Some Aspects of the Design of Power Transistors*

N. H. FLETCHER†

Summary—**This paper discusses some factors which have to be taken into account in the design of high power transistors. An effect of great importance is the reduction of emitter bias caused by transverse current flow in the base region. This effect is examined in some detail and the results of the discussion are applied to the design of improved transistor types. Finally, a short discussion of thermal stability and mechanical design is given.**

Introduction

TRANSISTORS are mostly considered as very low level small signal devices and much of the design theory of transistors has assumed their use in such applications. It will be our purpose in this paper to point out some of the considerations which enter design theory when the transistors are explicitly intended for high level operations.

Low Level Operation

As we pointed out above, most design theory so far developed is for the low level case. An excellent account of this subject is given by Early[1] who summarizes the current position in small signal terms.

The basic assumption of low level theory is that all perturbations are small and the theory is thus linear in all important respects. Among the important assumptions and approximations made in Early's theory are the following:

1. Junctions are assumed parallel and planar.
2. Surface effects are neglected.
3. Emitter current densities are assumed small.

It happens that in high power design we cannot validly make any of these approximations so that to this extent the theory is more complicated. However, at the moment we are not trying to develop a complete theory but merely to investigate some important problems which arise in high level operation, and for our present purpose, many of the factors involved in small signal operation can be recognized and neglected as of secondary importance. In this way, we shall be able to get a semi-quantitative idea of those design considerations which are of most importance in our present work, namely, the design and construction of power transistors with output ratings greater than one watt, to operate in the audio frequency range.

The Design Problems

These may be divided into two main classes:

1. Electrical Problems
2. Thermal and Mechanical Problems

* Original manuscript received by the IRE, December 10, 1954; revised manuscript received, February 9, 1955.

† Transistor Products, Inc., Waltham, Mass. (On leave from Division of Radiophysics, CSIRO, Sydney, Australia.)

[1] J. M. Early, "Design theory of junction transistors," *Bell Sys. Tech. Jour.*, vol. 32, p. 1271; 1953.

Electrical problems relate to the fact that we can no longer make the approximation that the injected carrier density is small enough to be considered a linear perturbation. Along with this, we have to recognize that in fact none of the electrical quantities involved is small and some approximations are no longer valid. It is with problems of this type that we shall be mainly concerned in an attempt to see what we should do to optimize performance at high power levels.

The second class of problems is in a sense incidental but nevertheless of considerable importance. It relates mainly to suitable mechanical design so that heat generated in the transistor can be adequately removed. Other mechanical considerations are of stability to thermal and mechanical shocks and of ease and economy of fabrication. These problems will be much more briefly dealt with, though this should not be understood as minimizing their importance.

Electrical Design

High power output from a device clearly involves high currents, high voltages, or both, and each brings its own problems. If a transistor is to operate with a collector voltage V_c, then it must withstand, in general, a voltage $2V_c$. Whilst diodes can be made with reverse characteristics extending to several hundreds of volts, operation of transistors in this region is not usually of advantage since stability is sacrificed. The power gain is, however, increased at high voltages since higher load impedances can be used. For this reason, we should like to operate at as high a collector voltage as is compatible with other considerations.

Since direction of increasing voltage (first by stability considerations and ultimately by the breakdown voltage attainable) is limited, let us examine the extension we can make in the direction of increasing current.

It is well known that the current transfer ratio α for a junction transistor first increases as emitter current increases, reaches a peak and then falls off to low values. Low level transistors can be conveniently operated near the peak, which occurs at emitter currents of less than 1 ma for ordinary small transistors. However, for moderate output, we are already operating with tens or hundreds of milliamperes and, for high powers, amperes are required. We must therefore find some way to shift the peak of the curve to higher current values and to minimize α fall-off at high currents.

A careful analysis of this effect has been made by Webster.[2] We shall give a brief account of his findings and illustrate them on a simple physical basis.

[2] W. M. Webster, "On the variation of junction transistor current amplification factor with emitter current," Proc. I.R.E., vol. 42, pp. 914–920; June, 1954

Consider a p-n-p transistor with a bar-like structure so that the junctions are plane and parallel. In the low level case, holes are injected into the base region from the emitter and flow by diffusion down a concentration gradient to the collector, a small fraction recombining with electrons which flow in as base current. This is the small signal theory and does not predict any variation of α with emitter current.

Now consider what happens when injected hole densities become large enough to become comparable with the equilibrium electron density of the base region. We still have a concentration gradient of holes in the base region and this yields a diffusion current of holes toward the collector, as illustrated in Fig. 1.

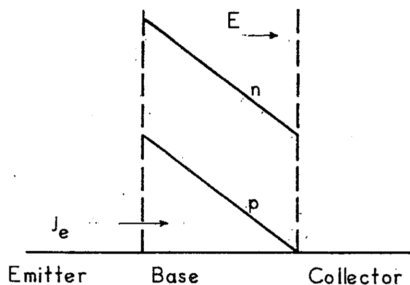

Fig. 1—Distribution of carriers in the base region for large injected current density. (p-n-p transistor.)

However, we must have a similar gradient in the electron density in the base region in order that approximate electrical neutrality may prevail ($n-p=$ Constant). This is also shown in Fig. 1. However, no electron current can flow across the collector barrier so that if Fig. 1 is to represent a steady-state situation, there must be an electric field E generated in the direction shown to maintain the electron distribution (by contributing an electron drift current equal and opposite to the electron diffusion current). However, this field acts also upon the holes and is in a direction to aid the flow of hole current to the collector. In view of the Einstein relation

$$\frac{\mu_p}{D_p} = \frac{q}{kT} = \frac{\mu_n}{D_n}$$

this field ultimately doubles the hole current due to diffusion alone. Thus as the injected hole current density becomes large, we effectively double the diffusion coefficient for holes in the base region. This reduces the hole density at the emitter necessary to cause a given current to flow and thus decreases the proportion of holes lost by surface recombination. This leads to an increase in α as j_e is increased.

As the emitter current becomes larger still, other effects begin to enter. The injected holes increase the density of electrons in the base region adjacent to the emitter junction. This effectively lowers the resistivity of the base region near the emitter junction and this in turn decreases the emitter efficiency, and hence α.

As the density of injected holes increases, the rate of recombination increases nonlinearly. This rate is ordinarily proportional to the product np which is linear for $p \ll n$ but more nearly quadratic when p is comparable to n due to electrical neutrality requirements. It is not clear yet what the exact dependence is at very high currents, but in any case, the trend is toward a lowered value of α.

Webster derives an expression giving the complete dependence of α on emitter current and calculates a numerical case for a typical low level p-n-p alloy junction transistor, with emitter area 10^{-3} square centimeters. His curve is illustrated in Fig. 2 and shows the increase, peak, and fall-off we discussed above. The curve agrees well with experiment.

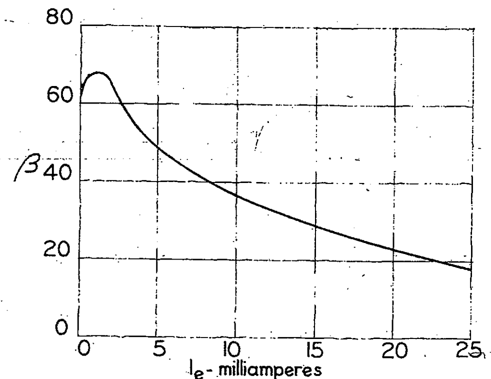

Fig. 2—Behavior of grounded emitter current gain, β, with I_e for typical low level transistor (after Webster,[2] Fig. 6).

Some of the effects of variations in electrode geometry have been investigated by Moore and Pankove[3] for the case of a typical low level p-n-p alloy transistor. They find that for such a transistor, the main loss of injected carriers is due to surface recombination in an annular region around the emitter, volume recombination being comparatively unimportant for base layer widths less than a few thousandths of an inch. On the basis of their calculations and an analog plotting technique for determining hole flow paths, they were able to predict the optimum emitter diameter for a given collector diameter, this optimum representing a compromise between most efficient collection and an excessive ratio of emitter circumference to emitter area. The optimum diameter is clearly a function of surface recombination velocity in an annular region surrounding the emitter.

In discussing the design of high power transistors, we shall make use of many of the ideas discussed above, modifying or extending them as necessary in view of the large currents involved.

Base Resistance Bias Effects

We shall now proceed to discuss an effect of very great importance in the design of power transistors which has been validly neglected in low power theory.[4]

[3] A. R. Moore and J. I. Pankove, "The effect of junction shape and surface recombination on transistor current gain," PROC. I.R.E., vol. 42, pp. 907–913; June, 1954.

[4] This effect has been recognized by Early, op. cit., but his treatment is confined to the small signal case.

In a high power transistor the current which flows transversely through the base region to the base lead is of considerable size, ranging to tens of milliamperes—an order of magnitude larger than is encountered in low level transistors. Because of the small thickness and finite resistivity of the base layer, this current causes an ohmic voltage gradient which is in such a direction as to reduce the effective forward bias on parts of the emitter distant from the base lead. This in turn causes a large drop in the current density injected by these parts of the emitter, and consequently a drop in the efficiency of the transistor. To obtain an expression for the magnitude of this effect we shall examine a simple idealized case.

Consider what we can call a "sem-infinite one-dimensional transistor," as shown in Fig. 3. Suppose that we have a semi-infinite region of P-type germanium containing an N-type region (the base) of width W bounded by two parallel planes normal to the boundary of the region. An ohmic base contact is made to the whole exposed area of the base region as shown. For purposes of calculation, we shall consider a slice of this semi-infinite region normal to its boundary and to the boundaries of the base region, the slice being of unit thickness.

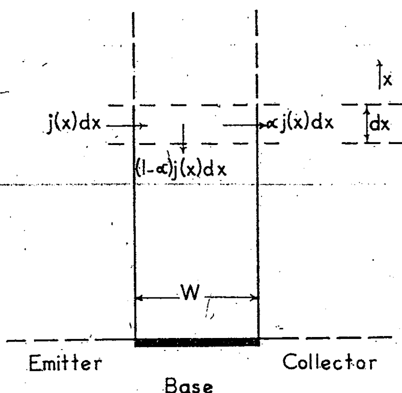

Fig. 3

We shall make the following assumptions:

1. The conductivity of the emitter region is very large so that emitter efficiency approaches unity.
2. The conductivity of the base region is sufficiently high that it is not severely altered by the presence of injeeted carriers at the levels considered.
3. Flow of minority carriers across the base region is by field-aided diffusion, as discussed by Webster, the effective diffusion coefficients being g times the normal coefficient where $1 \leqq g \leqq 2$.
4. Recombination of injected holes in the base region is approximately monomolecular and described by an effective lifetime τ which is independent of injected carrier concentration in range considered.

Before proceeding further, let us examine these assumptions to see that they are reasonably valid and applicable to practical cases. Consider an alloyed junction transistor with base width 0.0025 cm. Typically, emitter and collector region resistivities will be ~0.001 ohm-cm and base region resistivity ~1 ohm-cm. Assumption (1) is thus fairly valid and any departure of emitter efficiency from unity can be indicated in the effective α. For an injected emitter current density of 10 amps per square cm, the injected hole density at the emitter junction [making assumption (3) with $g=1.7$] is about 2×10^{15} per cubic centimeter which corresponds to a resistivity of about 1 ohm-cm for the associated electron density. The injected hole density of course decreases approximately linearly across the base region and is zero at the collector junction so that the effective value is about one-half of this. Thus assumption (2) is valid up to injected current densities of a few amps per square cm. We shall examine the case where this assumption is not valid at a later stage in the discussion.

The value of g to be used in assumption (3) can be found from Webster's paper.[2] In the present example $g>1.5$ for emitter current densities greater than about 5 amps per square cm. The exact value will not concern us for the moment.

Assumption (4) is just a simplifying assumption which appears very reasonable in the light of assumption (2). Its effect will be examined later.

Now let us proceed with our derivation, referring to Fig. 3:

Suppose a current density $j(x)$ is injected by the emitter into the base region at x in the element dx. This represents a current $j(x)dx$ of which $\alpha j(x)dx$ flows into the collector and $(1-\alpha)\,j(x)dx$ flows through the base region towards the base lead. Here α is the effective α of this idealized transistor and includes current due to non-unity emitter efficiency and to recombination of carriers within the base region. The first effect contributes $(1-\gamma)j(x)dx$ and the second $(1-\alpha')j(x)dx$, where from the usual theory

$$\alpha' \approx 1 - \frac{1}{2}\left(\frac{W}{L_p}\right)^2, \tag{1}$$

where $L_p=\sqrt{D_p\tau_p}$ is the diffusion length for holes injected into the base region. For typical values, $W\approx 0.0025$ cm, $\tau_p\approx 100\mu$sec, $\alpha'\approx 0.9995$, so that it may well be the $(1-\gamma)$ term which gives the major contribution to $(1-\alpha)$. We should point out here that these very high values of α are due to the fact that as yet we have not considered surface effects.

We now make the assumption, a result of assumptions (1)–(4) discussed above, that α is constant over the whole of the base region. The injected current I_e can be expressed in the usual way as a function of emitter-base voltage V as

$$j(x) = j_0(e^{qV(x)/kT} - 1), \tag{2}$$

where j_0 is the reverse saturation current of the emitter junction.

Now the base region has a resistivity ρ which we have assumed constant so that its linear resistance in the x direction is ρ/W ohm-cm. The emitter region is in effect

equipotential and we take the emitter to be at a potential V_0 with respect to the base contact at $x=0$. The base current at x is

$$i_b(x) = \int_x^\infty (1-\alpha)j(x)dx,$$

whence

$$\frac{di_b(x)}{dx} = -(1-\alpha)j(x).$$

The current $i_b(x)$ causes a voltage drop given by

$$\frac{dV(x)}{dx} = -\frac{\rho}{W} i_b(x),$$

whence

$$\frac{d^2V(x)}{dx^2} = -\frac{\rho}{W}\frac{di_b(x)}{dx},$$

so that we have

$$\frac{d^2V(x)}{dx^2} = +\frac{\rho}{W}(1-\alpha)j(x),$$

and using our previous expression (2) for $j(x)$

$$\frac{d^2V(x)}{dx^2} = \frac{\rho}{W}(1-\alpha)j_0(e^{qV(x)/kT}-1). \quad (3)$$

In (3) let $y=dV/dx$ then

$$\frac{d^2V}{dx^2} = y\frac{dy}{dV}$$

and we have

$$y\frac{dy}{dV} = \frac{\rho}{W}(1-\alpha)j_0(e^{qV/kT}-1). \quad (4)$$

Integrate from the general point considered to $x=\infty$ using the fact that $y(\infty)=0$, $V(\infty)=0$.

$$\tfrac{1}{2}y^2 = \frac{\rho}{W}(1-\alpha)j_0\left(\frac{kT}{q}e^{qV/kT} - \frac{kT}{q} - V\right)$$

$$\therefore \left(\frac{dV}{dx}\right)^2 = \frac{2\rho}{W}(1-\alpha)j_0\frac{kT}{q}\left(e^{qV/kT} - 1 - \frac{qV}{kT}\right),$$

$$\therefore \int_{x=0}^{x}\left(e^{qV/kT} - 1 - \frac{qV}{kT}\right)^{-1/2} dV = \pm\left(\frac{2\rho}{W}(1-\alpha)j_0\frac{kT}{q}\right)^{1/2}\int_0^x dx, \quad (5)$$

and using the boundary condition $V=V_0$ at $x=0$ this is the complete solution in terms of an integral. For our purposes, an approximate solution will be sufficient. At room temperature $(q/kT)\approx 39$ volts^{-1} and in a power transistor $V\sim 1$ volt. We therefore make the approximation $qV/kT \gg 1$ and neglect the term $[-1-qV/kT]$ in comparison with the exponential. The solution so obtained will be valid in regions not too far distant from the base connection. Making this approximation, (5) can be integrated directly to give

$$-\frac{2kT}{q}(e^{-qV/2kT} - e^{-qV_0/2kT}) \approx \pm\left(\frac{2\rho}{W}(1-\alpha)j_0\frac{kT}{q}\right)^{1/2} x,$$

whence

$$V(x) \approx -\frac{2kT}{q}\log_e\left[e^{-qV_0/2kT} \mp \frac{q}{2kT}\sqrt{\frac{2\rho}{W}(1-\alpha)j_0\frac{kT}{q}}\cdot x\right]. \quad (6)$$

We shall be more interested in emitter current density and this is given by

$$j(x) = j_0(e^{qV/kT}-1)$$

$$\therefore j(x) \approx j_0\{e^{qV_0/kT}[1 \pm Axe^{qV_0/2kT}]^{-2} - 1\}, \quad (7)$$

where

$$A = \frac{q}{2kT}\sqrt{\frac{2\rho}{W}(1-\alpha)j_0\frac{kT}{q}}. \quad (8)$$

Now $A>0$ and we know that $j(x)$ decreases as x increases from zero so that the ambiguous sign should be plus, giving

$$j(x) \approx j_0\{e^{qV_0/kT}[1 + Axe^{qV_0/2kT}]^{-2} - 1\}. \quad (9)$$

The expression (9) together with (8) then gives an approximate description of the fall-off of emitter current density away from the base lead connection. Under our assumptions, this solution is only valid fairly close to the base connection where the approximation $[qV(x)/kT] \gg 1$ is valid. We can thus simplify (9) and write to sufficient approximation

$$j(x) \approx j_0 e^{qV_0/kT}[1 + Axe^{qV_0/2kT}]^{-2}$$

$$\approx j(0)\left[1 + x\sqrt{\frac{\rho}{2W}(1-\alpha)\frac{q}{kT}j(0)}\right]^{-2}, \quad (10)$$

and, if we assume $\gamma=1$, then using (1)

$$j(x) \approx j(0)\left[1 + x\sqrt{\frac{\rho}{4}\frac{q}{kT}\frac{W}{D_p\tau_p}j(0)}\right]^{-2}. \quad (11)$$

We should notice, in looking at this result, that the scale in x is determined by the factor

$$\frac{\rho}{4}\frac{q}{kT}\frac{W}{D_p\tau_p}j(0),$$

and the significant thing to note is that this varies as $W^{1/2}\tau_p^{-1/2}$, so that we should strive for as long a lifetime and as small a base layer thickness as possible, to achieve minimum variation. This requirement of a thin base layer is at first sight a little surprising until we remember that though the base resistance varies as $1/W$, the factor $(1-\alpha)$ varies as W^2.

We should point out that there are two reasons why we wish to minimize this variation in emitter current density. The first is the obvious desire to pass as much current as possible, which clearly presents fewer problems if its density is approximately constant. The second is to preserve a large value of $\beta=\alpha/(1-\alpha)$ at high currents. Since to a first approximation β varies as $1/j$, the integrated value of β for the transistor as a whole is greatest for a uniform current distribution.

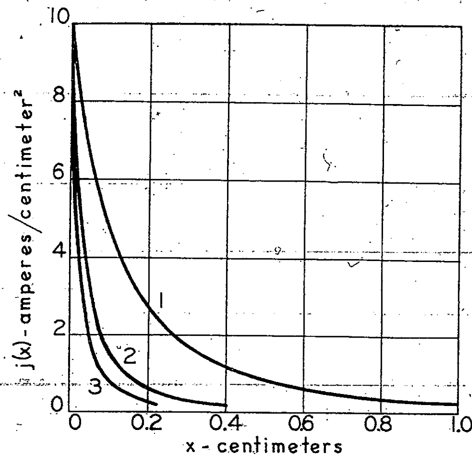

Fig. 4—Calculated fall-off of injected current density $j(x)$ as a function of distance x from the edge of the emitter for the three cases discussed in the text.

Numerical Examples

To give concrete realization to some of our ideas, we shall plot the dependence shown in (11) for three more or less typical cases. The transistor structure is as shown in Fig. 3.

1. $W = 0.002$ cm

   $\tau_p = 100$ μsecs

   $\rho = 0.5$ ohm-cm

   $j(0) = 10$ amps per square cm

then

$$j(x) = 10[1 + 4.7x]^{-2} \text{ amps per square cm.}$$

2. $W = 0.002$ cm

   $\tau_p = 10$ μsecs

   $\rho = 0.5$ ohm-cm

   $j(0) = 10$ amps per square cm

then

$$j(x) = 10[1 + 14.9x]^{-2} \text{ amps per square cm.}$$

3. $W = 0.006$ cm

   $\tau_p = 10$ μsecs

   $\rho = 0.5$ ohm-cm

   $j(0) = 10$ amps per square cm.

then

$$j(x) = 10[1 + 25.8x]^{-2} \text{ amps per square cm.}$$

These three current distributions are plotted in Fig. 4. (Numbers correspond.)

More Refined Theory for Very High Level Injection

We noted above when discussing the assumptions of our simple theory that only at quite moderate injection levels can we assume that emitter efficiency is unity and that the base region resistivity is not changed by the injected carrier density. We shall now attempt to formulate a more complete theory where these effects are not neglected.

We shall make the same assumptions as in our earlier theory except that we shall not neglect the effect of injected carriers on the base resistivity.

For clarity, consider a *p-n-p* transistor, then for an emitter current $j(x)$ and base width W, the average injected carrier density in the base region is $\frac{1}{2}p$ where

$$i(x) = -gqD_p\frac{p}{W} \quad \text{where} \quad 1 \leqq g \leqq 2.$$

This injected hole density carries with it an equal electron density which contributes to the majority carrier conductivity. The conductivity due to this cause is

$$\sigma' = \frac{1}{2}\,\frac{\mu_n W}{gD_p}j(x),$$

so that the total conductivity is

$$\sigma = \sigma_0 + \sigma' = \sigma_0 + \frac{1}{2}\,\frac{\mu_n W}{gD_p}j(x), \tag{12}$$

where σ_0 is the original conductivity of the base region. We shall write

$$\rho(x) = \frac{1}{\sigma(x)}\cdot \tag{13}$$

Referring to Fig. 3, the current $(1-\alpha)j(x)dx$ originating in dx at x flows through a resistance

$$\int_0^x \frac{\rho(y)}{W}dy,$$

and so contributes a voltage drop

$$[(1-\alpha)+(1-\gamma)]j(x)dx\int_0^x \frac{\rho(y)}{W}dy.$$

Summing all such voltage drops, the emitter-base potential at point x_0 is thus

$$V(x_0) = V_0 - \int_0^{x_0}[(1-\alpha)+(1-\gamma)]j(x)\int_0^x\frac{\rho(y)}{W}dy\cdot dx$$
$$-\int_{x_0}^{\infty}[(1-\alpha)+(1-\gamma)]j(x)dx\int_0^{x_0}\frac{\rho(y)}{W}dy. \tag{14}$$

This integral equation, together with (12) and (13), and Webster's expression for γ, is sufficient to determine $V(x)$.

For the particular case we considered before when α, γ and ρ are assumed constant, (14) reduces immediately to (4) upon differentiating twice with respect to x_0. This reduction cannot, however, be easily performed in the general case.

For our present purposes we shall not require this general solution, and it will be sufficient to observe some general properties of the current distribution which it yields. The distribution will be quite similar to that given by (11); it will, however, tend to be more steep all over because of the appreciable magnitude of $(1-\gamma)$, though for small x this will be compensated for somewhat by the increased base region conductivity. It is hoped to obtain a detailed solution of (14) at some later date.

We should perhaps emphasize again at this stage that the result (10) was derived neglecting completely the effects of surface recombination. The expression $(1-\alpha)$ used in (10) includes only contributions from nonunity emitter efficiency and from transport loss by volume recombination. In the case where surface recombination is considered, (10) is still valid, and $(1-\alpha)$ still has the value given above, to a first approximation. There is, however, an additional loss of carriers to the surface—mostly carriers injected near $x=0$ so that they have little effect on the phenomenon we have been discussing—which lowers the over-all α of the transistor in a manner which is approximately independent of the current.

Conclusions—Design of Emitter Region

From the discussion presented above and the curves plotted in Fig. 4, it is clear that we cannot achieve a large emitter current merely by increasing the area of the emitter since only those parts of the emitter closest to the base electrode connection will carry appreciable current. In fact, in even the best case considered, the emitter current density has fallen by a factor of 2 by the time we have gone in 1 mm from the edge of the emitter nearest to the base lead.

These remarks make the next step obvious—we must arrange the emitter-base geometry so that the electrical resistance between any point on the emitter and the base connection is as uniform as possible. It is also desirable that this resistance be as small as possible. A convenient geometry at once suggests itself. Let us make the emitter a very long thin bar, and use for a base electrode two long bars flanking the emitter and as close to it as possible. The collector, of course, is on the opposite side of the semiconductor as is usual with alloy junction transistors. This arrangement is shown in Fig. 5. The geometry is really similar to that discussed previously and illustrated in Fig. 3 except that the transistor structure extends only a distance d (the width of the emitter bar) in the x direction and another base contact is applied at this edge.

The results of the calculation made above cannot strictly be applied to this case since now the boundary condition is

$$\frac{dV}{dx} = 0 \quad \text{at} \quad x = \frac{d}{2}.$$

The analysis for this case is, however, more difficult and (11) can be used to give a first approximation to the current distribution. The true current distribution will vary less than that predicted by (11) but it is still a useful approximation.

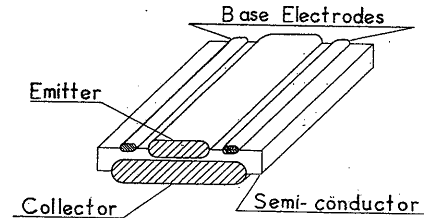

Fig. 5—Proposed electrode configuration for high power transistor.

Standard Emitter-Base Configuration

Optimum dimensions could of course be worked out for each particular transistor design, but the state of the art is not yet such that transistor dimensions (particularly base layer thickness W) and germanium properties can be controlled with great precision. It is therefore convenient in practice to design a standard configuration which can be modified to yield various desired transistor characteristics. In this section, we shall propose such a configuration and show how it can be modified to give transistors of different ratings.

As far as materials are concerned, germanium is available in a large range of resistivities and lifetime can be held to greater than 100 μsecs. The resistivity to be used is dictated by breakdown requirements in general, though often no very good correlation appears to exist for large area junctions. We wish to use as low a resistivity as is compatible with sufficient breakdown voltage. Usually 0.5–3 ohm-cm germanium will satisfy most common requirements, selected units having breakdowns (which are soft in any case) in excess of 100 volts.

From the point of view of our bias cut-off effect, the smaller we make the width d of our emitter bar, the more efficiently do we use this area. However, several other factors enter in the opposite sense and so we can in fact find a finite optimum width. The first of these "other factors" is surface recombination which we mentioned earlier. The number of minority carriers lost at the surface is approximately proportional to emitter perimeter, so that very thin emitters, having a large ratio of perimeter to area, are bad from this point of view. We must also remember that our transistor will be made by an alloying process and depth of penetration is finite. In order that the base region be as uniformly thin as possible, we require the emitter width to be large compared to the alloying depth. We should finally mention that very narrow emitters will be mechanically weak and difficult to fabricate.

The relative importance of these various effects depends upon many things, surface recombination velocity (which is closely dependent on manufacturing technique) being one of the most important. Ideally an analysis similar to that of Moore and Pankove[3] should be carried out for this transistor design and for particular surface conditions. In the absence of such an analysis, experience shows that for typical transistor requirements and manufacturing methods an emitter width of 1–2 mm is close to optimum. The base electrodes are equally spaced and as close to the emitter as possible, their proximity being dictated primarily by mechanical considerations.

We have said very little about the collector as yet, and indeed little can be said at the moment. From the viewpoint of collection efficiency, the collector should be as large as possible, but on the other hand, increasing the collector area increases the saturation current I_{co} and also increases the likelihood that the collector region contain some serious imperfection. In practice, if the collector is made to extend one or two diffusion lengths sideways past the emitter, the collection will be sufficiently good.

Possible Electrode Configurations

So far, we have merely designed a "standard" emitter-base configuration. We shall now show how it can be modified to yield several important and useful transistor types.

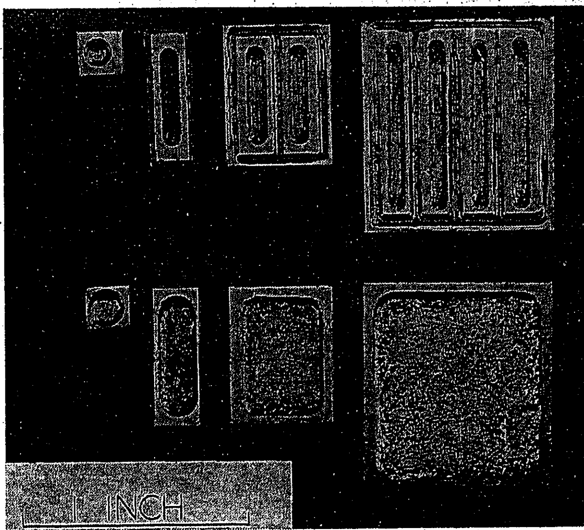

Fig. 6—Transistor types: (left to right)—(1) X-78 standard power transistor (without base electrode); (2) single-bar transistor; (3) double-bar transistor; (4) multi-bar transistor. The upper row shows the emitter-base side of the transistors and the lower row the collector side.

Single-Bar Type: For medium power applications where currents up to 2 amps or so are required, it is convenient to make a transistor consisting merely of a short length (say 1 cm) of our standard configuration. The dimensions of such a structure are mechanically reasonable and units made to this design behave as expected. A unit of this type is shown in Fig. 6, while

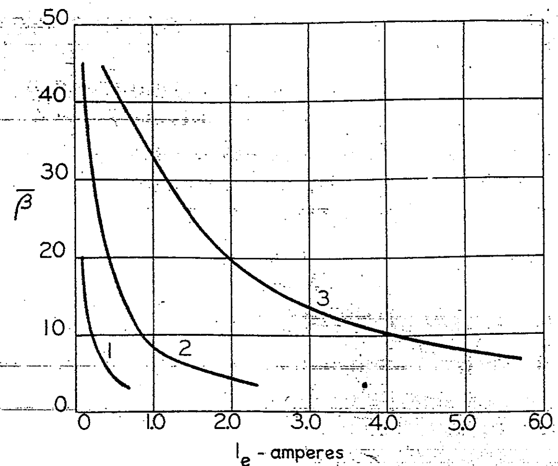

Fig. 7—Average grounded emitter current gain $\bar{\beta}=I_c/I_b$ as a function of I_e for (1) X-78 standard power transistor, (2) single-bar unit, (3) double-bar unit. All *p-n-p* types with $V_c=-12$ volts.

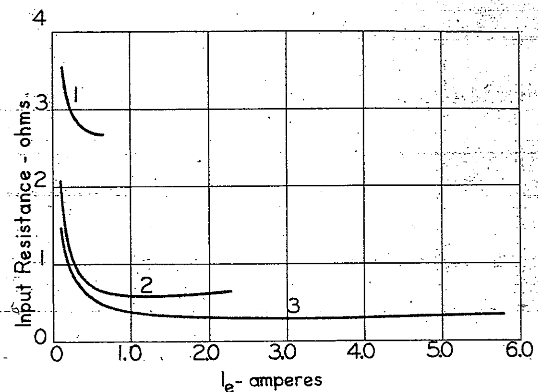

Fig. 8—Average grounded base input resistance V_e/I_e as a function of I_e for (1) X-78 standard power transistor, (2) single-bar unit, (3) double bar unit. All *p-n-p* types with $V_c=-12$ volts.

a typical plot of $\bar{\beta}=I_c/I_b$ vs I_e is shown in Fig. 7. A similar curve for input resistance is shown in Fig. 8.

Multi-Bar Type: For higher current requirements, the standard configuration is inconveniently long, so a convenient modification is to place two or more of these configurations parallel to each other. Units of this type are shown in Fig. 6. If many bars are used, they may be conveniently connected together into a comb-like structure, the base leads being connected similarly.

It should be pointed out that despite a superficial resemblance to the comb-like power transistor reported by Hall,[5] this unit is of a basically different geometry, Hall's transistor having base and collector geometry interchanged with respect to our unit.

One of the most successful units of this type is a double-bar unit, 1.5 cm in length, which has an efficient emitter current rating of about 5 amps. A $\bar{\beta}$ vs I_e curve for such a unit is shown in Fig. 7 and a similar curve for input resistance in Fig. 8.

Annular Type: As a further variation, a length of standard configuration can be bent to form a ring so

[5] R. N. Hall, "Power rectifiers and transistors," Proc. I.R.E., vol. 40, pp. 1512–1518; November, 1952.

that the emitter is an annulus. Several concentric rings of this type can then be built up into a larger unit. Units of this type have some slight disadvantage in base resistance in the innermost ring but are otherwise approximately equivalent to a multi-bar unit. It appears that fabrication methods are not so simple for a unit of this type as for a bar-type unit.

Thermal and Mechanical Design

Thermal Design

We have completed our brief survey of some of the electrical considerations in power transistor design. We turn now to have a brief look at some of the thermal and mechanical problems for which solutions must be sought.

It is well known that as its temperature increases, the number of free carriers in a semiconductor also increases until the number of holes and electrons become comparable, at which stage the semiconductor is said to be in its intrinsic range, and many of the attributes necessary for transistor action disappear. Even before the intrinsic range is reached, many undesirable phenomena appear, chief among these being the increase in the saturation current I_{co} flowing in the collector circuit. For germanium, the temperature at which these effects become severe is ordinarily less than 100 degrees C.; for silicon, it is considerably higher.

In operation, a power transistor usually has a considerable amount of power dissipated within its structure in the form of heat, and unless this heat is removed efficiently, the temperature will rise and the undesirable effects discussed above will set in. In practice, we must therefore provide a suitable low resistance path through which heat can be removed from the transistor. This may be done by means of conduction, convection or radiation, or a combination of these three. The usual method is to mount the transistor on a piece of metal to which is attached a suitable system of cooling fins. For the small temperature rises encountered, Newton's law of cooling is valid and the maximum allowable dissipation is

$$P = \frac{T - T_A}{R}, \tag{15}$$

where T is the maximum permissible device temperature, T_A is the ambient temperature and R is the thermal resistance between the device and the ambient.

Design of a proper fin system is very important, since if we make R small, we increase P the allowable dissipation. The subject of fin design falls outside the scope of this paper so we shall not consider it here.

Another thermal problem which we must consider is that of stability. Since the saturation current I_{co} increases with temperature, the heat generated by this current is $V_c I_{co}$, where V_c is the collector voltage, and this similarly increases with temperature. The device may thus be temperature unstable unless it is properly designed.

Suppose the transistor is in equilibrium at a temperature T and that its collector voltage is V_c and its collector current I_c. Consider a small random increase in temperature dT. This causes a corresponding increase dI_c in I_c. The increased dissipation is thus

$$dP = V_c dI_c.$$

Suppose the thermal resistance between the transistor and the ambient is R then the increase dT allows an additional conduction of heat

$$dH = \frac{dT}{R}.$$

Then the equilibrium is stable and the system will return to its initial temperature state if

$$dP < dH$$

$$V_c dI_c < \frac{dT}{R}$$

$$V_c \frac{dI_c}{dT} < \frac{1}{R}. \tag{16}$$

Eq. (16) then determines the critical value of T for stability.

We should note that the thermal resistance R in (16) is, more properly speaking, an impedance. In a typical case, the collector of a transistor may be soldered to a large copper or aluminum fin structure. A first approximation to R is then given by the electrical analog shown in Fig. 9. Here, R_1 is the thermal resistance between the collector junction and the fin assembly, R_2 is the thermal resistance between the fin assembly and the ambient, and C is the heat capacity of the fin structure; ordinarily $R_2 \gg R_1$.

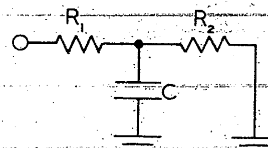

Fig. 9—Electrical analog for thermal impedance R.

For perturbations much less than R_1C in duration, the value of R to be used in (16) is essentially R_1 while for steady-state considerations ($t \gg R_1C$), the appropriate thermal resistance is $R_1 + R_2$. Electrical analogs of this type may be helpful in the investigation of more complicated cases, such as switching circuits, where there is a very great difference between peak and average dissipations.

We should note that a higher temperature limit can be achieved by

1. Decreasing R.
2. Decreasing I_{co}.
3. Decreasing V_c.

All these should be borne in mind in designing transistor structures and in using transistors in circuits. When we

have done the best we can with the transistor along the above lines, we can improve stability further by incorporating suitable thermally-sensitive elements in the electrical circuit.

We noted above three means by which thermal stability could be improved. The first two of these, namely decreasing R and decreasing I_{co}, can be considered a little further. We should note that by (15) the maximum allowable dissipation will also be increased if we decrease R.

R can be minimized by using materials of high thermal conductivity in the assembly. Indium ($K=0.057$ cal/cm sec degrees C.) is particularly bad in this respect and R can be reduced considerably if the indium collector (for a p-n-p transistor) is reduced in thickness as much as possible and then soldered to a piece of copper. The major contribution to R comes, however, between the fin structure and the ambient, and can be reduced by proper fin design. We shall not consider this further here.

I_{co} can be minimized by using as small a collector area as is compatible with other requirements, and, other things being equal, by using a low resistivity semiconductor with large energy gap. Since we are mostly concerned with germanium, we cannot alter the energy gap but we should note that semiconductors with larger energy gaps, for example silicon or silicon-germanium alloys, should give improved performance in this respect. I_{co} in this context includes the leakage current as well as the true saturation current. The temperature dependence of this leakage component is not as great as that of the saturation current, but we should try to minimize its value by choice of germanium (orientation may be important) and by proper etching technique.

Mechanical Design

This is a subject which we will discuss briefly without intending to ignore its importance in final transistor design. The mechanical package must be such as to maintain those desirable electrical characteristics which have been built into the transistor and to preserve them from environmental influences such as humidity, shock, vibration, etc.

The transistor package may either be integral with the fin structure (e.g., Transistor Products type X-107), or may be a separate unit which is subsequently attached to the fin structure. This latter procedure has much to recommend it in the case of high power transistors where the fin structure required may vary considerably depending upon the application. A package of this type, housing a double-bar transistor of about 20-watts dissipation rating is shown in Fig. 10. Also shown is an X-107 transistor package rated at 3-watts dissipation in a 25 degrees C. ambient and an experimental package of intermediate rating. The fin structure for the 20-watt unit may typically be a larger version of that shown on the intermediate design, or alternatively, the transistor may be bolted to a chassis or other convenient "heat sink."

The package may be either plastic filled or hermetically sealed to afford humidity protection. Plastic filling is cheaper and easier and provides a very good seal if proper encapsulants are used, but in some cases the more certain protection of an hermetic seal may be desired, the container then either being evacuated or filled with some inert substance.

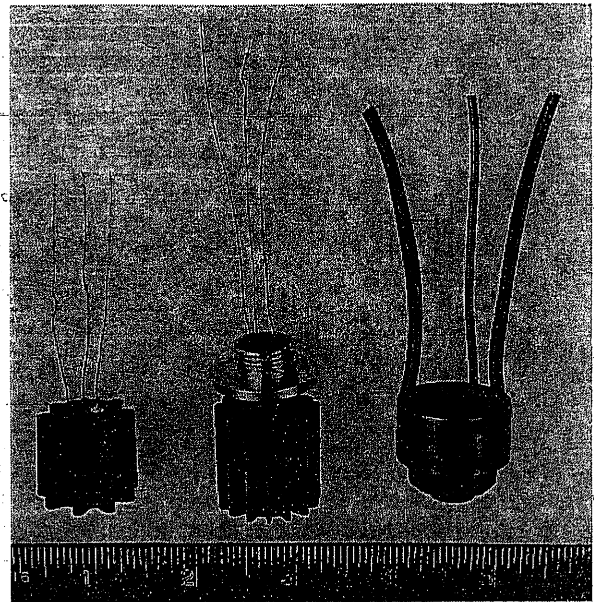

Fig. 10—Power transistor assemblies: (1) X-107 (3 watts), (2) experimental unit (about 5 watts), (3) high power unit, cooling fins not shown (up to 100 watts).

Conclusion

It has been the aim of this paper to present certain considerations which are important in the design of high power transistors. The theory is of quite a general nature and shows how units can be designed to have improved performance at high currents and at high dissipations. Although the theory is generally useful, it has been here only specifically applied to alloyed junction transistors. The results reported are for p-n-p units but theoretical considerations as well as measurements indicate that n-p-n units should be even more suitable for high current applications. By following the precepts outlined above, transistors with class A output ratings over 50 watts have been successfully designed and fabricated.

Acknowledgment

I should like to thank the members of the Engineering Department of Transistor Products, Inc., who helped make this study possible, both by discussion, and by experimental assistance. I should also like to thank the the Australian Commonwealth Scientific and Industrial Research Organization for giving me the opportunity to do this work.

The study was supported by Signal Corps Contract No. DA-36-039-SC-63135.