Iowa State University
Digital Repository

Retrospective Theses and Dissertations

Iowa State University Capstones, Theses and Dissertations

1963

Secondary breakdown in transistors

Paul Riley Bond
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Electrical and Electronics Commons

Recommended Citation
Bond, Paul Riley, "Secondary breakdown in transistors " (1963). *Retrospective Theses and Dissertations*. 2335.
https://lib.dr.iastate.edu/rtd/2335

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

المنارة للاستشارات

www.manaraa.com

This dissertation has been 63–5169
microfilmed exactly as received

BOND, Paul Riley, 1930–
SECONDARY BREAKDOWN IN TRANSISTORS.

Iowa State University of Science and Technology
Ph.D.,1963
Engineering, electrical

University Microfilms, Inc., Ann Arbor, Michigan

SECONDARY BREAKDOWN IN TRANSISTORS

by

Paul Riley Bond

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Electrical Engineering

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1963

TABLE OF CONTENTS

	Page
INTRODUCTION	1
REVIEW OF THE LITERATURE	3
THEORETICAL CONSIDERATIONS	15
Thermal Effects	15
The Mechanism	26
EXPERIMENTAL RESULTS	35
Transistors	35
Diodes	50
SUMMARY AND CONCLUSIONS	52
LITERATURE CITED	54
ACKNOWLEDGMENT	56
APPENDIX: DEVELOPMENT OF ELECTRICAL ANALOG FOR HEAT FLOW PROBLEM	57

INTRODUCTION

Figure 1 shows an oscillogram of the swept collector characteristic of a transistor in the grounded-emitter connection, with the base open. The collector current is low until collector voltage reaches an apparent breakdown voltage, where the voltage will rise no higher (except for current-resistance drop in the connections) with increasing current. As current is increased further, a secondary breakdown occurs, and the voltage drops to a much lower value, and shows a negative dynamic resistance to any further increase in current. As the current decreases, the characteristic shows a hysteresis effect. A similar secondary breakdown is observed when the base is either forward or reverse biased.

The objective of this dissertation is to investigate this secondary breakdown, and to set forth a probable mechanism for this phenomenon. As a result of the investigation there is strong evidence that the secondary breakdown is almost entirely a thermal mechanism, and several previously proposed mechanisms were not observed.

The measurements described in this dissertation were undertaken to understand the secondary breakdown mechanism in germanium alloy-junction power transistors. However, the analysis developed should be equally applicable to other types and materials.

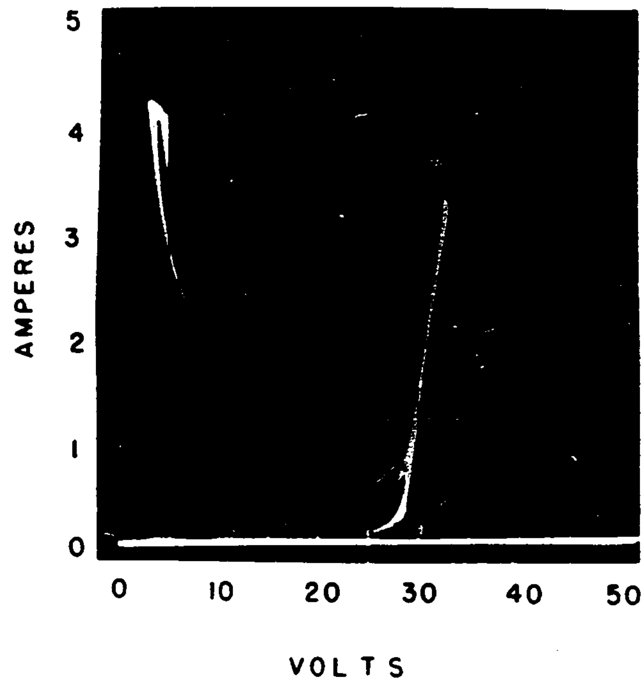

Figure 1. Swept collector characteristic

REVIEW OF THE LITERATURE

Two mechanisms for the primary breakdown of reverse-biased p-n junctions have been proposed. The first is the "Zener" breakdown, attributed to C. Zener (1), which postulates a spontaneous disruption of covalent bonds among atoms located in the immediate vicinity of the junction as a consequence of the high field that exists there. This is analogous to cold emission from metallic surfaces in the presence of a high field. The second explanation, termed "avalanche" breakdown by K. G. McKay and K. B. McAfee (2) and (3), proposes that the thermally generated electrons and ions are sufficiently accelerated in the high field to give rise to ionizing collisions. The secondary electrons may cause additional ionization products, and the ions may be accelerated to the cathode and produce secondary electron emission by bombardment. This is similar to the Townsend discharge in a gas. The accelerating potential at which the secondary effects become self-maintaining has been termed the "avalanche" voltage.

S. L. Miller (4) has shown that, in germanium transistors at least, the primary breakdown must be due to the avalanche mechanism. The proof follows from the observed dependence of the breakdown voltage on resistivity and barrier width. Practical considerations are such that the avalanche breakdown

occurs at a lower voltage than does the Zener breakdown in germanium transistors (and most other semi-conductors).

The secondary breakdown in the swept collector characteristic of p-n-p junction transistors was first reported by C. G. Thornton and C. D. Simmons (5). The collector volt-ampere characteristic of Figure 1 was studied, with the critical current at which the voltage drop takes place designated as I_m. I_m was observed to decrease for greater reverse bias on base-emitter junction. I_m was observed to decrease with lower sweep frequency. I_m was not observed to depend on temperature. Four possible results of repeated sweeping of the collector characteristic were reported:

1. Figure 1 may be completely stable and reproducible.
2. The avalanche voltage may drop to a lower value and then become stable. This was thought to be a surface effect, since re-etching often restores the unit.
3. I_m may progressively decrease to a lower current on each trace until I_m stabilizes at a lower value, then unit is stable. The original condition cannot be recovered.
4. The unit may suddenly develop a collector-to-emitter short, although the individual collector and emitter diode are still good.

Their microscopic examination of the base region of units which had developed collector-emitter shorts revealed a hole penetrating the base near the center. The appearance of the hole suggested melting due to localized heating.

Thornton and Simmons advanced the "pinch-in" effect as a possible mechanism for secondary breakdown. A radial potential gradient in the base would concentrate the current in the center of the transistor. An excessive concentration of mobile charge carriers would affect the resistivity of the base, and the depletion layer would shrink in width. Since the voltage remains constant, the field will increase. The secondary breakdown occurs when a certain critical field is reached. Because of the localized current path, destructive heating may take place.

The manner in which the radial potential gradient in the base might arise was shown with the aid of Figure 2. Suppose the emitter junction is biased off and reverse collector voltage is raised until avalanche is reached. Electron current in the base will flow outward radially so that a voltage gradient will exist in the base tending to forward bias the center of the transistor.

Thornton and Simmons did not mention secondary breakdown for zero or forward base current. Indeed, H. C. Lin, A. R. Hlavacek, and B. H. White (6) state that secondary breakdown does not occur with open base. They propose that the high

reverse base current described above may be due to p-n-p-n action. At high current levels the collector contact may form an injector, so that a transistor becomes a four-layer device. J. L. Moll, M. Tanenbaum, J. M. Goldey, and N. Holonyak (7) describe the action of such devices. In the four-layer device as shown in Figure 3, there are effectively two transistors. The upper n-p-n transistor, when biased as in Figure 2, has terminal b as its collector, and this collector current is the reverse base current that causes the "pinch-in" according to these writers. When the current is increased until the sum of the two current amplification factors equals one, the unit suddenly breaks over into the low-resistance mode.

H. A. Schafft and J. C. French (8), in a complete review of the subject, refuted both the "pinch-in" effect and p-n-p-n action as an explanation of secondary breakdown. Their observations indicated that every transistor with leads capable of carrying the required current could be driven into secondary breakdown. This includes p-n-p and n-p-n polarities, alloy-junction, double-diffused, and drift resistivities, and planar, disk, mesa, and epitaxial geometries, using both germanium and silicon. This indicates that secondary breakdown is a characteristic behavior of the transistor structure. Furthermore, in every transistor examined, secondary breakdown occurred not only for reverse

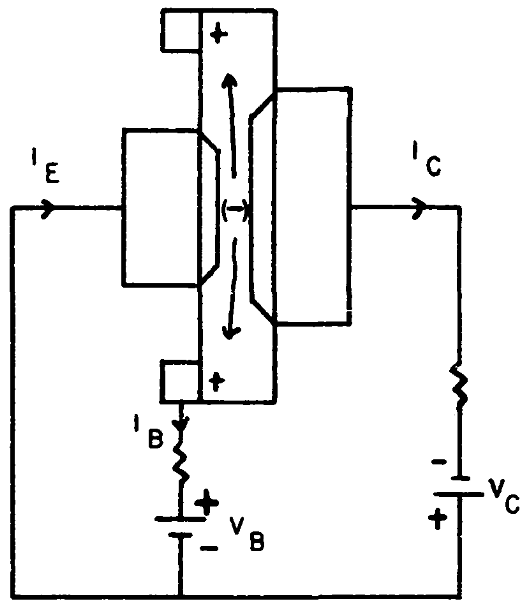

Figure 2. Voltage gradient in base for negative base current

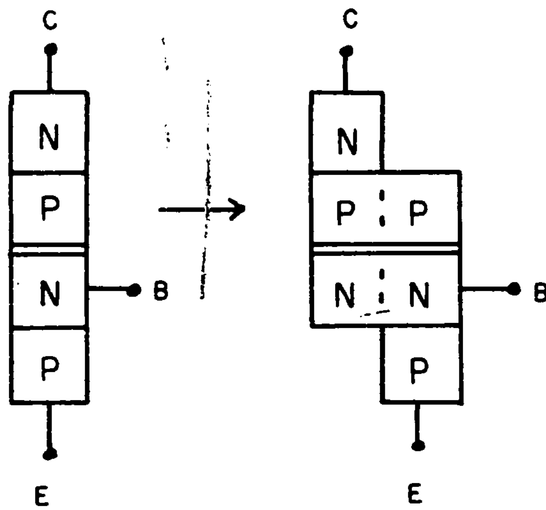

Figure 3. P-N-P-N structure

base drive, but also open base and forward base drive. This contradicts the previously cited work, and destroys the argument for the "pinch-in" effect. The "pinch-in" effect described earlier for reverse-bias conditions would not be expected to be important at open-base conditions, where potential gradients due to radial base currents are minimal. Yet secondary breakdown was found under these conditions, and also under forward-bias conditions, where the current would be expected to concentrate near the emitter edge. The site at which short-circuit failure due to secondary breakdown occurred continued to be located near the center of the base, even with forward bias. The dependence of I_m was monotonic throughout the range of base currents.

I_m was found to be a function of temperature, which conflicts with previous findings. I_m was also found to vary with the parameters of the circuit for obtaining the swept characteristic. For this reason these authors decided it was misleading to characterize secondary breakdown by the single parameter I_m without a complete understanding of the factors which affect I_m and a specification of the conditions of measurement.

Another method of measurement employed by Schafft and French was as follows: with the emitter grounded, and the desired base current flowing, a high dc voltage was applied through a current limiting-resistance to the collector.

Secondary breakdown would occur after a certain delay time if the operating current was sufficient. It was found that the delay time was a function of operating current, ambient temperature, and base drive in a way which could generally be correlated qualitatively with the behavior of the swept secondary breakdown characteristic.

Figures 4 and 5 illustrate the dependence of delay time on current, temperature and base drive as observed by Schafft and French.

It was also observed that the energy dissipated in the transistor before secondary breakdown was initiated exhibited the same kind of dependence on temperature, operating current, and base drive as did the delay time. For this reason some sort of thermal phenomenon was suggested, although the thermal time constants of these transistors appeared to be orders of magnitude greater than observed delay times. The authors concluded that some critical energy must be dissipated in the transistor before secondary breakdown occurs.

From this assumption it is immediately clear why I_m should be a function of sweep rate and sweep circuit parameters, and also how a transistor can be protected from secondary breakdown and its associated failure modes. If a minimum energy must be dissipated in a transistor before secondary breakdown occurs, then as long as the energy dissipated is less than this the transistor is safe.

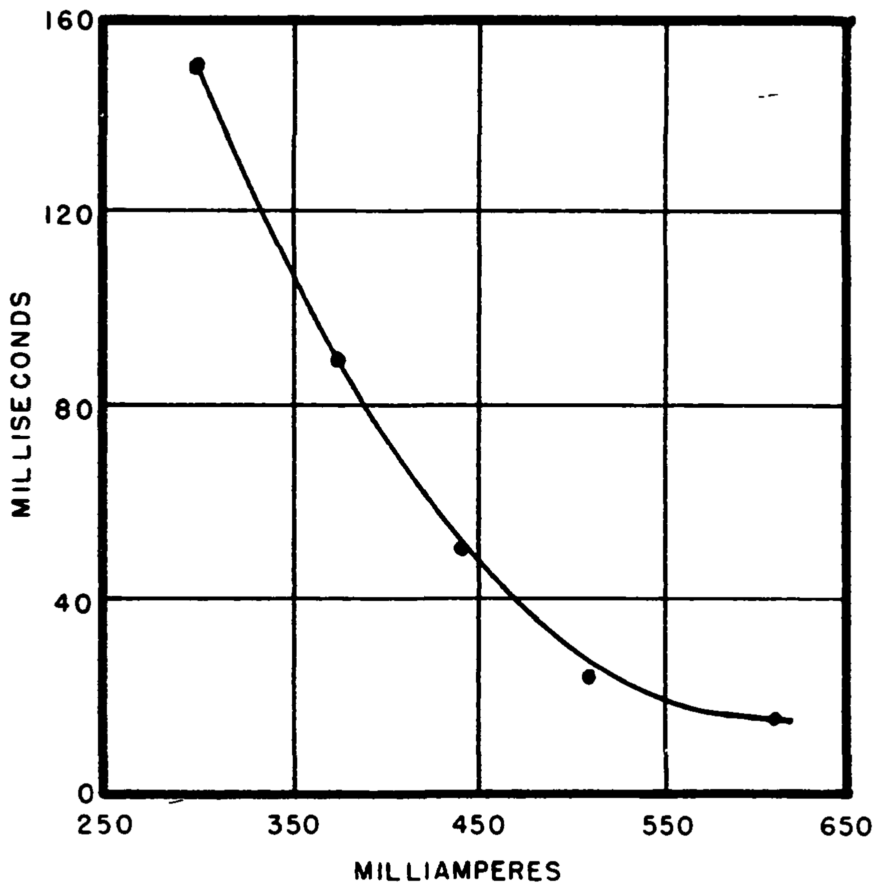

Figure 4. Dependence of delay time on collector current for a 30-watt p-n-p transistor with base floating and a case temperature of 25°C

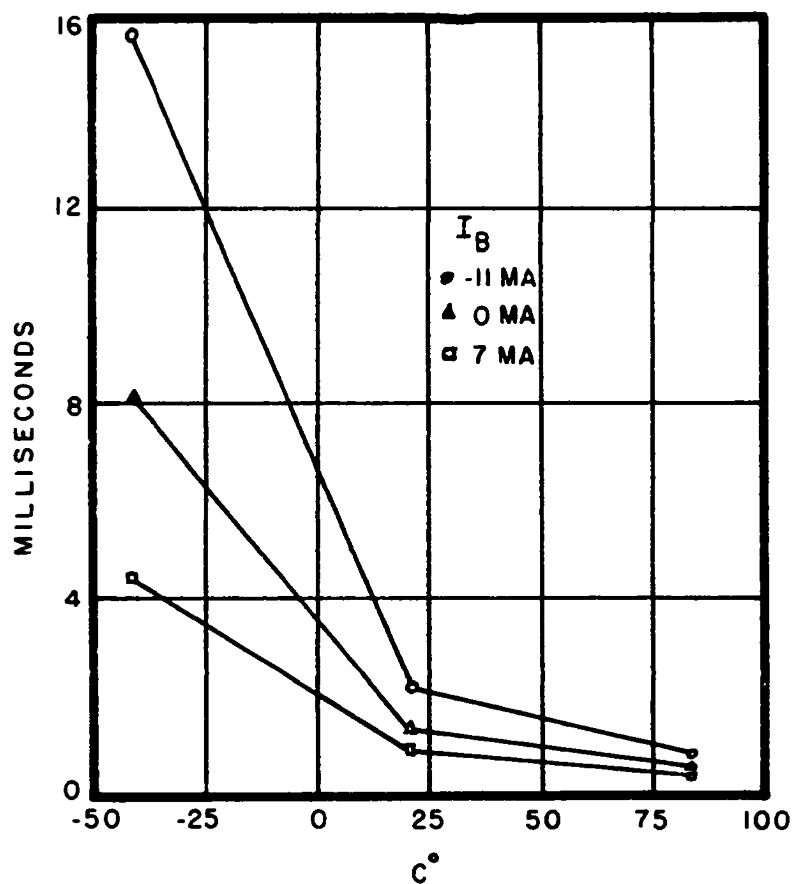

Figure 5. Dependence of delay time on ambient temperature at three base drives with a constant collector current of 2.8 amperes for a 40-watt p-n-p transistor

A close look was taken at transistors which had failed due to collector-to-emitter shorts. Typically the short developed in a region a few mils in diameter where the emitter and collector regions had alloyed together. Transistor action could be restored by drilling out the region around the hole. The holes always appeared near the center of the base region, even though the base drive was varied widely. It appears that the melting was caused by a constriction of current to a narrow channel as was suggested earlier.

Schafft and French also rejected the suggestion that secondary breakdown might be explained by p-n-p-n action. Their reasoning will be set forth here, since many current textbooks make no distinction between secondary breakdown and p-n-p-n action.

1. The theory of the four-layer structure predicts that in the "on" condition the device will behave like a forward-biased diode (7). Thus a constant positive dynamic resistance is expected in the low voltage-mode. But reference to Figure 1 will show the dynamic resistance to be negative and decreasing in the low-voltage mode.
2. The voltage drop in the low-voltage mode in secondary breakdown can range from 5 volts to 20 volts depending on device and current. This is considerably higher than is expected for a p-n-p-n device.

3. Positive base current flowing into the base of a p-n-p-n device (Figure 3) will force the device to a higher current before the collector junction will be forward biased. But for a p-n-p transistor, I_m will decrease if the base current is changed this way.
4. Time needed to switch a p-n-p-n device is of the order of the transit time of the injected carriers that are needed to forward bias the collector junction. Under the most adverse conditions the switch-on time should be about one microsecond. This is several orders of magnitude less than delay times that have been observed for secondary breakdown.
5. Secondary breakdown is observed in transistors whose physical structure appears to prevent p-n-p-n action, even if the collector contact should inject minority carriers.

Since secondary breakdown as a surface phenomenon has already been considered and rejected (5), no adequate explanation for secondary breakdown exists at present.

THEORETICAL CONSIDERATIONS

Thermal Effects

After experimental verification of the cited work, it was decided to proceed on the assumption that the secondary breakdown mechanism is entirely thermal. Certainly heating takes place due to the dissipation which occurs in the transistor after primary (avalanche) breakdown. Once the geometry of the semi-conductor is specified, the thermal resistance of the material known, and the heating calculated, the temperatures within the unit may be predicted. The effect of temperature on the physical processes within the structure must, therefore, be examined.

The Fermi-Dirac statistical model predicts that the number of electrons in the conduction band and the number of holes in the valence band of a crystal are given by the following expressions, subject to the restriction that E_f, the Fermi energy level, is separated several kT from the bottom of the conduction band or top of valence band.

$$n = A\,(kT)\;e^{-\frac{(E_c - E_f)}{kT}} \quad \text{per unit volume} \qquad (1)$$

$$p = B\,(kT)\;e^{-\frac{E_f - E_v}{kT}} \quad \text{per unit volume} \qquad (2)$$

where n is number of free electrons, p is number of holes, A and B are constants of the crystal, k is Boltzmann's constant, T is absolute temperature, E_c is energy level of the bottom of the conduction band, and E_v is the energy level of the top of the valence band. The product of hole and conduction-electron densities

$$pn = KT^3 \; e^{-\frac{E_g}{kT}} \; \text{volume}^{-2} \qquad (3)$$

is independent of E_f and dependent only on temperature and the gap energy E_g. This energy gap changes with temperature because of lattice expansion and for germanium at temperatures above 50° K can be approximated by the expression (9)

$$E_g = .782 - .00039T \;\; \text{electron volts} \qquad (4)$$

Combining Equations 3 and 4 and inserting the numerical value of K from measured data we obtain

$$pn = 3.1 \times 10^{32} \; T^3 e^{-\frac{9100}{T}} \; cm^{-6} \qquad (5)$$

for germanium. For intrinsic germanium the number of holes and conductor-electrons are equal so that

$$p = n = 1.76 \times 10^{16} \; T^{3/2} \; e^{-\frac{4550}{T}} \; cm^{-3} \qquad (6)$$

The conductivity of the semiconductor will be

$$\sigma = p\mu_p q + n\mu_n q \tag{7}$$

where μ_p and μ_n are mobilities of the hole and electron, respectively. This factor indicates the ease with which the carrier drifts under the influence of an electric field. Mobility is proportional to mean free path divided by mean thermal velocity. Both of these factors vary with temperature in such a way that mobility falls with temperature. Empirical approximations of Morin and Maita (9) and Prince (10), good in the range 100-300° K, give for pure germanium

$$\mu_n = 4.9 \times 10^7 \; T^{-1.66} \quad cm^2 \,/\, \text{volt-sec} \tag{8}$$

$$\mu_p = 1.05 \times 10^9 \; T^{-2.33} \quad cm^2 \,/\, \text{volt-sec} \tag{9}$$

The above relationships will enable the calculation of the conductivity of intrinsic germanium, but for conductivity of doped crystals the impurity concentration must be specified. The ionization energy required to produce a free carrier from the impurity atoms is much lower than that of germanium, so that activation of impurity carriers is complete at low temperatures before thermal generation of electron-hole pairs becomes appreciable. For example, ionization energy in germanium of indium (valence 3) is .0112

electron volts and of arsenic (valence 5) is .0127 electron volts.

The mobility of carriers is also affected by the presence of impurity atoms. Collisions of electrons with lattice atoms out of their equilibrium positions because of thermal vibration would be expected to occur more often at higher temperatures. Thus mobility will drop with temperature rise as indicated in Equations 8 and 9 when thermal vibrations dominate the scattering process. But the scattering effect of charged impurity atoms is different because the deflecting effectiveness of the charged ion is greater, the lower the speed of the approaching charged carrier. The concentration of impurities and the temperature both affect mobility of doped crystals then, and analytic prediction becomes difficult. Much experimental data is available in the literature; for example, see Adler and Longini (11, p. 134).

An example of the effect of temperature on the conductivity of n-type germanium is shown in Figure 6. Several interesting points can be observed. The first positive slope indicates that impurity carriers are being activiated as thermal energy reaches ionization level. Then, as most of the impurities are activated, conductivity drops because of decreasing mobility. Finally, conductivity rises because hole-conduction electron pairs are being generated thermally. The change in slope occurs at 380° K. The heavy line at the

right is a plot of Equation 7, using Equations 6, 8, and 9. It is apparent that the semiconductor is intrinsic above 410° K. (Conduction is almost entirely due to thermally generated hole-electron pairs, so that p = n). These are the important facts to be established by the preceding discussion.

Figure 7(a) shows conditions in a p-n-p transistor. Abrupt junctions are assumed, and the doping level of the p material is assumed much higher than that of the n material. This corresponds to a realistic situation in alloy-junction power transistors. The Fermi level, which is defined as the energy level at which the probability is one-half that an available state at that level is occupied, is different for the two types of material. For n material the Fermi level is slightly above the center of the energy gap because there are a number of available states for the electrons near the conduction band contributed by the donor atoms. For p material the Fermi level is farther below the center of the energy gap because of a larger number of available states down near the valence band, contributed by the larger number of acceptor atoms.

When the two materials are brought into contact, the Fermi levels will coincide if the material is in thermal equilibrium, but the other energy levels adjust to account for any flow of charge that occurs. Some electrons move from the

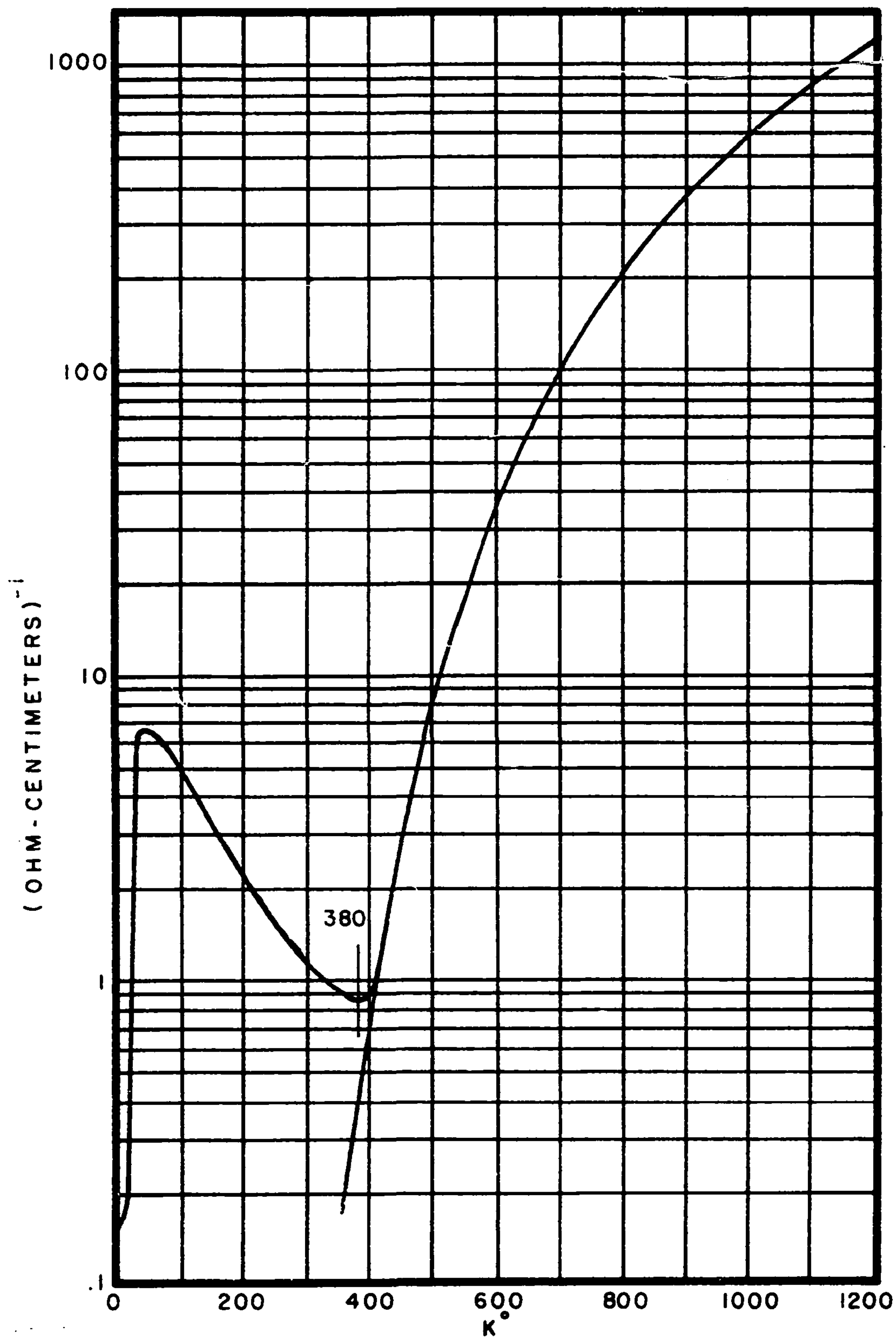

Figure 6. Conductivity of germanium crystal doped with 10^{15} arsenic atoms per cubic centimeter

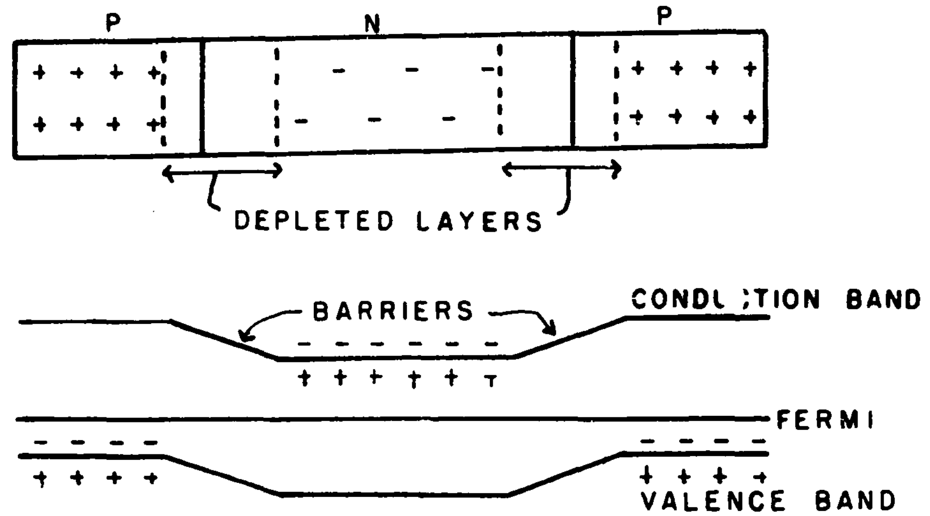

Figure 7(a). Energy level diagram for unbiased p-n-p junctions

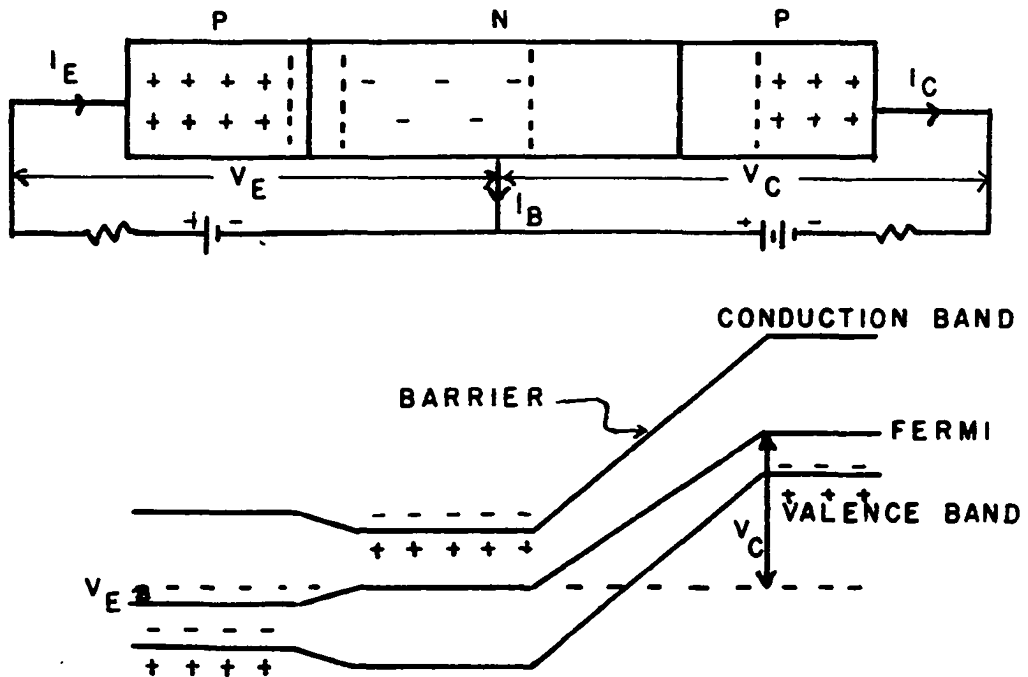

Figure 7(b). Energy level diagram for biased p-n-p junctions

n-type material to the p-type material, leaving the "uncovered" positive charge on the donor atoms in the n-type material and, by filling holes, cause some of the acceptor atoms in the p-type material to take on a negative charge. The change in energy levels creates a barrier. This barrier as seen from the n-type material is a barrier to the motion of electrons in the conduction band into the p-type material, and when viewed from the p-type material represents a barrier to the flow of holes in the valence band into the n-type material. The barrier extends farther into the material with lowest impurity concentration.

Application of external voltage to the junction may encourage or discourage this "uncovering" of charges, depending upon the polarity. Figure 7(b) indicates an application of a negative voltage at the collector with respect to the base, and a positive voltage on the emitter. Note the change in barrier heights. The emitter junction will be "forward" biased, and the collector junction will be "reverse" biased, with almost all of the voltage drop appearing across the collector junction. I_c will consist of thermally generated holes from the base region diffusing to the collector barrier and swept across to the collector barrier. A small fraction of the injected holes recombine with electrons in the base and so constitute an electronic base current.

$$I_c = \alpha I_e + I_{co} \tag{10}$$

$$I_b = (1 - \alpha)\, I_e + I_{co} \tag{11}$$

From Figure 7(b)

$$I_e = I_c + I_b \tag{12}$$

Combining Equations 10 and 12, we obtain

$$I_c = \frac{\alpha}{1 - \alpha} \quad I_b + \frac{I_{co}}{1 - \alpha} \tag{13}$$

From Equation 13 it is obvious that there is a current through the collector junction even with the base open ($I_b = 0$), so secondary breakdown under this simplified condition will be examined first. The factor α in the above equations takes into account the injection efficiency of the emitter and the transport efficiency of the base. There is another term, called collector efficiency, which takes into account the multiplication of carriers during their transit of the high field region in the depletion layer of the reverse biased collector junction.

Of great importance to this discussion is the effect of collector multiplication at high collector voltages.

Equation 10 must be rewritten

$$I_c = \frac{\alpha\ I_e + I_{co}}{1 - (V_c/V_a)^n} \qquad (14)$$

to account for the variation in collector efficiency with collector voltage. The denominator in Equation 14 is an empirical expression developed by Miller (4) where V_c is collector voltage, V_a is the avalanche voltage of the collector junction, and the exponent n should be approximately three for p-n-p germanium types of transistors. Equations 12 and 14 can be combined to give

$$I_c = \frac{\alpha\ I_b + I_{co}}{1 - \alpha - (V_c/V_a)^n} \qquad (15)$$

which indicates that an apparent breakdown will occur when

$$(V_c/V_a)^n = 1 - \alpha \qquad (16)$$

This means that an apparent avalanche breakdown will occur at a collector voltage much lower than the collector junction avalanche voltage when the transistor is connected grounded emitter. Refer to Figure 9.

The apparent breakdown voltage is a direct function of the avalanche voltage, however. Miller (4) has also shown that the avalanche voltage is related to the resistivity ρ of the depletion layer by an empirical expression

$$V_a = 87\rho^{.725} \tag{17}$$

We have already seen how the resistivity varies with temperature, so we can predict the variation of avalanche voltage with temperature. For the germanium sample shown in Figure 6, avalanche voltage will increase in the range 50 to 380° K, and then decrease with temperature. A large amount of data was taken to verify this statement.

This voltage drop across the depletion layer, multiplied by the collector current, gives rise to a power dissipation in the depletion layer that can only result in heating. Given the initial temperature distribution, the thermal properties of the materials, the geometry and dimensions, and the distribution and rate of heat generation, the temperature distribution in the base region as a function of time can be found.

Figure 6 gives us an idea of how the resistivity of the base material changes with temperature, and it is apparent that the increased thermal generation of hole-electron pairs has a drastic effect on the ratio of electron carriers to hole carriers in the material. Indeed the thermally

generated carriers so dominate the impurity carriers that the material is intrinsic (p = n). This affects the Fermi level, since it must be midway between the bottom of the conduction band and the top of the valence band for this condition.

The Mechanism

The mechanism proposed to account for the secondary breakdown proceeds as follows. Assume a transistor with open base, and a high reverse voltage applied to the collector through a current-limiting resistance. For high currents the voltage drop across the collector junction will approach the apparent breakdown voltage described above. The heating in the depleted region will raise the temperature, but the temperature will be uniform throughout the depletion layer because of a "negative feedback" situation. Should the current tend to concentrate, the relative temperature will rise at this point, raising the avalanche voltage of the junction at the point, which in turn will raise the voltage drop across the layer and force the current elsewhere.

As the temperature of the depletion zone rises, then, the entire layer will be constrained to be of uniform temperature until that critical temperature is reached where the temperature dependence of resistivity and avalanche voltage changes sign. At that point the feedback will become positive, and the constraint of uniform temperature in depletion layer

will be relaxed. Geometry of the transistor will dictate where the temperature will be greatest, and the current will rapidly concentrate there. The time required for this thermal "pinch-in" to take place is expected to be negligibly small compared to the total heating cycle because of the regenerative effect.

At high temperatures, where the semiconductor is basically intrinsic, the avalanche voltage drops to such a low value as to be of small magnitude compared to the current-resistance drop in the material which is effectively in series with any such barrier. Some typical values for avalanche voltage, calculated from Equation 17 and Figure 6, are 5 volts at 400° C and 1.3 volts at 900° C. The actual voltage drops, then, are reduced from this value by carrier multiplication. Localized heating of the depletion layer evidently creates a small region where the voltage drop is much less than the surrounding area. This "punch-through" establishes a low resistance-path through the base, and the voltage drops. This is the secondary breakdown. The geometry of the transistor is such that the temperature rises most rapidly at the center of the heated area, so that it is not surprising that the punch-through usually occurs near the center. In addition, the fabrication process of alloy-junction, diffused-junction, and other types of transistors,

is such that the base region tends to be thinner at the center if the junctions are not absolutely parallel.

At the point where punch-through occurs, the current is constricted to a narrow channel through the base region. The thermal time constant of this narrow channel is extremely short, so that an equilibrium is immediately established between power dissipated in this path and the temperature (resistivity). The voltage drop will be controlled by the current permitted by the external circuit parameters, and the semiconductor will exhibit negative incremental resistance. This negative resistance can be demonstrated with the aid of Figure 6. Assume temporarily that the resistance is constant when current increases. Then the power dissipated will increase, and temperature will rise. But Figure 6 shows that resistivity, and thus resistance, must decrease with a temperature increase in the range above 380° K.

After the switch to the low voltage mode, the total power dissipation will decrease, but since the current is constricted to a narrow channel, the temperature of this narrow path may jump to a high value, depending on the current. The temperature and elapsed time in this condition will determine whether or not the characteristics of the transistor are permanently changed.

The melting point of germanium is 937° C, so this temperature will not be reached without an immediate fusion

together of the emitter and base regions. The melting point of indium, the most widely used acceptor material, is 156° C. At any temperature much higher than this the molten indium from the p material will attack the germanium in the base region and form a liquid germanium solution. The germanium continues to dissolve into the indium until the saturation limit appropriate to the given temperature is reached. Figure 8 gives the equilibrium concentration of liquid indium-germanium solutions in contact with solid germanium at the temperature indicated (12, p. 7.17). It is seen that the saturation solution of indium-germanium at 500° C contains about 10 per cent germanium. The result of temperatures somewhat less than 937° C in the narrow current path through the base region will be a path of acceptor atoms extending in from both the emitter and the collector junctions. If the heating continues at high temperatures for a sufficient length of time, the saturated solution will extend completely across the base. This will either lower the doping level of the n-type material or change it to p-type material, depending upon the relative concentrations. But the characteristics of the transistor would be permanently changed either way. The failure of the transistor due to a collector-emitter short, with the base-emitter and base-collector diodes still operative is an example of this unwanted alloying process.

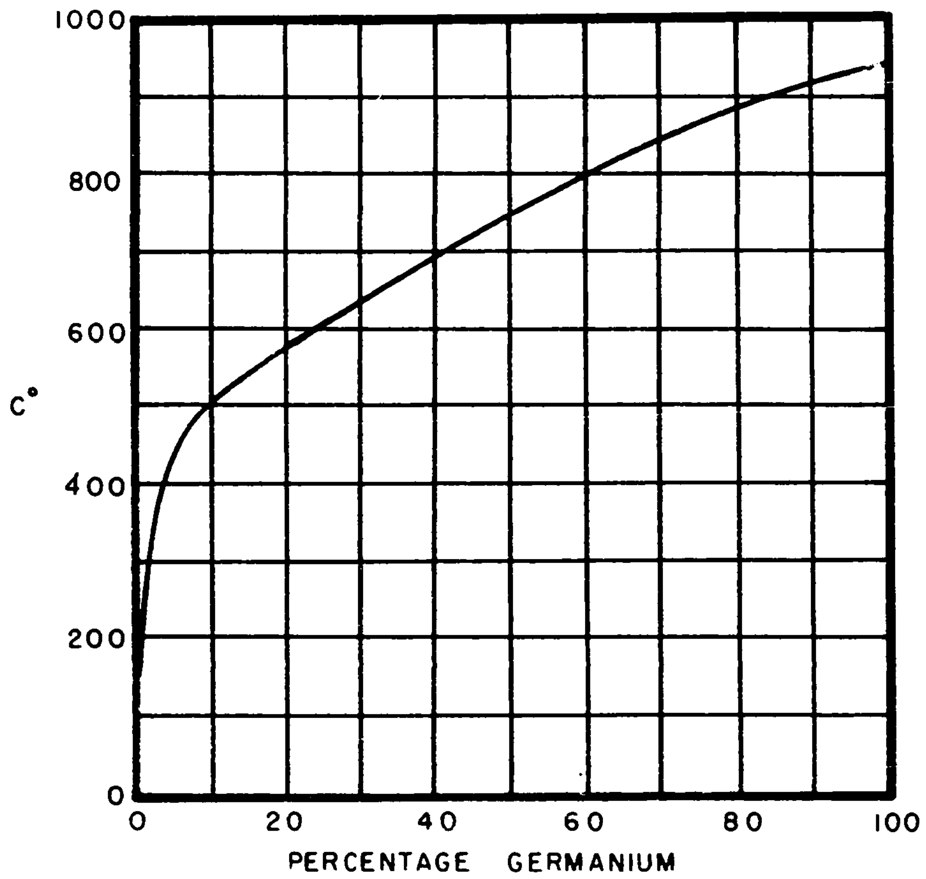

Figure 8. Phase diagram of the In-Ge system

If the heating process is interrupted before the short develops, the base width may have been narrowed at the site of secondary-breakdown current without alloying completely across. Such an event might not affect normal transistor operation at all, but would facilitate secondary breakdown the next time conditions were favorable. Such a change in transistor characteristics has also been observed in the literature (5).

Solid state diffusion of p-type impurities into the hot region of the base is also a possibility. But the depth of penetration is limited and it is a relatively slow process. It takes 3 hours at 575° C for indium to penetrate germanium to a depth of 800 angstroms with an impurity concentration of 10^{14} atoms/cm^3 (12, p. 7.14).

The mechanism described above for secondary breakdown for transistors with open-base connection can be extended to cover the cases of either forward or reverse base current. Reference to Equation 15 will reveal the dependence of voltage across the collector junction on base current for any particular collector current. Figure 9 illustrates the relation of collector characteristic to base current in the region of collector multiplication.

For constant collector current, the collector voltage is greater for reverse base current, and less for forward base current than that value expected for open-base condition.

Thus the power dissipated in the depletion layer increases for larger reverse base current, and the secondary breakdown should occur sooner. In like manner, the secondary breakdown should occur later for forward base current. This concurs with the monotonic dependence of delay time on base drive observed in the literature (8). The "pinch-in" effect of the base current does not seem to be important for typical transistor parameters.

Initial temperature distribution in the transistor, and the temperature of the heat sink are important in determining the time required for critical temperatures to be attained, and secondary breakdown initiated. Certainly higher initial temperature, and higher ambient temperature will decrease the delay before thermal breakdown.

Solution of the diffusion equation

$$cp \frac{\partial U}{\partial t} = k \left[\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} \right] + \Theta \qquad (18)$$

would give us the temperature U as a function of space co-ordinates and time. The rate of internal generation of heat Θ is also a function of space co-ordinates and time. The density of the material is p, c is the specific heat, and k is the thermal conductivity. Analytic solution is rendered difficult because the equation is not homogeneous, and the boundary conditions are very cumbersome expressions.

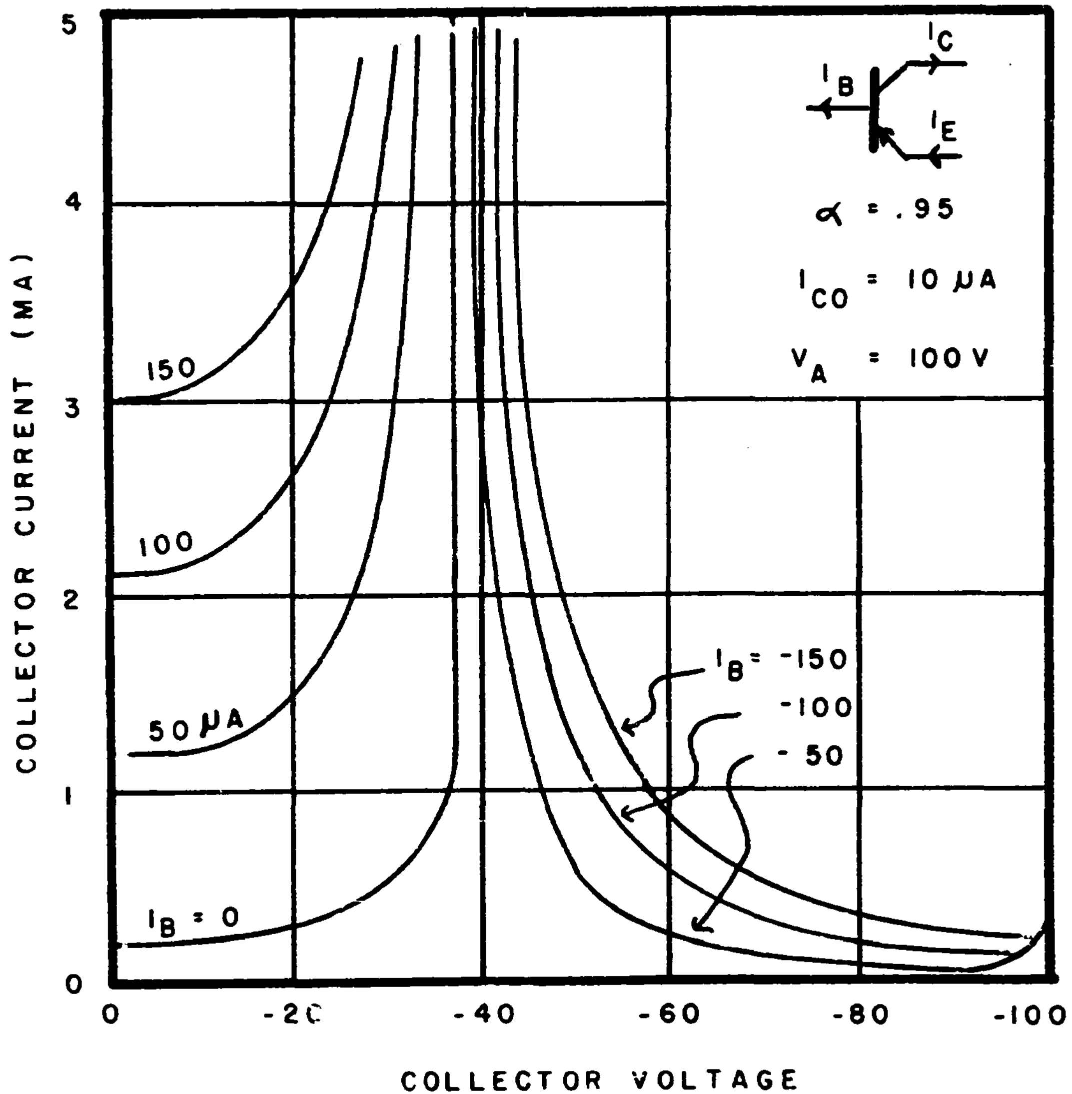

Figure 9. The collector characteristic families extended into the region of collector multiplication

An electrical analog for the thermal situation would facilitate the solution however.

The mechanism proposed here for the secondary breakdown of transistors would seem to indicate that this same secondary breakdown could occur in a reverse-biased p-n junction diode also. In the case of a diode, the depletion layer extends farther into the region of lowest impurity concentration (as in the transistor). High currents due to avalanche breakdown would heat the depleted zone at uniform temperature until resistivity changes slope. At that point the regenerative condition develops where increasing temperature lowers avalanche voltage, and lower voltage causes increased concentration of current with higher local temperature. The barrier will be punctured, a low-resistance path established, and the diode will switch to the low-voltage mode. Such a phenomenon has not been reported in the literature.

It should be pointed out that most semiconductor diodes are constructed of silicon to reduce thermal effects at moderate temperatures. The temperature required for a 5 ohm-cm sample of silicon to become essentially intrinsic would be near 1200° K. At this temperature other effects would be important; for instance, the contacts may melt, so the secondary breakdown is not always easily recognized.

EXPERIMENTAL RESULTS

Transistors

The next phase of the investigation was an effort to determine whether calculations made using the proposed mechanism for secondary breakdown were substantiated by experimental data. The units tested were germanium alloy-junction power transistors, in order that the physical dimensions and geometry might be determined with greater accuracy than would be possible for smaller devices. These units have the additional advantages of widespread application and low cost.

The experimental results were obtained as follows: the transistors were connected to a large heat sink and submerged in a water bath which could be maintained at the desired temperature. With the emitter grounded, and the desired base current flowing, a step pulse of current was applied to the collector. The collector supply was effectively a constant-current source, being obtained from a high direct-current voltage source through a current-limiting resistance. The time delay from the application of the current pulse until the secondary breakdown of the transistor was obtained with the aid of an oscilloscope which was triggered at the initiation of the pulse. Figure 10 is an example, showing a delay time of 93 milliseconds. The initial avalanche voltage was 28 volts, rising slightly with temperature, and the voltage after

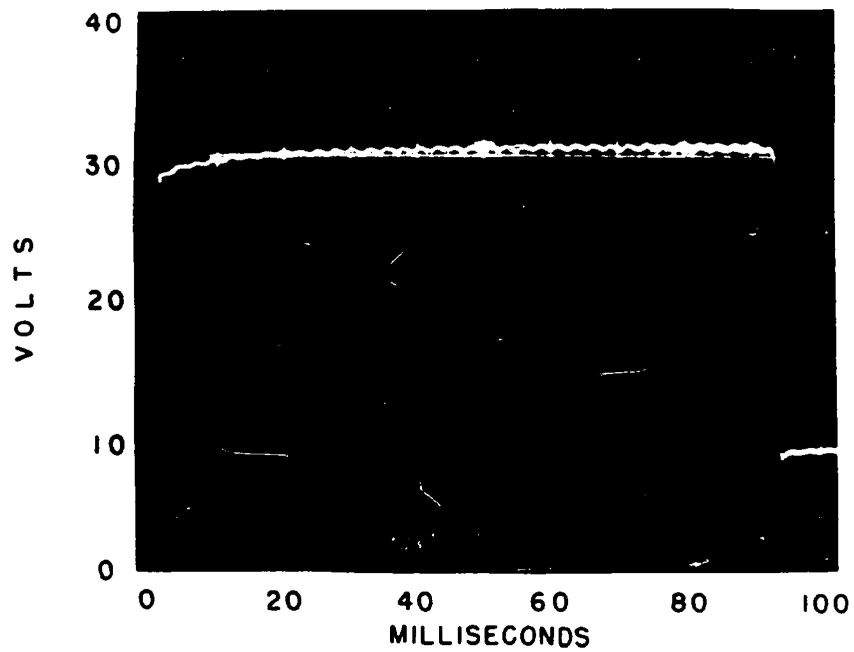

Figure 10. Collector voltage, showing secondary breakdown

breakdown about 9 volts. The ripple observed is a result of inadequate filtering of the power source.

As soon as possible after the secondary breakdown occurred, the current was removed to avoid permanent alteration of the transistor characteristics. Delay times were recorded for different values of collector current, base current, and heat sink temperature.

After electrical measurements were completed, the transistors were dismantled and physical geometry and dimensions determined.

The procedure for one of the transistors is reproduced here as an example, and then results of all tests are tabulated.

Table 1. Measured and calculated delay times of transistor A under specified conditions

Temperature	I_b	I_c	V_c	Time (observed)	Time (calculated)
295° K	0	3a	30-33v	3.5 ms	3 ms
295	0	2	30-33	10	9
295	0	1	30-33	35	35
373	0	1	33	.1	.1
200	0	3	20-33	25	23
200	0	2	20-33	50	47
295	+10 ma	1	27-30	45	43
295	-10 ma	1	33-35	25	27

The data in Table 1 were obtained on transistor A, under a variety of operating conditions designed to test the postulates stated in the preceding section. The two voltages given for collector voltage indicate the increase in avalanche voltage as the temperature increased. The calculated times were found later from Figure 13, which is yet to be developed.

The transistor was then disassembled to determine the physical structure. Figure 11 shows a cross-section through a radius of the cylindrical structure. Generally accepted values of the physical constants at 300° K for the materials of this transistor are listed in Table 2. The solder used was nearly all tin.

Table 2. Physical constants of materials used

Physical property	Units	Germanium	Indium	Tin
Specific heat	cal/gm-°C	.074	.057	.054
Thermal conductivity	cal/sec-cm-°C	.14	.057	.16
Density	gm/cm^3	5.33	7.31	7.3

In order to determine the temperature variation and distribution with heating, an electrical analog of the thermal situation was constructed. The following assumptions were made to simplify the model:

1. The thermal conductivities remain constant with temperature. The assumption is valid for the

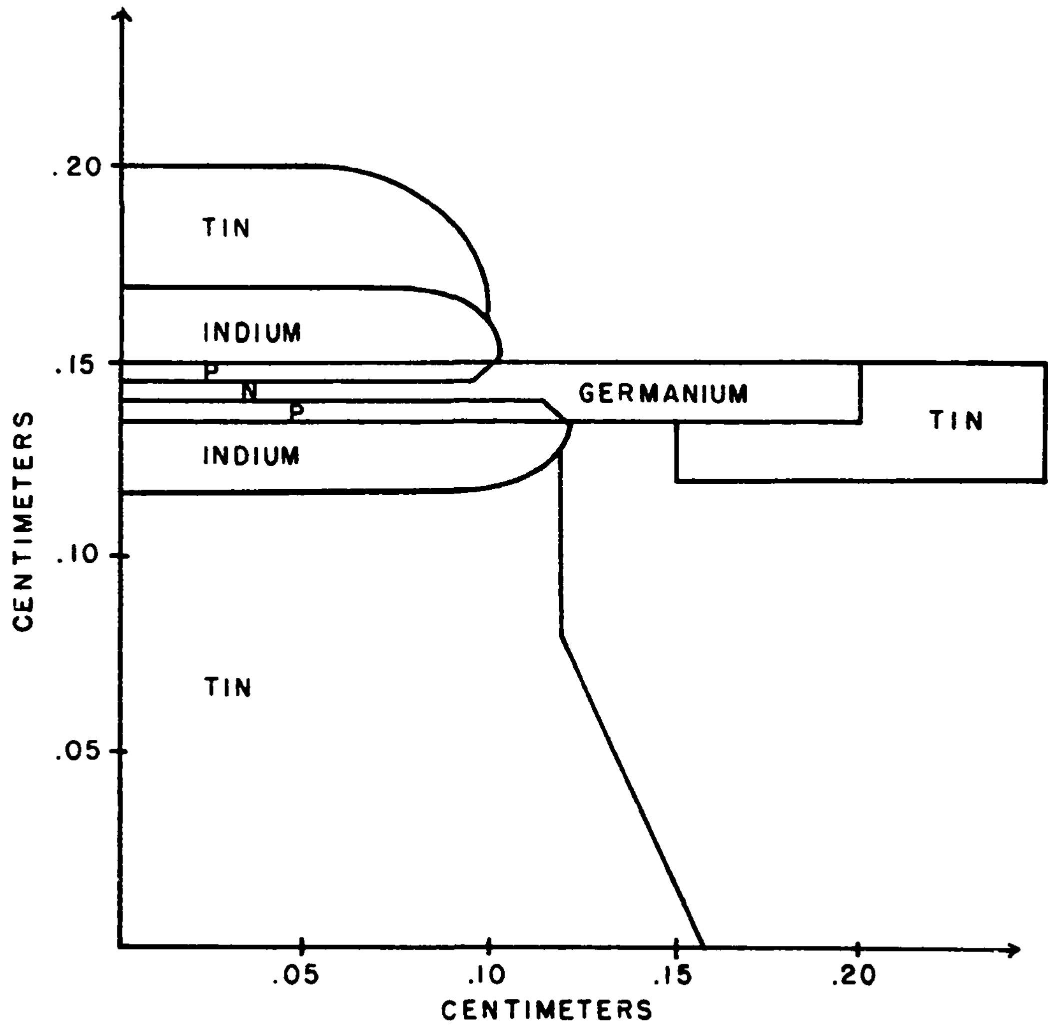

Figure 11. Radial section of the transistor cited

metals. The conductivity of germanium drops slightly with temperature, and is down to .12 cal/sec-cm-^{o}C at 400° K (13).

2. Thermal capacitance, the product of specific heat and density, is constant with temperature. In the range from 200-500° K the true variation is less than 10 per cent for these materials.
3. The base can be maintained at constant temperature. No problem was encountered here, since this was intermittant operation, and the average power was very low. Care was used in attaching the heat sink.
4. The heat loss by convection and radiation at the surfaces is negligible. The sealed atmosphere eliminates air currents to control convection. The temperature of the surfaces does not exceed 400° K in the useful range of the model, and radiation could not exceed .05w according to the law of Stefan-Boltzmann.

The electrical analog for the model of Figure 11, using the constants of Table 2, appears in Figure 12. A development of this analog may be found in the Appendix. If currents proportional to the internally generated heat are applied at the appropriate nodes, then temperatures within the transistor can be determined by a voltage measurement at the corresponding point in the electrical circuit.

As pointed out earlier, the avalanche voltage increases with temperature to a value where the intrinsic conductivity begins to dominate. This temperature is about 380° K for the problem at hand (see Figure 6). The temperature throughout the depletion layer will be maintained at a uniform level until this temperature is reached, and this constraint is inserted into the analog by tying nodes 1, 2, 3, 4, and 5 together until this temperature (voltage) is reached.

Figure 13 is a plot of the results obtained using this analog. Temperature rise of the depleted layer above the heat sink reference temperature is plotted as a function of time, for a constant power dissipation. Thus, knowing the power dissipated in the transistor depletion layer, one can immediately predict the time required for this layer to reach the critical temperature of 380° K. The thickness of the depletion layer decreases somewhat as the temperature rises, which should decrease the value of capacitance associated with this node in the analog. Fortunately, however, the value of this particular capacitance is not at all critical in the problem at hand, since its effect is completed very early in the transient. The time constant associated with the entire base region (all nodes tied together) is only .25 milliseconds.

When temperature of the junction reaches 380° K, the constraint of constant temperature throughout the depleted region is removed. The temperature at the center rises above

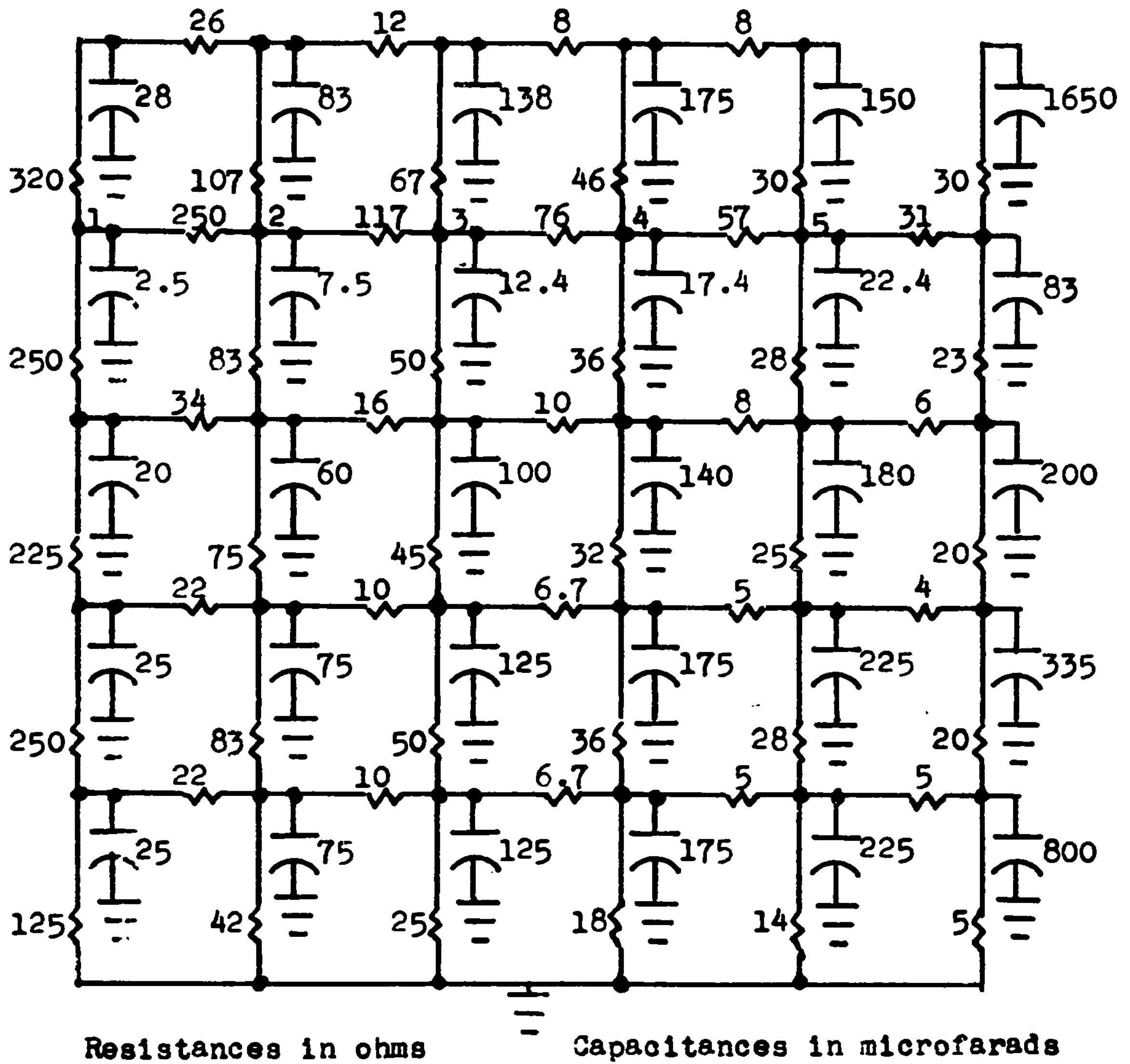

Figure 12. Electrical analog for thermal problem

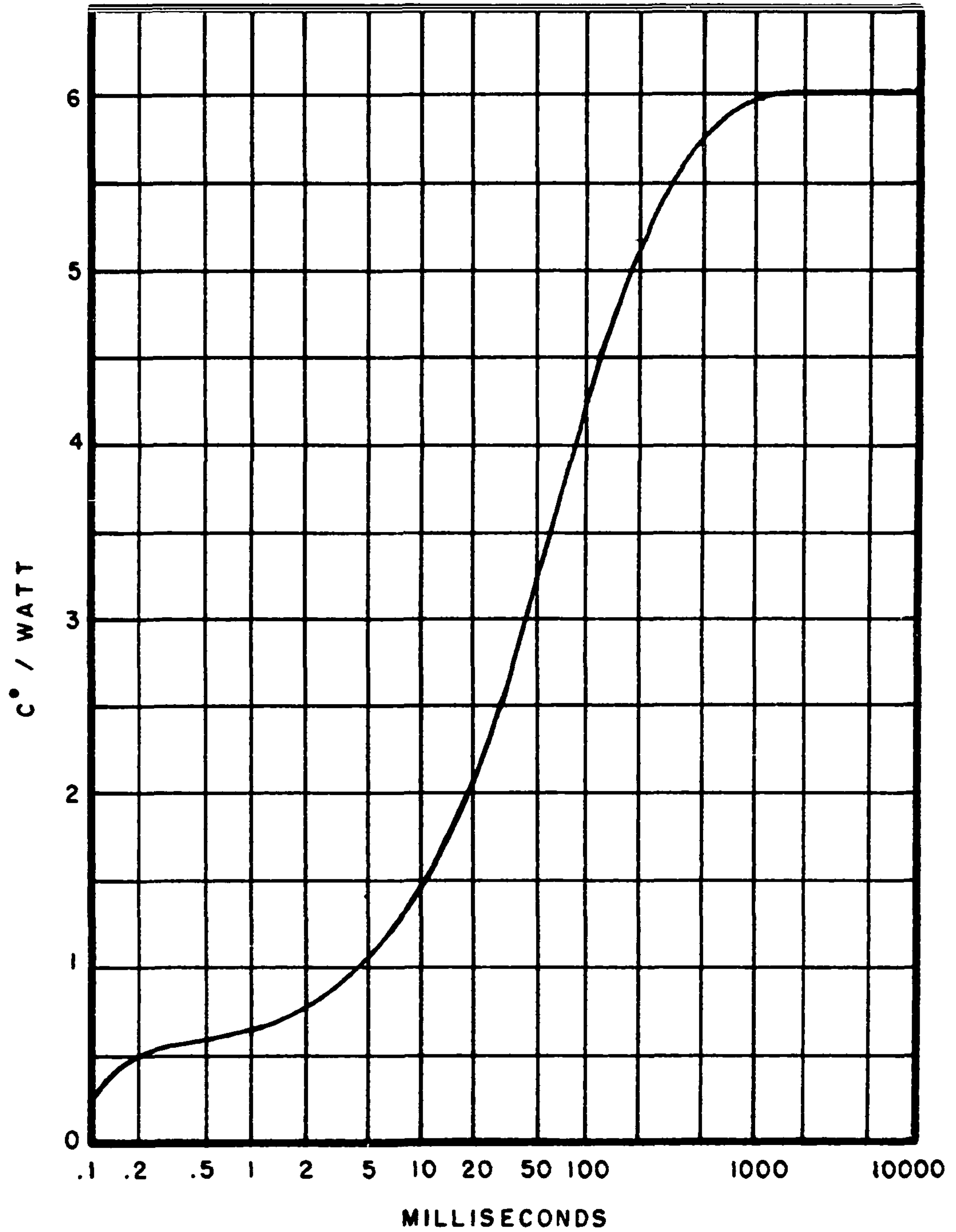

Figure 13. Temperature rise in transistor base per watt dissipated (reproduced from oscilloscope trace)

that of the periphery and the effect is cumulative, since now the avalanche voltage is lower at higher temperatures, and the current is "pinched in". The time constant associated with node 1 alone is also .225 milliseconds, and the total time required for the "pinch in" to take place is observed to be less than 10 micro-seconds. An analog with a controlled source could be constructed to simulate this regenerative action, but in view of the times involved in the rest of the heating cycle, this switching time will be neglected.

The calculated values inserted in Table 1 assume that the delay time is that time required for the transistor base region to reach 380° K. In view of the coarseness of the analog, there is excellent correlation between calculated and experimental values.

A similar analysis was performed on each of the following transistors, with experimental and calculated results tabulated. A variety of germanium alloy-junction power transistors is represented.

Table 3. Delay times for several different transistors with zero base current at a heat sink temperature of 295° K. Common-emitter parameters are specified for current amplification and cut-off frequency

Unit	V_c	I_c	β	f_{co}	time (observed)	time (calculated)
A	30v	3a	35	11kc	3.5 ms	3 ms
B	32	3	80	10	5	4.5
C	35	3	50	7	50	58
D	31	1	60	100	1	1.5
E	34	2	54	12	6	6

Low-power, medium-power, high-power, and high-frequency transistors are represented by B, E, C, and D respectively. The variation between observed and calculated values is not considered to be excessive in view of the assumptions that had to be made in the construction of their models. Unit D, especially, offered the handicap of very small physical dimensions.

Forty units of the same manufacturers type as unit A were subjected to a test for delay time under the identical conditions of open base, 3 amperes collector current and heat sink at 295° K. The results are shown in Table 4. A course breakdown of results into three groups was made.

1. Twenty-four units had delay times between 2 and 8 milliseconds. This compares favorably with the calculated 3 milliseconds for unit A, which must have geometry typical of the production run. These transistors are low cost units and standardization of geometry evidently is not maintained. Also, the avalanche voltage varied between 26 and 36 volts, which has its effect on delay times.
2. Five units had delay times of less than .5 millisecond. All of these were disassembled and examined in detail. Two units had an unusually poor thermal path to the heat sink, but this in itself is not believed to be the cause, since increasing the

Table 4. Variation of delay times for transistors of same type
$I_c = 3a$ $I_b = 0$ $T = 295^{\circ}$ K

Unit	V_c	β	f_{co}	I_{co}(10v)	delay time
1	35 v	50	10 kc	1.2 ma	8 ms
2	32	70	11	.06	5
4	30	60	9	.07	5
5	35	54	11	.08	3
6	28	52	11	.03	5
8	28	20	11	.03	8
9	35	80	11	.05	.1
12	28	34	12	.2	8
13	25	78	8	.7	3
14	33	54	9.5	.1	4
15	32	40	12	.04	.6
16	27	54	12	.12	16
17	34	34	11	.2	6
18	29	58	9.5	.06	4
19	27	76	12	.65	4
20	28	48	10	.05	.1
21	31	48	11	.06	12
22	34	64	9.5	.06	25
23	35	52	12	.06	.3
25	34	54	13	.09	6
26	26	56	8	.2	6
27	32	56	10.5	.04	13
29	30	48	12	.07	15
30	35	54	8.5	.14	7
31	30	44	11.5	.04	20
32	32	52	10	.05	.1
34	34	64	8.5	.07	30
35	35	56	10.5	.07	6
36	33	62	10	.04	2
37	30	62	10	.04	14
38	28	60	9	.06	4
39	30	80	9	.03	35
40	34	52	9	.06	18
41	33	48	11	.18	5
42	33	64	9.5	.08	2
43	32	60	11	.04	25
44	33	61	10	.18	4
45	28	64	10	.06	8
46	29	54	11.5	.08	4
47	30	60	12	.06	8

appropriate values of resistance in the analog did not affect the first millisecond of the voltage transient to any great extent. Peeling away the indium and regrowth regions was not always successful, but in two cases flaws were found in the base region, so that it was extremely narrow in one or more places. This could account for a shorter delay, especially if the resistivity at the fault were somewhat lower to concentrate the current.

3. Eleven units had delay times of 10 milliseconds or more. Without exception, these transistors were found to have either a larger diameter than A, a shorter distance between the base region and the heat sink than A, or both. This group really represents the more desirable type of transistors, with the lowest thermal resistance.

An investigation of the thermal situation arising after secondary breakdown was also carried out. Using the model of Figure 11, and assuming the heat source to be concentrated near the center of the base region, one may make several observations at once. The temperature near the heat source varies roughly as the reciprocal of the distance from it if one assumes a point source. This, of course, means that in the actual situation the current is concentrated into a path of finite radius, or the temperature would be infinite.

The actual radius, inside of which the temperature exceeded the melting point of germanium, was estimated from experimental data. For example, a transistor of similar construction to A typically had a voltage immediately after secondary breakdown of 8 volts for 1 ampere current, 7 volts for 2 amperes, and 6 volts for 3 amperes for ambient temperature of 295° K. The higher currents could not be maintained for many milliseconds or the unit would fail. To observe the maximum power that could be dissipated in the steady state before failure occurred, it was necessary to obtain secondary breakdown with minimum current, and then slowly (over a period of several seconds) increase the current until failure occurred. Typically the failure occurred at about 1.5 amperes and 6 volts. Using this power dissipation, together with the thermal resistance, one is able to estimate the geometry of the isotherm T = 1210° K (melting point of germanium). The thermal conductivity of germanium drops from .14 cal/sec-cm-°C at 300° K to .05 at 1200° K in an approximately linear fashion (13). The thermal conductivity of indium is .057, so a value of .06 was chosen as an average value of conductivity in the vicinity of the hot spot.

As a first approximation, spherical symmetry was assumed, with the temperature at r = .14 cm being 295° K. With the heat crossing the 1210° K isothermal sphere at a

rate of 9 watts = 2.16 cal/sec in the steady state,

$$1210 - 295 = 2.16 \frac{1}{k4\pi} \left(\frac{1}{r} - \frac{1}{.14} \right) \qquad (17)$$

the radius is .003 cm. In actuality the heating takes place in a cylinder about .015 cm long (the width of the germanium layer, in which the resistivity is several orders of magnitude greater than that of indium). So the hot spot is a cylinder .015 cm long and approximately .001 cm in radius, found by equating surface area of the cylinder and the sphere.

Evidently the temperature does not obey the 1/r relationship inside this radius, or failure would have occurred sooner at a smaller radius.

The electrical resistance of such a current path is $L/\sigma A$, so at 1200° K when σ is 1000, R = 4.8 ohms. This compares well with the 4 ohms observed. Such a correlation, considering the approximations made, is very likely a result of some compensating errors, but indicates that the assumptions were reasonable.

The radius of the hole melted through the base region of this type of transistor during failure has always been observed to be approximately .002 cm in diameter (8).

It is doubtful whether the temperature must always reach 1210° K for the unit to fail, however, because the indium atoms will diffuse readily at somewhat lower temperatures,

and, given sufficient time, will complete a path of p-type material across the base. Such failures can be distinguished from those where complete melting takes place, since they have a much higher resistance at low currents.

Where diffusion begins, but heating stops before the unit fails, the characteristics of the unit will be permanently altered.

Diodes

Secondary breakdown was observed to occur in p-n diodes, as well as transistors. The phenomenon appears to be identical to that of the transistor, except that the heating in the diode takes place in the region between the junction and the metallic contact of the higher resistivity material. If this observation had been reported earlier, several previously proposed explanations of secondary breakdown would not have developed. The probable reason for the delay in recognizing secondary breakdown in diodes is that, for most diodes, the dissipation is such that the unit fails almost immediately after secondary breakdown occurs. The delay is often in the order of 10 microseconds. The data in Table 5 was obtained from a high current germanium diode, using a heat sink temperature of 295° K, and a constant current source to protect the diode after secondary breakdown. Avalanche voltage is 90 volts. The relatively long delay

times shown in the table are a result of the low current and the good heat sink.

Table 5. Delay time for secondary breakdown of p-n junction diode

Reverse voltage	Reverse current	Delay time	Voltage after secondary breakdown
70 volts	.05 amps	2000 millisec.	20 volts
80	.133	4	11
85	.175	2.5	10
87	.33	.8	9

SUMMARY AND CONCLUSIONS

A thermal mechanism has been proposed to explain the secondary breakdown observed in transistors. It was stated that heating in the depletion layer of the collector junction causes a uniform temperature rise in that region, until a critical temperature is reached. At this point the avalanche voltage suddenly decreases with temperature rise, and a regenerative situation develops. The effect of this thermal "pinch-in" is to quickly concentrate the collector current into a narrow path through the base. As a result of the ensuing localized high temperature, the base material in the current path becomes intrinsic, punching through the emitter barrier. The resulting release of carriers results in a voltage drop dependent only on the resistance of the high temperature current path.

It is felt that the experimental results verify these conclusions. Further, it is thought that the tests justify the belief that secondary breakdown is a characteristic basic to the p-n junction, rather than to the transistor.

Factors important in guarding against failures occurring in transistors due to secondary breakdown include those precautions taken to reduce the possibility of thermal runaway in transistors. These include the maintenance of a low heat sink temperature, and the construction of transistors with

as low a thermal resistance between collector junction and heat sink as possible.

It has been suggested that the negative resistance associated with the secondary breakdown phenomenon will allow use of the device as a bistable switch (5). This author would caution that, unless minimal currents are used, any device so used may show a "fatigue" due to the gradual diffusion of impurity atoms into the base region that would inevitably occur at high temperatures, even though they are quite localized.

LITERATURE CITED

1. Zener, C. A theory of the electrical breakdown of solid dielectrics. Proceedings of the Royal Society (London). 145: 523-529. 1934.

2. McKay, K. G. and McAfee, K. B. Electron multiplication in silicon and germanium. Physical Review. 91: 1079-1084. 1953.

3. McKay, K. G. Avalanche breakdown in silicon. Physical Review. 94: 877-944. 1954.

4. Miller, S. L. Avalanche breakdown in germanium. Physical Review. 99: 1234-1241. 1955.

5. Thornton, C. G. and Simmons, C. D. A new high current mode of transistor operation. Institute of Radio Engineers Transactions on Electronic Devices. ED-5-1: 6-10. 1958.

6. Lin, H. C., Hlavacek, A. R., and White, B. H. Transient operation of transistor with inductive load. Institute of Radio Engineers Transactions on Electronic Devices. ED-7-3: 174-178. 1960.

7. Moll, J. L., Tanenbaum, M., Goldey, J. M., and Holonyak, N. P-N-P-N transistor switches. Proceedings of the Institute of Radio Engineers. 44: 1174-1182. 1956.

8. Schafft, H. A. and French, J. C. Second breakdown in transistors. Institute of Radio Engineers Transactions on Electronic Devices. ED-9-2: 129-136. 1962.

9. Morin, F. J. and Maita, J. P. Conductivity and Hall effect in the intrinsic range of germanium. Physical Review. 94: 1525-1529. 1954.

10. Prince, M. B. Drift mobilities in semiconductors. Physical Review. 92: 681-687. 1953.

11. Adler, R. B. and Longini, R. L. Introduction to semiconductor physics. New York, New York, John Wiley and Sons. 1962.

12. Hunter, L. P. Handbook of semiconductor electronics. New York, New York, McGraw-Hill Book Company. 1956.

13. Abeles, B., Beers, D. S., Cody, G. D., and Dismukes, J. P. Thermal conductivity of Ge-Si alloys at high temperatures. Physical Review. 125: 44-46. 1962.

ACKNOWLEDGMENT

The author is indebted to Mr. Henry Ablin for his suggestion of the need for work in this area. He also wishes to thank Dr. R. G. Brown for his advice in the preparation of this manuscript.

APPENDIX: DEVELOPMENT OF ELECTRICAL ANALOG FOR HEAT FLOW PROBLEM

The diffusion of heat in a solid is governed by the following laws:

1. Heat flows in the direction of decreasing temperature.
2. The rate of heat flow through an area is proportional to the area and to the temperature gradient normal to the area. The proportionality constant is thermal conductivity k.
3. The quantity of heat required to produce a given temperature change in a body is proportional to the mass of the body and to the temperature change. This proportionality is the specific heat c.

If one were interested in making measurements within this solid, it might be desirable to replace the original distributed parameter device with a model consisting of a number of smaller sections, so that the necessary measuring devices might be inserted between the "lumped" parameters.

Figure 14 is an illustration of one method for sectioning the transistor shown. Cylindrical symmetry is assumed, so that each section represents a toroid. Measurements are made to the center of each area.

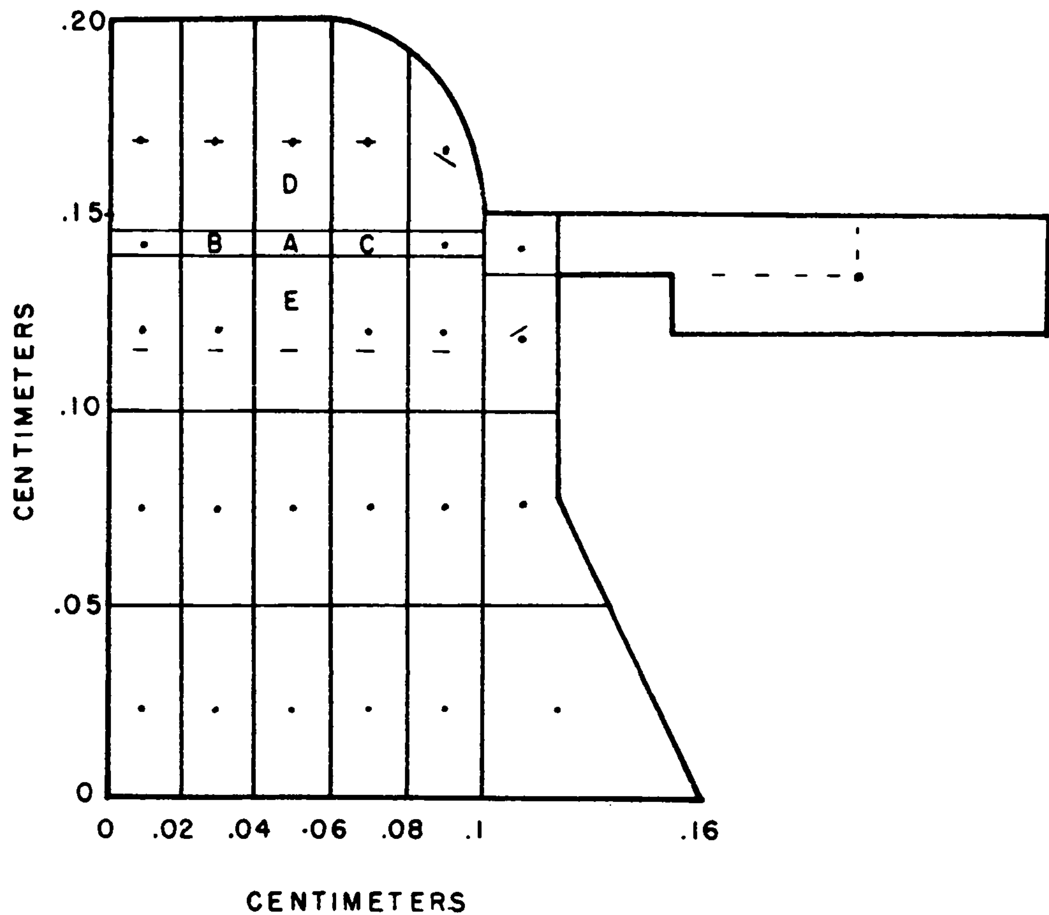

Figure 14. Radial section of transistor showing subdivision used in determining lumped constant model

Consider the element marked A. The rate at which heat flows into A from B is (using k for germanium)

$$\frac{dQ_1}{dt} = (T_b - T_a)\frac{2\pi hk}{\ln r_a/r_b} = (T_b - T_a)\frac{2\pi(.005)(.14)}{\ln 5/3}$$

$$= (T_b - T_a).0086 \qquad (18)$$

and the rate from C into A

$$\frac{dQ_2}{dt} = (T_c - T_a)\frac{2\pi hk}{\ln r_a/r_c} = (T_c - T_a)\frac{2\pi(.005)(.14)}{\ln 7/5}$$

$$= (T_c - T_a).0131 \qquad (19)$$

From D to A the rate is (using k for indium)

$$\frac{dQ_3}{dt} = (T_d - T_a)\frac{kA}{d} = (T_d - T_a)\frac{k\pi(r_{a+}^2 - r_{a-}^2)}{d}$$

$$= (T_d - T_a)\frac{\pi(.06^2 - .04^2)(.06)}{.025} = (T_d - T_a).015 \qquad (20)$$

and from E to A

$$\frac{dQ_4}{dt} = (T_e - T_a)\frac{kA}{d} = (T_e - T_a)\frac{k\pi(r_{a+}^2 - r_{a-}^2)}{d}$$

$$= (T_e - T_a)\frac{\pi(.06^2 - .04^2)(.06)}{.02} = (T_e - T_a).019 \qquad (21)$$

The rate of temperature change in element A (using c and density p of germanium)

$$\frac{dQ}{dt} = cp\ h(r_a{}^2 - r_a{}^2)\ \frac{dV_a}{dt} = (.074)(5.33)(.005)(.06^2 - .04^2)\frac{dV_a}{dt} =$$

$$12.4(10^{-6})\ \frac{dV_a}{dt} \qquad (22)$$

is proportional to the sum of all the rates of heat entering the element plus Θ, the rate at which heat is generated within the element.

$$12.4(10^{-6})\ \frac{dV_a}{dt} = \frac{T_b - T_a}{117} + \frac{T_c - T_a}{76} + \frac{T_d - T_a}{67} + \frac{T_e - T_a}{52} + \Theta \qquad (23)$$

The units on each term in Equations 18-23 are calories/second. Equation 22 resembles an equation summing currents at a node,

$$C\ \frac{dE}{dt} = \frac{E_b - E_a}{R_1} + \frac{E_c - E_a}{R_2} + \frac{E_d - E_a}{R_3} + \frac{E_e - E_a}{R_4} + I \qquad (24)$$

where E represents voltage relative to the reference (indicated as ground) and suggests the use of an electrical analog to facilitate construction of the model and measurements of variables. Each of the values of C and R in Figure 12 were determined in the manner outlined above. Current, which is analogous here to rate of heat flow, was injected into the

nodes where internal heat is being generated. Voltage measurements at any node indicate the temperature variation (from heat sink reference) in the model. The electrical circuit is completely analogous to the lumped model of the thermal problem, and the lumped model approximates the actual situation.