PHYSICAL REVIEW B VOLUME 8; NUMBER 8 15 OCTOBER 1973

Thermal Bulk Emission and Generation Statistics and Associated Phenomena in Metal-Insulator-Semiconductor Devices under Non-Steady-State Conditions*

J. G. Simmons and H. A. Mar
Electrical Engineering Department, University of Toronto, Toronto 5, Ontario
(Received 26 February 1973)

The non-steady-state statistics is derived for traps in the depletion layer of semiconductors. These results are applied to traps in the depletion region of metal-oxide-semiconductor systems, to obtain the emission current vs temperature (I_e-T) characteristic associated with the emission of electrons (in the case of n-type semiconductors) from traps in the upper half of the band gap. It is shown that the I_e-T characteristic is a direct image of the trap distribution in the upper half of the band gap. The generation current vs temperature (I_g-T) characteristic associated with the electron-hole generation process that occurs subequent to the emission process is also derived using the statistics. The I_g-T characteristic is shown to contain a single pronounced maximum from which the so-called carrier lifetime may be determined. Furthermore, if the leading edge of the peak is plotted in the form $\log_e I_g$ vs T^{-1} the resulting characteristic is a straight line, the slope of which provides the activation energy of the generation process. It is predicted that the donor density of the semiconductor and its activation energy can be obtained by applying the technique at low temperatures. The extension of this technique to reverse-biased p-n junctions, Schottky-barrier diodes, and highly defect semiconductors is discussed.

I. INTRODUCTION

Most studies[1-5] on the interface trap parameters of metal-oxide-semiconductor (MOS) structures have been carried out under quasiequilibrium or steady-state conditions because the statistics for such conditions are well understood. Although the non-steady-state processes are usually more difficult to formalize and solve than the corresponding steady-state processes, they do offer further insight into the defect nature of the device and the trapping processes *per se*. Furthermore, although the theoretical aspects may be complex, the actual implementation of the theory to experiment can be quite straightforward; this is certainly the case in the study to be presented here.

Most of the quasiequilibrium and steady-state studies have been concerned with alternating current characteristics of the device. Here, we are concerned with the thermal *direct-current* characteristics of the device. We will show that the method provides the trap distribution in the band gap of the semiconductor and, in particular, vivifies the bulk-generation process as applied to the MOS structure.

II. PHYSICAL PRINCIPLES

The procedure that is used to obtain the emission I-T curve and the generation I-T curve for the MOS structure is as follows: A positive voltage bias is applied to the metal electrode of the device so the surface of the semiconductor is in the accumulation mode, as shown in Fig. 1(a). In this mode all traps in the semiconductor located below the equilibrium Fermi level are occupied, while those above it are empty. The semiconductor is now cooled to a low temperature T_0 in this condition [see Fig. 1(b)]. A reverse voltage is then applied to the metal electrode, and the semiconductor goes into the non-steady-state deep-depletion mode because bulk generation, which is responsible for inverting the sample, is inhibited at the low temperature.

Now, at room temperature the traps in the upper half of the band gap located in the depletion region will normally be empty. However, when the device is biased into the deep-depletion mode at low temperatures, traps in the depletion region remain filled. Nevertheless, if the temperature of the device is now raised at a uniform rate, traps in the upper half of the band gap begin to emit their electrons to the semiconductor conduction band [see Fig. 1(c)], and an *emission* current is observed to flow in the circuit as a result of the electrons being swept out of the depletion region. When all the traps in the upper half of the band gap in the depletion region have been emptied, the emission current ceases to flow. Here, we have assumed that the density of states at the silicon–silicon-dioxide interface is sufficiently small so that the effect of these states on the total current that flows in the external circuit is negligible. However, when the density of states at the Si-SiO_2 interface is comparable to the density of traps in the depletion layer of the semiconductor, the contribution to the total current cannot be neglected. Nevertheless, by judicious biasing of the device, for example, by initially applying a large reverse voltage bias so that the device is in the strong inversion mode before cooling it, the effects of states

at the $Si-SiO_2$ interface may be eliminated. When the emission current ceases to flow, generation of electron-hole pairs in the depletion region then begins to be significant, as indicated in [Fig. 1(d)], and a generation current is observed. This generation supplies the charge required to invert the surface of the semiconductor.

In the sections that follow, we will derive the non-steady-state occupancy function f for traps throughout the band gap of the semiconductor, and use this occupancy function to obtain the non-steady-state emission and generation currents, from which we will show how the trap parameters may be obtained.

III. NON-STEADY-STATE OCCUPANCY

The processes that determine the extent to which traps are filled, that is, the occupancy of the traps, are indicated in Fig. 2. Processes r_a and r_c represent the rate of emission of electrons and holes, respectively, from traps in the band gap,

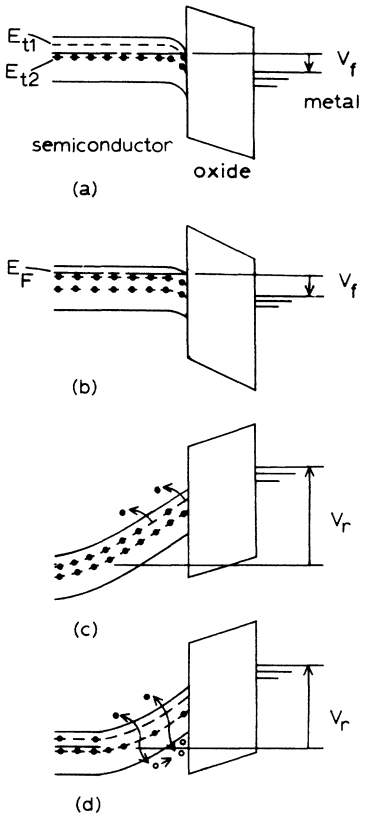

FIG. 1. Energy-band diagrams of the MOS structure for an n-type semiconductor illustrating (a) the accumulation mode at room temperature, (b) accumulation at the initial low temperature, (c) the deep-depletion mode showing electron emission, (d) the generation of electron-hole pairs in the depletion layer of the semiconductor.

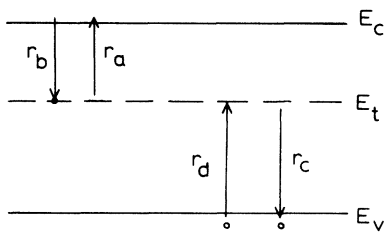

FIG. 2. Schematic illustrating capture and release of electrons and holes by traps. The capture of electrons from the conduction band and holes from the valence band are denoted by r_b and r_d, respectively. The corresponding emission processes are denoted by r_a and r_c.

and are given by

$$r_a = n_t e_n, \tag{1}$$

$$r_c = (N_t - n_t) e_p, \tag{2}$$

where N_t is the trap density, n_t is the density of trapped electrons, and e_n and e_p are the emission probability rates for electrons and holes, respectively, and are given by

$$e_n = v\sigma_n N_c e^{(E_t - E_c)/kT} \tag{3}$$

and

$$e_p = v\sigma_p N_v e^{-E_t/kT}, \tag{4}$$

where v is the thermal velocity, σ_n and σ_p are the electron and hole cross sections, respectively, and E_t is the energy of the trapped level measured with respect to the top of the valence band. Processes r_b and r_d represent the capture by the traps of free electrons (n) and holes (p), respectively, and are given by

$$r_b = nv\sigma_n(N_t - n_t) \equiv \bar{n}(N_t - n_t), \tag{5}$$

$$r_d = pv\sigma_p n_t \equiv \bar{p} n_t. \tag{6}$$

Hence the rate equation for the system is given by

$$\frac{dn_t}{dt} = (e_p + \bar{n})N_t - (e_n + e_p + \bar{n} + \bar{p})n_t. \tag{7}$$

The trap occupancy f given by the relation

$$f = n_t/N_t, \tag{8}$$

is determined by solving the rate equation (7) for n_t. In principle, this can be done provided that the spatial distributions of the free carriers, $n(x)$ and $p(x)$, are known. In practice, these expressions often cannot be obtained in closed forms. However, for systems with which we are concerned here, that is, the depletion layer in a semiconductor, any generated free carriers are immediately swept out of the depletion region by the fields therein; hence n and p are negligibly small in this region, and Eq. (7) reduces to the simpler form

$$\frac{dn_t}{dt} + (e_n + e_p)n_t = e_p N_t. \tag{9}$$

The technique with which we are concerned involves heating the semiconductor from a low temperature in a controlled manner, and a linear relation between temperature T and time is assumed:

$$T = \beta t + T_0, \tag{10}$$

where T_0 represents the initial low temperature of the semiconductor, and β denotes the heating rate. Thus, by making use of Eq. (10), Eq. (9) can be rewritten

$$\frac{df}{dT} + \beta^{-1}(e_n + e_p)f - \beta^{-1}e_p = 0, \tag{11}$$

where f is the occupancy of the trap at energy E. The solution of the above differential equation can be readily shown to be

$$f = e^{-\lambda}(\beta^{-1}\int_{T_0}^{T} e_p e^{\lambda} dT + f_0), \tag{12}$$

where

$$\lambda = \beta^{-1}\int_{T_0}^{T} (e_n + e_p) dT, \tag{13}$$

and $f_0 = 1$ is the initial occupancy of the filled traps as determined by the bias condition. When T is sufficiently high so that λ is greater than unity, the traps begin to empty and to evaluate (12) we write it in the following form:

$$f = e^{-\lambda}\left(\beta^{-1}\int_{T_0}^{T} \frac{e_p}{(e_n + e_p)} (e_n + e_p)e^{\lambda} dT + 1\right). \tag{14}$$

Note that $\lambda \ll 1$ the integral in (12) becomes $\beta^{-1}\int_{T_0}^{T} e_p\, dT \ll f_0$, which means that $f \simeq f_0$, that is, the traps have not yet begun to empty. Now, $(e_n + e_p)e^{\lambda}$ is a much more rapidly varying function of temperature than $e_p/(e_n + e_p)$ when $\lambda \geq 1$, so the latter factor may be taken outside the integral, without introducing significant error:

$$f = e^{-\lambda}\left(\beta^{-1}\frac{e_p}{e_n + e_p}\int_{T_0}^{T} (e_n + e_p)e^{\lambda} dT + 1\right). \tag{15}$$

Equation (15) integrates to

$$f = \left(1 - \frac{e_{p0}}{e_{n0} + e_{p0}}\right)e^{-\lambda} + \frac{e_p}{e_n + e_p}. \tag{16}$$

The function $e^{-\lambda}$ is shown plotted as a function of energy for various values of temperature in Fig. 3(a). It will be noted that in the *upper* half of the band gap it has a similar functional dependence on energy[6] as that of a Fermi-Dirac distribution function, in that above an energy E_{tn}^* defined by $\lambda = 1$, that is,

$$\int_{T_0}^{T_f} [e_n(E_{tn}^*) + e_p(E_{tn}^*)] dT \simeq \int_{T_0}^{T} e_n(E_{tn}^*) dT = 1, \tag{17}$$

traps are essentially empty, and below E_{tn}^* they are essentially filled. Therefore, it follows that for $E_{tn}^* > \frac{1}{2}E_g$ the second term on the right-hand side of (16) may be neglected in comparison to the first,

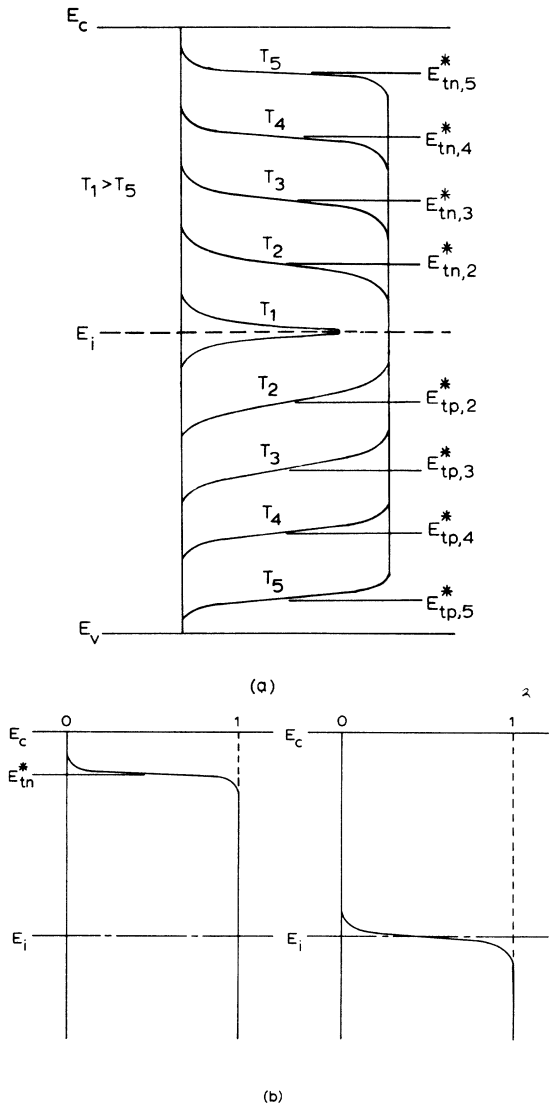

FIG. 3. (a) Illustration of the term $e^{-\lambda}$ plotted as a function of energy with temperature as the parameter. The quasi-Fermi levels E_{tn}^* and E_{tp}^* are shown approaching the midgap with increasing temperature. (b) $e_n e_p/(e_n + e_p)$ plotted as a function of E and negligible in comparison with $e^{-\lambda}$ for $E_{tn}^* \gg \frac{1}{2}E_g$.

as shown in Fig. 3(b). Thus in the upper half of the band gap

$$f \cong e^{-\lambda}, \quad E_{tn}^* > \tfrac{1}{2}E_g, \tag{18}$$

and the statistic is governed solely by the emission process, the generation process being negligible. This may be seen by solving the rate equation (9) with emission only included, that is, for the case $e_p = 0$. As the temperature is increased, E_{tn}^* approaches midgap [see Fig. 3(a)], which means that electrons in energy levels above E_{tn}^* have been emptied by the emission process. In the process of

being swept out of the depletion layer the emitted electrons give rise to the *emission* current.

At some temperature, say, T_1, when E_{tn}^* reaches midgap, $e^{-\lambda}$ becomes a sharply peaking function about E_i, (here ν_n is assumed to be equal to ν_p), which collapses rapidly about E_i for any further increase in temperature. Hence above T_1 the first term in (16) becomes negligible compared to the second one, which means that the emission process ceases and the generation process takes over. It is interesting to note that when emission has ceased, i.e., $e^{-\lambda} \simeq 0$, the statistic is identical to the *steady-state* generation statistic, that is, the non-steady-state generation rate is equal to the steady-state rate. Let us now consider the emission and generation processes quantitatively.

IV. EMISSION CURRENT

In this section we will be concerned with current that arises from the emission of electrons from traps in the upper half of the band gap to the conduction band. (In the case of a p-type semiconductor the process would involve the emission of holes from traps in the lower half of the band gap to the valence band.) The effect of both distributed traps and discrete traps will be treated.

A. Distributed Traps

For the case of distributed traps above the midgap, the emission current I_e is proportional to the rate at which traps are emptied:

$$I_e \propto qAN(E_{tn}^*)\frac{dE_{tn}^*}{dt} \tag{19}$$

or (see Appendix)

$$I_e = \frac{C_0 AL}{C_d + C_0}\ \tfrac{1}{2} q\,\beta N(E_{tn}^*) \times [1.92\times 10^{-4}\log_{10}(\nu/\beta) + 3.2\times 10^{-4}]. \tag{20}$$

In practical MOS devices in the depletion mode $C_0 \gg C_d$, so that (20) reduces to

$$I_e = (\tfrac{1}{2}qAL)\,\beta[1.92\times 10^{-4}\log_{10}(\nu/\beta) + 3.2\times 10^{-4}]N_t(E_{tn}^*), \tag{21}$$

which shows that the emission current is proportional to the trap density at E_{tn}^*. Also, from (17) it can be shown[7] that

$$E_c - E_{tn}^* = T[1.92\times 10^{-4}\log_{10}(\nu/\beta) + 3.2\times 10^{-4}] - 0.0155; \tag{22}$$

hence the temperature axis can be transformed into an axis representing energy below the conduction-band edge. Thus, it will be apparent from (21) and (22) that *an I_e-T plot is a direct image of the trap distribution in the upper half of the forbidden gap*. A similar[7,8] conclusion was deduced for the emission current from trap states in the upper half of the band gap in metal-insulator-metal structures, although the formulation used there was more rigorous.

In order to demonstrate the efficacy of this technique, consider the trap distribution shown by the full line in Fig. 4. Using the transformation equations (21) and (22), the trap distribution axis and the energy axis convert to current and temperature axes, respectively. Hence the full curve represents the approximate I-T characteristic when referenced with respect to the current and temperature axes. Now, to demonstrate the accuracy of approximation (21), we calculate the exact I-T characteristic for the trap distribution considered above using the exact expression for the current [cf. Eq. (2) of Ref. 6]:

$$J = q\int_{E_i}^{E_c} N_t(E)\,e_n(E,T) \times \exp\left(-\beta^{-1}\int_{T_0}^{T} e_n(E,T)\,dT\right)dE. \tag{23}$$

The resulting I-T characteristic is shown by the dotted line of Fig. 4, in which it is seen that the approximate and exact I-T characteristics correlate extremely well, illustrating that the functional form of the I-T characteristic is a true reflection of the trap distribution.

B. Discrete Trap Level

The approximate nature of (19) makes it valid only for trap distributions more than about $2kT$ wide in energy. In particular, it is invalid for discrete traps. However, the emission current from traps at a discrete energy level in the upperhalf of the forbidden gap can be shown[9] to be given by

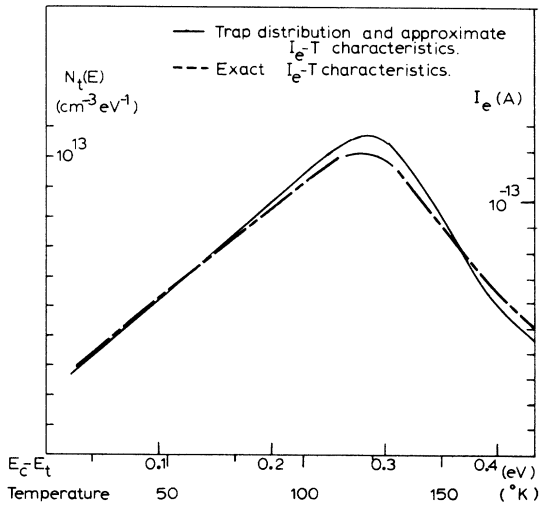

FIG. 4. Comparison of approximate and exact I-T characteristics for the given trap distribution.

$$I_e = qAN_t e_n e^{-\lambda}, \tag{24}$$

where

$$\lambda = \beta^{-1} \int_{T_0}^{T} e_n dT.$$

The area under the emission I_e-T peak, denoted by Q, is proportional to the charge released from the traps and is given by

$$Q = \int_{T_0}^{T_f} I dT = \tfrac{1}{2} q \beta N_t \Delta L, \tag{24a}$$

where ΔL represents the resultant increase in the width of the depletion layer when the reverse voltage bias is applied. The trap density may thus be evaluated from (24a). The energy E_t of the trap level may be determined quite readily from the I_e-T characteristic. This is because (24) exhibits a pronounced peak at a temperature T_m which is related to the energy level of the trap by

$$E_t = E_c - T_m(1.92\times 10^{-4} \log_{10}(\nu/\beta) + 3.2\times 10^{-4}) + 0.0155, \tag{25}$$

where T_m is the temperature at which the peak in the I_e-T curve occurs.

It should be noted that distributed trap levels as well as a discrete level may result in a sharp peak in the I_e-T curve. Therefore it becomes important to obtain a criterion to distinguish between the I_e-T characteristics for the two cases. Such a criterion is provided by the fact that the half-width of the I_e-T curve associated with the *discrete* trap is given by the relationship[9]

$$\Delta T = 2kT_m^2/(E_c - E_t). \tag{26}$$

If the width of the peak is greater than this, then the traps involved extend over a range of energies.

V. GENERATION CURRENT

A. Initial Considerations

When emission ceases and generation begins, we have seen that the occupancy is determined by

$$f = e_p/(e_n + e_p). \tag{27}$$

Thus, the rate of generation of electron-hole pairs through traps distributed *throughout* the band gap per unit volume is given by

$$G = \int_{E_v}^{E_c} \frac{e_n e_p}{e_n + e_p} N_t(E) dE. \tag{28}$$

Now, the function $e_n e_p/(e_n + e_p)$ peaks sharply at an energy E^* given by $e_n(E^*) = e_p(E^*)$, which, for $\nu_n = \nu_p$, is at the midgap (that is, $E^* = \frac{1}{2}E_g$). Hence, provided $N_t(E)$ is not a strong function of energy, that is, $dN(E)/dE < N(E)/kT$, the function $e_n e_p/(e_n + e_p)$ can be represented by the δ function:

$$e_n e_p/(e_n + e_p) = B\delta(E - E^*), \tag{29}$$

where B is the area of the δ function, and given by[10]

$$B = \int_{E_v}^{E_c} \frac{e_n e_p}{e_n + e_p} dE \cong \tfrac{1}{2}\pi kT (\nu_n \nu_p)^{1/2} e^{-E_g/2kT}. \tag{30}$$

Substituting (29) into (28) and integrating yields

$$G(E^*) = BN_t(E^*). \tag{31}$$

Now, the rate of generation of electron-hole pairs per unit volume through a *discrete* trap level at energy E_t and density N_t is given by

$$G(E_t) = \frac{e_n(E_t) e_p(E_t)}{e_n(E_t) + e_p(E_t)} N_t(E_t). \tag{32}$$

By a comparison of (31) and (32), it is seen that the generation rate through distributed-trap levels is equivalent to that of the discrete-trap case of energy E^* and density $\bar{N}_t(E^*)$ given by

$$\bar{N}_t(E^*) = \frac{BN_t(E^*)[e_n(E^*) + e_p(E^*)]}{e_n(E^*) e_p(E^*)}. \tag{33}$$

Thus in the following we will assume that the system contains a single discrete-trap level at energy E_t and density N_t, and use the transformation (33) when the results are applied to distributed traps.

B. Formulation of Generation Current-Temperature Characteristic

The rate of change in charge at the surface of the semiconductor as a result of generation of electron-hole pairs in the depletion layer is given by

$$\dot{Q}_s = \frac{q e_n(E_t) e_p(E_t) N_t L}{e_n(E_t) + e_p(E_t)}, \tag{34}$$

where N_t and E_t are the density and energy of the traps through which generation takes place, L is the width of the depletion layer, and the rate of generation of carriers is related to the current by

$$J = C_0 \dot{Q}/(C_0 + C_d). \tag{35}$$

In the typical MOS device, the insulator capacitance C_0 is larger than the depletion-layer capacitance C_d in which case (35) reduces to

$$J \cong Q_s. \tag{35a}$$

J is also equal to the rate of change of the charge on the gate electrode $\dot{Q}_g$, where Q_g is, from Gauss's Law

$$Q_g = C_0(V_g - qN_d L^2/2\epsilon_s - \phi_{ms}), \tag{36}$$

where V_g is the applied voltage bias, N_d is the donor density, ϵ_s is the permittivity of the semiconductor, and ϕ_{ms} is the metal-semiconductor work function. Hence

$$\dot{Q}_g = J = -\frac{qC_0 N_d L}{\epsilon_s} \frac{dL}{dt}. \tag{37}$$

From (34), (35), and (37) we have

$$\frac{dL}{dT} = -\frac{e_n(E_t) e_p(E_t)}{e_n(E_t) + e_p(E_t)} \frac{\epsilon_s N_t(E_t)}{\beta C_0 N_d}, \tag{38}$$

where β is the rate at which the temperature is increased. If $e_n(E_t) \simeq e_p(E_t)$ then the trap level is positioned close to midgap, and (38) reduces to

$$\frac{dL}{dT} \simeq -\frac{e_n(\frac{1}{2}E_g)\epsilon_s N_t(\frac{1}{2}E_g)}{2\beta C_0 N_d} . \tag{39}$$

On the other hand, if the trap level is in the upper half of the band gap then $e_n \gg e_p$; or in the lower half of the gap, $e_p \gg e_n$. Assuming the former to be the case, (38) reduces to

$$\frac{dL}{dT} = -\frac{e_p(E_t)\,\epsilon_s N_t(E_t)}{\beta C_0 N_d} = -\frac{\nu_p \epsilon_s}{\beta C_0}\frac{N_t}{N_d} e^{-E_t/kT}. \tag{40}$$

Hence, the functional dependence of (39) and (40) on temperature are identical in form. Thus, we need to solve only (40), from which the solution of (39) may be deduced.

The approximate solution of (40) is readily shown (see Appendix) to be given by

$$L = L_0 - \frac{\upsilon\sigma_p N_c \epsilon_s N_t}{\beta C_0 N_d}\frac{kT^2}{E_t} e^{-E_t/kT} , \tag{41}$$

where L_0 represents the width of the depletion layer when the reverse voltage bias is increased by V_R at the low temperature. The equivalent solution for the case of trap levels at the midgap is given by

$$L = L_0 - \frac{\upsilon\sigma_p N_v \epsilon_s N_t}{2C_0 N_d \beta}\frac{kT^2}{E_t} e^{-E_g/2kT} . \tag{42}$$

Substituting (40) and (41) into (37) yields

$$I = JA = qe_p(E_t)N_t A\left(L_0 - \frac{\epsilon_s kT^2 N_t}{\beta C_0 N_d E_t} e_p(E_t)\right) , \tag{43}$$

where I is the current and A represents the area of the device. Similarly, for the special case in which the energy level of the traps is at midgap, the generation current is given by

$$I = \frac{qN_t A e_p(\frac{1}{2}E_g)}{2}\left(L_0 - \frac{\epsilon_s kT^2 N_t}{2\beta C_0 N_d} e_p(\tfrac{1}{2}E_g)\right) . \tag{44}$$

When the traps are distributed throughout the band gap, generation takes place through traps at the midgap, as argued in Sec. V A. Thus, using (33), (39), and (42) in (37), the resultant expression for the generation current is given by

$$I = \frac{q\bar{N}_t(E^*)A e_p(\frac{1}{2}E_g)}{2} \times\left(L_0 - \frac{\epsilon_s kT^2 \bar{N}_t(E^*)}{2\beta C_0 N_d} e_p(\tfrac{1}{2}E_g)\right) . \tag{45}$$

VI. COMPUTED CHARACTERISTICS

The family of large peaks shown in Fig. 5(a) are the theoretical *generation* I-T curves (43) obtained for different values of V_R for a discrete trapping level positioned in energy $E_c - E_t = 0.3$ eV below the bottom of the conduction band; the same values of trap density and cross sections are used in each case. The small peak occurring at low temperatures is the *emission* I-T characteristic corresponding to the discrete trap. The emission I-T characteristic appears at lower temperature than the generation peak because the activation energy associated with the former is simply $E_c - E_t$, whereas the *limiting* activation energy associated with the generation process is the energy of the trap E_t. Thus, the closer the trap moves to midgap, the smaller the temperature difference between the two peaks [cf. Figs. 5(a) and 5(b)]. When the discrete trap is at midgap [Fig. 5(c)] or below [Fig. 5(d)] one cannot distinguish between the emission and the generation I-T curves since, in this case, the activation energy associated with each process is the same. When the discrete trap is displaced below midgap the generation I-T curve shifts back to higher temperatures. This is because the limiting activation energy for this generation process is now $E_c - E_t$ ($> E_t$); thus, for example, the generation I-T curve for a trap level that is 0.3 eV above the top of the valence band will peak at the same temperature as that for a trap level positioned

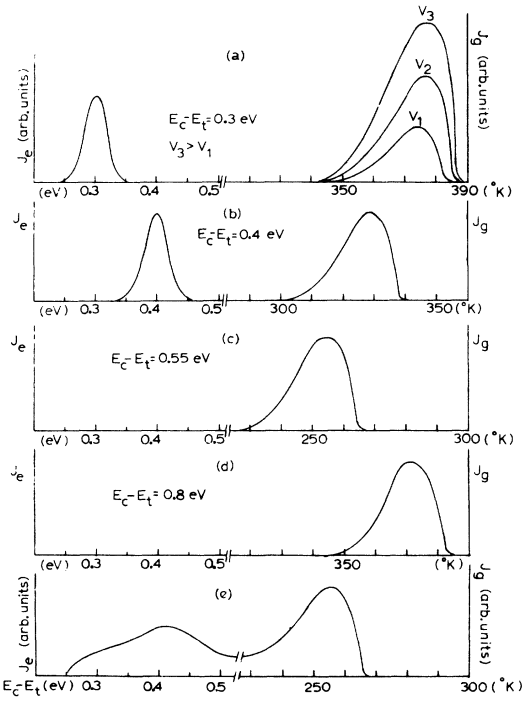

FIG. 5. (a) Theoretical I-T curves using different values of V_R for a discrete trap 0.3 eV below the edge of the conduction band. Theoretical I-T curves for discrete trap levels with energies (b) 0.4 eV, (c) 0.55 eV (midgap), (d) greater than 0.55 eV below the edge of the conduction band. (e) Theoretical I-T curves for the case of distributed traps in the band gap.

0.3 eV below the bottom of the conduction band [cf. Figs. 5(a) and 5(d)].

Figure 5(e) illustrates the I-T characteristics for distributed traps. The I-T characteristic occurring at the lower temperatures [Fig. 5(e)] corresponds to the emission process, generation currents in this temperature range being negligibly small, and is thus a direct reflection of the trap distribution (cf. Fig. 4). The large peak corresponds to generation through traps at midgap, and was discussed in Sec. V.

A. Determination of Trap Energy and Carrier Lifetime

For temperatures such that the second term within the large parentheses of (43) is much less than L_0, that is, the rising portion of the I_g-T curve, a plot of $\log_e I_g$ vs $1/T$ is a straight line, the slope of which is $-E_t/k$, thus providing the energy of the trap. The intercept provides the quantity $\frac{1}{2}qL_0N_vA\tau$ where $\tau(=v\sigma_nN_t)$ is the so-called carrier lifetime.

A crosscheck on the trap energy and carrier lifetime is provided by the maximum in the I_g-T characteristic. The temperature at which the maximum in the I_g-T characteristic occurs is determined by taking the derivative of the current with respect to temperature and equating it to zero yielding

$$E_t = kT_m \ln \frac{2v\sigma_p \epsilon_s N_v N_t k T_m^2}{\beta C_0 L_0 N_d E_t} \tag{46}$$

when $E_t > \frac{1}{2}E_g$, or

$$E_t = E_c - kT_m \ln \frac{2v\sigma_n \epsilon_s N_c N_t k T_m^2}{\beta C_0 L_0 N_d (E_c - E_t)} \tag{47}$$

when $E_t < \frac{1}{2}E_g$. Thus, (46) or (47), as the case may be, provides a crosscheck for E_t and N_t.

B. Significance of Area Under Curve

It should be noted that the functional form of the generation I-T curve is the same for all trap levels. Furthermore, the area under the curve corresponds to the net increase in the charge on the gate due to the increase in the reverse voltage bias under steady-state conditions, that is,

$$\int_{T_0}^{T_f} I\,dT = C_0 V_R, \tag{48}$$

where T_f is the temperature at which the generation current ceases to flow. Hence (48) provides a means for determining the oxide capacitance.

VII. DISCUSSION

The discussions presented so far have been centered around the MOS structure. Nevertheless, the theory for the non-steady-state emission current is equally applicable to other structures, for example, the p-n junction and the Schottky diode. Experimental I-T curves have been obtained[11-14] using reverse-biased p-n junctions, and they indicate the presence of a significant density of traps in the semiconductor of the devices used for the experiments. In these cases the experimental technique is as follows: First, the device is forward biased, in order to fill the traps, and then cooled in this condition so that the traps are still full at the low temperature. The device is then reverse biased, and the temperature increased uniformly. The resulting I-T curve provides the trap distribution in the upper half of the band gap. It should be noted that the non-steady-state current is observed provided that the reverse current which flows through the device as a result of the voltage bias is negligible in comparison with the emission current. For the case of the Schottky diode, the barrier height should be greater than one-half the band gap if traps in the upper half are to be examined. It is also noted that unlike the generation I-T curve for the MOS structure, which manifests a peak, that of the p-n junction and the Schottky diode increases montonically with temperature, for a discrete trap, according to

$$I = \frac{qALN_t e_n(E_{t1}) e_p(E_{t1})}{e_n(E_{t1}) + e_p(E_{t1})}\,. \tag{49}$$

A. Practical Limitations

At the starting temperature T_0, a trap, of, say, energy E_0, that satisfies the condition

$$e_n > 10^{-2}\,,$$

that is,

$$(E_c - E_0) < kT \ln 100 v\sigma_n N_c\,,$$

has a relatively high emission probability for electrons. What this means is that the technique is limited to traps positioned deeper than $E_c - E_0$ below the bottom of the conduction band, since those positioned closer to the conduction-band edge can discharge quickly by *isothermal* emission when the MOS capacitor is reverse biased.[13] Thus, at liquid-nitrogen temperature assuming $\sigma_n = 10^{-15}$ cm^{-3}, $v \cong 10^7$ cm sec^{-1} and $N_c \cong 5.4\times10^{15}T^{3/2}$ cm^{-3} the technique can be applied only to traps more than about 0.19 eV below the bottom of the conduction band. However, at liquid-helium temperature where $kT \simeq 0.0036$ eV, the trap distribution can be probed to within about 0.01 eV of the bottom of the conduction band.

B. Temperature below 20 °K—Effect of Donor Centers

Starting the heating process at temperatures below about 20 °K requires special consideration. The majority of the donor centers at such temperatures are deionized, which means that since there are very few free electrons in the semiconductor it is not possible to charge the traps at low temperatures by positively biasing the semiconductor. Hence, they must be charged by positively biasing

the semiconductor at higher temperatures and then cooling the device to the desired low temperatures with the bias applied; as a consequence of the biasing conditions described above, all the donors in the semiconductor will be *filled* at the low temperature.

In Sec. VII A it was noted that by cooling the semiconductor to liquid-helium temperature ($T_0 = 4.2\,^\circ$K) the trap distribution may be examined to within about 0.01 eV of the bottom of the conduction band. Since donor ionization energies in silicon are typically 0.04–0.06 eV, the above observation means that the properties of the donor states can be investigated using the thermal technique. To illustrate this point, consider Fig. 6(a) which shows the energy diagram after the device is negatively biased at the starting low temperature. In this case, the field is uniform throughout the semiconductor and remains so, since the donor centers cannot readily release their electrons. However, during the heating process the donor centers release their electrons and the device relaxes to the deep-depletion mode [see Fig. 6(b)]. Hence, on the emission I-T curve associated with the distributed traps in the band gap will be superimposed the I-T curve associated with the release of charge from the donor centers during the formation of the depletion region, as shown in Fig. 7. This donor I-T peak will, of course, have all the characteristic features of a *discrete* trap level (Sec. IV B), namely, the maximum of the peak will occur at a temperture given by

$$E_c - E_d = T_m[1.92\times10^{-4}\log_{10}(\nu/\beta) + 3.2\times10^{-4}] - 0.0155, \quad (50)$$

and the half-width of the peak will be given by (26). Furthermore, the area under the donor peak will be equal to $\beta N_d \lambda_d$, where λ_d is the width of the semiconductor in the deep-depletion region for the reverse voltage bias on the semiconductor. Hence the greater the negative voltage bias on the semiconductor, the higher will be the donor peak.

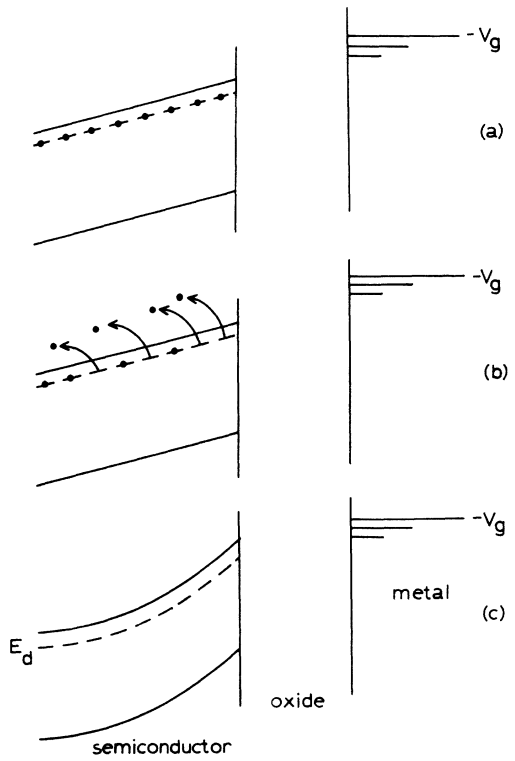

FIG. 6. (a) Energy-band diagram of the MOS structure at temperatures below 20 °K. (b) As the temperature is increased the donor centers release the trapped electrons. (c) At temperatures above 20 °K the donor centers have been inonized and the semiconductor is in the deep-depletion mode.

C. MOS Structures having Semiconductors with High Trap Densities

In the above sections we have been concerned with device grade semiconductors in which the density of traps are sufficiently low so that the effect of trapped charge can be ignored when compared with that due to the donor ions. In this section we will assume that the reverse is true, that is, we will consider the effect of a high trap density such that $N_t \gg N_d$. In particular, let us consider a semiconductor with a trap density $N_t = 10^{16}\,\mathrm{cm}^{-3}$ ($N_d = 10^{14}\,\mathrm{cm}^{-3}$) at a discrete energy level in the upper half of the band gap. Furthermore, let us assume that traps in the upper half of the band gap are acceptor type while those below midgap are donor type. Then, for the case of an n-type semiconductor, it can be shown by using Shockley-Read statistics that essentially all the donor electrons exist in the traps; thus the position of the Fermi level is given by

$$E_t - E_F = kT\ln(N_t/N_d), \quad (51)$$

as shown in Fig. 8(a). For simplicity, we assume that the energy bands are initially flat throughout when the device is short circuited. Now, let us suppose that a reverse voltage bias is applied at low temperatures. Hence the energy bands are as shown in Fig. 8(b), since at the low temperatures the traps cannot readily release their electrons so that the semiconductor behaves as an ideal insulator because the system is compensated. However, as the temperature is increased, the electrons are released and a non-steady-state emission current is observed to flow, and I-T characteristics given by (24) result. A deep-depletion space-charge region [Fig. 8(c)] thus forms as a result of the release of the trapped electrons. The space-charge density is constant and equal to qN_d, and the width of the depletion region d is given by

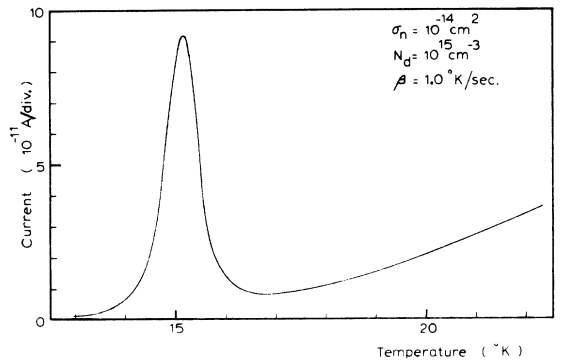

FIG. 7. Typical I_e-T curve to be expected from the emission of electrons from donor centers.

$$d = (2\epsilon_s \phi_s / qN_d)^{1/2}, \tag{52}$$

where ϕ_s is the surface potential.

As the temperature is further increased, the generation of electron-hole pairs through the trap level begins and a generation current begins to flow. Simultaneously, the width of the space-charge layer decreases because of the presence of the generated free holes at the interface, which causes the device to go from the deep-depletion mode to the equilibrium inversion mode. When the device reaches the equilibrium inversion mode the current ceases to flow. Thus, the over-all nonsteady I-T characteristic curves for MOS structures using semiconductors with high trap densities are similar to those obtained when device grade semiconductor is used. From the emission I-T curve the energy level E_t and cross section of the trap can be obtained. The value of E_t can be checked by using the generation I-T curve as discussed in Sec. VI A. In addition, the trap density may be determined by using (46) or (47) provided that both E_t and the cross section of the traps are known.

VIII. CONCLUSIONS

The non-steady-state statistics for traps in the depletion layer of semiconductors have been derived and have been used to obtain the non-steady-state emission and generation I-T curves for the MOS structure. It has been shown that the I_e-T characteristic obtained is a result of the release of electrons from traps in the upper half of the band gap, and, furthermore, is a direct image of the energy distribution of the traps. The activation energy of a discrete trap level obtained from the $\log_{10} I_g$-vs-$1/T$ plot was shown to be related to the temperatures at which the peaks in the emission I-T curve and the generation I-T curve occurred, thus providing a means for cross checking the results. In addition, it has been shown that the so-called carrier lifetime in the semiconductor can be derived from the generation I-T curve.

APPENDIX A: RELATION BETWEEN EMISSION CURRENT AND TRAP DENSITY IN MOS STRUCTURE

If L denotes the width of the depletion layer from which electrons are emitted, then the rate of increase in the effective charge at the silicon-oxide interface $\dot{Q}_n$ is given by

$$\dot{Q}_n = (A/L) \int_0^L q e_n e^{-\lambda} N_t x\, dx, \tag{A1}$$

which on integrating gives

$$\dot{Q}_n = \tfrac{1}{2} q e_n e^{-\lambda} N_t L A. \tag{A2}$$

It can be shown[7] that $e_n e^{-\lambda}$ can be represented by a δ function $B\delta(E - E_{tn}^*)$ at an energy level E_{tn}^* defined by $T\beta^{-1} \int_{E_v}^{E_c} [e_n(E_{tn}^*)/T]\, dT = 1$. The position of E_{tn}^* below the bottom of the conduction band is linearly dependent on time according to

$$E_c - E_{tn}^* = (\beta t + T_0)[1.92 \times 10^{-4} \log_{10}(\nu/\beta) + 3.2 \times 10^{-4}] - 0.0155.$$

Thus, (A2) simplifies to

$$Q_n = \tfrac{1}{2} q N_t(E_{tn}^*) L A B \delta(E - E_{tn}^*). \tag{A3}$$

Now, the charge on the gate Q_g is related to the effective charge at the silicon-oxide interface by

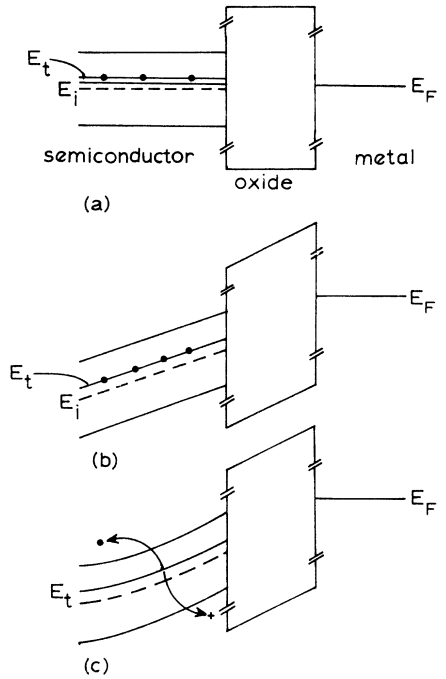

FIG. 8. Energy-band diagrams of the MOS structure illustrating (a) the equilibrium condition, (b) the non-steady-state mode at the initial low temperatures, and (c) the deep-depletion mode illustrating the generation process.

$$LQ_n = Q_g[L + (\epsilon_s/\epsilon_I)x_I], \tag{A4}$$

where L and x_I represent the widths of the depletion layer and the insulator, respectively, and ϵ_s and ϵ_I are the relative permittivities of the semiconductor and insulator. Using the relations $C_d = \epsilon_s/L$ and $C_0 = \epsilon_I/x_I$, (A4) can be rewritten in the form

$$Q_g = [C_0/(C_d + C_0)]Q_n. \tag{A5}$$

Thus taking the derivative of (A5) with respect to time, and substituting (A3) into the result gives the following expression for the current in the external circuit:

$$I_e = \frac{C_0}{C_d + C_0}[\tfrac{1}{2}qLAN_t(E_{tn}^*)] \times \beta[1.92\times 10^{-4}\log_{10}(\nu/\beta) + 3.2\times 10^{-4}]. \tag{A6}$$

APPENDIX B: WIDTH OF DEPLETION REGION AS A FUNCTION OF TEMPERATURE

The equation that describes the dependence of the depletion width (L) (through which carriers are generated) on temperature is given by (40):

$$\frac{dL}{dT} = -\frac{\nu_p \epsilon_s N_t}{\beta C_0 N_d} e^{-E_t/kT} \tag{B1}$$

or

$$\int_{L_0}^{L} dL = -\frac{\nu_p \epsilon_s N_t}{\beta C_0 N_d} \int_{T_0}^{T} e^{-E_t/kT} dT. \tag{B2}$$

Consider the integral

$$\int e^{-E_t/kT} dT. \tag{B3}$$

Using the substitution $x = E_t/kT$ and integrating by parts we obtain

$$\int \frac{e^{-x}}{x^2}(1 + 1/x)dx = -\frac{e^{-x}}{x^2}, \tag{B4}$$

or, after substitution

$$\int \left(1 + \frac{kT}{E_t}\right) e^{-E_t/kT} dT = \frac{kT^2}{E_t} e^{-E_t/kT}. \tag{B5}$$

For $kT \ll E_t$, which is normally the case,

$$\int e^{-E_t/kT} dT \simeq \frac{kT^2}{E_t} e^{E_t/kT}. \tag{B6}$$

Substituting (B6) into (B2) yields

$$L = L_0 - \frac{\nu_p \epsilon_s N_t kT^2}{\beta C_0 N_d E_t} e^{-E_t/kT}. \tag{B7}$$

---

*Supported in part by the Defense Research Board and the National Research Council of Canada.

[1]C. Jund and R. Poirier, Solid-State Electron. **9**, 315 (1966).
[2]J. Grossvelet, C. Jund, C. Motach, and R. Poirier, Surf. Sci. **5**, 49 (1966).
[3]G. Clement and M. Voss, Surf. Sci. **11**, 147 (1968).
[4]F. P. Heiman, IEEE Trans. Electron Devices ED-14, 781 (1967).
[5]H. Preier, IEEE Trans. Electron Devices ED-15, 990 (1968).
[6]Note that the function $e^{-\lambda}$ is symmetrical about the middle of the band gap when $\nu_n = \nu_p$ [Fig. 3a]. However, because the substrate is n type, there are no holes in the traps in the lower half of the band gap, which means that, hole emission cannot exist. Hence, the actual details of $e^{-\lambda}$ in the lower half of the band gap is of no concern to us here until E_{tp}^* and E_{tn}^* (see below) approach midgap, in which case the symmetrical nature of $e^{-\lambda}$ causes the function per se to peak sharply at midgap and then to ultimately collapse [see Fig. 3(a)].
[7]J. G. Simmons, G. W. Taylor, and M. C. Tam, Phys. Rev. B **7**, 3714 (1973).
[8]J. G. Simmons, and G. W. Taylor, Solid-St. Electron. (to be published).
[9]J. G. Simmons and G. W. Taylor, Phys. Rev. B **5**, 1619 (1972).
[10]The integral $B = \int_{E_v}^{E_c} e_n e_p (e_n + e_p)^{-1} dE$ is readily solved using the substitution $x = e^{E_t/kT}$ yielding $B = kT(\nu_n \nu_p)^{1/2} xe^{-E_g/2kT}[\pi/2 - \tan^{-1}(\nu_p \nu_n)^{1/2} e^{-E_g/2kT}]$.
[11]L. R. Weisberg and H. Schade, J. Appl. Phys. **39**, 5149 (1968).
[12]H. Schade and D. Herrick, Solid-State Electron. **12**, 857 (1969).
[13]D. Eirug Davies and S. Roosild, Appl. Phys. Lett. **18**, 548 (1971).
[14]R. Williams, J. Appl. Phys. **37**, 3411 (1966).