CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION, CFSTI
DOCUMENT MANAGEMENT BRANCH 410.11

LIMITATIONS IN REPRODUCTION QUALITY

Accession #

1. We regret that legibility of this document is in part unsatisfactor Reproduction has been made from best available copy.

2. A portion of the original document contains fine detail which may make reading of photocopy difficult.

3. The original document contains color, but distribution copies are available in black-and-white reproduction only.

4. The original distribution copies contain color which will be shown in black-and-white when it is necessary to reprint.

5.

RESEARCH AND INVESTIGATION OF INVERSE EPITAXIAL UHF POWER TRANSISTORS

Technical Documentary Report No. AL TDR 64-207
September 1964

AF Avionics Laboratory
Research and Technology Division
Air Force Systems Command
Wright-Patterson Air Force Base, Ohio

Project No. 4460, Task No. 446001

(Prepared under Contract No. AF 33(657)-11403
by Clevite Corporation, Semiconductor Division,
Shockley Research Laboratory Palo Alto, Calif.)

Foreword

This report was prepared by Clevite Corporation, Semiconductor Division, Shockley Research Laboratory under USAF Contract No. 33(657)-11403. The contract was initiated under Project No. 4460, Task No. 446001. The work was administered under the direction of the Air Force Avionics Laboratory Research and Technology Division, Neil DiGiacomo, Project Engineer.

This report covers work conducted from July 1963 to June 1964.

Dr. A Goetzberger and Dr. R. M. Scarlett were principal investigators on the program. This report was compiled by R M. Scarlett. Substantial contributors to the work were R. Finch, R. Gereth, A. Goetzberger, W. Shockley, V. Williams and N Zetterquist.

Abstract

This report describes development work toward a 10 watt 500 Mc silicon npn epitaxial transistor. An extensive design theory permits calculation of important material and geometrical parameters to realize a given performance. Diffusion studies produced a greater understanding of the emitter dip effect, but lacked the control necessary for thin, heavily doped layers. Ion bombardment doping and epitaxial base growth are described, both these methods are promising but require further development. Devices were produced giving up to 7 watts at 500 Mc with 4 db gain.

An important device principle, the shorted emitter, is described. Methods of evaluating contact resistance and designing control structures for diffusion evaluation are discussed.

Publication Review

Publication of this technical documentary report does not constitute Air Force approval of the report's findings or conclusions. It is published only for the exchange and stimulation of ideas.

TABLE OF CONTENTS

List of Tables

List of Figures

INVERSE EPITAXIAL UHF POWER TRANSISTORS

1. INTRODUCTION

This report describes work performed toward the development of an Ultra High Frequency (UHF) power transistor intended to deliver 10 watts at 500 Mc with a power gain of at least 6 db.

An npn silicon epitaxial structure was chosen, and the principal effort went into producing the required base and emitter regions by diffusion of boron and phosphorus. The diffusion techniques could not be developed to the point where the design values for both the base sheet resistance and cutoff frequency could be met simultaneously, or where good reproducibility in the very thin layers was achieved. Other methods of doping were investigated but not perfected. Ion bombardment was shown to be capable of producing hard junctions, but has severe surface erosion. Epitaxial growth of base or emitter layers is promising, but still plagued with non-uniform growth and poor control of resistivity.

The best performance achieved was a power output of 7 watts at 500 Mc with 4 db gain and 40% efficiency. There are two main factors limiting this performance. One is high base sheet resistance which impairs the maximum power capability, and the other is excess capacitance between the aluminum overlay bonding areas and the collector, which reduces power gain. This capacitance could be reduced by various means, but none of these could be fully put into effect during the contract period.

Manuscript released 20 August 1964 for publication as an RTD Technical Documentary Report by author.

Although the desired goal was not reached, a number of contributions were made. An extensive design theory was developed which takes into account important large signal effects and provides a basis for choosing the necessary emitter-base junction periphery in terms of the attainable base sheet resistance. A number of important experiments were performed to elucidate the mechanism of the "emitter dip" effect (enhanced base diffusion under the emitter) which it is important to understand and control if very thin base layers are to be made reproducibly. Initial steps were taken in the development of ion bombardment doping and epitaxial base techniques. A new transistor structure, the "shorted-emitter" was proposed and analyzed. This structure looks promising for further developments in the UHF transistor field.

2. DEVICE DESIGN

2.1 Preliminary Design

The following design considerations are presented to indicate the feasibility of making the required device and to establish some of the main parameters. A power output of 10 watts at 500 Mc with an efficiency of 40% and a power gain of at least 6 db is desired. The supply voltage is to be about 25 volts, the collector breakdown BV_{CEX} should be greater than 60 volts, and the thermal resistance less than $7^{\circ}C$/watt since 15 watts may be dissipated, and no more than a $100^{\circ}C$ temperature rise is desirable. An $NPN^{-}N^{+}$ silicon structure is assumed, the P base and N emitter region being diffused or grown epitaxially on an $N^{-}N^{+}$ epitaxial collector region.

a) Collector breakdown

Experience has shown that to achieve a 60 volt breakdown in a shallow diffused planar junction (which behaves like an abrupt junction) a collector resistivity in the underlying N^{-} layer must be at least 1 ohm-cm, corresponding to an impurity concentration of 5×10^{15} cm^{-3}. The space-charge layer at maximum voltage is calculated to be 4.4 μ wide, so the N^{-} layer should be at least this wide. To allow for the base layer and for diffusion from the N^{+} collector bulk, a nominal N^{-} layer width of 10μ will be assumed. The breakdown in a shallow planar junction occurs around the edges which have sharp curvature; because the effective current gain of the transistor in this region is small, BV_{CEO} is not much less than $BV_{CEX} \simeq BV_{CBO}$.

b) Thermal

The dissipated power in the device is spread over an effective area A. It is supposed that this heat must flow through 100μ of silicon

to the mounting tab, and through the equivalent of another 50μ of silicon to the heat sink. Since the thermal conductivity of silicon is 0.8 watts/cm $^{o}C^{-1}$, the required area is at least

$$A = \frac{1.5 \times 10^{-2}}{0.8 \times 7} \simeq 2.7 \times 10^{-3} \text{ cm}^2 \qquad (1)$$

The device to be proposed will have a collector area of $A_c = 3 \times 10^{-3} \text{cm}^2$ and an emitter area of $A_e = 1 \times 10^{-3} \text{cm}^2$. To allow for the spreading of heat from the narrow emitter stripes into the collector body, the effective area is closer to A_c than to A_e. Accordingly, the proposed design will likely meet the thermal resistance requirement.

c) Cutoff frequencies

Transistor power gain at high frequencies is ordinarily considered to vary as $(f_{max}/f)^2$, where f_{max} is the well-known figure of merit frequency:[2]

$$f^2_{max} = f_t/8\pi r_b C_c \qquad (2)$$

Here, f_t is the frequency beyond which the common-emitter current gain magnitude drops below one, r_b is the effective resistance in the base layer through which the input (base) current flows, and C_c is the direct collector-base feedback capacitance.

To achieve a gain of 6 db (factor of 4) at 500 Mc requires $f_{max} > 1000$ Mc. Extensive experiments with diffused layers have shown that it is difficult to produce base layers reliably which are thin enough (< 0.4μ) over such relatively large areas so that $f_t > 600$ Mc. (By using epitaxial layers, as discussed later, this limitation may be less serious.)

Hence, from Eq. (2):

$$r_b C_c < 2.4 \times 10^{-11} \text{ sec} \tag{3}$$

The cutoff frequency f_t is composed of the frequency f_t' resulting from the transit-time of carriers across the base layer, and the frequency $f_c = 1/wt_c$, where t_c is the transit time across the collector space-charge layer: $1/f_t = 1/f_t' + 1/f_c$. At an assumed peak voltage of 50 volts, the collector space charge layer, see a) above, is about 4.4μ wide. Since the limiting carrier velocity[3] is around 5×10^6 cm/sec, the maximum collector transit time t_c is $4.4 \times 10^{-4} / 5 \times 10^6 = 0.9 \times 10^{-10}$ sec, corresponding to $f_c \simeq 3500$ Mc. This means that to realize $f_t = 600$ Mc, the base-layer cutoff must be nearly:

$$f_t' = 720 \text{ Mc} \tag{4}$$

a value considered feasible. The base layer width required for this frequency is of the order of 0.35μ, based on experience with diffused layers having a range of cutoff frequencies from 300 to 1200 Mc (see Section 4).

d) Collector capacitance

With the 1 ohm-cm collector resistivity needed for a 60 volt breakdown rating, the space charge layer width is calculated to be 4.4μ wide at the peak operating voltage of 50 volts. The appropriate capacitance to use for large signal calculations is twice the capacitance at peak voltage (for an abrupt junction), thus $C = 2\epsilon / 4.4\mu = 2 \times 10^{-12} / 4\ 4 \times 10^{-4} = 4500$ pf/cm^2. According to Eq (1), the area of this junction must be at least 2.7×10^{-3} cm^2; allowing a slightly larger value, the total collector capacitance becomes.

$$C_c = 3 \times 10^{-3} \times 4500 \simeq 14\,\text{pf} \tag{5}$$

e) Base resistance

From Eqs. (3) and (5), one finds.

$$r_b < 1.7 \text{ ohms} \tag{6}$$

With this value. an estimate can be made of the minimum fineness of geometry required, in the following manner. Consider a geometry consisting of parallel nterdigitated emitter and base stripes. Fig. 1 shows a cross section through a single pair of stripes extending from the center line of the base to the center line of the emitter, a distance of L_c. This structure is assumed to repeat periodically in mirror image fashion with total period $2L_c$. It is further assumed that the spacing L_b is equal to $L_c/3$ and that emitter stripe half-width L_c is equal to $L_c/2$, these being approximately convenient ratios to produce. It is evident that within an area A_c, the total emitter-base junction periphery is $L_p = A_c/L_c = A_c/3L_b$.

The base current, represented by B in Fig. 1, has to flow a distance L_b through the extrinsic base sheet resistance R_s' ohms per square. A further base current flows under the emitter through the intrinsic base sheet resistance R_s to charge the collector capacity there. By considering the power loss in these resistances, it can be shown that the effective total base resistance is.

$$r_b \simeq R_s' L_b / L_p + R_s L_b / 4L_p \tag{7}$$

Conservative values for R_s' and R_s are 500 and 4000 ohms per square respectively for the layer thicknesses under consideration, thus

$r_b = 1500\ L_b / L_p = 4500\ L_b^2 / A_c$. Inserting $r_b = 1.7$ ohms from Eq. (6) and $A_c = 3 \times 10^{-3}$ cm^2 results in $L_b = 11\mu$. Consideration of large signal effects discussed in subsections 2.2, 2.3, and 2.4 below will show that L_b should be of the same order of magnitude. Thus the preliminary design value for base resistance should easily be realized. A dimension of 10μ for L_b implies an emitter-base junction periphery of $L_p = A_c / 3L_b = 1$ cm, according to the above assumptions.

2.2 Power Output and Efficiency

a) Load resistance and output resistance

To achieve maximum power gain at maximum power output, the necessary load resistance must approximately match the transistor output resistance. It is well known,[2] and consideration of the equivalent circuit of Fig. 2b will show, that the output resistance is $1/\omega_t C_c \simeq 19$ ohms for $f_t = 600$ Mc and $C_c = 14$ pf. To estimate the load resistance, first consider ideal Class A operation. With a 25 volt supply, the peak-to-peak ac output voltage is 50 volts. For 10 watts output, the maximum collector current (peak-to-peak ac current) must be $8 \times 10/50 \simeq 1.6$ amps. The load resistance required is evidently $50/1.6 = 31$ ohms. To achieve greater efficiency, an operating mode closer to Class B is used in practice, requiring a smaller load resistance. This is satisfactorily close to the expected 19 ohm output resistance.

b) Collector body drop

By this term is meant the minimum voltage drop occurring in the collector region at the time of maximum current through the device. This voltage is similar in nature to the commonly specified saturation voltage measured under dc conditions, except that it can be considerably larger

in large-signal high-frequency operation because of the increased current concentration along the edge of the emitter. The length of emitter periphery required is generally determined by a limit set on this minimum collector voltage at the peak ac current and by base widening effects caused by high current density in the space-charge layer. Such a periphery is generally more than adequate to provide a suitably low dc saturation voltage.

Typical voltage and current waveforms observed experimentally show nearly sinusoidal current, and collector voltage clipping at a minimum value of V_{min}. Although the circuit is usually arranged so that "Class B" operation is expected (see Section 5). at a frequency near f_t it is difficult to cut off the current completely for reasons outlined below in subsection 2.4. A minimum current I_{min} exists. Greater than 50% efficiency can often be obtained by allowing the voltage waveform to clip considerably. Evidently, the values of V_{min} and I_{min} greatly influence efficiency, and must be kept within reasonable limits. Estimates of these quantities follow, and are considered in more detail in the subsequent subsections.

An approximate calculation of V_{min} is given in subsection 2.3 below, with reference to the device cross section in Fig. 1. The current J_n represents the peak carrier current per unit emitter periphery, crossing the base layer and flowing into the collector body. This current is in units of amps/cm. Because of the well known transverse biasing effect caused by the base current B, J_n is emitted principally from the edge of the emitter nearest the base contact over an effective width L_f. This effect is much more important at high frequencies than at dc, because the base current is much larger. As L_f becomes small, very high current densities J_n/L_f result, causing high electric fields in the collector bulk. To avoid "stagnant regions"[4] and consequent base widening effects,[5] the

maximum field in the collector must be kept below values where hot-electron effects set in, and the carrier mobility is seriously reduced [6] With this restriction, a good working approximation to the peak voltage drop in the collector is:

$$V_{min} \leq 2\rho_c J_n \tag{8}$$

In the present example, tentative values are $\rho_c = 1$ ohm-cm and $J_n = 1.6$ amps/cm (perimeter $L_p = 1$ cm, peak current = 1.6 amps) so that $V_{min} \leq 3.2$ volts, which is not excessive.

c) Minimum current

As the collector-base voltage swings from minimum to maximum during the rf cycle, charge associated with the collector capacitance is pushed into the base layer. This charge consists of holes which are forced out of the widening collector space-charge layer, and it appears uniformly over the collector-base junction. That portion of it under the base contact flows out the base terminal as ordinary capacitive charging current. The charge appearing under the center of the emitter, however, may not be able to flow out through the high resistance base layer, particularly as this layer is effectively shunted by the distributed emitter junction capacity. Any charge remaining in the base layer will require injected charge from the emitter for neutralization, and accompanying this charge is a current flow from emitter to collector occurring at a time when the collector voltage is maximum. This unwanted current is represented by I_{min}, and it is given by $\omega_t Q$ according to the charge control transistor model, where Q is the charge remaining in the base layer. This effect is represented by the dotted resistance $1/\omega_t C_3$ in Fig. 2a, where C_3 is the capacitance of the collector junction under the region in question

An exact calculation of I_{min} is not practical because of the non-linearity of the problem. However, an estimate of whether I_{min} is serious can be made in the following manner. The charging current from the collector capacity flows through a region of sheet resistance R_s having a distributed capacitance C_e (the capacitance of the emitter junction) to ground. The behavior of a signal on such an RC transmission line is governed by a diffusion equation. At frequency ω, the signal penetrates a distance of the order of

$$L_\omega = (1/\omega R_s C_s)^{1/2} \tag{9}$$

This is roughly the distance under the emitter from which the collector charging current can emerge and flow to the base contact. Measurements suggest that a reasonable average value of C_e in the region of low forward bias is 4×10^{-7} F/cm^2. Using R_s = 5000 ohms/square and $\omega = 2\pi \times 500$ Mc, L_ω from Eq. (9) comes out to be 4μ. This is considerably less than the proposed emitter halfwidth L_e of 10μ (refer to Fig. 1 and see 2.1e above, where $L_b \simeq 10\mu$ is derived to provide the required base resistance; under the assumptions of this analysis, $L_e \simeq L_b$). As a result, the collector charge coming from under the emitter in a region 6μ wide will not be able to flow out of the base terminal. The collector charge on a per unit area basis is $Q_c = V_c C = 25$ volts x 4500 pf/cm^2 (see 2.1d) $\simeq 10^{-7}$ c/cm^2. The total amount flowing into the emitter capacitance is $Q_c L_p \times 6\mu$, where L_p is the emitter periphery of 1 cm derived in part 2.1e above. Probably less than one-half of this is effective in provoking injection from the emitter, thus the minimum collector current is expected to be roughly $I_{min} < \omega_t Q_c L_p \times 3\mu = 0.11$ amps under maximum signal conditions. This is less than 10% of the peak current of 1.6 amperes. Furthermore, it will be shown below that a base sheet resistance less than 5000 ohms/square will be desirable, so that I_{min} will be even less

2.3 Maximum Field in the Collector Region, and the Selection of Emitter Periphery

a) The peak ac base current B

To produce the working current J_n amps/cm shown in Fig. 1 (this is the current composed of emitted charge carriers, electrons for an npn transistor, which are collected at the reverse biased collector and do useful work in the output circuit), a base current B amps/cm must be provided at the emitter periphery as shown. This base current has two components; one to supply the charge of holes necessary to neutralize the electron charge associated with J_n, and the other to supply charge to the collector capacitance lying under the emitter stripe.

From the charge-control model of the transistor, the charge of excess electrons in the base layer at maximum current J_n is J_n/ω_t c/cm, where ω_t is the characteristic frequency of the base layer including transit time effects across the collector space charge region (see 2.1d). This represents the peak-to-peak ac value, so that the peak ac value is $J_n/2\omega_t$. The peak current necessary to establish this charge and withdraw it at an angular frequency ω is $\omega J_n/2\omega_t$ amps/cm.

We represent the charge per cm^2 on the collector capacitance at maximum voltage by $Q_c = CV_c$ where C is an average collector capacitance per unit area having a value, for an abrupt junction, of twice the capacitance at maximum voltage. Since for efficient operation, the peak-to-peak ac charge is nearly Q_c, its peak value is $Q_c/2$. Evidently the charge flowing to the collector capacitance under the emitter is $L_e Q_c/2$ per unit of emitter junction periphery (see Fig. 1), and the corresponding current at frequency ω is $\omega Q_c L_e/2$ amps/cm This is true if $L_e < L_\omega$, (see Eq. 9). For $L_e > L_\omega$, the value of L_ω should be used.

The total base current according to the above discussion can be written:

$$B = \omega J_n / 2\omega_t + \omega Q_c L_e / 2 = (\omega / 2\omega_t)(J_n + J_t) \tag{10}$$

where $J_t \triangleq \omega_t Q_c L_e$

The current J_t has been introduced to describe the important combination of parameters which it represents: the transistor cutoff frequency, the collector capacitance charge and the emitter stripe width.

b) The collector charge Q_c

The collector charge depends only upon the fraction of the available voltage swing which is utilized. The relationships between this charge, the voltage V_c, the maximum field E_c, the space charge layer width W_c, and the collector resistivity ρ_c are:

$$Q_c = \epsilon E_c = 2\epsilon V_c / W_c \tag{11}$$

$$\mu_n Q_c = W_c / \rho_c \tag{12}$$

where $\epsilon = 10^{-12}$ farad/cm is the dielectric constant for silicon. From (11) and (12)

$$V_c = \rho_c Q_c^2 \mu_n / 2\epsilon \tag{13}$$

Now at breakdown, the field designated by E_B is very roughly a constant, independent of the breakdown voltage V_B.[8] The breakdown charge Q_B from Eq. (11) is also nearly constant - suitable values for silicon are $E_B \sim 3 \times 10^5$ volts/cm and $Q_B \sim 3 \times 10^{-7}$ c/cm^2. From Eq. (13),

one can express Q_c as a fraction of Q_B:

$$Q_c/Q_B = (V_c/V_B)^{1/2} \tag{14}$$

c) The effective width for carrier current flow

The peak ac base current B of Eq. (10) produces a transverse bias in the base layer under the emitter in a direction to intensify emission at the outer edge of the emitter. Consequently, the current J_n flows in a region of effective width L_f, and by various forms of analysis it can be shown that this width is such that the current B flowing through R_s a distance L_f would produce a voltage drop of $2V_\theta = 2kT/q$:

$$L_f = 2V_\theta/BR_s \tag{15}$$

where $V_\theta = kT/q \sim 1/40$ volt
and R_s = base sheet resistance under emitter

A further limiting effect may occur for very high injection levels. For high injection levels the formula for L_f must be reduced because of high injection effects. Under these conditions, the holes in the p-type base layer may be considered to move with approximately twice their normal diffusion constant in no electric field. (Under these conditions there is no change in the imref for electrons along the base layer and consequently electron drift and diffusion carrents along the base layer exactly cancel.) The high injection condition is reached when the minority carrier charge in the base layer is comparable to the charge on the acceptors (p-type base). The condition in which high injection occurs can be found to be given by the relationship

$$J_n^2 = 2Q_b^2 \mu_p V_\theta \omega_t^2/\omega \tag{16}$$

in which Q_b is the charge of uncompensated acceptors per square centimeter of the base layer, μ_p is the hole mobility, ω_t is the base layer omega and is given approximately (neglecting effects of drift fields in the base layer and other refinements) by

$$\omega_t = 2D_n / W_b^2 \tag{17}$$

in which W_b is the thickness of the base layer. The relation between Q_b and the base resistance R_s is

$$R_s = 1/\mu_p Q_b \tag{18}$$

This expression for maximum current can be conveniently rewritten by making use of the reference value Q_B introduced in connection with Eq. (14). This leads to an expression for the peak collector current at the high injection condition given by

$$J_n = 0.06 (\omega_t/\omega)^{1/2} (Q_b / Q_B) \text{ amp/cm} \tag{19}$$

In a desirable design, this limit will probably not be greatly exceeded. The L_f corresponding to high injection is independent of current, and is given by

$$L_{f\,min} = (2D_p/\omega)^{1/2} \tag{20}$$

in accordance with the discussion preceding Eq. (16).

d) Collector voltage drop at peak current

When the peak current J_n amp/cm flows across the base collector junction, it can produce a large drop in the collector body. This will be

ohmic unless the electric field in the (n^-) collector body reaches a value E_{nh} which produces "hot electron" effects. At this field the drift velocity is v_{nh}, and above it

$$E = E_{nh}(v/v_{nh})^2 \tag{21}$$

in accordance with an approximation[3] in which the drift velocity varies as $E^{1/2}$ above E_{nh}. Typical values for E_{nh} and v_{nh} for electrons in silicon are 5×10^3 v/cm and 5×10^6 cm/sec respectively.

The maximum current J_n will flow at minimum collector voltage preferably such that the drop across the base-collector junction just keeps from vanishing. The maximum current density is evidently J_n/L_f, and in the ohmic region would produce a peak collector field of $J_n\rho_c/L_f$. In the non-ohmic region ($E > E_{nh}$) the current density can be expressed as

$$J_c/L_f = k_h E_{nh}/\rho_c \tag{22}$$

where k_h represents the factor by which the current density exceeds the hot carrier or "Ryder" limit of E_{nh}/ρ_c.

The voltage drop from base-collector junction to W_c (see Fig. 1) may be calculated in a straightforward way by considering the current as spreading from a cylinder of radius L_f/π to a cylinder of radius W_c. The result is called minimum collector voltage:

$$V_{min} = (\rho_c J_n/\pi)F_h \tag{23}$$

$$\text{where } F_h = k_h - 1 + \ln(\pi W_c/k_h L_f) \tag{24}$$

For the ohmic condition of $k_h < 1$, F_h is simply $\ln(\pi W_c/L_f)$.

For $\ln \pi W_c / L_f = 4$, for example, F_h has values 4, 4.3, and 5.6 for $k_h = 1$, 2, and 4. Evidently, F_h does not vary rapidly with k_h and only logarithmically with L_f. Large values of k_h (>4) are not possible for reasons discussed below. As a useful working approximation, one can assume F_h will be less than 6, so that

$$V_{min} \lesssim 2 \rho_c J_n \tag{25}$$

which is the result quoted in Eq. (8).

e) Maximum collector current density

To achieve good high frequency performance at low collector voltage, it is desirable to limit the mobile carrier charge density to a value less than that of the fixed charge density in the collector space charge layer. Otherwise, the net charge density changes sign, and the position of the space-charge region shifts in such a way as to greatly widen the effective base region.[5]

We shall assume that a sufficient field exists at maximum current for the electrons to be moving with their limiting velocity $v_{n\,max}$, which we shall take to be about 2×10^7 cm/sec. In terms of the peak mobile electron density n_c, the current density can be written

$$J_n / L_f = q n_c v_{n\,max} \tag{26}$$

If the acceptor density in the n^- collector region is N_A, we require

$$n_c \leq N_A = 1 / q \mu_n \rho_c \tag{27}$$

where μ_n is the low-field electron mobility, taken to be 10^3 cm^2/v-sec.

Now from Eqs. (26) and (27), the limiting current density is

$$J_n / L_f \leq v_{n\,max} / \mu_n \rho_c \simeq 2 \times 10^4 / \rho_c \qquad (28)$$

Comparison of Eqs. (28) and (22) show that the above limit corresponds to $k_h < 4$ (if $E_{nh} = 5 \times 10^3$ v/cm) which justifies the approximation of Eq. (25).

f) Determination of the emitter periphery

The above restriction on current density, Eq. (28), serves to determine the emitter periphery in the following manner: substituting for L_f from Eq. (15) and for B from Eq. (10) and introducing a convenient reference current J_o amps/cm, Eq. (28) can be written

$$J_n / J_o \left(J_n / J_o + J_t / J_o\right) \leq 1 \qquad (29)$$

where $J_t \triangleq \omega_t Q_c L_e$ (introduced in Eq. 10)

and $J_o \triangleq \left(4 \omega_t V_\theta v_{n\,max} / \omega \mu_n \rho_c R_s\right)^{1/2}$

$$\simeq \left(2 \times 10^3 \omega_t / \omega \rho_c R_s\right)^{1/2}$$

The current J_o is the critical parameter in determining the current per unit periphery, and hence the periphery if the total current is known. The current J_o depends on the ratio of f_t to the operating frequency and on the product of collector resistivity ρ_c and base sheet resistance R_s. Low values of these resistances obviously permit high current.

Figure 3 shows two representations of Eq. (29) taken with the equality sign. The ratios J_n / J_o and J_n / J_t are plotted vs the ratio J_t / J_o. The smaller the latter ratio the closer to the reference current

J_o can the actual working current J_n be made. To realize the maximum power gain of a given structure, it is desirable for reasons discussed below that J_t not be large compared with J_o. This means that J_n is of the same order of magnitude as J_o, and J_o is hence a good estimate of the current which can be carried per cm of emitter-base periphery.

Figure 3 also shows the manner in which the power gain (common-emitter) depends on the current ratio J_t / J_o with respect to the maximum possible power gain at the given frequency. This dependence arises as follows. It is assumed that the dimensions of the transistor are such that $L_c = 2L_e$ in Fig. 1, which means that the collector capacitance under the emitter, represented by C_1 in Fig. 2a, is equal to the remaining collector capacitance C_2 in the figure. For this ratio of capacities, the load conductance for optimum power gain (see subsection 2.5) turns out to be

$$G_{L\,opt} = \omega_t \, (C_1 + C_2) / \sqrt{2} = \sqrt{2} \, \omega_t \, C_1 \tag{30}$$

For any other G_L, the power gain is reduced from the maximum value by a factor of

$$PG_{max} / PG = (x + 1)^2 / 4x \tag{31}$$

where $x = G_L / G_{L\,opt}$

Now the working current J_n passes through G_L, and the capacitive current is supplied by a parallel inductor or its equivalent. Consider a device of emitter-base periphery L_p. Evidently

$$J_n = G_L V_c / L_p \tag{32}$$

Furthermore, $J_t = \omega_t Q_c L_e$ and $Q_c = C_1 V_c / L_e L_p$ so that

$$J_t = \omega_t C_1 V_c / L_p$$

$$= G_{L\ opt} V_c / \sqrt{2}\ L_p \qquad (33)$$

using Eq. (30). From the above two equations

$$J_n / J_t = \sqrt{2}\ G_L / G_{L\ opt} \qquad (34)$$

Thus the ratio J_n / J_t is controlled by the load conductance and is basically related to the ratio of conductive to capacitive current in the transistor. Eq. (34) says that the maximum gain is obtained when $J_n / J_t = \sqrt{2}$, and this corresponds to $J_t / J_o = 0.53$ and $J_n / J_o = 0.76$ according to Eq. (29) and Fig. 3. This point is indicated by a dotted line in Fig. 3. Other points on the curve referring to power gain are computed similarly from Eqs. (31), (34), and (29).

In Fig. 4 are shown curves of J_n vs the collector resistivity ρ_c for various values of the parameter $(\omega/\omega_t)R_s$. These have been calculated from the expression for J_o in Eq. (29) with the assumption that $J_n = 0.76\ J_o$ as determined from the maximum gain condition discussed in the preceding paragraphs. Also shown (in dotted lines) are the values which J_n assumes, denoted by J_{min}, when L_f reaches its minimum value given by Eq. (20). The values are calculated from Eq (28) for different $L_{f\ min}$. At any given ρ_c, the appropriate J_n is the lower of the two values determined for the relevant parameters $(\omega/\omega_t)R_s$ and $L_{f\ min}$. The injection level is estimated by the ratio of the two values.

The minimum collector voltage V_{min} is determined for any point on the chart of Fig. 4 by calculating $V_{min} \leq 2\rho_c J_n$ according to Eq (8)

The power output per unit periphery can be estimated from

$$P \simeq J_n V_c / 8 \tag{35}$$

where V_c is the maximum collector voltage. Conservatively, we may take V_c to be $V_B/2$, and the breakdown voltage V_B depends on ρ_c approximately as [9]

$$V_B \simeq 95\ \rho_c^{0.7} \tag{36}$$

so that

$$P \simeq 6 J_n \rho_c^{0.7} \tag{37}$$

On Fig. 4, Eq. (37) would be represented by lines sloping up to the right with slope 0.2 for J_n corresponding to the solid lines, and sloping down with slope 0.2 for J_n corresponding to the dotted lines.

g) Maximum emitter stripe width

The maximum gain condition derived above in connection with Eq. (33) requires a definite ratio of conduction (J_n) current to capacitive (J_t) current. This in turn requires that $J_t \sim 0.5\ J_o$. Substituting for J_t and J_o from Eq. (29) and supposing that to achieve large power output, $Q_c \sim 2 \times 10^{-7}$ c/cm^2 (see part b above) results in:

$$L_e \simeq 10^8 / \sqrt{\omega \omega_t \rho_c R_s} \tag{38}$$

This equation furnishes an estimate of the fineness of geometry required to realize maximum performance. If this L_e comes out to be greater than L_ω of Eq (9), the analysis is no longer exactly valid and the

value of L_ω should be used in calculating J_t. In that event. the minimum current effect discussed in the next subsection should be taken into account.

h) Parameters required for 500 Mc transistor

Starting with the preliminary design of subsection 2.1, we find that the current per unit periphery arrived at was J_n = 1 6 amp/cm from 2.1 e) and 2.2 a). Using the chart of Fig 4. it is seen that for ρ_c = 1 ohm-cm a value of $(\omega/\omega_t)R_s$ = 500 ohms is necessary for this current. The assumed f_t is 600 Mc; thus R_s = 600 ohms. This low value will be rather difficult to achieve. (A calculation of $L_{f\,min}$ from Eq. (20) gives 1.7μ; this point on Fig. 4 lies well above 1.6 amp/cm.)

For optimum power gain (see Fig. 3), J_n / J_t = 1.4. so that $J_t \simeq$ 1.1 amps/cm = $\omega_t Q_c L_e$. Taking $Q_c = 2 \times 10^{-7}$ c/cm^2 gives L_e = 14μ. The proposed design value of L_e = 10μ is satisfactory. The above value also follows from Eq. (38).

2.4 The Minimum Current Effect, and the Possibility of Current Cutoff

The minimum current effect is discussed in subsection 2 2 c) A further discussion is given below, and additional remarks are made concerning a possible operating mode which may avoid this effect

An important efficiency loss effect may be associated with the collector charge $Q_c = V_c C$. As the collector voltage rises during a cycle from nearly zero to its maximum V_c the charge $Q_c L_e$ coulombs appears per cm of perimeter length into the figure (see Fig. 1). If this charge cannot flow to the base terminal, but provokes an equal charge

of injected carriers in the base layer, a current of electrons $\omega_t Q_c L_e$ amperes per cm flows to the collector in accordance with the charge control model of a transistor.[7] This current was denoted by J_t in the previous subsection, and since it is comparable to the carrier current, it is obvious that it may represent a serious power loss. This power loss can be eliminated in various ways.

One recognized way of eliminating the $Q_c L_e$ effect is by the well-known Wallace tetrode structure. This however, eliminates using at least half of the emitter stripe. Another preferable procedure is to control emitter width and average emitter junction bias so that the charge Q_c does not forward bias the emitter junction.

The conditions which avoid unwanted forward bias due to Q_c can be analyzed as follows. We shall concentrate attention on the case of a sinusoidal base current drive of the form

$$B \cos \omega t \text{ amp/cm} \tag{39}$$

which introduces a charge into the base layer denoted by Q.

$$Q(t) = (B/\omega) \sin \omega t + Q_o \tag{40}$$

where Q_o is a constant of integration determined by the d.c. bias level. We first consider the case of no a.c. voltage on the collector. Then Q may be regarded as made up of two terms, one arising chiefly from injection capacitance where the emitter junction is forward biased (Q_f) and another of space charge where there is reverse bias (Q_r). The maximum value of the former is denoted by Q_f (the subscript "f" for forward) and produces the maximum carrier current to the collector J_n amp/cm

$$J_n = \omega_t Q_f \tag{41}$$

This situation is represented in Fig. 5. Here $Q(x, t)$ is the charge per unit area of the base emitter junction taking thermal equilibrium and no collector voltage as $Q = 0$. Two curves are shown corresponding 90^o and 270^o in (39). The change in charge $(2B/\omega)$ is represented by the sum of the two shaded areas.

The penetration of the voltage wave is also shown on Fig. 5 For the reverse biased region this is an attenuated RC wave with a penetration depth given by Eq. (9), repeated below

$$L = (R_s C_e \omega)^{-1/2} \tag{42}$$

where R_s is the resistance per square of base layer. and C_e is the reverse biased emitter junction capacitance.

For L_e much greater than the penetration depth. the middle of the emitter layer should have a reverse bias such that

$$C_e V_b > Q_c \tag{43}$$

so that when the collector puts a charge Q_c on the base. forward bias of the emitter junction does not occur. This charge Q_c appears on the base at the moment of maximum collector voltage which comes about 180^o after the maximum Q_f and J_n condition It alters the Q and V_b situation to the dotted lines in Fig. 5.

Before analyzing Fig. 5 further we note that a desirable operating mode is class B, where the carrier current flows for half a cycle and is cut off for half a cycle. This implies that the charge Q_f should

be built up in one quarter cycle and removed in the next quarter; thereafter Q_r is built up and removed in the next two quarter cycles. This implies reverse bias on the base terminal for about one half cycle so the average value of the negative V_b and negative charge Q at x = 0 is about $1/\pi$ times its maximum negative value. The result is seen by inspection to give approximately

$$Q_r = \pi Q_c L_\omega \equiv Q_c L_r \tag{44}$$

(the subscript "r" representing reverse) where L_r is an effective depth for which reverse charge must be stored.

If on the other hand L_e is much less than the penetration depth, the charge which must be removed in Q_r is evidently

$$Q_r = Q_c L_e \tag{45}$$

If the Class B condition is desired, it follows as discussed above, that $Q_f = Q_r$. Actually since both the forward and reverse charge distributions are represented by damped progressive waves, the dividing line is not sharp and the same B(t) current may be simultaneously affecting both. However, this probably does not significantly affect the main conclusions. We should also note that the voltage lags the current by about 45° for such waves so that peak input power is about 0.7 times maximum V_b and B values for the wide case.

The desirable Class B operating mode described above requires an average negative bias on the base-emitter junction. Experimentally it is usually observed that the introduction of such a bias improves efficiency, as expected but degrades power gain to the extent that the trade-off is not generally worthwhile

2.5 Power Gain

The preliminary design carried out in subse tion 2.1 was directed toward realizing a minimum gain at 500 Mc of 6 db. An additional important factor which can strongly influence power gain is the package load inductance, principally in that lead which is common to input and output where it is not feasible to tune out the inductance. The effect can be evaluated from the equivalent circuit of Fig. 2b. This circuit omits certain refinements included in Fig. 2a, but permits simple and sufficiently accurate estimates to be made of the influence of lead inductance and resistance.

Consider first the common-emitter connection. The optimum load resistance is about $1/\omega_t C_c$ as mentioned in 2.2 a) A shunt parallel inductance or its equivalent across the output to tune out C_c is assumed. With this load impedance, the power gain has the following form:

$$PG\big|_e = \frac{(\omega_t/\omega)^2}{4\,\omega_t R\, C_c} \tag{46}$$

$$\text{where } R = r_b + r_e/2 + \omega_t L_e/2$$

Here, R is an effective resistance including a contribution from the emitter resistance and the effect of emitter lead inductance. From previous calculations in section 2.1 c) and e), R must be less than 1.7 ohms to obtain 6 db gain. From Eq. (7) and previously established dimensions, $r_b \sim 0.5$ ohms. An additional 0.2 ohms is probably present in contact resistance. This leaves about 1 ohm for $\omega_t L_e/2$ which means that L_e must be less than $2/\omega_t = 0.53$ nhy to obtain the desired power gain. This value becomes a critical parameter in package design.

As a simple comparison, the common-base power gain has been calculated for the circuit of Fig. 2b when the same load impedance assumed for the common-emitter case is used. The result is:

$$PG\big|_b = \frac{(\omega_t/\omega)^2}{\omega_t R_e C_e} \tag{47}$$

$$\text{where } R_e = r_e\left[1 + (\omega_t/\omega)^2\right] + 2(r_b - \omega_t L_b)$$

An important conclusion is that base lead inductance L_b introduces a negative resistance into the input circuit. For example with the values used above in the common emitter calculation, and assuming $\omega L_b = 2$ ohms, R_e comes out to be -2 ohms indicating that the circuit would be unstable. However, by providing extra resistance in the emitter (at least 0.67 ohm is needed), stability could be obtained. The common-base arrangement is then capable of greater power gain than the common emitter.

A more accurate expression for power gain can be derived from the equivalent circuit of Fig. 2a. In this circuit, C_1 represents the collector capacitance under the emitter and C_2 is the balance of the collector capacitance. (C_3 is intended to represent that region which extends under the emitter for a distance of greater than L_ω as discussed in 2.2c. Its effect on gain is the same as if it were included in C_1). Contact resistance to the base layer is represented by r_{b_2}, and r_{b_1} is the resistance in the base layer proper. We shall write the total base resistance as $r_b = r_{b_1} + r_{b_2}$ and the total capacitance as $C_c = C_1 + C_2$. The maximum power gain (common-emitter) is then determined to be

$$PG_{max} = \frac{(\omega_t/\omega)^2}{\omega_t C_c \left[R_E^{1/2} + R_B^{1/2}\right]^2} \tag{48}$$

$$\text{where } R_E = r_e + r_b + \omega_t L_e$$

$$R_B = r_{b_2} + r_{b_1} C_1/(C_1 + C_2)$$

This gain is achieved when the load conductance has its optimum value of

$$G_{L\,opt} = \left(\frac{R_B}{R_E}\right)^{1/2} \omega_t C_c \qquad (49)$$

Large departures of G_L from the above value can be tolerated before the gain drops seriously. Typically, a factor of more than 6 is required for a 3 db drop.

The resistance R_E and R_B in the above expression can be physically interpreted and measured as follows. R_E is the common-emitter input resistance (real part of h_{11e}), and R_B is the reverse common-base transfer resistance (real part of z_{12b}). It is noted that a critical component in R_E is the emitter lead inductance L_e, so that measurements must be made in a physical arrangement like that of the operating circuit.

From the discussion in connection with 2.1e, it is evident that to realize 6 db gain the quantity $[R_E^{1/2} + R_B^{1/2}]^2$ must be less than $4 \times 1.7 = 6.8$ ohms. Allowing 0.2 ohms for r_{b_2} and 0.5 ohms for r_{b_1} (from Eq. 7, using $R_s = 600$ from 2.3h, $R_s = 300$ $L_b = 10\mu$, $L_p = 1$ cm), and assuming $C_1 = C_2$, $R_B = 0.45$ ohms. From the above relation, one finds $R_E = 3.75$ ohms. Since $r_e + r_b = 0.9$ ohms $\omega_t L_e = 2.85$ ohms or $L_e = 0.75$ mhy. This is somewhat more favorable than the estimate made on the basis of the simplified power gain equation.

2.6 Summary

The following list summarizes the principal design parameters derived in the preceding sections. Some of them have been adjusted

slightly to correspond to the geometry finally used.

With reference to Fig. 1:

$W_c = 10\mu$

$= 1$ ohm-cm

$V_{c\,max} = 50$ volts

$W_b = 0.4\mu$

$L_e = 10\mu\, L_b = 10\mu L_c = 25\mu$

L_p = emitter periphery = 1 cm

A_e = emitter area = 1×10^{-3} cm^2

A_c = collector area = 3×10^{-3} cm^2

R_s = 600 ohms per square

R'_s = 200 ohms per square

J_n = 1.6 amps/cm

With reference to Fig. 2b:

f_t = 600 Mc

r_b = 0.7 ohms

r_e < 0.2 ohms (contact resistance)

C_c = 14 pf (50 volts)

L_e < 0.75 nhy

3. SPECIFIC STRUCTURES AND GEOMETRY

3.1 Figures of Merit

According to the summary in subsection 2.6 above. an emitter periphery (L_p) of 1 cm is to be realized with an equivalent emitter stripe half-width (L_e) of 10μ or less. For non-stripe geometries. the quantity L_e can be defined as:

$$L_e = A_e / L_p \tag{50}$$

where A_e is the emitter area. This is one figure of merit for the fineness of geometry; another one of equal importance is an effective "collector stripe" half-width represented by L_c in Fig 1

$$L_c = A_c / L_p \tag{51}$$

where A_c is the collector area. The latter quantity relates to the size of the total collector capacitance in relation to the useful working portion of the transistor.

3.2 Length of Fingers

Consider an interdigitated geometry with diffused emitter fingers 20μ wide and aluminum contact stripes 10μ wide (these may overlap the actual contact region). Because of the resistance of these metal stripes. there is a limit on the length of the fingers which can be used before excessive voltage drop down a finger reduces the effectiveness of the remote end. This limit assumes that current is fed in from one end of the finger.

The maximum length is estimated as follows. Suppose the aluminum is 0.3μ thick so that its sheet resistance is 0.1 ohms/square. The resistance down a finger of length L is $R = 0.1\ L / 10^{-3}$ ohms. The total current flowing down a finger is $I = 2J_n L$, where $J_n = 1.6$ amps/cm. Since this tapers off toward the end of the finger, the voltage drop is of the order of $IR/2 = 0.1\ J_n L^2 / 10^{-3}$. Requiring this to be less than $kT/q \simeq 1/40$ volt means that $L < 125\mu$. Choosing $L = 100\mu$ will require approximately 40 fingers.

The inductance down an emitter finger can be estimated from formulas relating to a bar above a ground plane [10] (the collector N^+ substrate). For the above dimensions, the inductive reactance down a finger at 500 Mc is estimated to be between 0.15 and 0.2 ohms, negligible compared to the resistance of 1 ohm. The skin depth in aluminum at 500 Mc is 3.75μ, so the entire cross-section of the 0.3μ stripes is effective.

3.3 Geometries

An ideal form for the required transistor would consist of a single emitter stripe 1 cm long paralleled by a base stripe. Connections would be made to emitter and base by means of ribbons 1 cm wide, the current being fed in perpendicular to the length of the finger. In this way minimum inductance would be achieved and, furthermore, a rather small thermal resistance of the order of 1 to 2°C / watt would result because of the cylindrical spreading of the heat from an effective line source. However, such a structure would give rise to severe handling and mounting problems. It was decided that the maximum feasible length would be about 2 mm This means that the 1 cm periphery has to be obtained within a region one-fifth as long. If we consider an array of parallel emitter fingers each of length L and center to center spacing s, the ratio of the emitter base junction periphery to the length which the

array occupies is 1 + 2L/s. For a ratio of 5 the quantity 2L/s is required to be 4. The minimum spacing s which could be easily maintained was judged to be 50μ (this allows for a 20μ wide finger plus 30μ intervals between fingers) so that the finger length L comes out to be 100μ. This length is within the limit calculated in 3.2 above The resultant geometry consists of 40 such emitter fingers with one end of each finger connected to a common bus bar region for purposes of making contact. The appearance of the structure is that of two one-sided interdigitated combs as can be seen in the photographs of Fig. 11

Figure 6a illustrates a rough sketch, not to scale. of the basic stripe geometry and a number of modifications. To realize the required emitter perimeter of 1 cm in an area of 3×10^{-3} cm^2 (see section 3 above), 40 emitter stripes 100μ long and 20μ wide are used in a planar npn structure with epitaxial collector regions. These stripes are spaced by 30μ, so that the overall array is about 2 mm long. Aluminum stripes connect the emitter stripes to a common bonding area lying over an oxide layer, and interdigitated stripes contact the base layer. connecting it to a similar bonding area on the other side of the stripe array. Gold ribbons 2 mm wide contact the bonding areas to provide minimum inductance connections.

Variations on the basic stripe geometry which could be incorporated on the same mask are shown in Fig. 6. In Fig. 6b the emitter stripes are broken up into squares, giving approximately the same total periphery but one-half the total emitter area This reduces the minimum current effect discussed in 2 2c. but the registration difficulties are slightly greater. A further advantage is that the J_t current, introduced in connection with Eq 10. is reduced by a factor of 2 since the effective L_e is reduced by the same factor

Another variation, illustrated in Fig. 6c. consists of a lattice emitter (or base squares) Again, the same periphery as the stripe geometry is obtained. One possible advantage of the lattice emitter is that it lends itself well to incorporating the shorted emitter principle discussed in Appendix A. By this means, the effective direct collector-base capacitance is reduced by a factor of order two, improving the common-emitter power gain, and the transistor should be more resistant to burn-out by formation of hot spots. Without the shorted emitter construction. this geometry might have excessive I_{min} and J_t currents. A further disadvantage is a higher base resistance.

3.4 Masks for the Final Design

In designing the mask for this transistor a number of different devices were incorporated. In Fig. 11 is shown the "unit cell" of the masks as evidenced by a group of finished devices metallized but not contacted or mounted. The standard design is the long 40 emitter finger structure having emitter and base aluminum bonding areas 250μ wide extending out on an insulating oxide area over the collector body. In addition to the standard device there is an identical device of one tenth the length with four emitter fingers. This was included in order to get information on the yields obtainable from similar devices of different areas and to observe the effects of different areas on the electrical performance. Evidently the emitter lead inductance problem will be somewhat less serious with the small device. Figure 12 shows an enlarged view of this small device in which there can plainly be seen the planar collector base junction, the diffused emitter region, and the areas which are opened up to provide aluminum contacting. There is a 7μ tolerance which must be maintained in the registration procedures. In the event that this tolerance gives excessive difficulty a device similar to the small structure but with doubled lateral dimensions was

also included on the mask as can be seen in Fig. 11 This could also give information on the effects of different emitter finger widths on the electrical performance, but as it turned out, these devices were not used. Finally there is a structure consisting of a series of annular rings so that by measuring between the various rings the resistance of the base layer both under the emitter and external to the emitter can be determined directly This serves as a valuable check on the diffusion procedures and gives a direct measurement of the important sheet resistance R_s (see subsection 2.3 above).

Figures 7 through 11 are scale drawings of the masks used for the 40 finger (10 watt) transistor. These are photographic masks used to expose photoresist which in turn acts as a mask for the selective etching of oxide windows Finally, the oxide windows define the area to be diffused or contacted.

The base mask of Fig. 7 is first applied to produce a rectangular planar base area. During the base diffusion, an oxide layer builds up or else is deposited after diffusion (see section 4) This oxide is etched out in the pattern of Fig. 8, and the emitter diffusion performed

During the emitter diffusion, further oxide is produced. This must be etched away in stripes running down the emitter fingers so that the subsequent aluminum stripes can contact the silicon At the same time, stripes are etched for the base contact. The contact window etching mask is shown in Fig. 9.

Finally, aluminum is evaporated all over the slice, alloyed, and the unwanted areas removed by the use of photoresist and the mask of Fig 10. This mask allows bonding areas 250μ wide to extend over the oxide covering the collector body While these areas are convenient for

bonding the necessary wide ribbon for low inductance connections, they are also a source of extra capacitance between the collector and the emitter and base terminals. This extra capacitance proved very troublesome to reduce to the point where power gain was not seriously affected.

The minimum line width on these masks is 7μ. The emitter diffusion and contact window masks must register to 5μ. These dimensions were the minimum feasible with the mask-making and registration equipment available at Shockley Laboratory during the duration of the contract. Note that this tolerance must be maintained over the 2 mm length of the structure.

The definition achieved on the aluminum contacts is indicated in Fig. 12, which shows fingers nominally 10μ wide lying over emitter and base contact slots of width 7 and 10μ respectively.

Recent improvements in the equipment used for cutting the original master and in registration fixtures would now permit line widths of 5μ and tolerances of 2.5μ. Thus it would be quite feasible to use 15μ wide fingers spaced by 15μ, giving over 60% more periphery with a consequent gain in performance.

4. PROCESS DEVELOPMENT AND FABRICATION

4.1 Starting Material

The collector material for the proposed transistor is to be 1 ohm-cm according to the design outlined in section 2 An epitaxial collector is desirable to reduce the collector resistance to an absolute minimum (a collector of 1 ohm-cm throughout would introduce a resistance of about 3 ohms). At the beginning of the contract, "inverse epitaxy" was used, consisting of an epitaxially grown n^+ region on original single-crystal 1 ohm-cm n^- material which has been subsequently lapped down to the required thickness (about 10μ). In this way, the device is diffused into high quality material, and the epitaxy merely serves as a low resistance handle.

As the work developed, it was found that straight epitaxy, both commercially available and grown in-house, had improved in quality to the point where it could be considered for fabrication of this device The results obtained were nearly equivalent to those on inverse epitaxy, and since the latter is considerably more expensive, it was discontinued

Epitaxy of 1 ohm-cm (20% tolerance) and a thickness of 10μ (±2μ) was initially obtained from Merck When this became unavailable, further slices were procured from Monsanto, Texas Instruments, and Allegheny No consistent differences in the quality of the junctions obtainable on epitaxy from these various suppliers were noted, but slightly different diffusion schedules were necessary to achieve the same base and emitter thickness The reason for this different diffusion behavior is not understood.

Many diffusion experiments were run on slices of straight 1 ohm-cm material to check layer thicknesses and resistivities prior to diffusing in epitaxy. Unfortunately, the behavior was not always the same. Further remarks on diffusion anomalies are found in subsection 4.3 below.

4.2 Diffusion

a) Preliminary experiments on straight material

Standard diffusion techniques were tried for the production of base layers between 0.25 and 0.5μ wide. Previous experience as well as theoretical considerations indicated that best control of thin base layers is obtained by making very shallow structures. Therefore a base diffusion between 0.5 and 1μ and an emitter diffusion between 0.25 and 0.5μ was chosen. The first experiments were carried out with transistor masks available from other projects until the specially designed masks were completed. Junction depths were measured by angle lapping and staining.

Slices of 1.0 ohm-cm resistivity, n-type were cleaned in solvents, deionized water, and hydrofloric acid. They were predeposited (erfc. distribution) in a closed platinum box containing boric acid anhydride (B_2O_3), and subsequently diffused (Gaussian distribution) in a steam atmosphere, giving a surface concentration of approximately 10^{20} cm^{-3}. An emitter layer was then put on by predepositing the slice with P_2O_5 (phosphorus pentoxide) using the two zone furnace technique to obtain a base layer thickness (W_b) of approximately 0.3 micron.

Mesas were made on the slices and V - I characteristics of the emitter-collector junctions (base floating) were checked. The junctions had breakover voltages of approximately 40 volts at 5 mA current.

After the initial diffusions on non-device slices, it was decided that an all planar type transistor would be made using simple rectangular masks and similar diffusion techniques to those used for the non-planar slices. Base layer widths (W_b) were less than 0 3μ and emitter depths 0.3μ.

V - I characteristics showed "soft" (high leakage current) emitter-base junctions and base-collector junctions along with very soft emitter-collector junctions. Angle lapping and staining of the structure, Fig. 13, showed a very pronounced "emitter dip" effect which looked potentially troublesome. The influence of various parameters on the formation of the emitter dip was investigated. It was shown that this effect can be masked if a guard ring is employed and certain diffusion conditions are maintained. The guard ring used for this purpose is shown in Fig. 14. The emitter edge, where the dip normally occurs extends over a deeper portion of the base layer. This expedient may eliminate possible difficulties caused by extra thin regions at the edge of the emitter, but it does not improve the control of base layer thickness and uniformity in the thin working region. The diffusion was the same originally used for the non-planar slices, except that the guard ring was put in initially to a depth of 4.0 microns with a low surface concentration The resulting devices had very bad V - I characteristics which showed shorted emitter-collector junctions and very soft base-collector junctions Attempts to harden the "guard ring" junctions were not successful.

Meanwhile experiments were continued to determine if the emitter dip could be eliminated. Experiments using approximately the same temperatures for oxidation, base diffusion, and emitter diffusion were tried, and it appeared that the emitter dip became less pronounced Further investigations are reported in subsection 4 3

Because of the lack of any clear advantage, it was decided that it would not be worthwhile incorporating the guard ring on the set of masks that were designed for this device.

Next the prototype masks were delivered and were used with the following diffusion schedule. These masks were discussed in section 3 above, and are illustrated in Fig. 11. An initial oxide was deposited using a steam atmosphere. Base areas were opened by standard Kodak Photo Resist (KPR) techniques, and the base layer was put in using a low temperature (950°C) boron predeposit (platinum box with B_2O_3) followed by a dry oxygen diffusion at a higher temperature. (Dry oxygen was used instead of wet oxygen or steam to prevent depletion of the boron layer). A pyrolytic oxide was then deposited and the emitter areas were exposed by KPR techniques. The emitter layer was put in by predepositing the slice at a high temperature for a short time using a phosphorus pentoxide source. After the emitter diffusion, the slice was protected by steam oxidizing at a low temperature so that the KPR could be used to open the contacting areas. The resulting devices had base layer widths (W_b) of less than 0.3 micron and V-I characteristics showing "punch through" (Fig. 15A).

In the next run, the base layer was diffused (Gaussian distribution) longer to widen the base layer (W_b). The resulting devices gave fairly good V-I characteristics and were therefore metallized, and the small and medium size devices mounted in TO-18 packages for testing. (The small device has the same dimensions as the large device has, except that it is one-tenth the length and has 4 fingers instead of 40. The medium device has twice the lateral dimensions of the small device.) Unfortunately the devices were destroyed in mounting. The structure was angle lapped and junction depths were measured to give W_e = 0.45 micron and W_b = 0.3 micron (Fig. 15B).

The same diffusion schedule was continued with minor changes for adjusting the junction depths, refining the metallizing, and perfecting the mounting procedure. Finally with run number 16, devices got through all the procedures and were tested. The principal results are given in Table II. The letter S following the run number indicates small devices.

b) Further experiments on epitaxial material.

Inverse epitaxy and regular epitaxy slices of 1.0 ohm-cm n-type of approximately 10 microns thickness on heavily doped n-type substrates were used for making devices. The diffusion schedules worked out for non-epitaxy material were used initially. The first epitaxy devices taken through to the testing (in packages) were those of runs number 22 and 23 (see Table II). These generally had low collector breakdown voltage and low f_t.

It was decided that a deeper structure should be tried for possible better control of the layer depths. The first experiments were done on non-epitaxy material and in run number 26c, devices were obtained on epitaxy material that were testable. The diffusion schedule used was as follows (see Table II). A base predeposit and diffusion was done to give a base depth of 1.05 micron and a C_s (surface concentration) of 5×10^{18} cm^{-3}. Then an emitter predeposit and short steam oxidation were done to give a high emitter concentration, W_e of 0.75 microns and W_b of 0.6 micron. The deeper structure was continued along with the shallow one, but it became evident that f_t values obtainable were lower for the deeper structures, and the control problem no better.

c) Uniformity of diffusion

To estimate the degree of uniformity, a diffusion study using the platinum B_2O_3 box technique for the base doping procedure was done

Table I (data taken by four-point probe and by angle lapping and staining) shows the reproducibility of the schedules that were tested. At the time, the control seemed satisfactory, and it was decided to continue using the platinum box method. Further devices were made largely with the two workable diffusion schedules coming out of this study. Unfortunately, only two kinds of material were used, and later experience showed that this degree of uniformity was not generally obtained.

Devices were fabricated using the schedules obtained from the diffusion study from runs 35 through 44. The results are reported in Table II. These results show that the shallower device is generally better for high frequency performance as estimated by the ratio f_t/R_s.

d) Anodic oxide doping.

The idea was to dope anodic oxides on silicon slices with phosphorus or boron compounds. This oxide is then effectively a diffusion source, and good uniformity may be possible. Solutions of different percentages of phosphoric acid and methyl alcohol were used as the anodizing solutions for n-type doping. Figure 16 shows the apparatus for anodizing. The procedure of doping was as follows: the slice was cleaned and then anodized in the solution to obtain an oxide layer of approximately 500°A thick. The slice was diffused at a low temperature to obtain a base layer of less than 1μ. The resulting surface concentrations were approximately 5×10^{20} cm^{-3} on high resistivity material (greater than 1.0 ohm-cm). On low resistivity material the doping was not at all uniform; therefore, it was decided to discontinue using this technique for n-type doping. p-type doping of n-type slices was tried using solutions of sodium borate and boric acid, but the surfaces of the slices became very rough and chewed up at high voltages; therefore, this technique was pursued no further.

e) Boron diffusion with diborane source

By using diborane (B_2H_6) instead of the platinum box (B_2O_3), it was thought to obtain more uniform base layer depths and more uniform and higher base concentrations Initial experiments were done on non-epitaxy slices of 1.0 ohm-cm n-type. They were diffused (erfc. distribution) with 100 P.P.M. diborane (diluted in hydrogen and nitrogen) to obtain junction depths of less than 1.0 micron and surface concentrations of 5×10^{20} cm^{-3}. The slices were then diffused with phosphorus to form emitter layers. The resulting devices showed base layer thicknesses (W_b) of less than 0.5 micron. This approximate schedule was used to make planar transistors and the resulting devices showed very high base concentrations resulting in thin emitter layers, low current gains and soft reverse emitter-base junction characteristics.

The need for a simple and highly reliable method of doping the base layer became more necessary since the platinum box method appeared to be giving increasingly non-uniform layers. It was decided that the concentration profile of the base layer obtained by the platinum box method would be reproduced using the diborane method. Further experiments with different dilutions of diborane gave base layer concentration profiles of approximately the same as those obtained with the platinum box with somewhat more reliable results The platinum box method was therefore discontinued

Devices using the diborane technique of base doping are represented in Table II starting from run number 45A.

f) Phosphorus nitride emitter diffusions

Because of the necessity for a short, high temperature predeposit when using the phosphorus pentoxide procedure, it was thought that the

use of phosphorus nitride (P_3N_5) instead of phosphorus pentoxide might be advantageous in reducing the temperature required or increasing the time. Experiments were run using the same base layer diffusion as used for the phosphorus pentoxide emitter, and it was found that a much shallower initial base depth was needed because the emitter drove the base in too deeply (see subsection 4.3). Experiments using a shallower base showed that the phosphorus nitride doping was not uniform when used at a short time and a high temperature. The work was discontinued in favor of the more reliable phosphorus pentoxide procedure.

g) Decreasing extra capacitance

(i) Thicker oxide.

The capacitance between the aluminum bonding overlays and the collector must be kept small. Means of evaluating this capacitance are given in section 5. If the original oxide thickness covering the collector body and under the aluminum overlays were increased, then the capacitance would be decreased. Experiments were done to increase this oxide from about 1μ to 2μ, but the thickness that could be used was limited by registration and definition difficulties caused by the small size of the contacting fingers. A lesser increase in oxide thickness was not considered good enough, so it was decided that other ways should be tried.

(ii) P-plugs under the aluminum overlays.

This method provides an isolating junction under the bonding area. The procedure for putting the plugs in is to open the original oxide and diffuse P-type dopant into a depth of approximately 5 microns isolated from the device, under the area where the aluminum overlays will go. The remaining procedure is the same as the regular devices

Since this is a floating junction, its capacitance is relatively large (zero bias) and the degree of isolation is small.

h) Final diffusion schedule

(i) Non-plug devices

The slice is oxidized to give an oxide layer 1.0 micron thick. The base areas are then opened up by KPR techniques and the base layer is put in by a diborane predeposit and diffusion in dry O_2 to a depth of approximately 0.8 micron with a surface concentration of 3.5×10^{19} cm^{-3}. The slice is then protected with a pyrolytic oxide and given a KPR treatment to expose the emitter area. The emitter is put in by predepositing from a phosphorus pentoxide source and then steam oxidizing to give protection and a better surface for KPR adherence. The slice is given a further KPR treatment to open the contact areas and is then aluminized by the standard procedure (see below). If the devices are mountable, then heavy aluminum is put on the overlay areas for bonding purposes and the devices are alloyed to the heat sink (stud mount).

(ii) Plug devices

The slice is oxidized to obtain an oxide layer of 1.0 micron, in which the plug areas are exposed by KPR. The plug is put in by predepositing with diborane, cleaning in hydrofloric acid, and steam diffusing to give a thickness of 1.0 micron of oxide and a plug diffused depth of 4 micron. This oxide is then given the KPR treatment to expose the base areas. The rest of the procedure is the same as for the non-plug devices. Further details on the schedules are shown for the runs toward the end of Table II.

i) General remarks on the diffusion procedure and results

All of the experimental devices produced during the period covered by this report were made by standard boron and phosphorus diffusion techniques using oxide masking, with the geometry controlled by photoresist procedures. The layers are diffused for the most part into 1 ohm cm epitaxy from various suppliers. The standard mask is that shown in Fig. 11, corporating in each "unit cell" the prototype 10 watt 500 Mc device, a smaller device of one-tenth the length, another small device where all lateral tolerances are doubled, and an annular ring structure for measuring the sheet resistance of the base layer both under the emitter and external to the emitter. The former sheet resistance is denoted by R_s and the latter by R'_s, as indicated in Fig. 1 in connection with discussion of the design procedure in section 2.

The diffusion of the base layer (boron) for runs subsequent to no. 37 is carried out in dry oxygen instead of steam. Insufficient oxide is built up in this way to mask against the emitter diffusion, so additional oxide is deposited by the decomposition of ethyl silicate. This procedure insures against excessive loss of boron into a steam oxide, resulting in a more heavily doped base layer.

The principal results obtained on the various runs are indicated in Table II. The designation P refers to an initial plug diffusion as described in part g) above.

Quite similar diffusion schedules were used on runs after 44B where the diborane base doping method was introduced (see part 3) above). It is evident that even with this method, a serious lack of uniformity exists from run to run, and there is considerable nonuniformity within some

of the runs. The former condition is particularly evident in the measurements of base sheet resistances, and the latter in the lack of correlation sometimes evident between the base sheet resistance R_s, the measured base width, and the cutoff frequency f_t measured on devices from the same run as the base resistance samples. On one run (28S) cutoff frequencies up to 1100 Mc were obtained which is above the design value of 1000 Mc. There is a large spread on this run, however, which also appears in β with high values of β corresponding to high f_t, indicating a non-uniform base layer. It is believed that the principal source of non-uniformity is in the base diffusion, and that a large uncontrolled influence is exerted by the emitter dip effect discussed in subsection 4.3 following.

The measured values of R_s indicate that it would be extremely difficult to achieve the design value of about 500 ohms/square for the base resistance under the emitter (see section 2.3h) while retaining a high cutoff frequency. Systematic variation of these parameters produced results summarized in Table III. The quantity f_t/R_s is a form of figure of merit; if R_s referred to the actual base resistance of the device, this ratio would be proportional to the power gain expected at any given frequency. It is noted that a much higher value is obtained for 56 than for 39B, despite the larger f_t of the latter. In other words, attempts to thin the base layer beyond a value corresponding to f_t = 500 Mc results in R_s increasing faster than f_t. Run 61 has such a low R_s that the extrinsic resistance R_s' largely determines base resistance, and the figure of merit is inapplicable. The optimum for this range of variation appears to be close to the values of run 56. It is seen that values of the current gain β follow the expected pattern, with high β corresponding to high f_t (thin base layer)

Cross sections of these diffusion runs with thin, intermediate and thick base layers are shown in Fig. 17 a, b, c.

The sheet resistance R_s not only affects power gain but is essential in determining maximum power capability according to the discussion in subsection 2.3. Although the design value of R_s = 500 ohms/square is thought to be quite conservative, an increase of nearly 5 times is certain to have an adverse effect, particularly since even this can only be achieved by reducing f_t from its design value. The excessive base sheet resistance is a principal reason why the desired power output at 500 Mc was not realized.

The collector-base breakdown voltage is in all cases (not more than 50 to 60 volts) considerably less than that expected for 1 ohm-cm material. This effect is probably caused by enhancement of the electric field in the space charge layer at the edges of the thin planar base layer where a large curvature exists. One means of reducing this effect, if desired, at the expense of adding more capacitance, is to employ a guard-ring structure.

4.3 The "Emitter Dip" Effect*

a) Introduction

Localized enhanced diffusion of the base collector junction underneath the emitter regions is generally observed in double diffused high frequency transistors.[14-16] This phenomenon is often referred to as the "emitter dip effect" (EDE). Not much attention has been given to the EDE in the relevant literature during recent years, apparently because of lack of reproducibility complicating the study of the EDE

* This work was performed by R. Gereth, P. van Loon, and V. Williams. Much of it was supported by ARO (Durham). The principal results are summarized here because of their relevance; further details are in the final report on Contract DA 04-200-ORD-1166.

and its theoretical explanation. The EDE was first reported by Miller.[14] He described the phenomenon and suggested that this effect may be caused by impurity interactions such as published by Reiss, et al[17] for lithium in germanium. Later on the EDE was discussed by Baruch and coworkers.[15] They developed an hypothesis which explains the EDE by a vacancy enhanced diffusion mechanism which partially accounted for effects observed in an npn structure made with gallium base diffusion. Baruch's group was prevented from making conclusive statements by large and unpredictable variations from crystal to crystal.

Further experiments tending to confirm Baruch's model are summarized here. After a schedule had been found which gave reproducible results from run to run, intensive studies were made to establish the experimental conditions for the creation of the EDE. Most emphasis was put on a shallow diffused silicon npn transistor similar to that used for the 500 Mc transistor. Later, the experiments were extended to pnp structures.

b) Starting material and standard diffusion schedule.

The starting material used in the majority of the experiments was n^-n^+ epitaxial material obtained from Merck, Monsanto, Alleghény and Texas Instruments. The vapor grown n^- silicon film was about 10 μ thick and had a resistivity of 1 ohm cm, corresponding to a carrier concentration of 5×10^{15} cm^{-3}. The resistivity of the n^+ type substrate ranged from 0.002 to 0.04 ohm-cm in a typical slice. Exactly the same results were obtained (with respect to junction depth) when experiments were carried out using crucible grown silicon material of equivalent doping level. The schedule followed in most of the work is described below.

After standard cleaning steps, the base was diffused in at 1050°C for 15 minutes using 1% diborane in nitrogen as the dopant source and dry nitrogen as the carrier. This procedure gave a junction depth of 0.7 micron and a surface concentration of 1.8×10^{20} cm^{-3}.

Following the base diffusion a surface layer of silicon dioxide was grown by means of the thermal decomposition of ethyl silicate in a vacuum at 750°C for 45 minutes. Emitter windows were then opened in this oxide layer.

The emitter region was produced by diffusion of phosphorus at 1000°C for 7.5 minutes. The source was P_2O_5 at a temperature of 240°C; the ambient was dry nitrogen.

Beveling and staining the slice following the emitter diffusion revealed the "standard" EDE. An enlarged illustration of the EDE is shown in Fig. 18, which defines certain distances. The emitter base junction depth is defined as X_1, the base collector junction under the oxide X_2 and the base collector junction under the emitter X_3. The magnitude of the dip is equal to $X_3 - X_2 = \Delta X$. The distances X_3, X_2 and ΔX are independently measurable and since $X_3 - X_2 = \Delta X$, a check is provided for the value of ΔX.

An advantage of this schedule is that comparatively low temperatures are involved, so that complicating factors such as thermal diffusion, out diffusion through the opened windows, and diffusion into the oxide mask during the time involved are minimized.

c) Influence of emitter diffusion on EDE

It was shown that the presence of a phosphorus doped emitter was necessary to produce the EDE. Under the conditions described

above, a dip was never observed unless staining revealed n type doping in the emitter window. The magnitude of the EDE, however, was independent of the phosphorus concentration. If windows were opened in the oxide over the base layer and the slice diffused for up to 15 minutes at 1000°C in a clean tube, no dip was observed. A further experiment showed that diffusion of boron into the oxide which would result in the EDE, did not play an appreciable role at 1000°C.

Evidence for the observation that the dip is formed in time intervals which are short compared to the overall diffusion times at 1000°C was provided by the following experiment. A 200 μ wide window was opened in the oxide coating of a slice which contained the standard base. An emitter was then predeposited for 10 minutes at 1000°C. Following this treatment, the window was widened, and the slice returned to the predeposit furnace for another 10 minutes. The window was again widened and the slice predeposited for a final 10 minutes. Beveling and staining the material revealed a dip after each diffusion. The initial dip in this particular slice was 0.2 μ and each successive predeposit produced a dip although it became smaller. The base under the oxide remained constant at 0.7 μ throughout the experiment.

d) Variation of base layer

Most of the work discussed above was done on material containing a 0.6 - 0.7 μ deep base junction. The surface concentration of acceptors in this material as determined by four point probe measurements was 1.8×10^{20} cm^{-3}.

Base layers were diffused to depths of 2.5 μ, 3.9 μ and 6 μ. After oxide growth and 7.5 minutes of phosphorus diffusion at 1000°C, an EDE of approximately 0.3 μ was visible in the material containing

the 2.5 μ base. A slight rounding was visible under the emitter of the material containing the 3.9 μ base. No EDE was visible in the 6.0 μ base layer.

Using a series of epitaxially deposited base layers (see subsection 4.4 below) of different doping concentrations the influence of the base doping level on the EDE was further investigated. The substrate was 1 ohm-cm, n type silicon in all cases. Only epitaxial layers having doping greater than 3×10^{18} cm^{-3} show the EDE. This and similar experiments clearly demonstrated that the magnitude of the base concentration close to the base collector junction and not the surface concentration is important for the formation of the EDE. The influence of the base surface concentration is only indirect since it obviously determines the impurity profile in a diffused layer and thus the concentration in the vicinity of the base collector junction.

e) The cooling rate and EDE

Baruch[15] observed a drastic influence of the cooling rate on the EDE. He reported an increase of a factor of 2 in the base region of a silicon slice containing a gallium doped base and phosphorus emitter when the slice was cooled at the rate of 5°/minute compared to much faster cooling. To test this observation, two slices containing the standard base were diffused at 1000° in a phosphorus predeposit furnace. One was cooled normally, and the other at an average rate of 3°/minute. from 1000° to 980°C. It was found that the EDE dip in the material which had not been cooled slowly was the standard 0.3 μ. On the other hand, the EDE in the slowly cooled material was 0.6 μ. The depth of the base under the oxide was exactly the same in both slices. In the normal procedure, the average time required to remove a slice from the phosphorus diffusion furnace was 11 seconds. A modification in procedure resulted

in decreasing this time to about 1 second. Phosphorus diffusion for 4 minutes at 1050°C into silicon containing the standard base followed by this rapid cooling resulted in a 27% decrease in the dip as compared to a slice cooled at the ordinary rate.

Material containing the standard EDE was given 4 heat treatments of 4 minutes each at 1000° in a clean tube without source. The EDE was found to increase after each heat treatment.

f) Determination of concentration profiles

All results on the EDE described thus far were obtained from cross section evaluation of diffused npn structures. In order to gain more information on the EDE, the concentration profile in the standard test structure was determined using the technique of anodic sectioning.[18] Two epitaxial silicon slices (n type, 1 ohm-cm resistivity) were prepared for this experiment. The first slice contained the standard base layer formed at 1050°C for 15 minutes as described in part b) above. The second slice contained the standard base plus a shallow phosphorus layer over the entire slice which was produced by a 4 minute phosphorus predeposition step at 1000°C. This additional 4 minute heat treatment at 1000°C was also applied to the first slice after the base diffusion, however, in this case no phosphorus source was present.

Each slice was anodically sectioned separately. The silicon was removed in increments of 0.05 to 0.06 μ as determined by microbalance weighings. Sheet conductance calculations were made using the results of four point probe measurements. Figure 19 reproduces the experimental results. The sheet conductance is plotted as a function of distance from the surface. The "circles" correspond to the data obtained from the slice which contained only the base layer. In the second slice

the sheet conductance of the base (denoted by "diamonds" in Fig. 19) could be measured only after the emitter was removed. Since the emitter junction depth was 0.4 μ, the first reliable sheet conductance measurement of the base underneath the emitter could be performed at a depth of 0.46 μ.

The emitter dip effect is clearly visible in Fig. 19 as the difference between the extrapolated base collector junction depth of the two slices. The junction depths as revealed by the sectioning procedure were in good agreement with those determined by staining.

The concentration profiles of the diffused boron base layers are plotted in the upper part of Fig. 20. These curves were prepared by determining the differential sheet resistivity from Fig. 19. The concentration of the boron atoms was then found by referring to Irvin's paper.[19] The notation of Fig. 20 is the same as that of Fig. 19. The boron concentration as a function of distance from surface was calculated in steps of 0.5 μ. At the surface, the boron impurities reach a concentration of 2×10^{20} atoms per cm^{-3}.

In addition to the profiles of the diffused boron layers, the profile observed in an epitaxially grown base layer is shown together with a theoretical erfc distribution for a diffused layer with a surface concentration of 2×10^{20} atoms per cm^3 and a junction depth of 0.65 μ.

The results of the concentration profile measurements illustrated in Fig. 20 suggest that (1) there is no pile-up of boron impurities directly in front of the phosphorus emitter (no "snowplow effect") and (2) the EDE is not caused by a normal thermal diffusion process. The fact that "snow plow effect" is not operative is also indicated by the identical curvature which both base layer sheet conductance curves (Fig. 19) exhibit at distances between 0.45 μ and 0.55 μ.

At a depth of 0.6 μ the concentration of the base without emitter drops off rapidly (see Fig. 20) while the profile of the base with emitter shows a more gradual decrease. It was found that the experimental results in this region can be fitted by an erfc function assuming a constant "surface concentration" of 3×10^{19} cm^{-3} and $\sqrt{D_t} = (0.12 \pm 0.01) \times 10^{-4}$ cm, where D is the effective diffusion constant and t the time interval for the diffusion. These data allow the estimation of the D which will be necessary for the formation of the EDE. In part c) above it was reported that the EDE is created during emitter predeposition steps as short as 3 minutes. Therefore, a minimum value for D is given by

$$D_{min} = \frac{1.44 \times 10^{-10}}{1.8 \times 10^{2}} = 8 \times 10^{-13} \text{ cm}^2 \text{ sec}^{-1} \qquad (52)$$

This value is 30 times larger than that published for normal boron diffusion at 1000°C by Fuller and Ditzenberger.[20] The time interval used to derive D_{min} is an upper limit. The slow cooling studies reported in part 2) actually indicate that the EDE is formed while the slice is taken out of the diffusion furnace. This step normally requires from 10 to 11 seconds. The apparent diffusion constant which would govern the dip formation under these circumstances must then be at least 500 times larger than normal. These estimates clearly show that the EDE cannot be explained by a normal thermal diffusion process.

Another interesting feature of the data plotted in Fig. 20 is the fact that both profiles of the diffused base layers coincide at a distance of 0.5 μ from the surface. This observation together with results of spreading resistance measurements[16] underneath the emitter seem to indicate that the EDE is caused by an enhanced diffusion only of the boron impurities present in the vicinity of the base collector junction.

g) g) EDE in pnp structures

To our knowledge, the EDE in pnp structures has never been reported. Attempts were therefore made to produce an EDE in this type of structure with concentrations and junction depths as nearly the same as in the npn structure as possible.

A slice of boron doped silicon ($C_B = 1.5 \times 10^{15}$ cm^{-3}) was diffused for 15 minutes at 800°C with a P_2O_5 source followed by diffusion in a clean tube for 15 minutes at 950°C. This treatment resulted in a junction depth of 0.6 μ and a surface concentration of about 2×10^{20} cm^{-3} which is in rough agreement with the junction depth and surface concentration of the base in the npn structure.

After pyrolytic oxide growth, boron was diffused into the opened windows using a platinum box and B_2O_3 as the source for 5 minutes at 1100°C. Beveling and staining did not reveal a dip.

h) Summary

The main results of the above investigations can be briefly summarized as follows:

(1) The EDE is observed only in npn structures.

(2) The EDE occurs only if the phosphorus emitter doping is sufficient to overcompensate the base doping, and if the surface concentration is much greater than n_i at the diffusion temperature.

(3) Multiple emitter diffusion causes multiple dips.

(4) Slow cooling from the diffusion temperature enhances the dip.

(5) An enhanced diffusion constant for boron near the collector junction apparently causes the dip, rather than rejection of boron by the region of high phosphorus concentration.

(6) The base concentration near the collector junction prior to emitter diffusion affects the magnitude of the EDE. The dip is not appreciable in lightly doped ($< 10^{18}$ cm^{-3}) or wide ($> 2\ \mu$) base layers.

The above evidence tends to confirm Baruch's[15] model, which is based on the fact that heavy donor doping increases the vacancy concentration[21] assuming that vacancies act like acceptors. Thus the heavily doped emitter could act as a source of excess vacancies during the time the slice is being cooled after emitter diffusion. Some of these excess vacancies diffuse toward the base collector junction and greatly enhance the boron diffusion there. A theoretical objection to this mechanism is that few of the negatively charged vacancies could surmount the built-in potential hill between emitter and base, and most of them would diffuse to the surface.

An obvious practical conclusion to be drawn from this work is that to obtain reproducible results in diffusing thin-layer transistors, it is necessary (but not sufficient) to control carefully the rate of cooling, particularly after the emitter diffusion. This conclusion was only reached toward the end of the present contract work; otherwise more consistent diffusion runs would probably have been obtained.

4.4 Epitaxial Base Layers

The idea of using epitaxially grown base layers was discussed. It was thought that such layers would be useful for making thin high concentration base layer devices if uniformity could be maintained. In principle, a layer of given thickness should have a lower sheet resistance than is possible by diffusion.

Initial experimens used an HCl etching technique to clean the surface of the slice prior to growing the base layer. The first experiments were done on non-planar devices, and when reasonable results were obtained, the techniques were used on planar devices with oxide masking.

The attempt to make epitaxy base planar devices was hampered by the non-uniformity of the HCl etching and subsequent growing (Fig. 21A). Some planar devices showing fairly good characteristics were made on non-epitaxy material. An outline of the procedure used is as follows: An initial oxide was put on the slice and areas for base layers were opened using KPR. The base layers were grown to a thickness of 1.1 μ in the center of the device. Figure 21B shows a cross-section through one of the base layers, in which it is evident that the initial HCl etching is deeper around the edges of the base region, producing a thicker layer there. This effect is not necessarily undesirable if the emitter can be restricted to a region where the base is of uniform thickness.

The resistivity of the above base layer was 0.06 ohm-cm. The slice was deposited with a pyrolytic oxide and the emitter areas opened. Emitter layers were put in using the phosphorus pentoxide technique followed by a short steam treatment for KPR adherence. The devices

were contacted with aluminum by the standard method, and tested. The results showed (small 4-finger devices):

R_{s1}	=	1045 ohms/□	R_{s2}	=	5900 ohms/□
W_e	=	0.9 μ	W_b	=	1.4 μ
V_{EBO}	≅	8 volts at 1 mA	f_t	≃	300 Mc
V_{CBO}	≅	50 volts at 1 mA			
V_{ECO}	≅	40 volts at 1 mA			
β	=	2 to 5 at 25 mA			

More planar epitaxy devices were made using a base resistivity of 0.0022 ohm-cm, but this gave devices with very low betas and breakdown voltages. The sheet resistances were

$R_{s1} = 32$ and $R_{s2} = 103$ ohms per square.

Considerably more experimental work would be needed to overcome the severe problems of resistivity, control and non-uniform growth.

4.5 Ion Bombardment Doping

Ion bombardment[13] may be a very useful tool for the diffusion of devices containing very thin layers. The principle of this technique is enhancement of impurity diffusion by an excess density of vacancies generated by ion bombardment. A glow discharge in hydrogen is used for this purpose (see Fig. 22). For doping purposes a small quantity of diborane of phosphine is added to the hydrogen. The sample is heated by the glow discharge. The bombardment of the silicon surface by positive ions creates vacancies which have the effect of enhancing the diffusion

rate of impurities. Since the vacancies have a well determined diffusion length of 0.3 μ, very abrupt shallow junctions can be made this way. Diffusion temperatures are lower than for thermal diffusion.

Simple three layer structures were made at the beginning of the investigation. It became obvious that surface erosion would pose a serious problem. More work will be necessary to solve the problem of surface erosion and of oxide masking. It was found that even short exposure to a glow discharge will destroy a layer of SiO_2.

The use of silicon oxide films for selective masking against impurity vapors during thermal diffusion is well known for semiconductor device fabrication. Silicon oxide films covered with molybdenum were found to be useful for selective masking against ion bombardment doping in a dc glow discharge. The evaporated molybdenum layer shorts the electric field across the silicon oxide and absorbs the sputtering damage.

Silicon slices were prepared for bombardment by growing a thermal oxide layer in steam at 1200°C for two hours. The oxide was removed on the back side of the wafer and a molybdenum layer of about 1.0 micron thickness was evaporated on the freshly oxidized surface. Standard photo-resist (KMER) techniques were used to etch windows in the molybdenum-oxide layer. The molybdenum was removed with a solution containing 1 part of a saturated ferric chloride solution at 25°C and 5 parts of water.

The prepared silicon slices were ion bombarded in the hollow-cathode gas discharge apparatus described by Strack.[13] Temperatures ranged from 400°C to 1000°C, as measured by a total radiation pyrometer to about ± 10°C. Bombardment time varied from 5 minutes to 120 minutes.

The cathode fall was approximately 1 KV and the current density ranged from 10 mA/cm^2 to 50 mA/cm^2, depending on the slice temperature. The gas ambient was hydrogen at 1.7 Torr, with admixtures of 1 vol % of either phosphine PH_3 or diborane B_2H_6.

During the bombardment the slice is simultaneously heated and doped by the impinging low energy ions. The silicon oxide film prevents the diffusion of dopant impurities into the silicon surface. The oxide surface is protected from damage by the molybdenum layer which is partially sputtered away. The molybdenum covers the entire surface, except for windows, providing a good electrical path across the oxide surface to the cathode pedestal. This conduction prevents electrical breakdown of the oxide in the high field of the cathode fall region.

After bombardment the molybdenum was removed and a clean undamaged oxide surface remained. The slices were beveled and stained to delineate the p^+n or n^+p junctions. Impurity surface concentrations and junction depths under the bombarded windows correspond to the values reported by Strack.[13]

4.6 Metallizing

The metallizing procedure initially used was as follows: After the contact areas were opened, the slice had aluminum evaporated over its entire surface to a thickness of 0.3 μ. The slice was exposed to KPR with the contact mask of Fig. 10, and etched in NaOH solution to remove the unwanted aluminum. The aluminum was then alloyed in vacuum at 630°C.

The metallizing procedure was changed because of difficulty in aluminum adherence. The aluminum was evaporated while the

slice was at 300°C and no further alloying was done. This procedure gave good aluminum adherence to the slice.

By means of measurements like those discussed in Appendix B, it was determined that without alloying, an undesirably high resistance contact was produced. Accordingly the final procedure decided upon was as follows: The aluminum is evaporated on the silicon surface which is at 200°C (At 300°C the aluminum tends to be too granular). The KPR procedure would be used to remove the unwanted aluminum, and the slice is alloyed at 630°C. If the initial testing of the devices is favorable, a thick 1 μ layer of aluminum is evaporated, through a metal mask onto the bonding overlay areas.

These techniques were used and good aluminum contacts were generally obtained.

4.7 Mounting and Packaging

It was decided at the outset that no effort would be put into developing a permanent package until a feasible device design had been demonstrated. A temporary mounting procedure was devised in which the device is alloyed to a beryllia disk having two separated metallized areas on the top surface. This beryllia disk in turn is previously alloyed to a stud header for heat dissipation. A gold ribbon 2 mm wide and less than 2 mm long connects the emitter bonding region on the chip directly to the header, so that a reasonably low inductance lead is achieved (grounded emitter operation). A ribbon of the above dimensions is calculated to have an inductance of about 0.6 nhy, which is within the inductance limit calculated in subsection 2.5.

The base bonding area is connected to one of the metallized regions on the beryllia by means of a number of 3 mil gold leads, while the second region, to which the chip is bonded, forms the collector contact. The stud header mounts in a test jig in which connections are made to the collector and base regions by metal spring fingers forming the ends of 50 ohm-strip transmission lines.

Figure 24 is a sketch of the mounting arrangement for the device and of the test jig. The collector and base are brought out to General Radio 50 ohm coaxial connectors, to which coaxial circuitry is readily attached. RF chokes for biasing are provided. A number of small (4 finger - see Fig. 11) devices were mounted in standard TO-18 packages for testing.

The above mounting scheme provides for simple and adequate electrical evaluation of the transistors. Further packaging development was not carried out; however it is evident that a double-ended stud-mounted package with internally grounded emitter could be provided using standard techniques.

5. ELECTRICAL EVALUATION

In this section, the methods of measurement used for the most important electrical parameters are briefly described. Circuits for 500 Mc performance testing are shown. Typical results are quoted to illustrate the results achieved. The run numbers referred to are those whose diffusion parameters are given in Table II. Typical electrical parameters for most of the runs that were carried through to mounting are shown in Table IV. These are representative of the group of from 2 to 10 devices in each run.

5.1 DC Parameters

The breakdown voltages BV_{CBO}, BV_{EBO} and BV_{CBS} were measured on a Tektronix Type 575 curve tracer at a current of 20 mA (2 mA for the small device) which was generally higher than any excess leakage current so that the true breakdown region was reached. A collector-base breakdown above 55 volts was rarely obtained, whereas in 1 ohm-cm material one expects 80 to 90 volts. The lower value is probably due to the curvature at the edges of the shallow base layer acting to concentrate the electric field there. Because the collector junction avalanche tends to take place in regions removed from the emitter junction, the collector-emitter breakdown voltage is very close to the collector-base breakdown. Considerable softness and frequent shorts were usually observed in the collector-emitter characteristic; this can be attributed to the presence of "pipes".[22]

The current gain β (incremental) and saturation voltage V_{SAT} were also measured on the Tektronix curve tracer. Beta generally correlates well with base sheet resistance measurements (see Table II); high beta goes along with the high sheet resistance corresponding to low total base layer doping with consequent high injection efficiency.

Saturation voltage is a useful indicator of excessive contact resistance or grossly non-uniform operation. A value in excess of 0.3 volts was considered high enough to impair performance.

5.2 Capacitance

The collector capacitance was measured at various collector bias voltages on a Boonton 75A-S8 capacitance bridge. Table IV lists values, taken at 10 volts reverse bias, of the total collector capacitance obtained by measuring from collector to base with the emitter floating. This total capacitance consists of three different components, represented in Fig. 2B. Here, C_c is the capacitance of the collector junction proper, while C_{ec} and C_{bc} represent capacitances of the emitter and base bonding areas, respectively, to the collector. The emitter junction capacitance C_e (not shown) is much larger than C_{ec}, consequently when one measures collector capacitance with floating emitter the value obtained is very nearly $C_c + C_{bc} + C_{ec}$.

It is very desirable to determine each of the above component capacitances separately. This may be done by making use of the fact that C_c varies as $V_c^{-\frac{1}{2}}$ (abrupt junction) while C_{ec} and C_{bc} do not depend on collector voltage (determined by making similar oxide capacitors). Thus if the total capacitance C is plotted against $V_c^{-\frac{1}{2}}$ and extrapolated to zero, the value remaining is $C_{ec} + C_{bc}$ as illustrated in Fig. 25 for the "emitter open" curve. Furthermore, the Boonton bridge provides for a three-terminal capacitance measurement, so the direct collector-base capacity $C_c + C_{bc}$ can be determined and plotted in a similar way. Extrapolation then gives C_{bc} as shown in Fig. 25 for the "emitter ground" curve.

From Fig. 25, it is seen that each bonding area has a capacity to the collector of about 18 pF for devices of run 26 C, consistent with the oxide thickness of about 0.75μ under the bonding areas. The capacity C_{ec} is not

necessarily detrimental (except that it produces extra current through the emitter lead inductance) and can in principle be incorporated in the load circuit. The capacity C_{bc}, however, effectively appears as part of C_2 in Fig. 2A, and hence acts to reduce power gain. The value of 18pF is already more than the design limit of 14 pF for collector capacitance listed in subsection 2.6.

Various expedients were attempted to reduce C_{bc} as described in subsection 4.2.g) above, but in no case were values less than 10pF reached. This excess capacity is one of the major reasons why the desired power gain of 6 db was not obtained.

5.3 Cutoff frequency f_t

The cutoff frequency f_t was obtained by measuring the short-circuit common-emitter current gain h_{fe} directly at 500 Mc on a General Radio Transfer Function and Immittance Bridge. A mount for the device similar to that shown in Fig. 24 is used to maintain 50 ohm impedance connections as close to the device as possible.

Principally because of the influence of emitter lead inductance, the external current gain h_{fe} is different from the internal current gain $\beta \simeq \omega_t/j\omega$. An approximate treatment of this difference is obtained by assuming that the input current divides between the input resistance R (which is about $r_b + \omega_t L_e$) and the collector base capacitance C. Only that portion flowing in R is amplified by the internal gain, while the portion in C is shunted directly to the output. The external current gain is then found to be

$$h_{fe} = \frac{(\omega_t/j\omega) - j\omega RC}{1 + j\omega RC} = \text{Re} + j\,\text{Im} \qquad (53)$$

and it is readily determined that

$$\text{Im} = -\frac{(\omega_t/\omega) + \omega RC}{1 + (\omega RC)^2}$$

$$\text{Re}/\text{Im} = \omega RC \qquad (54)$$

The desired frequency ratio f_t/f is calculated from the measured quantities Re and Im according to the following formula derivable from Eq. (54)

$$f_t/f = \omega_t/\omega = -\text{Im} - (\text{Re}/\text{Im})(\text{Re} + 1) \qquad (55)$$

The values of f_t in Table I were calculated from this formula. They were measured at a bias level of 10 volts 50 mA (small device) and 10 volts 500 mA (large device). The apparent f_t rises somewhat with increasing current at this bias level because of the influence of emitter capacitance.

A discussion of the relation of f_t to base sheet resistance is given in subsection 4.2 i).

5.4 Base resistance r_b

The common method of determining base resistance from input impedance (h_{11}, common emitter) measurements is not applicable here because of the relatively large additional resistance $\omega_t L_e$ introduced into the input circuit by the emitter inductance L_e. The procedure adopted to get an estimate of r_b was to measure z_{11} on the General Radio Transfer Function and Immittance Bridge (collector open-circuited for a c) at 500 Mc and a bias of V_c = 10 volts, I_c = 500 mA (z_{11} is essentially bias independent). The real part of z_{11} was attributed entirely to base resistance and the imaginary part is largely lead inductance.

The base resistance determined in this manner was generally not much larger than the design value of 0.7 ohms (subsection 2.6), however

the method probably is not particularly sensitive to contributions to base resistance arising from the high resistance portion of the base under the emitter. For this purpose, the direct measurement of base sheet resistance on the annular ring structure illustrated in Fig. 11 is much more valuable.

5.5 Emitter inductance L_e

A convenient means of determining L_e is by measurement of the reverse transfer impedance z_{12} (common emitter), the imaginary part of which is essentially $j\omega L_e$. This measurement should be performed in a test jig similar to the amplifier circuit arrangement. A typical value for z_{12} at 500 Mc is j 2 ohms, from which L_e is less than 0.6 nhy, within the design requirement of 0.75 nhy.

5.6 Power output and gain

Two types of amplifier circuit were used to measure power output and power gain. One of these, shown in Fig. 26, can be considered a "lumped" circuit. A mounting arrangement like that of Fig. 24 is employed. Simple L pads formed of short lengths of strip transmission line and lumped capacitors are used for impedance matching. Bias is applied through RF chokes, and external capacitors are required to prevent shorting the bias. Normally, the base is operated at zero (dc with respect to emitter) bias, and the transistor is in effect turned on by the positive peaks of the incoming signal.

In an attempt to neutralize part of the collector-base capacitance, a variable capacitor C_n was connected as shown in Fig. 26 through a $\lambda/4$ line so that its capacitance was transformed into an inductance. Not unexpectedly, this expedient usually resulted in the circuit oscillating. In

some cases, oscillator performance measurements were made, indicated in table IV by "osc".

Small (4 finger) devices in TO-18 cans were tested in a smaller version of the circuit of Fig. 26.

The other type of circuit used, which generally gave somewhat better performance and is far more flexible, is illustrated in Fig. 27. The device mount of Fig. 24 is used with externally connected double-stub tuners at input and output. These tuners are conveniently assembled with standard General Radio coaxial components. The usual procedure was to tune the input for a good match to 50 ohms (indicated through the directional coupler by a minimum reading in the reflected voltage E_r) and the output for maximum power. Power input is calculated from E_i, the coupler having previously been calibrated by connecting it directly to the power meter. An alternative procedure, if variable attenuators are available, is as follows. After tuning the amplifier and setting the desired power output, remove the amplifier and connect the coupler output directly to the attenuators. Maintaining the same E_i, adjust the attenuators so that the same reading is obtained on whatever scale was used on the power meter. The gain is then simply the difference in attenuator settings.

The base bias voltage V_b was generally zero. The gain fell off severely for negative V_b, and the transistors were much more sensitive to burn-out if positive (forward bias) V_b was used.

The best performance (run 64) obtained at 500 Mc was a gain of 4 db at a power output of 7 watts. Higher gains were sometimes observed on the small devices, and in one case (run 28) the z parameters were measured, and the "intrinsic" gain (excluding package impedances) was

calculated to be 7 db. Unfortunately, it has not been possible to repeat the results of run 64 due to increasingly large non-uniformities in diffusion whose cause is still unknown. The principal reasons the desired 10 watts at 6 db was not obtained are that the base sheet resistance could not be reduced to its design value of 600 ohms per square (see subsection 2.6) without also seriously reducing f_t, and that excess capacitance of the bonding areas also proved difficult to reduce. These problems are discussed elsewhere in this report.

6. CONCLUSIONS

An extensive design theory now permits estimating the most important transistor geometrical and material parameters required for a given frequency and power performance. Present mask-making and photoresist techniques are probably adequate for the 500 Mc 10 watt device; however, conventional boron and phosphorus diffusion techniques were found to lack the necessary control for the thin layers required. It also appears to be extremely difficult to produce a sufficiently low base sheet resistance by diffusion.

Diffusion calculations are complicated by the emitter dip effect. Considerable progress was made in understanding this effect. It appears to be a vacancy enhanced diffusion of boron in the vicinity of the collector junction, evidently taking place during cooling after the emitter diffusion. It is thus important to control this cooling phase closely to achieve reproducible results.

Alternative means of base doping are by ion bombardment and by epitaxial growth. The former is hampered by severe surface erosion, so that the most promising method currently appears to be the latter. Preliminary work encountered trouble with non-uniform growth, but some good devices (having thick base layers) were produced, and the method is well worth continued investigation.

It is essential in UHF transistors that the capacitance under areas used only for bonding be minimized. The thickness of oxide that can be used in the usual planar process is limited by masking and etching resolution difficulties — not much more than 1.5μ could be obtained on the present transistor design, which was insufficient to reduce capacitance by the

desired amount. Hence a thick oxide isolation technique, possibly by etching mesas and filling these with oxide, is greatly needed.

The shorted-emitter principle discussed in Appendix A was not applied to 500 Mc transistors, but it should have important advantages, particularly at the higher powers, and is recommended for further work.

Lack of diffusion control, the inability to achieve very heavily doped base layers, and excess bonding area capacitance did not allow the desired performance to be reached. With better understanding of the emitter dip effect and development of epitaxial base layers and oxide isolation, there should be nothing inherently difficult about obtaining this performance.

Table I - Studies of Diffusion Uniformity

Experiment No.	Material Used	Original V/I (Ω)	Boron Predep.	V/I after Boron Predep. (Ω)	Boron Diff.	Boron Steam Oxide	V/I after Boron Diff. (Ω)	Emitter Diff.	Steam Oxide	V/I after Emitter Diff (Ω)	X_j of Emitter (μ)	X_j of Base (μ)
3	*	-	900°C 90 min.	14.0 ± 4%	1075°C 50 min.	~1075°C 50 min.	64 ± 6%	1050°C 5 min.	950°C 5 min.	1.43 ± 1%	0.81	1.62
4	*	-	"	16.0 ± 30%	"	1200°C 50 min.	52 ± 15%	-	-	-	-	-
5	*	8.1 ± 7%	"	15.0 ± 3%	"	1000°C 50 min.	66 ± 6%	1050°C 5 min.	950°C 2 min.	1.45 ± 3%	0.75	1.35
6	*	8.6 ± 5%	"	13.5 ± 2%	"	"	53 ± 2%	"	"	1.45 ± 2%	0.75	1.35
7A	*	8.2 ± 10%	"	12.9 ± 1%	1075°C 30 min.	"	53 ± 4%	"	"	1.35 ± 1%	0.75+	1.05
7B	**	-	"	12.6 ± 4%	"	"	51 ± 6%	"	"	1.35 ± 4%	0.75	1.05
8A	*	8.7 ± 14%	"	13.5 ± 4%	"	"	55 ± 5%	"	-	1.45 ± 3%	0.75	1.05
8B	**	-	"	13.3 ± 17%	"	"	52 ± 11%	"	-	1.40 ± 4%	0.75	1.05
9A	*	8.2 ±10%	"	12.9 ± 1%	1075°C 50 min.	"	55 ± 4%	1050°C 7.5 min	950°C 2 min.	1.15 ± 0%	1.05-	1.35
9 B	**	-	"	12.6 ± 4%	"	"	55 ± 4%	"	"	1.05 ± 3%	1.05	1.35
10A	*	8.7 ± 14%	"	13.5 ± 4%	"	"	43 ± 7%	"	-	1.15 ± 0%	1.05	1.35
10B	**	-	"	13.3 ± 17%	"	"	52 ± 33%	"	-	1.10 ± 4%	1.05	1.35

* **Shockley Non-epitaxy**

** Monsanto Epitaxy

Table II - Schedules and Parameters for Principal Diffusion Runs

Run No.	R'_s	R_s	Base Predep. Temp. Min.	Base Diff. Temp. Min.	Emitter Diff. Temp. Min.	f_t Mc	β	W_e μ	W_B μ	C_e pf	Mat.
16	205 232	2497 3145	950°C - 30	1100°C - 10 O_2	1000°C - 5	160 350	4	.5	.8	300	1Ω-cm
22	132 137	1762 1742	" "	1100°C - 15 O_2	" "	350 470	5	.5	1.0	350	I
25	168 132 145	600 612 441	950°C - 30	1100°C - 30 O_2	1000°C - 5	200 250	3 3.5	.6	1.5	250	I
26C	336 363	2718 2347	900°C - 90	1075°C - 50 O_2 1000°C - 50 S	1050°C - 5	400 500	14 22	.75	1.35	160	I
28	118 127	717 902	950°C - 30	1100°C - 22 O_2	1000°C - 5	400 1100	3 20	.5	1.0	350 280	I
29	270	843	" "	" 18 O_2	" "	750	6	.4	.9	260	I
31B	127 136	1262 1531	" " '	" 15 O_2	" " " "	650	5	?	?	250 320	I
32A	127 127	992 947	" "	" 18 O_2	" "	625	4 5	.5	1.0	-	I
32B	148 148	2130 2120	" "	" 15 O_2	" "	600 1100	10 16	.5	.9		I
34A	175 208	1027 1580	" "	" "	" "	400 600	6 8	.5	..9	430	
34B	108 153	365 901	" "	" 12	" "	320 600	2.1 6	.5	.8	500	
35A	328 373	2119 2436	900°C 90	1000°C 50S 1075°C 50 O_2	1050°C 5			0.75	1.35		Mo

I = inverse
Mo = Monsanto

Table II - Continued

Run No.	R'	R_s	Base Predep. Temp. Min.	Base Diff. Temp. Min.	Emitter Diff. Temp. Min.	f_t Mc	β	W_e μ	W_B μ	C_e pf	Mat.
35A	291 382 300	5351 9075 5351	900°C 90	1075°C 30 O_2 1000°C 50 S^2	1050°C 5			0.75	1.05		Mk
36A	565 684	7614 6787	900°C 90	1075°C 30 O_2 1000°C 50 S^2	1050°C 5	1100	60	0.75	1.05	250	Mo
36B	Poor Contacts		900°C 90	1075°C 50 O_2 1000°C 50 S^2	1050°C 7.5	1000	50	1.05	1.35	170	Mk
37A	309 282 318	4900 3672 3170	900°C 90	1075°C 40 1000°C 50	1050°C 5	900	25	0.75	1.05	190	Mk
37A	255 282 264	4634 3762 5363	900°C 90	1075°C 40 1000°C 50	1050°C 5	980	35	0.75	1.05+	240	Mk
38B	199	1010	900°C 90	1075°C 60 O_2 *	1050°C 7.5	500					
39A	315 1640 191	3078 5625 2070	900°C 90	1075°C 40 O_2	1050°C 5	920	18				
39B	437 1640 328	12200 17000 12000	900°C 90	1075°C 60 O_2	1050°C 7.5	1200	50	1.05	1.35+	290	A
45A	211 291 182	4170 5280 1830	Diborane 950°C 20	1075°C 40 O_2	1050°C 5			0.6+	0.9+		Mk
48	394 ±29%	5680 ±29%	Diborane 950°C 30	1075°C 30	1050°C 5	600	15	0.75	1.05-	270	TI

A = Allegheny
Mk = Merck
Mo = Monsanto
TI = Texas Instrument

* Pyrolitic O_2 introduced in subsequent runs

Table II - Continued

Run No.	R'_s	R_s	Base Predep. Temp. Min.	Base Diff. Temp. Min.	Emitter Diff. Temp. Min.	f_t Mc	β	W_e μ	W_B μ	C_e pf	Mat.
57	209 219 200	3080 3270 2900	Diborane 950°C 30	1075°C 35 O_2 "	1050°C 5 950°C 5	480	12	-	-	300	TI
56	164 164 155	2400 2810 1450	"	"		500	14	0.75	1.10	300	TI
60	187 291 155	1340 1630 1270	"	"	1050°C 5 950°C 5	180	6	0.75	1.15	230	
61	62 63 60	355 398 334	"	"	"	200	5	.75-	1.2-		TI
64	71 82 46	601 695 297	"	"	1050°C 5 950°C 10	?	5	0.70	1.05	280	TI
66	179 228 173	3080 3360 2810	S. O. at 1200°C 60 Diborane at 950° 30 min.	1075°C 30 O_2	"	?	25	0.75	1.1	260	TI
69	189 200 173	4450 4640 3720	"	1075°C 35 O_2	"	?	60	0.75	1.1	260	TI
71AP	196 220 192	5590 5890 4260	Diborane 950°C 30	1075°C 35 Pyrolytic O_2 30	"	-	15	0.75	1.1	-	TI
77AP	191 200 182	5300 5640 4820	"	1075°C 40 Pyrolytic O_2 30	"	-	50	0.7	1.05	200	A
78B	164 164 164	2060 3270 1450	"	"	" Steam	380	15	0.7	1.05	250	A

TI = Texas Instrument
A = Allegheny

Table II (continued)

NOTES ON TABLE II

a) R_s' and R_s are the sheet resistances, in ohms per square, of the base layer external to the emitter and under the emitter, respectively. They are measured on the annular ring structure evident in Figure 11. The values given represent high and low, or high average and low readings on a single slice.

b) Under "base diff", O_2 and S represent dry oxygen and steam ambients, respectively.

c) Some entries under f_t and β show high and low values from devices on a single slice. Both f_t and β are measured at a current of 500 ma.

d) W_e is the depth of the emitter junction and W_B is the depth of the base-collector junction, both measured from the surface.

e) C_e is the zero-bias emitter capacitance for the large (40 finger) device.

f) All material (except for run 16) is nominally 1 ohm-cm epitaxy.

Table III - Extremes of Base Sheet Resistance and Cut-off Frequency

Run No.	β	R'_s	R_s	f_t	f_t/R_s
		Ohms/square	Ohms/square	Mc	
39B	50	440	12,200	1200	0.098
56	14	195	2,400	500	0.21
61	5	62	350	200	0.57

Table IV - Principal Electrical Parameters

Run No.	BV CBO	BV EF	BV CES	β	f_t Mc	V_{SAT}	C pF	Other Data
16*	35 v	4.2 v	35 v	4	350	—	9	
22*	25	4.8	25	5	420	0.3	7.5	
25	50	4.9	50	3	250	0.4	30	
26C	58	5.0	58	18	450	3.5	32	r_b = 1Ω, 5 watts, 2 db gain
28	45	4.7	45	3 20	400 1100	0.2	30	4.5 db gain, 0.3 w, r_b = 12Ω, intrinsic gain 7 db (small device); r_b = 0.6Ω, gain = 2.6 db, 5 w, 350 Mc 12 watts osc (large)
29*	40	4.5	40	6	750	0.25	6	
31B*	40	4.6	20	5	650	0.2	—	
32A	35	4.5	28	4	625	0.7	—	
34A	39	4.7	33	5	500	0.25	—	
36A*	43	5.5	34	60	1100	—	8.5	0.15 w osc
37A*	45	5.5	45	25	900	—	9	0.12 w osc
38B*	40	5	40	5	500	0.2	8.2	0.45 w osc

Table IV - Continued

Run No.	BV CBO	BV EBO	BV CES	β	f_t Mc	V_{SAT}	C pF	Other Data
39A	18 v	4.5 v	18 v	18	920	0.8	47	
39B*	36	5	16	50	1200	1.2	7	
44	50	5	50	5	500	0.9	42	r_b = 1 Ω 5 w 4 db
48*	19	5.5	—	15	600	0.2	7	
56	40	5	40	14	500	0.3	50	
57*	48	5.2	48	12	480	0.3	5.8	
60*	31	5	31	6	180	1.4	6	
64	45	5.4	45	5	—	0.15	39	7 w 4 db
66	60	5.6	58	25	—	0.3	33	5 w 3 db
69	38	5.8	38	60	—	0.25	40	4 w 3 db
71A	20	0.2	18	15	—	0.25	—	4 w 3 db
75A	50	3.6	50	40	—	0.4	35	
77A	50	5.9	50	50	—	0.35	35	3 w 2 db
78B	49	5.8	49	15	380	0.35	35	4 w 2 db

Table IV (continued)

NOTES ON TABLE IV

a) * indicates measurements made on small (4 finger) devices.

b) Breakdown voltages measured at 2 and 20 mA for small and large devices, respectively.

c) β measured at V_C = 5 volts, I_c = 100 mA (small), 1 A (large).

d) f_t measured at V_C = 10 volts, I_c = 50 mA (small), 500 mA (large). Missing values are caused by difficulties in oscillation in the bridge.

e) V_{SAT} measured at I_c = 100 mA (small), 1 A (large), $I_b = I_c/10$ or $I_c/5$ depending on β.

f) C is total collector capacity (collector to base, plus collector to emitter) at V_C = 10 volts.

g) Under "Other Data", measurements are at 500 Mc unless otherwise stated.

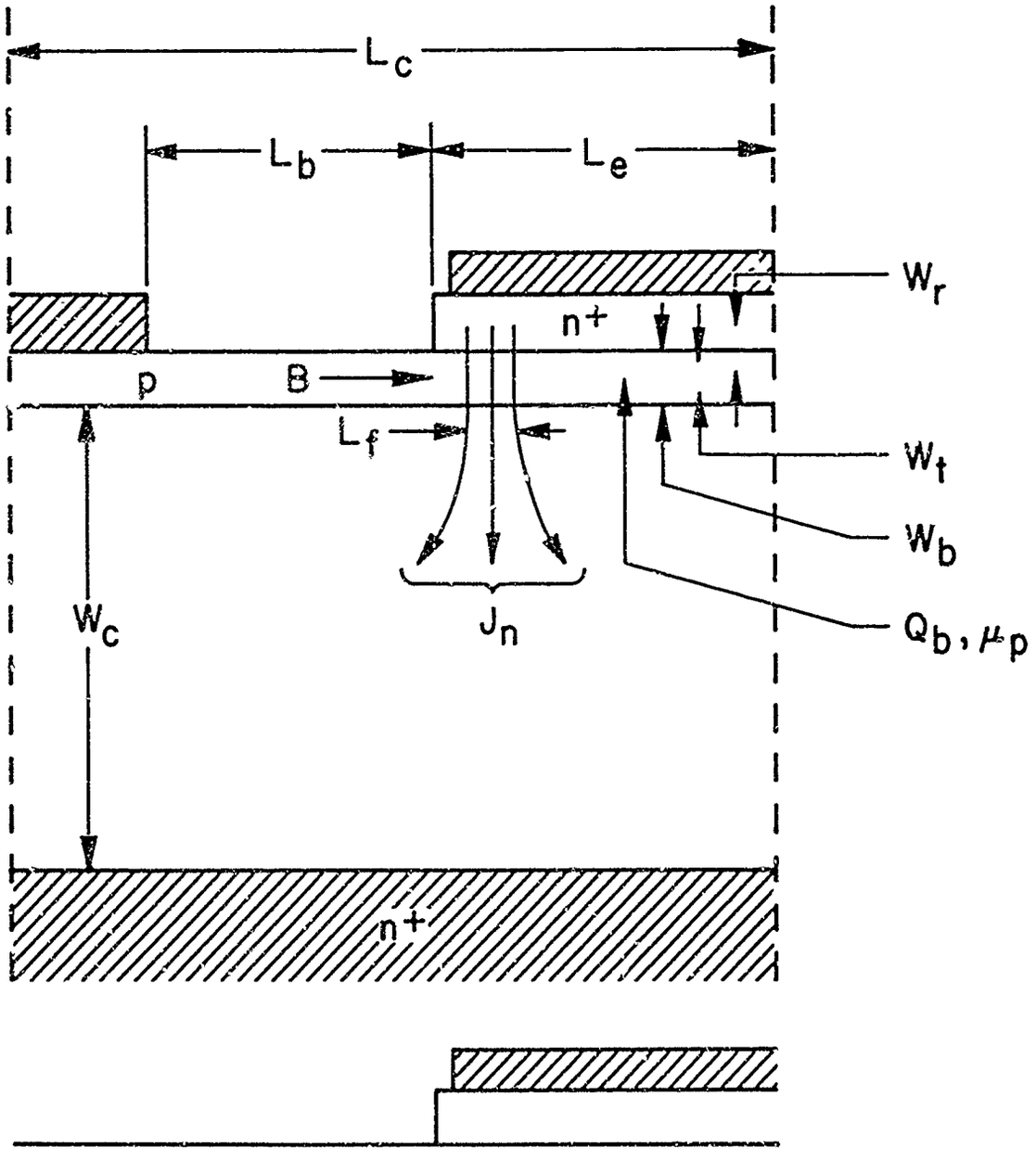


Fig. 1 Cross-section of transistor structure showing principal parameters and current flow.

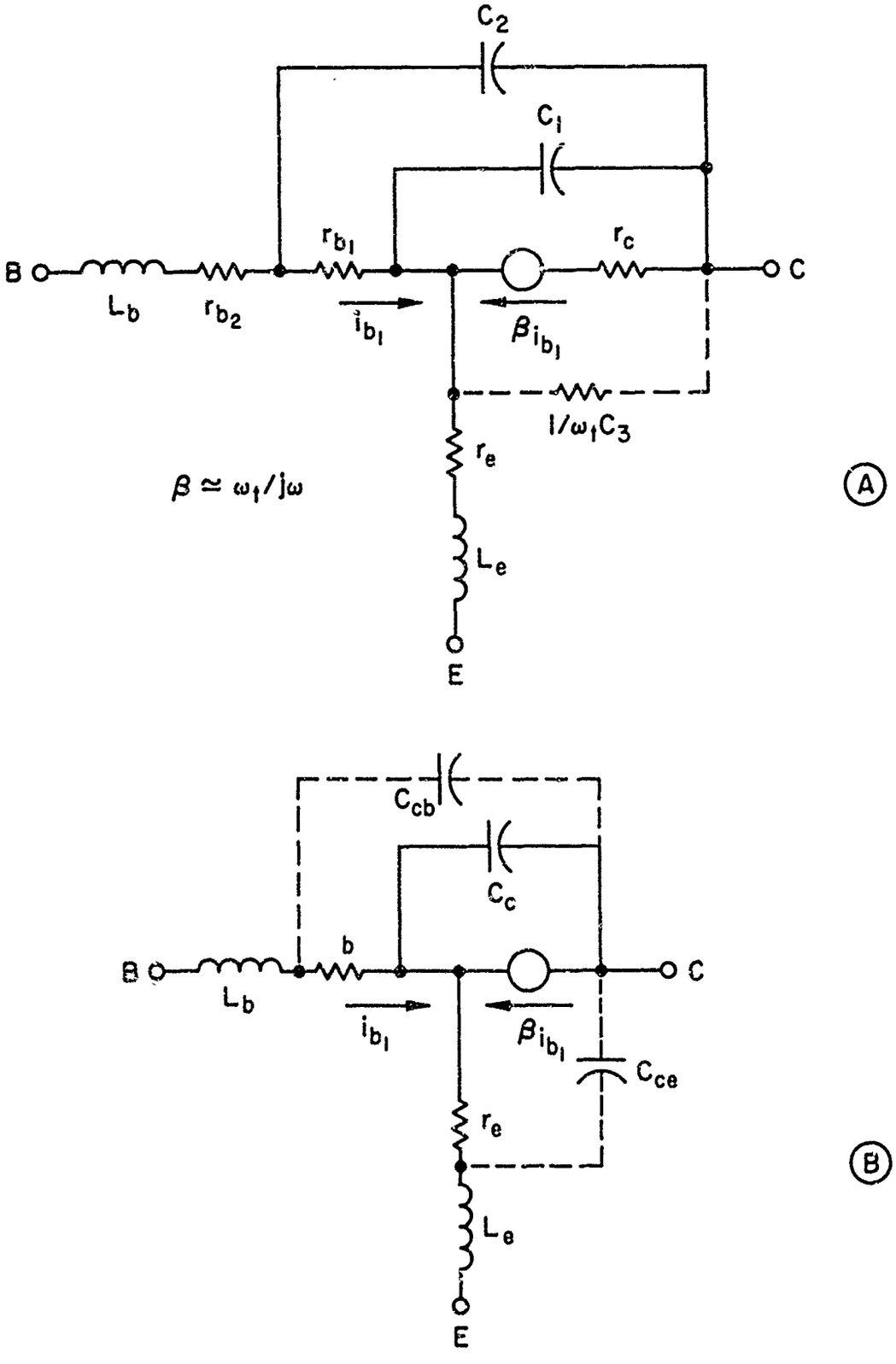


Fig. 2 "Exact" and simplified high-frequency equivalent circuits.

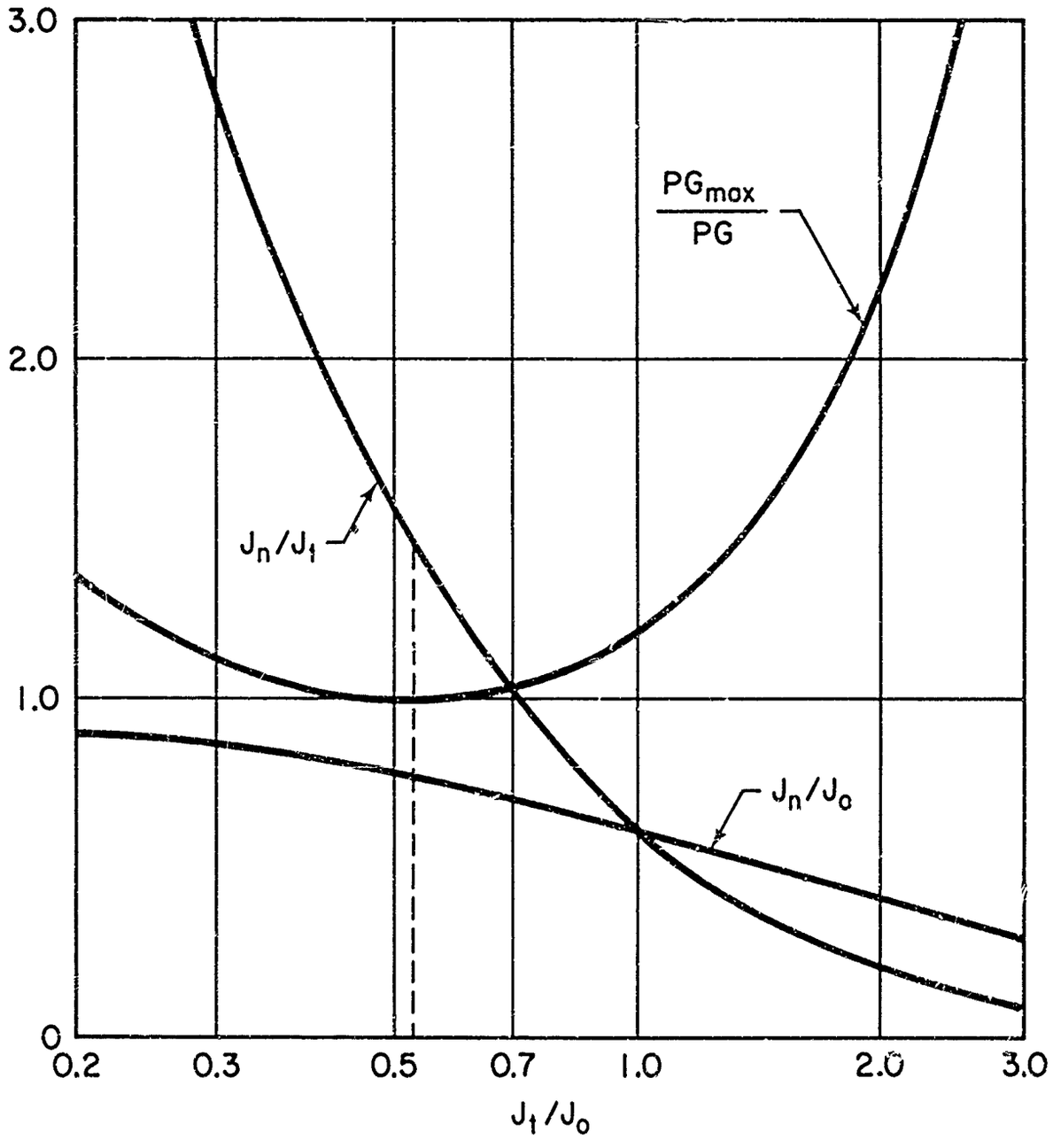


Fig. 3 Chart showing relations between conduction and displacement currents.

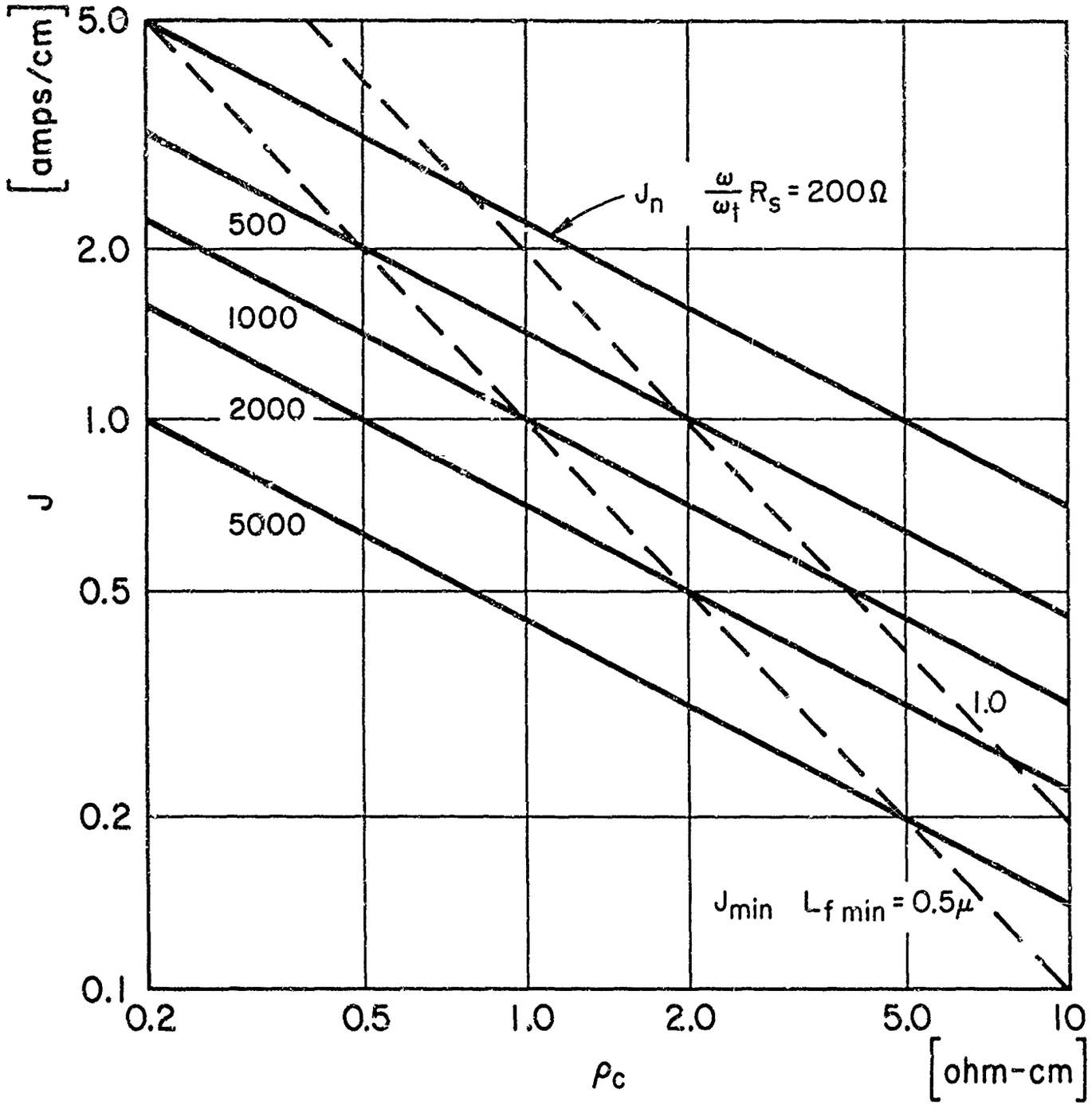


Fig. 4 Current per unit periphery as a function of collector resistivity and base sheet resistance.

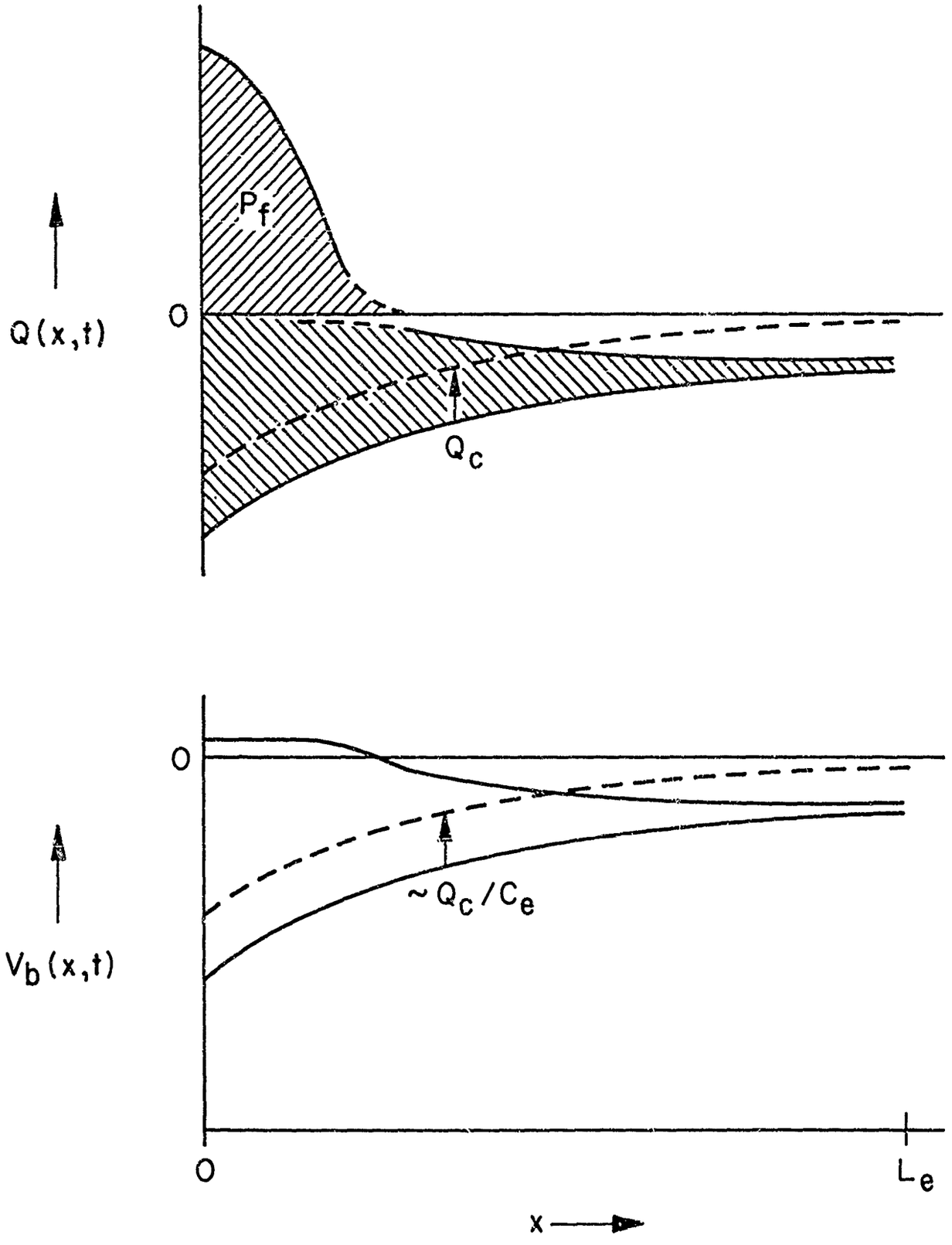


Fig. 5 Charge distributions and base-emitter voltage for maximum and minimum collector current conditions.

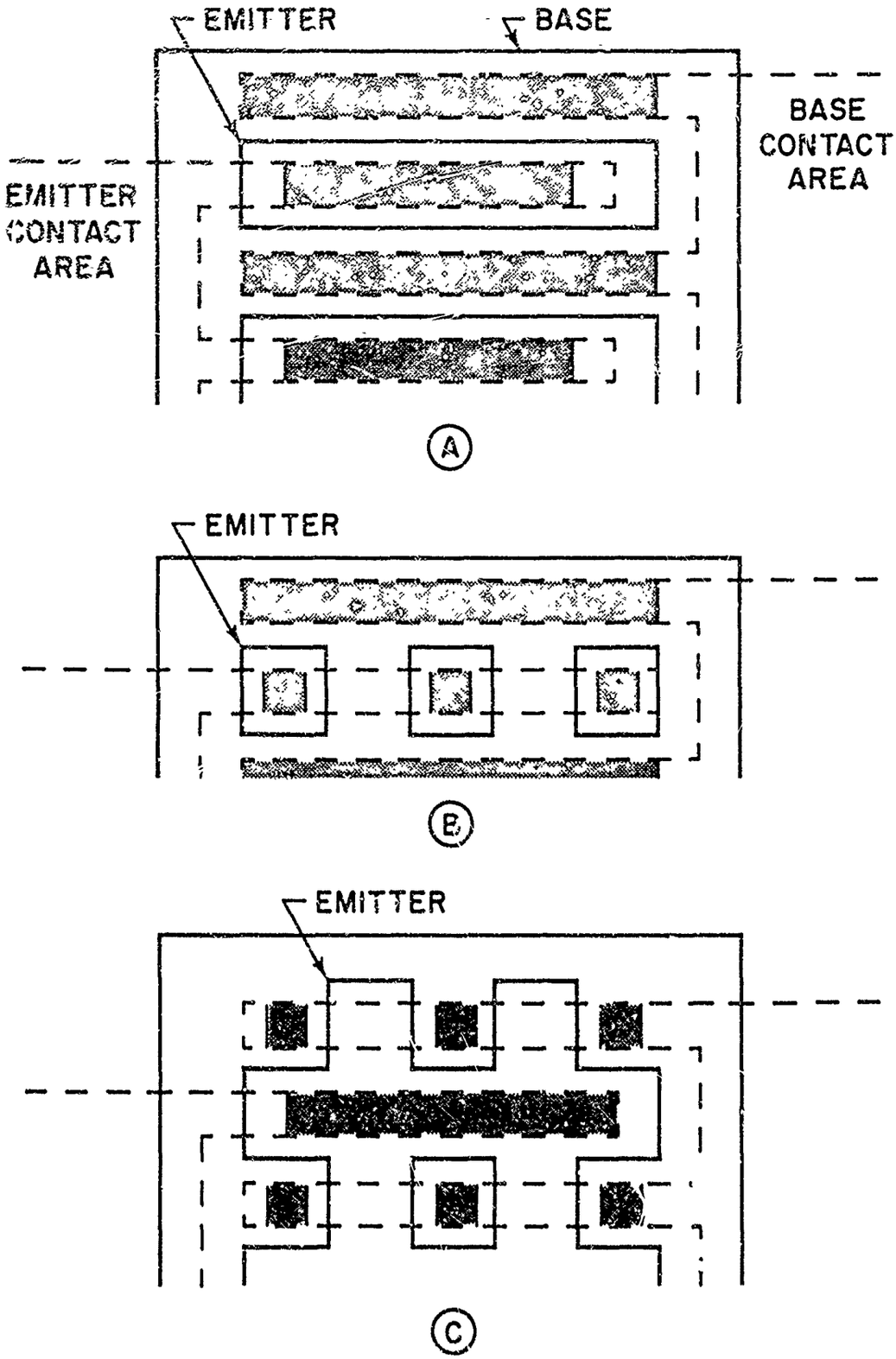


Fig. 6 Basic stripe geometry and variations.

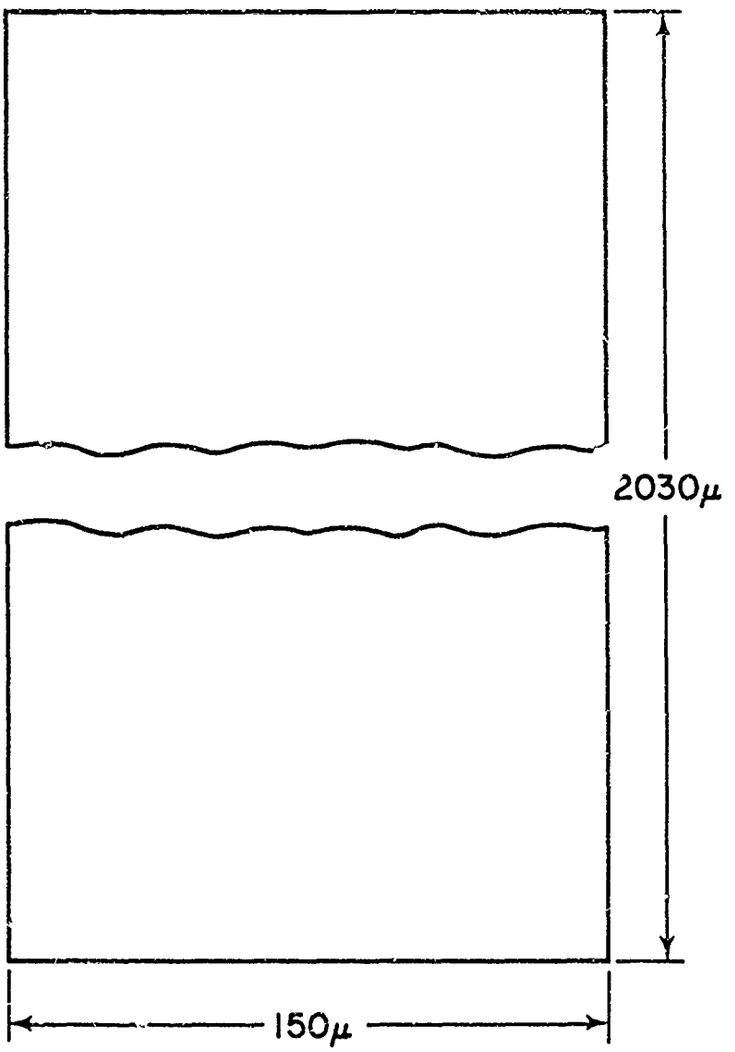


Fig. 7 Base diffusion mask.

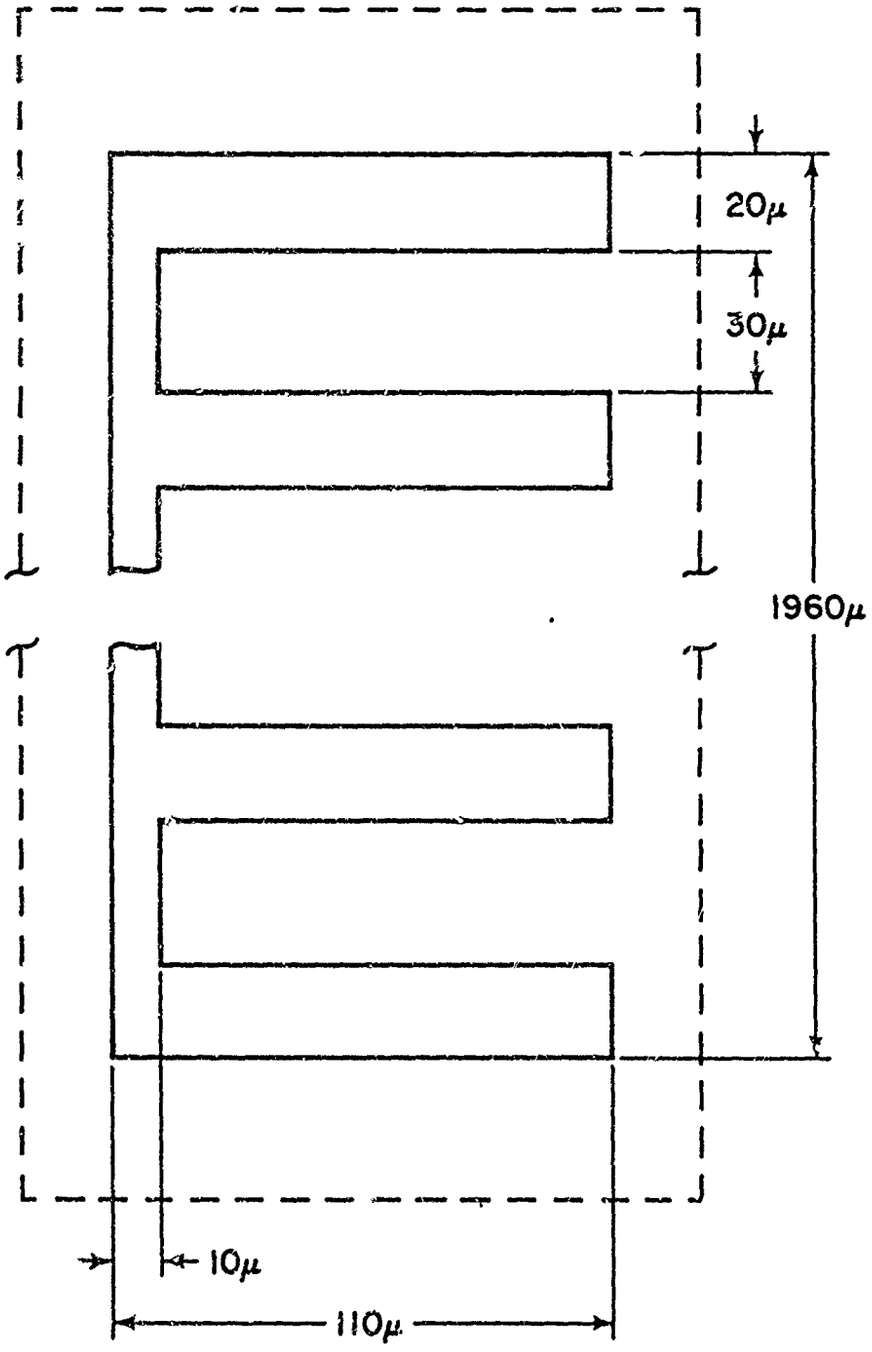


Fig. 8 Emitter diffusion mask.

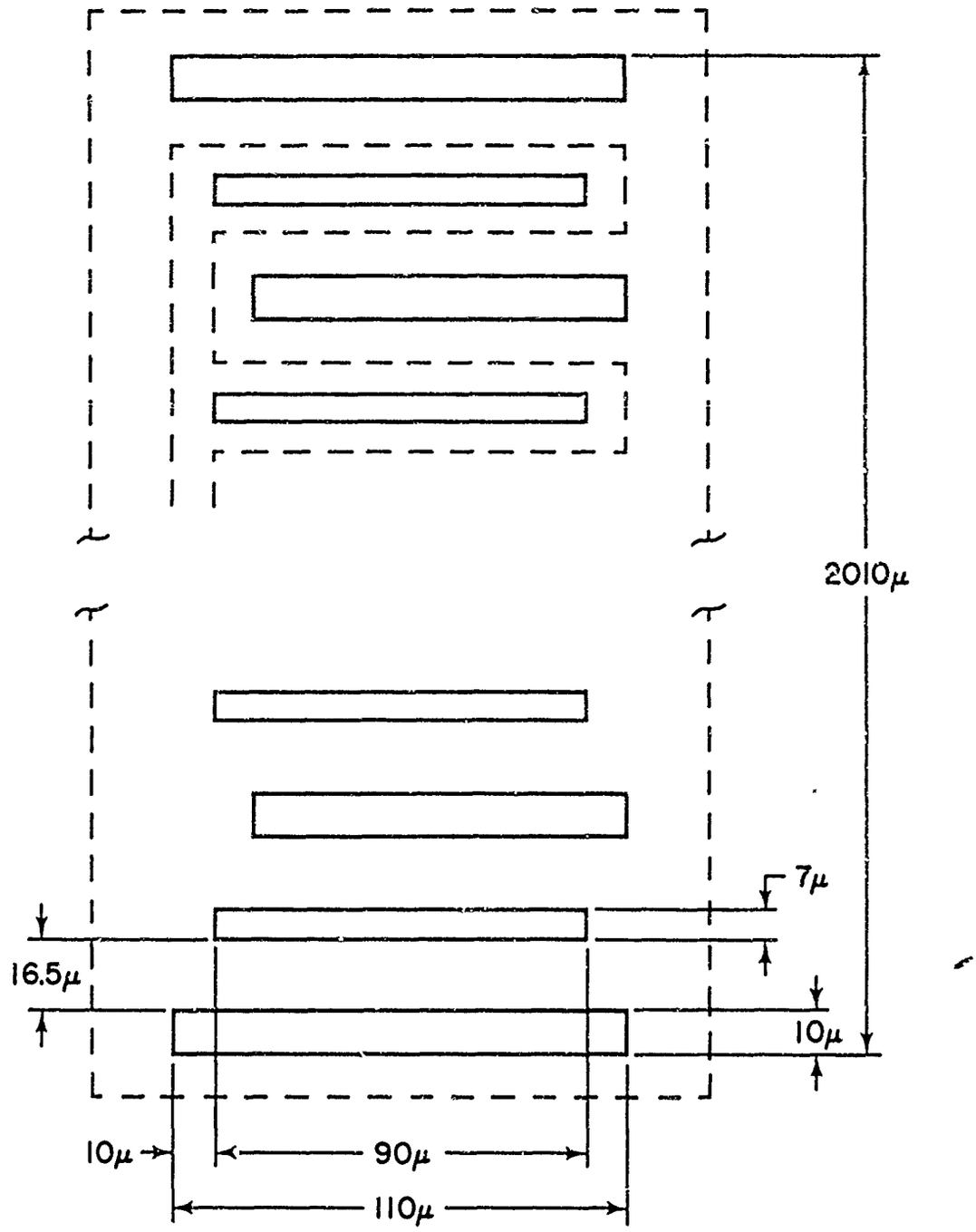


Fig. 9 Mask for opening contact areas.

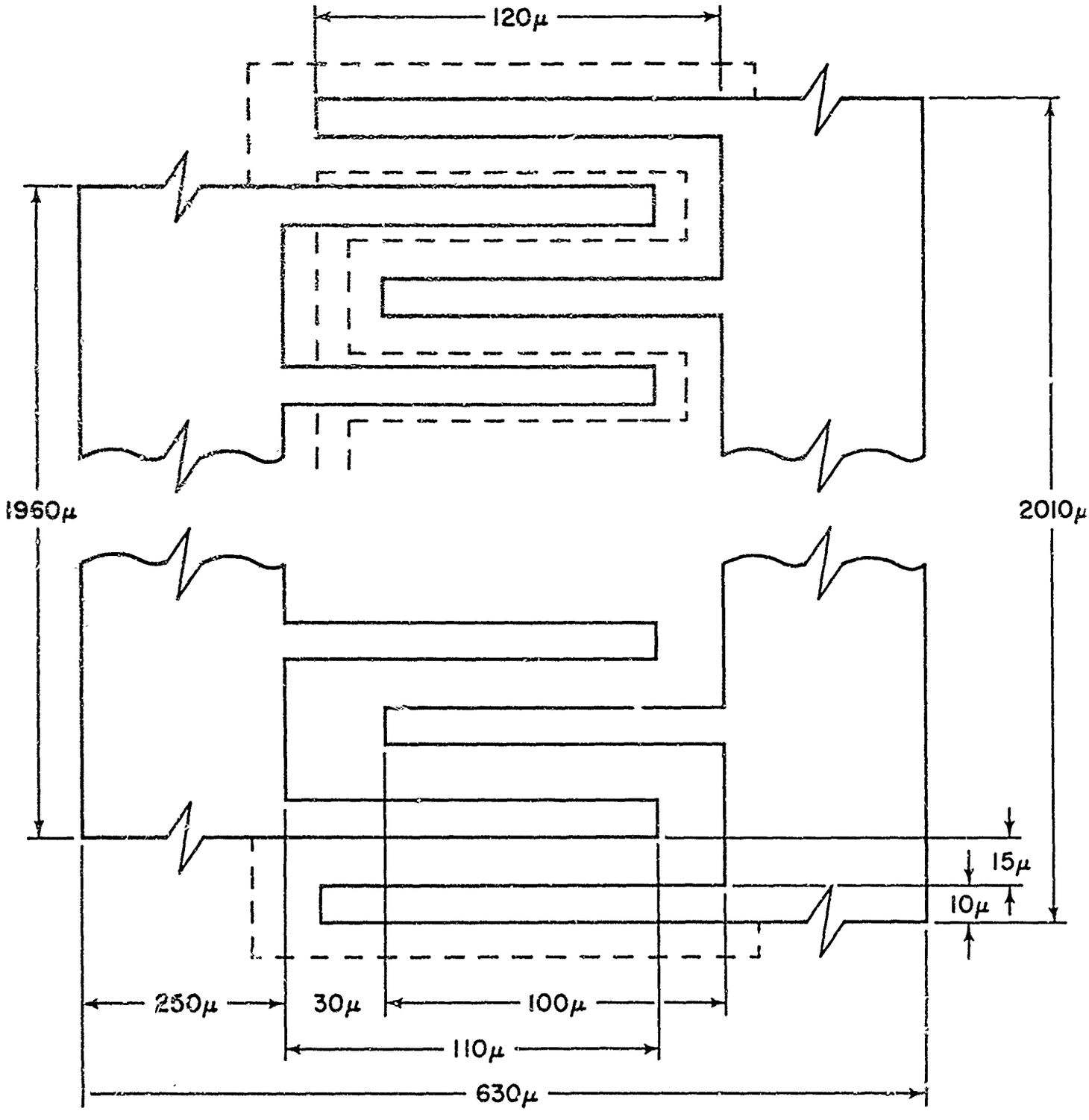


Fig. 10 Mask for aluminum contacts.

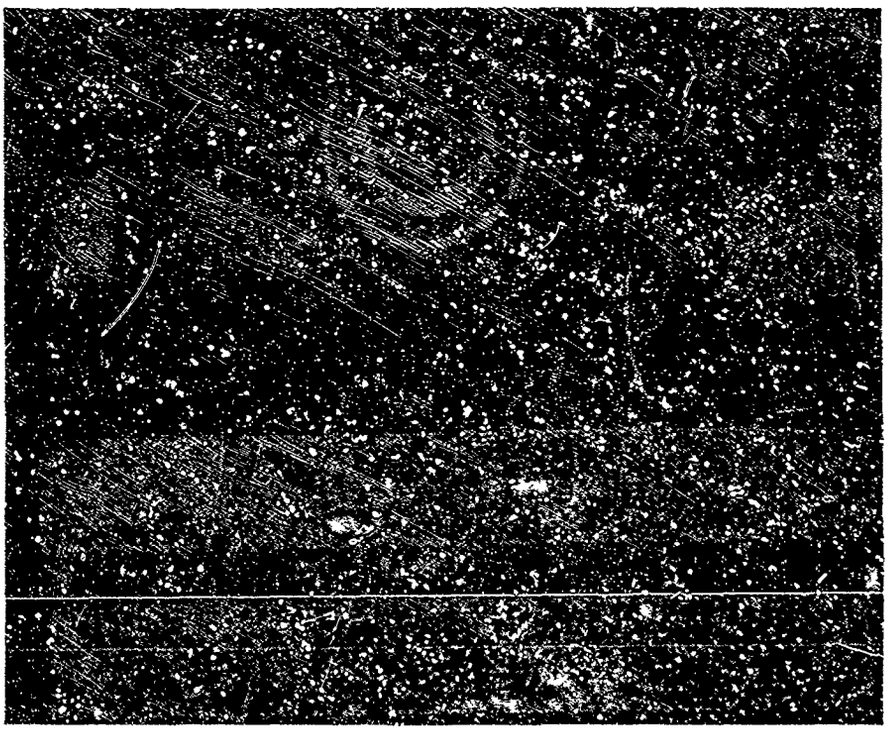


Fig. 11 "Unit cell" on final mask showing metallized devices.

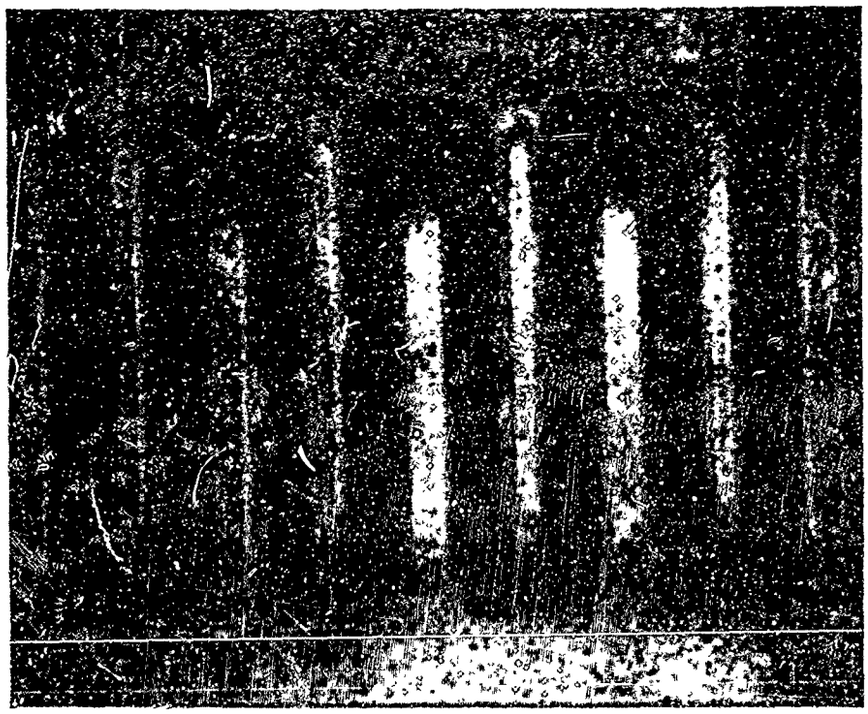


Fig. 12 Enlarged view of metal on small device.

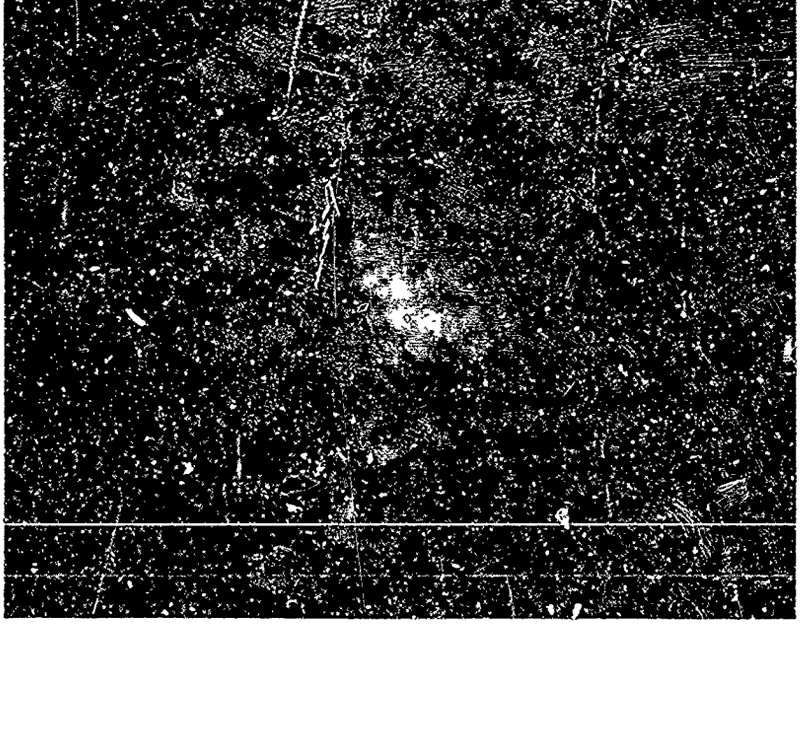


Fig. 13 Cross-section through diffused device with 0.2 μ base width showing emitter dip.

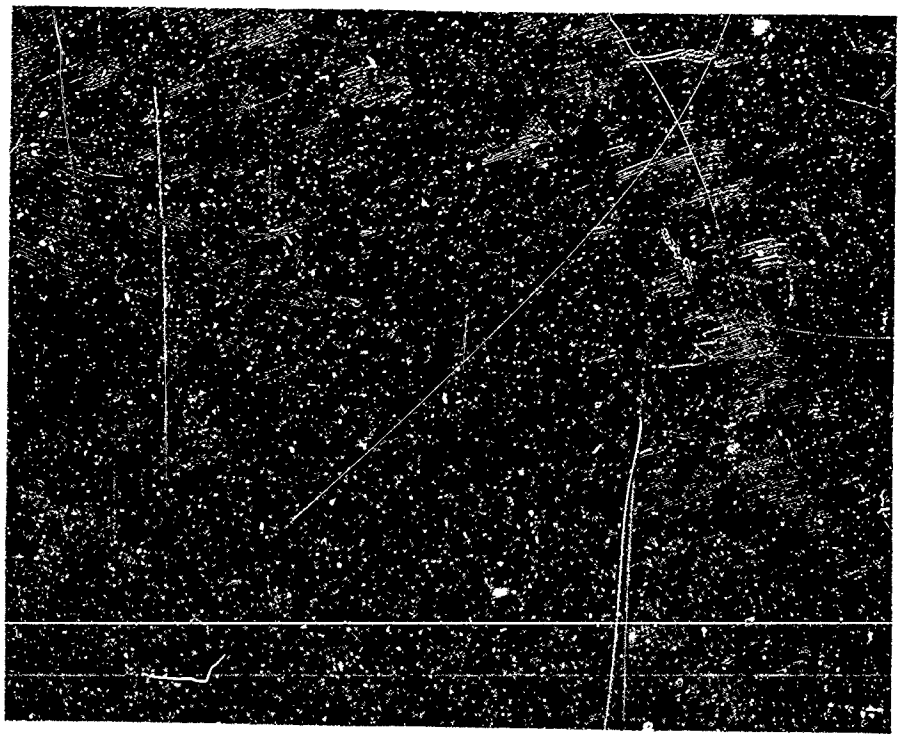


Fig. 14 Cross-section through diffused guard ring.

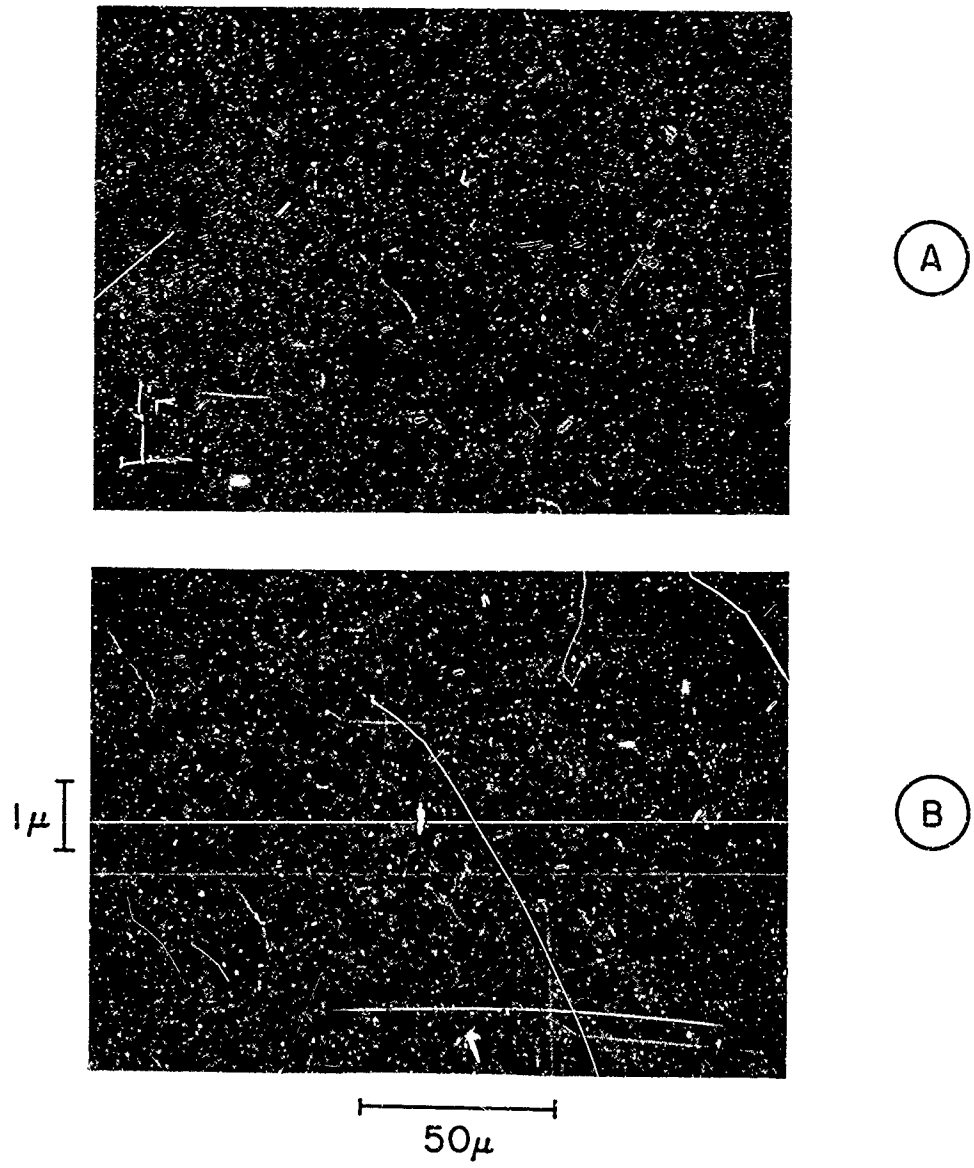


Fig. 15 Photographs of bevelled and stained sections.

A. Very thin base exhibiting punch through.

B. 0.3μ base with severe emitter dip.

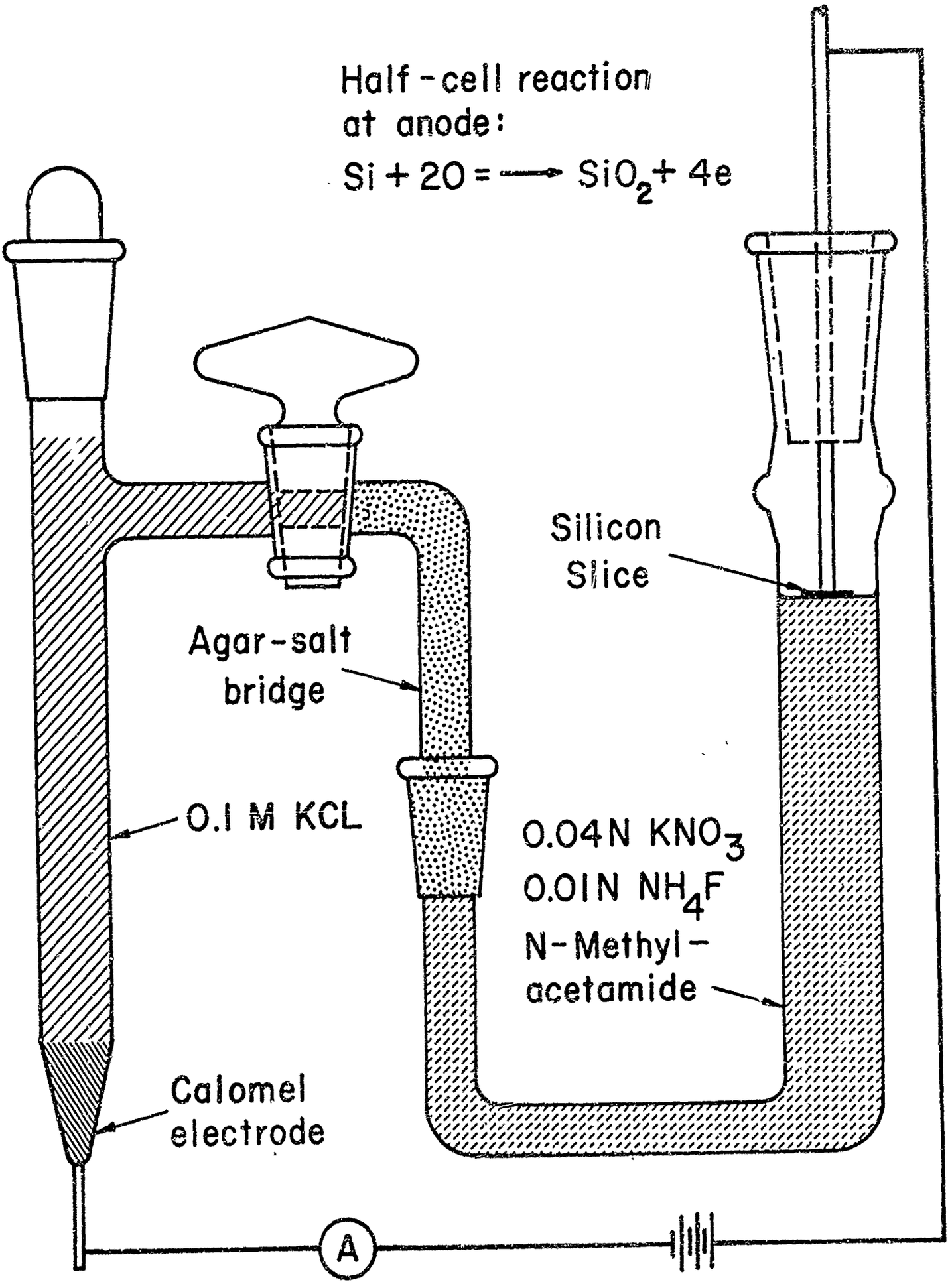


Fig. 16 Anodic oxidation apparatus.

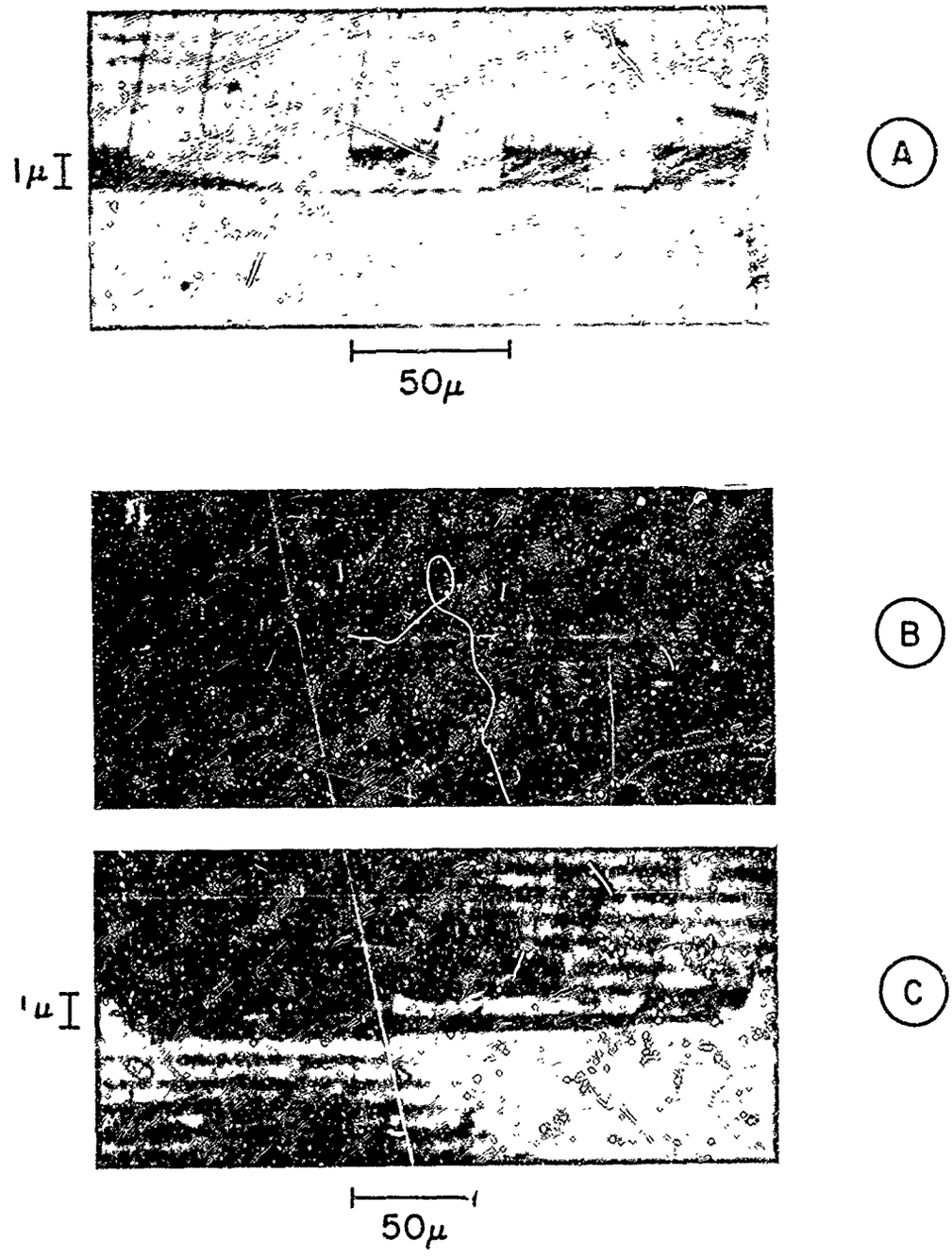
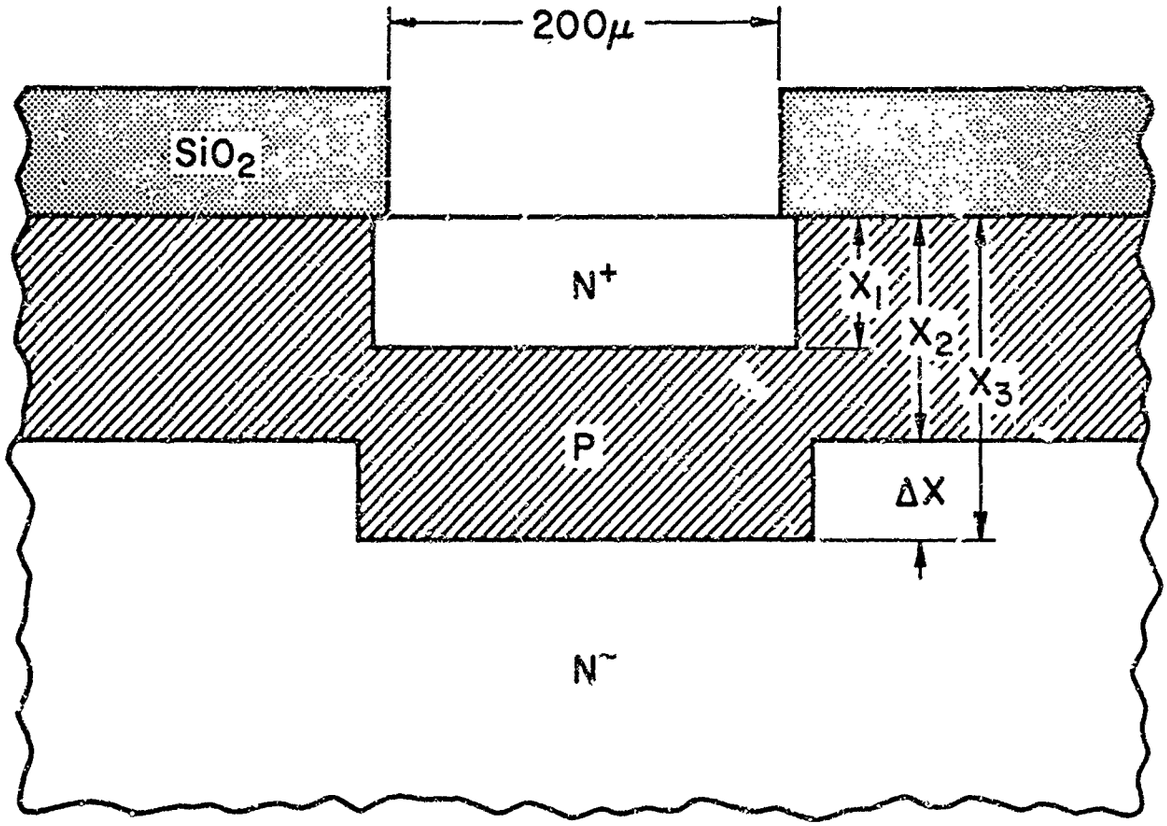


Fig. 17 Sections showing thin, intermediate and thick base layers (see Table III).



STANDARD NPN STRUCTURE

$X_1 = 0.4\mu$ $X_3 = 1.0\mu$

$X_2 = 0.7\mu$ $\Delta X = 0.3\mu$

Fig. 18 Cross-section defining distances for emitter dip studies.

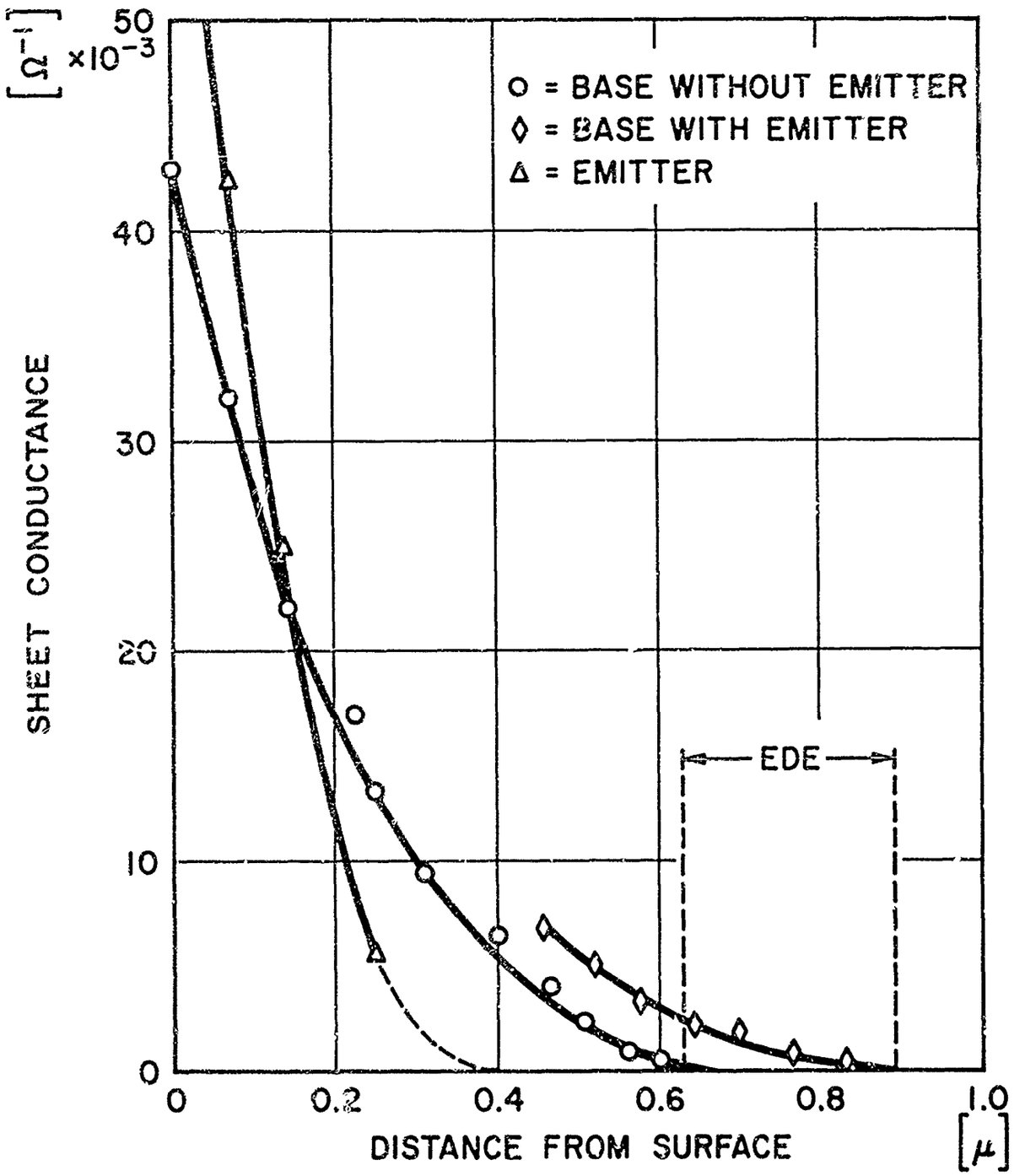


Fig. 19 Sheet conductance profile through standard emitter-dip structure.

6-1-64
51028

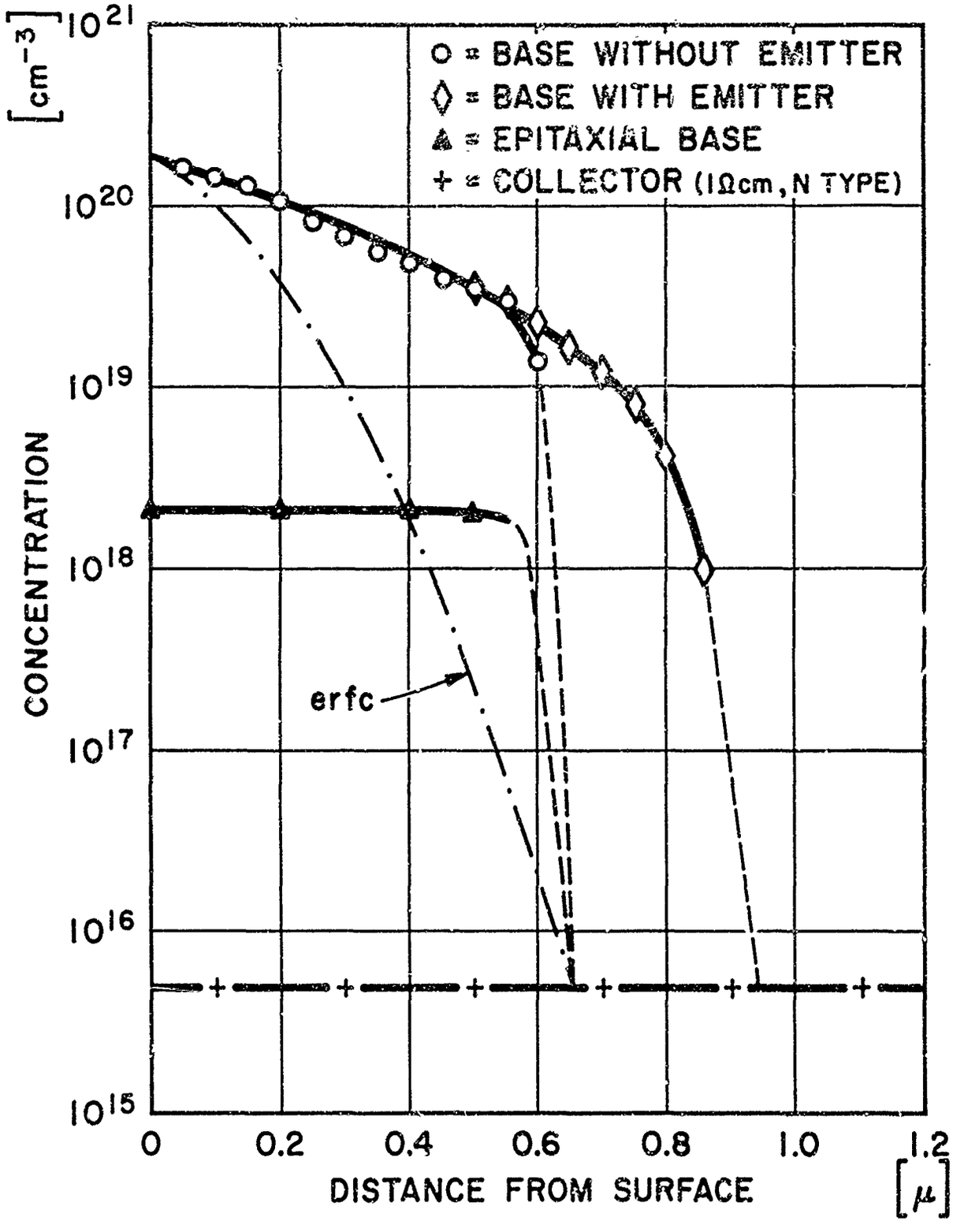
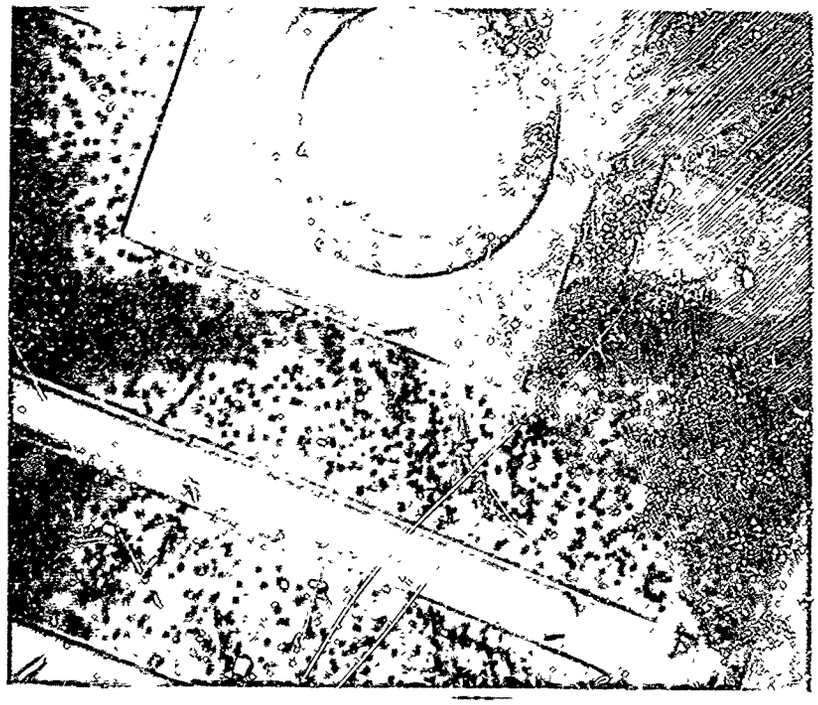


Fig. 20 Concentration profile through emitter-dip structures.



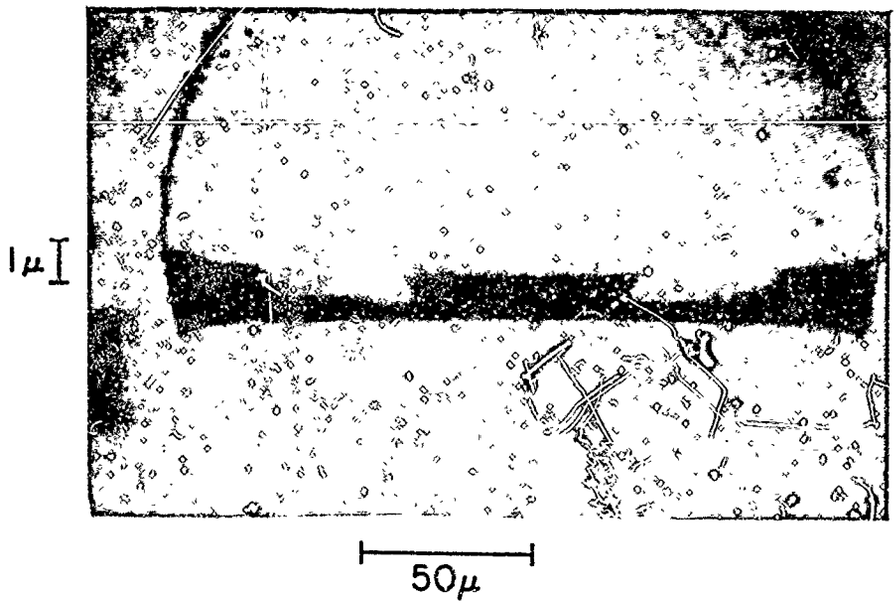


Fig. 21 Epitaxial base layers. A. Top view of slice showing growth irregularities. B. Cross-section showing non uniform penetration.

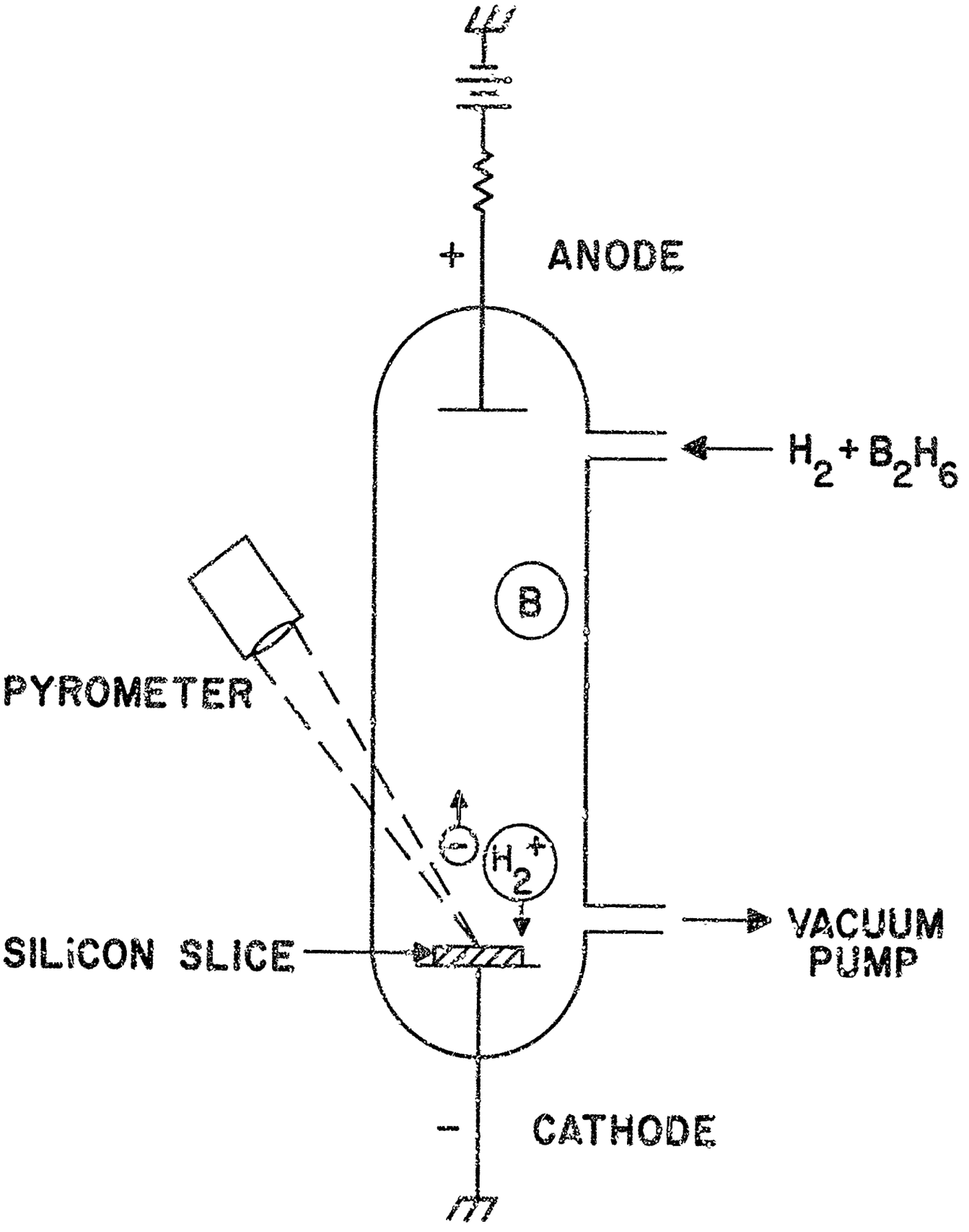


Fig. 22 Apparatus for ion bombardment.

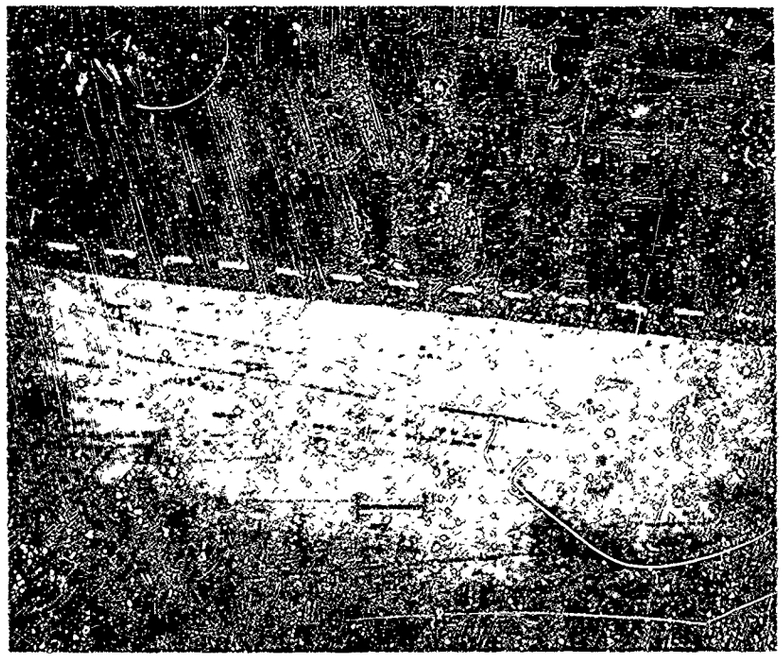


Fig. 23 Finger structure with n^+ emitter regions produced by ion bombardment.

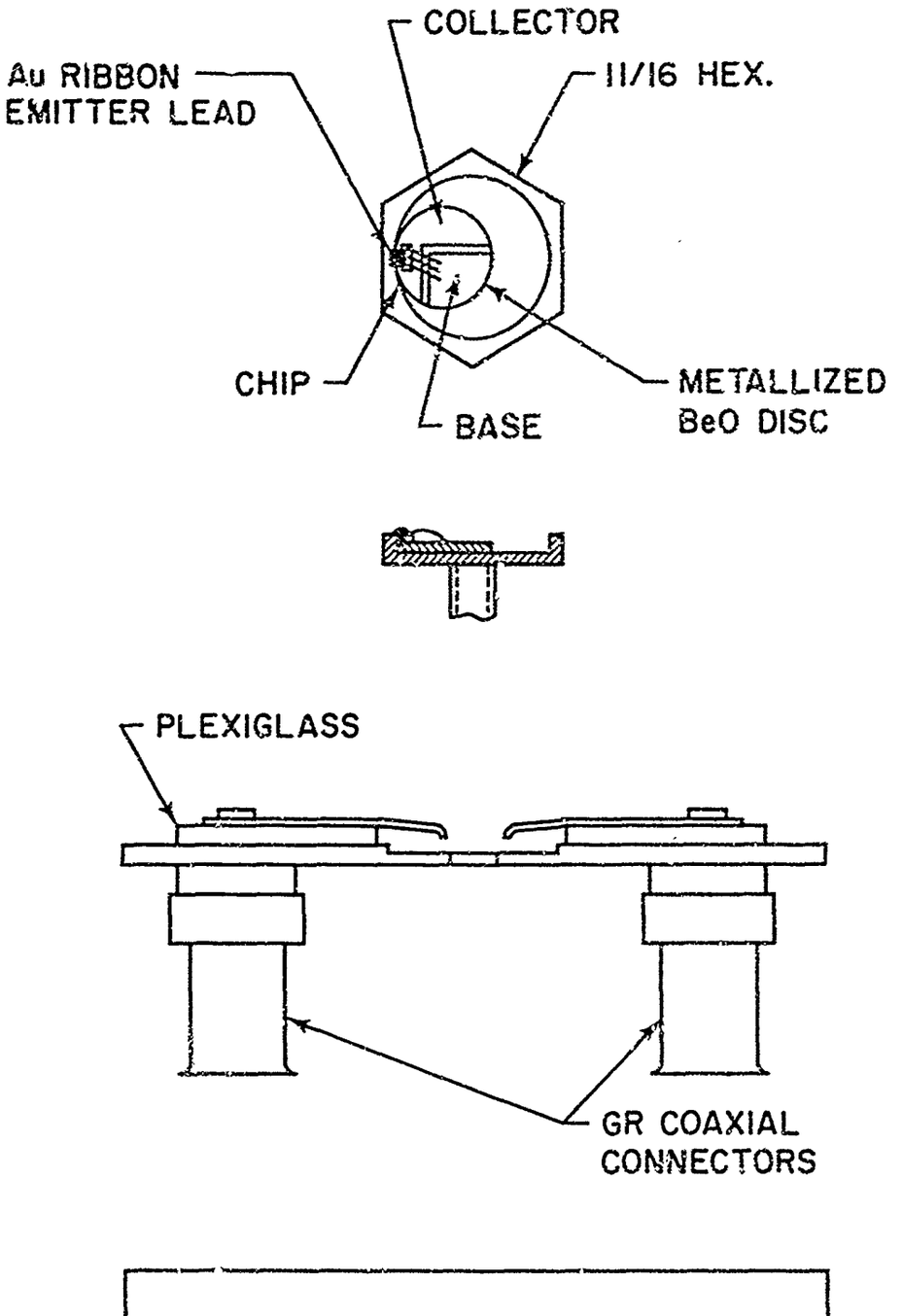


Fig. 24 Sketch of mounting base and test jig.

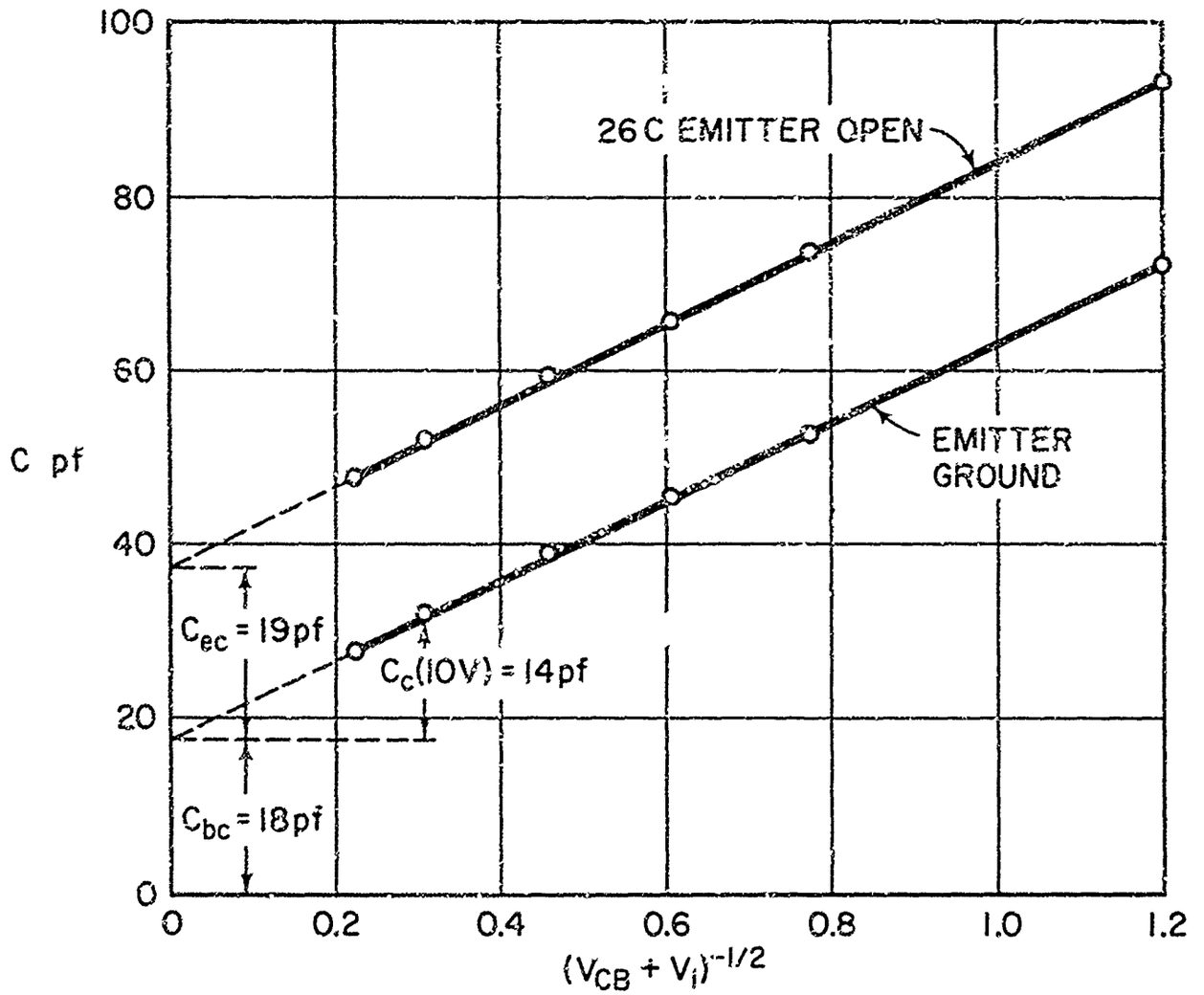
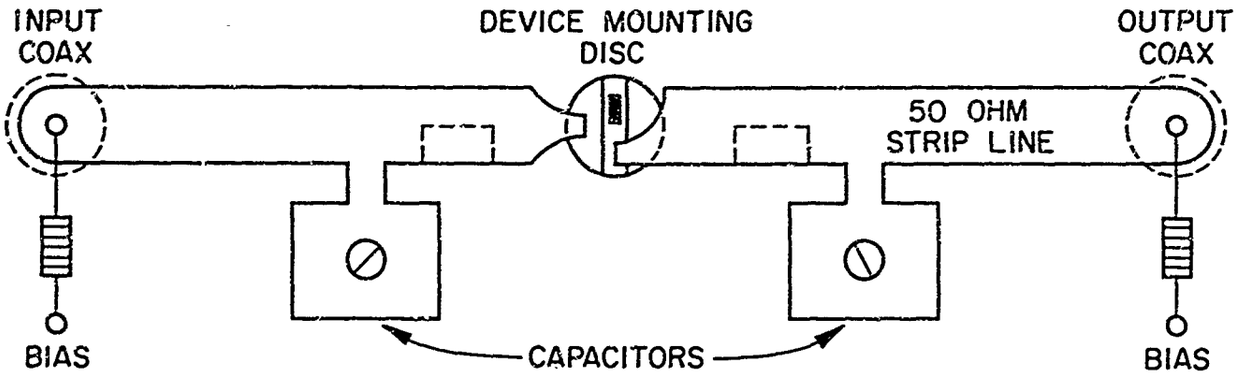


Fig. 25 Capacitance evaluation.



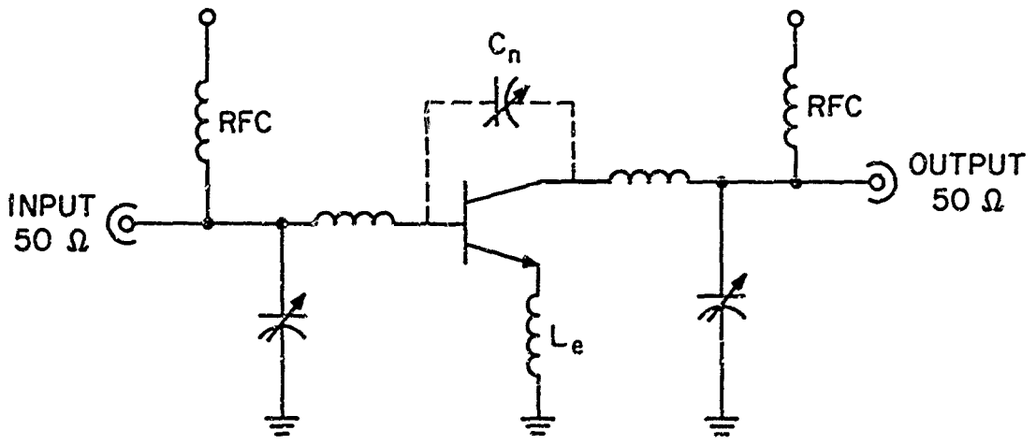


Fig. 26 "Lumped" 500 Mc circuit.

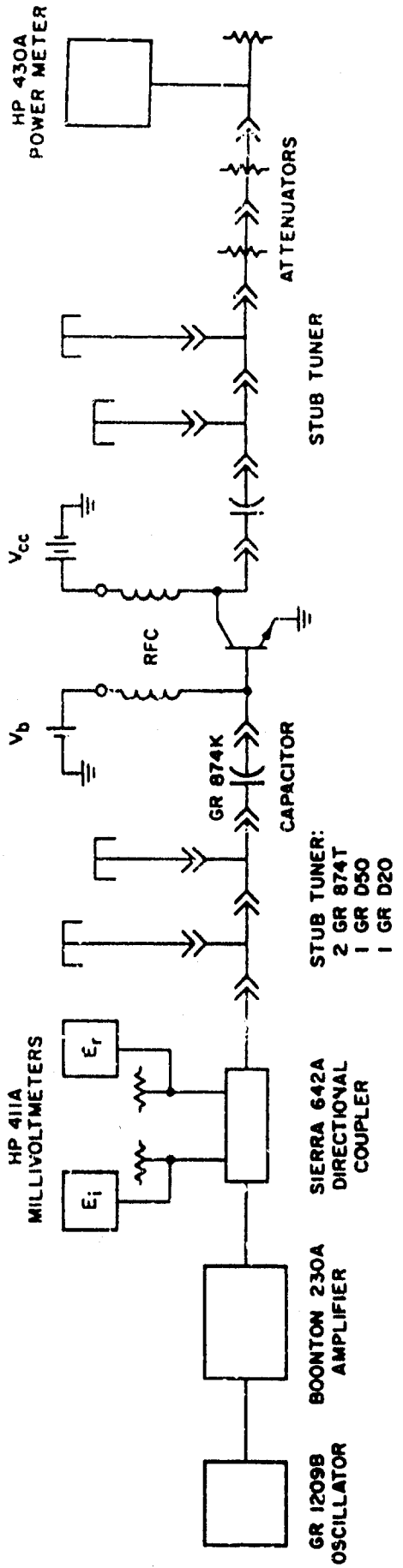


Fig. 27 Distributed 500 Mc circuit.

References

1. M. G. Holland, Phys. Rev., 134 A471, (1964); see also C. J. Glassbrenner and G. A. Slack, Phys. Rev., 134, A1058 (1964) These authors give values of about 1.6 w/cm°C at 30°C, decreasing to 0.8 w/cm°C at 230°C.

2. R. L. Pritchard, Proc. IRE, 43, 1075 (1955)

3. E. J. Ryder, Phys. Rev , 90, 766 (1953)

4. W. Shockley, BSTJ, 33, 799 (1954); see also U. S. Patent No. 2,790,037

5. C. T. Kirk, IRE Trans. on Electron Devices, ED-9, 164 (1962)

6. W. Shockley, BSTJ, 30, 990 (1951)

7. J. J. Sparkes, Proc IRE, 48, 1696 (1960)

8. W. Shockley, Solid State Elect., 2, 35 (1961)

9. K. G. McKay, Phys. Rev., 94, 877 (1954)

10. Terman, Radio Engrs. Handbook, pp. 51-52, McGraw Hill, (1943)

11. R. H. Finch and H. J. Queisser, Electrochem. Tech., 2, 114 (1964)

12. H. Strack, Bull. Am. Phys. Soc., 7, 543, (1962)

13. H. Strack, J. Appl. Phys., 34, 2405 (1963)

14. L. E. Miller, Properties of Elemental and Compound Semiconductors, p. 321, H. Gatos ed, Interscience, New York (1960)

15. P. Baruch, C. Constantin, J. C. Pfister and R. Saintespring, Disc. Faraday Soc., 31, 76 (1961)

16. G. E. Moore, Microelectronics, p. 262, E. Koenijian ed, McGraw Hill (1963)

17. H. Reiss, C. S. Fuller and A. J. Pietruszkiewicz, J. Chem. Phys., 25, 650 (1956)

18. E. Tannenbaum, Solid State Electronics, 2, 123 (1961)

19. J. C. Irvin, Bell Syst. Tech. J., 41, 387 (1962)

20. C. S. Fuller and J. A. Ditzenberger, J. Appl. Phys., 27, 544 (1956)

21. H. Reiss and C. S. Fuller, Semiconductors, pp. 222-268, Hannay ed., Reinhold, New York, (1956)

22. A. Goetzberger, Solid State Elect., 5, 61 (1962)

APPENDIX A

56)

THE SHORTED EMITTER STRUCTURE

This Appendix reports some preliminary considerations of the "shorted-emitter" transistor structure first proposed by W. Shockley. This principle enables emitter stripes to be made whose effective width is much less than their actual physical width. Deleterious effects associated with charging the base layer by the collector capacitance under the emitter (discussed in section 2.3.3 of the preceding report)may thereby be reduced.

The shorted-emitter principle can be illustrated by means of Fig. A1, which is drawn in a parallel manner to the non-shorted structure of Fig. 1. The difference is that in Fig. A1, the base layer extends up to the surface in the center of the emitter stripe (the emitter is not diffused here) and the emitter metal contact connects emitter and base layers together at this point.

Forward bias is applied between base and emitter in the usual way. This path is prevented from being completely shorted out by the relatively high sheet resistance R_s of the base layer under the emitter, across which the forward bias appears. It is evident that a potential will exist in the base layer ranging from 0.7 - 0.8 volts at the active edge of the emitter to zero volts at the location of the "short", and hence most of the emitter-base junction will be prevented from injecting. This potential gradient prevents losses due to the base layer charging effect referred to above. That portion of the collector capacitive current arising from the strip $(L_e - L'_e)$ and from the strip approximately $L'_e/2$ flows harmlessly to the emitter contact. Since this portion does not have to be supplied from the base contact, it has no influence on the power gain. The effective emitter stripe width is thus less than $L'_e/2$, which

is narrower than can be achieved by conventional construction by a factor of the order of two to three.

We shall make a rough estimate of the optimum width of the emitter L_e', assuming other dimensions remain fixed and that $R_s \gg R_s'$. The resistance of the "short" under the emitter per unit periphery is $I/R_s L_e'$. Assuming that an ac voltage v_f is developed across this resistance in Class B operation, the power absorbed in it is $v_f^2/R_s L_e'$. The only other component of input power dependent on L_e' comes from the peak collector capacitive current density ωQ_c ($Q_c = \overline{C}_c V_c$) flowing through a strip of width $L_e'/2$, giving a power of about $(\omega Q_c)^2 R_s L_e'^3/8$. The total relevant input power is thus

$$P_{in}(L_e') = v_f^2/R_s L_e' + (\omega Q_c)^2 R_s L_e'^3/8 \qquad \text{(A 1)}$$

This power is minimized by making

$$L_e' \text{ opt} = 1.28\left(\frac{v_f}{\omega Q_c R_s}\right)^{1/2} = 1.28\left(\frac{v_f}{V_c}\,\frac{C_e}{\overline{C}_c}\right)^{1/2} L_\omega \qquad \text{(A 2)}$$

where C_e is the emitter capacity and L_ω is the charging depth discussed in the previous report. Generally, the charge swing on the emitter ($v_f C_e$) will be considerably less than that on the collector ($V_c \overline{C}_c$), so L_e' opt will be less than L_ω by a factor perhaps of the order of two. This implies that the shorted emitter principle may offer little advantage with respect to the base layer charging effect at frequencies where L_ω is larger than the width easily achieved in ordinary emitter stripes.

The simple analysis just presented omits consideration of the fact that the capacitive current from the strip $L_e - L_e'/2$ flows to the emitter and does not have to be supplied through the base resistance. Higher power gain should thus result; the expected improvement should

be calculable from the power gain expression given in section 1. The capacitance C_1 is effectively reduced by a factor of the order of $L_e'/2L_e$. To first order, the power gain will be improved by an amount equal to the factor by which the total capacitance $C_1 + C_2$ is effectively reduced; however this ignores the power lost in the shorting resistance which could become appreciable in Class B operation.

If the input signal is required to drive the transistor into conduction (as, for example, in the circuits discussed in section 4, and commonly referred to as "Class B"), substantial power loss could occur in the shorting resistance as the above analysis attempts to account for. However the initial bias current required through the shorting resistance to reach approximately 0.6 volts forward bias on the emitter edge could be provided by a separate dc current source. Experimentally, this increased gain has not been observed, and the discrepancy is not understood.

In preliminary experiments, the shorted structure was applied to an existing 50 - 100 Mc power transistor having an emitter stripe width of 75 μ. This width is of the order of the applicable L_ω length. The shorted structure resulted in a power gain improvement of 2 - 3 db and a slight improvement in efficiency. These results are roughly as expected, and serve to demonstrate the principle. The feasibility of applying this principle to the 500 Mc design and the improvements to be expected are more doubtful. The reason is that the emitter stripes are much narrower, leaving little room for the shorted region. The distance L_e' and consequently the resistance of the short become quite small unless compensated for by high sheet resistance R_s. But high sheet resistance will result in a smaller power handling capability according to the discussion in section 1. Perhaps the shorted-emitter principle could be applied to geometries like that shown in Fig. 2c without introducing excessively low shorting resistances.

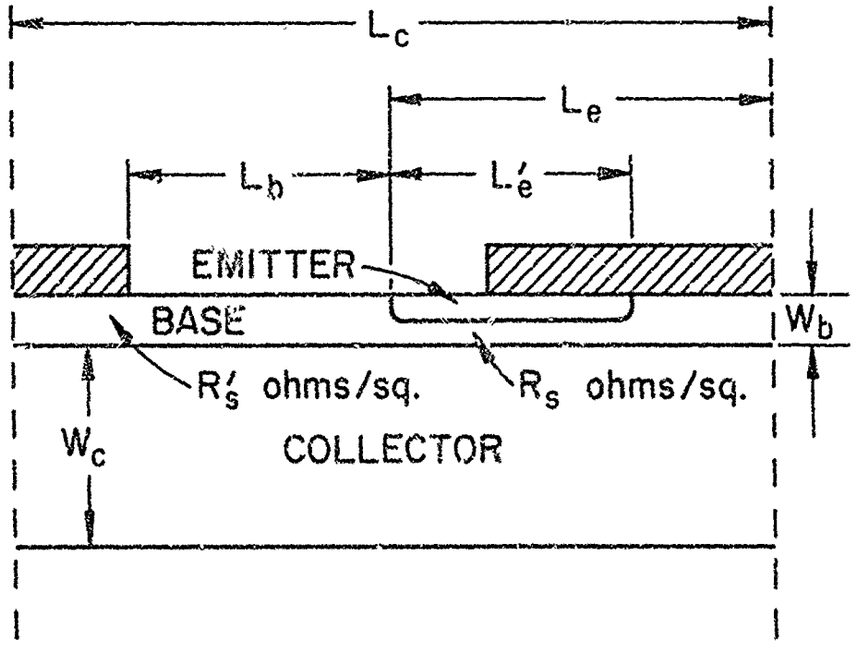


Fig. A1 Cross section through shorted-emitter device

APPENDIX B*

THEORY AND EXPERIMENT ON CURRENT TRANSFER FROM ALLOYED CONTACT TO DIFFUSED LAYER

B1. The Transfer Length L_t for the Continuously Distributed Model of the Aluminum-Silicon Contact

The occurrence of a contact resistance between metal contact and diffused layer in a silicon transistor can be detected by electrical probing. One method of carrying out such probing to obtain a quantitative result is illustrated in Fig. B1. It is assumed that a contact now shown on the figure introduces a current J amp/cm flowing from right to left in the figure. This current is removed from the aluminum contact. If voltage is probed between the aluminum contact and various positions at coordinate x on the diagram, then a voltage plot like that shown in the lower part of the figure is obtained. It is found that there is an abrupt "step" in voltage at the edge of the aluminum.

As will be shown below, this step can best be described in terms of the quantity L_t, defined as the "transfer length". This is the effective length over which current must flow in the diffused layer before it transfers through the alloy contacts into the metal layer. As is represented in the diagram, there may be uncertainty in the measurement of the quantity L_t depending upon the resolution of the probing apparatus. If data like that shown in part (b) of the figure is obtained, then the value of L_t has an uncertainty equal to the interval between measured points. This is a consequence of the fact that from the electrical measurement it is not possible to say exactly at what point the aluminum plating stops, and all that can be said with certainty is that it ceased between one measured value and the other.

---

* Work performed by W. Shockley

The uncertainty in L_t values is in general of little practical importance. The reason for this is that for contact resistance to be troublesome the distance which current must flow in the diffused layer must be larger than the resolution possible in probing. Consequently, the actual resistance in the transistor structure is unimportant even if it produces a voltage drop corresponding to an L_t value somewhat larger than the resolution of the probing experiment.

For the interpretation of the quality of alloy contacts, however, it is desirable to obtain L_t values with a maximum of precision.

In terms of a continuous distributed model of the properties of the contacts, the value of L_t can be visualized in terms of the diagram shown in Fig. B2. This represents the same situation as in Fig. B1 but in terms of an equivalent circuit. It is supposed that the resistance per square of the diffused layer is R_s. It is also supposed that the particular structure extends perpendicular to the plane of the diagram for a distance M. On the diagram two resistances are represented. One of these is the resistance which would arise if the current were to flow from the edge of the metal through the diffused layer alone for a distance L_t. The resistance of this rectangle, as shown on the figure, is $R_s L_t/M$, in which L_t/M is the number of squares in series for this current path.

The other resistance represented in the diagram is the resistance for current flow from the diffused layer to the metal plate, assuming for purposes of this estimate that the diffused layer is an equipotential. In this case, the area for current flow considered is again L wide and M deep. The contact resistance arising from the alloying is described in terms of an infinite distribution of parallel resistors which will carry G_a amperes per square centimeter when the voltage difference is 1 volt

In other words, G_a, called the "alloy transfer conductance," has the dimensions of mhos per square centimeter. This leads to the resistance $1/G_a$ LM shown on the figure.

Physical reasoning, which is given in mathematical detail in the First Interim Report, leads to the conclusion that the current flow will distribute itself under the metal plate so that the resistance in the diffused layer will be approximately equal to the resistance in flowing across the contacts. This leads to the relationship (B3) given below in terms of the previously defined quantities:

$$R_s = \text{ohms per square of diffused layer} \qquad \text{(B1)}$$

$$G_a = \text{conductance in mho/cm}^2 \text{ between diffused layer and contact metal} \qquad \text{(B2)}$$

$$1/L_t M G_a = (L_a/M)R_s \qquad \text{(B3)}$$

This equation is readily solved for L_t and gives

$$L_t^2 = 1/R_s G_a \qquad \text{(B4)}$$

The same relationship results from an exact analysis.

We next consider the situation prevailing when a metal contact finger, having width L_f, lies over the diffused layer. In general, equal currents will flow towards this metal strip from both sides, as represented in part (a) of Fig. B3. As is represented in part (b) of the figure, this leads to a voltage step at the edge of the layer which can be described in terms of a conductance per centimeter of edge length such that the current J in amperes per centimeter is equal to this current

times the alloy transfer voltage V_a. Mathematical analysis given leads to the conclusion that the voltage disturbance under the alloyed layer must vary exponentially with distance, as represented in part (b) of the figure, so that the value for the finger conductance in mhos per centimeter is given by

$$G_f(L_f) = (1/R_s L_t) \tanh (L_f/2L_t) \tag{B5}$$

Evidently, if the finger is very wide a maximum value of G_f will be obtained. This value is represented by $G_{f\infty}$ and is the limiting form given by

$$G_{f\infty} = 1/R_s L_t = (G_a/R_s)^{1/2} = G_a L_t \tag{B6}$$

On the other hand, if the width of the finger is very small compared to the alloy transfer length L_t, then the formula for G_f reduces to

$$G_f \cong L_f/2R_s L_t^2 = L_f G_a/2 \quad ; \quad L_f < L_t \tag{B7}$$

It is seen that this equation reduces to a simple expression which says that the conductance is simply equal to the conductance of the contacts for a strip whose width is half that of the finger width.

It is useful to calculate the additional resistance which will arise from the contact as compared to the resistance that would exist in a transistor if the contact were perfect and had no resistance. In the latter case the important resistance is that from the metal contact through the diffused layer to the nearest junction. This is usually given by half the width L_{df} of the diffused finger, as represented in

part (a) of Fig. B3. For the case in which the metal finger L_f is narrow compared to the alloy transfer length L_t, the extra resistance from metal contact to junction is boosted by a factor given by

$$\text{Boost in resistance to junction} = 1 + (2L_t^2 / L_{df} L_f) \qquad (B8)$$

It should be noted that the boost factor in resistance, which is associated with the decrease in G_f of Eq. (B7) for small values of L_t, varies as L_t^2. For this reason it is quite possible that boost factors of an order of magnitude can occur in devices which have large values of alloy transfer length.

A convenient method for measuring L_t under some circumstances is represented in Fig. B3. Here current is assumed to flow only in one side of the metal contact. For this case the voltage V_0 which appears on the side where no current flows is given by

$$V_0 / J = R_s L_t / L_t) \qquad (B9)$$

For large values of L_f / L_t, this equation limits the accuracy of measurement to the accuracy with which the width L_f of the metal finger can be determined. For the situation $L_f / L_t >> 1$, the important term in this equation will be the hyperbolic sign function. This has the consequence that inaccuracies in measurement of the other quantities in (1.9) will affect the determined value of L_t only logarithmically, and thus L_t's accuracy is almost completely determined by the accuracy with which L_f can be measured.

B2. The Discrete Contact Model of the Transfer Conductance G_a

In this section we shall consider some consequences of potential theory for the situation in which the diffused layer makes only occasional contacts at discrete small regions with the metal layer. The essential quantities involved are represented in Fig. B4 and are defined as follows:

$$D = 2r = \text{diameter of ohmic contact} \tag{B10}$$

$$S = \text{center to center contact spacing} \tag{B11}$$

$$W = \text{effective layer thickness} \tag{B12}$$

In terms of these quantities the sheet resistance R_s in ohms per square of the diffused layer is related to an effective thickness of the diffused layer and a resistivity. This resistivity ρ and its reciprocal, the conductivity σ_s are the values corresponding to the surface of the layer, and the layer is treated as if it were uniform in conductivity, having the same value as the surface. This approximate model will introduce small errors when the contacts are small in diameter compared to the effective thickness W. Under these conditions the principal voltage drop of current flowing into the contact occurs within a few contact radii of the center of the contact, and in this region the resistivity will be fairly uniform and will have the value given at the surface. For this case the relationship

$$1/R_s = \sigma_s W = W/\rho_s \tag{B13}$$

holds for W and the conductivity and resistivity at the surface.

If we assume that the resistance of one of the small areas of

contact shown in Fig. B4 to the layer may be represented by R_a, then the value of G_a, the alloy contact conductance, will evidently be given by

$$R_a = \text{resistance of contact to layer in ohms} \tag{B14}$$

$$G_a = (1/R_a) / S^2 = 1/R_a S^2 \ (G_a \text{ in mhos/cm}^2) \tag{B15}$$

This leads to a relationship between the center to center spacing of the contacts and the alloy transfer length given by

$$L_t^2 / S^2 = (1/R_s G_a) / S^2 = R_a / R_s \tag{B16}$$

The concept of a contact resistance to the layer is mathematically a difficult one, since the diffused layer is not an equipotential in this case. A mathematical treatment leads to the conclusion that for a discrete contact structure like Fig. B4 there is a characteristic alloy transfer length L_t which specifies how the current flowing in the layer decays exponentially and is transferred to the metal contact. The relationships given above are based on this mathematical analysis.

Fig. B4 illustrates the case in which the alloy transfer length is larger than the center to center spacing by approximately the square root of 2. As a consequence of this, it works out that the current flowing in the diffused layer decays by approximately a factor of 2 in each period S. This decay is represented in Fig. B4, where it is seen that 50% of the current flowing in from the right of the figure decays on the first row of contacts, 50% of the remainder on the second row, and 50% of the subsequent remainder on the third row, and so on. It is also noted in Fig. B4 that the actual contact area for the

representative situation shown is extremely small.* Only one part in 1600 of the area is actually covered by contacts, and yet the first row of contacts stops 50% of the current, so that the effective depth of penetration is, as shown in the diagram, less than 3/2 row spacings. This illustrates how effective very intermittent contacts can be in transferring current.

Conversely, the diagram suggests that where a relatively large value such as 20 or 30 microns is obtained for L_t, the fraction of the area which has actually been contacted in the alloying process is extremely minute.

In terms of the quantity G_a, expressed here in terms of the contact resistance (B14), and the analysis of Section 1, which leads to values of G_a from measured values of R_s and L_t, it is possible to apply the theory discussed in Section 1 to actual situations. The mathematical analysis presented in the Appendix shows how values for the contact resistance R_a may be calculated from values of S, W, D and R_s. Two cases are of importance, and the distinction is represented in Fig. B5. Either the contact area is large compared to the effective layer thickness or it is small. In the event that it is large compared to the effective layer thickness, a two dimensional potential theory model is adequate and the resistance required for the current to penetrate through the distance W of the layer is unimportant compared to the resistance in spreading further outwards. In this case it turns out from the analysis presented in the First Interim Report that

---

*Eq. (2.14) gives $D/W = 1/4$ for $L_t/S = 1.4$.

the appropriate value of R_a to use is given by

$$R_a = (R_s/2\pi)\ \ln\ (S/\pi D) \qquad D > 2W \tag{B17}$$

This value of R_a corresponds to current flow to an outer cylinder whose diameter is the center to center spacing divided by π.

On the other hand, if as represented in Fig. B5 the area of the contact is actually small compared to the thickness, then the principal voltage drop occurs within a few radii of the contact area (it being assumed, of course, that the actual regions of contact themselves have negligible resistance and simply establish an equipotential on the surface of the semiconductor). This leads to an approximate expression for the total resistance consisting of the sum of two terms

$$R_a \cong (R_s/2\pi)\ln\ (S/2\pi W)$$

$$+\ (R_s/4)\ [(2W/D) - 1] \qquad D < 2W \tag{B18}$$

This form is not exact but has been chosen so as to join correctly to (B17) when the diameter is equal to twice the effective layer thickness W and to vary properly for a contact spreading resistance when the diameter is small compared to W. The transition region in which (B18) must be used to obtain accurate results is probably relatively small, and when D is considerably less than W it is adquate to approximate Eq. (B18) by the much simpler expression

$$R_a \cong R_s\ W/2D \qquad D << 2W \tag{B19}$$

The case of Fig. B10 is based on this approximate formula.

In terms of Eqs. (B18) and (B19) when combined with (B16), one may obtain an expression relating the diameter of the contact to the measured quantity L_t and an assumed quantity S for the contact interval. These relationships are represented as follows

$$D = (S/\pi) \exp(-2\pi L_t^2/S^2) \qquad D > 2W \qquad \text{(B20)}$$

$$D = (S/L_t)^2 W/2 \qquad D << 2W \qquad \text{(B21)}$$

Experimentally it may be difficult to determine the quantity S unless the contact areas are large enough to be seen. In general, when this is the case the values of L_t are sufficiently small to be of little practical interest.

It is, however, possible experimentally to make an estimate for an upper limit for the value of S. This can be done by probing transistor structures and looking for current sinks. For example, if the value of S were as large as 25 microns in a typical transistor structure, then there would be substantial discontinuities in the current as measured along a metal contact finger. Observations made so far do not indicate the presence of such pronounced localized current sinks for the current flowing in the metal contact areas. Consequently, for the case studied it is concluded that the effective value for S is substantially less than 25 microns.

This reasoning leads to the conclusion that contact areas may be very small indeed. For example, in one case a value of approximately 90 microns was obtained for L_t. Assuming that S was only about 10 microns, using a value of 0.25 microns for W, one concludes that the diameter of the contact was less than 100 angstroms or 10^{-6} cm in diameter.

B3. Some Experimental Values and Discussion of Measurements

The large variation that may occur in the contact situation is suggested by the photographs shown in Fig. B6. These were made by etching the aluminum from a completed transistor structure by using sodium hydroxide. This structure has 75μ wide emitter diffused fingers, but the contacting procedure is identical to that used on the 500 Mc transistor. It is observed that in some cases large numbers of spots in the etching pattern are shown in regions where the metal was alloyed to the semiconductor. (It should be noted that there is currently no objective proof that the etch pattern actually corresponds to ohmic contact between the aluminum and the silicon. However, the correlation between such etch patterns and the observed alloy transfer lengths suggests that it is a very reasonable assumption that these etch pits do represent actual intimate metallic alloying between the aluminum and the silicon.)

The structure shown in part (b) of the figure was a transistor extensively measured by electrical probing of potentials from point to point. On the basis of these measurements and by using mathematical analysis like that discussed in Section 1, it was concluded that the alloy transfer lengths for emitter and base contacts were of the order of 60 microns and 90 microns. If the contacts for such a case were as small as the 100 angstroms discussed at the end of Section 2 (a value which corresponds to these alloy transfer lengths according to theory), then it is not surprising that no etch pits can actually be seen in the photograph.

The resistance boost factors associated with structures of this sort, estimated from Eq. (B8), are about an order of magnitude

and can account for resistances as measured at low frequencies appearing in series with the emitter-base junction of 6 to 12 ohm-cm (reciprocal of the quantity G_f discussed in connection with Eqs. (B5), (B6) and (B7) for the base resistance and 0.3 ohm-cm for the emitter resistance. Such relatively large values can account for extra saturation voltages and reduced power gain.

In the course of carrying out these contact investigations a transistor using a metal contact consisting of 250 angstroms of aluminum, followed by 950 angstroms nickel, followed in turn by 1.5μ of aluminum was made. This contact, which can be made only with a metal mask, allows higher temperature alloying of the aluminum without penetration through the base layer. Such transistors gave low saturation voltages; from these and from the logarithm of collector current versus emitter-base voltage it was concluded that the effective series resistance with the emitter was less than 0.10 ohm-cm.

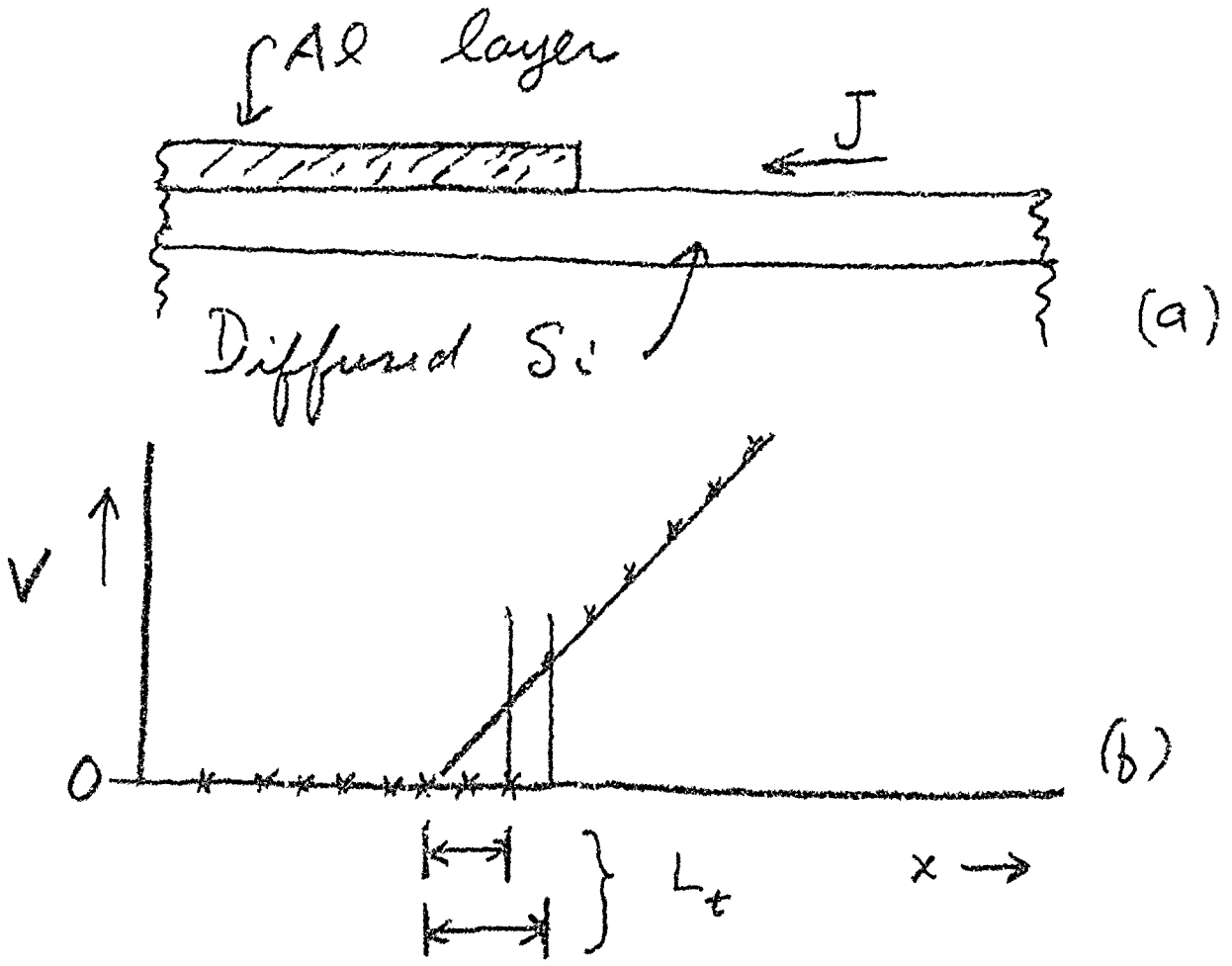


Fig. B1 Method of measuring "transfer length".

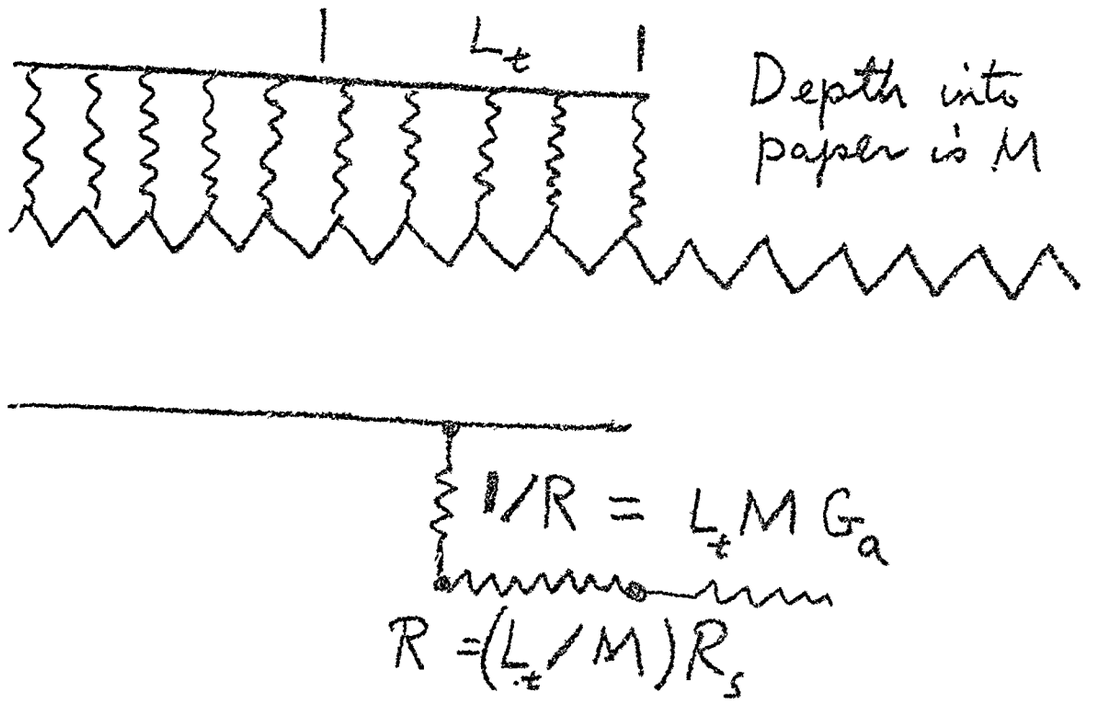


Fig. B2 Definition of L_t.

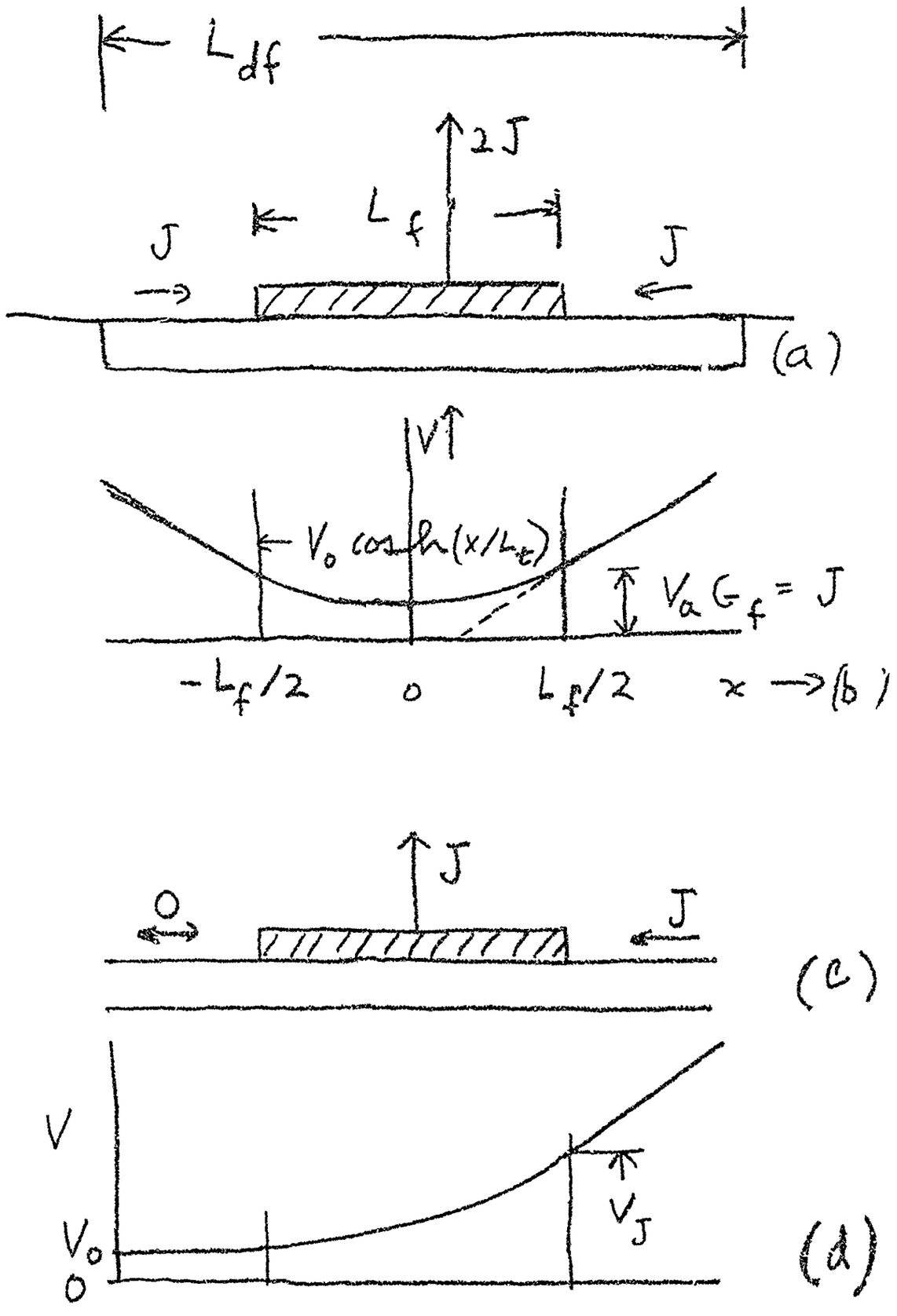


Fig. B3 Penetration of voltage under finger.

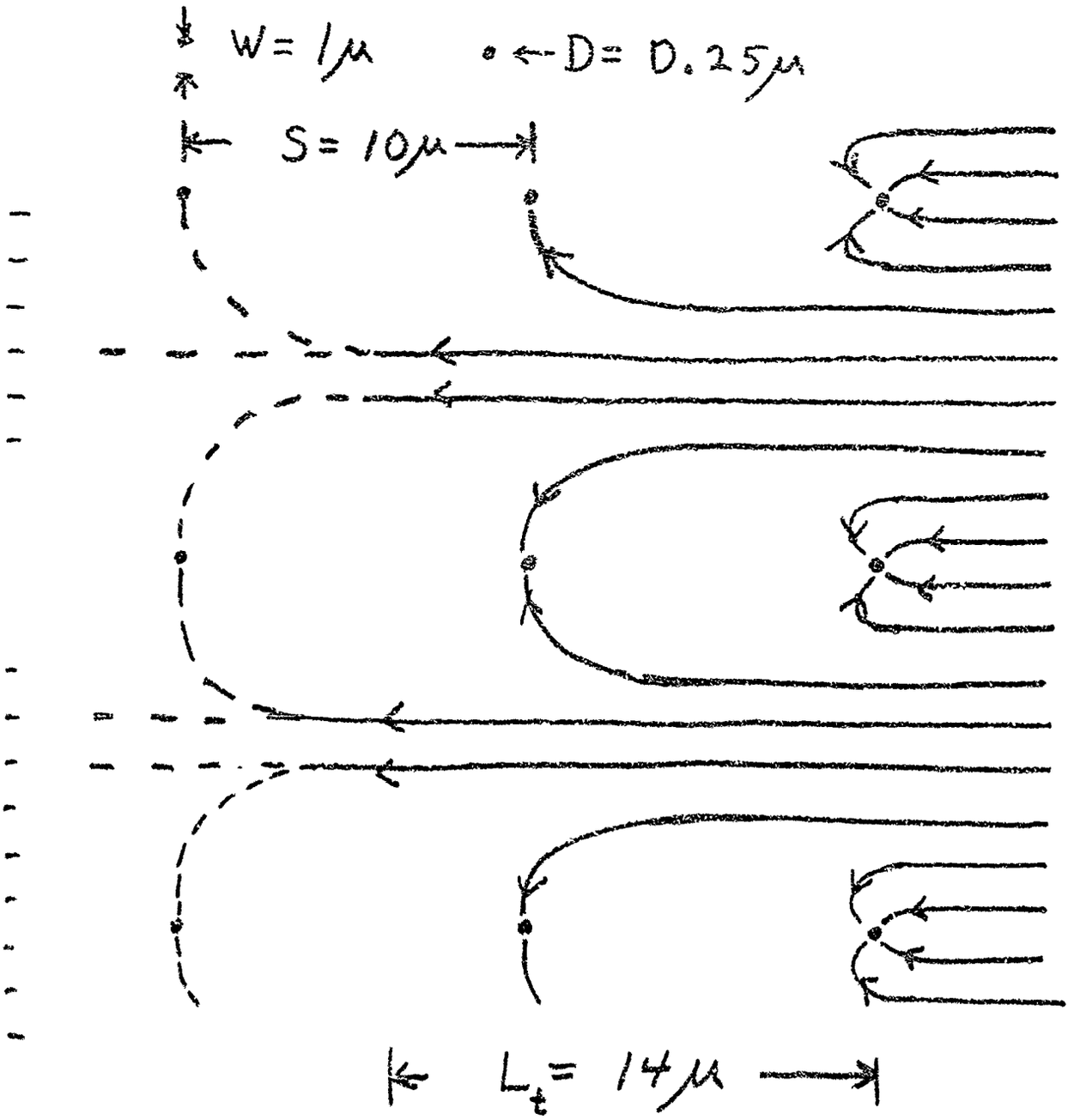


Fig. B4 Contact illustration.

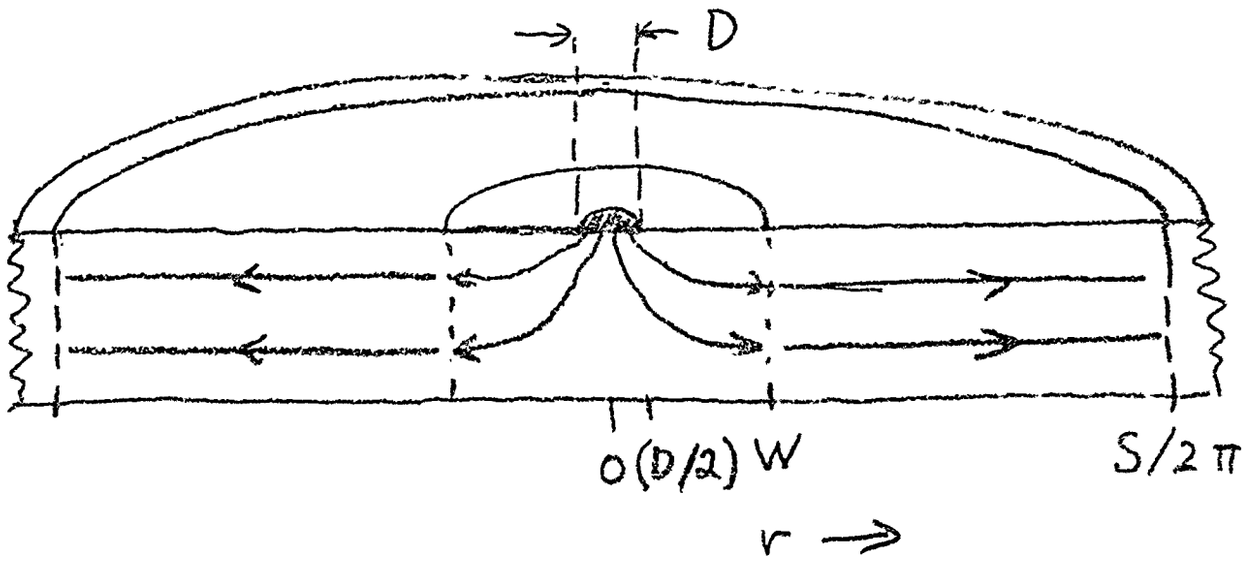
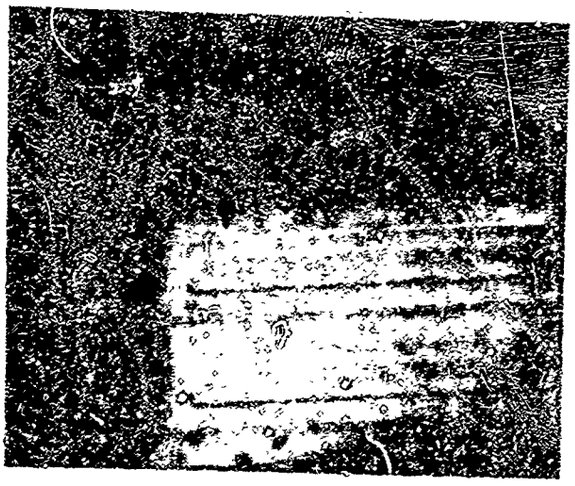
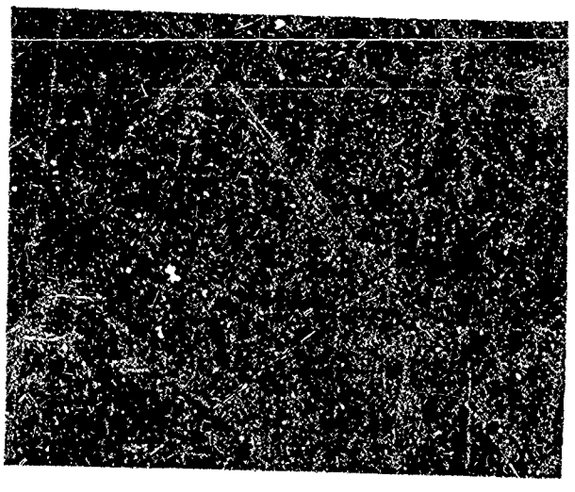


Fig. B5 Contact spreading resistance.



(a)

(b)



(c)

Fig. B6 Photographs of alloyed contacts.

APPENDIX C

CONTROL MASK STRUCTURES*

C1 Introduction

During the course of this contract, it became evident that it would be desirable to monitor as many critical process variables as possible. This can be done by separate control slices, or more desirably, by incorporating special structures in the same masks as the final transistors. The initial mask design anticipated this by including a concentric ring structure for measurement of base resistance under the emitter, shown in Fig. 10. This structure could also be used to estimate contact resistance by probing methods suggested in Appendix B. Although it was not necessary to design a completely different set of masks, the following considerations were developed for the design of suitable control structures in future masks. The discussion is in terms of a slightly different geometry than that used elsewhere in this report, but it will be evident how the principles apply to any geometry.

C2 Principal Objectives

In fabricating power transistor structures, it is desirable to know the values obtained for various process steps involved. This discussion deals with five parameters. One particular parameter of the process which may cause deterioration in transistor performance is the resistance between the metal contact and the emitter layer and between the metal contact and the base layer. Such resistances have been discussed in Appendix B.

---

* Work performed by W. Shockley

There are also three sheet resistances, which may be expressed in ohms/square, for diffused layers. These are for the emitter diffusion, the base diffusion as a whole, and the base diffusion as it lies under the emitter layer.

Each of the set of masks discussed here is intended to measure essentially one of the five physical quantities discussed above, namely the contact resistance between the aluminum and the base layer, and the three sheet resistances. In fact each mask individually measures essentially just one of these five parameters.

The design of the masks was suggested by an observation that the evaporated metal structures of a transistor could be used as electrodes for contact to control structures. Although more reliable electrical information can be obtained by using three electrodes, it is possible to design masks in which a high degree of control information can be obtained by using two electrodes only, i.e. those normally serving as base and emitter contacts. The advantage of using only two electrodes is that the measurements may be made on a slice simultaneously with measurements made of the transistor parameters. The control areas are identical in their contact geometry with the actual transistors. The electrical measurements consist simply of measuring the resistance between the emitter and base contact areas.

The cases of particular practical interest are those for which the current transfer length L_t (see Appendix B) is comparable to the dimensions of the metal contact fingers. If the transfer length is substantially less than the width of the contact fingers, then contact resistance is unimportant. On the other hand if the transfer length is very large compared to the metal contact finger width, then in general large

additional contact resistances will be present. Consequently, the design of each control structure is made so as to permit relatively simple and straightforward interpretation in the range in which the transfer length is comparable to the dimensions denoted by "a" of the metal fingers. In general this same dimension "a" represents the spacing between the metal fingers and the emitter-base junction.

Another design consideration for the control structures discussed here is that the normal measured resistances should be preferably of the order of 100 to 1000 ohms. This is because the resistance between the electrical pressure contacts used in a test set-up and the aluminum layers may be of the order of 1 to 2 ohms. The resistance of the metal contact layer itself varies from about 0.2 to 0.02 ohms per square. For the structures considered the maximum number of squares in series will be less than 10 so that relatively unimportant resistances of less than one ohm are to be expected. Thus if the measured resistances are larger by 2 orders of magnitude, the error made due to variations in the metal contact resistance will be unimportant.

C3 Two Terminal Transistor Control Structure Designs

The designs are based on a typical set of values for the sheet resistance which we denote by the letter S with an appropriate subscript. Typical values for the three diffused layers and the number of squares that should be incorporated in series or parallel in order to give desirable resistances are shown below:

$S_e \doteq 2$ ohms ; requires 10 to 100 squares in series (C1)

$S_{bb} \doteq 30$ to 300 ohms requires 1 to 10 squares in series (C2)

$$S_b \doteq 5000 \text{ ohms} \quad ; \quad \text{requires about 10 squares in parallel} \tag{C3}$$

The resistance per square S_{bb} stands for the thick base layer which does not lie under the emitter; the other symbol S_b represents the working region of the base layer lying under the emitter.

Sheet Resistances

Figure C1 shows the structure used for measuring the quantity S_e. Part (a) represents the diffused and contacted structure and Part (b) of the figure indicates where it lies in respect to the transistor structure. It is seen that the metal areas where contact is made to the diffused emitter region lie under the massive bus bar portions of the transistor structure. In some designs, it may be necessary to use a group of emitter fingers for one contact, and a group of base fingers at the opposite end of the structure for the other.

For ease of measurement it is desirable to make the resistance between the two contacts the order of approximately a hundred ohms so that the ratio of the widths W to the length H of the connecting diffusion strip should give the order of 10 to 100 squares. If necessary, a slanting or zig-zag path can be used.

Sheet Resistance S_{bb}

Figure C2 represents the structure used for measuring S_{bb}. In this case the desired resistance corresponds to approximately one square of this diffusion. This means that the spacing between contacts and widths of the region measured should be approximately equal. Figure C2, like Figure C1 exaggerates the relative dimension of the transfer length for the contact and also the registration tolerance As represented in

Part (a) of Fig. C2 the base layer is surrounded by a region in which the emitter is diffused. Under this region the sheet resistance of the base material is S_b which is one or two orders of magnitude higher than the resistance within the rectangle where the emitter is not diffused. Consequently only a small correction need be made for the conductance of this surrounding region and for practical purposes it can be neglected.

It should be remarked that for the chief purpose of these masks a high degree of accuracy is not required. Instead the chief desideratum is that control will be obtained so that significant variations from one slice to another or one region to another can be detected.

Sheet Resistance S_b

Figure C3 shows the structure planned for measuring S_b the resistance of the base layer where it lies under the emitter. In this case it is desirable to have the order of 10 to 100 squares in parallel. The structure shown has 30 squares in parallel. As is represented on the figure, an emitter diffusion extends over a "picture frame" or rectangular region lying in the area of the interdigitated emitter and base fingers. Marked on the edge of the rectangle are intervals of magnitude "a" so as to illustrate tolerance considerations that may be important in designing the structure. The structure has been based upon permitting a tolerance of 2a in registration vertically and 1.5a horizontally so that the demands on this structure will be somewhat less than those placed on making and registering the transistor proper. The metal finger electrodes used for making contact are chosen in an alternating fashion so as to produce a much larger interval for the emitter diffusion and to decrease registration tolerance.

Contact Resistance and Current Transfer Length

We now come to a consideration of the contact resistance measurements. These are represented to a first approximation by a certain conductance per cm^2 of metal semiconductor area and may be represented by the symbol G:

$$G_e = \text{emitter contact conductance mho/cm}^2 \tag{C4}$$

$$G_b = \text{base contact in conductance mho/cm}^2 \tag{C5}$$

These quantities are related to the corresponding current transfer lengths by the analysis given in Appendix B:

$$L_{te}^2 G_e S_e = 1 = L_{tb}^2 G_b S_b \tag{C6}$$

As discussed in the reference, various methods have been used and proposed for measuring this transfer length. One of the most effective can be accomplished with a three-contact arrangement in which the penetration of the electric field under a contact stripe is measured. This method is not available if only two contacts are to be used. Consequently the scheme proposed here involves making a very small area of contact so as to exaggerate the resistance.

The structure for measuring the contact conductance or transfer length is represented in Fig. C4 with its relative position in respect to an interdigitated structure shown. As for Fig. C3 a situation is used in which additional tolerances are allowed over that involved in making the transistor structure. In Fig. 5 the tolerance has been approximately doubled so that the contact areas represented as two small circles can be displaced from their desired positions in respect to subsequent contact

window openings by more than the tolerance "a". The contact resistance is enhanced in the structure of Fig. 5 by making contact windows which are smaller in diameter than "a". Since these do not have to be registered the only requirement is the ability to cut the oxide so as to permit the contact to reach through to the diffused layer. For the cases involved the layer may be either the base layer or a large region of emitter diffusion centered about the neighborhood of the contacts.

In the analysis of the L_t structure of Fig. C4, we consider two terms to the resistance between the contact regions. For one of these we regard the contacts as being perfect and consider the sheet resistances. For this case the potential distribution V produced by current I is as represented in Fig. C5 so that the potential in the unplated region satisfies Laplace's equation:

$$V = (SI/2\pi) \ln (r_2/r_1) \tag{C7}$$

At the edges of the contacted region we should take the radii to be

$$r_2 \doteq c \qquad r_1 = d/2 \tag{C8}$$

This leads to a difference in potential between the peripheries of the two contact areas which in turn leads to a resistance given by

$$R(S) = (S/\pi) \ln 2\, c/d \tag{C9}$$

This resistance will be altered if the measurement is made on a diffused emitter area which does not extend infinitely far away in all directions. The actual area will be terminated at about 10a from the right hand point contact. This will mean in effect a small variation in the effective

center to center separation c and will slightly affect the logarithm term in Eq. (C9). As we shall see below in considering Eq. (C14) a small variation in the logarithm term is not important in estimating the value of the contact transfer lengths.

For the particular case represented in Fig. C4 the center to center distance is six times the minimum spacing "a" and the diameter is a/2. This leads to Eq. (C10)

$$c = 6a \quad , \quad d = a/2 \quad , \quad \ln 2c/d = \ln 24 = 3.17 \doteq \pi \qquad \text{(C10)}$$

so that for this case the contribution R(S) due to the sheet resistance can be taken to be practically equal to S

$$R(S) = 1.01\ S \doteq S \qquad \text{(C11)}$$

We now consider the contact resistance. It has been supposed, as discussed above, that d is made smaller than the minimum "a" so that we are concerned with transfer length problems only if the transfer length is comparable to d or larger. For this case the resistance of the metal to the diffused layer can simply be taken to be the reciprocal of the conductance G per unit area. This leads to a contact resistance for the two contacts in series given by

$$R(G) = 2/G\,\pi(d/2)^2 = (8/\pi)/Gd^2$$

$$= (8S/\pi)\,L_t^2/d^2 \qquad \text{(C12)}$$

The sensitivity of the measurement to L_t can be expressed as the radio of R(G) to R(S). This gives

$$\frac{R(G)}{R(S)} = \frac{8}{\ln(2c/d)} \frac{L_t^2}{d^2} \qquad \text{(C13)}$$

so that the relative importance of R(G) is seen to vary inversely as d^2 in the assumed range that L_t is greater than d. For $L_t = d$ and the dimensions of Eq. (C10), R(G) is 2.5 times larger than R(S). The measured resistance for the structure of Fig. C4 will be the sum of the contact and the sheet resistances, and this leads to

$$R(L_t) = (8S/\quad)\,(L_t/d)^2 + (S/\pi)\ln(2c/d) \qquad \text{(C14)}$$

In order to interpret the experimental data the measured value for $R(L_t)$ obtained by a measurement on the structure of Fig. C4 should be compared with the appropriate S value obtained from Fig. C1 or Fig. C2. It is seen that small errors in geometry which affect the logarithm term will not have an appreciable influence upon the measured value of L_t provided that the range of experimentation is such that L_t is comparable to or larger than the diameter d of the contact windows open in the G-structure mask of Fig. C4. For the particular case in which L_t is equal to the diameter d of the contact region, the ratio of the measured resistance $R(L_t)$ to the sheet conductance in ohms/square for the geometry of Fig. C4 is given by

$$L_t = d \qquad R(L_t)/S = 3.55 \qquad \text{(C15)}$$

For control purposes, L_t values of the order of d or less should be satisfactory if d is as small as a/2. Consequently ratios of 3 or less for $R(L_t)/S$ obtained by measurement of the L_t structure and the S_{bb} or S_e structures should mean that the processes are satisfactory.

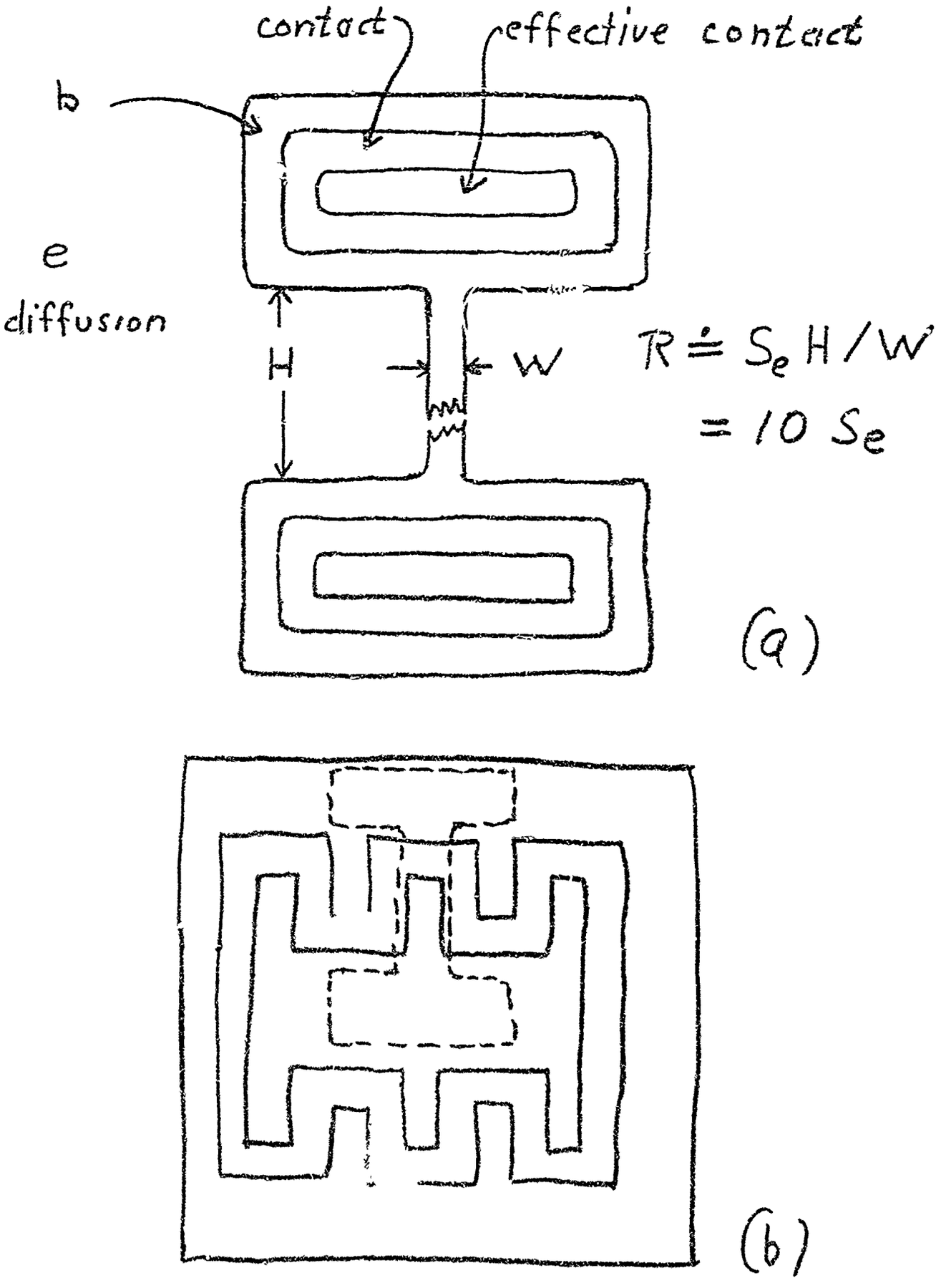
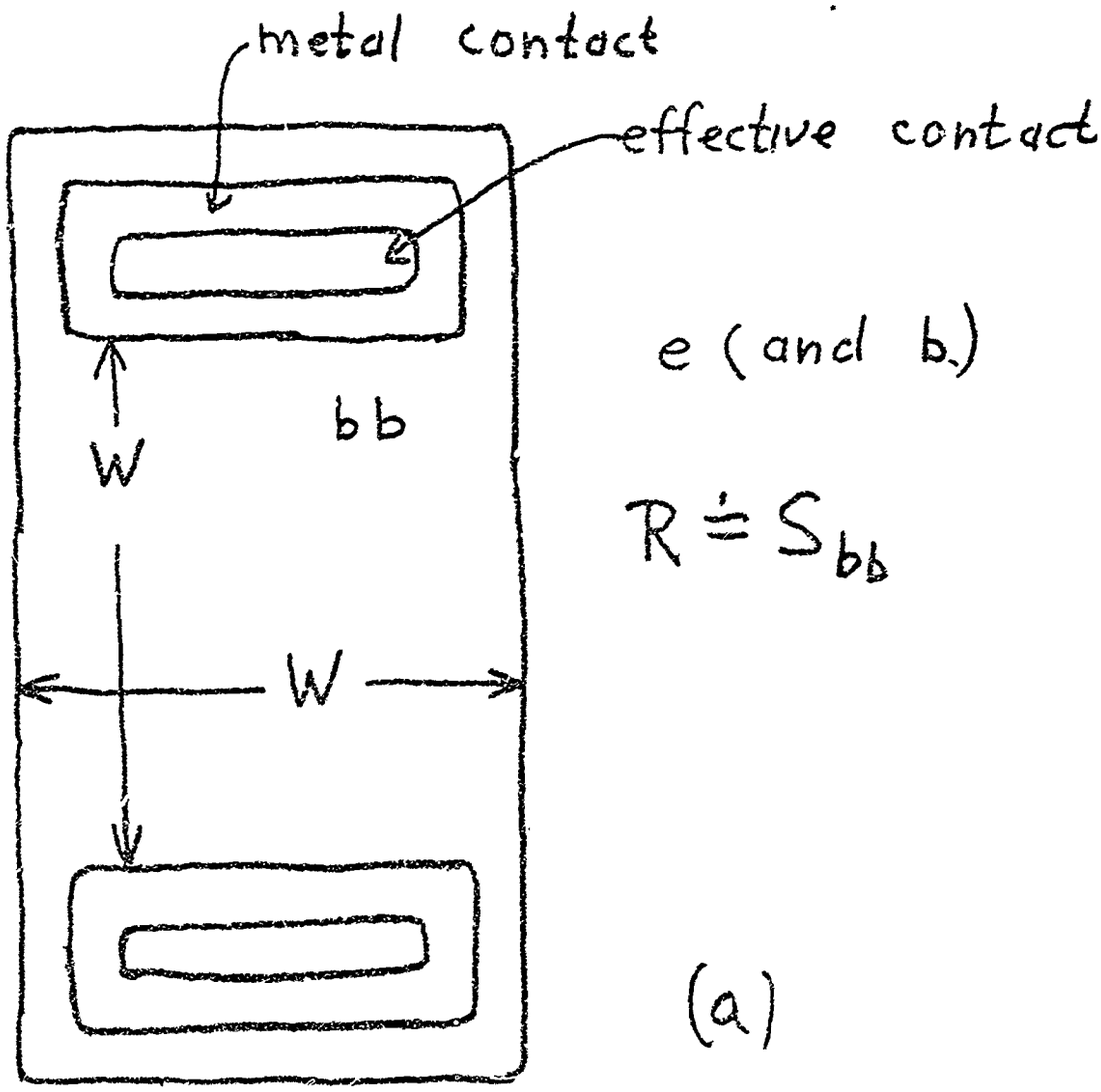


Fig. C1 The S_e structure.



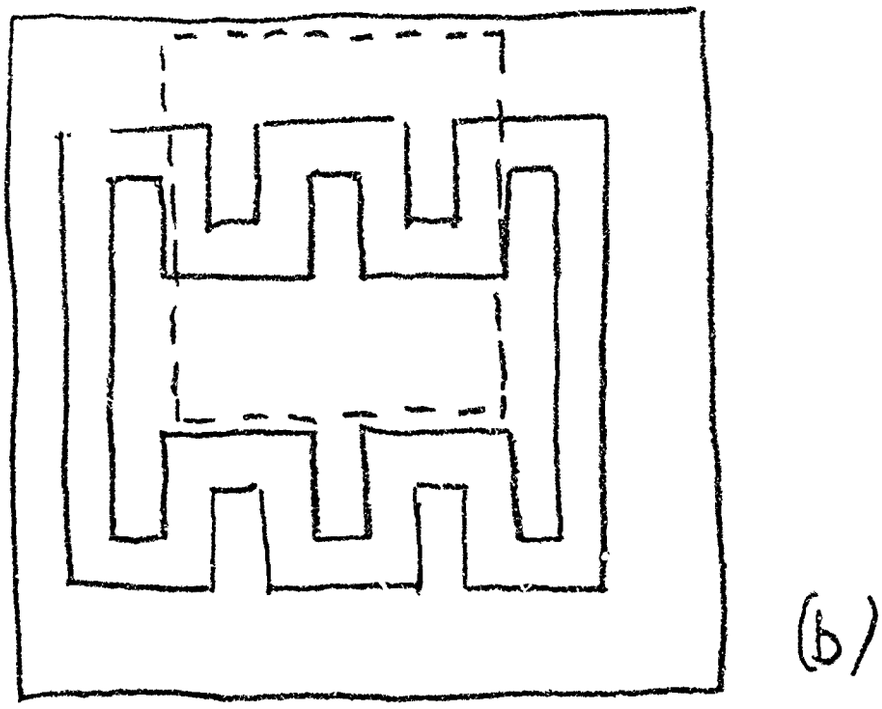


Fig. C2 The S_{bb} structure.

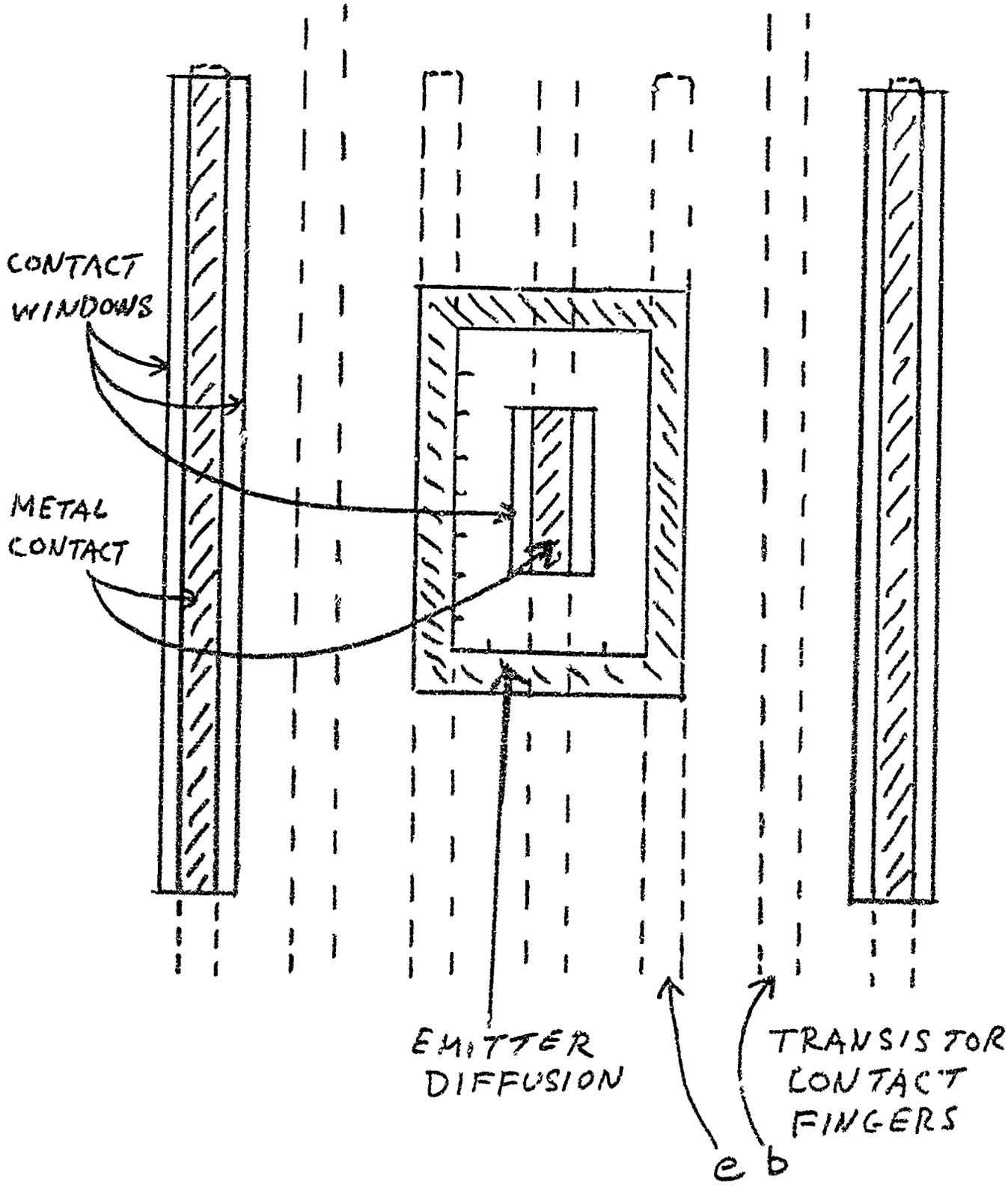


Fig. C3 The S_b structure.

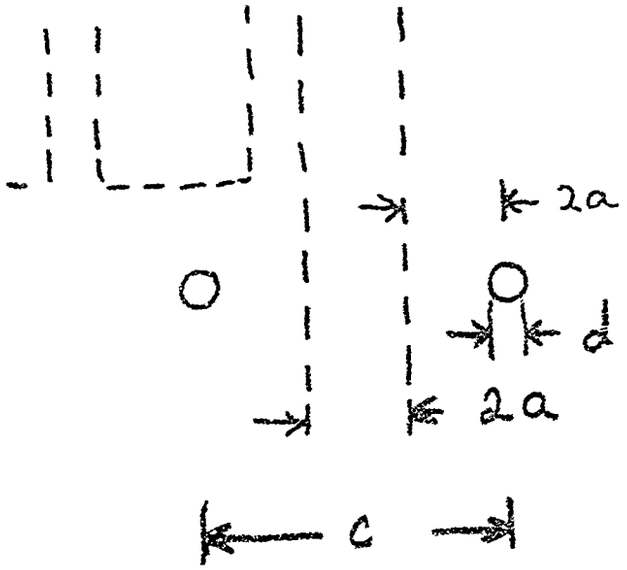
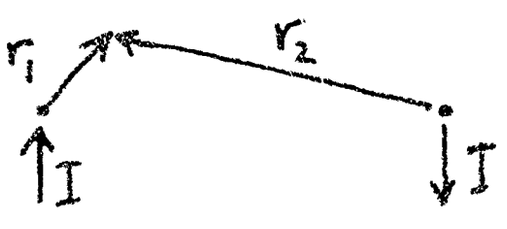


Fig. C4 The L_t structure.



$$V = (SI/2\pi) \ln r_2/r_1$$

Fig. C5 Potential function for two point sources.