Internal Field Emission at Narrow Silicon and Germanium *p-n* Junctions

A. G. CHYNOWETH, W. L. FELDMANN, C. A. LEE, R. A. LOGAN, AND G. L. PEARSON
Bell Telephone Laboratories, Murray Hill, New Jersey

AND

P. AIGRAIN*
Ecole Normale Superieure, 24 Rue Lhomond, Paris France

(Received November 5, 1959)

A detailed study has been made of the reverse characteristics of several silicon and germanium alloyed *p-n* junctions with breakdown voltages in the range of about 0.1 to 0.8 volt. In these junctions the reverse current is generated almost entirely by internal field emission (tunneling). The reverse bias characteristics are insensitive to the dislocation density present so that the tunneling current occurs mainly in undistorted material. From capacitance studies it is established that these narrow junctions are very close to being ideal step junctions. The room-temperature reverse characteristics are analyzed in terms of the usual tunneling probability expressions and in particular, good agreement, both qualitative and quantitative, is found between experiment and theory. The tunneling probability exp $(-\alpha\epsilon^{\frac{1}{2}}/E)$, when compared with experiment, yields values for $\alpha\epsilon^{\frac{1}{2}}$ in agreement with the theoretical ones to within a factor of less than 2 for both silicon and germanium.

The critical voltage (the reverse bias voltage necessary to maintain a constant tunneling current) was measured as a function of temperature from 4.2°K up to temperatures as high as 700°K. In germanium, the critical voltage drops monotonically as the temperature increases whereas in silicon, there is considerable structure in the curve. This is shown to be consistant with the tunneling being by direct transitions in germanium and by indirect transitions (involving phonon emission and absorption) in silicon. In germanium, the temperature dependence of the critical voltage arises from that of the direct energy gap while in silicon, it is determined, primarily, by the available phonon density. From an analysis of the temperature data for silicon that invokes the transverse acoustic phonons, the estimate of $\alpha\epsilon^{\frac{1}{2}}$ that is obtained is in excellent agreement with that found, independently, from the analysis of the reverse characteristics.

INTRODUCTION

CONSIDERABLE interest has been rekindled recently in the phenomenon of internal field emission because of the Esaki effect.[1] However, though theories of internal field emission have been extant for 25 years or more, and though its occurrence at sufficiently high fields is well established, there has not yet been a convincing quantitative experimental verification of the theoretical laws. The work to be described in this paper was undertaken with a view to obtaining such verification.

The most reliable systems in which to study internal field emission are sufficiently narrow *p-n* junctions. It has been shown previously[2] that reverse bias (positive voltage applied to the *n*-type side) breakdown in such junctions in silicon is by internal field emission. The diffused junctions used in this early work, however, were still relatively wide, having width constants of about 400 A. (The width constant is defined as the width of the space charge region with a total potential drop across it, built-in plus applied, of one volt.) Breakdown occurred in these units around five or six volts so that the total breakdown current arose from both field emission and avalanche mechanisms. More recently, narrower junctions have been obtained in silicon and germanium by alloying to very low resistivity materials. These junctions have breakdown voltages as low as a few tenths of a volt. Thus, the total potential across the junction is appreciably less than that required for pair production (which requires 2.3 volts for silicon[3,4] and 1.5 volts for germanium,[4,5] not allowing for energy-losing collisions made by the carriers in crossing the junction). It was, therefore, of considerable interest to make a detailed study of the behavior of these narrow silicon and germanium junctions in the hope that it would provide experimental confirmation of the theoretical expression for the internal field emission current.

EXPERIMENTAL

(i) Fabrication of Junctions

A series of *n*-type silicon-arsenic alloys with resistivities ranging between 0.001 and 0.10 ohm-cm were used as the starting materials for the silicon junctions and a germanium-arsenic alloy of 0.0035 ohm-cm for the germanium junctions. Individual units consisted of a 0.040-inch cube of the crystal to which, first, an ohmic contact was made by alloying to it an antimony-doped gold wire, 0.010 inch in diameter. The junction was then formed by alloying a 0.003-inch diameter aluminum wire to the cube. In every case, the junction alloying was done on a (111) face of the crystal. Alloying was done in a nitrogen atmosphere at about 850°C and immediately after the bonding had occurred, the crystal was rapidly cooled to minimize diffusion effects.

(ii) Voltage-Current Characteristics

Figure 1 shows the voltage-current characteristics for silicon and germanium units of various starting

* Guest of Bell Telephone Laboratories, summer, 1959.

[1] L. Esaki, Phys. Rev. **109**, 603 (1958).

[2] A. G. Chynoweth and K. G. McKay, Phys. Rev. **106**, 418 (1957).

[3] A. G. Chynoweth and K. G. McKay, Phys. Rev. **108**, 29 (1957).

[4] V. S. Vavilov, J. Phys. Chem. Solids **8**, 223 (1959).

[5] J. Tauc, J. Phys. Chem. Solids **8**, 219 (1959).

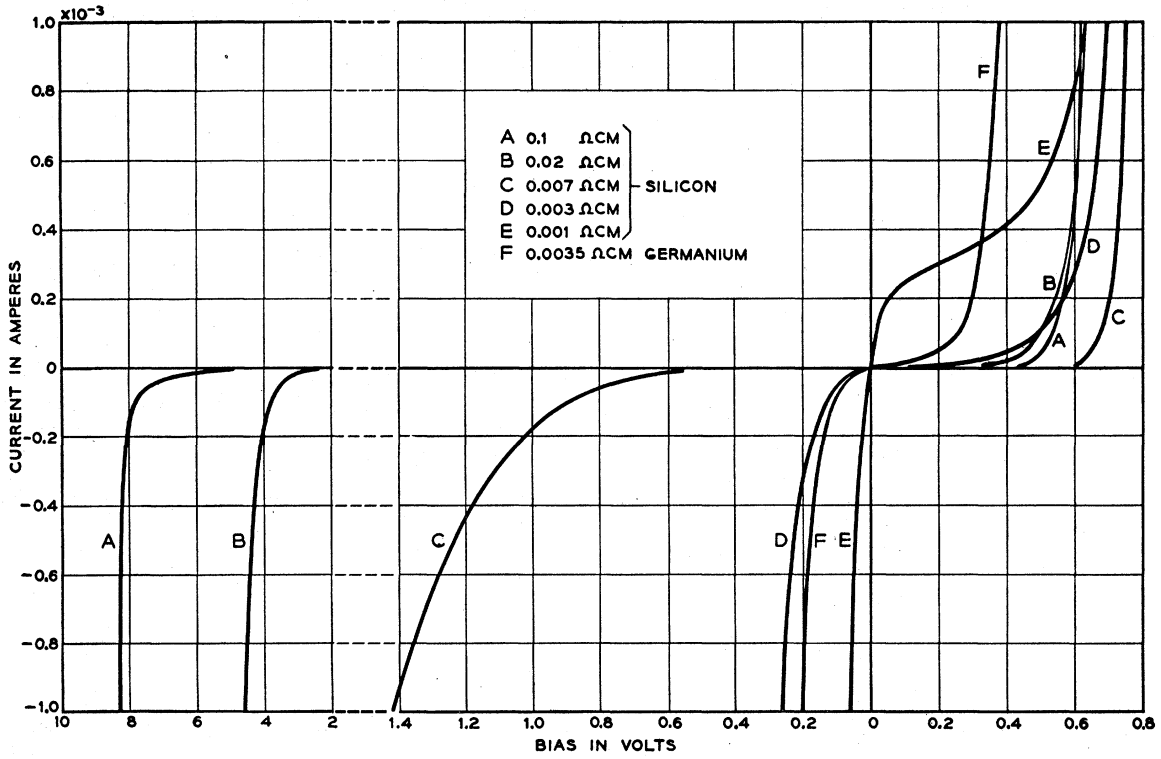

FIG. 1. Room-temperature current-voltage characteristics for typical examples of the silicon and germanium junctions made from materials of different resistivities.

resistivities. For the silicon junctions, the forward characteristics tend to saturate at about 0.7 volt, in common with conventional junctions. The forward impedance tends, however, to become less as the resistivity is decreased. The 0.001 ohm-cm junction shows a considerable amount of forward current and, in fact, is close to demonstrating the Esaki effect (a region of negative resistance in the forward characteristic). The reverse characteristics vary considerably, the breakdown voltages clearly covering the range that includes breakdown due entirely to field emission[2,3] (breakdown voltage less than about one volt for silicon) and breakdown due mainly to avalanching (breakdown voltage greater than about 6 volts). The former is exemplified by the units made from 0.001, 0.003, and 0.007 ohm-cm material while the latter refers to the unit made from 0.1 ohm-cm material. The unit made from 0.02 ohm-cm material breaks down at about 5 volts and, no doubt, exhibits a mixture of the two breakdown mechanisms.[2] The rest of this paper is concerned with silicon units made from the three lowest resistivity materials and the germanium units for which it is reasonably certain that essentially all of the reverse current originates by internal field emission.

(iii) Effect of Temperature

The reverse characteristics of the three silicon field emission junctions and the germanium junction are shown in Fig. 2 at 300°K and 78°K. It should be noted that in all cases this temperature difference does not cause any great change in the current (less than an order of magnitude). Also, the breakdown voltage increases as the temperature decreases. Both of these features are expected for silicon field emission junctions[2] (and probably for germanium field emission junctions also) and are not characteristic of avalanche junctions.

(iv) Effect of Dislocations

The reverse characteristics of field emission junctions are notable for their "softness." One hypothesis attri-

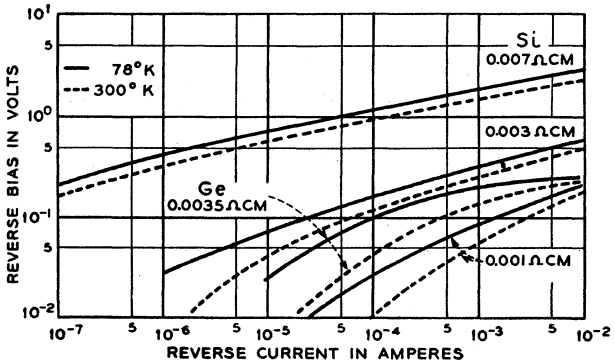

FIG. 2. Logarithmic plots of the room temperature, and liquid nitrogen temperature reverse characteristics of the field emission junctions.

buted at least some of this softness to dislocations. When edge dislocations pass through the crystal the distortions in the crystal lattice cause a local reduction in the width of the band gap thus making field emission much easier at the places where the dislocations run through the high field region of the junction.[6] It might be expected, therefore, that as the reverse bias is increased, most of the field emission current is generated at first near the dislocation. However, the current through these "weak" spots would eventually become saturated on account of the spreading resistance of the end regions of the junction so that a spectrum of weak spots would give rise to an abnormally soft reverse characteristic.

To test this hypothesis a low resistivity (0.002 ohm-cm) n-type crystal was grown with a low dislocation density. Subsequently, dislocation etch pit studies yielded a dislocation density of the order of 50 cm^{-2} so that the possibility of any dislocations being present within the very small area of the junction was virtually zero. Alloyed junctions were made on (111) faces in sets of four: (i) on the material as grown (ii) on material that had been heated to 750°C without bending (iii) on material that had been heated to about 750°C and bent to a radius of curvature of 3.3 cm about the [111] bending axis, (iv) on material that had been heated and bent to a curvature of 0.75 cm. Such bending treatments can be expected to introduce dislocation arrays with densities greater than 10^7 cm^{-2}. However, the voltage-current characteristics of all four junctions of each of several sets were, in every case, virtually identical. To check the possibility that the alloying process itself might introduce a high dislocation density, some junctions were etched after fabrication but they revealed no excess amounts of dislocation etch pits. It is concluded, therefore, that the shapes of the reverse characteristics of these narrow junctions are not determined to any significant extent by the presence of dislocations.

(v) Capacitance Studies

There was some question as to how valid it was to assume that these alloyed junctions could be described as step junctions with field strengths that varied linearly with distance through the space charge region. Accordingly, the bias dependence of the capacitance was determined experimentally using a substitutional bridge method. The measurements were made at a frequency of 400 kc/sec with a driving voltage of not more than 10 mv. Figure 3 shows the results of these measurements plotted in the form C^{-2} (where C is the capacitance in arbitrary units, differently scaled for each junction) against V_a, the applied potential. Data were taken over the widest range possible for each junction, the range being limited in both the forward and reverse

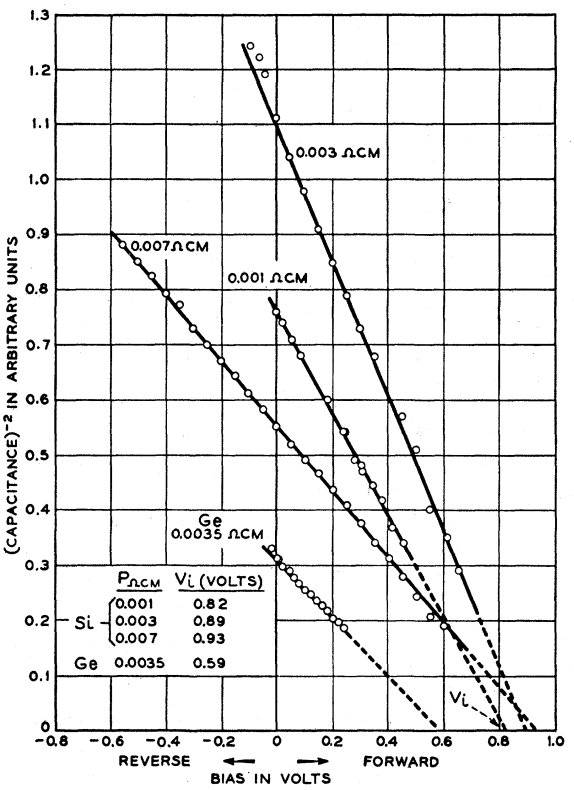

FIG. 3. Capacitance plots showing that the narrow junctions behave as good step junctions over fairly wide ranges of bias.

directions by the "lossiness" of the junction at high currents. The three silicon junctions all fit straight lines satisfactorily, and over a somewhat limited range of bias, the germanium junction fits a straight line also. Deviations from the straight line occur in all units at appreciable currents due to the injected carrier capacitance but within those limits, the junctions, especially the silicon ones, behave as reasonably good step junctions.

The intercept on the bias axis is usually taken to be the built in voltage, V_i. However, the values of V_i determined from Fig. 3 are unusual in two respects: (i) they are much lower than is to be expected for junctions with both sides degenerate, and (ii) V_i apparently increases with the resistivity of the starting material whereas it would be expected to decrease. These anomalies have been considered in detail by Aigrain,[7] who has shown that they can probably be accounted for by various corrections to the normal diode theory that have to be taken into account when dealing with degenerate semiconductors.

(vi) Estimates of Junction Widths

To fully analyze the reverse characteristics it was necessary to know the widths of the space charge regions. These could be estimated knowing the resis-

[6] A. G. Chynoweth and G. L. Pearson, J. Appl. Phys. **29**, 1103 (1958).

[7] P. Aigrain (private communications).

TABLE I. Summary of data pertinent to the analysis of the field emission junctions.[a]

	Silicon			Germanium
Impurity, n^+	As	As	As	As
Resistivity, ohm-cm	0.001	0.003	0.007	0.0035
n concentration cm^{-3}	7×10^{19}	2.5×10^{19}	8×10^{18}	2.7×10^{18}
p^+ alloying impurity	Al	Al	Al	Al
p concentration cm^{-3}	1.3×10^{19}	1.3×10^{19}	1.3×10^{19}	4×10^{20}
V_{in} ev	0.06	0.03	0.01	0.01
V_{ip} ev	0.04	0.04	0.04	0.76
$V_i=V_{in}+V_{ip}+\epsilon_G$ ev	1.25	1.22	1.20	1.43
V_i expt.	0.82	0.89	0.92	0.59
$\langle W_1\rangle_{av}$ measured, $\times10^{-8}$ cm volt$^{-\frac{1}{2}}$	90	110	125	310
W_1, calculated, $\times10^{-8}$ cm volt$^{-\frac{1}{2}}$	110	125	160	260

[a] See reference 8.

tivities of the end regions but they were also determined experimentally. To do this, the capacitances of several junctions made from each starting material were measured at zero bias. The aluminum wires were then etched out of the alloyed junctions so that the shape of the alloyed junction could be studied under the microscope. In some cases it was not possible to make very reliable estimates of the junction areas as they were too irregular. However, in several units the alloying had produced a relatively clean shape which enabled the area to be estimated to within $\pm30\%$. From the capacitance (corrected to unity potential using the established square law and calculated values for V_i) and the area measurements, the width constants were obtained.

(vii) Summary of Quantitative Data

Quantitative data pertinent to these junctions are summarized in Table I. The figures for the donor and acceptor concentrations were known from resistivity and solid solubility studies. The values of V_{in} and V_{ip} were estimated from the formula for degenerate semiconductors,

$$V_{in}=\epsilon_F-\epsilon_C=38\times10^{-16}n^{\frac{2}{3}}(m_0/m^*)\text{ ev},$$

(and similarly for V_{ip}) where ϵ_F, ϵ_C are the positions of the Fermi level and the bottom of the conduction band, respectively, n is the carrier concentration,[8] and (m_0/m^*) is the ratio of the free electron mass to the effective electron mass. The values so obtained for $V_i=V_{in}+V_{ip}+\epsilon_G$, where ϵ_G is the energy gap expressed in volts, are seen to be considerably in excess of the experimental values, as noted above.

The width constants were estimated from the formula for a step junction:

$$W_1^2=\kappa(p+n)/2\pi epn,$$

where p and n are the majority carrier concentrations, κ is the dielectric constant, and e is the electron charge. The predicted value is seen to be in good agreement with the experimental values considering the uncertainties in the estimates of the junction areas.

(viii) Effect of Temperature on Critical Field

It has been shown in Fig. 2 how the reverse voltage, for any given current, increases as the temperature is lowered. This property was examined in more detail by using, at first, an X-Y recorder to trace the voltage-current characteristics at various temperatures. The critical voltage is defined as the reverse bias for a given current, and by interpolation of the voltage-current characteristics, this quantity could be plotted as a

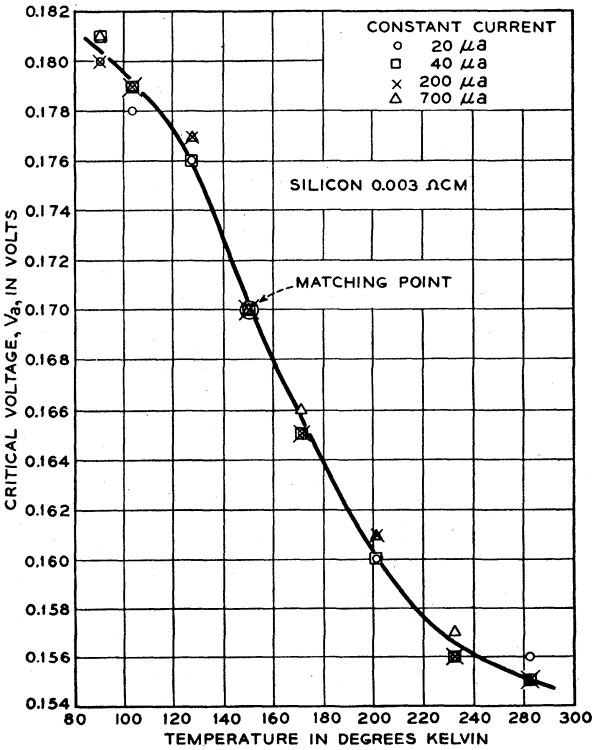

FIG. 4. The critical voltage (reverse bias) versus temperature for a silicon junction in 0.003 ohm-cm material, for various values of the constant current. The data are scaled on the V_a axis only so as to match the point for a current of 40 μa at 150°K.

[8] In the p-type materials, the carrier concentrations were given by the known solid solubilities [F. A. Trumbore, Bell System Tech. J. (to be published)] of aluminum in Si and Ge. For n-type Ge, the source was W. W. Tyler and T. J. Soltys, General Electric Research Report P-193 (unpublished), and for the n-type Si, it was G. Backenstoss, Phys. Rev. **108**, 416 (1957).

function of temperature. The results are shown in Fig. 4 for four different values of the constant current, the critical voltages all being scaled so as to match the data taken at a current of 40 μa at 150°K. The evident superposition of the experimental points proves that the shape of the reverse characteristic does not vary with temperature for reverse currents ranging, at least, between 20 and 700 μa. Thus, for studies of the critical voltage versus temperature, there was nothing particularly critical about the value of the constant current.

Typical variations of the critical voltage as a function of temperature are shown in more detail in Fig. 5, where the results pertain to a constant current of 45 μa for both the silicon and germanium junctions. The voltages were measured with a potentiometer by the four terminal method. The germanium junctions show V_a decreasing slowly at first as T increases, but then more and more rapidly in a monotonic fashion. The silicon crystal, on the other hand, shows a similar slight variation at low temperatures followed by a fairly rapid drop after which, the decrease in V_a slows up. After a region where V_a is again relatively insensitive to the temperature, V_a quite abruptly begins to drop, this drop becoming catastrophic in the temperature range 600 to 620°K. In some cases, the measurements were extended until V_a was very close to zero, and there was no indication of a second step appearing in the curve. The temperature runs were virtually reversible and reproducible, even when the crystal temperature had previously reached 620°K, so that the catastrophic fall in V_a appears to be a fundamental effect rather than due to deterioration of the junction. Essentially the same V_a versus T variations were shown by silicon junctions from each of the three starting materials, and all of them showed the rapid drop of V_a between 600°K and 650°K.

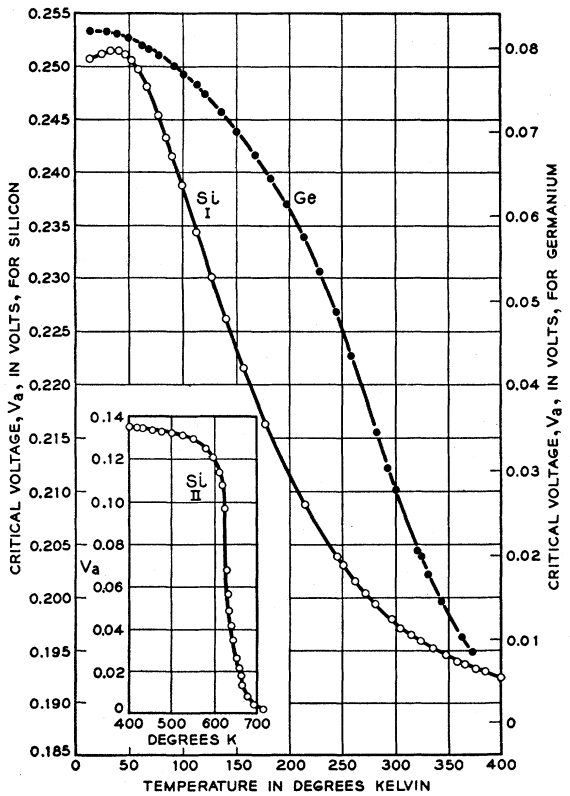

FIG. 5. The critical voltage versus temperature for typical silicon and germanium junctions. The two curves for silicon for the high- and low-temperature ranges refer to two different, though typical silicon junctions.

DISCUSSION

The purpose of this section is to analyze and discuss the experimental results in terms of the theory of internal field emission.

Until recently, all theoretical attempts to derive a law for internal field emission ignored the effects of phonons and were concerned solely with the probability of direct transitions of electrons from the valence band to the conduction band. This defect has been remedied by the excellent theoretical study of Keldysh[9] who, in particular, has pointed out the role that phonons can play in promoting indirect transitions. From the recent experimental work of Holonyak et al.[10] and Esaki[11] it is now known that both direct and indirect tunneling transitions can occur in various semiconducting materials but that for any particular material, one form of transition usually dominates the other. In particular if the semiconductor has its minimum band gap at $k=0$, then direct transitions are much more probable than indirect ones. On the other hand, if the band gap minimum is not at $k=0$, indirect transitions will usually dominate, even for very low phonon densities. In general, it is the value of the band gap that is critical rather than the phonon density since the tunneling probability is a rapidly varying function of the band gap.

Holonyak et al.[10] showed that at liquid helium temperatures, indirect transitions predominate in silicon. Germanium, on the other hand, was an intermediate case, probably because the band gap at $k=0$ is only slightly larger than the actual band gap (0.90 ev compared with 0.74 ev for undoped crystals at 4°K). For germanium crystals highly doped with arsenic or phosphorus, they found no evidence of phonon cooperation but did in the case of antimony doped crystals. It will be shown that the results of the present work for arsenic doped silicon and germanium crystals are consistent with these conclusions, and furthermore, confirm the Keldysh theory for phonon cooperation.

A. Analysis of Reverse Characteristics

For direct transitions, all theories of internal field emission predict that the unidirectional tunneling

[9] L. V. Keldysh, J. Exptl. Theoret. Phys. (U.S.S.R.) **33**, 994 (1957) and **34**, 962 (1958) [translations: Soviet Phys. JETP **6**, 763 (1958) and **7**, 665 (1958), respectively].

[10] N. Holonyak et al., Phys. Rev. Letters **3**, 167 (1959).

[11] L. Esaki (private communication).

current, I, is given by an expression of the form:

$$I=A_D E^r P_D, \tag{1}$$

where A_D is a factor containing, among other quantities, the number of carriers available to attempt the tunneling transition and their frequency of attempting it. The field, E, appears to some power r, where r lies usually between 1 and 3, depending on the various corrections included in the theory such as the effects of image fields or Coulombic forces.[12] The factor, P_D, represents the direct transition tunneling probability and is given by[9]:

$$P_D=\exp[-\pi(m^*)^{\frac{1}{2}}\epsilon^{\frac{3}{2}}/2e\hbar E]=\exp(-\alpha_D\epsilon^{\frac{3}{2}}/E), \tag{2}$$

with

$$\alpha_D=\pi(m^*)^{\frac{1}{2}}/2e\hbar, \tag{3}$$

where m^* is the reduced effective mass, ϵ is the forbidden energy gap, e is the electron charge, and $\hbar=(h/2\pi)$, where h is Planck's constant. For indirect transitions, the effect of phonon cooperation has to be included and we have, essentially:

$$I=A_I E^n[NP_A+(N+1)P_E], \tag{4}$$

where A_I is another constant, similar in purpose to A_D, E is raised to some power n, N is the phonon density, and P_A and P_E are the indirect tunneling probabilities with phonon absorption and phonon emission, respectively. [Note that the above expression applies to the case of one phonon energy only. To include the effects of phonons of other energies, the expression for the current would contain additional terms of similar form. Further terms would also be needed if multiphonon processes are important.] The tunneling probabilities are now given by[9]

$$\begin{aligned} &P_A=\exp[-\alpha_I(\epsilon-\hbar\omega)^{\frac{3}{2}}/E], \\ \text{and} \\ &P_E=\exp[-\alpha_I(\epsilon+\hbar\omega)^{\frac{3}{2}}/E], \end{aligned} \tag{5}$$

where $\hbar\omega$ is the phonon energy and $\alpha_I=4(2m^*)^{\frac{1}{2}}/3e\hbar$. Though $\hbar\omega$ is small compared with ϵ the value of α is such as to make P_A appreciably larger than P_E, even for the smallest phonon energies present in silicon. Equation (4) then becomes, approximately,

$$\begin{aligned} I&=A_I E^n P_E[N(P_A/P_E)+N+1] \\ &\simeq A_I E^n P_E[N(P_A/P_E)+1]. \end{aligned} \tag{6}$$

As will be seen below, at room temperature we have $N(P_A/P_E)\gg 1$ so that, under these conditions

$$I\simeq A_I E^n N P_A. \tag{7}$$

Equations (1) and (7), which are essentially of the same form, are not sufficient, however, for describing the *measured* current-voltage characteristics at a given temperature as they do not take into account the tunnel current in the opposite direction. Specifically, at zero applied bias the net current is zero because the tunnel current from the valence band to the empty states of the conduction band is exactly balanced by a tunnel current from the occupied states in the conduction band on the degenerate n-type side to empty states in the valence band. When an increasing reverse bias is applied, more and more electrons in the valence band on the p side lie opposite empty states on the n side and fewer electrons on the n side lie opposite empty states on the p side. This change in populations can be taken into account exactly in terms of Fermi-Dirac distributions and the densities of states. For present purposes, however, it seems sufficient to put the measured current,

$$I_m=AV_a{}^p E^q P, \tag{8}$$

where A is a constant for a given temperature, P is the appropriate tunneling probability, $q=r$ or n, and the $V_a{}^p$ term represents, in an empirical fashion, the effect of the bias dependence of the net current due to the carrier energy distributions. In particular, the $V_a{}^p$ term, where p is some power, ensures that the measured current, I_m, is zero when $V_a=0$, where V_a is the applied bias. As will become apparent, the actual form of the carrier distribution term is not critical as long as it depends on the field more slowly than does P.

It is appropriate, at first, to put $E=E_M$, where E_M is the maximum field in the step junction. Then

$$E=2(V_a+V_i)/W=2(V_a+V_i)^{\frac{1}{2}}/W_1, \tag{9}$$

where V_i is the built-in voltage and W_1 is the width constant. We note that for the junctions under study, V_a ranges from 0 to a few tenths of a volt whereas V_i is greater than 1 volt so that the variation in E is very slight compared with that of V_a. From (2), (8), and (9) one obtains

$$\begin{aligned} d(\ln I_m)/d(\ln V_a)=p&+qV_a/2(V_a+V_i) \\ &+(\alpha\epsilon^{\frac{3}{2}}W_1/4)V_a(V_a+V_i)^{-\frac{3}{2}}. \end{aligned} \tag{10}$$

Putting $B=(\alpha\epsilon^{\frac{3}{2}}W_1/2)$, where α and ϵ are appropriate to the direct or indirect energy gap with or without phonon cooperation, and since $V_a\ll V_i$, we have to a good approximation,

$$d(\ln I_m)/d(\ln V_a)=p+(B/2V_i{}^{\frac{3}{2}})V_a. \tag{11}$$

Thus, a plot of $d(\ln I_m)/d(\ln V_a)$ against V_a should be a straight line of gradient $B/2V_i{}^{\frac{3}{2}}$. Such plots are shown in Fig. 6 for typical examples of each of three silicon resistivities as well as the germanium junctions. In view of the possible experimental errors and the approximations involved in the analysis, it is felt that satisfactorily straight lines do ensue and that these are consistent with the form of the tunneling probability. In all cases, however, it was found that the plots started to deviate from straight lines for currents in excess of about 100 μa. It is noteworthy that the common factor among the junctions was the critical current of about 100–200 μa rather than a critical voltage, as the applied voltage

[12] W. Franz, *Handbuch der Physik* (Springer-Verlag, Berlin, 1956), Vol. 17, p. 155.

corresponding to such currents varied considerably among the various junctions. Thus, the deviations are not due simply to failures of the approximations, such as putting $V_a \ll V_i$, in the analysis. The deviations all occur for current densities greater than about 1 amp cm^{-2} and at such levels, the injected space charge is becoming comparable with the built-in space charge. Also it is in this same region that the capacity data generally show deviations from the $C^2V =$ constant law. So it seems reasonable to conclude, tentatively, that the prime cause of the deviations is the failure of the effective field to vary as $(V_a+V_i)^{\frac{1}{2}}$ at the higher currents and in fact, in some cases it was found that the form of the deviations in the capacity data paralleled the deviations in the field emission plots fairly closely.

The slopes of the field emission plots, as determined experimentally, can be compared with the values predicted by theory. The slopes yield $B/2V_i^{\frac{1}{2}}$, and using the estimated values of V_i and the measured values for W_1, the quantity $\alpha\epsilon^{\frac{1}{2}}$ was evaluated for each of the junctions. The results are tabulated in Table II where it will be seen that the data from the three silicon junctions agree to within $\pm 20\%$. This is regarded as satisfactory. In deriving the theoretical values, ϵ was put equal to 1.1 ev for silicon and to 0.80 ev (the direct band gap) for germanium. There is considerable uncertainty, as to the correct choice of effective mass to be used in estimating α. Keldysh[9] defines m^*, the reduced effective mass, as

$$m^* = m_C^* m_V^* / (m_C^* + m_V^*),$$

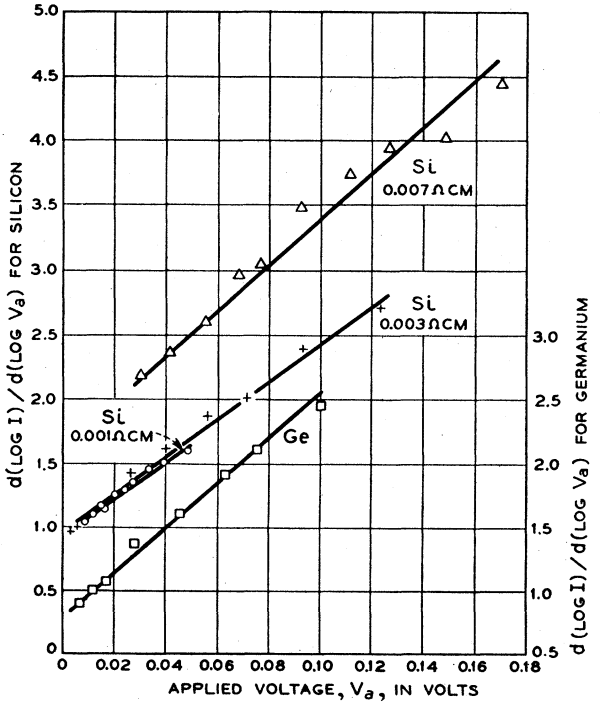

FIG. 6. Plots of $d(\log I)/d(\log V_a)$ versus the reverse bias, V_a, for typical examples of the various field emission junctions.

TABLE II. Experimental and theoretical values of $\alpha\epsilon^{\frac{1}{2}}$.[a]

	Experiment A	Experiment B	Theory
Si 0.001 ohm-cm	9.6×10^7	Si	
Si 0.003 ohm-cm	6.5×10^7	average	7.2×10^7	4.8×10^7
Si 0.007 ohm-cm	7.0×10^7	7.7×10^7	
Ge 0.0035 ohm-cm	2.8×10^7		1.8×10^7

[a] $\alpha\epsilon^{\frac{1}{2}}$ is expressed in volts cm^{-1}, appropriate to E being expressed in the same units. Experiment A refers to the analysis of the reverse characteristics, Experiment B to that of the temperature characteristics analyzed in Sec. B. The experimental value given for Ge is that obtained after making the correction to the field given by Eq. (12). Without this correction the result is 3.5×10^7 units.

where m_C^*, m_V^* refer to electrons and holes, respectively, and are given by expressions of the form,

$$m^{-1} = \sum_i [(\cos^2\gamma_i)/m_i].$$

Here, the quantities, m_i, are the principal values of the tensor m_{ik}^{-1} and γ_i are the angles between the directions of the field and the principal axes of the tensor, which in general, do not coincide with the principal crystal axes. As an approximation, therefore, m_C^* and m_V^* were taken as the density of states masses, defined by

$$m_C^* = \nu^{\frac{2}{3}}(m_L m_T^2)^{\frac{1}{3}},$$

and

$$m_V^* = [m_l^{\frac{3}{2}} + m_h^{\frac{3}{2}}]^{\frac{2}{3}},$$

where ν is the valley degeneracy, m_L and m_T are the longitudinal and transverse electron masses, m_l and m_h are the light and heavy hole masses, and the holes in the spin orbit split-off band are disregarded. Using the values for m_L, m_T, m_l, and m_h given by Herman,[13] we obtain for the reduced effective mass,

$$m^* = 0.36 m_0 \text{ for Si},$$

and

$$m^* = 0.19 m_0 \text{ for Ge},$$

where m_0 is the free electron mass. With these mass values, the calculated values for $\alpha\epsilon^{\frac{1}{2}}$ agree with the experimental values to within a factor of 2 for both germanium and silicon. Considering the experimental and theoretical uncertainties, this is felt to be good agreement. For the calculated values to agree with the experimental ones would require m^* to be about 0.95 m_0 for Si and 0.46 m_0 for Ge, and both these values seem rather high as it is to be expected that the lightest carriers will completely dominate the tunneling probability. A further fact is that the field is not uniform so that the effective field may be significantly different from E_M. However, a calculation of the barrier transparency of a step junction using the WKB method[14] shows that the effective field is given by Eq. (9) to a good approximation. On the other hand, there will be an effect on the field due to the high degeneracy of the p side. This results in the electron actually

[13] F. Herman, Proc. Inst. Radio Engrs. 43, 1703 (1955).

[14] See, for example, N. F. Mott and J. N. Sneddon, *Wave Mechanics and Its Application* (Oxford, Clarendon Press, 1948), Chap. I.

leaving the valence band at a point in the junction where the field is somewhat less than that given by Eq. (9). In particular, the effective field,

$$E_{\text{eff}}=E_M[1-(V_{ip}/V_i)]^{\frac{1}{2}}, \tag{12}$$

for $V_a \ll V_i$. For silicon, this correction is negligible but it makes $E_{\text{eff}}=0.71E_M$ for germanium. This, in turn, brings the experimental value, now 2.8×10^7 volts cm^{-1}, in better agreement with the calculated value.

Though the experimental and theoretical values of $[\alpha\epsilon^{\frac{3}{2}}]$ are gratifyingly close there does seem to be a systematic discrepancy between their values. This discrepancy can be expressed by

$$[\alpha\epsilon^{\frac{3}{2}}]_{\text{expt.}}=f[\alpha\epsilon^{\frac{3}{2}}]_{\text{theory}},$$

where f is a number equal to about 1.6 for both silicon and germanium. The origin of the discrepancy factor, f, is not understood though it may be in the estimates of the effective built-in potentials or the width constants. Regarding the latter, it is quite conceivable that owing to the randomness in the impurity distribution, the actual junction width will vary appreciably from place to place in the junction. As it is to be expected that field emission will take place preferably at the narrowest parts of the junction, the effective width could, therefore, be significantly less than the measured (average) width.

It should also be noted that the plots in Fig. 6 generally give a value for p of slightly less than unity, a value that is not physically unreasonable.

B. Analysis of Temperature Characteristics

Figure 5 demonstrates that the critical field dependence on temperature for germanium is a monotonic function while that for silicon shows structure. In agreement with the conclusions of Holonyak et al.[10] it is reasonable to interpret these results, qualitatively, as evidence of the dominance of indirect transitions in silicon and direct transitions in germanium. It is easy to verify that in both materials, the cause of the temperature variation of the critical voltage at constant current is almost entirely contained in the temperature dependence of the tunneling probability. It should be emphasized that in all the temperature characteristics, the values of V_a were considerably greater than kT (except where V_a drops rapidly at the highest temperatures) so that we are concerned, essentially, with only the tunnel current from the valence band to the conduction band. As the temperature dependence is quite different depending on whether there is, or is not, phonon cooperation, the two cases will be considered separately.

(a) *Germanium*

To maintain constant current requires that the tunneling probability remains essentially constant. Thus,

$$\alpha\epsilon^{\frac{3}{2}}/E_{\text{eff}}=\alpha\epsilon^{\frac{3}{2}}W_1/1.42(V_a+V_i)^{\frac{1}{2}}=k, \tag{13}$$

where k is a constant. It will be supposed that as T is varied, W_1, V_{in}, and V_{ip} stay constant, assumptions that are probably valid down to very low temperatures in these degenerate materials. Thus, the temperature variation of V_a reflects that of α through the effective mass and ϵ, the direct band gap. Note that ϵ also appears in V_i. Hence plotting $(V_a+\epsilon)$, where ϵ is expressed in volts, against ϵ^3 (or ϵ^4 if the temperature variation of the effective mass is to be included in an approximate fashion) should yield a straight line. For the temperature dependence of ϵ, six points have been given by Macfarlane, McLean, and Quarrington.[15] Figure 7 shows $(V_a+\epsilon)$ plotted against ϵ^4 and it is apparent that a fairly good straight line results. This is taken to confirm that the temperature variation of V_a in germanium reflects, primarily, that of the direct band gap.

(b) *Silicon*

Considering the effect of one phonon energy only, Eq. (6) predicts that the temperature dependence of V_a should be determined, primarily, by

$$P_E[N(P_A/P_E)+1]=k', \tag{14}$$

where k' is a constant. The ratio

$$P_A/P_E \simeq \exp 3\alpha\epsilon^{\frac{1}{2}}\hbar\omega/E, \tag{15}$$

can be regarded as essentially constant. (It is estimated that between 0°K and 400°K, P_A/P_E decreases by only 5%.)

By inspection of Eq. (14) we see that the temperature characteristics can be divided into two regions.

(i) At low temperatures, N is very small, so that the constant current condition becomes:

$$P_E=k'. \tag{16}$$

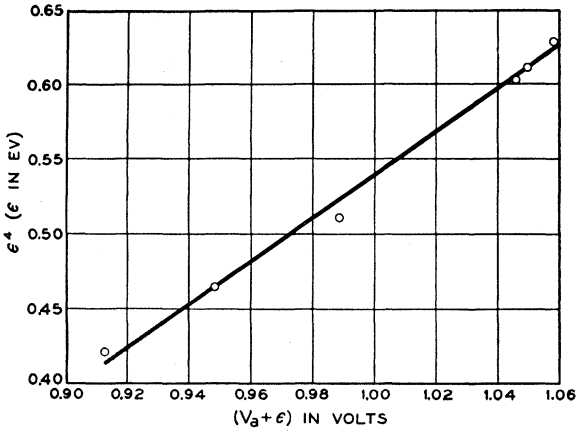

FIG. 7. Plot showing the correspondence between the temperature variation of the energy gap, ϵ, and the critical voltage, V_a, for the germanium junctions.

[15] G. G. Macfarlane, T. P. McLean, and J. E. Quarrington, Proc. Phys. Soc. (London) **71**, 863 (1958).

(ii) At high temperatures, $N(P_A/P_E)\gg 1$, so that the constant current condition becomes:

$$NP_A=k'. \tag{17}$$

There will be a transition temperature determined by

$$N(P_A/P_E)\simeq 1. \tag{18}$$

Equation (16) is equivalent to Eq. (13) in analyzing the germanium data. It predicts that the temperature dependence of V_a will reflect that of ϵ (the indirect band gap) at low temperatures and, therefore, will show only a very slight variation with temperature, in accordance with the experimental data of Fig. 5.

By substituting the energy 0.019 ev, appropriate to the transverse acoustic phonon[16] (the lowest energy phonon) into Eq. (16) one obtains a transition temperature of about 90°K, using the experimental value for $\alpha\epsilon^{\frac{3}{2}}$. This is clearly quite consistent with the experimentally determined V_a versus T variation.

At sufficiently high temperatures, $kT\gg\hbar\omega$, and Eq. (17) becomes, $TP_A=$ constant, so that V_a will again be relatively slowly dependent on the temperature. This could correspond to the leveling portion of the curve at temperatures greater than about 300°K. To understand the experimental results through the transition region in more detail, though, it is better to use the less approximate relation in (6), which yields the condition

$$\ln(\lambda N+1)-\alpha(\epsilon+\hbar\omega)^{\frac{1}{2}}W_1/2(V_a+V_i)^{\frac{1}{2}}=\text{constant},$$

where

$$\lambda=(P_A/P_E)+1\simeq 5,$$

when the experimental value for $\alpha\epsilon^{\frac{3}{2}}$ is used. The temperature dependence of N is known from the relation

$$N=N_0[\exp(\hbar\omega/kT)-1]^{-1}.$$

Noting again that $V_a\ll V_i$, neglecting $\hbar\omega$ relative to ϵ, and noting that $V_i=V_{ip}+V_{in}+\epsilon\simeq\epsilon$, where, in this relation, ϵ is expressed in volts, we have

$$\ln(\lambda N+1)-(\alpha W_1/4)(2\epsilon-V_a)=\text{constant}.$$

Thus, a plot of $\ln(\lambda N+1)$ against $(2\epsilon-V_a)$ should be a straight line whose gradient should yield an independent value for α. Such plots, for three different silicon junctions, all made from 0.003 ohm-cm resistivity material, are shown in Fig. 8. All three plots are very similar and show a slight curvature rather than following straight lines. If the best straight lines are drawn, they all have similar slopes and yield a value of 7.2×10^7 volts cm^{-1} for $\alpha\epsilon^{\frac{3}{2}}$. This is in very good agreement with the value obtained independently from the reverse characteristics.

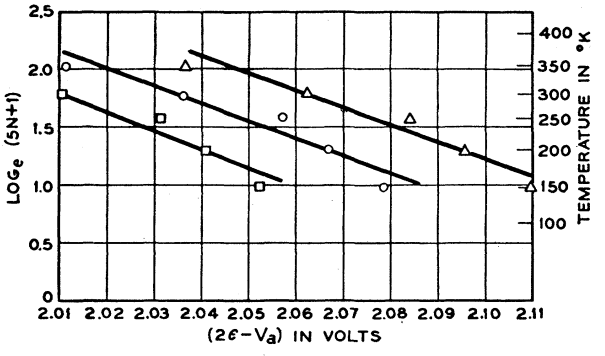

FIG. 8. Plots deduced from the temperature variation of the critical voltage for three different 0.003 ohm-cm silicon units. The forms of the plots are those given by the theory that takes into account the influence of the phonon population on the indirect tunneling probability.

[16] J. R. Haynes (private communication).

The cause of the curvature in the plots is not properly understood. The approximation involved in the analysis appear to be justified and insufficient to explain the observed departures. On the other hand, the analysis has been based on phonons of one energy only. It is possible that at higher temperatures especially, significant contributions to the current come from phonons with higher energies or even multiphonon processes.

The cause of the final plunge of V_a to zero at about 620°K is likewise not understood at present. It cannot be caused by the crystals becoming intrinsic—for these heavily doped materials this would require a temperature of 1200°K or more. It is of interest, therefore, to examine the consequences of adding a second term to Eq. (14) so that the effects of two phonon energies, $\hbar\omega_1$ and $\hbar\omega_2$, are now included. If the two give equally weighted contributions to the current, then we can determine the temperature at which the two phonons give comparable contributions to the current from the relation:

$$(P_{E1}/P_{E2})(\lambda_1N_1+1)\simeq(\lambda_2N_2+1), \tag{19}$$

where the subscripts 1 and 2 refer to the two phonon energies. Putting $\hbar\omega_1=0.019$ ev (the transverse acoustic phonon, TA) and $\hbar\omega_2=0.058$ ev (the transverse optical phonon, TO, according to Haynes[16]), we have $\lambda_1=5.7$, $\lambda_2=110$, and $(P_{E1}/P_{E2})=5$. Graphically, it was found that equality (19) held at a temperature of about 700°K. This is sufficiently close to the temperature at which V_a plunges towards zero to suggest that there is some basis to believing the turn over at 600°K is due to the influence of the second phonon. However, in optical work, Haynes has found that recombination transitions involving the TO phonon are about 29 times more common than transitions involving TA phonons.[16] While this would seem to compare with the experimental result that the second drop in V_a is very much greater than the first, it would also mean, if true, that the TO phonon contribution to the current should dominate the TA phonon contribution at all but the lowest temperatures. It is true that the analysis of the reverse characteristics and the temperature characteristics does not depend very critically on the actual phonon energy involved, except for the condition given in Eq. (18),

but nevertheless, the work of Holonyak et al. and Esaki shows that the TA phonon is operative in the tunneling transitions and that there is little reason to disregard it in favor of the TO phonon. The question of the final fall of V_a, therefore, remains unresolved as the data are not yet sufficient to differentiate between various models invoking phonons of different energies or multiphonon processes. On the other hand, it is possible that V_a drops at about 600°K because the normal reverse bias diode current becomes dominant over the tunneling current.

CONCLUSIONS

The implications of the theoretical form of the tunneling probability, particularly as modified by Keldysh to take into account phonon cooperation, have been examined experimentally. It is felt that relatively good qualitative and quantitative agreement has been reached between theory and the measured dependence of the tunnel current on the field in both germanium and silicon, where indirect transitions are dominant, the temperature variation of the critical field up to about 500°K is also in accord with the theory which includes phonon cooperation by a single phonon (at 0.019 ev). However, the analysis is not critically dependent on the phonon energy so the selection of this particular phonon energy rather than a larger one or a combination of phonons, should be regarded to some extent as tentative at this stage. In germanium, where direct transitions are dominant, the temperature variation of the critical field can be almost entirely accounted for by the temperature variation of the direct energy gap.

If is felt that temperature studies of the sort described in this paper afford a useful general method for determining whether direct or indirect transitions are responsible for tunneling in any given junction.

Note added in proof. After this paper was submitted for publication, a report was published by J. V. Morgan and E. O. Kane, Phys. Rev. Letters 3, 466 (1959), which described the observation of the onset of direct tunneling transitions in narrow germanium *p-n* junctions as the reverse bias was increased. The *p*-type material used by these authors was much more lightly doped (8×10^{18} carriers cm^{-3}) than that used by the present authors so that the built-in potential was very close to the magnitude of the band gap. Thus, only indirect transitions could be accommodated at zero applied bias. As the reverse bias was increased, the point was eventually reached where the total potential across the junction was sufficient to promote direct tunneling transitions, and this manifested itself as the onset of a current that rose more steeply with the bias. The germanium junction described in the present paper had a built-in potential of 1.46 volts so that it was certainly energetically possible for tunneling to be by direct transitions for all values of the reverse bias.