LB-956

SATURATION CURRENT IN ALLOY JUNCTIONS

RADIO CORPORATION OF AMERICA
RCA LABORATORIES DIVISION
INDUSTRY SERVICE LABORATORY

LB-956

1 OF 7 PAGES

AUGUST 12, 1954

RADIO CORPORATION OF AMERICA

RCA LABORATORIES DIVISION

INDUSTRY SERVICE LABORATORY

LB-956

Saturation Current in Alloy Junctions

Approved

Stuart Wm Seeley

Saturation Current in Alloy Junctions

Introduction

According to theory, the current which flows across a p-n junction when it is biased in the reverse direction should be nearly independent of the applied voltage. This current is called the saturation current, I_s. In actual practice, the current often increases with voltage (due to leakage paths across the junction at the surface) and eventually large currents flow as the breakdown voltage of the junction is approached. However, for "good" junctions well below their breakdown voltage, the reverse current follows theory so that the magnitude of I_s is of considerable practical interest.

A previous equation[1] for saturation current has been given which applies where the effect of free surfaces may be ignored and where there are quite thick (greater than a diffusion length) layers of material on either side of the junction. For alloy junctions made on thin pieces of semiconductor, these conditions are violated.

In this bulletin a new equation is developed which applies specifically to diodes made by alloying circular junctions on thin wafers. Over its range of applicability, it gives accurate values of I_s. It is shown that most of the saturation current comes from thermal generation at the free surfaces of the base wafer. The equation for I_s consists of a geometrical term and a coefficient which depends on the physical constants of the material. The principle dependences are these: I_s increases linearly with base wafer resistivity and exponentially with temperature. It also increases, but more slowly, with wafer thickness and surface recombination velocity.

For the collector of an alloy junction transistor, the basic equation is the same as for a diode, except for a small correction involving the emitter area. For the emitter junction, reverse saturation current is often not of great interest but can also be calculated; this requires a modification of the basic equation, which is also given in this bulletin.

---

[1] Shockley, W., HOLES AND ELECTRONS IN SEMICONDUCTORS, p. 314, Van Nostrand, New York (1950).

Saturation Current in Alloy Junctions

General Discussion

Saturation current is due to thermal generation of hole-electron pairs on both sides of a p-n junction. The pairs diffuse toward the junction where they are separated by the applied field so that current flows. Generation occurs at surfaces and in the volume just as does recombination. This is not surprising since the two phenomena are intimately connected. Under equilibrium conditions, holes and electrons are constantly generated and recombining. These two processes must take place, on the average, at equal rates. Otherwise, the hole and electron concentrations would eventually become either zero or infinite. From measurements of the lifetime of excess hole-electron pairs (produced, for example, by illuminating a sample), the rate of recombination at equilibrium can be computed. This is also the rate of generation at equilibrium. The net generation is defined as the difference between this total generation and whatever recombination may be taking place.

Analytically, generation may be looked upon as "negative" recombination and treated with the same mathematical techniques. For example, in n-type material, the volume recombination rate per unit volume is commonly given as $(p-p_o)/\tau_v$ where p is the hole density (minority carriers), p_o is the equilibrium value of p, and τ_v is the average hole lifetime for holes lost through volume recombination. When p is less than p_o, the recombination rate becomes negative and, in fact, represents the net rate of thermal generation. Similarly, surface recombination and generation are both represented by $(p-p_o)s$ (per square cm second) where s is the surface recombination velocity. The saturation current, I_s, is calculated by integrating these terms over the volume and surface which contribute reverse current. In the present work, calculations are made for an n-type base wafer, such as would be used for a p-n-p alloy junction transistor. The same form applies to p-type material by appropriate exchange of symbols.

Saturation Current in Diodes

Consider the diode geometry shown in Fig. 1. Four regions contribute to I_s: (1) the volume of n-type wafer within a hole diffusion length, λ_p, of the junction, (2) the wafer surfaces within the same distance, λ_p, (3) the volume of the recrystallized material within an electron diffusion length, L_e, in the p-type material, and (4) the surface of the p-type material within L_e.

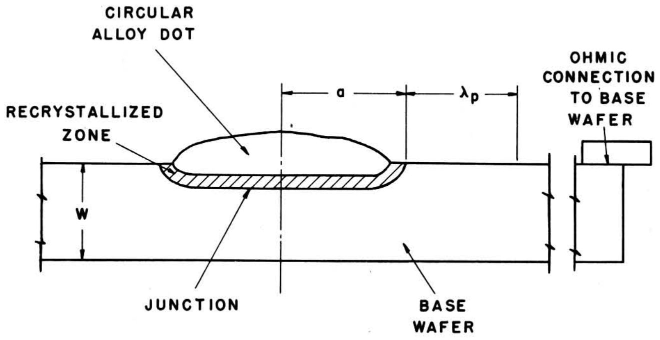

Fig. 1 - Alloy junction diode geometry consisting of circular alloy dot on thin base wafer.

λ_p is determined by the actual hole lifetime in the n-type wafer which, in the present case, is almost totally controlled by surface recombination. As such, it differs from the usual diffusion length (generally given the symbol L_p) which is determined only by the volume lifetime and applies where surface effects may be neglected. To avoid confusion and emphasize the difference, λ_p is used herein to designate the surface controlled diffusion length. If the actual hole lifetime is called τ_m, then λ_p is given by the expression

$$\lambda_p = \sqrt{D_p \tau_m} \qquad (1)$$

where D_p is the diffusion coefficient for holes in the n-type base material. For thin wafers such as those used for alloy junction devices,

$$1/\tau_m = 1/\tau_v + \frac{2s}{W} \qquad (2)$$

where τ_v is the mean lifetime of holes lost by volume recombination, s is the surface recombination velocity and W is the wafer thickness.[2] Usually, the most convenient way to compute λ_p is from direct measurement of τ_m made on the alloy junction sample itself. This may be done by measuring the transient response as described in *LB-934*.[3]

---

[2] *LB-953, The Effective Surface Recombination of a Germanium Surface with a Floating Barrier.*

[3] *LB-934, Measurement of Minority Carrier Lifetime and Surface Effects in Junction Devices.*

For simplicity, the analysis is limited to the case where $2s/W >> 1/\tau_v$. This means that volume recombination and generation are neglected. Typical values of s and τ_v are 300 cm/sec and 200 μsec, respectively, in germanium. The wafer thickness, W, is usually less than 2.5×10^{-2} cm (0.010"). For such conditions, the contribution from the volume is less than 20 per cent of the contribution from the surface so that the approximation is not a serious limitation. Now, λ_p may be written as

$$\lambda_p = \sqrt{\frac{WD_p}{2s}} \qquad (3)$$

In addition, the two contributions to I_s from the recrystallized region are of negligible importance in practice as can be shown by the following argument: The ratio of electron current, I_n, which flows from the alloy region to the base wafer, to the hole current, I_p, which flows in the opposite direction, is given by the familiar relation

$$\frac{I_n}{I_p} = \frac{\sigma_b \lambda_p}{\sigma_e L_e} \qquad (4)$$

where σ_b and σ_e are the conductivities of the base and p-type regions, respectively. Eq. 4 is related to the injection efficiency of the junction. From measured values of emitter efficiency of p-n-p alloy junction transistors, one may compute values of I_n/I_p which are generally much less than 0.1.

Thus, the only important contribution to I_s is from the surface of the base wafer. An approximate expression for I_s results from assuming that all holes generated at the free surfaces bounded by the radius $(a + \lambda_p)$ are collected and that the number of holes generated per square centimeter of free surface within this range is $p_o s$. Then,

$$I_s = qp_o\, s[2\pi(a+\lambda_p)^2 - \pi a^2] \qquad (5)$$

where q is the electronic charge.

Eq. 5 may be put into more familiar terms by replacing s with $D_pW/2\lambda_p^2$ and p_o with

$$p_o = \frac{\rho}{\rho_i^2} \cdot \frac{b}{(1+b)^2} \cdot \frac{kT}{qD_p}$$

Here, ρ is the resistivity of the base wafer, ρ_i the resistivity of the intrinsic semiconductor and b is the ratio of electron to hole mobilities.[4] The result is

$$I_s = \frac{kT}{q}\frac{\rho}{\rho_i^2}\frac{b}{(1+b)^2}\pi W\left[1+\frac{2a}{\lambda_p}+\frac{1}{2}\left(\frac{a}{\lambda_p}\right)^2\right] \qquad (6)$$

where

$$\lambda_p = \sqrt{D_p\tau_m} \approx \sqrt{\frac{D_pW}{2s}}$$

ρ_i decreases exponentially with temperature so that I_s increases rapidly. In germanium, the rate of increase of I_s is approximately 10 per cent per degree Centigrade in the neighborhood of room temperature.

Eq. 6 may be evaluated for the case of an n-type germanium wafer where $b = 2$, $\rho_i = 47$ ohm-cm at 25 degrees C, and $D_p = 44$ cm²/sec:

$$I_s = 7.6\,\rho W e^{\frac{\Delta T}{10}}\left[1+\frac{2a}{\lambda_p}+\frac{1}{2}\left(\frac{a}{\lambda_p}\right)^2\right] \text{ microamperes,}$$

$$\lambda_p = 4.7\sqrt{\frac{W}{s}} \text{ cm.}$$

ΔT is the difference between operating temperature and 25 degrees C. The exponential in ΔT accounts for the dependence of ρ_i on temperature. For a typical case where $\rho = 4$ ohm cm, $W = .013$ cm, $a = 0.06$ cm, and $s = 300$ cm/sec, $\lambda_p = 0.03$ cm and $I_s = 2.8$ μa at 25 degrees C.

Eq. 6 may also be applied to p-type base material by substituting $\lambda_n = \sqrt{D_n\tau_m}$ for λ_p where D_n is the electron diffusion coefficient. Since $b/(1+b)^2$ has the same value if $1/b$ is substituted for b (i.e., when hole and electron mobilities are interchanged), no other changes need be made. In germanium λ_n will be greater than λ_p by about 40 per cent for the same value of τ_m. Thus, for the same resistivity, I_s should be a little lower in p-type base material than in n-type.

The analysis was initially limited to the case where $W << 2s\tau_v$ so that volume generation is neglected. For thicker wafers, (or smaller s

---

[4] The origin of this expression is clear from Ref. 1. Shockley's expression for saturation current and the result of the present derivation have the same coefficient; they differ in the geometrical terms.

or τ_v), the contribution from the volume must be included. This is automatically accomplished if λ_p is computed from lifetime measurements made on the sample itself, using Eq. 1 as suggested above.

In germanium diodes of about five mils thickness, λ_p is in the neighborhood of ten mils. For "a" less than 4λ (0.040") the term $(1+2a/\lambda)$ dominates the geometrical factor. For large junctions, the term $a^2/2\lambda_p^2$ dominates. Since λ_p^2 is approximately proportional to W, this means that I_s increases with wafer thickness more rapidly for small junctions than for large ones. Also, in small junctions I_s increases approximately as $\sqrt{s}$ while the dependence approaches the first power for large junctions.

A more rigorous calculation of the geometrical terms involves integration of the net generation, $(p_o - p)$ s, over the surface. We may make the assumption that $p \approx o$ on the surface region opposite the junction when W is small but in the circular ring where $a < r < (a+\lambda_p)$, strong radial hole density gradients exist and this assumption is certainly untrue. Rightfully, the terms $(1+2a/\lambda_p)$ should be replaced by the solution of the equations

$$J_p(r) = -qD_p \frac{dp}{dr} \qquad (7)$$

and

$$2qs \int_{\infty}^{r} [p_o - p(r)] \quad r dr = -rWJ_p(r) \qquad (8)$$

The first of these equations simply states that holes flow in the wafer by diffusion. The second demands continuity of hole current by equating the total hole current flowing toward the junction through a circle of radius r to the net generation from that point outward. These equations may be combined to give a Bessels' equation of zero order whose solution has an imaginary argument:

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dy}{dr}\right) = \frac{2s}{WD_p}\, y = \frac{y}{\lambda_p^2} \qquad (9)$$

where $y = p_o - p$.

Without going into further detail, the term needed to replace $(1+2a/\lambda_p)$ in Eq. 6 turns out to be

$$f\left(\frac{a}{\lambda_p}\right) \equiv \frac{2a}{\lambda_p}\left[\frac{-H_1^{(1)}(ia/\lambda_p)}{iH_0^{(1)}(ia/\lambda_p)}\right] \cdot \qquad (10)$$

The terms $-H_1^{(1)}$ and $iH_0^{(1)}$ are tabulated functions[5] and $f(a/\lambda_p)$ is plotted with $(1+2a/\lambda_p)$, against a/λ_p in Fig. 2. Only for a/λ_p less than about 0.3 do the two solutions depart appreciably. In most practical cases a/λ_p is greater than 0.3 so that the approximate expression is sufficiently good. If $a<W$, Eq. 6 will give values of I_s which are somewhat high since density gradients across the wafer thickness (which were neglected in the calculation) will become appreciable. Within the usual range of application, however, this is not a likely source of error.

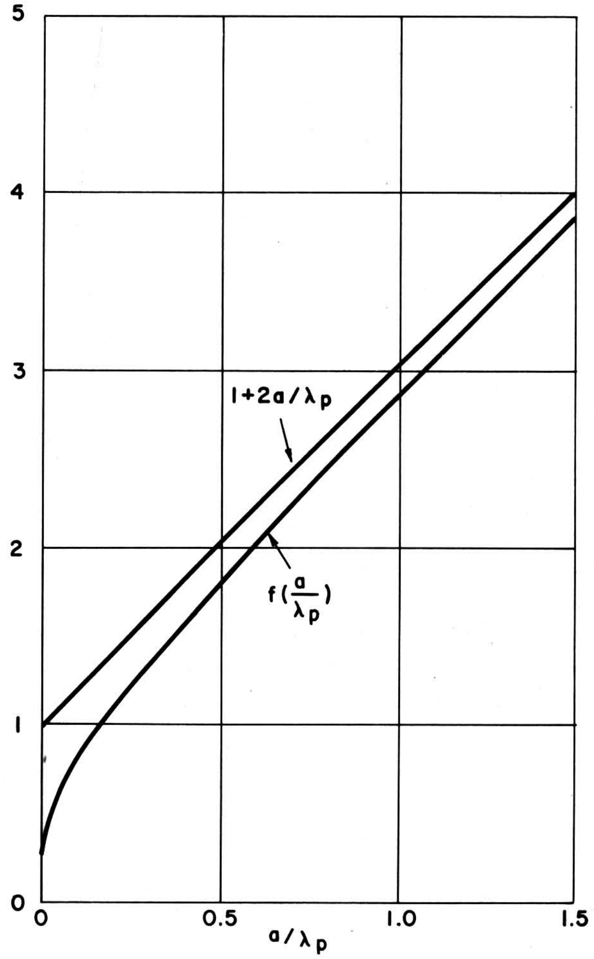

Fig. 2 - Approximate solution $(1+2a/\lambda_p)$ compared to Bessel function solution $f(a/\lambda_p)$.

---

[5] Jahnke, E., and Emde, F., TABLES OF FUNCTIONS, G. E. Stechert & Co., New York (1938).

Experimental Verification

In order to verify Eq. 6, I_s was measured and computed for a series of 48 alloy junction diodes of different geometry made on n-type germanium wafers of various resistivities. λ_p was computed from τ_m measurements according to Eq. 1. The bulk resistivity was measured on the original ingots and also checked for local variations in the following manner: The capacitance of the junction when biased in the reverse direction by a voltage, V, is

$$C \approx 4.10^{-8} \frac{a^2}{\sqrt{V\rho}} \text{ farads.}$$

"a" and C were determined by direct measurement (traveling microscope and capacitance bridge respectively). From these, ρ was computed. In this experiment, I_s varied from 0.8 to 23 µamps, ρ from 0.5 to 15 ohm cm, τ_m from 1 to 30 microseconds and "a" from 10 to 53 mils. All samples were five mils in thickness. About 80 per cent of the measured values fell within ±50 per cent of the calculated values and the average of the ratio of calculated to measured values is within 1 per cent of unity. This agreement is even better than one would expect considering the possible errors. The accuracy of measurement of resistivity alone would account for the observed spread.

Saturation Current in Transistors

The above development applies principally to diodes since it assumes a free surface opposite the dot. For the collector of a transistor whose emitter is floating, the emitter area should be subtracted out of the geometrical term. This amounts to replacing $(a/\lambda_p)^2$ with $(a^2-c^2)/\lambda_p^2$ in Eq. 6, where c is the radius of the emitter. This is usually a small correction since the emitter area is usually much smaller than the collector area. For the emitter, however, Eq. 6 gives much lower values of I_s than those one measures. This is because the floating collector is able to enlarge the collecting ability of the emitter by what is called the feed-in feed-out effect.[2] By this mechanism, holes are collected by the collector near its periphery and translated to its center where they are re-emitted to contribute to the emitter saturation current. Thus, the emitter saturation current is related to the collector saturation current. An approximate expression for the emitter saturation current, I_{es}, in terms of the collector saturation current, I_{cs}, (where both currents are measured on the same transistor with the opposite junction floating) is:

$$I_{es} = \frac{I_{cs}}{1+\frac{WW_b}{c^2}\left[1+\frac{2a}{\lambda_p}+\frac{1}{2}\left(\frac{a}{\lambda_p}\right)^2\right]} \qquad (11)$$

where W_b is the average emitter-collector spacing, "c" is the emitter radius and "a" the collector radius as before. For example, when a = 0.060 cm, c = 0.020 cm, W = 0.013 cm, W_b = 0.005 cm, and λ_p = 0.030 cm, I_{es}/I_{cs} should equal 0.67. For a transistor of this geometry, the experimental value was 0.69 which suggests that Eq. 11 is valid.

W. M. Webster

W. M. Webster