Avalanche Breakdown in Silicon

K. G. McKAY
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received December 23, 1953)

An avalanche theory of breakdown at room temperature is proposed for semiconductors based on the assumption of approximately equal ionization rates for electrons and positive holes. The problem of obtaining ionization rates from data obtained in inhomogeneous fields is solved exactly for two specific field distributions. Ionization rates for silicon thus calculated from experimental data on breakdown voltage and on prebreakdown multiplication for both linear-gradient and step junctions are in good agreement. The temperature coefficient of the ionization rate exhibits a similar internal consistency. It is concluded that internal field emission has not been observed in silicon.

Detailed observations are reported of the pulse-type noise associated with breakdown. It is shown that this noise represents the unstable onset of breakdown and that, for the junctions studied, all of the current flow in the breakdown region can be attributed to the current carried by the noise pulses.

INTRODUCTION

THE Townsend theories of avalanche breakdown in gases have, in general, not been applied to intrinsic electric breakdown in solids. The reason for this is that no comparable regenerative mechanism for the electron avalanche has been postulated for solids. In gases, regeneration was first assumed through ionization by positive ions. This hypothesis was later abandoned in favor of electron emission from the cathode by positive ion bombardment. However, the role played by positive holes in intrinsic breakdown in solids has usually been ignored. Consequently, the principal theories[1] have dealt only with electron multiplication by electron impact or electron production by internal field emission. The former mechanism, represented by the work of von Hippel and Fröhlich, has been reasonably successful when applied to insulators such as the alkali halides. The latter mechanism, proposed by Zener, has been identified with the high-current region of the reverse voltage characteristic of very narrow germanium *n-p* junctions.[2] However, the recently measured charge multiplication at room temperature in the prebreakdown region of germanium and silicon *n-p* junctions[3] shows that avalanche formation can also occur at attainable fields and that electrons and positive holes have approximately equal ionization rates. This makes possible the application of a modified form of simple gas discharge theory to breakdown in *n-p* junctions of germanium and silicon.

The distinction between what is here proposed and what has been observed on narrow germanium junctions[2] should be emphasized. In a narrow junction, it is proposed that the field across the junction eventually attains such a value that internal field emission takes place from the region of the highest field. The observed current is determined by the field strength in that region (and possibly by space charge effects). As the field increases, the current increases but breakdown, in the sense that it is defined for gases, cannot take place solely through this mechanism. In the strict sense, breakdown defines a discontinuity in behavior; it requires some form of positive feedback which, at breakdown, may result in instability. Internal field emission does not result in instability except through secondary effects. In broader junctions with the fields insufficient to produce Zener emission, the field distribution may be such as to permit multiplication by injected carriers, i.e., an injected electron interacts with a valence electron to produce an electron-hole pair and so on. This is a cumulative process and can result in a rigorously defined breakdown. Its existence depends on the fact that *both* holes and electrons can ionize, thus providing what is essentially positive feedback.

A problem associated with the study of breakdown in semiconductors is that the experimental data are obtained from fairly narrow regions across which the fields are not uniform. To compare theory with experiment, it is necessary to transform the observed behavior into what would be observed in an infinite solid with a uniform field. In the following it will be demonstrated how rates of ionization can be derived from observed data for certain specific field distributions.

THEORY

Following Townsend's derivation of his "β" mechanism in gases,[4] we define the rate of ionization α_i as the number of electron-hole pairs produced by an electron per centimeter travelled in the direction of the electric field E. We assume that the rates of ionization for electrons and for holes are equal. The results do not depend critically on this assumption, i.e., if the ionization rates differ by 10 percent the results will be in error by about this amount. As shown in Fig. 1, the barrier has a width W and plane parallel geometry. The field is assumed to be solely a function of the position coordinate x with the E and x vectors coincident. The

[1] These are reviewed by S. Whitehead, *Dielectric Breakdown in Solids* (Oxford University Press, London, 1951).

[2] McAfee, Ryder, Shockley, and Sparks, Phys. Rev. 83, 650 (1951).

[3] K. G. McKay and K. B. McAfee, Phys. Rev. 91, 1079 (1953).

[4] L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases* (John Wiley and Sons, Inc., New York, 1939), p. 372 ff.

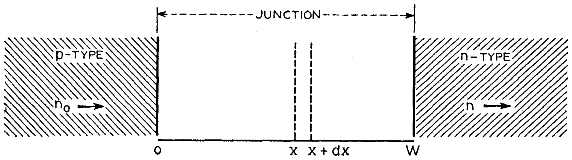

FIG. 1. The geometry assumed for the calculation of avalanche multiplication.

number of electrons injected into the junction at $x=0$ is n_0. These electrons can produce more carriers, both holes and electrons. However, we need only to follow the history of carriers of one sign, the others being automatically taken care of; we choose here to calculate the electron avalanche. Let the number of electrons produced by electrons or holes between 0 and x be n_1 and the number of electrons produced between x and W be n_2. Then the number of electrons produced between x and $x+dx$ is

$$dn_1=(n_0+n_1)\alpha_i dx+n_2\alpha_i dx=n\alpha_i dx,$$

where $n=n_0+n_1+n_2=$ the number of electrons collected at the anode.

Integrating with the boundary conditions $n_1=0$ at $x=0$ and $n=n_1+n_0$ at $x=W$, we obtain

$$1-\frac{1}{M}=\int_0^W \alpha_i dx, \qquad (1)$$

where $M=n/n_0=$ the multiplication factor.

When the integral in Eq. (1) attains unity, $M\rightarrow\infty$ and breakdown occurs. We should note three assumptions that are implicit in the use of Eq. (1). These are:

(1) $\alpha_i(E)$ is solely a function of E. This neglects the influence of the past history of the ionizing carrier, which assumption seems reasonable if the carrier loses energy primarily through lattice collisions rather than through ionizing collisions. This is probably invalid for very narrow junctions where we should approach the situation found in the cathode fall region in gases.

(2) We neglect the loss of carriers in the junction by recombination. Since the time required for an electron to traverse a typical junction with a field appropriate for appreciable multiplication is of the order of 10^{-10} sec or less, this is negligibly small compared with recombination times of greater than 10^{-6} sec.

(3) We neglect the mutual interaction between conduction electrons (such as postulated by Fröhlich) on the grounds that at any given instant of time, the actual number of electrons in the barrier region is insignificantly small. Although this assumption is not always valid, it is for the experimental data to be considered here.

We shall now consider the direct solution of Eq. (1) for certain specific field distributions.

A. The Step Junction

The step junction considered here is one in which the impurity concentration varies abruptly from p type to n type. If the acceptor concentration on the p side is much greater that the donor concentration on the n side, essentially all of the space charge region will be on the n side. This leads to the Schottky-type barrier[5] as shown in Fig. 2(a). The following relations hold between the field E, the junction voltage V and the width W:

$$E=E_M(1-x/W), \qquad (2)$$

$$E_M=2VW^{-1}, \qquad (3)$$

$$W=W_1\sqrt{V}=0.5W_1^2E_M, \qquad (4)$$

where $E_M=$ the maximum field in the junction, $W_1=$ the width constant for a given step junction

$$W_1=\left[\frac{1.317\times 10^7}{N_D-N_A}\right]^{\frac{1}{2}} \text{ for silicon,}$$

$V=V_a$, the applied voltage, plus V_i, the built-in voltage. V_i normally ranges from about 0.5 to 0.7 volt at room temperature in silicon.

Using Eq. (2) to change the variable of integration in Eq. (1) and noting that $W/E_M=W_1^2/2$, we obtain

$$1-1/M=\int_0^{E_M}\frac{W}{E_M}\alpha_i(E)dE=\frac{W_1^2}{2}\int_0^{E_M}\alpha_i(E)dE. \qquad (5)$$

Differentiating (5) with respect to E_M leads to

$$\alpha_i(E_M)=\frac{2}{W_1^2}\frac{d(1-1/M)}{dE_M}-\frac{4}{W_1^3}(1-1/M)\frac{dW_1}{dE_M}. \qquad (6)$$

The two terms of the right-hand side of Eq. (6) disclose two different ways in which experiments can be

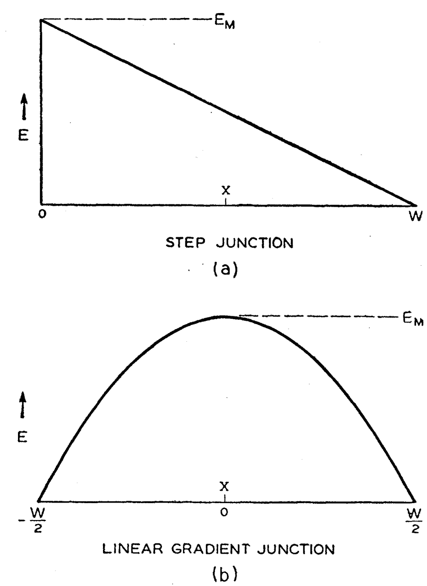

FIG. 2. Idealized field distributions for n-p junctions.

[5] W. Shockley, Bell System Tech. J. 28, 449 (1949).

performed to determine α_i since the terms can be considered separately in the following manner:

Case I. If we measure M *versus* E_M on a single junction, W_1 is constant and

$$\alpha_i(E_M)=\frac{2}{W_1^2}\frac{d(1-1/M)}{dE_M}. \qquad (7)$$

Case II. If we measure E_B, the maximum field in a junction at breakdown for various junction widths W_1, $M=\infty$ and we have

$$\alpha_i(E_B)=-\frac{4}{W_1^3}\frac{dW_1}{dE_B}. \qquad (8)$$

Equations (7) and (8) permit a point by point evaluation of $\alpha_i(E)$ from either multiplication data or breakdown data obtained on step junctions.

B. The Linear-Gradient Junction

The second common simple junction is the linear-gradient junction[5] in which the charge density of donors N_D and acceptors N_A is given by

$$N_D-N_A=ax,$$

where the zero of the x coordinate is taken in the center of the junction and a is a constant for a given junction. This leads to the field distribution shown in Fig. 2(b). The following relations hold between E, V, and W:

$$E=E_M[1-(2x/W)^2], \qquad (9)$$

$$E_M=1.5V/W, \qquad (10)$$

$$W=W_1V^{\frac{1}{3}}=[\tfrac{2}{3}W_1^3E_M]^{\frac{1}{2}}, \qquad (11)$$

where E_M= the maximum field in the junction and W_1= the width constant for a given linear-gradient junction.

Using Eq. (9) to change the variable of integration in Eq. (1) and noting that $W/E_M^{\frac{1}{2}}=(W_1^3/1.5)^{\frac{1}{2}}$, we obtain

$$1-1/M=2\int_0^{E_M}\frac{W\alpha_i(E)}{4E_M^{\frac{1}{2}}(E_M-E)^{\frac{1}{2}}}dE$$
$$=\frac{1}{2}\left[\frac{W_1^3}{1.5}\right]^{\frac{1}{2}}\int_0^{E_M}\frac{\alpha_i(E)dE}{(E_M-E)^{\frac{1}{2}}}. \qquad (12)$$

By the use of Abel's integral theorem[6] this can be transformed to

$$\alpha_i(E_M)=-\frac{1}{\pi}\frac{d}{dE_M}\int_0^{E_M}\frac{(1-1/M)}{(E_M-E)^{\frac{1}{2}}}2\left[\frac{1.5}{W_1^3}\right]^{\frac{1}{2}}dE. \qquad (13)$$

As before, two separate types of experimental data can be used to evaluate α_i as follows.

[6] E. C. Titchmarsh, *Theory of Fourier Integrals* (Clarendon Press, Oxford, 1948), p. 331. The use of this transformation was suggested by G. H. Wannier.

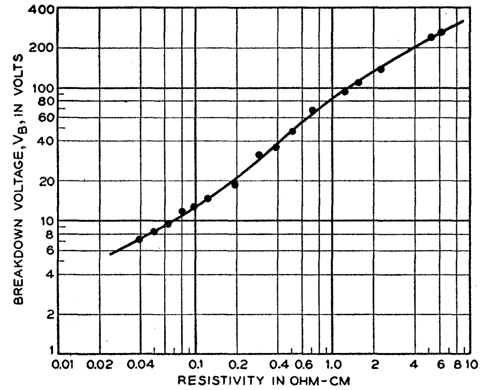

FIG. 3. The breakdown voltage V_B as a function of resistivity ρ of the n-type base of silicon alloy diodes.

Case I. If we measure M *vs* E_M on a single junction, W_1 is constant and

$$\alpha_i(E_M)=\frac{2}{\pi}\left(\frac{1.5}{W_1^3}\right)^{\frac{1}{2}}\frac{d}{dE_M}\int_0^{E_M}\frac{(1-1/M)}{(E_M-E)^{\frac{1}{2}}}dE. \qquad (14)$$

Case II. If we measure E_B, the maximum field in a junction at breakdown for various junction widths W_1, $M=\infty$ and Eq. (13) becomes

$$\alpha_i(E_B)=\frac{2}{\pi}(1.5)^{\frac{1}{2}}\frac{d}{dE_B}\int_0^{E_B}\frac{dE}{W_1^{\frac{3}{2}}(E_B-E)^{\frac{1}{2}}}. \qquad (15)$$

Again, Eqs. (14) and (15) permit a point by point evaluation of $\alpha_i(E)$ from either multiplication or breakdown data obtained on linear-gradient junctions.

No exact solutions of Eq. (1) have been obtained for field distributions other than those presented above. However, fairly good approximations can be obtained for some commonly encountered distributions, e.g., the extension to a composite junction consisting of part linear-gradient and part step junction is accomplished by assuming that the entire junction has a linear-gradient field distribution with a width constant that varies in an appropriate manner with applied voltage. Some experimental results have been so obtained that are in substantial agreement with those to be discussed later.

EXPERIMENT

A basic assumption of the preceding theory is that both holes and electrons have the same ionization rates for a given field. Photoconduction scanning curves taken with various voltages applied to the junctions indicate that this is a valid assumption for both silicon and germanium within the experimental accuracy limits of about 15 percent as previously published.[3] More accurate measurements over a wide range of field strengths are needed. However, the theory is not highly sensitive to small differences between ionization rates for electrons and holes so it appears reasonable to maintain the assumption of equal rates of ionization.

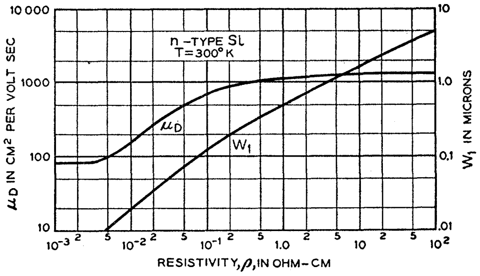

FIG. 4. The mobility of electrons μ_D and the width constant W_1 for step junctions in n-type silicon as a function of resistivity.

One set of data on ionization rates was obtained from the measurement of breakdown voltages in alloyed step junctions. A plot of E_B as a function of W_1 is required to determine α_i from Eq. (8). However, the areas of these particular junctions were too small to make accurate determinations of junction widths from capacity measurements so the data appeared as a plot of breakdown voltage V_B *vs* ρ_n where ρ_n is the resistivity of the n-type silicon. The p-type side of the junction had such a high acceptor concentration that essentially all of the space charge region is assumed to extend through the n-type end. The plot of V_B *vs* ρ_n is shown in Fig. 3.[7] It is now necessary to transform this into a plot of E_B *vs* W_1. Since both the junction width and the resistivity depend on the net donor concentration, we can use the definition of W_1, the width at unit voltage, and that of the resistivity to yield for silicon

$$W_1 = 1.44\times10^{-6}(\rho_n\mu_n)^{\frac{1}{2}}, \qquad (16)$$

where μ_n is the mobility of electrons in the n-type material. The only data available on mobility as a function of resistivity consisted of Hall mobility measurements.[8] These mobility values were uniformly reduced by 25 percent to provide coincidence with the known electron drift mobility for high-purity silicon. The result is shown in Fig. 4. Combining this with Eq. (16) we obtain the relation between W_1 and ρ_n also shown in Fig. 4. The value E_B, the maximum field in the junction at breakdown, was obtained from Eqs. (3) and (4) as

$$E_B = 2(V_B+V_i)^{\frac{1}{2}}/W_1, \qquad (17)$$

where V_i, the built-in voltage, was taken as 0.7 volt. The resultant plot of E_B as a function of W_1 is shown in Fig. 5. It should be noted that this is analogous to the relation between breakdown field and plate to cathode separation obtained in a gas at constant pressure. Differentiation of the E_B/W_1 curve yields dW_1/dE_B which, inserted in Eq. (8), gives the ionization rate α_i as a function of E_B. The values of α_i so obtained are shown as triangles in Fig. 6. This method of determining E_B presupposes that the onset voltage and the sustaining voltage for breakdown are equal. Within the experimental error this is a valid assumption for these junctions.

The use of breakdown voltage for these particular step junctions to obtain values of α_i, is open to some objection. A junction is never perfectly uniform so breakdown will always occur at a somewhat lower voltage than the averaged junction properties would predict. This will tend to yield values of α_i that are too large. A more serious objection is that no check could be made of the actual junction field distribution. It was assumed to be a step junction, but actually, resistivity changes have been noted in the n-type body as a result of junction forming. This probably means that the excess donor concentration is not uniform and that the maximum junction field is probably less than calculated. Consequently the remainder of the data to be presented here on ionization rates has been obtained from studies of the multiplication of injected carriers in the prebreakdown region. In all these cases the junctions were in grown single crystals. Careful measurements were made of junction capacity as a function of bias to ascertain the field distribution. The multiplication measurements were then made by injecting carriers by alpha-particle bombardment and determining the resultant charge transferred across the barrier as a function of bias. This technique has been described previously.[3]

A typical multiplication curve is shown in Fig. 7 for a linear-gradient junction. The voltage scale has been normalized by dividing by V_B, the breakdown voltage, as determined by the current-voltage characteristic of the unit. In this particular case $V_B = 11.18$ volts at room temperature. In general it is not possible to distinguish between linear-gradient and step junctions solely through inspection of the multiplication curve. However, the field distribution can be deter-

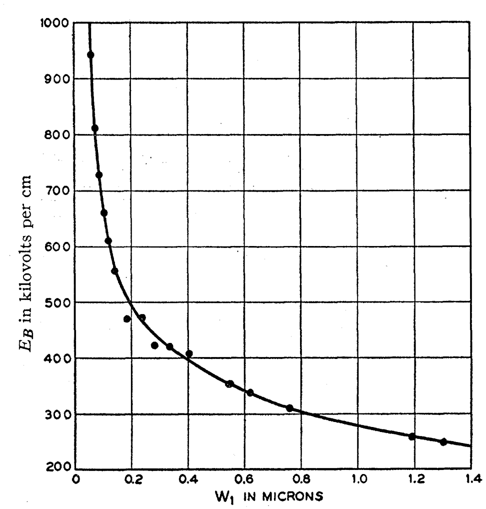

FIG. 5. The maximum field in silicon step junctions at breakdown as a function of the width constant W_1.

[7] The data shown in Fig. 3 were kindly supplied by D. K. Wilson who presented them at the Institute of Radio Engineers Transistor Research Conference, The Pennsylvania State College, July 6–8, 1953 (unpublished).

[8] These data were kindly supplied by F. Morin.

mined from the capacity data and then either Eq. (7) or (14) is used to obtain values for the ionization rate. Experimentally determined multiplication curves have been analyzed in conjunction with capacity data for a number of junctions and the resulting values of the ionization rate are shown in Fig. 6. A few remarks about the properties of the different junctions studied are in order.

The junctions represented by open circles and pluses were linear gradient junctions, i.e., the capacity varied as the cube root of the applied voltage within an accuracy of 2 percent from one volt to the breakdown voltage. The junctions represented by crosses, solid circles, and solid squares were approximately step junctions. In plotting $\log(V_A+V_i)$ *vs* $\log C$, a relation $V=KC^{-n}$ would be obtained where, for various junctions, $2.0<n<2.2$ instead of $n=2.0$ for a true step junction. In obtaining values of α_i, a step junction field was assumed for each value of the applied voltage together with the actual width measured at this voltage. The results so obtained did not differ significantly from those obtained by assuming an ideal step junction characterized by a single width constant W_1.

By calculating the results for each junction with several different approximations, it was shown that for fields of less than 500 kilovolts/cm, the values for α_i are quite consistent; the principal source of error lies in the determination of the appropriate value of the field. For example, the points shown in Fig. 6 for any given junction do not usually follow the averaged curve for α_i. It is believed that this is a consequence of the fact that the exact field distribution of any junction cannot be known in detail merely from capacity measurements. Thus there is some uncertainty as to what value of field should be associated with a given value of α_i and, for this reason, it is not possible at this stage to determine any significant fine structure in the field dependence of α_i.

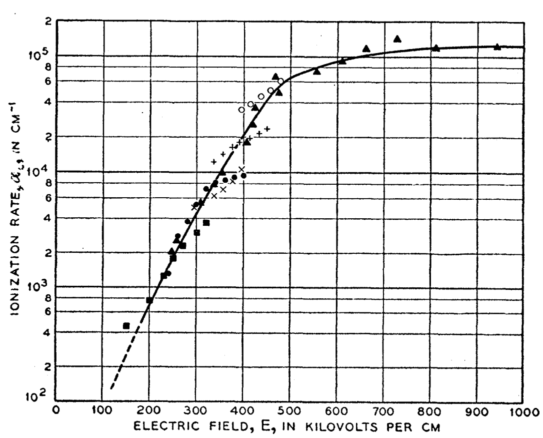

FIG. 6. The ionization rate α_i as a function of field in silicon. △ obtained from step-junction breakdown data. ○+ obtained from linear-gradient multiplication data. ×■● obtained from step-junction multiplication data.

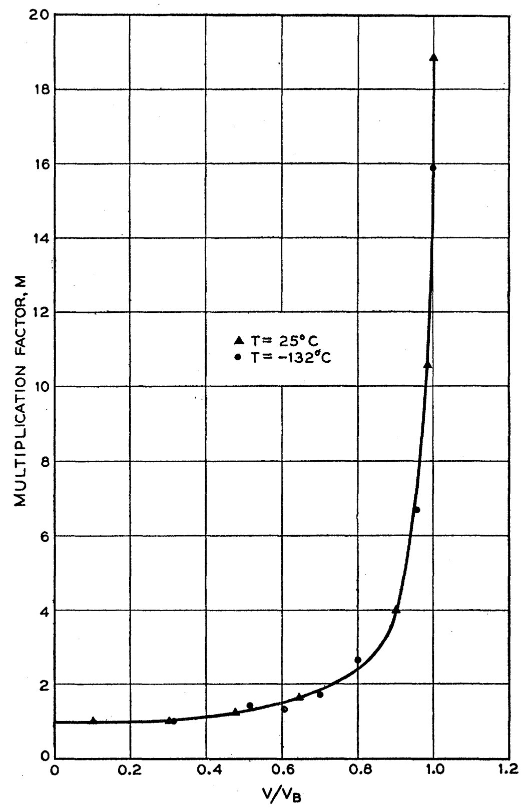

FIG. 7. Multiplication curves for a linear-gradient junction with the voltage scale normalized to the breakdown voltage for different temperatures.

Temperature Coefficient of α_i

One significant parameter in any theory of breakdown is the temperature coefficient. In this section it will be shown that breakdown and prebreakdown multiplication measurements yield consistent and reasonable values for the temperature coefficient of α_i. Neither the data nor the analysis are as extensive or as accurate as that presented in the previous section. However, its qualitative importance warrants its inclusion at this time.

The first significant observation is that the multiplication curve of a junction has the same temperature coefficient as the breakdown voltage. This is illustrated in Fig. 7, where multiplication curves are plotted against V/V_B for a silicon junction at 25°C and −132°C. At the lower temperature, the breakdown voltage had decreased from 11.18 volts to 10.03 volts and the multiplication curve had shifted quite appreciably from the room temperature curve. However, with the normalized voltage as abscissa, the two multiplication curves are indistinguishable. Thus the temperature coefficient of α_i can be determined either from breakdown data or multiplication data provided the appropriate relations are established.

Pearson and Sawyer[9] have shown that for certain silicon step junctions, the breakdown voltage varies linearly with temperature between −196°C and 25°C. This can be expressed in the form

$$V_B(T)=V_B(T_0)[1+\beta'(T-T_0)], \tag{18}$$

where $V_B(T_0)$ = the breakdown voltage at room temperature T_0 and $V_B(T)$ = the breakdown voltage at the temperature T. For the junction reported by Pearson and Sawyer, $V_B(T_0)=32.0$ volts and $\beta'=8.8\times10^{-4}/°\mathrm{C}$. It can be shown that the extent of the variation of the junction width with temperature is usually negligible. Thus, neglecting higher-order terms in β', we have, from Eq. (4),

$$E_B(T)=E_B(T_0)[1+\beta(T-T_0)], \tag{19}$$

where $E_B(T)$ = the maximum field at breakdown at temperature T, and $\beta=0.5\beta'$. To derive from this the temperature coefficient of α_i,[10] we let $M\to\infty$ in Eq. (5), differentiate with respect to T and then again with respect to E_B. By evaluating the differentials with the aid of Eqs. (8) and (19), it can be shown that

$$\frac{1}{\alpha_i}\left[\frac{\partial\alpha_i}{\partial T}\right]=-\beta\left\{1+\frac{E_B}{\alpha_i}\left[\frac{\partial\alpha_i}{\partial E_B}\right]_T\right\}, \tag{20}$$

where all quantities are evaluated at $T=T_0$. $\partial\alpha_i/\partial E_B$ can be obtained by differentiating the curve given in Fig. 6 so that if β is known, we can determine $\partial\alpha_i/\partial T$ or *vice versa*. It should be noted that as derived, Eq. (20) is applicable only to step junctions and is subject to the assumptions underlying Eq. (1).

The first measurements to be considered are of the breakdown voltage as a function of temperature for three-step junctions. These are shown in Table I and plotted in Fig. 8. $E_B(T_0)$ has been obtained from Figs. 4 and 5. $(1/\alpha_i)(\partial\alpha_i/\partial T)$ has been calculated from Eq. (20).

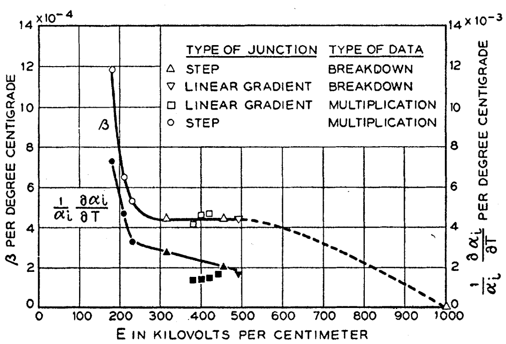

FIG. 8. The temperature coefficient β (open points) of the breakdown field and $(1/\alpha_i)(\partial\alpha_i/\partial T)$ (solid points) as a function of field.

A second approach is that of the measurement of multiplication curves at temperatures other than room temperature. From these α_i and hence $\partial\alpha_i/\partial T$ can be determined directly. Measurements were so made on two linear-gradient junctions at −130°C and α_i determined from Eq. (15). It was assumed that $\Delta\alpha_i/\Delta T=\partial\alpha_i/\partial T$ in calculating the latter. There is some question about the validity of this assumption for such a large temperature differential and the results, plotted on Fig. 8, should be accepted cautiously, particularly those for low field strengths. From these results, values of β could be calculated from Eq. (20) and these are also shown. It is interesting to observe that fairly close agreement is obtained with the values determined from step-junction breakdown voltage.

Finally, a set of measurements of breakdown voltage *versus* temperature for a linear-gradient junction was available.[11] This junction followed a relation of the form given by Eq. (18) from 150°K to 450°K with $V_B(T_0)=11.24$ volts and $\beta'=6.7\times10^{-4}/°\mathrm{C}$. Since the field and voltage bear a different relation in a linear-gradient junction from that in a step junction, we must use Eq. (11) to obtain the coefficient for the temperature variation of the breakdown field. Here $\beta=\frac{2}{3}\beta'=4.4\times10^{-4}/°\mathrm{C}$. The maximum field in this junction at breakdown is 490 kv/cm. Thus the temperature coefficient of the breakdown field for this linear junction is essentially the same as that for a step junction operated at approximately the same field. If further measurements show that this is generally true, it means that the temperature coefficient β' of breakdown *voltage* is 50 percent greater in a step junction than in a linear-gradient junction. That it is probably true is shown by the reasonably good agreement obtained from data from multiplication in linear-gradient junctions and data from breakdown in step junctions.

It is evident that much more data must be acquired before these temperature coefficient parameters can be considered to have more than qualitative significance. The behavior for fields greater than 5×10^5 volts/cm is certainly doubtful. It is probable that in this region,

TABLE I. The temperature variation of breakdown voltage of step junctions.

$V_B(T_0)$ volts	$E_B(T_0)$ kv/cm	β' (°C)$^{-1}$	β (°C)$^{-1}$	$\frac{1}{\alpha_i}\frac{\partial\alpha_i}{\partial T}$ (°C)$^{-1}$	Ref.
7.5	1000	0	0	0	a, b
32.0	455	8.8×10^{-4}	4.4×10^{-4}	-21×10^{-4}	c
125	315	8.9×10^{-4}	4.45×10^{-4}	-28×10^{-4}	a

[a] See reference 7.
[b] This value of β' has been determined after correcting for the temperature dependence of V_i, the built-in voltage.
[c] See reference 9.

[9] G. L. Pearson and B. Sawyer, Proc. Inst. Radio Engrs. **40**, 1348 (1952).

[10] I am indebted to P. A. Wolff who derived the relation between the temperature coefficient of α_i and that of E_B for a step function as given in Eq. (20). This equation neglects one term involving the field dependence of β. This term can be shown to be completely negligible over the range that is considered here.

[11] These data were kindly supplied by G. L. Pearson.

the basic assumption that α_i is solely a function of E is violated. Consequently Eq. (20) is not applicable here. The large rise in temperature coefficient for very low fields should be further substantiated. Nevertheless, the general consistency of the present meager data is quite encouraging.

BREAKDOWN INSTABILITY

One feature of breakdown in silicon that has been noted by many observers is the frequent occurrence of a peculiar form of noise just at the onset of breakdown.[9] The noise sometimes appears "clipped" and has a rather uniform spectrum. This distinguishes it from the kinds of noise normally encountered in semiconductors which have a spectrum that varies approximately as the inverse of the frequency. The noise amplitude decreases as the breakdown current is increased although sometimes it goes through several maxima and minima until, at larger currents, it disappears entirely. The amplitude also decreases as the temperature decreases. These are known facts; to these must be added some detailed observations as follows.

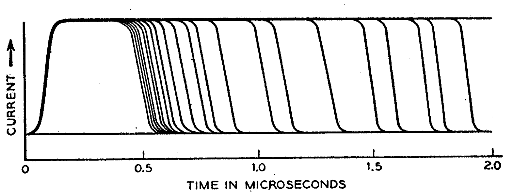

FIG. 9. "Noise" pulses of junction current at onset of breakdown in silicon traced from C.R.T. photograph.

Figure 9 is a sketch of a cathode-ray tube representation of some of the "noise" pulses, observed on a silicon n-p junction just at the onset of breakdown. The traces represent noise pulses, each of which has triggered off the sweep so they all started at the same point. The junction was not irradiated or bombarded during these measurements. The amplifier input impedance was set at 100 ohms. This is orders of magnitude smaller than the junction impedance at this bias so the pulses really represent *current* pulses through the junction. The following facts are to be noted. (1) All pulses are in the same direction. (2) All pulses have a leading edge and trailing edge whose widths are determined by the amplifier characteristic, i.e., both the onset and decay of the pulse takes place in less than 0.02 μsec. (3) The pulses are of various lengths but of constant amplitude. In the example shown, the current during the pulse corresponds to 50 μa and does not increase or decrease during the life of the pulse. By varying the sweep triggering bias it was shown that there were no pulses that did not have this current value of 50 μa. (4) Measurements on two other silicon junctions from the same crystal but of different cross-sectional areas yielded the same results with essentially the same pulse current within 10 percent. This was also true of tests on several silicon step junctions.

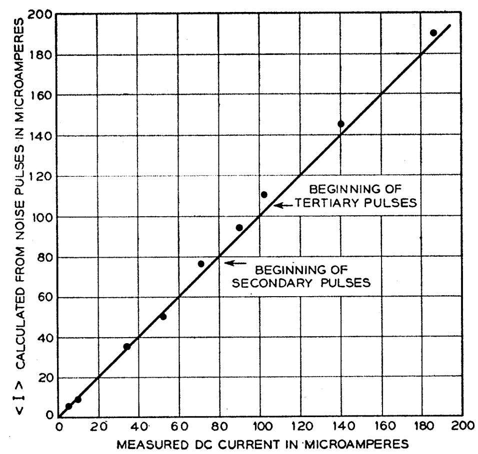

FIG. 10. Averaged current carried by noise pulses as a function of total dc current in the breakdown region.

It is instructive to examine the variation of the pulse characteristics with applied voltage in a representative junction. For a voltage just below breakdown and a junction current of about 10^{-7} amp, the pulses first appear. They are very short, <0.1 μsec, and with an amplitude of about 25 μa. An increase of about 0.05 volt in applied voltage produces the characteristic 50 μa flat-topped pulse with an average length of about 0.5 μsec. No two pulses coincide although the pulse lengths and the distance between pulses both vary in a random manner. As the applied voltage is further increased, the pulse amplitudes increase slightly. However, the principal effect is a lengthening of the pulses and a reduction of the time between pulses. Eventually the pulses are "on" most of the time, giving the appearance of short pulses in the opposite direction. By the time the pulses are on 90 percent of the time, the pulse amplitude has risen to 80 μa. At this stage, a new set of pulses appears, of amplitude about 50 μa, and resembling the initial appearance of the first set. As the voltage is increased further, these go through the same evolution as did the first set. Further sets of pulses appear at still higher voltages and evolve similarly.

Since the amplifiers do not pass dc, it is possible to estimate from the oscilloscope, the average dc current flowing through the crystal solely as a result of the pulses. By suitable bookkeeping, it is possible to carry this up to the point where the first set of pulses is on all the time. Beyond this one cannot go with reasonable accuracy since the steady current, corresponding to the first pulse set, will increase somewhat with voltage in an unknown way. Figure 10 shows a comparison between the average current carried by the current pulses and the actual total dc current through the crystal as measured on a dc ammeter. The agreement between the two currents is well within the experimental accuracy and shows that essentially all of the current in the breakdown region is carried by the pulse currents and by no other mechanism.

The interpretation of these observations is that the first set of pulses represents the unstable onset of breakdown in one region of the junction. Here, the breakdown is essentially bistable; it is either on or off. At this delicate point it can be switched from one condition to the other by thermal fluctuations. However, as the voltage becomes larger, it prefers to remain on and eventually it stays on steadily. As another region breaks down we go through a similar cycle. The increase of current in the breakdown region is then accounted for entirely by (a) the successive breakdown of various regions of the junction and, (b) by an increase in breakdown current with voltage for those regions that have broken down. The actual breakdown current in any region is probably determined by the boundary conditions and space charge considerations which should be rather difficult to compute at the present stage.

Other observations tend to confirm this picture. As the temperature of the silicon is lowered, the pulses reduce in amplitude but increase in length as does also the time between pulses. Thus a reduction in thermal agitation tends to reduce the fluctuation rate of the discharge. This separates the different regions more clearly so that the first set of pulses completes its life cycle before the second set starts, etc. This accounts for the observation of the peculiar cyclic behavior of the noise as a function of breakdown current. Even more striking is the fact that in some step junctions, several slope discontinuities in the current-voltage characteristic have been observed in the breakdown region. Wilson has pointed out that separate groups of noise pulses have been observed at the onset of each slope discontinuity in a single characteristic.

CONCLUSIONS

Data from both the breakdown region and the prebreakdown multiplication region have been used to calculate ionization rates and their temperature coefficients. The results show sufficiently good agreement to establish that, throughout the range studied, breakdown in silicon is a direct result of multiplication by collision. The fact that the results obtained from different field distributions could be analyzed without approximation to yield substantially the same results, appears to vindicate the underlying assumptions. Minor modifications of the assumptions may be required to account for the fact that values of α_i calculated for a given junction often exhibit considerable variation from the averaged values. However, it seems most probable that this is really due to a lack of detailed knowledge of the actual field distributions in these junctions. For fields above 500 kv/cm, the effective ionization rate levels off, and the temperature coefficient of the breakdown field decreases. It is probable that in this region, the ionization rate is a function, not only of the field, but also of the position of the charge carrier in the junction. In that case, the theory presented here does not apply and the plot of α_i *vs* E in that region should be considered as nothing more than a convenient method of displaying the experimental results obtained from breakdown of step junctions.

The above conclusions are substantiated by several other points. The temperature coefficient of the ionization rate has the proper sign to result from the effect of electron-lattice interaction and the magnitude is reasonable. The equations governing avalanche formation lead to instability at breakdown although they do not predict that such instability can be observed. The fact that instability is actually observed confirms the existence of a "feed-back" mechanism rather than a smoothly increasing electron density with field as postulated in the Zener theory of internal field emission. Such instability has been observed in different junctions throughout the entire voltage range studied which indicates that the same breakdown mechanism is operative throughout this range. Moreover, the variation of breakdown field with junction width shows no evidence of a constant breakdown field even for very narrow junctions. This evidence suggests that internal field emission has not been observed at all in silicon.

ACKNOWLEDGMENTS

It is evident that this work has been made possible only through the cooperative effort of a number of individuals. In addition to the specific footnote acknowledgments, I wish to thank D. K. Wilson and G. L. Pearson who have supplied most of the breakdown data, D. J. Rose and P. A. Wolff for many stimulating discussions concerning the theory, E. Buehler through whose efforts most of the junctions were grown, and H. Meier who prepared and mounted the crystals and maintained the equipment in operating order. I wish also to express my appreciation to R. W. Hamming for the numerical analysis of the linear-gradient junction data.