backscattered electrons is roughly twice as large as that of the primaries. Furthermore, since the backscattered electrons escape with a cosine distribution in angle, the average path lengths of the backscattered electrons in the secondary escape zone can be shown to be twice the zone depth. Therefore, the ratio between the rate of energy dissipation of the backscattered and incident electrons near the surface is roughly 4. A more accurate estimate using the known energy distribution of backscattered electrons as a weight function leads to a ratio of 4.3 at $E_p=10$ kev. This is in reasonable agreement with the experimentally observed value of $\beta=4.9$.[15]

Thus, both theory and experiment indicate that even for materials of low atomic number, the backscattered electrons make an important contribution to secondary electron emission from solids. Therefore, as pointed out by Sternglass[1] and more recently by Dekker and Lye,[16] any theory of the secondary electron emission under electron bombardment must take into account the diffusion and backscattering of the incident electrons.

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Dr. E. J. Sternglass for helpful discussions and to Mr. R. Matta for preparing the thin films.

[15] It should be pointed out that if the average energy of the backscattered electrons were a constant fraction of the initial energy, the ratio between the rate of energy dissipation of the backscattered and incident electrons would increase somewhat with increasing initial energy due to the logarithmic term in Eq. (2). However, as shown in reference 4, the average energy of the backscattered electrons also increases continuously relative to the incident energy with increasing E_p. As a result, the ratio of the rate of energy dissipation for backscattered and incident electrons changes less rapidly than would be expected on the basis of Bethe's formula alone. Since the backscattering coefficient of aluminum does not vary with the incident energy, it follows that the ratio between the energy dissipation density or secondary electron production of incident and backscattered electrons is about constant with energy, in agreement with the results obtained for the two different aluminum films.

[16] A. J Dekker and R. G. Lye, Wright Air Development Division Technical Report WADD 57-760 (unpublished).

PHYSICAL REVIEW VOLUME 121, NUMBER 3 FEBRUARY 1, 1961

Excess Tunnel Current in Silicon Esaki Junctions

A. G. Chynoweth, W. L. Feldmann, and R. A. Logan
Bell Telephone Laboratories, Murray Hill, New Jersey

(Received June 6, 1960; revised manuscript received October 20, 1960)

At low forward biases, a high current flows in Esaki junctions due to band-to-band tunnelling. At sufficiently high biases the current flows by normal forward injection. Between these two bias ranges, the current is unexpectedly high and has been called the excess current. A comprehensive experimental study has been made of this excess current in silicon junctions. It is shown that the properties of the excess current observed so far can be accounted for by a mechanism originally suggested by Yajima and Esaki, in which carriers tunnel by way of energy states within the forbidden gap. Based on this model, the following expression for the excess current, I_x, is proposed:

$$I_x \sim D_x \exp\{-(\alpha_x W_1 \epsilon^{\frac{1}{2}}/2)[\epsilon - eV_x + 0.6e(V_n+V_p)]\},$$

where D_x is the density of states in the forbidden gap at an energy related to the forward bias, V_x, and the Fermi energies on the n and p sides are V_n and V_p, respectively, e is the electron charge, ϵ is the energy gap, W_1 is the junction width constant, and α_x is a constant containing a reduced effective mass, m_x. This formula describes the observed dependence of I_x (i) on D_x, observed by introducing states associated with electron bombardment, (ii) on ϵ, studied by the temperature variation of the diode characteristics, (iii) on V_x, verified from semilogarithmic plots of the forward characteristics, and (iv) on W_1, tested by using junctions of different widths. From these experiments, $m_x=0.3m_0$ to within a factor of 2.

The origins of the states in the band gap are not known for certain though they are most likely the band edge tails inherent to heavily doped semiconductors. It is probable that the tunnelling-via-local-states model for the excess current in silicon is applicable to excess currents in other materials.

INTRODUCTION

THE occurrence of so-called excess current in Esaki junctions[1] is now well known.[2] It occurs at forward biases in the range where the electrons in the degenerate donor levels in the n side have been raised to energies greater than those of the degenerate acceptor levels on the p side. Ideally, tunnelling of electrons from the conduction band to the valence band in a single energy-conserving transition should then be impossible and only the normal diode current due to the forward injection of minority carriers should flow. In practice, as Yajima and Esaki first noted,[2] the actual current at such biases is considerably in excess of the normal diode current; hence the term, excess current.

It was apparent that the excess current is primarily a tunnelling process since its behavior generally parallels that of the peak current; the peak and excess currents exhibit much the same dependence on pressure,[3] on temperature,[4] and on the donor and acceptor concentra-

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

[2] T. Yajima and L. Esaki, J. Phys Soc. Japan **13**, 1281 (1958).

[3] S. L. Miller, M. I. Nathan, and A. C. Smith, Phys. Rev. Letters 4, 60 (1960).

[4] R. A. Logan and A. G. Chynoweth, Bull. Am. Phys. Soc. **5**, 160 (1960).

tions that make up the junction. (The last fact comes from our studies of silicon Esaki junctions where it was found that the ratio of the peak and excess currents was very insensitive to the doping concentration even though both currents could be changed by orders of magnitude in this way.)

As interband tunnelling transitions are energetically impossible at biases in the excess current range, a mechanism has to be found whereby the electrons can dissipate their excess energy either "during" the tunnelling transition or separately. In the former it has been suggested that electrons could lose energy by emitting photons, or phonons, or plasmons, or in Auger processes. Kane has considered several of these hypotheses from a theoretical standpoint[5–7] and has concluded that they are unlikely to account for any appreciable fraction of the excess current. Further, the photon emission hypothesis has been tested experimentally. Using lead sulfide and cadmium telluride photocells, no light could be detected from silicon Esaki junctions in the appropriate wavelength range even though the detecting sensitivity was many orders of magnitude greater than that which would be needed if a large fraction of the excess current were carried by the photon-emitting mechanism.

Instead, there is now a growing amount of evidence that the excess current is caused by electrons tunnelling not completely through the energy gap, but only part of the way, making use of more or less localized imperfection energy levels present in the energy gap. This mechanism was first suggested by Yajima and Esaki[2] and, subsequently, several authors have obtained strong confirmation of it by showing that the magnitude of the excess current can be altered by deliberately changing the imperfection content of the crystal, either by suitable doping[8] or by radiation damage.[4,9,10]

The present paper is concerned primarily with the properties of the excess current in silicon Esaki junctions. The dependence of the excess current upon the imperfection density has been studied using electron irradiation and these experiments are complemented by studies of the excess current as a function of bias, of temperature, and of the concentration of the donors and acceptors used for forming the junction. It was found that the interpretation of the experimental results was aided considerably by a simple phenomenological model for the excess current which will be presented in the next section.

[5] E. O. Kane, Bull. Am. Phys. Soc. **5**, 160 (1960).
[6] E. O. Kane, International Conference on Semiconductor Physics, Prague, 1960 (unpublished).
[7] E. O. Kane (private communication).
[8] C. T. Sah, IRE Solid-State Device Research Conference, Pittsburgh, June, 1960 (unpublished).
[9] T. A. Longo, Bull. Am. Phys. Soc. **5**, 160 (1960).
[10] R. R. Blair and J. W. Easley, J. Appl. Phys. **31**, 1772 (1960).

MODEL FOR EXCESS CURRENT

Figure 1 shows some examples of possible routes whereby electrons can pass via local levels from the conduction band on the *n* side to the valence band on the *p* side when the junction is biased into the excess-current region. An electron starting at *C* in the conduction band might tunnel to an appropriate local level at *A* from which it could then drop down to the valence band, *V*. Alternatively, the electron could drop down from *C* to an empty level at *B* from which it could tunnel to *V*. A third variant is a route such as *CABV*, where the electron dissipates its excess energy in a process which could be called impurity band conduction between *A* and *B*. A fourth route which should also be included is a staircase route from *C* to *V* which consists of a series of tunnelling transitions between local levels together with a series of vertical steps in which the electron loses energy by transferring from one level to another, such a process being possible if the concentration of intermediate levels is sufficiently high. The route *CBV* can be regarded as the basic mechanism, the other routes being simply more complicated modifications which are not likely, of themselves, to produce radically different behavior. In this paper, it will be shown that the simple model appears adequate to account for the observed behavior of the excess current in silicon Esaki junctions.

It has long been regarded as possible for electrons to tunnel from impurity levels within the forbidden gap to the conduction band; this is called field-ionization. This is conceptually easy to accept as the electrons then move in the direction appropriate to the electric field. However, if one regards tunnelling as simply a barrier transparency problem, the direction of the field is immaterial; the role of the field can be thought of as simply to determine the barrier width. In particular, the electrons may just as readily tunnel against the electric field as with it and indeed, the peak current in Esaki junctions is proof that electrons may tunnel against the electric field.

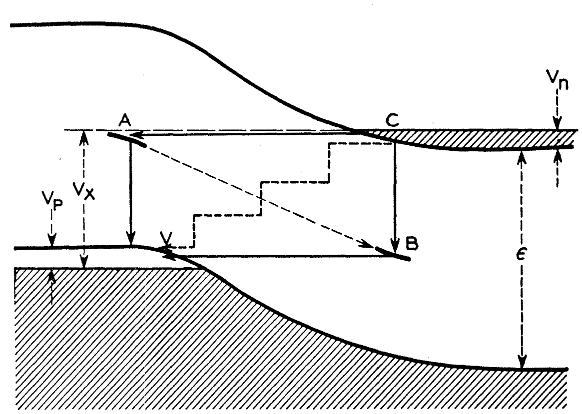

FIG. 1. Diagram illustrating proposed mechanisms for the excess current flow.

A theoretical treatment of the conventional field ionization process in which the electron tunnels from a discrete state to the conduction band has been given by Franz.[11] He finds that the tunnelling probability factor has the same form as that for interband tunnelling transitions except that the energy gap is replaced by the ionization energy of the impurity, namely, the height of the barrier through which the electron must tunnel. (There are also minor changes in the numerical factor contained in the exponent of the tunnelling probability.) By direct analogy, it is reasonable to employ a similar tunnelling probability factor for the case in which the electron tunnels to the valence band, except that here the height of the energy barrier through which the electron tunnels is now the difference between the energy gap and the ionization energy of the impurity level.

Let $V_x \equiv$ the forward bias, in volts, in the bias range where the excess current, I_x, is dominant; $V_n \equiv$ the potential difference (in volts) between the Fermi level on the n-type side and the bottom of the conduction band; $V_p \equiv$ the potential difference (in volts) between the Fermi level on the p-type side and the top of the valence band; $\epsilon \equiv$ the forbidden energy gap, in ev, of relatively pure material; and $e \equiv$ the electron charge.

Let the junction be at a bias V_x, and consider an electron making a tunnelling transition from B to V in Fig. 1. The energy gap through which it must tunnel, ϵ_x, is then given by

$$\epsilon_x \simeq \epsilon - eV_x + e(V_p + V_n), \quad (1)$$

assuming that the electron ends up near the top of the valence band. The potential difference across the junction promoting this transition is

$$V_i - V_x = [\epsilon/e + 0.6(V_n + V_p) - V_x], \quad (2)$$

where V_i is the built-in potential. For a step junction the maximum field is given by

$$E = 2(V_i - V_x)^{\frac{1}{2}}/W_1, \quad (3)$$

where W_1 is the width constant of the junction, the junction width for unity potential across it given by

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

where K is the dielectric constant and p and n are the acceptor and donor concentrations, respectively.

The tunnelling probability, P_x, for the electron on the level at B will be, following Franz,[11]

$$P_x = \exp(-\alpha_x \epsilon_x^{\frac{3}{2}}/E), \quad (4)$$

where α_x is defined by

$$\alpha_x = [4(2m_x)^{\frac{1}{2}}/3e\hbar]\theta. \quad (5)$$

In Eq. (5), $h(=2\pi\hbar)$ is Planck's constant and θ is a numerical factor of the order of unity. By analogy with the case of interband tunnelling[12] it may be reasonable to regard the tunnelling mass, m_x, as a reduced mass, the dominant component of which will be the hole mass in the valence band. It will be assumed that Eq. (4) is sufficiently accurate for present purposes, though strictly it applies only to the case of a uniform electric field.

Let the volume density of occupied levels at an energy, ϵ_x, above the top of the valence band (such as levels B) be D_x. It has already been stated that the excess current behaves in a broadly similar way to the peak current, implying that the tunnelling transition, BV, is rate-controlling rather than the replenishment mechanism along the route CB. Thus, the excess current will be given by

$$I_x = AD_xP_x, \quad (6)$$

where the reasonable assumption is made that the excess current will vary predominantly with the parameters in the exponent of P_x rather than with those in the factor A. Substituting expressions (1), (2), (3), and (4) into (6), it is therefore possible to write the following approximate expression for the excess current:

$$\ln I_x = \ln A + \ln D_x - (\alpha_x W_1 e^{\frac{1}{2}}/2)[\epsilon - eV_x + e(V_p + V_n)]. \quad (7)$$

This expression is susceptible to detailed experimental testing since the parameters D_x, W_1, ϵ, and V_x can be varied independently and their effect on I_x studied. Such experiments will be described in the next section.

EXPERIMENTAL RESULTS

A. Junction Fabrication

The silicon junctions were made by alloying boron-rich aluminum wires into n-type samples of resistivities ranging between 0.0018 and 0.0006 ohm-cm using a strip heater with a quick heat cycle. The resulting silicon diodes had Esaki characteristics which generally yielded values between 2 and 3 for the ratio of the peak current to the valley current at room temperature. In some experiments where germanium Esaki junctions were used, junctions with roughly comparable peak-to-valley current ratios were obtained by alloying aluminum wire into n-type material of 0.0006 ohm-cm resistivity. Alloying to germanium was carried out by means of an electrical discharge through the aluminum wire and the crystal while the two were held in contact.

The acceptor concentration obtained in the silicon junction was estimated to be 3×10^{20} cm^{-3}, using a distribution coefficient for boron of about unity and taking into account the solubility of silicon in the alloy as given in the Al-Si phase diagrams. It is evident that there could be considerable error in this estimate. The Fermi energy on the p side was estimated to be 0.18 ev and for

[11] W. Franz, *Handbuch der Physik*, edited by S. Flügge (Springer-Verlag, Berlin, 1956), Vol. 17, p. 155.

[12] L. V. Keldysh, J. Exptl. Theoret. Phys. (U.S.S.R.) **34**, 962 (1958) [translation: Soviet Phys.—JETP **7**, 665 (1958)].

the highest donor concentrations used (2.1×10^{20} cm^{-3}) the Fermi energy was 0.06 ev on the n side.

B. Excess Current Versus Forward Bias

Figure 2 shows a semilogarithmic plot of the forward current versus forward bias for a silicon Esaki junction typical of those used in the present study. In the range between biases corresponding to the peak current, I_P, and the valley current, I_V, the junction exhibits a negative resistance. At sufficiently high forward biases, the main component of the current becomes the conventional forward minority-carrier injection current and on the semilogarithmic plot, the characteristic enters a line of slope close to e/kT, as is to be expected from the usual rectifier equation. Here, e is the electron charge, k is Boltzmann's constant, and T is the absolute temperature. The fact that the slopes were close to e/kT shows that series resistance effects were not important in these junctions. By extrapolating the normal diode current to lower biases, it is clear that in the bias range between the valley point and the point where the normal diode current becomes dominant (at about 0.7 volt at room temperature) the actual junction current can be several orders of magnitude higher than the conventional diode current. This is the excess current which forms the subject of this paper. It is clear that the dependence of I_x on bias is very close to exponential, which is in accordance with Eq. (7) (a similar dependence has also been observed in germanium junctions by Yajima and Esaki[2]). Furthermore, the lack of any structure in the semilogarithmic plot implies that if the proposed model for the excess current is correct, there must be a more or less continuous distribution of B states through the energy gap. In particular, the density of these states, D_x, must either be constant or vary exponentially with the energy difference, eV_x.

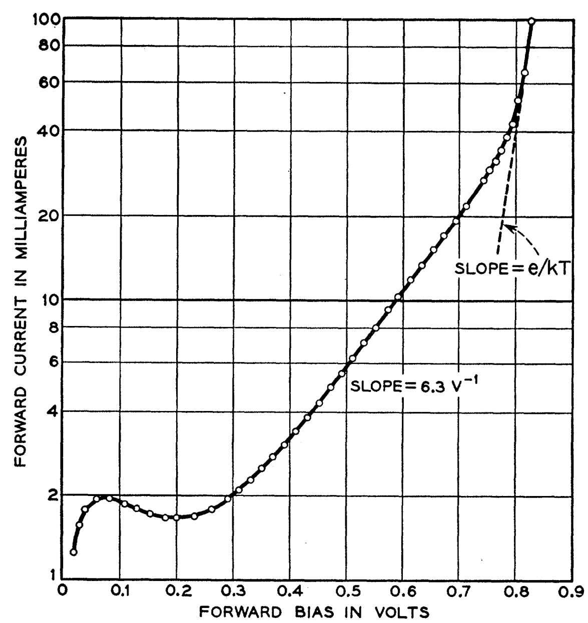

FIG. 2. Semilogarithmic plot of the forward characteristic of a silicon Esaki junction possessing a low peak current density.

Figure 3 displays two more semilogarithmic plots of the forward characteristics of a junction, obtained at 300°K and 4°K. It will be seen that the linear excess current range becomes much more prominant at low temperatures, due to the decreased normal diode current. However, there is very little change in the slope of this characteristic, as noted above. In particular, it is to be noted that in the silicon junction used in these experiments, there is no evidence of structure in the semilogarithmic plots even at 4°K. This is in contrast to the work of Esaki who observed broad maxima in the excess current versus bias curves at low temperatures in silicon junctions.[13]

C. Excess Current Versus Imperfection Density

It has been shown that both electron and neutron irradiations result in an increase in the excess current

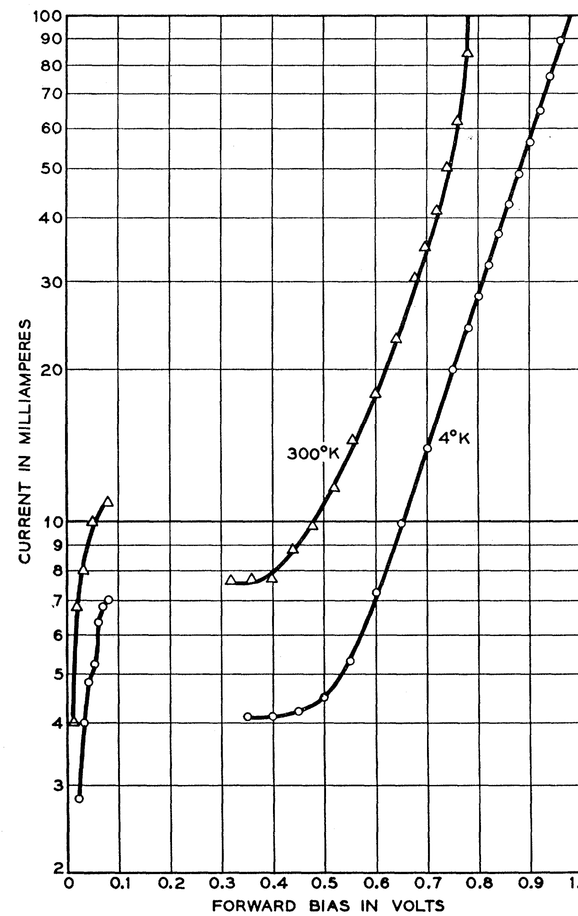

FIG. 3. Semilogarithmic plots of the forward characteristics of a silicon Esaki junction at 300°K and 4°K.

[13] L. Esaki and Y. Miyahara, Solid-State Electron. 1, 13 (1960).

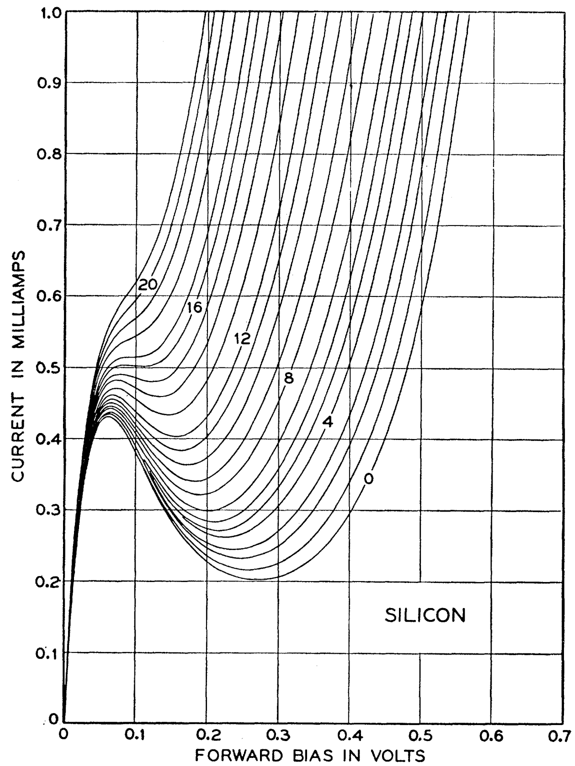

FIG. 4. Changes in the Esaki characteristic of a silicon Esaki diode, rather similar to that of Fig. 1, brought about by bombardment with 800-kev electrons. Curve 0 refers to the junction before bombardment and the series of curves up to curve 21 represent the junction characteristics at various stages of the bombardment.

in germanium and silicon Esaki junctions.[9,10] To investigate this effect further, several Si Esaki junctions were subjected, in vacuum, to a beam of 800-kev bombarding electrons from a van der Graaff generator. The samples were soldered to a copper block for a heat sink and no effects due to heating of the samples could be detected. The bombardments were carried out at room temperature and they were halted at frequent intervals while the forward bias characteristics were traced on an X-Y recorder. Figure 4, which applies to a relatively low current density junction, shows a typical set of data obtained in this way and it is immediately obvious that the main effect of the bombardment is to increase the excess current while leaving the peak current relatively unchanged. This is in agreement with the findings of Blair and Easley[10] and is consistent with the proposed model if the volume density of the local energy levels giving rise to the excess current is much less than the impurity concentration which determines the number of electrons available for band-to-band tunnelling. (Note that the curves do not extend into the bias range where the normal diode current becomes important. This can be seen by comparison with Fig. 2, for example.) A variety of junctions were subjected to the electron bombardments, their initial excess current densities at 0.3 volt forward bias covering a range of two orders of magnitude. Nevertheless, all the junctions yielded data basically similar to those shown in Fig. 4.

Figure 5 shows semilogarithmic plots of some of the curves of Fig. 3 at various stages of the bombardment. It is clear that prior to the bombardment, a straight line is obtained in the excess current range, as in Fig. 2, in accordance with Eq. (7) and that the slope of this line does not change very much during the bombardment, to within the limits of experimental error. Referring to Eq. (7), this implies that neither α_x nor W_1 are significantly affected by the bombardment. Relatively little change could be expected in α_x as the only relevant quantity it contains is the reduced effective mass. The conclusion that W_1 is practically constant is consistent with the observation that I_P does not change appreciably, implying that in these experiments the radiation damage was not sufficient to alter appreciably the free carrier density.

Before analyzing the data further, it is necessary to elaborate on Eq. (7) so as to take into account the effects of changes in D_x brought about by the bombardment. Let

$$D_x = D_{0x} + D_{Bx},$$

where D_{0x} is the initial level density at a position in the energy gap corresponding to a bias V_x, and D_{Bx} is the

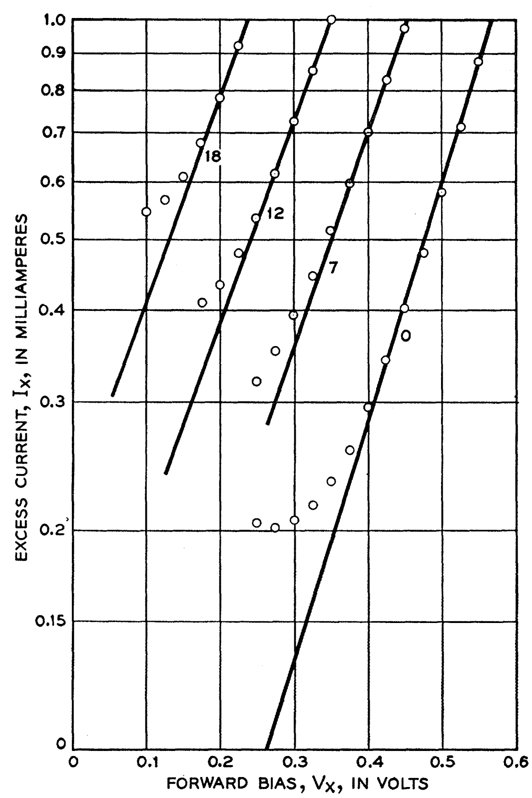

FIG. 5. Semilogarithmic plots of curves 0, 7, 12, and 18 of Fig. 3.

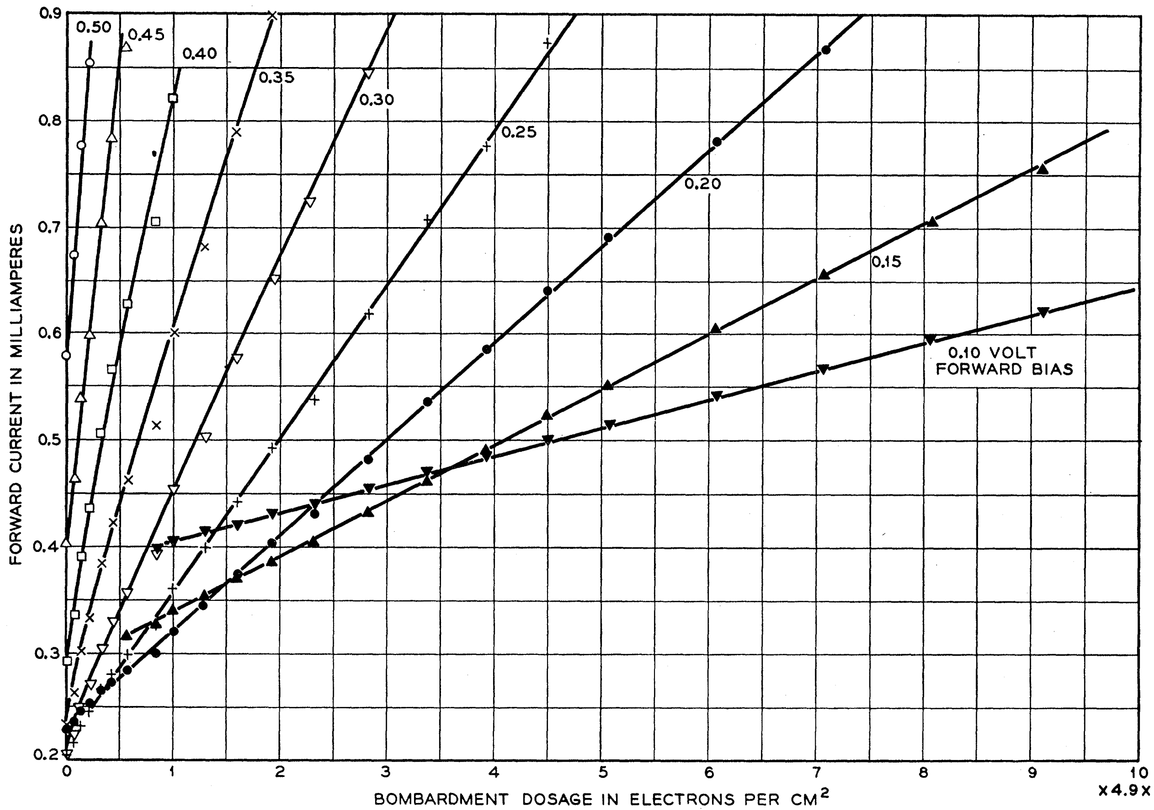

FIG. 6. Curves of total current versus bombardment dose for various values of the applied bias, obtained by interpolation of the curves of Fig. 4.

level density at the same energy introduced by the bombardment. Then

$$D_{Bx}=\varphi t\rho_x,$$

where φ is the electron flux, t is the bombardment time, and ρ_x is the level density at the energy corresponding to V_x introduced by a unit bombardment dose. Then

$$I_x=B(D_{0x}+\varphi t\rho_x)\exp(\alpha_x W_1 e^{\frac{1}{2}} V_x/2), \qquad (8)$$

where

$$B=A\exp(-\alpha_x W_1 e^{\frac{1}{2}}/2)[\epsilon+e(V_p+V_n)].$$

At a given bias, a plot of I_x against (φt) should be a straight line of gradient

$$S_x=B\rho_x\exp(\alpha_x W_1 e^{\frac{1}{2}} V_x/2). \qquad (9)$$

This expression, in turn, predicts that if ln S_x is plotted against V_x, a straight line should result if ρ_x is only slowly dependent on V_x compared with the exponential term. Furthermore, the slope of this line gives another determination of $\alpha_x W_1$ and this can be compared with the value obtained independently from the slopes of the linear regions of the curves in Fig. 5.

By interpolation of the curves of Fig. 4, plots were obtained of I_x (or I at biases lower than those in the excess current range) against the bombardment dose, (φt). These are shown in Fig. 6 and it is apparent that excellent straight lines result at all biases, in agreement with Eqs. (7) and (8). Similar plots made for other junctions followed much the same pattern, though the higher the initial excess current, the greater the bombardment dose required to produce the same relative change in the excess current. It was observed, however, that in those junctions that required the heavier bombardments, the plots of I versus φt, though linear at low doses, tended to fall below the extrapolated linear portions. Such behavior could arise either through saturation effects in the rate of introduction of defects by the bombardment, or more likely, through carrier removal becoming sufficient to cause changes in the tunnelling probability because of the resulting increase in W_1.

It was useful to determine the "doubling-dose," Δ, defined as the total number of electrons per square centimeter that introduces a level density equal to that originally present at a given bias. The doubling dose at 0.3 volt bias was estimated for a variety of junctions from curves such as those shown in Fig. 6 and the results are summarized in Table I. It is clear that the doubling dose increases roughly with the initial excess current, the average doubling dose being 3.7×10^{16} electrons cm^{-2} ma^{-1} for the actual junctions studied. These junctions had roughly equal areas, with an average value of about 3.5×10^{-4} cm^2, so that a more useful way of expressing the doubling dose is that 1.3×10^{16} 800-kev

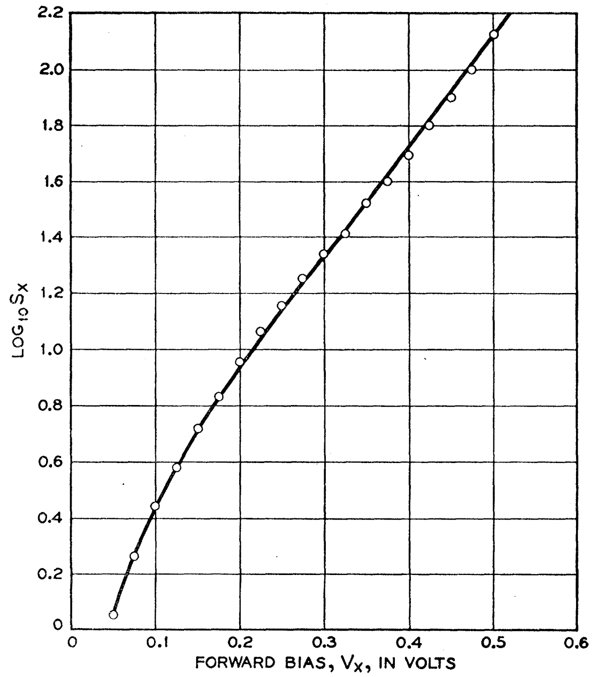

FIG. 7. The logarithms of the slopes of the lines in Fig. 6 plotted against the forward bias.

electrons per square centimeter increase the excess current density by 1 amp cm^{-2}. It is clear that typical junctions with a few milliamperes of excess current are as susceptible to radiation damage as many other semiconductor devices.

The fact that $\Delta(\equiv D_{0x})$ is roughly proportional to the excess current originally present in the junction strongly suggests that the B states are related to the donor and acceptor concentrations used in making the junction rather than to other possible causes which would be expected to be relatively constant from junction to junction. Furthermore, the increase in excess current with doping concentration cannot arise solely from the effect of changes in the junction width and the built-in potential on the tunnelling probability.

It remains to test Eq. (9). In Fig. 7, the logarithms of the slopes of the lines in Fig. 6 are plotted against the forward bias and a relatively good straight line results, as predicted by Eq. (9). The deviation from the straight line at the lowest biases occurs in the region where the band-to-band tunnelling current is very much larger than the excess current as extrapolated from the linear portion of the $\log I_x$ versus V_x plot. In this region it is unrealistic to invoke the actual existence of an excess current.

The relatively good straight line in Fig. 7 indicates that ρ_x is either approximately constant or an exponential function of bias (i.e., position in the energy gap). If we put $\rho_x=\rho_0e^{-\mu V_x}$, where ρ_0 and μ are constants, then the slope of the line in Fig. 7 will be $[-\mu+(\alpha_xW_1e^{\frac{1}{2}}/2)]$. Similarly, the straight lines in Figs. 2 and 4 imply that we can write $D_x=D_0e^{-\lambda V_x}$, where D_0 and λ are constants, and that the slopes of the lines in these figures will then be $[-\lambda+(\alpha_xW_1e^{\frac{1}{2}}/2)]$. In Fig. 4 the forward slope prior to bombardment was 7.8 volt^{-1}, while that of the line in Fig. 6 was 9.1 volt^{-1}. Hence, $\lambda-\mu=1.3$ volt^{-1}, a value that is perhaps comparable with the experimental error. Thus, $\lambda\simeq\mu$ though at this stage it is not possible to draw any conclusions as to their actual magnitudes, which may even be zero. More information on the values of these slopes will be given in the succeeding sections.

It is of interest to report the results of other attempts to modify the excess current in silicon junctions. Sah[8] has been able to increase the excess current by doping the junctions with gold, the gold presumably introducing deep-lying energy levels in the forbidden gap. We have tried unsuccessfully two other methods for increasing the excess current. In the first, two sets of silicon Esaki junctions were made that were similar in all respects except that one set was made using silicon that contained copper at a concentration of 3×10^{17} cm^{-3}, the copper having been introduced by diffusion at 1000°C. Associated with the copper impurities are deep donor and acceptor levels, at 0.24 and 0.49 ev above the valence band.[14] There was no perceptible difference between the Esaki characteristics of the two sets of junctions.

In another experiment, sets of junctions were made using silicon that had been plastically deformed (either by bending or by twisting) until they possessed dislocation densities of as much as 10^8 cm^{-2}. Again there was no noticeable difference between these junctions and the control junctions made using undeformed material. Low-frequency noise measurements throughout the Esaki region also failed to reveal any difference (to within a factor of 2 for the magnitudes of the excess currents) between the junctions with and without dislocations.

D. Excess and Peak Currents Versus Temperature

In general, it was found that the biases at which the peak currents occurred were insensitive to the temperature, thus making I_P a convenient quantity for studying

TABLE I. Doubling dose, Δ, at 0.3 volt bias for several different silicon junctions.

Junction number	Initial I_x (ma)	Δ (electrons cm^{-2})	Δ/I_x (electrons cm^{-2} ma^{-1})
198−1	0.21	10.1×10^{15}	4.8×10^{16}
198−2	0.21	8.3×10^{15}	4.0×10^{16}
529	2.5	9.2×10^{16}	3.7×10^{16}
473	3.1	7.7×10^{16}	2.5×10^{16}
1254	24	8.2×10^{17}	3.4×10^{16}

Average value of $\Delta/I_x=3.7\times10^{16}$ electrons cm^{-2} ma^{-1}
Junction areas $=3.5\times10^{-4}$ cm^2
Average doubling dose $=1.3\times10^{16}$ electrons amp^{-1}

[14] C. B. Collins and R. O. Carlson, Phys. Rev. 108, 1409 (1957).

the effect of temperature on the band-to-band tunnelling. To study the effect of temperature on the excess current tunnelling mechanism, the excess currents at various biases and the valley currents were measured versus temperature for both silicon and germanium Esaki junctions. The peak and excess currents as a function of temperature are shown for Si in Fig. 8 (the excess currents being measured at two biases, 0.35 and 0.60 volt, the former being close to the valley point and the latter being well within the bias range where the excess current dominates at all temperatures). The peak and valley currents for Ge versus temperature are shown in Fig. 9. For Si, as the temperature is increased, I_x shows little change at first but then increases more and more rapidly, this being true at both biases where I_x was measured. For germanium, though the initial temperature-insensitive portion is a little less prominent, I_V shows a very similar trend with temperature. On the other hand, the behavior of the peak currents differs considerably between the two materials. For germanium, the variation of I_P with T is very similar to that of I_V. For silicon, I_P behaves qualitatively much in the same way as I_x until T reaches 100 to 150°K. In this temperature range the rate of increase of I_P with T reaches a maximum, and as T increases further the rate of increase of I_P decreases, in contrast to the behavior of I_X.

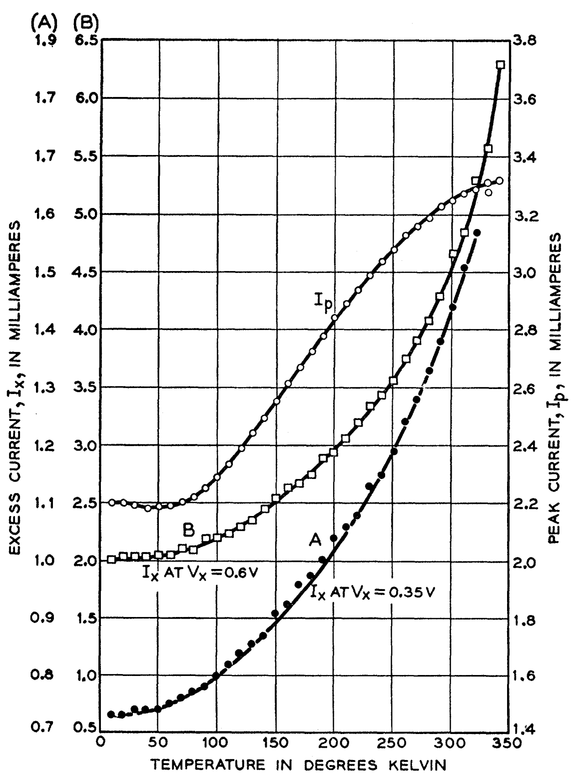

FIG. 8. The peak current and the excess current at two different biases, plotted against temperature for a silicon Esaki junction.

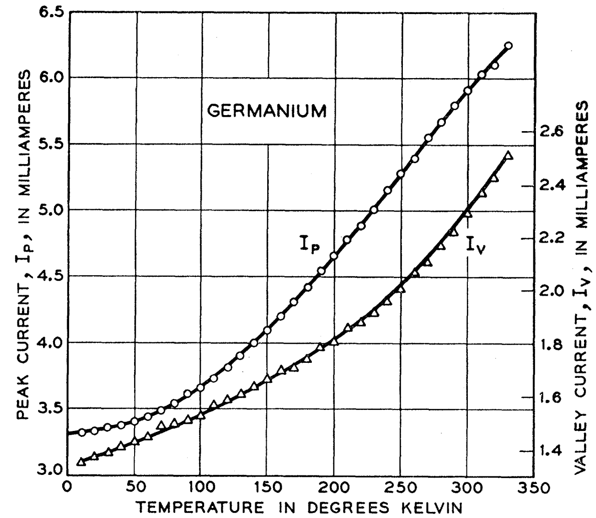

FIG. 9. The peak and the valley currents plotted against temperature for a germanium Esaki junction.

There is a particularly close correspondence between the curves of Figs. 8 and 9 for the peak currents and the results of earlier work[15] in which the temperature variation of the reverse bias required to maintain a constant current in the tunnelling breakdown region was measured. The distinct shape of the I_P versus T curve for Si arises through the necessity for phonon cooperation in the band-to-band tunnelling transitions. In Ge, where it has been shown that usually, phonon cooperation is not necessary for band-to-band tunnelling,[16] there is no S-bend shape to the I_P versus T curve. Instead, the current increases steadily with temperature in a way that mirrors the steady decrease in the direct energy gap as the temperature increases.

The temperature studies can be made more quantative by means of Eq. (7). This equation predicts that as the temperature is varied, the logarithm of the excess, or valley, current should vary as ϵ_I (the indirect energy gap) at constant V_X, if all the other quantities are essentially insensitive to T, and that the slope of this line should be $(-\alpha_x W_1 e^{\frac{1}{2}}/2)$. From studies of the bias dependence of the excess current at different temperatures, it was established that the quantity, $(\alpha_x W_1)$ was almost independent of temperature except for an increase of a few percent that occurred mainly below 20°K. Such an increase could be most plausibly accounted for by a very slight amount of carrier freezeout at the lowest temperatures with an effective ionization energy of the order of 10^{-4} ev. Plots of log I_x versus ϵ_I for Si and Ge are shown in Fig. 10 and the fact that relatively good straight lines result is in qualitative agreement with the form of Eq. (7) and a confirmation of the relative constancy of the quantity, $(\alpha W_1 e^{\frac{1}{2}}/2)$. The actual slopes were 17.3 ev^{-1} at 0.60 volt bias, and

[15] A. G. Chynoweth *et al.*, Phys. Rev. **118**, 425 (1960).
[16] N. Holonyak *et al.*, Phys. Rev. Letters **3**, 167 (1959).

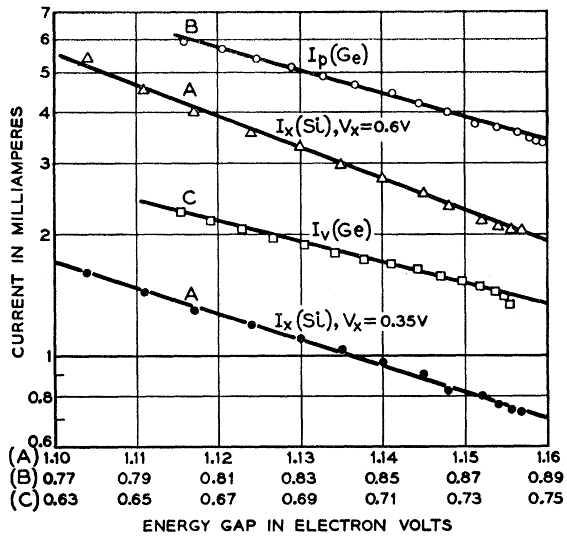

FIG. 10. The logarithm of the peak and valley currents for germanium and the excess currents for silicon plotted against the appropriate energy gaps.

14.6 ev^{-1} at 0.35 volt, for junctions with peak currents of magnitudes not too different from those of the junctions used for the detailed radiation damage studies. Also shown in Fig. 10 is the curve for the valley current versus indirect energy gap in Ge. It also is a relatively good straight line and has a slope of 5.8 ev^{-1}. These temperature studies of the excess and valley currents again confirm that the excess and valley currents in both Si and Ge are primarily tunnelling in origin and that the temperature variation of the excess current is determined to a large extent by that of the energy gap.

It has been shown that if the n side of the junction is formed using As or P, as is the case here, the tunnelling transitions in Ge do not involve phonons.[16] The lack of phonon structure in the curve of I_P versus T shown in Fig. 9, in contrast to that in silicon diodes, corroborates this result. Morgan and Kane[17] have pointed out that the germanium band structure suggests that forward tunnelling is by indirect transitions with momentum conserved by impurity scattering. On the other hand, there exists the possibility that in degenerate germanium, the impurities modify the band structure sufficiently to populate the $k=0$ minimum in the conduction band, making direct transitions possible. The actual energy gap used in the analysis of the temperature variation of I_P is not critical since the difference between the two energy gaps $(\epsilon_D-\epsilon_I)$ is only about 0.15 ev and almost independent of temperature. The field promoting the transition will be

$$2e^{-\frac{1}{2}}[\epsilon_I-eV_P+0.6e(V_p+V_n)]^{\frac{1}{2}}/W_1.$$

If V_p and V_n are sufficiently small, then a plot of $\ln I_P$ versus ϵ_I or ϵ_D should be a straight line of slope $(-\alpha_P W_1 e^{\frac{1}{2}}/2)$, where α_P is an expression, similar to Eq. (5) but appropriate to the band-to-band tunnelling process in Ge. The I_P curve in Fig. 10 is reasonably straight, in agreement with the formulation, and the slope of the straight line has a value, 6.5 ev^{-1}. Though there is appreciable uncertainty as to the value of (V_p+V_n) for the Ge junctions, it was found that the slope of the peak current plot changed by less than 10% in the estimated range of variation of these parameters.

The possibility that the excess current could be caused by simple modifications to normal diode theory was easily removed. It is conceivable that the forward diode characteristic could be of the form

$$I_{sn}\exp[(eV_x/nkT)-1],$$

where the number $n(\geq 1)$ arises, for example, when carrier generation occurs via recombination centers within the space-charge region,[18] or if the potential distribution within the junction is suitably distorted, and I_{sn} is the appropriate saturation current. The dominant factor determining the temperature dependence of I_{sn} will be the intrinsic carrier density which is proportional to $\exp(-\epsilon_I/kT)$. This model therefore predicts that in the excess current range, where $(eV_x/nkT)\gg 1$, $\ln I_x$ will vary as $[2\epsilon+(eV_x/n)]/kT$. Thus, an approximately straight line should result if $\ln I_x$ is plotted against T^{-1}. When such plots were made for both Si and Ge it was quite impossible to fit them to straight lines, the variation of T^{-1} being far too great compared with that of the current.

It is concluded that the observed temperature variation of the excess current is in agreement with the model for the excess current that involves tunnelling via local levels in the forbidden gap, as formulated in Eq. (7).

E. Excess Current Versus Junction Width

To study the effects of varying W_1, the junction width constant defined above, a series of junctions was made in which all the p-type sides were similar but the doping on the n-type side was varied. Semilogarithmic plots were made of the forward characteristics at room temperature and the slope, σ_x, was obtained for each junction in the excess current range. The value of p was estimated as described above. The values of n were determined from resistivity measurements (made on bars of the material before the alloyed junctions were formed on them), and the known dependence of the resistivity on the donor concentration. From these data a plot was made of σ_x versus $[(p+n)/pn]^{\frac{1}{2}}$. This is shown in Fig. 11 and to within experimental error a satisfactory straight line results, again providing confirmation of Eq. (7). The sources of error lay principally in the lack of uniformity in the boron concentration in the aluminum wire and in the estimates of the active

[17] J. V. Morgan and E. O. Kane, Phys. Rev. Letters 3, 466 (1959).

[18] C. T. Sah, R. N. Noyce, and W. Shockley, Proc. Inst. Radio Engrs. 45, 1228 (1957).

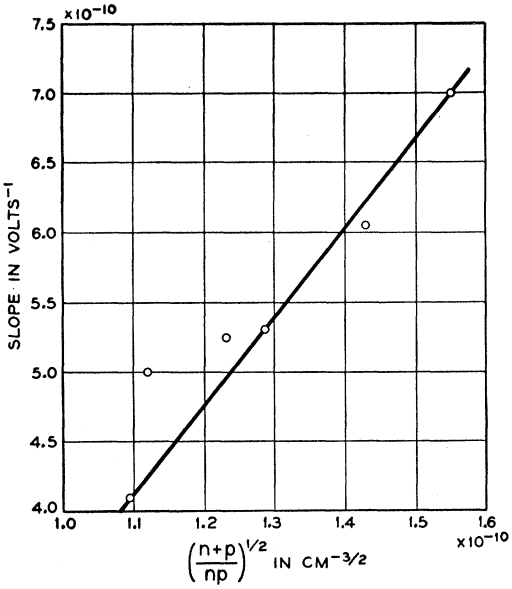

FIG. 11. Plot establishing the relation between the slope of the semilogarithmic plot of the forward characteristic in the excess current range, and the junction width.

junction areas. These two uncertainties could very well account for the scatter of the experimental points in Fig. 11.

F. Quantitative-Considerations

In the preceding sections, experiments have been described which, when analyzed in the ways dictated by Eq. (7), lead to good qualitative confirmation of the local level model for the excess current. The slopes of the straight lines resulting from the analyses of these experiments contain the quantity $(\alpha_x W_1)$. Unfortunately, owing to the lack of knowledge of the various functions that are contained in the coefficient of the tunnelling probability factor, the slopes may, in general, contain other quantities as well. We have already noted that the functions, D_x and ρ_x can have bias dependences expressed by $\exp(-\lambda V_x)$ and $\exp(-\mu V_x)$, respectively. Likewise, in the experiments where I_x is measured as a function of T, and therefore, ϵ, the function D may again contain a factor, $\exp(-\nu\epsilon)$, where ν is a constant. Putting $(\alpha_x W_1 e^{\frac{1}{2}}/2)=S$, we then have $(-\lambda+S)=7.8$ volt^{-1}, $(-\mu+S)=9.1$ volt^{-1}, and $(-\nu-S)=-14.6$ volt^{-1}. The last quantity refers to the results obtained at a bias of 0.35 volt in temperature experiments performed on junctions very similar in doping to those used for the radiation damage studies. However, in Ge, the slope of the peak current curve in the temperature experiment should not contain a quantity, ν, yet the slope of the peak current curve is comparable with that of the valley current curve. Assuming that the (αW_1) quantities are comparable for band-to-band and excess current tunnelling, this implies that ν is negligible and if the same is true for Si, then $S\simeq 14.6$. However, not too much weight should be attached to this estimate of S derived from the temperature experiments, since the variation of the energy gap with temperature in silicon is only about 6%, thus making the magnitude of the slope very susceptible to slight changes with temperature in other quantities contained explicitly or implicitly in Eq. (7). Thus, about the best that can be done at the present stage is to take simply an average value for S, which assumes that all the quantities λ, μ, and ν are small. Then $S\simeq 10.5\pm 3$ volts^{-1}, for junctions whose width constants were estimated to be 54 ± 3 A. By means of Eq. (5) an estimate was made of m_x, the reduced effective mass in the excess current tunnelling process. The value so obtained for m_x was $0.33m_0$, where m_0 is the free electron mass. This result, though it could possibly be in error by as much as a factor of 2 either way due to the uncertainty in the value of S, is an extremely satisfactory one, comparing well with the value of $0.36m_0$ which is obtained for the reduced effective mass for band-to-band tunnelling using the electron and hole density of states masses,[15] and the value of $0.28m_0$ obtained using the slope of Fig. 11. Since only the temperature experiments were made in the Ge junctions the possible error in the estimate of S for Ge will be considerably greater than that for Si. However, using $S=6.0$ volts^{-1} for Ge yields a value for the reduced effective mass of about $0.05m_0$, which is considered to be a reasonable value, lying between the value, $0.02m_0$, quoted by Hall for the reduced effective mass in band-to-band tunnelling,[19] and the theoretical estimate of $0.19m_0$ based on the densities-of-state masses.

DISCUSSION

The proposed model for the mechanism of the excess current, where tunnelling occurs by way of fairly localized levels in the forbidden region, leads to predictions which are well confirmed by the various experiments described above. Particularly convincing evidence that local levels in the energy gap are the prime cause of excess current is provided by the radiation damage experiments.

It is not possible, at present, to give a specific origin for the local levels in the energy gap. However, from the experiments that have been described it is clear that the local levels are introduced during the junction fabrication process itself and that they are caused in some way by the donor and acceptor impurities. Inherent to these narrow *p-n* junctions there are at least two possible sources for these deep energy levels:

(1) As the materials used for these junctions are necessarily very impure, distortions to the band edges can be expected,[20,21] resulting in a tail of energy levels extending into the forbidden region from both the conduction and valence bands.

---

[19] R. N. Hall, Bull. Am. Phys. Soc. **5**, 38 (1960).
[20] M. Lax and J. C. Phillips, Phys. Rev. **110**, 41 (1958).
[21] P. Aigrain, Physica **20**, 978 (1954).

(2) At the high doping densities present in Esaki junctions there must clearly be a large proportion of donor-acceptor pairs formed[22]; at acceptor (say) densities of the order of 10^{20} cm^{-3}, the average spacing between acceptors will be only about 20 A so that every donor will be within 10 A of an acceptor. Thus, even the average donor-acceptor spacings can be regarded as loosely-coupled ion pairs tending to give rise to electrically neutral impurities. Furthermore, the range in donor-acceptor separations due to chance alone will give rise to a spectrum of energy levels for the neutral complexes and it is not unreasonable to suppose that such a system of levels will range over a large fraction of the forbidden energy gap. Also, it is clearly possible for these local energy levels to be present in the necessary concentrations at the doping densities used in Esaki junctions.

The success of the local level model resulted by assuming that the tunnelling step was the rate-controlling one rather than the process whereby the field-ionized (or filled) level is replenished (or emptied). This feature imposes some conditions on the natures of the A and B levels (Fig. 1) though these conditions appear to be fulfilled as a natural consequence of the band-edge-tail model. The B levels are then a tail of donor levels extending to energies below the normal bottom of the conduction band. Electrons on these levels can tunnel to the valence band and the empty B level is then refilled by an electron falling from higher in the conduction band. Similarly, the A levels lie in a tail of acceptor levels extending to energies above the top of the valence band. Electrons can then tunnell directly to the A levels and then drop to the Fermi level. Both of the energy dissipation mechanisms, CB, AV, will be fast, occurring in times of the order of the dielectric relaxation times for the majority carriers, thus satisfying the condition that the tunnelling transition be the rate controlling one. It is conceivable that other sources of A and B levels, such as donor-acceptor pairs, could act in much the same way under certain conditions. For the present, however, the band-edge-tail model seems the simplest concept. In addition to satisfying the above requirement, it provides a smooth function, D_x, for Eq. (6), and also, D_x will increase with the doping density in accordance with the experimental results. To test the hypothesis further in the light of the accumulation of experimental results requires a theoretical determination of the nature of the D_x function, information which is almost entirely lacking at present. We may note, however, that band-edge-tailings or donor-acceptor pairs need not be the only sources of the A and B levels. It has been demonstrated for example, that additional levels, which will be different from those introduced by the original donor and acceptor impurities, are introduced by radiation damage.

[22] C. S. Fuller, in *Semiconductors*, edited by N. B. Hannay (Reinhold Publishing Corporation, New York, 1959), p. 192.

CONCLUSIONS

In silicon Esaki junctions, all of the properties of the excess current that have been observed so far can be satisfactorily accounted for by a current flow mechanism which involves tunnelling by way of local energy levels in the forbidden gap. By assuming the tunnelling transitions to be the rate-controlling ones in the electron transport process and by using a simple phenomenological model, an analytical expression for the excess current was derived [Eq. (7)], all the features of which were subjected to experimental test. The model satisfactorily accounts for the observed dependence of the excess current on the local energy level density, the temperature, the forward bias, and the junction width. A parameter bearing on all of these observations is the reduced effective mass for which a reasonable estimate was obtained.

The exact origins of the local levels in the forbidden gap are not known at present, though the concept of band-edge tails due to the high impurity concentrations seems to have the most appeal. If so, the proposed mechanism for the excess current would be directly applicable to excess currents in Esaki junctions made in other semiconductors. In any event, it is felt that the local level model is applicable, in principle, to other materials.

It is hoped that more refined and detailed experiments will lead to a knowledge of the nature and distributions of the local energy levels through the forbidden gap, though as yet there are no published theoretical studies of this problem with which experimental data can be compared.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to the numerous people who have aided them in their studies; to P. A. Wolff, M. Lax, J. Klauder, D. Kleinman, and G. K. Wertheim for many useful discussions of fundamental problems, to F. A. Trumbore and R. M. Atkins for the supply of suitable silicon crystals, to E. Dickten, P. W. Foy, J. J. Schott, and D. J. Walsh for fabricating many of the junctions used in connection with the studies of excess currents, to W. Augustyniak for performing the electron bombardment experiments, to H. C. Montgomery for performing noise measurements, and to A. A. Murray and A. J. Peters for general technical assistance.