usual traveling-wave tube would seem to have a serious competitor for power applications in the traveling-wave magnetron amplifier, which is discussed briefly in a later chapter.

CHAPTER XIII

TRANSVERSE MOTION OF ELECTRONS

Synopsis of Chapter

SO FAR WE HAVE taken into account only longitudinal motions of electrons. This is sufficient if the transverse fields are small compared to the longitudinal fields (as, near the axis of an axially symmetrical circuit) or, if a strong magnetic focusing field is used, so that transverse motions are inhibited. It is possible, however, to obtain traveling-wave gain in a tube in which the longitudinal field is zero at the mean position of the electron beam. For a slow wave, the electric field is purely transverse only along a plane. The transverse field in this plane forces electrons away from the plane and preferentially throws them into regions of retarding field, where they give up energy to the circuit. This mechanism is not dissimilar to that in the longitudinal field case, in which the electrons are moved longitudinally from their unperturbed positions, preferentially into regions of more retarding field.

Whatever may be said about tubes utilizing transverse fields, it is certainly true that they have been less worked on than longitudinal-field tubes. In view of this, we shall present only a simple analysis of their operation along the lines of Chapter II. In this analysis we take cognizance of the fact that the charge induced in the circuit by a narrow stream of electrons is a function not only of the charge per unit length of the beam, but of the distance between the beam and the circuit as well.

The factor of proportionality between distance and induced charge can be related to the field produced by the circuit. Thus, if the variation of V in the x, y plane (normal to the direction of propagation) is expressed by a function Φ, as in (13.3), the effective charge ρ_E is expressed by (13.8) and, if y is the displacement of the beam normal to the z axis, by (13.9) where Φ' is the derivative of Φ with respect to y.

The equations of motion used must include displacements normal to the z direction; they are worked out including a constant longitudinal magnetic focusing field. Finally, a combined equation (13.23) is arrived at. This is rewritten in terms of dimensionless parameters, neglecting some small terms, as (13.26)

$$j\delta - b = \frac{1}{\delta^2} + \frac{\alpha^2}{(\delta^2 + f^2)}.$$

Here δ and b have their usual meanings; α is the ratio between the transverse and longitudinal field strengths, and f is proportional to the strength of the magnetic focusing field.

In case of a purely transverse field, a new gain parameter D is defined. D is the same as C except that the longitudinal a-c field is replaced by the transverse a-c field. In terms of D, b and δ are redefined by (13.36) and (13.37), and the final equation is (13.38). Figures 13.5–13.10 show how the x's and y's vary with b for various values of f (various magnetic fields) and Fig. 13.11 shows how x_1, which is proportional to the gain of the increasing wave in db per wavelength, decreases as magnetic field is increased. A numerical example shows that, assuming reasonable circuit impedance, a magnetic field which would provide a considerable focusing action would still allow a reasonable gain.

The curves of Figs. 13.6–13.10 resemble very much the curves of Figs. 8.7–8.9 of Chapter VIII, which show the effect of space charge in terms of the parameter QC. This is not unnatural; in one case space charge forces tend to return electrons which are accelerated longitudinally to their undisturbed positions. In the other case, magnetic forces tend to return electrons which are accelerated transversely to their undisturbed positions. In each case the circuit field acts on an electron stream which can itself sustain oscillations. In one case, the oscillations are of a plasma type, and the restoring force is caused by space charge of the bunched electron stream; in the other case the electrons can oscillate transversely in the magnetic field with cyclotron frequency.

Let us, for instance, compare (7.13), which applies to purely longitudinal displacements with space charge, with (13.38), which applies to purely transverse fields with a longitudinal magnetic field. For zero loss ($d = 0$), (7.13) becomes

$$1 = (j\delta - b)(\delta^2 + 4QC)$$

While

$$1 = (j\delta - b)(\delta^2 + f^2) \tag{13.38}$$

describes the transverse case. Thus, if we let

$$4QC = f^2$$

the equations are identical.

When there is both a longitudinal and a transverse electric field, the equation for δ is of the fifth degree. Thus, there are five forward waves. For an electron velocity equal to the circuit phase velocity ($b = 0$) and for no attenuation, the two new waves are unattenuated.

If there is no magnetic field, the presence of a transverse field component merely adds to the gain of the increasing wave. If a small magnetic field is

imposed in the presence of a transverse field component, this gain is somewhat reduced.

13.1 Circuit Equation

Consider a tubular electrode connected to ground through a wire, shown in Fig. 13.1. Suppose we bring a charge Q into the tube from ∞. A charge Q will flow to ground through the wire. This is the situation assumed in the analysis of Chapter II. In Fig. 2.3 it is assumed that all the lines of force from the charge in the electron beam terminate on the circuit, so that the whole charge may be considered as impressed on the circuit.

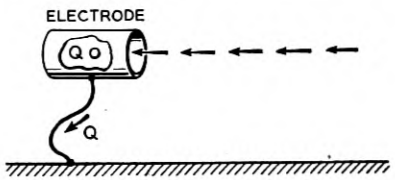


Fig. 13.1—When a charge Q approaches a grounded conductor from infinity and in the end all the lines of force from the charge end on the conductor, a charge Q flows in the grounding lead.

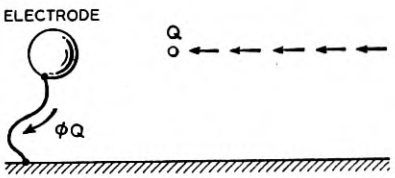


Fig. 13.2—If a charge Q approaches a conductor from infinity but in the end only part of the lines of force from the charge end on the conductor, a charge ΦQ flows in the grounding lead, where $\Phi < 1$.

Now consider another case, shown in Fig. 13.2, in which a charge Q is brought from ∞ to the vicinity of a grounded electrode. In this case, not all of the lines of force from the charge terminate on the electrode, and a charge ΦQ which is smaller than Q flows through the wire to ground.

We can represent the situation of Fig. 13.2 by the circuit shown in Fig. 13.3. Here C_2 is the capacitance between the charge and the electrode and C_1 is the capacitance between the charge and ground. We see that the charge ΦQ which flows to ground when a charge Q is brought to a is

$$\Phi Q = QC_2/(C_1 + C_2) \tag{13.1}$$

Now suppose we take the charge Q away and hold the electrode at a potential V with respect to ground, as shown in Fig. 13.4. What is the potential V_a at a? We see that it is

$$V_a = [C_2/(C_1 + C_2)]V = \Phi V \tag{13.2}$$

Thus, the same factor Φ relates the actual charge to the "effective charge" acting on the circuit and the actual circuit voltage to the voltage produced at the location of the charge.

We will not consider in this section the "space charge" voltage produced by the charge itself (the voltage at point a in Fig. 13.4).

The circuit voltage V we consider as varying as $\exp(-\Gamma z)$ in the direction of propagation. The voltage in the vicinity of the circuit is given by

$$V(x, y) = \Phi V \tag{13.3}$$

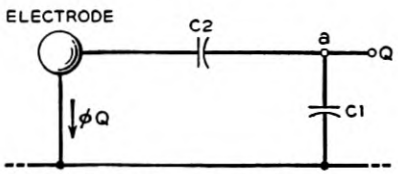


Fig. 13.3—The situation of Fig. 13.2 results in the same charge flow as if the charge were put on terminal a of the circuit shown, which consists of two capacitors of capacitances C_1 and C_2.

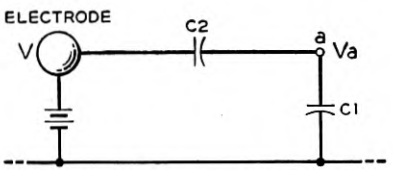


Fig. 13.4—A voltage V inserted in the ground lead divides across the condensers so that $V_a = \Phi V$, where Φ is the same factor which relates the charge flowing in the ground lead to the charge Q applied at a in Figs. 13.2 and 13.3.

Here x and y refer to coordinates normal to z and Φ is a function of x and y. We will choose x and y so

$$\partial\Phi/\partial x = 0 \tag{13.4}$$

Then

$$E_y = -V\partial\Phi/\partial y = -\Phi' V \tag{13.5}$$

$$\Phi' = \partial\Phi/\partial y \tag{13.6}$$

In (13.3), Φ will vary somewhat with Γ, but, as we are concerned with a small range only in Γ, we will consider Φ a function of y only.

From Chapter II we have

$$V = \frac{-\Gamma\Gamma_1 Ki}{(\Gamma^2 - \Gamma_1^2)} \tag{2.10}$$

and

$$\rho = \frac{-j\Gamma i}{\omega}. \tag{2.18}$$

So that

$$V = \frac{-j\omega\Gamma_1 K\rho}{(\Gamma^2 - \Gamma_1^2)}. \tag{13.7}$$

In (13.7), it is assumed that $\Phi = 1$. If $\Phi \neq 1$, we should replace ρ in (13.7) by the a-c component of effective charge. The total effective charge ρ_E is

$$\rho_E = \Phi(\rho + \rho_0) \tag{13.8}$$

The term ρ_0 is included because Φ will vary if the y-position of the charge varies. To the first order, the a-c component ρ_E of the effective charge is,

$$\rho_E = \Phi\rho + \rho_0\Phi' y \tag{13.9}$$

$$\rho_E = \Phi\rho - (I_0/u_0)\Phi' y \tag{13.9}$$

Here y is the a-c variation in position along the y coordinate. Thus, if $\Phi \neq 0$, we have instead of (13.7)

$$V = \frac{-j\omega\Gamma_1 K(\Phi\rho - (I_0/u_0)\Phi' y)}{(\Gamma^2 - \Gamma_1^2)}. \tag{13.10}$$

This is the circuit equation we shall use.

13.2 Ballistic Equations

We will assume an unperturbed motion of velocity u_0 in the z direction, parallel to a uniform magnetic focusing field of strength B. As in Chapter II, products of a-c quantities will be neglected.

In the x direction, perpendicular to the y and z directions

$$d\dot{x}/dt = -\eta B\dot{y} \tag{13.11}$$

Assume that $\dot{x} = 0$ at $y = 0$. Then

$$\dot{x} = \eta B y \tag{13.12}$$

In the y direction we have

$$d\dot{y}/dt = \eta(B\dot{x} - E_y) \tag{13.13}$$

From (13.5) this is

$$d\dot{y}/dt = \eta(B\dot{x} + \Phi' V) \tag{13.14}$$

$$d\dot{y}/dt = \partial\dot{y}/\partial t + (\partial\dot{y}/\partial z)(dz/dt) \tag{13.15}$$

$$(d\dot{y}/dt) = u_0(j\beta_e - \Gamma)\dot{y} \tag{13.16}$$

We obtain from (13.16), (13.14) and (13.12)

$$(j\beta_e - \Gamma)y = -u_0\beta_m^2 y + \eta\Phi' V/u_0 \tag{13.17}$$

$$\beta_m = \eta B/u_0 \tag{13.18}$$

Here ηB is the cyclotron radian frequency and β_m is a corresponding propagation constant.

Now

$$\dot{y} = \partial y/\partial t - (\partial y/\partial z)(\partial z/\partial t) \tag{13.19}$$

$$\dot{y} = u_0(j\beta_e - \Gamma)y \tag{13.20}$$

From (13.20) and (13.17) we obtain

$$y = \frac{\Phi' V}{2V_0[(j\beta_e - \Gamma)^2 + \beta_m^2]}. \tag{13.21}$$

It is easily shown that the equation for ρ can be obtained exactly as in Chapter II. From (2.22) and (2.18) we have

$$\rho = \frac{I_0 \Gamma^2 \Phi V}{2u_0 V_0(j\beta_e - \Gamma)^2}. \tag{13.22}$$

13.3 Combined Equation

From the circuit equation (13.10) and the ballistical equations (13.21) and (13.22) we obtain

$$1 = \frac{-j\beta_e \Gamma_1 \Gamma^2 \Phi^2 K I_0}{2V_0(\Gamma^2 - \Gamma_1^2)}\left[\frac{1}{(j\beta_e - \Gamma)^2} - \frac{(\Phi'/\Phi)^2}{\Gamma^2[(j\beta_e - \Gamma)^2 + \beta_m^2]}\right]. \tag{13.23}$$

The voltage at the beam is Φ times the circuit voltage, so the effective impedance of the circuit at the beam is Φ^2 times the circuit impedance. Thus

$$C^3 = \Phi^2 K I_0/4V_0 \tag{13.24}$$

It will be convenient to define a dimensionless parameter f specifying β_m and hence the magnetic field

$$f = \beta_m/\beta_e C \tag{13.25}$$

We will also use δ and b as defined earlier

$$-\Gamma = -j\beta_e + \beta_e C\delta$$

$$-\Gamma_1 = -j\beta_e - j\beta_e C b$$

After the usual approximations, (13.23) yields

$$j\delta - b = \frac{1}{\delta^2} + \frac{\alpha^2}{(\delta^2 + f^2)} \tag{13.26}$$

$$\alpha^2 = (\Phi'/\beta_e\Phi)^2 \tag{13.27}$$

It is interesting to consider the quantity $(\Phi'/\beta_e\Phi)^2$ for typical fields. For

instance, in the two-dimensional electrostatic field in which the potential V is given by

$$V = Ae^{-\beta_e y}e^{-j\beta_e z} \tag{13.28}$$

$$\partial V/\partial y = -\beta_e V \tag{13.29}$$

and everywhere

$$\alpha^2 = (\Phi'/\beta_e\Phi)^2 = 1. \tag{13.30}$$

Relation (13.30) is approximately true far from the axis in an axially symmetrical field.

Consider a potential giving a purely transverse field at $y = 0$

$$V = Ae^{-j\beta_e z} \sinh \beta_e y \tag{13.31}$$

$$\frac{\partial V}{\partial y} = \beta A e^{-j\beta_e z} \cosh \beta_e y. \tag{13.32}$$

In this case, at $y = 0$

$$\alpha^2 = (\Phi'/\beta_e\Phi)^2 = \infty \tag{13.33}$$

In the case of a purely transverse field we let

$$D^3 = \frac{I_0\Phi'^2 K}{4V_0\beta_e^2} \tag{13.34}$$

$$D^3 = (E_y^2/\beta^2 P)(I_0/8V_0) \tag{13.35}$$

In (13.35), E_y is the magnitude of the y component of field for a power flow P, and β is the phase constant.

We then redefine δ and b in terms of D rather than C

$$-\Gamma = -j\beta_e + \beta_e D\delta \tag{13.36}$$

$$-\Gamma_1 = -j\beta_e - j\beta_e Db \tag{13.37}$$

and our equation for a purely transverse field becomes

$$1 = (j\delta - b)(\delta^2 + f^2) \tag{13.38}$$

In (13.38), δ and b are of course not the same as in (13.26) but are defined by (13.36) and (13.37).

13.4 Purely Transverse Fields

The case of purely transverse fields is of interest chiefly because, as was mentioned in Chapter X, it has been suggested that such tubes should have low noise.

In terms of x and y as usually defined

$$\delta = x + jy$$

equation (13.38) becomes

$$x[(x^2 - y^2 + f^2) - 2y(y + b)] = 0 \tag{13.39}$$

$$(y + b)(x^2 - y^2 + f^2) + 2x^2y + 1 = 0 \tag{13.40}$$

From the $x = 0$ solution of (13.39) we obtain

$$x = 0 \tag{13.41}$$

$$b = \frac{1}{y^2 - f^2} - y. \tag{13.42}$$

It is found that this solution obtains for large and small values of b. For very large and very small values of b, either

$$y \doteq -b \tag{13.43}$$

or

$$y \doteq \pm f \tag{13.44}$$

The wave given by (13.43) is a circuit wave; that given by (13.44) represents electrons traveling down the tube and oscillating with the cyclotron frequency in the magnetic field.

In an intermediate range of b, we have from (13.39)

$$x = \pm\sqrt{2y(y + b) - (f^2 - y^2)} \tag{13.45}$$

and

$$b = -2y \pm \sqrt{f^2 - 1/2y}. \tag{13.46}$$

For a given value of f^2 we can assume values of y and obtain values of b. Then, x can be obtained from (13.45). In Figs. 13.5–13.10, x and y are plotted vs. b for $f^2 = 0$, .5, 1, 4 and 10. It should be noted that x_1, the parameter expressing the rate of increase of the increasing wave, has a maximum at larger values of b as f is increased (as the magnetic focusing field is increased). Thus, for higher magnetic focusing fields the electrons must be shot into the circuit faster to get optimum results than for low fields. In Fig. 13.11, the maximum positive value of x is plotted vs. f. The plot serves to illustrate the effect on gain of increasing the magnetic field.

Let us consider an example. Suppose

$$\lambda = 7.5 \text{ cm}$$

$$D = .03$$

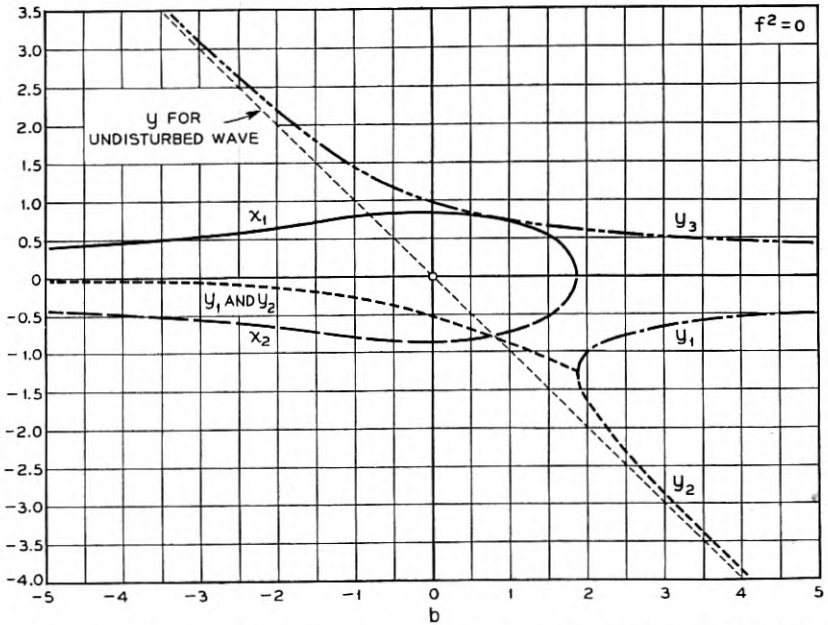


Fig. 13.5—The x's and y's for the three forward waves when the circuit field is purely transverse at the thin electron stream, for zero magnetic focusing field ($f^2 = 0$).

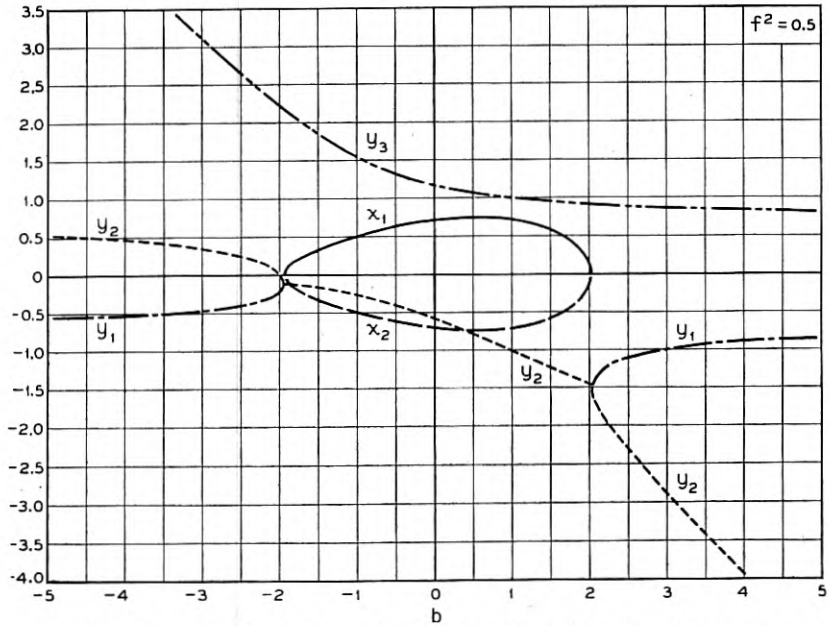


Fig. 13.6—Curves similar to those of Fig. 13.5 for a parameter $f^2 = 1$. The parameter f is proportional to the strength of the magnetic focusing field.

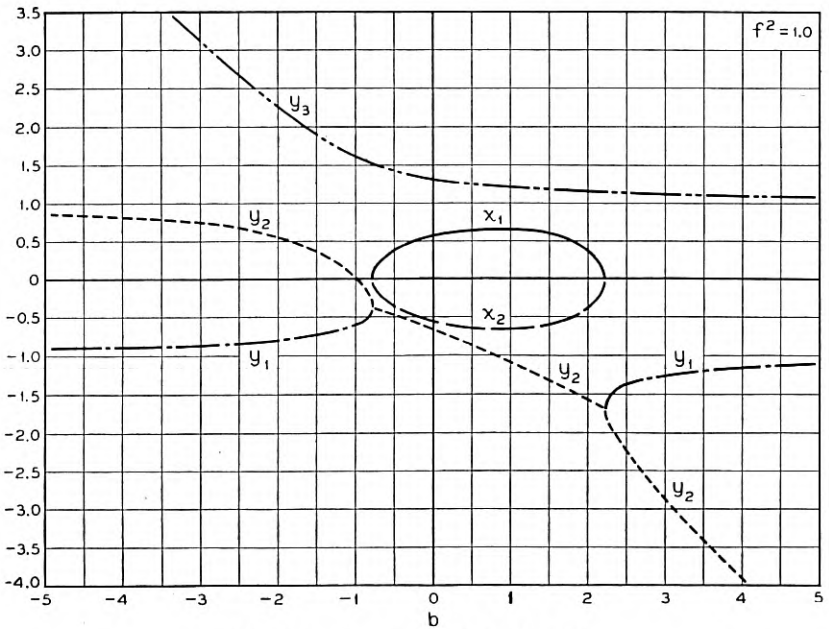


Fig. 13.7—The x's and y's for $f^2 = 1.0$.

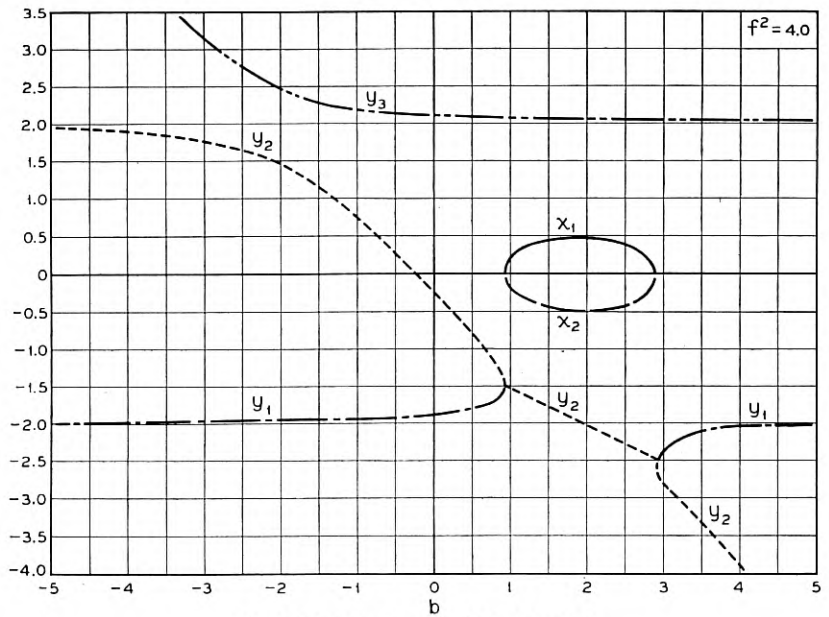


Fig. 13.8—The x's and y's for $f^2 = 2.0$.

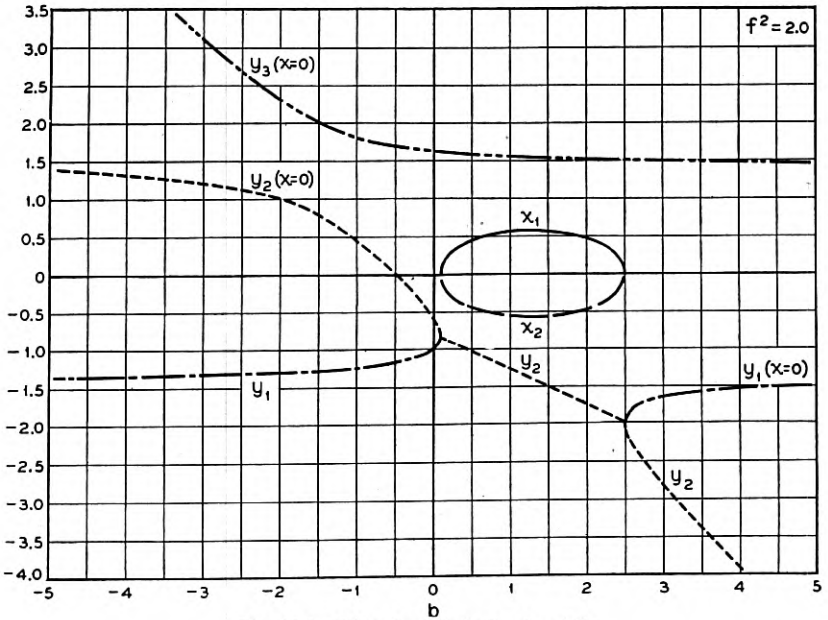


Fig. 13.9—The x's and y's for $f = 4.0$.

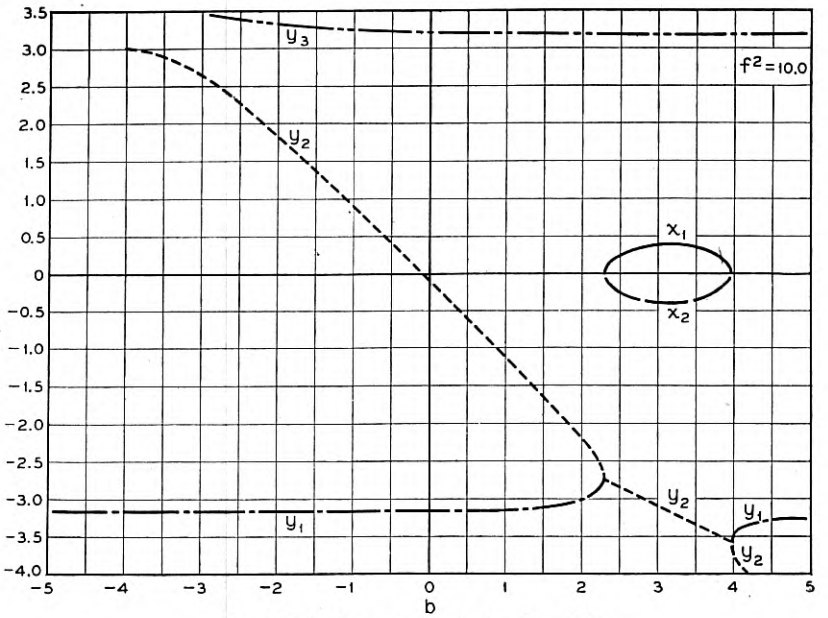


Fig. 13.10—The x's and y's for $f^2 = 10.0$.

These values are chosen because there is a longitudinal field tube which operates at 7.5 cm with a value of C (which corresponds to D) of about .03. The table below shows the ratio of the maximum value of x_1 to the maximum value of x_1 for no magnetic focusing field.

Magnestic Field in Gauss	f	x_1/x_{10}
0	0	1
50	1.17	.71
100	2.34	.50

A field of 50 to 100 gauss should be sufficient to give useful focusing action. Thus, it may be desirable to use magnetic focusing fields in transverse-

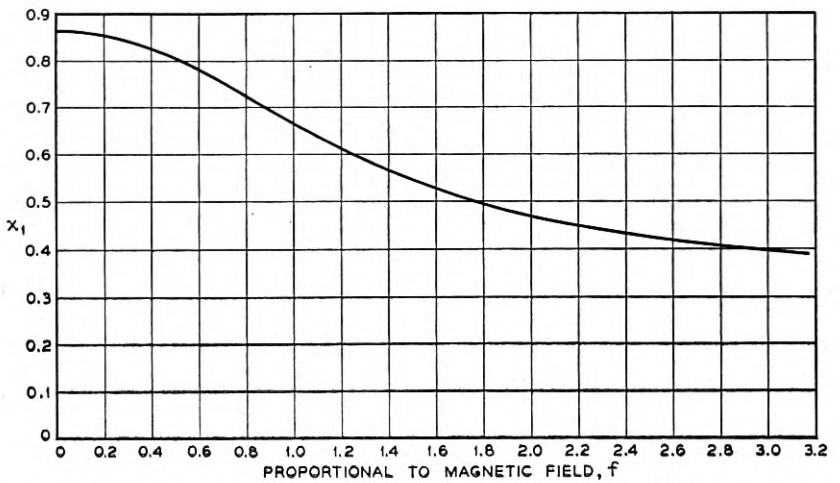


Fig. 13.11—Here x_1, the x for the increasing wave, is plotted vs f, which is proportional to the strength of the focusing field. The velocity parameter b has been chosen to maximize x_1. The ordinate x_1 is proportional to gain per wavelength.

field traveling-wave tubes. This will be more especially true in low-voltage tubes, for which D may be expected to be higher than .03.

13.5 Mixed Fields

In tubes designed for use with longitudinal fields, the transverse fields far off the axis approach in strength the longitudinal fields. The same is true of transverse field tubes far off the axis. Thus, it is of interest to consider equation (13.26) for cases in which α is neither very small nor very large, but rather is of the order of unity.

If the magnetic field is very intense so that f^2 is large, then the term containing α^2, which represents the effect of transverse fields, will be very small and the tube will behave much as if the transverse fields were absent.

Consideration of both terms presents considerable difficulty as (13.26) leads to five waves (5 values of δ) instead of three. The writer has attacked the problem only for the special case of $b = 0$. In this case we obtain from (13.26)

$$\delta = -j\left[\frac{1}{\delta^2} + \frac{\alpha^2}{\delta^2 + f^2}\right] \tag{13.47}$$

MacColl has shown[1] that the two "new" waves (waves introduced when $\alpha = 0$) are unattenuated and thus unimportant and uninteresting (unless, as an off-chance, they have some drastic effect in fitting the boundary conditions).

Proceeding from this information, we will find the change in δ as f^2 is increased from zero. From (13.47) we obtain

$$d\delta = j\left[\frac{2d\delta}{\delta^3} + \frac{2\alpha^2\,\delta d\delta}{(\delta^2 + f^2)^2} + \frac{\alpha^2\,df^2}{(\delta^2 + f^2)^2}\right] \tag{13.48}$$

Now, if $f = 0$

$$\delta^3 = -j(1 + \alpha^2) \tag{13.49}$$

If we use this in connection with (13.48) we obtain

$$d\delta = -\frac{\alpha^2}{3\delta}\,df^2 \tag{13.50}$$

For an increasing wave

$$\delta_1 = (1 + (\Phi'/\beta_e\Phi)^2)(\sqrt{3}/2 - j/2) \tag{13.51}$$

Hence, for the increasing wave

$$d\delta_1 = \frac{\alpha^2(-\sqrt{3}/2 - j/2)}{3(1 + \alpha^2)}\,df^2 \tag{13.52}$$

This shows that applying a small magnetic field tends to decrease the gain. This does not mean, however, that the gain with a longitudinal and transverse field and a magnetic field is less than the gain with the longitudinal field alone. To see this we assume that not f^2 but $(\Phi'/\beta_e\Phi)^2$ is small. Differentiating, we obtain

$$d\delta = -j\left[-\frac{2d\delta}{\delta^3} - \frac{2\alpha^2\,\delta d\delta}{(\delta^2 + f^2)^2} + \frac{d\alpha^2}{\delta^2 + f^2}\right] \tag{13.53}$$

If $\alpha = 0$

$$\delta^3 = -j \tag{13.54}$$

[1] J. R. Pierce, "Transverse Fields in Traveling-Wave Tubes," *Bell System Technical Journal*, Vol. 27, pp. 732–746.

and we obtain

$$d\delta = \frac{1}{3}\frac{\delta^3}{(\delta^2 + f^2)}\,d\alpha^2 \tag{13.55}$$

$$d\delta = \frac{-j}{3(\delta^2 + f^2)}\,d\alpha^2 \tag{13.56}$$

If we have a very large magnetic field ($f^2 \gg |\,\delta^2\,|$), then

$$d\delta = \frac{-j}{3f^2}\,d\alpha^2 \tag{13.57}$$

and the change in δ is purely reactive. If $f = 0$ (no magnetic field), from (13.55)

$$d\delta = \frac{\delta}{3}\,d\alpha^2 \tag{13.58}$$

Adding a transverse field component increases the magnitude of δ without changing the phase angle.

CHAPTER XIV

FIELD SOLUTIONS

Synopsis of Chapter

SO FAR, it has been assumed that the same a-c field acts on all electrons. This has been very useful in getting results, but we wonder if we are overlooking anything by this simplification.

The more complicated situation in which the variation of field over the electron stream is taken into account cannot be investigated with the same generality we have achieved in the case of "thin" electron streams. The chief importance we will attach to the work of this chapter is not that of producing numerical results useful in designing tubes. Rather, the chapter relates the appropriate field solutions to those we have been using and exhibits and evaluates features of the "broad beam" case which are not found in the "thin beam" case.

To this end we shall examine with care the simplest system which can reasonably be expected to exhibit new features. The writer believes that this will show qualitatively the general features of most or all "broad beam" cases.

The case is that of an electron stream of constant current density completely filling the opening of a double finned circuit structure, as shown in Fig. 14.1. The susceptance looking into the slots between the fins is a function of frequency only and not of propagation constant. Thus, at a given frequency, we can merely replace the slotted circuit members by susceptance sheets relating the magnetic field to the electric field, as shown in Fig. 14.2. The analysis is carried out with this susceptance as a parameter. Only the mode of propagation with a symmetrical field pattern is considered.

First, the case for zero current density is considered. The natural mode of propagation will have a phase constant β such that H_x/E_z for the central region is the same as H_x/E_z for the finned circuit. The solid curve of Fig. 14.3 shows a quantity proportional to H_x/E_z for the central space vs $\theta = \beta d$ (d defined by Fig. 14.1), a quantity proportional to β. The dashed line P represents H_x/E_z for a given finned structure. The intersections specify values of θ for the natural active modes of propagation to the left and to the right, and, hence, values of the natural phase constants.

The structure also has passive modes of propagation. If we assume fields which vary in the z direction as exp $(\Phi/d)z$, H_x/E_z for the central

opening varies with Φ as shown in part in Fig. 14.4. A horizontal line representing a given susceptance of the finned structure will intersect the curve at an infinite number of points. Each intersection represents a passive mode which decays at a particular rate in the z direction and varies sinusoidally with a particular period in the y direction.

If the effect of the electrons in the central space is included, H_z/E_z for the central space no longer varies as shown in Fig. 14.3, but as shown in Fig. 14.5 instead. The curve goes off to $+\infty$ near a value of θ corresponding to a phase velocity near to the electron velocity. The nature of the modes depends on the susceptance of the finned structure. If this is represented by P_1, there are four unattenuated waves; for P_3 there are two unattenuated waves and an increasing and a decreasing wave. P_2 represents a transitional case.

Not the whole of the curve for the central space is shown on Fig. 14.5. In Fig. 14.6 we see on an expanded scale part of the region about $\theta = 1$, between the points where the curve goes through 0. The curve goes to $+\infty$ and repeatedly from $-\infty$ to $+\infty$, crossing the axis an infinite number of times as θ approaches unity. For any susceptance of the finned structure, this leads to an infinite number of unattenuated modes, which are space-charge waves; for these the amplitude varies sinusoidally with different periods across the beam. Not all of them have any physical meaning, for near $\theta = 1$ the period of cyclic variation across the beam will become small even compared to the space between electrons.

Returning to Fig. 14.1, we may consider a case in which the central space between the finned structures is very narrow (d very small). This will have the effect of pushing the solid curve of Fig. 14.5 up toward the horizontal axis, so that for a reasonable value of P (say, P_1, P_2 or P_3 of Fig. 14.5) there is no intersection. That is, the circuit does not propagate any unattenuated waves. In this case there are still an increasing and a decreasing wave. The behavior is like that of a multi-resonator klystron carried to the extreme of an infinite number of resonators. If we add resonator loss, the behavior of gain per wavelength with frequency near the resonant frequency of the slots is as shown in Fig. 14.7.

One purpose of this treatment of a broad electron stream is to compare its results with those of the previous chapters. There, the treatment considered two aspects separately: the circuit and the effect of the electrons.

Suppose that at $y = d$ in Fig. 14.1 we evaluate not H_x for the finned structure and for the central space separately, but, rather, the difference or discontinuity in H_x. This can be thought of as giving the driving current necessary to establish the field E_z with a specified phase constant. In Fig. 14.8, y_1 is proportional to this H_x or driving current divided by E_z. The dashed curve y_2 is the variation of driving current with θ or β which we have

used in earlier chapters, fitted to the true curve in slope and magnitude at $y = 0$. Over the range of θ of interest in connection with increasing waves, the fit is good.

The difference between H_x/E_z for the central space without electrons (Fig. 14.3) and H_x/E_z for the central space with electrons (Fig. 14.5) can be taken as representing the driving effect of the electrons. The solid curve of Fig. 14.9 is proportional to this difference, and hence represents the true effect of the electrons. The dashed curve is from the ballistical equation used in previous chapters. This has been fitted by adjusting the space-charge parameter Q only; the leading term is evaluated directly in terms of current density, beam width, β, and variation of field over the beam, which is assumed to be the same as in the absence of electrons.

Figure 14.10 shows a circuit curve (as, of Fig. 14.8) and an electronic curve (as, of Fig. 14.10). These curves contain the same information as the curves (including one of the dashed horizontal lines) of Fig. 14.5, but differently distributed. The intersections represent the modes of propagation.

If such curves were the approximate (dashed) curves of Figs. 14.8 and 14.9, the values of θ for the modes would be quite accurate for real intersections. It is not clear that "intersections" for complex values of θ would be accurately given unless they were for near misses of the curves. In addition, the complicated behavior near $\theta = 1$ (Fig. 14.6) is quite absent from the approximate electronic curve. Thus, the approximate electronic curve does not predict the multitude of unattenuated space-charge waves near $\theta = 1$. Further, the approximate expressions predict a lower limiting electron velocity below which there is no gain. This is not true for the exact equations when the electron flow fills the space between the finned structures completely.

It is of some interest to consider complex intersections in the case of near misses by using curves of simple form (parabolas), as in Fig. 14.11. Such an analysis shows that high gain is to be expected in the case of curves such as those of Fig. 14.10, for instance, when the circuit curve is not steep and when the curvature of the electronic curve is small. In terms of physical parameters, this means a high impedance circuit and a large current density.

14.1 The System and the Equations

The system examined is a two-dimensional one closely analogous to that of Fig. 4.4. It is shown in Fig. 14.1. It consists of a central space extending from $y = -d$ to $y = +d$, and arrays of thin fins separated by slots extending for a distance h beyond the central opening and short-circuited at the outer ends. An electron flow of current density J_0 amperes/m^2 fills the open space. It is assumed that the electrons are constrained by a strong magnetic field so that they can move in the z direction only.

We can simplify the picture a little. The open edges of the slots merely form impedance sheets.

From 4.12 we see that at $y = -d$

$$\frac{H_x}{E_z} = \frac{j\omega\epsilon}{\beta_0} \cot \beta_0 h \tag{14.1}$$

$$\frac{H_x}{E_z} = -jB \tag{14.2}$$

$$B = -\sqrt{\epsilon/\mu} \cot \beta_0 h \tag{14.3}$$

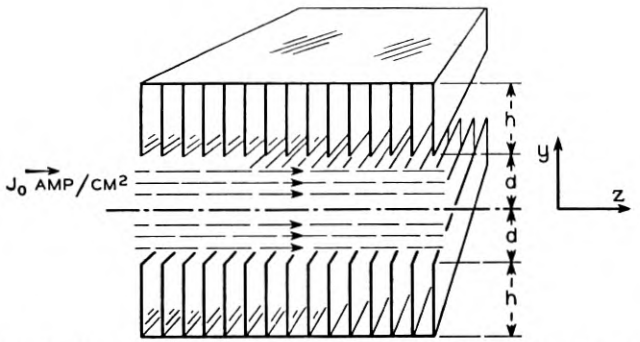


Fig. 14.1—Electron flow completely fills the open space between two finned structures. A strong axial magnetic field prevents transverse motions.

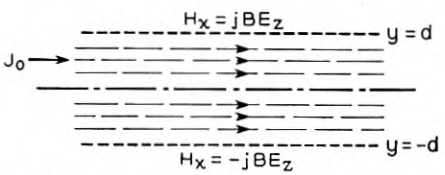


Fig. 14.2—In analyzing the structure of Fig. 14.1, the finned members are regarded as susceptance sheets.

for

$$\beta_0/\omega\epsilon = 1/c\epsilon = \sqrt{\mu/\epsilon} = 377 \text{ ohms} \tag{14.4}$$

Similarly, at $y = +d$,

$$\frac{H_x}{E_z} = jB \tag{14.5}$$

We can use B as a parameter rather than h. Thus, we obtain the picture of Fig. 14.2. This picture is really more general than Fig. 14.1, for it applies for any transverse-magnetic circuit outside of the beam.

Inside of the beam the effect of the electrons is to change the effective dielectric constant in the z direction. Thus, from (2.22) we have for the electron convection current

$$i = \frac{jJ_0\beta_e\Gamma V}{2V_0(j\beta_e - \Gamma)^2} \tag{2.22}$$

Now

$$E_z = -\frac{\partial V}{\partial z} = \Gamma V \tag{14.6}$$

so that

$$i = \frac{jJ_0\beta_e E_z}{2V_0(j\beta_e - \Gamma)^2} \tag{14.7}$$

The appearance of a voltage V in (2.22) and (14.6) does not mean that these relations are invalid for fast waves. In (2.22) the only meaning which need be given to V is that defined by (14.6), as it is the electric field as specified by (14.6) that was assumed to act on the electrons in deriving (2.22).

Let us say that the total a-c current density in the z direction, J_z, is

$$J_z = j\omega\epsilon_1 E_z \tag{14.8}$$

This current consists of a displacement current $j\omega\epsilon E_z$ and the current i, so that

$$J_z = j\omega\epsilon_1 E_z = j\omega\epsilon E_z\left(1 + \frac{J_0\beta_e}{2\epsilon\omega V_0(j\beta_e - \Gamma)^2}\right) \tag{14.9}$$

Hence

$$\epsilon_1/\epsilon = \left(1 + \frac{J_0\beta_e}{2\epsilon\omega V_0(j\beta_e - \Gamma)^2}\right) \tag{14.10}$$

This gives the ratio of the effective dielectric constant in the z direction to the actual dielectric constant. We will proceed to put this in a form which in the long run will prove more convenient.

Let us define a quanity β

$$\Gamma = j\beta \tag{14.11}$$

and a quantity A

$$A = \frac{J_0 d^2}{2\epsilon u_0 V_0} \tag{14.12}$$

And quantities θ and θ_e

$$\theta_e = \beta_e d = (\omega/u_0)d \tag{14.13}$$

$$\theta = \beta d \tag{14.14}$$

We recognize d as the half-width of the opening filled by electrons. Then

$$\epsilon_1/\epsilon = 1 - \frac{A}{(\theta_e - \theta)^2} \tag{14.15}$$

We can say something about the quantity A. From purely d-c considerations, the electron flow will cause a fall in d-c potential toward the center of the beam. Indeed, this is so severe for large currents that it sets a limit to the current density which can be transmitted. If we take V_0 and u_0 as values at $y = \pm d$ (the wall), the maximum value of A as defined by (14.12) is 2/3, and at this maximum value the potential at $y = 0$ is $V_0/4$. This is inconsistent with the analysis, in which V_0 and u_0 are assumed to be constant across the electron flow. Thus, for the current densities for which the analysis is valid, which are the current densities such as are usually used in traveling-wave tubes

$$A \ll 1 \tag{14.16}$$

In the a-c analysis we will deal here only with the symmetrical type of wave in which $E_z(+y) = E_z(-y)$. The work can easily be extended to cover cases for which $E_z(+y) = -E_z(-y)$. We assume

$$H_x = H_0 \sinh \gamma y e^{-j\beta z} \tag{14.17}$$

From Maxwell's equations

$$j\omega\epsilon E_y = \frac{\partial H_x}{\partial z} = -j\beta H_0(\sinh \gamma y)e^{-j\beta z}$$

$$E_y = -\frac{\beta}{\omega\epsilon} H_0(\sinh \lambda y)e^{-j\beta z} \tag{14.18}$$

Similarly

$$j\omega\epsilon_1 E_z = -\frac{\partial H_x}{\partial y} = -\gamma H_0(\cosh \gamma y)e^{-j\beta z}$$

$$E_z = \frac{j\gamma}{\omega\epsilon_1} H_0(\cosh \gamma y)e^{-j\beta z} \tag{14.19}$$

We must also have

$$-j\omega\mu H_x = \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}$$

$$-j\omega\mu H_0 e^{-j\beta z} \sinh \gamma y = \frac{j\gamma^2}{\omega\epsilon_1} H_0 e^{-j\beta z} \cosh \gamma y - \frac{j\beta^2}{\omega\epsilon} H_0 e^{-j\beta z} \sinh \gamma y$$

$$\gamma^2 = (\epsilon_1/\epsilon)(\beta^2 - \beta_0^2) \tag{14.20}$$

$$\beta_0^2 = \omega^2\mu\epsilon = \omega^2/c^2 \tag{14.21}$$

Now, from (14.17), (14.19) and (14.20)

$$\frac{H_x}{E_z} = \frac{-j\omega\epsilon(\epsilon_1/\epsilon)\tanh\left[(\epsilon_1/\epsilon)^{1/2}(\beta^2-\beta_0^2)^{1/2}y\right]}{(\epsilon_1/\epsilon)^{1/2}(\beta^2-\beta_0^2)^{1/2}} \tag{14.22}$$

But

$$\omega\epsilon = (\omega/c)(c\epsilon) = \beta_0\sqrt{\epsilon/\mu} \tag{14.23}$$

Hence

$$\frac{H_x}{E_z} = \frac{-j\sqrt{\epsilon/\mu}(\epsilon_1/\epsilon)^{1/2}\beta_0\tanh\left[(\epsilon_1/\epsilon)^{1/2}(\beta^2-\beta_0^2)^{1/2}y\right]}{(\beta^2-\beta_0^2)^{1/2}} \tag{14.24}$$

At $y = d$, (14.5) must apply. From (14.24) we can write

$$P = -\frac{(\epsilon_1/\epsilon)^{1/2}\tanh\left[(\epsilon_1/\epsilon)^{1/2}(\theta^2-\theta_0^2)^{1/2}\right]}{(\theta^2-\theta_0^2)^{1/2}} \tag{14.25}$$

Here θ is given by (14.14)

$$\theta_0 = \beta_0 d = (\omega/c)d \tag{14.26}$$

and P is given by

$$P = B/\beta_0 d\sqrt{\epsilon/\mu} = B/\theta_0\sqrt{\epsilon/\mu} \tag{14.27}$$

Thus, θ_0 expresses d in radians at free-space wavelength and P is a measure of the wall reactance, the susceptance rising as B rises.

14.2 Waves in the Absence of Electrons

In this section we will consider (14.25) in the case in which there are no electrons and $\epsilon_1/\epsilon = 1$. In this case (14.25) becomes

$$P = -\frac{\tanh(\theta^2-\theta_0^2)^{1/2}}{(\theta^2-\theta_0^2)^{1/2}} \tag{14.28}$$

Suppose we plot the right-hand side of (14.28) vs θ for real values of θ_1 corresponding to unattenuated waves. In Fig. 14.3 this has been done for $\theta_0 = 1/10$. For $\theta_0 > \pi/2$ the behavior near the origin is different, but in cases corresponding to actual traveling wave tubes $\theta_0 < \pi/2$.

Intersections between a horizontal line at height P and the curve give values of θ representing unattenuated waves. We see that for the case which we have considered, in which $\theta_0 < \pi/2$ and $\theta_0 \cot\theta_0 > 1$, there are unattenuated waves if

$$P > -\tan\theta_0/\theta_0 \tag{14.29}$$

For $P = -\infty$ (no slot depth and no wall reactance) the system for $\theta_0 < \pi/2$ constitutes a wave guide operated below cutoff frequency for the type of

wave we have considered. If we increase P ($|P|$ decreasing; the inductive reactance of the walls increasing) this finally results in the propagation of a wave. There are two intersections, at $\theta = \pm\theta_1$, representing propagation to the right and propagation to the left. The variation of θ_1 with P is such that as P is increased (made less negative) θ_1 is increased; that is, the greater is P (the smaller $|P|$), the more slowly the wave travels.

There is another set of waves for which θ is imaginary; these represent passive modes which do not transmit energy but merely decay with distance. In investigating these modes we will let

$$\theta = j\Phi \tag{14.30}$$

so that the waves vary with z as

$$e^{(\Phi/d)z} \tag{14.31}$$

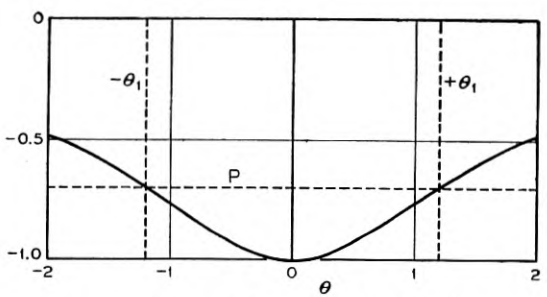


Fig. 14.3—The structure of Fig. 14.1 is first analyzed in the absence of an electron stream. Here a quantity proportional to H_x/E_x at the susceptance sheet is plotted vs $\theta = \beta d$, a quantity proportional to the phase constant β. The solid curve is for the inner open space; the dashed line is for the susceptance sheet. The two intersections at $\pm\theta_1$ correspond to transmission of a forward and a backward wave.

Now (14.28) becomes

$$P = -\tan(\Phi^2 + \theta_0^2)^{1/2}/(\Phi^2 + \theta_0^2)^{1/2} \tag{14.32}$$

In Fig. 14.4 the right-hand side of (14.28) has been plotted vs Φ, again for $\theta_0 = 1/10$.

Here there will be a number of intersections with any horizontal line representing a particular value of P (a particular value of wall susceptance), and these will occur at paired values of Φ which we shall call $\pm\Phi_n$. The corresponding waves vary with distance as $\exp(\pm\Phi_n z/d)$.

Suppose we increase P. As P passes the point $-(\tan\theta_0)/\theta_0$, Φ^n for a pair of these passive waves goes to zero; then for P just greater than $-(\tan\theta_0)/\theta_0$ we have two active unattenuated waves, as may be seen by comparing Figs. 14.4 and 14.3.

14.3 Waves in the Presence of Electrons

In this section we deal with the equations

$$P = \frac{-(\epsilon_1/\epsilon)^{1/2} \tanh [(\epsilon_1/\epsilon)^{1/2}(\theta^2 - \theta_0^2)^{1/2}]}{(\theta^2 - \theta_0^2)^{1/2}} \tag{14.25}$$

and

$$\epsilon_1/\epsilon = 1 - \frac{A}{(\theta_e - \theta)^2} \tag{14.15}$$

We consider cases in which the electron velocity is much less than the velocity of light; hence

$$\theta_e \gg \theta_0 \tag{14.33}$$

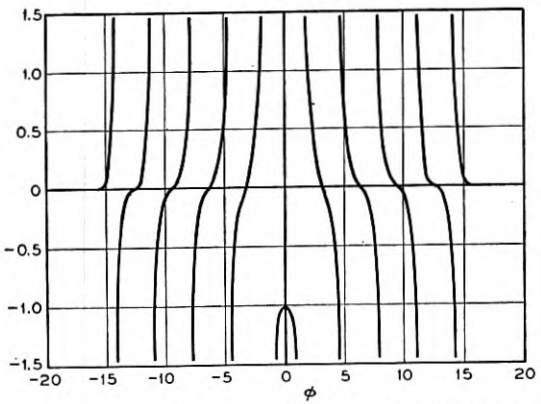


Fig. 14.4—If a quantity proportional to H_x/E_z at the edge of the central region is plotted vs $\Phi = -j\theta$, this curve is obtained. There are an infinite number of intersections with a horizontal line representing the susceptance of the finned structure. These correspond to passive modes, for which the field decays exponentially with distance away from the point of excitation.

In Fig. 14.5, the right-hand side of (14.25) has been plotted vs. θ for $\theta_e = 10\ \theta_0$, corresponding to an electron velocity 1/10 the speed of light. Values of $\theta = 1/10$ and $A = 1/100$ have been chosen merely for convenience.* The curve has not been shown in the region from $\theta = .9$ to $\theta = 1.1$, where ϵ_1/ϵ is negative, and this region will be discussed later.

For a larger value of $P(|\ P\ |$ small), P_1 in Fig. 14.5, there are 4 intersections corresponding to 4 unattenuated waves. The two outer intersections obviously correspond to the "circuit" waves we would have in the absence of electrons. The other two intersections near $\theta = .9\theta_e$ and $\theta = 1.1\theta_e$ we call electronic or space-charge waves.

* At a beam voltage $V_0 = 1{,}000$ and for $d = 0.1$ cm, $A = 1/100$ means a current density of about 330 ma/cm², which is a current density in the range encountered in practice.

For instance, increasing P to values larger than P_1 changes θ for the circuit waves a great deal but scarcely alters the two "electronic wave" values of θ, near $\theta = \theta_e(1 \pm 0.1)$. On the other hand, for large values of P the values of θ for the electronic waves are approximately

$$\theta = \theta_e \pm \sqrt{A} \tag{14.34}$$

Thus, changing A alters these values, but changing A has little effect on the values of θ for the circuit waves.

Now, the larger the P the slower the circuit wave travels; and, hence, for large values of P the electrons travel faster than the circuit wave. Our narrow-beam analysis also indicated two circuit waves and two unattenuated electronic waves for cases in which the electron speed is much larger than the speed of the increasing wave. It also showed, however, that, as the difference between the electron speed and the speed of the unperturbed

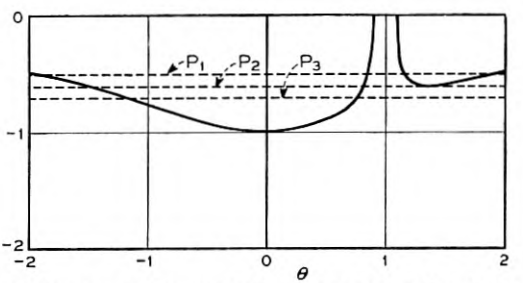


Fig. 14.5—When electrons are present in the open space of the circuit of Fig. 14.1, the curves of Fig. 14.3 are modified as shown here. The nature of the waves depends on the relative magnitude of the susceptance of the finned structure, which is represented by the dashed horizontal lines. For P_1, there are four unattenuated waves, for P_3, two unattenuated waves and an increasing wave and a decreasing wave. Line P_2 represents a transition between the two cases.

wave was made less, a pair of waves appeared, one increasing and one decreasing. This is also the case in the broad beam case.

In Fig. 14.5, when P is given the value indicated by P_2, an "electronic" wave and a "circuit" wave coalesce; this corresponds to y_1 and y_2 running together at $b = (3/2)(2)^{1/3}$ in Fig. 8.1. For a somewhat smaller value of P, such as P_3, there will be a pair of complex values of θ corresponding to an increasing wave and a decreasing wave. We may expect the rate of increase at first to rise and then to fall as P is gradually decreased from the value P_2, corresponding to the rise and fall of x_1 as b is decreased from $(3/2)(2)^{1/3}$ in Fig. 8.1.

It is interesting to know whether or not these increasing waves persist down to $P = -\infty$ (no inductance in the walls). When $P = -\infty$, the only way (14.25) can be satisfied is by

$$\coth\left((\epsilon_1/\epsilon)^{1/2}(\theta^2 - \theta_0^2)^{1/2}\right) = 0 \tag{14.35}$$

This will occur only if

$$(\epsilon_1/\epsilon)^{1/2}(\theta^2 - \theta_0^2)^{1/2} = j\left(n\pi + \frac{\pi}{2}\right)$$
$$(\epsilon_1/\epsilon)(\theta^2 - \theta_0^2) = -\left(n\pi + \frac{\pi}{2}\right)^2 \tag{14.36}$$

Let

$$\theta = u + jw \tag{14.37}$$

From (14.37), (14.36) and (14.15)

$$\left[1 - \frac{A}{((\theta_e - u) + jw)^2}\right]((u + jw)^2 - \theta_0^2) = -\left(n\pi + \frac{\pi}{2}\right)^2 \tag{14.38}$$

If we separate the real and imaginary parts, we obtain

$$[(A - 1)(\theta_e - u)^2 - (A + 1)w^2](u^2 - w^2 - \theta_0^2)$$
$$- 4Auw^2(\theta_e - u) = [(\theta_e - u)^2 + w^2]\left(n\pi + \frac{\pi}{2}\right)^2 \tag{14.39}$$

$$w(u[(\theta_e - u)^2 + w^2] - A[(\theta_e - u)^2 - w^2] + (\theta_e - u)(u^2 - w^2 - \theta_0^2)) = 0 \tag{14.40}$$

The right-hand side of (14.39) is always positive. Because always $A < 1$, the first term on the left of (14.39) is always negative if $u^2 > (w^2 + \theta_0^2)$, which will be true for slow rates of increase. Thus, for very small values of w, (14.39) cannot be satisfied. Thus, it seems that there are no waves such as we are looking for, that is, slow waves ($u \ll c$). It appears that the increasing waves must disappear or be greatly modified when P approaches $-\infty$.

So far we have considered only four of the waves which exist in the presence of electrons. A whole series of unattenuated electron waves exist in the range

$$\theta_e - \sqrt{A} < \theta < \theta_e + \sqrt{A}$$

In this range $(\epsilon_1/\epsilon)^{1/2}$ is imaginary, and it is convenient to rewrite (14.25) as

$$P = \frac{(-\epsilon_1/\epsilon)^{1/2} \tan [(-\epsilon_1/\epsilon)^{1/2}(\theta^2 - \theta_0^2)^{1/2}]}{(\theta^2 - \theta_0^2)^{1/2}} \tag{14.41}$$

The chief variation in this expression over the range considered is that due to variation in $(-\epsilon_1/\epsilon)^{1/2}$. For all practical purposes we may write

$$P = \frac{(-\epsilon_1/\epsilon)^{1/2} \tan [(-\epsilon_1/\epsilon)^{1/2}(\theta_e^2 - \theta_0)^{1/2}]}{(\theta_e^2 - \theta_0^2)^{1/2}} \tag{14.42}$$

Near $\theta = \theta_e$, the tangent varies with infinite rapidity, making an infinite number of crossings of the axis.

In Fig. 14.6, the right-hand side of (14.41) has been plotted for a part of the range $\theta = 0.90\,\theta_e$ to $\theta = 1.10\,\theta_e$. The waves corresponding to the intersections of the rapidly fluctuating curve with a horizontal line representing P are unattenuated space-charge waves. The nearer θ is to θ_e, the larger $(-\epsilon_1/\epsilon)$ is. The amplitude of the electric field varies with y as

$$\cosh\,(j(-\epsilon_1/\epsilon)^{1/2}(\beta^2 - \beta_0^2)^{1/2}y) = \cos\,((-\epsilon_1/\epsilon)^{1/2}(\beta^2 - \beta_0^2)^{1/2}y) \quad (14.45)$$

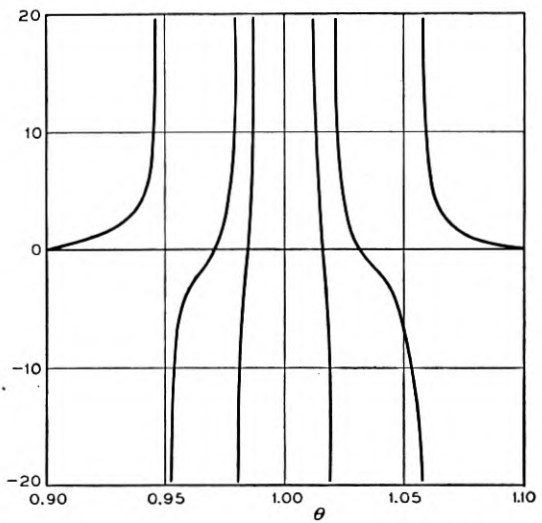


Fig. 14.6—The curve for the central region is not shown completely in Fig. 14.5. A part of the detail around $\theta = 1$, which means a phase velocity equal to the electron velocity, is shown in Fig. 14.6. The curve crosses the axis, and any other horizontal line, an infinite number of times (only some of the branches are shown). Thus, there is a large number of unattenuated "space charge" waves. For these, the amplitude varies sinusoidally in the y direction. Some of these have no physical reality, because the wavelength in the y direction is short compared with the space between electrons.

For small values of $|\theta - \theta_e|$ the field fluctuates very rapidly in the y direction, passing through many cycles between $y = 0$ and $y = d$. For very small values of $|\theta - \theta_e|$ the solution does not correspond to any actual physical problem: spreads in velocity in any electron stream, and ultimately the discrete nature of electron flow, preclude the variations indicated by (14.45).

The writer cannot state definitely that there are not increasing waves for which the real part of θ lies between $\theta_e - \sqrt{A}$ and $\theta_e + \sqrt{A}$, but he sees no reason to believe that there are.

There are, however, other waves which exhibit both attenuation and

propagation. The roots of (14.32) are modified by the introduction of the electrons. To show this effect, let Φ_n be a solution of (14.32), and $j(\Phi_n + \delta)$ be a solution of (14.25). The waves considered will thus vary with distance as

$$e^{[(\Phi_n+\delta)/d]z} \tag{14.43}$$

We see that we must have

$$\begin{aligned}(\epsilon_1/\epsilon)^{1/2}\,(\Phi_n^2 + \theta_0^2)^{1/2}\,\cot\,(\Phi_n^2 + \theta_0^2)^{1/2} \\ = ((\Phi_n + \delta)^2 + \theta_0^2)^{1/2}\,\cot\,[(\epsilon_1/\epsilon)^{1/2}((\Phi_n + \delta)^2 + \theta_0^2)^{1/2}]\end{aligned} \tag{14.44}$$

$$(\epsilon_1/\epsilon)^{1/2} = \left(1 - \frac{A}{(\theta_e - j\Phi_n + \delta)^2}\right)^{1/2} \tag{14.15a}$$

As $A \ll 1$, it seems safe to neglect δ in (14.15a) and to expand, writing

$$(\epsilon_1/\epsilon)^{1/2} = 1 - \alpha \tag{14.46}$$

$$\alpha = \frac{A}{2(\theta_e - j\Phi_n)^2} = \frac{A[(\theta_e^2 - \Phi_n^2) + 2j\theta_e\Phi_n]}{2(\theta_e^2 + \Phi_n^2)^2} \tag{14.47}$$

If $|\,\delta\,| \ll \Phi_n$, we may also write

$$((\Phi_n + \delta)^2 + \theta_0^2)^{1/2} = \frac{\Phi_n\,\delta}{(\Phi_n^2 + \theta_0^2)^{1/2}} + (\Phi_n^2 + \theta_0^2)^{1/2} \tag{14.48}$$

We thus obtain, if we neglect products of δ and α

$$\begin{aligned}(1 - \alpha)\,\cot\,(\Phi_n^2 + \theta_0^2)^{1/2} = \left[1 + \frac{\Phi_n\,\delta}{(\Phi_n^2 + \theta_0^2)^{1/2}}\right]\cot\,(\Phi_n^2 + \theta_0^2)^{1/2} \\ - \left(\frac{\Phi_n\,\delta}{(\Phi_n^2 + \theta_0^2)^{1/2}} - \alpha\right)\csc^2\,(\Phi_n^2 + \theta_0^2)^{1/2}\end{aligned} \tag{14.49}$$

Solving this for δ, we obtain

$$\delta = -\frac{(\Phi_n^2 + \theta_0^2)^{1/2}}{\Phi_n}\left[\frac{\cos\,(\Phi_n^2 + \theta_0^2)^{1/2} + \csc\,(\Phi_n^2 + \theta_0^2)^{1/2}}{\cos\,(\Phi_n^2 + \theta_0^2)^{1/2} - \csc\,(\Phi_n^2 + \theta_0^2)^{1/2}}\right]\alpha \tag{14.50}$$

$$\begin{aligned}\delta = \left[\frac{(\theta_e^2 - \Phi_n^2)}{\Phi_n(\theta_e^2 + \Phi_n^2)^2} + j\,\frac{2\theta_e}{(\theta_e^2 + \Phi_n^2)^2}\right] \\ \cdot \left[\frac{\csc^2\,(\Phi_n^2 + \theta_0^2)^{1/2} + \cos\,(\Phi_n^2 + \theta_0^2)^{1/2}}{\csc\,(\Phi_n^2 + \theta_0^2)^{1/2} - \cos\,(\Phi_n^2 + \theta_0^2)^{1/2}}\right]\frac{A(\theta_0^2 + \Phi_n^2)^{1/2}}{2}\end{aligned} \tag{14.51}$$

As the waves vary with distance as $\exp\,[(\pm\,\Phi_n + \delta)z/d]$, this means that all modified waves travel in the $-z$ direction, and very fast, for the imaginary part of δ, which is inversely proportional to the phase velocity, will be small.

These backward-traveling waves cannot give gain in the $+z$ direction, and could give gain in the $-z$ direction only under conditions similar to those discussed in Chapter XI.

14.4 A Special Type of Solution

Consider (14.25) in a case in which

$$\theta_0 \ll \theta_e \tag{14.52}$$

$$\theta_e \ll 1 \tag{14.53}$$

In this case in the range

$$\theta < \theta_e - \sqrt{A} \quad \text{and} \quad \theta > \theta_e + \sqrt{A} \tag{14.54}$$

we can replace the hyperbolic tangent by its argument, giving

$$P = -(\epsilon_1/\epsilon) = \frac{A}{(\theta_e - \theta)^2} - 1. \tag{14.55}$$

This can be solved for θ, giving

$$\theta = \theta_e \mp \sqrt{A/(P+1)} \tag{14.56}$$

If

$$P < -1$$

Then θ will be complex and there will be a pair of waves, one increasing and one decreasing. We note that, under these circumstances, there is no circuit wave, either with or without electrons.

What we have is in essence an electron stream passing through a series of inductively detuned resonators, as in a multi-resonator klystron. Thus, the structure is in essence a distributed multi-resonator klystron, with lossless resonators. If the resonators have loss, we can let

$$P = (-jG + B)/\theta_0\sqrt{\epsilon/\mu} \tag{14.57}$$

where G is the resonant conductance of the slots. In this case, (14.56) becomes

$$\theta = \theta_e \pm \left(\frac{A\theta\sqrt{\epsilon/\mu}}{-jG + (B + \theta_0\sqrt{\epsilon/\mu})}\right)^{1/2} \tag{14.58}$$

Near resonance we can assume G is a constant and that B varies linearly with frequency. Accordingly, we can show the form of the gain of the increasing wave by plotting vs. frequency the quantity g

$$g = \text{Im}(-j + \omega/\omega_0)^{-1/2} \tag{14.59}$$

In Fig. 14.7, g is plotted vs. ω/ω_0.

Thus, if we wish we may write (14.70) in the form

$$P_e = -\frac{\tanh \theta}{\theta} - P \tag{14.72}$$

where

$$P_e = (1/\theta)[(\epsilon_1/\epsilon)^{1/2} \tanh [(\epsilon_1/\epsilon)^{1/2} \theta - \tanh \theta] \tag{14.73}$$

The quantities on the right of (14.72) refer to the circuit in the absence of electrons; if there are no electrons $P_e = 0$ and (14.72) yields the circuit

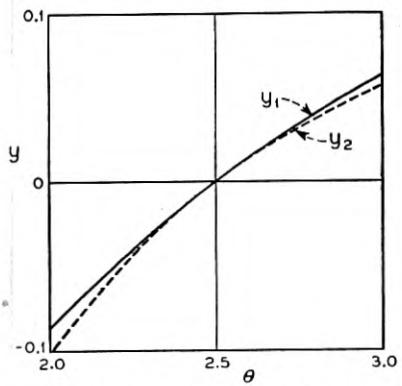


Fig. 14.8—Suppose we compare the circuit admittance for the structure of Fig. 14.1 with that used in earlier calculations. Here the solid curve is proportional to the difference of the H_x's for the finned structure and for the central space (the impressed current) divided by E_z. The dashed curve is the simple expression (6.1) used earlier fitted in magnitude and slope.

waves. Thus, P_e may be regarded as the equivalent of an added current i at the wall, such that

$$\frac{i}{jE_z} = \theta \sqrt{\epsilon/\mu}\, P_e \tag{14.74}$$

Now, the root giving the increasing wave, the one we are most interested in, occurs a little way from the pole, where $(\epsilon_1/\epsilon)^{1/2}$ may be reasonably large if θ is large. It would seem that one of the best comparisons which could be made would be that between the approximate analysis and a very broad beam case, for which θ is very large. In this case, we may take approximately, away from $\theta = \theta_e$

$$\tanh [(\epsilon_1/\epsilon)^{1/2} \theta] = \tanh \theta = 1 \tag{14.75}$$

$$P_e = (1/\theta)[(\epsilon_1/\epsilon)^{1/2} - 1$$

$$P_e = (1/\theta)\left[\left(1 - \frac{A}{(\theta_e - \theta)^2}\right)^{1/2} - 1\right] \tag{14.76}$$

Let us expand in terms of the quantity $A/(\theta_e - \theta)^2$, assuming this to be small compared with unity. We obtain

$$P_e = \frac{A}{2\theta(\theta_e - \theta)^2}\left[1 + \frac{A}{4(\theta_e - \theta)^2} + \cdots\right] \tag{14.77}$$

The theory of Chapter VII is developed by assuming that all electrons are acted on by the same a-c field. When this is not so, it is applied approximately by using an "effective current" or "effective field" as in Chapter IV; either of these concepts leads to the same averaging over the electron flow. An effective current can be obtained by averaging over the flow the current density times the square of the field, evaluated in the absence of electrons, and dividing by the square of the field at the reference position. This is equivalent to the method used in evaluating the effective field in Chapter III.

In the device of Fig. 14.2, if we take as a reference position $y = \pm d$, the effective current I_0 per unit depth

$$I_0 = \frac{J_0 \int_0^d \cosh^2(\gamma y)\, dy}{\cosh^2 \gamma d} \tag{14.78}$$

$$I_0 = (Jd/2)\left(\frac{\tanh \gamma d}{\gamma d} + \operatorname{sech}^2 \gamma d\right) \tag{14.79}$$

This is the effective current associated with the half of the flow from $y = 0$ to $y = d$. Here γ is the value for no electrons. For $\theta \ll \beta$, $\gamma = \beta$. For large values of θ, then

$$I_0 = J_0 d/2\theta \tag{14.80}$$

Now, the corresponding a-c convection current per unit depth will be:

$$i = -j\frac{I_0 \beta_e}{2V_0(\beta_e - \beta)^2} E \tag{14.81}$$

Here E is the total field acting on the electrons in the z-direction. From (7.1) we see that we assumed this to be the field due to the circuit (the first term in the brackets) plus a quantity which we can write

$$E_{z1} = \frac{j\beta^2}{\omega C_1} i \tag{14.82}$$

Accordingly

$$E = E_c + E_{z1} \tag{14.83}$$

and we can write i

$$i = -j \frac{I_0 \beta_e}{2V_0(\beta_e - \beta)^2} \left(E_z + \frac{j\beta^2}{\omega C_1} i \right) \tag{14.84}$$

$$i = \frac{jI_0 \theta_e \, dE_z}{2V_0[K - (\theta_e - \theta)^2]} \tag{14.85}$$

Here K is a parameter specifying the value of $\beta^2/\omega C_1$. As (14.85) need hold over only a rather small range of β, and C is not independent of β, we will regard K as a constant.

The parameter P_e corresponding to (14.85) is

$$P_e = \frac{I_0 \, d(\theta_e/\theta_0)}{2\sqrt{\epsilon/\mu}\, V_0} [K - (\theta_e - \theta)^2]^{-1} \tag{14.86}$$

Now, from (14.80), for large values of θ

$$\frac{I_0 \, d(\theta_e/\theta_0)}{2\sqrt{\epsilon/\mu}\, V_0} = \frac{J_0 \, d^2(\theta_e/\theta_0)}{4\sqrt{\epsilon/\mu}\, \theta V_0} \tag{14.87}$$

As

$$\sqrt{\epsilon/\mu} = \epsilon/\sqrt{\mu\epsilon} = \epsilon c,$$

$$\theta_e/\theta_0 = c/u_0,$$

and

$$A = \frac{J_0 \, d^2}{2\epsilon u_0 V_0} \tag{14.12}$$

$$P_e = \frac{A}{2\theta[K - (\theta_e - \theta)^2]} \tag{14.88}$$

Let us now expand (14.88) assuming K to be very small

$$P_e = \frac{A}{2\theta(\theta_e - \theta)^2} \left[1 + \frac{K}{(\theta_e - \theta)^2} + \cdots \right] \tag{14.89}$$

If we let

$$K = A/4 \tag{14.90}$$

we see that these first two terms agree with the expansion of the broad-beam expression, (14.77). The leading term was not adjusted; the space-charge parameter K was, since there is no other way of evaluating the parameter in this case.

In Fig. 14.9, the value of θP_e as obtained, actually, from (14.73) rather than (14.76), is plotted as a solid line and the value corresponding to the

earlier theory, from (14.86) with K adjusted according to (14.88), is plotted as a dashed line, for

$$A = 0.01$$

$$\theta_e = 8$$

We see that (14.88), which involves the approximations made in our earlier calculations concerning traveling-wave tubes, is a remarkably good fit to the broad-beam expression derived from field theory up very close to the points $(\theta_e - \theta) = A$, which are the boundaries between real and imaginary arguments of the hyperbolic tangent and correspond to the points where the ordinate is zero in Fig. 14.5.

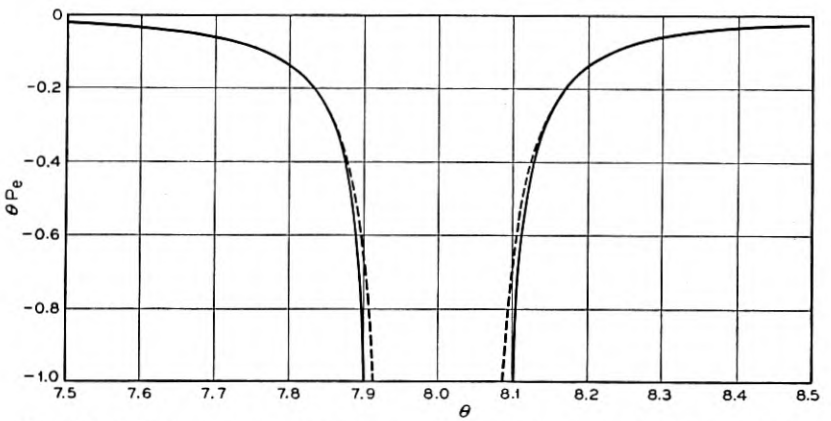


Fig. 14.9—These curves compare an exact electronic susceptance for the broad beam case (solid curve) with the approximate expression used earlier (dashed curve). In the approximate expression, the "effective current" was evaluated, not fitted; the space-charge parameter was chosen to give a fit.

Over the range in which the argument of the hyperbolic tangent in the correct expression is imaginary, the approximate expression of course exhibits none of the complex behavior characteristics of the correct expression and illustrated by Fig. 14.6. From (14.88) we see that the multiple excursions of the true curve from $-\infty$ to $+\infty$ are replaced in the approximate curve by a single dip down toward 0 and back up again. R. C. Fletcher has used a method similar to that explained above in computing the effective helix impedance and the effective space-charge parameter Q for a solid beam inside of a helically conducting sheet. His work, which is valuable in calculating the gain of traveling-wave tubes, is reproduced in Appendix VI.

14.5c *The Complex Roots*

The propagation constants represent intersections of a circuit curve such as that shown in Fig. 14.8 and an electronic curve such as that shown in Fig.

14.9. The propagation constants obtained in Chapters II and VIII represent such intersections of approximate circuit and electronic curves, such as the dotted lines of Fig. 14.8 and 14.9. Propagation constants obtained by field solutions represent intersections of the more nearly exact circuit and electronic curves such as the solid curves of Figs. 14.8 and 14.9.

If we plot a circuit curve giving

$$(1/\theta_0 \sqrt{\epsilon/\mu})(i/jE_z)$$

as given by (14.65) (the right-hand side of 14.75) and an electronic curve giving

$$(1/\theta_0 \sqrt{\epsilon/\mu})(i/jE_z) = P_e$$

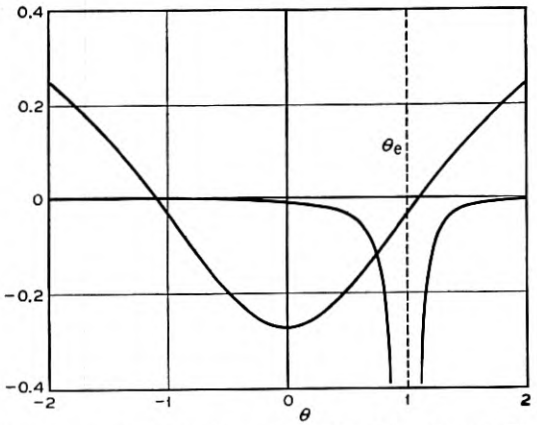


Fig. 14.10—The curves of Fig. 14.5 may be replaced by those of Fig. 14.6. Here the curve which is concave upward represents the circuit susceptance and the other curve represents the electronic susceptance (as in Fig. 14.9).

as given by (14.73) (the left-hand side of (14.72)), the plot, which is shown in Fig. 14.10, contains the same information as the plot of Fig. 14.5 for which θ_0, θ_e and A are the same. In Fig. 14.10, however, one curve represents the circuit without electrons and the other represents the added effect of the electrons.

We have seen that the approximate expressions of Chapter VII fit the broad-beam curves well for real propagation constants (real values of θ) (Fig. 14.8 and 14.9). Hence, we expect that complex roots corresponding to the increasing waves which are obtained using the approximate expressions will be quite accurate when the circuit curve is not too far from the electronic curve for real values of θ; that is, when the parameters (electron velocity, for instance) do not differ too much from those values for which the circuit curve is tangent to the electronic curve.

Unfortunately, the behavior of a function for values of the variables far

from those represented by its intersection with the real plane may be very sensitive to the shape of the intersection with the real plane. Thus, we would scarcely be justified by the good fit of the approximations represented in Figs. 14.8 and 14.9 in assuming that the complex roots obtained using the approximations will be good except when they correspond to a near approach of the electronic and circuit curves, as in Fig. 14.10.

In fact, using the approximate curves, we find that the increasing wave vanishes for electron velocities less than a certain lower limiting velocity. This corresponds to cutting by the circuit curve of the dip down from $+\infty$ of the approximate electronic curve (the dip is not shown in Fig. 14.9). This is not characteristic of the true solution. An analysis shows, however,

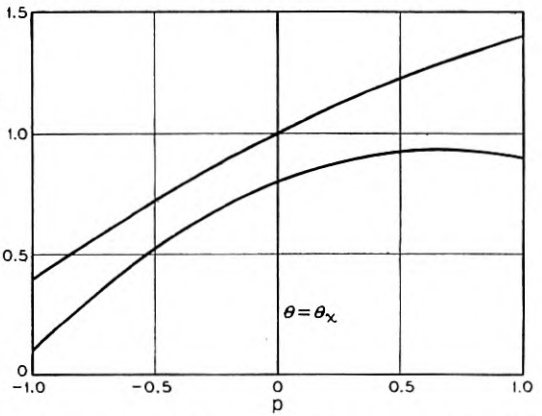


Fig. 14.11—Complex roots are obtained when curves such as those of Fig. 14.10 do not have the number of intersections required (by the degree of the equation) for real values of the abscissa and ordinate. In this figure, two parabolas narrowly miss intersect ng. Suppose these represent circuit and electronic susceptance curves. We find that the gain of the increasing wave will increase with the square root of the separation at the abscissa of equal slopes, and inversely as the square root of the difference in second derivatives.

that there will be a limiting electron velocity below which there is no increasing wave if there is a charge-free region between the electron flow and the circuit.

14.6 Some Remarks About Complex Roots

If we examine our generalized circuit expression (14.60) we see that the circuit impedance parameter $(E^2/\beta^2 P)$ is inversely proportional to the slope of the circuit curve at the point where it crosses the horizontal axis. Thus, low-impedance circuits cut the axis steeply and high-impedance circuits cut the axis at a small slope.

We cannot go directly from this information to an evaluation of gain in terms of impedance; the best course in this respect is to use the methods of

Chapter VIII. We can, however, show a relation between gain and the properties of the circuit and electronic curves for cases in which the curves almost touch (an electron velocity just a little lower than that for which gain appears). Suppose the curves nearly touch at $\theta = \theta_x$, as indicated in Fig. 14.11. Let

$$\theta = \theta_x + p \tag{14.91}$$

Let us represent the curves for small values of p by the first three terms of a Taylor's series. Let the ordinate y of the circuit curve be given by

$$y = a_1 + b_1 p + c_1 p^2 \tag{14.92}$$

and let the ordinate of the electronic curve be given by

$$y = a_2 + b_2 p + c_2 p^2 \tag{14.93}$$

Then, at the intersection

$$(c_1 - c_2)p^2 + (b_1 - b_2)p + (a_1 - a_2) = 0$$

$$p = -(1/2)\frac{b_1 - b_2}{c_1 - c_2} \pm j\sqrt{\frac{(a_1 - a_2)}{(c_1 - c_2)} - \frac{(b_1 - b_2)^2}{4(c_1 - c_2)^2}} \tag{14.94}$$

If we choose θ_x as the point at which the slopes are the same

$$b_1 - b_2 = 0 \tag{14.95}$$

$$p = \pm j\sqrt{\frac{(a_1 - a_2)}{(c_1 - c_2)}} \tag{14.96}$$

and we see that the imaginary part of p increases with the square root of the separation, and at a rate inversely proportional to the difference in second derivatives. This is exemplified by the behavior of x_1 and x_2 for b a little small than $(3/2)(2)^{1/3}$ in Fig. 8.1.

Now, referring to Fig. 14.10, we see that a circuit curve which cuts the axis at a shallow angle (a high-impedance circuit curve) will approach or be tangent to the electronic curve at a point where the second derivative is small, while a steep (low impedance) circuit curve will approach the electronic curve at a point where the second derivative is high. This fits in with the idea that a high impedance should give a high gain and a low impedance should give a low gain.

CHAPTER XV

MAGNETRON AMPLIFIER

Synopsis of Chapter

THE HIGH EFFICIENCY of the magnetron oscillator is attributed to motion of the electrons toward the anode (toward a region of higher d-c potential) at high r-f levels. Thus, an electron's loss of energy to the r-f field is made up, not by a slowing-down of its motion in the direction of wave propagation, but by abstraction of energy from the d-c field.[1]

Warnecke and Guenard[2] have published pictures of magnetron amplifiers and Brossart and Doehler have discussed the theory of such devices.[3]

No attempt will be made here to analyze the large-signal behavior of a magnetron amplifier or even to treat the small-signal theory extensively. However, as the device is very closely related to conventional traveling-wave tubes, it seems of some interest to illustrate its operation by a simple small-signal analysis.

The case analyzed is indicated in Fig. 15.1. A narrow beam of electrons flows in the $+z$ direction, constituting a current I_0. There is a magnetic field of strength B normal to the plane of the paper (in the x direction), and a d-c electric field in the y direction. The beam flows near to a circuit which propagates a slow wave. Fig. 15.3, which shows a finned structure opposed to a conducting plane and held positive with respect to it, gives an idea of a physical realization of such a device. The electron stream could come from a cathode held at some potential intermediate between that of the finned structure and that of the plane. In any event, in the analysis the electrons are assumed to have such an initial d-c velocity and direction as to make them travel in a straight line, the magnetic and electric forces just cancelling.

The circuit equation developed in Chapter XIII in connection with transverse motions of electrons is used. Together with an appropriate ballistical equation, this leads to a fifth degree equation for Γ.

[1] For an understanding of the high-level behavior of magnetrons the reader is referred to: J. B. Fisk, H. D. Hagstrum and P. L. Hartman, "The Magnetron as a Generator of Centimeter Waves," *Bell System Technical Journal*, Vol. XXV, April 1946.
"Microwave Magnetrons" edited by George B. Collins, McGraw-Hill, 1948.

[2] R. Warnecke and P. Guenard, "Sur L'Aide Que Peuvent Apporter en Television Quelques Recentes Conceptions Concernant Les Tubes Electroniques Pour Ultra-Hautes Frequences," *Annales de Radioelectricite*, Vol. III, pp. 259–280, October 1948.

[3] J. Brossart and O. Doehler, "Sur les Proprietes des Tubes a Champ Magnetique Constant: les Tubes a Propagation D'Onde a Champ Magnetique," *Annales de Radioelectricite*, Vol. III, pp. 328–338, October 1948.

The nature of this equation indicates that gain may be possible in two ranges of parameters. One is that in which the electron velocity is near to or equal to (as, (15.25)) the circuit phase velocity. In this case there is gain provided that the transverse component of a-c electric field is not zero, and provided that it is related to the longitudinal component as it is for the circuit of Fig. 15.3. It seems likely that this corresponds most nearly to usual magnetron operation.

The other interesting range of parameters is that near

$$\beta_e/\beta_1 = 1 - \beta_m/\beta_1 \tag{15.31}$$

Here β_e refers to the electrons, β_1 to the circuit and β_m is the cyclotron frequency divided by the electron velocity. When (15.31) holds, there is gain whenever the parameter α, which specifies the ratio of the transverse to the longitudinal fields, is not $+1$. For the circuit of Fig. 15.3, α approaccs $+1$ near the fins if the separation between the fins and the plane is great enough in terms of the wavelength. However, α can be made negative near the fins

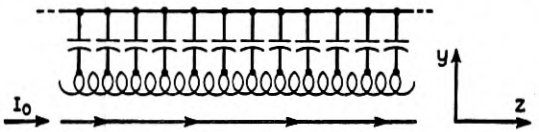


Fig. 15.1—In a magnetron amplifier a narrow electron stream travels in crossed electric and magnetic fields close to a wave transmission circuit.

if the potential of the fins is made negative compared with that of the plane, and the electrons are made to move in the opposite direction.

In either range of parameters, the gain of the increasing wave in db per wavelength is proportional to the square root of the current rather than to the cube root of the current. This means a lower gain than for an ordinary traveling-wave tube with the same circuit and current.

Increasing and decreasing waves with a negative phase velocity are possible when the magnetic field is great enough.

15.1 Circuit Equation

The circuit equation will be the same as that used in Chapter XIII, that is,

$$V = \frac{-j\omega\Gamma_1 K(\Phi\rho - (I_0/u_0)\Phi' y)}{(\Gamma^2 - \Gamma_1^2)} \tag{13.10}$$

It will be assumed that the voltage is given by

$$\Phi = (Ae^{-j\Gamma y} + Be^{j\Gamma y}) \tag{15.1}$$

so that

$$\Phi' V = -jV(Ae^{-j\Gamma y} - Be^{j\Gamma y}) \tag{15.2}$$

Atany$y$ − position we can write

$$\Phi' V = -j\Gamma\alpha\Phi V \tag{15.3}$$

$$\alpha = \frac{Ae^{-j\Gamma y} - Be^{j\Gamma y}}{Ae^{-j\Gamma y} + Be^{j\Gamma y}} \tag{15.4}$$

If Γ is purely imaginary, α is purely real, and as Γ will have only a small real component, α will be considered as a real number. We see that α can range from $+\infty$ to $-\infty$. For instance, consider a circuit consisting of opposed two-dimensional slotted members as shown in Fig. 15.2. For a field with a cosh distribution in the y direction, α is positive above the axis, zero on the axis and negative below the axis. For a field having a sinh distribution in the y

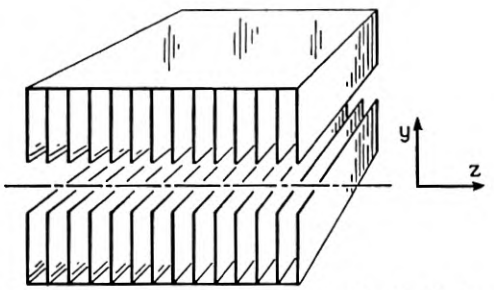


Fig. 15.2—If the circuit is as shown, the ratio between longitudinal and transverse field will be different in sign above and below the axis. This can have an important effect on th operation of the amplifier.

direction, α is infinite on the axis, positive above the axis and negative below the axis.

We find then, that, (13.10) becomes

$$V = \frac{-j\omega\Gamma_1\Phi K(\rho + j\alpha(I_0/u_0)\Gamma y)}{\Gamma^2 - \Gamma_1^2} \tag{15.5}$$

15.2 Ballistic Equations

The d-c electric field in the y direction will be taken as $-E_0$. Thus

$$\frac{d\dot{y}}{dt} = \eta\left[E_0 + \frac{\partial(\Phi V)}{\partial y} - B(z + u_0)\right] \tag{15.6}$$

In order to maintain a rectilinear unperturbed path

$$E_0 = Bu_0 \tag{15.7}$$

so that (15.6) becomes

$$\frac{d\dot{y}}{dt} = \eta\frac{\partial(\Phi V)}{\partial y} - \eta B^z \tag{15.8}$$

Following the usual procedure, we obtain

$$\dot{y} = \frac{-j\eta\Gamma\alpha\Phi V - \eta B^z}{u_0(j\beta_e - \Gamma)} \tag{15.9}$$

We have also

$$\frac{d\dot{z}}{dt} = \eta \frac{\partial \Phi V}{\partial x} + \eta B\dot{y}$$

$$\dot{z} = \frac{-\eta\Gamma\Phi V + \eta B\dot{y}}{u_0(j\beta_e - \Gamma)} \tag{15.10}$$

From (15.9) and (15.10) we obtain

$$z = \frac{-\eta\Gamma\Phi V[(j\beta_e - \Gamma) + j\alpha\beta_m]}{u_0[(j\beta_e - \Gamma)^2 + \beta_m^2]} \tag{15.11}$$

where

$$\beta_m = \omega_m/u_0 \tag{15.12}$$

$$\omega_m = \eta B \tag{15.13}$$

Here ω_m is the cyclotron radian frequency.

As before, we have

$$\rho = \frac{\Gamma\rho_0\dot{z}}{u_0(j\beta_e - \Gamma)} \tag{15.14}$$

whence

$$\rho = \frac{\Gamma^2\eta I_0\Phi V[(j\beta_e - \Gamma) + j\alpha\beta_m]}{u_0^3(j\beta_e - \Gamma)[(j\beta_e - \Gamma)^2 + \beta_m^2]} \tag{15.15}$$

We can also solve (15.9) and (15.10) for $\dot{y}$

$$\dot{y} = \frac{-j\eta\Gamma\Phi V[\alpha(j\beta_e - \Gamma) + j\beta_m]}{u_0[(j\beta_e - \Gamma)^2 + \beta_m^2]}. \tag{15.16}$$

Now, to the first order

$$\dot{y} = \frac{\partial y}{\partial t} + u_0 \frac{\partial y}{\partial z}$$

$$y = \frac{\dot{y}}{u_0(j\beta_e - \Gamma)} \tag{15.17}$$

and from (15.16) and (15.17)

$$y = \frac{-j\eta\Gamma\Phi V[\alpha(j\beta_e - \Gamma) + j\beta_m]}{u_0^2(j\beta_e - \Gamma)[(j\beta_e - \Gamma)^2 + \beta_m^2]}. \tag{15.18}$$

If we use (15.15) and (15.18) in connection with (15.5) we obtain

$$\Gamma^2 - \Gamma_1^2 = \frac{-j\beta_e \Gamma_1 \Gamma^2[(j\beta_e - \Gamma) + 2j[\alpha/(1+\alpha^2)]\beta_m]H^2}{(j\beta_e - \Gamma)[(j\beta_e - \Gamma)^2 + \beta_m^2]} \tag{15.19}$$

$$H^2 = \frac{(1+\alpha^2)\Phi^2 K I_0}{2V_0}. \tag{15.20}$$

Now let

$$-\Gamma_1 = -j\beta_1 \tag{15.21}$$

$$-\Gamma = -j\beta_1(1+p) \tag{15.22}$$

If we assume

$$p \ll 1 \tag{15.23}$$

and neglect p in sums in comparison with unity, we obtain

$$p(\beta_e/\beta_1 - 1 - p)[(\beta_e/\beta_1 - 1 - p)^2 - (\beta_m/\beta_1)^2] = -\frac{\beta_e}{2\beta_1}\left[(\beta_e/\beta_1 - 1 - p) + \frac{2\alpha\beta_m}{(1+\alpha^2)\beta_1}\right]H^2. \tag{15.24}$$

We are particularly interested in conditions which lead to an imaginary value of p which is as large as possible. We will obtain such large values of p when one of the factors multiplying p on the left-hand side of (15.24) is small. There are two possibilities. One is that the first factor is small. We explore this by assuming

$$\beta_e/\beta_1 - 1 = 0 \tag{15.25}$$

$$p^2\left(p^2 - \frac{\beta_m^2}{\beta_1^2}\right) = (1/2)\left(-p + \frac{2\alpha\beta_m}{(1+\alpha^2)\beta_1}\right)H^2, \tag{15.26}$$

If p is very small, we can write approximately

$$-p^2\frac{\beta_m^2}{\beta_1^2} = \frac{\alpha}{(1+\alpha^2)}\frac{\beta_m}{\beta_1}H^2$$
$$p = \pm j[\alpha/(1+\alpha^2)]^{1/2}(\beta_1/\beta_m)^{1/2}H \tag{15.27}$$

We see that p goes to zero if $\alpha = 0$ and is real if α is negative. If we consider what this means circuit-wise, we see that there will be gain with the d-c voltage applied between a circuit and a conducting plane as shown in Fig. 15.3.

Another possible condition in the neighborhood of which p is relatively large is

$$\beta_e/\beta_1 - 1 = \pm\,\beta_m/\beta_1 \tag{15.28}$$

In this case

$$p(\pm\beta_m/\beta_1 - p)(\mp 2(\beta_m/\beta_1)p + p^2) = -\left(1 \pm \frac{\beta_m}{\beta_1}\right)\left[(\pm\beta_m/\beta_1 - p) + \frac{2\alpha\beta_m}{(1+\alpha^2)\beta_1}\right]H^2. \tag{15.29}$$

As p is small, we write approximately

$$p^2 = \pm\frac{1}{4}\frac{(1 \pm \alpha)^2}{1+\alpha^2}\left(\frac{\beta_1}{\beta_m} \pm 1\right)H^2. \tag{15.30}$$

We see that we obtain an imaginary value of p only for the $-$ sign in (15.28) that is, if

$$\beta_e/\beta_1 = 1 - \beta_m/\beta_1 \tag{15.31}$$

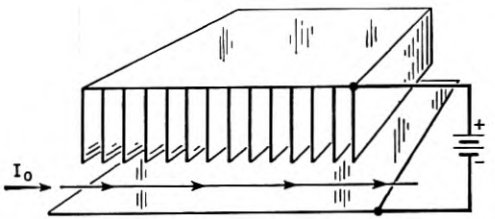


Fig. 15.3—The usual arrangement is to have the finned structure positive and opposed to a conducting plane.

In this case

$$p = \pm j\tfrac{1}{2}[(1-\alpha)/(1+\alpha)^{1/2}](\beta_e/\beta_m)^{1/2}H. \tag{15.32}$$

In this case we obtain gain for any value of α smaller than unity. We note that $\alpha = 1$ is the value α assumes far from the axis in a two-dimensional system of the sort illustrated in Fig. 15.2, for either a cosh or a sinh distribution in the $+y$ direction.

The assumption of $-\Gamma = -j\beta_1(1+p)$ in (15.22) will give forward ($+z$) traveling-waves only. In order to investigate backward traveling-waves, we must assume

$$-\Gamma = +j\beta_1(1+p) \tag{15.33}$$

where again p is considered a small number. If we use this in (15.19), we obtain

$$p\left(\frac{\beta_e}{\beta_1} + 1 + p\right)\left[\left(\frac{\beta_e}{\beta_1} + 1 + p\right)^2 - \frac{\beta_m^2}{\beta_1^2}\right] = -\frac{1}{2}\frac{\beta_e}{\beta_1}\left[\left(\frac{\beta_e}{\beta_1} + 1 + p\right) + \frac{2\alpha\beta_m}{(1+\alpha^2)\beta_1}\right]H^2. \tag{15.34}$$

As before we look for solutions for p where the terms multiplying p on the left are small. The only vanishing consistent with positive values of β_e and β_1 is obtained for

$$\frac{\beta_e}{\beta_1} + 1 = +\frac{\beta_m}{\beta_1}. \tag{15.35}$$

Under this condition (15.34) yields for p

$$p = \pm j \frac{1}{2} \frac{(1 + \alpha)}{(1 + \alpha^2)^{1/2}} \left(\frac{\beta_e}{\beta_m}\right)^{1/2} H. \tag{15.36}$$

Thus we can obtain backward-increasing backward-traveling waves for all values of α except $\alpha = -1$. For the situation shown in Fig. 15.3, with a backward wave, α is always negative, approaching -1 at large distances from the plane electrode, so that the gain is identical with that given by (15.32).

We note that (15.27), (15.32) and (15.36) show that p is proportional to the product of current times impedance divided by voltage to the $\frac{1}{2}$ power, while, in the case of the usual traveling-wave tube, this small quantity occurs to the $\frac{1}{3}$ power. The $\frac{1}{3}$ power of a small quantity is larger than the $\frac{1}{2}$ power; and, hence for a given circuit impedance, current and voltage, the gain of the magnetron amplifier will be somewhat less than the gain of a conventional traveling-wave tube.

CHAPTER XVI

DOUBLE-STREAM AMPLIFIERS

Synopsis of Chapter

IN TRAVELING-WAVE TUBES, it is desirable to have the electrons flow very close to the metal circuit elements, where the radio-frequency field of the circuit is strong, in order to obtain satisfactory amplification. It is, however, difficult to confine the electron flow close to metal circuit elements without an interception of electrons, which entails both loss of efficiency and heating of the circuit elements. This latter may be extremely objectionable at very short wavelengths for which circuit elements are small and fragile.

In the double-stream amplifier the gain is not obtained through the interaction of electrons with the field of electromagnetic resonators, helices or other circuits. Instead, an electron flow consisting of two streams of electrons having different average velocities is used. When the currents or charge densities of the two streams are sufficient, the streams interact so as to give an increasing wave.[1,2,3,4] Electromagnetic circuits may be used to impress a signal on the electron flow, or to produce an electromagnetic output by means of the amplified signal present in the electron flow. The amplification, however, takes place in the electron flow itself, and is the result of what may be termed an electromechanical interaction.[5]

While small magnetic fields are necessarily present because of the motions of the electrons, these do not play an important part in the amplification. The important factors in the interaction are the electric field, which stores energy and acts on the electrons, and the electrons themselves. The charge of the electrons produces the electric field; the mass of the electrons, and their kinetic energy, serve much as do inductance and magnetic stored energy in electromagnetic propagation.

[1] J. R. Pierce and W. B. Hebenstreit, "A New Type of High-Frequency Amplifier," *B.S.T.J.*, Vol. 28, pp. 33–51, January 1949.

[2] A. V. Hollenberg, "Experimental Observation of Amplification by Interaction between Two Electron Streams," *B.S.T.J.*, Vol. 28, pp. 52–58, January 1949.

[3] A. V. Haeff, "The Electron-Wave Tube—A Novel Method of Generation and Amplification of Microwave Energy," *Proc. IRE*, Vol. 37, pp. 4–10, January 1949.

[4] L. S. Nergaard, "Analysis of a Simple Model of a Two-Beam Growing-Wave Tube," *R.C.A. Review*, Vol. 9, pp. 585–601, December 1948.

[5] Some similar electromechanical waves are described in papers by J. R. Pierce, "Possible Fluctuations in Electron Streams Due to Ions," *Jour. App. Phys.*, Vol. 19, pp. 231–236, March 1948, and "Increasing Space-Charge Waves," *Jour. App. Phys.*, Vol. 20 pp. 1060–1066, Nov. 1949.

By this sort of interaction, a traveling wave which increases as it travels, i.e., a traveling wave of negative attenuation, may be produced. To start such a wave, the electron flow may be made to pass through a resonator or a short length of helix excited by the input signal. Once initiated, the wave grows exponentially in amplitude until the electron flow is terminated or until non-linearities limit the amplitude. An amplified output can be obtained by allowing the electron flow to act on a resonantor, helix or other output circuit at a point far enough removed from the input circuit to give the desired gain.

In general, for a given geometry there is a limiting value of current below which there is no increasing wave. For completely intermingled electron streams, the gain rises toward an asymptotic limit as the current is increased beyond this value. The ordinate of Fig. 16.3 is proportional to gain and the abscissa to current.

When the electron streams are separated, the gain first rises and then falls as the current is increased. This effect, and also the magnitude of the increasing wave set up by velocity modulating the electron streams, have been discussed in the literature.[6]

Double-stream amplifiers have several advantages. Because the electrons interact with one another, the electron flow need not pass extremely close to complicated circuit elements. This is particularly advantageous at very short wavelengths. Further, if we make the distance of electron flow between the input and output circuits long enough, amplification can be obtained even though the input and output circuits have very low impedance or poor coupling to the electron flow. Even though the region of amplification is long, there is no need to maintain a close synchronism between an electron velocity and a circuit wave velocity, as there is in the usual traveling-wave tube.

16.1 Simple Theory of Double-Stream Amplifiers

For simplicity we will assume that the flow consists of coincident streams of electrons of d-c velocities u_1 and u_2 in the z direction. It will be assumed that there is no electron motion normal to the z direction. M.K.S. units will be used.

It turns out to be convenient to express variation in the z direction as

$$\exp -j\beta z$$

rather than as

$$\exp -\Gamma z$$

[6] J. R. Pierce, "Double-Stream Amplifiers," *Proc. I.R.E.*, Vol. 37, pp. 980–985, Sept. 1949.

as we have done previously. This merely means letting

$$\Gamma = j\beta \tag{16.1}$$

The following nomenclature will be used

J_1 , J_2 d-c current densities
u_1 , u_2 d-c velocities
ρ_{01} , ρ_{02} d-c charge densities

$$\rho_{01} = -J_1/u_1 \text{ , } \rho_{02} = -J_2/u_2$$

ρ_1 , ρ_2 a-c charge densities
v_1 , v_2 a-c velocities
V_1 , V_2 d-c voltages with respect to the cathodes
V a-c potential
$\beta_1 = \omega/u_1$, $\beta_2 = \omega/u_2$

From (2.22) and (2.18) we obtain

$$\rho_1 = \frac{\eta J_1 \beta^2 V}{u_1^3(\beta_1 - \beta)^2} \tag{16.2}$$

and

$$\rho_2 = \frac{\eta J_2 \beta^2 V}{u_2^3(\beta_2 - \beta)^2} . \tag{16.3}$$

It will be convenient to call the fractional velocity separation b, so that

$$b = \frac{2(u_1 - u_2)}{u_1 + u_2} . \tag{16.4}$$

It will also be convenient to define a sort of mean velocity u_0

$$u_0 = \frac{2u_1 u_2}{u_1 + u_2} . \tag{16.5}$$

We may also let V_0 be the potential drop specifying a velocity u_0 , so that

$$u_0 = \sqrt{2\eta V_0} \tag{16.6}$$

It is further convenient to define a phase constant based on u_0

$$\beta_e = \frac{\omega}{u_0} . \tag{16.7}$$

We see from (16.4), (16.5) and (16.6) that

$$\beta_1 = \beta_e(1 - b/2) \tag{16.8}$$

$$\beta_2 = \beta_e(1 + b/2) \tag{16.9}$$

We shall treat only a special case, that in which

$$\frac{J_1}{u_1^3} = \frac{J_2}{u_2^3} = \frac{J_0}{u_0^3}. \tag{16.10}$$

Here J_0 is a sort of mean current which, together with u_0, specifies the ratios J_1/u_1^3 and J_2/u_2^3, which appear in (4) and (5).

In terms of these new quantities, the expression for the total a-c charge density ρ is, from (16.2) and (16.3) and (16.6)

$$\rho = \rho_1 + \rho_2 = \frac{J_0\beta^2}{2u_0 V_0} \cdot \left[\frac{1}{\left[\beta_e\left(1 - \frac{b}{2}\right) - \beta\right]^2} + \frac{1}{\left[\beta_e\left(1 + \frac{b}{2}\right) - \beta\right]^2}\right] V. \tag{16.11}$$

Equation (16.11) is a *ballistic* equation telling what charge density ρ is produced when the flow is bunched by a voltage V. To solve our problem, that is, to solve for the phase constant β, we must associate (16.11) with a *circuit* equation which tells us what voltage V the charge density produces. We assume that the electron flow takes place in a tube too narrow to propagate a wave of the frequency considered. Further, we assume that the wave velocity is much smaller than the velocity of light. Under these circumstances the circuit problem is essentially an electrostatic problem. The a-c voltage will be of the same sign as, and in phase with the a-c charge density ρ. In other words the "circuit effect" is purely capacitive.

Let us assume at first that the electron stream is very narrow compared with the tube through which it flows, so that V may be assumed to be constant over its cross section. We can easily obtain the relation between V and ρ in two extreme cases. If the wavelength in the stream is very short (β large), so that transverse a-c fields are negligible, then, from Poisson's equation, we have

$$\rho = -\epsilon \frac{\partial^2 V}{\partial z^2}$$
$$\rho = \epsilon\beta^2 V \tag{16.12}$$

If, on the other hand, the wavelength is long compared with the tube radius (β small) so that the fields are chiefly transverse, the lines of force running from the beam outward to the surrounding tube, we may write

$$\rho = CV \tag{16.13}$$

Here C is a constant expressing the capacitance per unit length between the region occupied by the electron flow and the tube wall.

We see from (16.12) and (16.13) that, if we plot ρ/V vs. β/β_e for real values of β, ρ/V will be constant for small values of β and will rise as β^2 for large values of β, approximately as shown in Fig. 16.1.

Now, we have assumed that the charge is produced by the action of the voltage, according to the ballistical equation (16.11). This relation is plotted in Fig. 2, for a relatively large value of J_0/u_0V_0 (curve 1) and for a smaller value of J_0/u_0V_0 (curve 2). There are poles at $\beta/\beta_e = 1 \pm \frac{b}{2}$, and a minimum between the poles. The height of the minimum increases as J_0/u_0V_0 is increased.

A circuit curve similar to that of Fig. 16.1 is also plotted on Fig. 16.2. We see that for the small-current case (curve 2) there are four intersections, giving *four real* values of β and hence *four unattenuated* waves. However, for

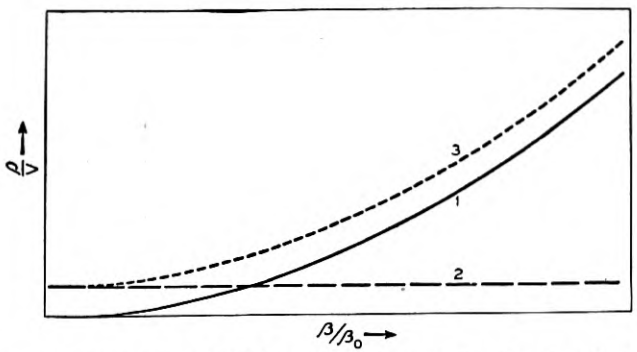


Fig. 16.1—Circuit curves, in which the ordinate is proportional to the ratio of the charge per unit length to the voltage which it produces. Curve 1 is for an infinitely broad beam; curve 2 is for a narrow beam in a narrow tube. Curve 3 is the sum of 1 and 2, and approximates an actual curve.

the larger current (curve 1) there are only two intersections and hence two unattenuated waves. The two additional values of β satisfying both the circuit equation and the ballistical equation are complex conjugates, and represent waves traveling at the same speed, but with equal positive negative attenuations.

Thus we deduce that, as the current densities in the electron streams are raised, a wave with negative attenuation appears for current densities above a certain critical value.

We can learn a little more about these waves by assuming an approximate expression for the circuit curve of Fig. 1. Let us merely assume that over the range of interest (near $\beta/\beta_e = 1$) we can use

$$\rho = \alpha^2\epsilon\beta^2 V \tag{16.14}$$

Here α^2 is a factor greater than unity, which merely expresses the fact that the charge density corresponding to a given voltage is somewhat greater than if there were field in the z direction only for which equation (16.12) is valid. Combining (16.14) with (16.11) we obtain

$$\frac{1}{\left(\beta_e\left(1-\frac{b}{2}\right)-\beta\right)^2}+\frac{1}{\left(\beta_e\left(1+\frac{b}{2}\right)-\beta\right)^2}=\frac{1}{\beta_e^2 U^2} \tag{16.15}$$

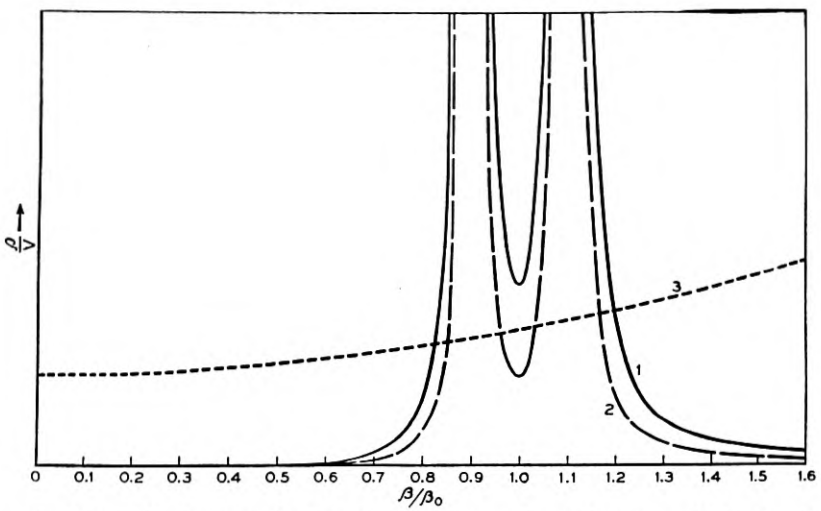


Fig. 16.2—This shows a circuit curve, 3, and two electronic curves which give the sum of the charge densities of the two streams divided by the voltage which bunches them. With curve 2, there will be four unattenuated waves. With curve 1, which is for a higher current density than curve 2, there are two unattenuated waves, an increasing wave and a decreasing wave.

where

$$U^2=\frac{J_0}{2\alpha^2\epsilon\beta_e^2 u_0 V_0}. \tag{16.16}$$

In solving (16.15) it is most convenient to represent β in terms of β_e and a new variable h

$$\beta=\beta_e(1+h) \tag{16.17}$$

Thus, (16.15) becomes

$$\frac{1}{\left(h-\frac{b}{2}\right)^2}+\frac{1}{\left(h+\frac{b}{2}\right)^2}=\frac{1}{U^2}. \tag{16.18}$$

Solving for h, we obtain

$$h = \pm\left(\frac{b}{2}\right)\left[\left(\frac{2U}{b}\right)^2 + 1 \pm \left(\frac{2U}{b}\right)\sqrt{\left(\frac{2U}{b}\right)^2 + 4}\right]^{1/2}. \quad (16.19)$$

The positive sign inside of the brackets always gives a real value of h and hence unattenuated waves. The negative sign inside the brackets gives unattenuated waves for small values of U/b. However, when

$$\left(\frac{U}{b}\right)^2 > \frac{1}{8} \quad (16.20)$$

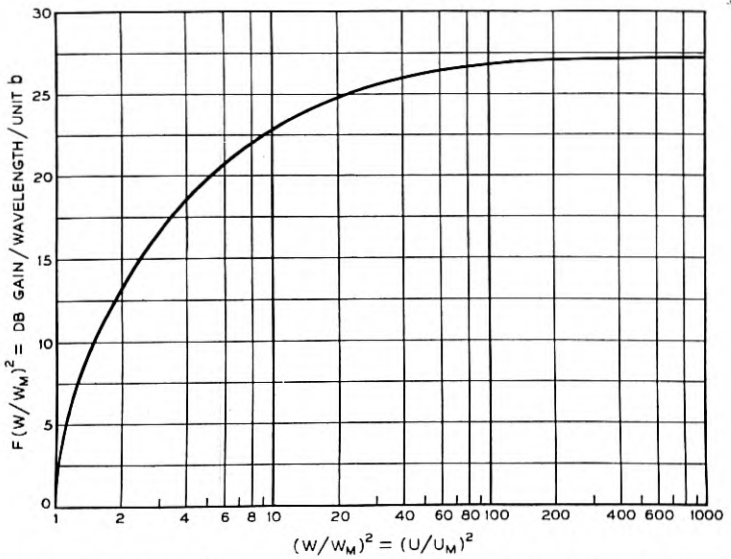


Fig. 16.3—The abscissa is proportional to d-c current. As the current is increased, the gain in db per wavelength approaches 27.3b, where b is the fractional separation in velocity. If the two electron streams are separated physically, the gain is lower and first rises and then falls as the current is increased.

there are two waves with a phase constant β_e and with equal and opposite attenuation constants.

Suppose we let U_M be the minimum value of U for which there is gain. From (16.20)

$$U_M^2 = b^2/8 \quad (16.21)$$

From (16.19) we have, for the increasing wave,

$$h = jb\left[\frac{U}{\sqrt{2}\,U_M}\sqrt{2\left(\frac{U}{U_M}\right)^2 + 1} - \left(\frac{U}{U_M}\right)^2 - 1\right]^{1/2}. \quad (16.22)$$

The gain in db/wavelength is

$$\text{db/wavelength} = 20(2\pi)\log_{10}e^{|h|}$$
$$= 54.6\,|\,h\,| \qquad (16.23)$$

We see that, by means of (16.22) and (16.23), we can plot db/wavelength per unit b vs. $(U/U_M)^2$. This is plotted in Fig. 16.3. Because U^2 is proportional to current, the variable $(U/U_M)^2$ is the ratio of the actual current to the current which will just give an increasing wave. If we know this ratio, we can obtain the gain in db/wavelength by multiplying the corresponding ordinate from Fig. 16.3 by b.

We see that, as the current is increased, the gain per wavelength at first rises rapidly and then rises more slowly, approaching a value $27.3b$ db/wavelength for very large values of $(U/U_M)^2$.

We now have some idea of the variation of gain per wavelength with velocity separation b and with current $(U/U_M)^2$. A more complete theory requires the evaluation of the lower limiting current for gain (or of U_M^2) in terms of physical dimensions and an investigation of the boundary conditions to show how strong an increasing wave is set up by a given input signal.[1, 6]

16.2 Further Considerations

There are a number of points to be brought out concerning double-stream amplifiers. Analysis shows[6] that any physical separation of the electron streams has a very serious effect in reducing gain. Thus, it is desirable to intermingle the streams thoroughly if possible.

If the electron streams have a fractional velocity spread due to space charge which is comparable with the deliberately imposed spread b, we may expect a reduction in gain.

Haeff[3] describes a single-stream tube and attributes its gain to the space-charge spread in velocities. In his analysis of this tube he divides the beam into a high and a low velocity portion, and assigns the mean velocity to each. This is not a valid approximation.

Analysis indicates that a multiply-peaked distribution of current with velocity is necessary for the existent increasing waves, and gain in a "single stream" of electrons is still something of a mystery.

CHAPTER XVII

CONCLUSION

ALTHOUGH THIS BOOK contains some descriptive material concerning high-level behavior, it is primarily a treatment of the linearized or low-level behavior of traveling-wave tubes and of some related devices. In the case of traveling-wave tubes with longitudinal motion of electrons only, the treatment is fairly extended. In the discussions of transverse fields, magnetron amplifiers and double-stream amplifiers, it amounts to little more than an introduction.

One problem to which the material presented lends itself is the calculation of gain of longitudinal-field traveling-wave tubes. To this end, a summary of gain calculation is included as Appendix VII.

Further design information can be worked out as, for instance, exact gain curves at low gain with lumped or distributed loss, perhaps taking the space-charge parameter QC into account, or, a more extended analysis concerning noise figure.

The material in the book may be regarded from another point of view as an introduction, through the treatment of what are really very simple cases, to the high-frequency electronics of electron streams. That is, the reader may use the book merely to learn how to tackle new problems. There are many of these.

One serious problem is that of extending the non-linear theory of the traveling-wave tube. For one thing, it would be desirable to include the effects of loss and space charge. Certainly, a matter worthy of careful investigation is the possibility of increasing efficiency by the use of a circuit in which the phase velocity decreases near the output end. Nordsieck's work can be a guide in such endeavors.

Even linear theory excluding the effects of thermal velocities could profitably be extended, especially to disclose the comparative behavior of narrow electron beams and of broad beams, both those confined by a magnetic field, in which transverse d-c velocities are negligible and in which space charge causes a lowering of axial velocity toward the center of the beam, and also those in which transverse a-c velocities are allowed, especially the Brillouin-type flow, in which the d-c axial velocity is constant across the beam, but electrons have an angular velocity proportional to radius.

Further problems include the extension of the theory of magnetron amplifiers and of double-stream amplifiers to a scope comparable with that of the

theory of conventional traveling-wave tubes. The question of velocity distribution across the beam is particularly important in double-stream amplifiers, whose very operation depends on such a distribution, and it is important that the properties of various kinds of distribution be investigated.

Finally, there is no reason to suspect that the simple tubes described do not have undiscovered relatives of considerable value. Perhaps diligent work will uncover them.

BIBLIOGRAPHY

1946

Barton, M. A. Traveling wave tubes, *Radio*, v. 30, pp. 11–13, 30–32, Aug., 1946.

Blanc-Lapierre, A. and Lapostolle, P. Contribution à l'étude des amplificateurs à ondes progressives, *Ann. des Telecomm.*, v. 1, pp. 283–302, Dec., 1946.

Kompfner, R. Traveling wave valve—new amplifier for centimetric wavelengths. *Wireless World*, v. 52, pp. 369–372, Nov., 1946.

Pierce, J. R. Beam traveling-wave tube, *Bell Lab. Record*, v. 24, pp. 439–442, Dec., 1946.

1947

Bernier, J. Essai de théorie du tube électronique à propagation d'onde, *Ann. de Radioélec.*, v. 2, pp. 87–101, Jan., 1947. *Onde Élec.*, v. 27, pp. 231–243, June, 1947.

Blanc-Lapierre, A., Lapostolle, P., Voge, J. P., and Wallauschek, R. Sur la théorie des amplificateurs à ondes progressives, *Onde Élec.*, v. 27, pp. 194–202, May, 1947.

Kompfner, R. Traveling-wave tuge as amplifier at microwaves, *I.R.E., Proc.*, v. 35, pp. 124–127, Feb., 1947.

Kompfner, R. Traveling-wave tube—centimetre-wave amplifier, *Wireless Engr.*, v. 24, pp. 255–266, Sept., 1947.

Pierce, J. R. Theory of the beam-type traveling-wave tube, *I.R.E., Proc.*, v. 35, pp. 111–123, Feb., 1947.

Pierce, J. R. and Field, L. M. Traveling-wave tubes, *I.R.E., Proc.*, v. 35, pp. 108–111, Feb., 1947.

Roubine, E. Sur le circuit à hélice utilisé dans le tube à ondes progressives, *Onde Élec.*, v. 27, pp. 203–208, May, 1947.

Shulman, C. and Heagy, M. S. Small-signal analysis of traveling-wave tube, *R.C.A. Rev.*, v. 8, pp. 585–611, Dec., 1947.

1948

Brillouin, L. Wave and electrons traveling together—a comparison between traveling wave tubes and linear accelerators, *Phys. Rev.*, v. 74, pp. 90–92, July 1, 1948.

Brossart, J. and Doehler, O. Sur les propriétés des tubes à champ magnétique constant. Les tubes à propagation d'onde à champ magnétique, *Ann. de Radioélec.*, v. 3, pp. 328–338, Oct., 1948.

Cutler, C. C. Experimental determination of helical-wave properties, *I.R.E., Proc.*, v. 36, pp. 230–233, Feb., 1948.

Chu, L. J. and Jackson, J. D. Field theory of traveling-wave tubes, *I.R.E., Proc.*, v. 36, pp. 853–863, July, 1948.

Döehler, O. and Kleen, W. Phénomènes non lineaires dans les tubes à propagation d'onde. *Ann. de Radioélec.*, v. 3, pp. 124–143, Apr., 1948.

Döehler, O. and Kleen, W. Sur l'influence de la charge d'espace dans le tube à propagation d'onde, *Ann. de Radioélec.*, v. 3, pp. 184–188, July, 1948.

Blanc-Lapierre, A., Kuhner, M., Lapostolle, P., Jessel, M. and Wallauschek, R. Étude et réalisation d'amplificateurs à hélice. *Ann. des Telecomm.*, v. 3, pp. 257–308, Aug.–Sept., 1948.

Blanc-Lapierre, A. and Kuhner, M. Réalisation d'amplificateurs à onde progressive à hélice. Résultats généraux, pp. 259–264.

Lapostolle, P. Les phénomènes d'interaction dans le tube à onde progressive, Théorie et vérifications expérimentales, pp. 265–291.

Jessel, M. and Wallauschek, R. Étude expérimentale de la propagation de long d'une ligne à retard en forme d'hélice, pp. 291–299.

Wallauschek, R. Détérmination expérimentale des caractéristiques d'amplificateurs à onde progressive, Résultats obtenus, pp. 300–308.

Lapostolle, P. Étude des diverses ondes susceptibles de se propager dans une ligne en interaction avec un faisceau électronique. Application à la théorie de l'amplificateur à onde progressive, *Ann. des Telecomm.*, v. 3, pp. 57–71, Feb., pp. 85–104, Mar., 1948.

Pierce, J. R. Effect of passive modes in traveling wave tubes, *I.R.E., Proc.*, v. 36, pp. 993–997, Aug., 1948.
Pierce, J. R. Transverse fields in traveling-wave tubes, *Bell Sys. Tech. Jl.*, v. 27, pp. 732–746, Oct., 1948.
Rydbeck, O. E. H. Theory of the traveling-wave tube, *Ericsson Technics*, no. 46, pp. 3–18, 1948.
Tomner, J. S. A. Experimental development of traveling-wave tubes, Acta Polytech., *Elec. Engg.*, v. 1, no. 6, pp. 1–21, 1948.
Nergaard, L. S. Analysis of a simple model of a two-beam growing-wave tube, RCA Rev. vol. 9, pp. 585–601, Dec. 1948.

1949

Doehler, O. and Kleen, W. Influence du vecteur électrique transversal dans la ligne à retard du tube à propagation d'onde, *Ann. de Radioélec.*, v. 4, pp. 76–84, Jan., 1949.
Bruck, L. Comparison des valeurs mesurees pour le gain lineaine du tube à propagation d'onde avec les valeurs indiquées par diverses theories. *Annales de Radioélectricete*, v. IV, pp. 222–232, July, 1949.
Döehler, O. and Kleen, W. Sur le rendement du tube à propagation d'onde. *Annales de Radioélectricete*, v. IV, pp. 216–221, July, 1949.
Döehler, O., Kleen, W. and Palluel, P. Les tubes à propagation d'onde comme oscillateurs à large bande d'accord électronique. *Ann. de Radioélec.*, v. 4, pp. 68–75, Jan., 1949.
Döhler, O. and Kleen, W. Über die Wirkungsweise der "Traveling-Wave" Röhre. *Arch. Elektr. Übertragung*, v. 3, pp. 54–63, Feb., 1949.
Field, L. M. Some slow-wave structures for traveling-wave tubes. *I.R.E., Proc.*, v. 37, pp. 34–40, Jan., 1949.
Guenard, P., Berterattiere, R. and Doehler, O. Amplification par interaction electronque dans des tubes sans circuits. *Annales de Radioélectricete*, v. IV, pp. 171–177, July, 1949.
Laplume, J. Théorie du tube à onde progressive. *Onde Élec.*, v. 29, pp. 66–72, Feb., 1949.
Loshakov, L. N. On the propagation of Waves along a coaxial spiral line in the presence of an electron beam. *Zh. Tech. Fiz.*, vol. 19, pp. 578–595, May, 1949.
Dewey, G. C. A periodic-waveguide traveling-wave amplifier for medium powers. *Proc. N.E.C.* (Chicago), v. 4, p. 253, 1948.
Pierce, J. R. and Hebenstreit, W. B. A new type of high-frequency amplifier, Bell System Technical Journal, v. 28, pp. 33–51, January, 1949.
Haeff, A. V. The electron-wave tube—a novel method of generation and amplification of microwave energy. *Proc. I.R.E.*, v. 37, pp. 4–10, January, 1949.
Hollenberg, A. V. The double-stream amplifier, *Bell Laboratories Record*, v. 27, pp. 290–292, August, 1949.
Guenard, P., Berterottiere, R. and Döehler, O. Amplification by direct electronic interaction in valves without circuits, *Ann. Radioelec.*, v. 4, pp. 171–177, July, 1949.
Rogers, D. C. Traveling-wave amplifier for 6 to 8 centimeters, *Elec. Commun.* (London), v. 26, pp. 144–152, June, 1949.
Field, L. M. Some slow wave structures for traveling-wave tubes, *Proc. I.R.E.*, v. 37, pp. 34–40, January, 1949.
Schnitzer, R. and Weber, D. *Frequenz*, v. 3, pp. 189–196, July, 1949.
Pierce, J. R. Circuits for traveling-wave tubes, *Proc. I.R.E.*, v. 37, pp. 510–515, May, 1949.
Pierce, J. R. and Wax, N. A note on filter-type traveling-wave amplifiers, *Proc. I.R.E.*, v. 37, pp. 622–625, June, 1949.

Technical Publications by Bell System Authors Other Than in the Bell System Technical Journal

*Circuits for Cold Cathode Glow Tubes.** W. A. DEPP[1] and W. H. T. HOLDEN.[1] *Elec. Mfg.*, v. 44, pp. 92–97, July, 1949.

Equipment for the Determination of Insulation Resistance at High Humidities. A. T. CHAPMAN.[2] *A.S.T.M. Bull.*, no. 165, pp. 43–45, Apr., 1950.

Twin Relationships in Ingots of Germanium. W. C. ELLIS.[1] *Jl. Metals*, v. 188, p. 886, June, 1950.

*Magnetic Cores of Thin Tape Insulated by Cataphoresis.** H. L. B. GOULD.[1] *Elec. Engg.*, v. 69, pp. 544–548, June, 1950.

Slip Markings in Chromium. E. S. GREINER.[1] *Jl. Metals*, v. 188, pp. 891–892, June, 1950.

New Porcelain Rod Leak. H. D. HAGSTRUM[1] and H. W. WEINHART.[1] *Rev. Sci. Instruments*, v. 21, p. 394, Apr., 1950.

Significance of Nonclassical Statistics. R. V. L. HARTLEY.[1] *Science*, v. 111, pp. 574–576, May 26, 1950.

Comment on Mobility Anomalies in Germanium. G. L. PEARSON,[1] J. R. HAYNES[1] and W. SHOCKLEY.[1] Letter to the editor. *Phys. Rev.*, v. 78, pp. 295–296, May 1, 1950.

Dislocation Models of Crystal Grain Boundaries. W. T. READ[1] and W. SHOCKLEY.[1] *Phys. Rev.*, v. 78, pp. 275–289, May 1, 1950.

Zero-Point Vibrations and Superconductivity. J. BARDEEN.[1] Letter to the editor. *Phys. Rev.*, v. 79, pp. 167–168, July 1, 1950.

Some Observations on Industrial Research. O. E. BUCKLEY.[1] *Bell Tel. Mag.*, v. 29, pp. 13–24, Spring, 1950.

Properties of Single Crystals of Nickel Ferrite. J. .K GALT,[1] B. T. MATHIAS[1] and J. P. REMEIKA.[1] Letter to the editor. *Phys. Rev.*, v. 79, pp. 391–392, July 15, 1950.

Photon Yield of Electron-Hole Pairs in Germanium. F. S. GOUCHER.[1] Letter to the editor. *Phys. Rev.*, v. 78, p. 816, June 15, 1950.

Data on Porcelain Rod Leak. J. P. MOLNAR[1] and C. D. HARTMAN.[1] *Rev. Sci. Instruments*, v. 21, pp. 394–395, Apr., 1950.

Magnetic Susceptibility of αFe_2O_3 *and* αFe_2O_3 *with Added Titanium.* F. J. MORIN.[1] Letter to the editor. *Phys. Rev.*, v. 78, pp. 819–820, June 15, 1950.

* A reprint of this article may be obtained on request to the editor of the B.S.T.J.

[1] B.T.L.

[2] W.E.Co.

Alternating Current Conduction in Ice. E. J. MURPHY.[1] Letter to the editor. *Phys. Rev.*, v. 79, pp. 396–397, July 15, 1950.

Ferromagnetic Resonance in Manganese Ferrite and the Theory of the Ferrites. W. A. YAGER,[1] F. R. MERRITT,[1] C. KITTEL[1] and C. GUILLAUD.[1] Letter to the editor. *Phys. Rev.*, v. 79, p. 181, July 1, 1950.

Conductivity Pulses Induced in Diamond by Alpha Particles. A. J. AHEARN.[1] *AEC, Brookhaven conference report, BNL-C-1 High speed counters and short pulse techniques,* Aug. 14–15, 1947. 1950. p. 7.

Behavior of Resistors at High Frequencies. G. R. ARTHUR[1] and S. E. CHURCH.[1] *T. V. Engg.*, v. 1, pp. 4–7, June, 1950.

Note on "The Application of Vector Analysis to the Wave Equation". R. V. L. HARTLEY.[1] Letter to the editor. *Acoustical Soc. Am. Jl.*, v. 22, p. 511, July, 1950.

*Number 5 Crossbar Dial Telephone Switching System.** F. A. KORN[1] and JAMES G. FERGUSON.[1] *Elec. Engg.*, v. 69, pp. 679–684, Aug., 1950.

Traveling-Wave Tube as a Broad Band Amplifier. J. R. PIERCE.[1] *AEC, Brookhaven conference report, BNL-C-1 High speed counters and short pulse techniques,* Aug. 14–15, 1947. 1950. p. 41.

* A reprint of this article may be obtained on request to the editor of the B.S.T.J.
[1] B.T.L.

Contributors to this Issue

JOHN BARDEEN, University of Wisconsin, B.S. in E.E., 1928; M.S., 1930. Gulf Research and Development Corporation, 1930–33; Princeton University, 1933–35, Ph.D. in Math. Phys., 1936; Junior Fellow, Society of Fellows, Harvard University, 1935–38; Assistant Professor of Physics, University of Minnesota, 1938–41; Prin. Phys., Naval Ordnance Laboratory, 1941–45. Bell Telephone Laboratories, 1945–. Dr. Bardeen is engaged in theoretical problems related to semiconductors.

A. E. BOWEN, Ph.B., Yale University, 1921; Graduate School, Yale University, 1921–24. American Telephone and Telegraph Company, Department of Development and Research, 1924–34. Bell Telephone Laboratories, 1934–42. U. S. Army Air Force, 1942–45. Bell Telephone Laboratories, 1945–48. With the American Telephone and Telegraph Company, Mr. Bowen's work was concerned principally with the inductive coordination of power and communications systems. From 1934 to 1942 he was engaged in work in the ultra-high-frequency field, particularly on hollow waveguides. He became a Major and later a Colonel while serving with the U. S. Army Air Force from 1942 to 1945 on a special mission to Trinidad and subsequently in the Pentagon. After returning to Bell Telephone Laboratories in 1945 he was engaged in the problems of microwave repeater research until his death in 1948.

M. E. HINES, B.S. in Applied Physics, California Institute of Technology, 1940; B.S. in Meteorology, 1941; M.S. in Electrical Engineering, 1946. U. S. Air Force Weather Service, 1941–45. Bell Telephone Laboratories, 1946–. Mr. Hines has been engaged in the development of vacuum tubes.

JACK A. MORTON, B.S. in Electrical Engineering, Wayne University, 1935; M.S.E., University of Michigan, 1936. Bell Telephone Laboratories, 1936–. Mr. Morton joined the Laboratories to work on coaxial cable and microwave amplifier circuit research; during the war he was at first a member of a group engaged in improving the signal-to-noise performance of radar receivers. In 1943 he transferred to the Electronic Development Department to work on microwave tubes for radar and radio relay. Since 1948 he has been Electronic Apparatus Development Engineer responsible for the development of transistors and other semiconductor devices.

WILLIAM W. MUMFORD, B.A., Willamette University, 1930. Bell Telephone Laboratories, 1930–. Mr. Mumford has been engaged in work that is chiefly concerned with ultra-short-wave and microwave radio communication.

J. R. PIERCE, B.S. in Electrical Engineering, California Institute of Technology, 1933; Ph.D., 1936. Bell Telephone Laboratories, 1936–. Dr. Pierce has been engaged in the study of vacuum tubes.

ROBERT M. RYDER, Yale University, B.S. in Physics, 1937; Ph.D., 1940. Bell Telephone Laboratories, 1940–. Dr. Ryder joined the Laboratories to work on microwave amplifier circuits, and during most of the war was a member of a group engaged in studying the signal-to-noise performance of radars. In 1945 he transferred to the Electronic Development Department to work on microwave oscillator and amplifier tubes for radar and radio relay applications. He is now in a group engaged in the development of transistors.

W. VAN ROOSBROECK, A.B., Columbia College, 1934; A.M., Columbia University, 1937. Bell Telephone Laboratories, 1937–. Mr. van Roosbroeck's work at the Laboratories was concerned during the war with carbon-film resistors and infra-red bolometers and, more recently, with the copper oxide rectifier. In 1948 he transferred to the Physical Research Department where he is now engaged in problems of solid-state physics.