# Interpretation of amorphous semiconductor threshold switching based on new decay-time data

Gary C. Vezzoli

*U.S. Army Armament Research and Development Command, Applied Sciences Division, Applied Physics Branch, Dover, New Jersey 07801*

(Received 7 May 1979)

Recent work on threshold switching amorphous semiconductors has shown that the on-state current-voltage characteristics are essentially linear provided the subholding voltage time ($\tau_{shv}$) does not exceed $\sim 12$ ns. For $\tau_{shv} > 12$ ns the $I$-$V$ curve displays the well-documented transient on characteristics (known as the TONC) having the familiar barrier offset voltage. In this paper an interpretation for the transition from linear to TONC behavior is given in terms of two alternative mechanisms for the initiation and sustenance of the on state—that of single recombinative injection and that of a Mott screening transition. Experimental data for the three major time constants $\tau_1 \approx 12$ ns, $\tau_2 \approx 150$–$200$ ns, and $\tau_3 \approx 1.8$ $\mu$s are reviewed and interpreted. Calculations are given for carrier concentration and decay and for Mott degeneracy. The disappearance of the blocked on state is shown to be predictable from the previously reported coupled-carrier equations.

## I. INTRODUCTION

It has been known for quite some time that if in an amorphous semiconductor diode, a bias less in magnitude than the bias sufficient to maintain the on state is applied after switching, then the device will exhibit a "blocked" on state of very high resistance from which the low-differential-resistance on state can be quickly obtained again without reswitching. The blocked on state of a threshold switch refers to the higher-impedance subregime of the current-voltage ($I$-$V$) characteristics as determined from short transient pulses applied to the post switching on state of the device. The resulting $I$-$V$ characteristics are referred to as the TONC and are shown in Fig. 1(a). Recent work has shown that, in an amorphous-alloy (Ovonic type) threshold switch, the on-state characteristics do not display a typical TONC behavior provided the subholding voltage time ($\tau_{shv}$) does not exceed about 12 ns,[1] which is the time required for the high-resistance on state to develop. In this regime the characteristics are instead essentially linear as depicted in Fig. 1(b), and do not reflect the blocked on state and offset voltage of the TONC shown in Fig. 1(a)[2] which are obtained only for $\tau_{shv} > 12$ ns. The purpose of this paper is to interpret these data in terms of two models for the mechanism that initiates and sustains the post switching high-conductance on state—the injection model and the field-induced phase-transition model.

## II. BACKGROUND TIME-CONSTANT DATA

From on-state studies of zero-voltage interruption and subsequent recovery,[3] and from normalized data on transient off-state conductance versus high frequency and versus ambient temperature,[4,5] three time constants were clearly distinguishable. The normalized conductance data taken on the transient on state suggested a fourth time constant considerably faster than the other three. The recovery work[3] yielded two of these four time constants. These represented the maximum zero-voltage time with which the initial on state is maintained without need of any reswitching transition and secondly, the zero-voltage interruption time which required a subsequent reswitching pulse of essentially the full initial threshold voltage. These are here, respectively, designated $\tau_2$ and $\tau_3$, and experimentally determined as[3] $\tau_2 \approx 2\times10^{-7}$ s, and $\tau_3 \approx 1.7\times10^{-6}$ s. High-frequency work[4] independently concurred with the above measurements and yielded $\tau_2 \approx 1.5\times10^{-7}$ s and $\tau_3 \approx 1.8\times10^{-6}$ s. The normalized conductance studies indicated the existence of two other time constants which were faster than both $\tau_2$ and $\tau_3$, and are here designated as $\tau_1$ and $\tau_0$. The data gave $\tau_1 \approx 10^{-8}$ s and $\tau_0 \leq 10^{-9}$ s, and corroborated $\tau_2 \approx 2.5\times10^{-7}$ s. Unlike $\tau_1$, $\tau_2$, and $\tau_3$, whose measurement involved passing through the blocked on state, $\tau_0$ pertains strictly to the high-conductance regime of the on state. At the time that this was first observed, the interpretation of the very fast time constant was not definitively discussed. Quoting from the relevant Ref. 4,

"The normalized curves and the experimental-interruption time study indicate two decay processes probably associated with free carrier decay in the transient off state, the room-temperature time constants being about $10^{-8}$ s and $2\times10^{-7}$ to $3\times10^{-7}$ s. This suggests that an event (bias change or interruption) of the order of $10^{-9}$ s should not affect decay in the on state sufficiently to cause a transition to the transient off state."

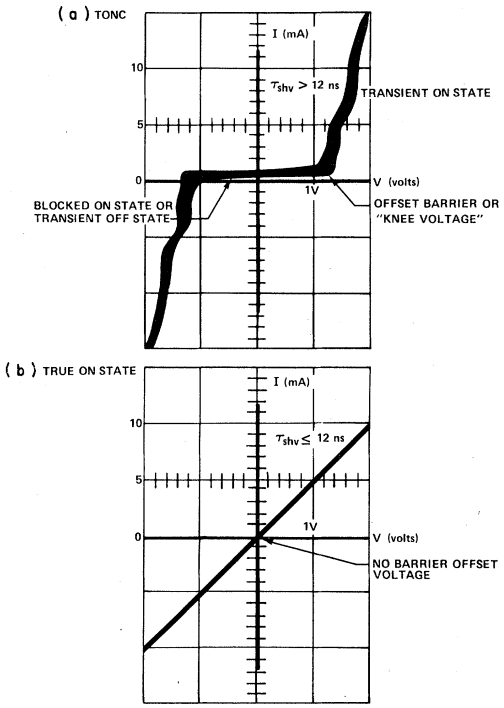


FIG. 1. On-state $I$–$V$ characteristics for (a) $\tau_{shv} > 12$ ns (blocked on state). ($\tau_{shv}$ is the subholding voltage time.) (b) $\tau_{shv} < 12$ ns (true on state).

(Note that in this reference what is referred to as "transient off state" is now called "blocked on state.")

A recent, more precise determination of $\tau_1$ gave the value of 12 ns as the maximum subholding voltage time interval prior to the development of the blocked on state or the offset voltage of the TONC.[1] Thus what are felt to be the most accurate measurements of the four time constants that are associated with Ovonic threshold switching are

$$\tau_0 \leq 1 \text{ ns}, \quad \tau_1 \simeq 12 \text{ ns}, \quad \tau_2 \approx 150 \text{ ns}, \quad \text{and} \quad \tau_3 \approx 1.8\ \mu\text{s}.$$

Thus the observed time constant $\tau_0$ that is associated with only the high-conductance on state[4] now becomes far more comprehensible as a decay constant which is faster than the interval that is required to develop the blocked on state $\tau_1$.

It is part of the purpose of this paper to evaluate the explanations of these time intervals in terms of injection and also in terms of a field-induced phase transition or Mott transition in order to compare the two models in regard to their explanation of the on state. Figure 2 reviews the data yielding the four time constants. From the standpoint of the Mott transition, it would seem that $\tau_1 \approx 12$ ns could represent the interval of subholding voltage time during which a sufficient number of carriers can recombine such that the existing field cannot mobilize free carriers in the conduction band.

## III. INJECTION

The injection models for describing the on state arose at least to some degree from the uncovering of the transient on-state behavior shown in Fig. 1(a), and was first reported by Henisch and coworkers.[2] This TONC behavior was very unlike that of any known thermistor curve. However, the observation of the offset voltage (often called the barrier or knee voltage) suggested the exploration of a barrier-tunneling or contact potential phenomenon. Such a behavior then suggests a possible injection state occurring after activation is complete. This would be in harmony with the very low average electric field present in the on state because the greatest proportion of the potential would necessarily fall in a very narrow metal-semiconductor interface near the electrodes (double injection[6]), or in a narrow region near the anode called the recombination front (single injection).[7] The former process is actually called trap-limited space-charge-saturation double injection and demands nonequilibrium electronic conditions throughout most of the interelectrode spacing with well-separated quasi-Fermi levels. This mechanism requires exactly equal injection (neutralization of positive and negative carriers) from junction contacts at the respective electrodes and is, moreover, dependent on the saturation or filling of all available trapping sites. The second process is referred to as recombinative space-charge injection and displays appreciable recombination only in the narrow recombination front. Materials exhibiting this behavior must possess in their off state a dielectric relaxation time which exceeds the mean carrier lifetime ($\tau_d > \tau_l$) —a usual characteristic of amorphous semiconductors of high resistivity. Figure 3 depicts the interelectrode voltage profile for double and single injection, comparing it to the simple linear thermal behavior and to the Mott transition.

In both injection theories the on state must be viewed as a semiconductor with either an abnormally high charge density or an unusually high carrier mobility. Since conventional four-terminal measurements, which avoid electrode-contact effects, cannot be conducted fruitfully in thin-film geometry, and since bulk samples show severe parasitic and obscuring thermal effects, the dependence of electrical conductivity on the field and the temperature is not yet precisely

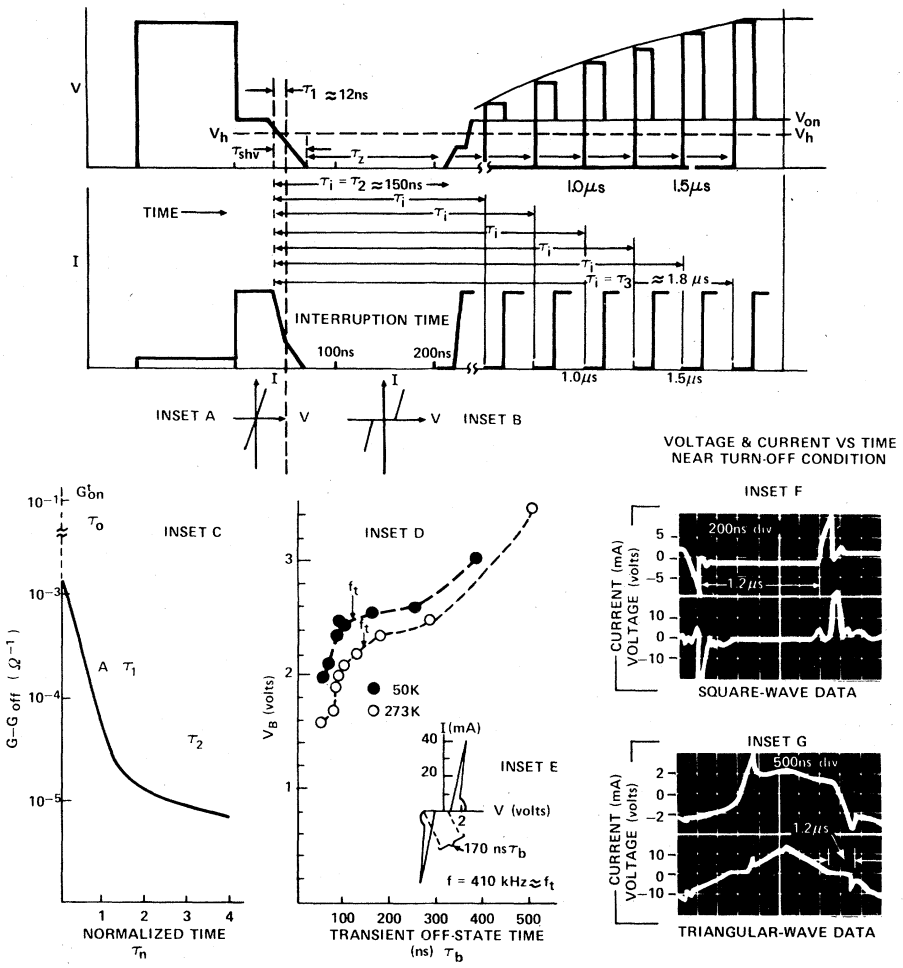


FIG. 2. Line drawing in Fig. 2 gives voltage vs time and corresponding current vs time obtained when on state is interrupted for time $\tau_i$. The interruption time is divided into two segments $2\tau_{shv}+\tau_z$ where $\tau_{shv}$ is the subholding voltage time or the time between the holding voltage and zero voltage, and $\tau_z$ the zero-voltage time. It is observed that when $\tau_{shv} < 12$ ns, the $I$-$V$ is linear (inset $A$) with no barrier offset voltage. However, when $\tau_{shv} > 12$ ns, the $I$-$V$ is that of the TONC (inset $B$) with a characteristic offset voltage which specifies the end of the blocked on state. The existence of a decay time $\tau_1$ of the order of 10 ns was shown in the normalized conductance data (Refs. 4,5) shown in inset $c$. Thus, these observations indicate a time constant $\tau_1$ of the order of 10–12 ns. Line drawing also shows that when $\tau_i \lesssim 150$ ns, no reswitching is necessitated to maintain the on state. As $\tau_i > 150$ ns the reswitching voltage threshold increases. Inset $D$ shows that when the time $\tau_b$ in the blocked on state (horizontal $I$-$V$ region in inset $B$) exceeds 150 ns, the curve of barrier or offset voltage vs $\tau_b$ shows an inflection, and inset $E$ shows that when $\tau_b \approx 170$ ns, a negative differential resistance region occurs in the $I$-$V$. The frequency for the inception of this region is designated as the turnover frequency $f_t$ in inset $D$. The normalized conductance data of inset $C$ show a distinct region corresponding to a normalized decay time of order $10^{-8}$–$10^{-7}$ s. Thus the corroboration of these data indicate that there is a time constant $\tau_2 \approx 150$–170 ns. Finally, when $\tau_i = 1.8\ \mu$s, the reswitching threshold becomes substantially equal to the initial threshold. Insets $F$ and $G$ show that at high frequency the maximum time allowed in the transient off state prior to a switchoff is 1.2 $\mu$s. This is because at $\tau_b = 1.2\ \mu$s the barrier voltage has risen to the maximum cw amplitude, thus precluding reswitching on. Extrapolation indicates that at about $\tau_b = 2\ \mu$s the barrier voltage would be equal to the original threshold, thus confirming another time constant $\tau_3$ of about $1.8 - 2.0\ \mu$s.

established. Thus, the quantitative separation of carrier charge distribution and carrier mobility has not been feasible. However, upon consideration of the reasonableness of localized states, of valence alternation pairs, and of tails in band structure, the hypothesis of a mobility gap and a mobility transition was suggested. This was called the Anderson transition and is treated within the brilliant theory of Anderson localization. This transition represents the mobility analog of the Mott carrier-concentration screening transition. (The antithetical thermal view-

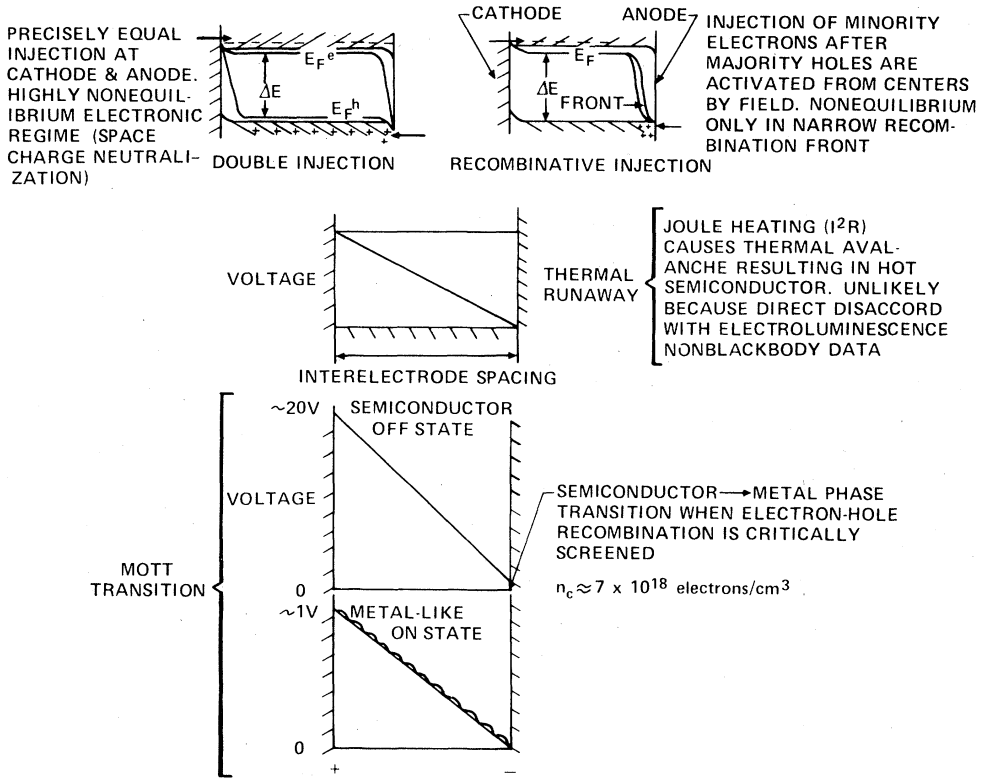


FIG. 3. Interelectrode potential profiles for injection, thermal, and Mott-transition mechanisms.

point for the mechanism of threshold switching would still view the on state as a semiconductor but a very hot one due to joule heating along the filament to about 500 °C, at which the ambient-temperature dependence of the conductivity is virtually negligible.)

Previous work in our laboratory[5,8] and elsewhere[9] has shown that appreciable recombination occurs during the threshold on state, and is indicated in terms of radiative emission of either full- or half-gap energy and of a nonblackbody character.[5,8,9] This suggested to us that if the explanation of threshold switching were cast in terms of injection, the recombinative-single-injection description advanced by van Roosbroeck[7] was given support. Further suggestion for the predominance of an effect favoring a single electrode and carrier—in particular, potential drop mostly at the anode and on-state conductivity due mostly to electrons—was provided by the increased intensity of electroluminescence when the transparent electrode was of positive polarity.[10] Because double injection is a very specific case of injection that requires balanced injection with near neutrality in the bulk, and since the on state of an Ovonic device appears, instead, to entail space charge, and because in its off state an Ovonic device appears to be a relaxation semiconductor (dielectric relaxation time $\tau_d$ is greater than the diffusion length lifetime $\tau_l$), the injection case will be discussed in terms of recombinative single injection.

The interpretation in terms of single injection is as follows: The interelectrode space in the single-injection regime is essentially one of space charge with near-zero recombination except in the narrow recombination front near the anode, where the Fermi level splits into two quasi-Fermi levels. The existence of a space-charge regime is supported by the observation that the offset barrier voltage in the TONC decreased *parabolically* as the time in the blocked on state $\tau_b$ decreased from 300 to 60 ns.[4] These data could not be smoothly extrapolated to the origin (or to zero offset voltage for zero blocked on-state time). This suggests the existence of an abrupt slope change in the curve of offset voltage versus subholding voltage time, $\tau_{shv} = \tau_b + \tau_{\text{fall time}}$. To reduce $\tau_b$ to about zero, an undistorted square wave of fall time about $10^{-8}$ s was required, thus forshadowing the more recent measurement of $\tau_1 \approx 12$ ns.

In terms of this recombinative space-charge injection, $\tau_1$ can be interpreted as the relaxation time of the on state, or the time required for the recombination front which is initially near the

anode to separate into two recombination fronts near the two contacts after the on-state bias or the holding voltage is suddenly removed. These fronts quickly become high-resistance fronts by decay of electrons out through the contacts. Then, $\tau_2$ (150 ns) can be viewed as the time required for the resulting electron distribution of the TONC to decay by further flow of electrons out through the two high-resistance fronts. Subsequent to interruption of holding voltage for ~150 ns, a reset pulse is required to restore the on state as described by van Roosbroeck.[11] Thus, the recombinative injection of electrons, having occurred during the switching, is invoked to explain the presence of the offset followed by its absence as $\tau_{shv}$ decreases. Presumably, then, $\tau_3$ is the relaxation time at thermal equilibrium.

In the above approach, the temperature *independence* of the low differential resistance of the true on-state $I$-$V$ regime[10] (as well as the temperature *dependence* of the blocked on-state differential resistance[4,5]) is explained in terms of a change in current producing a change in the positive space charge in the recombination front over which occurs nearly all of the potential drop. (The space charge has the sign of the less mobile carrier, since in the recombination front, current is mainly due to drift with electron and hole drift currents nearly equal, so that holes are the more numerous.) If one considers that the change in space charge under forward bias in the on state is due to changes in concentrations of mobile carriers, then the change in potential is precisely proportional to the change in forward current. The result is then a constant differential resistance $\tilde{R}$ in the front.[11] The blocked on state develops after the front splits into two fronts, near the respective contacts and generally after some initial decay with the time constant $\tau_0$ of electrons through these fronts to a certain critical concentration, their resistances increase, and large $\tilde{R}$ results.

Concerning the injection model, it is perhaps worth remembering that space charge in the on state gives rise to no depletion regions which could, as in certain semi-infinite cases, give instability of compensating space charge. This is because the material is itself essentially of minimum conductivity, and injection everywhere increases local conductivity.

## IV. PREDICTIONS FROM COUPLED-CARRIER EQUATIONS

The recent measurements of the disappearance of the TONC and barrier offset are in harmony with two-carrier coupled differential equations which can describe a wide range of threshold behavior.[12] These are the equations that apply if the usual simplifying assumption is made that spatial variations of all concentrations in the diode do not affect its circuit properties and may be neglected. Then the distance variable does not appear, and time is the only independent variable, and behavior in the circuit is ascribed simply to properties and behavior that are uniform in the bulk. For these equations in the present context, $n_1$ is the concentration of localized carriers, $n_{1c}$ is the concentration of localized carriers required to fill a critical number of traps leading to the switching transition, $n_2$ is the concentration of free carriers (which are for the most part created from $n_1$), and $n_{2c}$ is the critical concentration of free carriers when the on-state channel propagates from cathode to anode and the switching transition is initiated.

From the present known data we should expect the equations to predict that for decay times of the order of 10 ns or less, the differential conductance of the device should be that of the on state since the transient off state has not had time to appear or develop. It should also be expected that for very fast switching (incubation times of less than 10 ns) the carrier equations should provide for the possibility $n_1 = n_{1c} = 0$, that is, for the free carriers to be directly created from field emission from the valence band.

From Ref. 12 the conductance of the on state (state $A$) is given by

$$G_{on} = (\tau_{n2}/\tau_{n1})[G_{off}(T_0)], \tag{1}$$

where $\tau_{n1}$ is the emission time for trapped electrons and $\tau_{n2}$ is the capture time for free electrons.

The temperature $T_0$ (~490 K) is the experimentally observed temperature at which the conductance of the blocked on states and the normal off state become equal. This can be interpreted as the temperature at which nearly all of the electrons are activated from traps. As the temperature increases, the number of localized trapped carriers must decrease, and the number of activated free carriers will increase. With an independent interpretation, we note that the coupled carrier equations predict that the conductance of the on state is very nearly proportional to the electron concentration $n_2$, and consider $n_1$ as the concentration of electrons trapped at low temperature. Then $n_1$ would also represent the concentration of free electrons at $T_0$. Hence the conductance of an on state corresponding to some other free-electron concentration $n_2$ is simply represented by $(n_2/n_1)G_{off}(T_0)$.

If it is assumed that a quasi-steady state applied, so that the capture and release rates $R$ are approximately equal, then it follows that $n_1/\tau_{n1} = n_2/\tau_{n2} = R$. Rearranging gives us $n_2/n_1 = \tau_{n2}/\tau_{n1}$, and Eq. (1) follows. It is then clear in the context of the coupled equations that the blocked on state cannot develop unless the emission time $\tau_{n1}$ has elapsed so that trapped concentration $n_1$ can decay and no longer support the creation of mobile carriers $n_2$, and provide as well trapping vacancies for the $n_2$ carriers. The emission time $\tau_{n1}$ can with the model thus be identified with the time constant $\tau_1$.

Turning now to the condition of very fast switching and the hypothesis that complete trap filling is not necessarily required for the initiation of $n_2$, we write for the high-overvoltage–short-time-delay case [from Eq. (34) of Ref. 12(a)]

$$t_D = \tau_2(I_{1c}/I_v)[\ln(I_{2c}/I_{2i})]. \tag{2}$$

This equation should really be cast in terms of charge carriers $n$; it is easier to handle, however, in terms of measurable current. The critical and initial free carrier currents are $I_{2c}$ and $I_{2i}$, respectively. When $I_2$ becomes equal to $I_{2c}$, the charge carrier concentration of $n_2$ is sufficiently high to stabilize the on state across the interelectrode gap. $I_{1c}$ represents the virtual current connected with localized sites or states when the local carrier concentration $n_1$ has attained the level such that further field emission causes trapped carriers to empty into the conduction band. This would occur at a certain concentration because the localized internal field around trapping sites opposes further trapping unless an already trapped carrier is released. The intensity of the field at this condition causes the released carrier to be promoted into a conduction band as a free carrier. The term $I_v$ in Eq. (2) simply represents the total current response to a voltage $V$.

For very high overvoltage, $I_v$ will increase but remain finite. The argument of the natural logarithm in Eq. (2) will remain approximately constant or will increase because of the very high currents observed in the fast-switching regime.[13] $\tau_2$ refers to the lifetime of the distribution and not to the rise time of $n_2$, hence will not be affected by short pulses of high overvoltage. Hence, $t_D$ will effectively decrease if $I_{1c}$ also decreases. This could be interpreted in two ways. Either there is no localized conduction of holes because the traps are saturated and there is not ample time for electrons to be released from traps, thereby causing further electrons emitted from the valence band to populate directly the conduction band with $n_2$ carriers, or there is essentially no trapping and a localized or a hole current plays no part in switching in the fast regime. In either case if $I_{1c}$ approaches zero, then fast switching appears to bypass the blocked on state, and conversely, the very fast interruption mode $\tau_{shv} < 12$ ns should do the same. In reality, under the steady state the $n_1$ concentration should always be finite because after the on state is formed, some carriers must fall into localized states until a quasiequilibrium is reached where further recombination is opposed perhaps by the high $n_2$ carrier concentration. If electrons are created directly from the valence band in the very fast regime, then the on state is produced by a mechanism other than, though a consequence of, the normal switching mechanism. Then from Ref. 12(a), $V_h \cong G_{off}^{-1}(\tau_0)I_{1c}(\tau_0) \to 0$, or the holding voltage is inconsequential in the fast regime.

## V. CALCULATION OF NUMBER OF CARRIERS THAT MUST RECOMBINE OR BE IMMOBILIZED TO CAUSE DEVELOPMENT OF A BLOCKED ON STATE

We can now calculate the change in carrier density which is required to develop the blocked on state. This will occur when the generation rate minus the recombination rate fall to a critical level. The excess of the volume generation rate over recombination rate equals the time rate of increase of the free electron concentration:

$$G - R = \left(\frac{dn_2}{dV}\right)\left(\frac{dV}{dt}\right)$$

where $N_2/\mathcal{V} = n_2$. The blocked on state develops for $\tau_{shv} > 12$ ns, hence $dV/dt = \Delta V/\Delta t = V_h/12$ ns $= 8\times10^7$ V/s. We also know that the on state is ohmic or linear with typical differential resistance of about 20 Ω. Then, $dI/dV = \frac{1}{20}$ A/V $= 0.14$ (Coul/V sec) $= 10^{17}\ e/$(V sec). The current density can then be written as $dJ/dV = d(I/A)/dV$. The area $A$ of filament cross section for filament radius of 12 μm is about $5\times10^{-6}$ cm$^2$. Thus $dJ/dV = 2\times10^{22}\ e/$(V sec cm$^2$). This must now be multiplied by the voltage drop of 1V and the time interval of voltage drop, 12 ns. Then $(dJ/dV)(\tau_{shv} = 12\text{ ns})\ (V_h = 1\text{ V}) = 2.4\times10^{14}\ e/\text{cm}^2 = dN_2/dA$. We then rewrite as $dN_2/d\mathcal{V} = (dN_2/dA) \times (1/\Delta l)$, where $\Delta l$ is the filament length. Thus $dN_2/d\mathcal{V} = 2.4\times10^{18}\ e/\text{cm}^3$.

Therefore, for the blocked on state to be developed the mobile carrier concentration must fall a magnitude of $2.4\times10^{18}$ carriers/cm$^3$. That is, if recombination or immobilization reduces the free carrier concentration at a rate greater than the existing field can generate free carriers, such that the mobile carrier concentration falls about $2\times10^{18}$ carriers/cm$^3$ (when $\tau_{shv} > 12$ ns), the blocked on state will develop. In the time

interval of about 12 ns the free carriers cannot recombine into the valence band because from recovery and electroluminescence studies such a recombination requires 150 ns. Thus it is this author's contention that the recombination occurs into a trapping band or into a localized state. The sites in the trapping band are vacant to accept free electrons because the lifetime of the local $n_1$ carriers is about 10 ns, hence during $\tau_{shv}$ they ($n_1$) can decay into the valence band. The alternative explanation would suggest that the free carriers simply become a low-mobility space charge during $\tau_{shv} > 12$ ns and do not physically recombine (such as in an exciton state).

## VI. INTERPRETATION IN TERMS OF FIELD-INDUCED PHASE TRANSITION

### A. Calculation of degeneracy

In this section we examine the characteristics of threshold switching which are common to a field-induced phase transition such as a Mott transition.

In a Mott insulator the Coulomb potential $V(r)$ for the attraction of a single free electron and a single free hole is $V(r) = -(e^2/\epsilon_0 r)$. If the effect of screening were included, the above equation must be modified to $V(r) = (e^2/\epsilon_0 r)\exp(-\alpha r)$. The lowest-energy state of this interaction corresponds to an exciton or a hydrogenic bound state preventing positive and negative conduction. In this relationship the screening parameter $\alpha$ is not a constant but increases monotonically with free carrier concentration $n_2$. At the critical carrier concentration (which we call $n_{2c}$) the attraction between electrons and holes (resulting in recombination under low-concentration conditions) becomes so heavily screened that a large number of carriers flood the conduction band and maintain the on state. The carrier density in the on state then stabilizes at about $10^{19}$ cm$^{-3}$.

The equation for Mott degeneracy, giving the carrier concentration $n_c$ for critical screening is $n_c \gtrsim (4a_0\epsilon m/\epsilon_0 m^*)$, where $m$ and $m^*$ refer to the rest mass and the effective mass of the electron. $\epsilon$ and $\epsilon_0$ are the dielectric permittivity of the switching material and of a vacuum, respectively.

If a 3-coordinated selenium site or tellurium site in a chalcogenide glass is assumed to be the pertinent trap for the electron-hole attraction that is to be screened, then this site acts as a hydrogenic donor because of the single antibonding electron tending to lower its energy by being donated to a neighbor.[14] Peterson and Adler[14(b)] calculate that the Bohr radius for such a carrier is 13 Å and that the donor level is only 0.05 eV below the conduction-band minimum. This means that the sphere of influence is about equal to the short-range order of the glass, thus allowing the transition from localized to extended (Bloch-type) states, and it also means that transitions from donor to conduction band or recombinations are readily accomplished. On this basis the Mott criterion for degeneracy of the trapping band is $n_c \geq 7\times10^{18}$ carriers/cm$^3$. This is about the value of the charge carrier density in the on state ($10^{19}$ electrons/cm$^3$) where we postulate $n_2 = n_{2c}$, and is about a factor of 3 larger than the critical decrease in free carrier concentration required to develop the blocked on state.

If the above analysis is applicable to threshold switching in chalcogenide glasses, then the screening transition takes place in the band in which the $n_1$ carrier becomes trapped, and the buildup of $n_2$ remains in the impurity band until the concentration is sufficiently large to cause the screening-induced electron avalanche to the conduction band.

### B. The interpretation of the time constants using a Mott-transition model for threshold switching

The time constant $\tau_1$ should represent the emission time for electrons from localized states or from an intragap narrow band. Then at $\tau_{shv} \approx \tau_1 \simeq 12$ ns a critical number of the trapped carriers could recombine into the valence band and begin the development of the blocked on state because $n_1$ becomes less than $n_{1c}$, and the $n_2$ carrier creation rate will be decreased. The $n_2$ carrier density becomes smaller (in the conduction band), and recombination may occur to the trapping band. The major question at this point is whether or not the reentry of $n_2$ carriers into the trapping band causes the initiation of an inverse screening transition. Since at $12 < \tau_{shv} < 150$ ns there is no indication of a first-order discontinuity or catastrophic effect, it would appear that inverse screening does not occur and a *critical* recombination does not occur. The alternative is that the bulk of the $n_2$ carriers are temporarily populating a localized band where they can participate only in semiconduction under the influence of the subholding value of the electric field. The inverse of the screening transition does not then occur until a critical recombination, after which a reswitching transition is necessary to reinitiate the on state.

This interpretation leads us to believe that the $n_1$ carrier must exist during the on state because it is the decay of that carrier which triggers the transition to the blocked on state. Conversely the existence of the blocked on state without the appearance of a first-order discontinuity indi-

cates that the $n_2$ carrier is present during the blocked on state. Then in a Mott screening regime a high carrier concentration precludes either the critical recombination to the trapping band or the critical recombination to the valence band. Thus it appears that the concentration of the $n_1$ carrier governs switch-on characteristics, and the decay of the $n_1$ carrier governs the transition to the blocked on state.[15]

The time constant $\tau_2 \approx 150$ ns might then represent the zero-voltage time interval during which a critical amount of $n_2$ carrier decay and recombination to the valence band occurs and would then demand a reswitching pulse to reactivate electrons out of the valence band to reinitiate the threshold-switching process. The phrase "a critical number of traps" would really refer to a critical localized carrier concentration. Thus at $\tau_{shv} > \tau_2$, the inverse screening transition occurs.

The time constant $\tau_3$ could then be considered to represent the total zero-voltage time interval for complete decay such that the full magnitude of recharging energy is required to satisfy the conditions $n_1 = n_{1c}$ and $n_2 = n_{2c}$. In time intervals of subholding voltage intermediate between $\tau_2$ and $\tau_3$, there are still enough $n_2$ carriers such that only a partial recharging is necessary to initiate critical screening.

Thus in zero-voltage time interval $\tau_3$ the conduction and donor bands should become depopulated of $n_2$ carriers to result in the low-temperature insulating-energy-band configuration and behavior, which is what is clearly observed. In the zero-voltage time interval $\tau_1$ to $\tau_2$ the donor bands (or localized states) should be partially empty, and the valence band partially filled as is the case for semiconductors. This condition represents the observed semiconducting behavior of the blocked on state (in previous work called the transient off state). Actually there are an infinite number of these semiconducting states depending on the value of the subholding voltage or zero-voltage time interval, as has been described experimentally.[4] Each state has a typical semiconductive temperature dependence of conductivity.

Finally, in the time interval $0 < \tau_{shv} < 12$ ns, the electrical behavior of the material is similar to that of a metal, as has been shown by the virtually zero temperature coefficient of differential resistance over a range of at least 400 K. In a field-induced phase transition of the Mott type in materials where trapping is inescapably significant, the metallic state is not generated by a complete band overlap as in a pressure-induced transformation to a metallic polymorph. Accordingly, the electrically metallic on state can and does behave optically, as does a semiconductor, and exhibits electroluminescence.

The observed electroluminescence implies the existence of an optical gap. However, the temperature independence of differential resistance in a noninjection regime implies a gap of less than $12 \times 10^{-4}$ eV. Thus in the metallic regime, the population of the conduction band may occur isoenergetically from a donor-band maximum to a conduction band-minimum. This could be the explanation for the constant voltage intervals shown in Fig. 4 after subholding voltage times

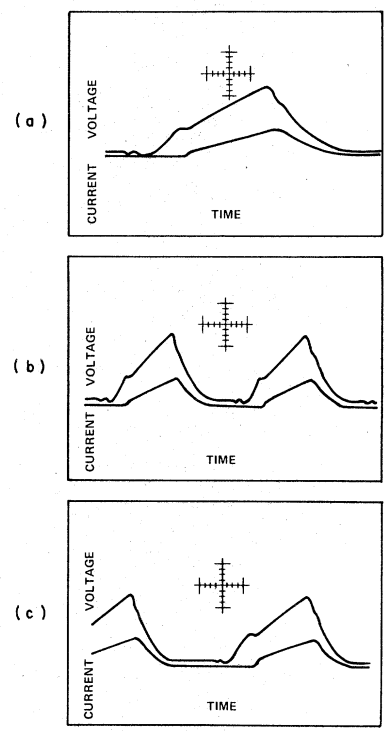


FIG. 4. Voltage (upper trace in each oscillogram, 5 V/div) vs time and current (10 mA/div) vs time for on state using subthreshold ramp cw in conjunction with set pulse. (Ramp characteristic allows variable subholding voltage time $\tau_{shv}$ simulating a zero-voltage interruption time.) (a) $\tau_{shv} = 150$ ns ($t = 100$ ns/div). Note 40-ns interval of constant voltage prior to increase in current. This constant-voltage interval may be interpreted as arising via tunneling from filled trapped band to conduction band in order to reinitiate the on-state conduction. (b) $\tau_{svh} \approx 300$ ns ($t = 200$ ns/div). Note constant-voltage interval remains at 40 ns. (c) $\tau_{shv} \approx 500$ ns ($t = 200$ ns/div). Note indication of high-frequency reswitching (falling voltage just prior to increasing current) because free-carrier lifetime has been exceeded. Reswitching does not show any tunneling appearance and may not be associated with any isoenergetic phenomenon. Note: the equivalent *zero*-voltage interruption time for cases (a) and (b) would be about 100 and 150 ns, respectively.

of about 150 to 300 ns. The interval of constant device voltage (even though applied voltage is increasing) may be a tunneling time. This is in harmony with the undulating-band model of amorphous materials advanced by Ingliss and Williams.[16]

The above suggests that there should not exist any discontinuity in $I$-$V$ characteristic at the point where the blocked on regime intersects the true on state $I$-$V$—what is called the knee voltage or barrier offset [Fig. 1(a)]. At the value of the knee voltage (the magnitude of which is slightly *less* than the critical holding voltage) the electric field is adequate to accelerate charges in the on state provided a critical time has not elapsed for those charges to decay. One would then predict that the $I$-$V$ condition at which the knee or the slope-change occurs is independent of frequency and is independent of ambient temperature provided $\tau_{shv} > 12$ ns (yet less than 150 ns). This is exactly what is observed and is reported in Ref. 4, Fig. 10(a) under the conditions that the time in the transient off state does not exceed 150 ns ($\tau_b$). For $\tau_b \geq \tau_2 = 150$ ns the knee voltage increases very slightly with increasing $\tau_b$ and with increasing temperature. This slight dependence probably occurs because a partial switch off has taken place during each ac cycle and as the cw frequency decreases (in order that $\tau_b$ increase) the level of voltage for $\tau_{shv} = 12$ ns is higher for the lower frequency carrier wave. Hence a carrier that is under the $I$-$V$ condition at the knee can belong to either the on state or the blocked on state depending upon whether it is in the conduction band or in the trapping band. This also suggests a possible tunneling-induced conduction. Thus the value of the electric field at conditions under the holding voltage is not critical; however, the time at which the field is less than the holding condition is quite critical. This, therefore, is why a carrier concentration condition is postulated to be inextricably connected to the on state.

On the other hand, it appears that a holding field or holding voltage is a significant parameter. This must be considered in the present context as the electric field that supplies the force or the energy that overcomes the electron-hole Coulombic attraction which acts on the $n_2$ carrier. Even with screening, if the $n_2$ carrier is not supplied with enough kinetic energy to overcome the attraction of the multitude of trapping sites, then that carrier will undergo decay (in an average time of about 150 ns).

At subholding voltage times just slightly less than 150 ns, a negative differential resistance appears in the true on state in the down-voltage direction of applied field [Ref. 12(a), Fig. 12(c)]. This is interpreted to be a consequence of direct decay of mobile carriers to the valence band just prior to a direct switch off to the true off state rather than a transition to the blocked on state, with which the absence of this negative resistance suggests that the carriers populate localized states.

From the above standpoints there may be a fairly simple explanation of how the high-resistance on state (blocked) comes about through decay and redistribution of the space charge in the device after a holding bias is removed. This mechanism does, in general, entail some decay of the space charge to a suitable concentration in the device such that high-resistance regions near the contacts are realized. It should be noted that without such regions, the space charge would continue to decay rapidly through the contacts. However, if the coupled equations are to give a good description, it must be because the redistribution takes place as quickly as, or more quickly than, the decay. This possibility is plausible in view of the large estimated carrier concentration in the on state. It is this concentration that is proposed to give a screening transition. It should also be noted that if the decay (through the contacts) is a dielectric-relaxation decay of mobile carriers, then it and the redistribution are essentially the same. Furthermore, near the blocked-on-state threshold, the space charge is largely that of trapped electrons, and its redistribution is controlled by the time constant $\tau_1$; which is the same as the electron emission time.

### C. Calculation of on-state carrier concentration from high-frequency on-state data

The on-state carrier concentration cannot be calculated from the data on off-state charging versus delay time because in the screening transition regime this would only calculate the carrier concentration required for the critical screening at the switching threshold. Such a calculation would neglect the increase in carrier concentration due to the liberation of trapped electrons. Thus the on-state concentration must be calculated from data during the fully charged on state, prior to the decay of $n_2$.

From our previous study of reinitiated time-integrated on-state mobile charge for conditions of $\tau_{shv}$ slightly less than 150 ns (most of this time interval corresponded to transient off time) we experimentally showed that the charge for a fully charged on state was equal to $\int_0^t (dq/dt)dt = 40 \times 10^{-9}$ Coulomb (Ref. 5, Fig. 13). By a fully

charged on state, it is meant that because $\tau_{shv} < 150$ ns, there is no requirement for a partial reswitching and the on state is in a quasisteady state for a certain time interval.

A charging of $40\times10^{-9}$ Coulomb in a filament of radius 25 $\mu$m and length 1 $\mu$m gives rise to a carrier concentration $n = N/\mathcal{V} \approx 2\times10^{20}$ $e/\text{cm}^3$.

This carrier concentration is about an order of magnitude larger than the generally accepted measurement of $10^{19}$ $e/\text{cm}^3$,[16] and is about four orders of magnitude greater than the concentration calculated from off-state data up to the switching transition. The low-field off-state carrier concentration is about[14(b)] $5\times10^{10}$ $\text{cm}^{-3}$. It was shown in Ref. 5 that the effect of decreasing ambient temperature caused an increase in the charge required to sustain the true on state for a given value of time in the blocked on state, which can be rewritten in terms of a given $\tau_{shv}$. The functional temperature dependence was a series of nested curves and indicated that a small proportion (about 10%) of the on-state mobile carrier concentration may be due to thermal excitation because of the relatively long periods of on-state maintenance (500 ns–1 $\mu$s). Thus it is suspected that the thermal contribution may have caused an order of magnitude increase in on-state carrier concentration as calculated above. (This includes time in blocked on state.)

In our most recent measurements[1] of the on state at or above the holding voltage conditions and for durations no longer than 240 ns, the above integrated calculation yields a value of $1.4\times10^{-9}$ Coulomb for minimum mobile charging, and hence a carrier concentration of $7\times10^{18}$ carriers/$\text{cm}^3$ for *true* on-state conditions. *This is in excellent agreement with the earlier calculation for Mott degeneracy in the trapping band.*

### D. The importance of critical temperature $T_0$

The data showing a critical temperature at which the conductances of the transient off states at frequencies from 1 MHz to 50 kHz all coalesce with that of the normal off state ($T_0 \approx 490$ K)[4,5] were one of the important observations leading to Walsh's suggestion that some type of phase transition accompanied threshold switching. The success of his original development of the coupled-carrier equations[12] to explain a wide spectrum of switching and on-state data in terms of the conditions $n_1 = n_{1c}$ and $n_2 = n_{2c}$ then suggested to him the possibility of a screening-type transition. The coupled-carrier differential equations can be derived from the current continuity equation $\vec{\nabla}\cdot\vec{J} = -\dot{\rho}$ (here $J$ is the displacement current part of total current density and $\rho$ is a space-charge density) and from Poisson's equation in the presence of two carriers,[9(b),11] rather than being simply phenomenological, as originally assumed. We are thus left with the necessity of explaining the existence of $T_0$ in a regime dominated by the electric field and also, in the author's opinion, by the carrier concentration.

The interpretation given herein leads to attributing $T_0$ to the temperature at which $n_1$ can build up to $n_{1c}$ thermally, but with no switching occurring because the zero-field condition does not mobilize trapped carriers, nor bend energy bands, nor create sufficient screening. The thermal release time fromthese traps for a value of $E_{\text{conduction}} - E_{\text{trap}}$ (assumed to be about 0.4 to 0.5 eV) would be about 0.1 ms to about 10 ms. This corresponds very nicely to the eventual asymptotic regime for the total recovery of the initial off state shown in the diagnostic probing pulse data and the rare-double-pulse data (see Fig. 2, top plateau region of recovery curve, and Refs. 2–4, 17).

Thus, because of thermal trap filling, the temperature $T_0$ is the value at which a linear field extrapolation of $\Delta E(r) = \Delta E - \beta T$ approaches zero (as stated in Ref. 12). This is not to be interpreted in the conventional sense of "wiping out" a gap, but in the sense that the imposition of an electric field, however small in magnitude, will nonetheless at $T = T_0$ populate the conduction band in a manner not affected by a further increase in ambient temperature (unless, in the amorphous case at hand, a devitrification occurs). The absence of further temperature effects reinforces the likelihood of a tunneling effect promoting the population of the conduction band and thus increasing $n_2$ until $n_{2c}$ is achieved, whereupon screening may be sufficient to flood the band with free carriers through tunneling or percolation.

## VII. THE RELATIONSHIP OF CARRIER LIFETIME TO THE DIELECTRIC RELAXATION TIME

The dielectric relaxation time $\tau_d$ is actually the time required for the net field in the material to decrease to zero. It is clear that in the off state the relaxation semiconductor condition $\tau_d > \tau_l$ should apply, since resistivity studies[2] indicate $\tau_{d\ \text{off state}} \approx 10^{-6}$ s. Furthermore, if we include mobile carriers and postulate that the average carrier diffusion length lifetime is between 12 and 150 ns, then the relaxation semiconductor condition still holds. As the increasing electric field continues to empty filled traps, the magnitude of $\tau_d$ will decrease while the resistivity is decreasing. In the on state the value of $\tau_d$ (taken from resistivity and capacitance data[1])

is generally accepted to be $10^{-10}$ to $10^{-11}$ s. Thus in the on state the relationship of $\tau_d$ and $\tau_l$ is reversed (i.e., $\tau_d < \tau_l$).

Two important implications can be derived from the on-state relationship of $\tau_d$ and $\tau_l$. First, a condition must have been passed through at which $\tau_d = \tau_l$ and this could indeed specify the switching condition. In terms of injection, this would mean that when $\tau_d = \tau_l$ the polarization distance diverges and electrons will separate from holes. This results in minority electron injection and the development of the true on state. In terms of a Mott transition, when enough carriers build up to screen the Coulombic potential, the resistivity will have decreased, forcing a decrease in $\tau_d$. At the condition $\tau_d = \tau_l$ the exciton attraction is overcome, and the electrons and holes become separated, with the electrons rapidly populating the conduction band.

The second implication is that because $\tau_{d(\mathrm{on})} < \tau_{l(\mathrm{on})}$, by at least two orders of magnitude, any voltage excursion from $V = V_h$ to $V = 0$ or vice versa, in time $\tau_1 \approx 12$ ns, should be visible. Thus if a contact potential does exist, then it must be somehow affected by a reduction of voltage beneath $V_h$ for a time interval $t$ such that $\tau_d < t < \tau_l$. This would, on the other hand, not be expected if $\tau_d > \tau_l$ in the on state; a change in potential in time interval $\tau_1$ would be invisible, instead, at the contacts, and ohmic behavior would result.

However, we find that even though the contact potential should redistribute itself in the time interval $\tau_d < t < \tau_l = \tau_1$, there is no evidence of nonohmic behavior in terms of development of an offset [Fig. 1(b)]. The underscore of this argument is thus the strong indication that the development of a contact potential should take a far shorter time than the observed $\tau_1 = 12$ ns. The above description has been developed to counter the argument that the observed subholding voltage time interval of 12 ns merely means a determination of the time interval for the barrier contact potential or resistance barriers to develop.

## VIII. THE TEMPERATURE INDEPENDENCE OF THE RECOVERY CURVE

It is well known that the recovery curve characteristics, or more explicitly the experimentally determined value of $\tau_3$, have been shown to be essentially temperature independent over the range of 4 to about 400 K. The number of filled traps should be considerably different at 4 than at 400 K, according to $n_{\mathrm{tr}} = N_{\mathrm{tr}}/[1 + (1/g)\exp(E_{\mathrm{tr}} - E_F/kT)]$, where $N_{\mathrm{tr}}$ is the trap concentration, $g$ the degeneracy factor, and $E_{\mathrm{tr}}$ the energy of the trap level. Thus as $T$ decreases, $n_{\mathrm{tr}}$ also decreases. In the double-injection regime complete trap-filling saturation is normally cited, and therefore it should require a different time interval to empty the saturated traps for each value of ambient temperature. However, the time for full carrier decay $\tau_3$ is independent of ambient temperature as shown in the recovery curve data. (The time interval of zero-voltage interruption which requires a full reswitching pulse is temperature independent.) Such a property is not at all semiconductorlike in behavior. Furthermore, the time interval $\tau_2$ is also temperature independent.

The above suggests that $\tau_3$ does not involve trapped carriers or trapping states but instead represents the total decay of the free carriers in the conduction band. Thus, in the time interval 1.8 $\mu$s, a calculated value of $2.4 \times 10^{18}$ carriers/cm$^3$ must recombine into the valence band. A sufficient number of these carriers recombine in $\tau_{\mathrm{shv}} > 150$ ns such that a critical screening concentration no longer exists and a reswitching transition is necessary. The actual recharging time for the subsequent switching events will be temperature dependent.

## IX. CONCLUSIONS AND SUMMARY

The major conclusion of this article is that the recently discovered time constant $\tau_1$ is the emission time of a trapped or localized carrier which is under the influence of an electric field less than the holding value. Experimentally, $\tau_1$ is the subholding voltage time required for the development of the blocked on state. Thus $\tau_1$ is clearly associated with localized carriers.

The further conclusions of this article are that the free-carrier concentration of the on state is approximately $10^{19}$ $e$/cm$^3$ as calculated from the $\tau_1$ data yielding the blocked on state, as well as from the high-frequency data. Correspondingly, the criteria for Mott degeneracy is also clearly calculated as approximately $10^{19}$ carriers/cm$^3$. This work also affirms that the other three time constants $\tau_0 \leq 1$ ns, $\tau_2 \approx 150$ ns, and $\tau_3 \approx 1.8$ $\mu$s are all associated with either free carriers in the conduction band $\tau_0$, or free-carrier recombination into the valence band (during the interval from $\tau_2$ to $\tau_3$).

## ACKNOWLEDGMENTS

The author wishes to express gratitude to Mr. L. W. Doremus, chief of the Device Physics Section at the Armaments Research and Develop-

ment Command, Professor Peter J. Walsh of Fairleigh Dickinson University, Teaneck, N. J., Professor David Adler of the Massachusetts Institute of Technology, and Mr. W. van Roosbroeck (retired) of Bell Telephone Laboratories, Murray Hill, N. J., for very beneficial discussions. The author is also indebted to the U.S. Army Reserve Mobilization Designation program for furnishing the time to correlate and write this paper while working on the theoretical high-field physics of electroceramics and glasses under Dr. James McCauley and Dr. Robert Katz at the U.S. Army Materials and Mechanics Research Center, Watertown, Massachusetts.

---

[1]G. C. Vezzoli, J. Appl. Phys. 50, 5810 (1979); Appl. Phys. Lett. 35, 288 (1979).

[2]R. W. Pryor and H. K. Henisch, J. Non-Cryst. Solids 7, 1181 (1972); H. K. Henisch, R. W. Pryor, and G. J. Vendura, J. Non-Cryst. Solids 8–10, 415 (1972).

[3]G. C. Vezzoli and L. W. Doremus, J. Appl. Phys. 44, 3245 (1973).

[4]G. C. Vezzoli, L. W. Doremus, and P. J. Walsh, in *Amorphous and Liquid Semiconductors*, edited by J. Stuke and W. Brenig (Taylor and Francis, London, 1974), Vol. I., p. 651. Note that in Refs. 4 and 5 the subscript nomenclature of the time constants is different from the nomenclature necessitated within this paper.

[5]G. C. Vezzoli, P. J. Walsh, and L. W. Doremus, J. Non-Cryst. Solids 18, 333 (1975).

[6]H. K. Henisch, E. A. Fagen, and S. R. Ovshinsky, J. Non-Cryst. Solids 4, 538 (1970).

[7]W. van Roosbroeck, Phys. Rev. Lett. 28, 1120 (1972).

[8]G. C. Vezzoli, P. J. Walsh, P. Kisatsky, and L. W. Doremus, J. Appl. Phys. 45, 4534 (1974).

[9](a) B. T. Kolomiets, E. A. Lebedev, N. A. Rogacher, and V. Kh. Shpunt, Fiz. Tekh. Poluprovodn. 6, 197 (1972) [Soviet Phys. Semicond. 6, 167 (1972)]. (b) P. J. Walsh, Fairleigh Dickinson University, Teaneck, N. J. (private communication concerning research conducted with D. Adler, Massachusetts Institute of Technology); and P. J. Walsh and M. J. Thompson, J. Non-Cryst. Solids 35/36, 1093 (1980). (c) P. J. Walsh (research conducted with M. Thompson, University of Sheffield, England), private communication; presented at Eighth International Conference on Amorphous and Liquid Semiconductors, Harvard University, Cambridge, Mass., 1979. See P. Walsh, D. Pooladdej, M. Thompson, and J. Allison, Appl. Phys. Lett. 34, 733 (1979).

[10]G. C. Vezzoli, P. J. Walsh, and L. W. Doremus, unpublished in detail, however, see Refs. 5 and 8; also see G. C. Vezzoli *et al.* Appl. Phys. Lett. 26, 234 (1975) for description.

[11]W. van Roosbroeck, J. Non-Cryst. Solids 12, 232 (1973).

[12](a) P. J. Walsh and G. C. Vezzoli, in *Amorphous and Liquid Semiconductors*, edited by J. Stuke and W. Brenig (Taylor and Francis, London, 1974), Vol. 2, p. 1391. (b) P. J. Walsh and G. C. Vezzoli, Appl. Phys. Lett. 25, 28 (1974).

[13]S. H. Holmberg and M. P. Shaw, in *Amorphous and Liquid Semiconductors*, edited by J. Stuke and W. Brenig, (Taylor and Francis, London, 1974), Vol. 1, p. 687.

[14](a) P. J. Walsh, S. Ishioka, and D. Adler, J. Appl. Phys. (to be published). (b) K. Petersen and D. Adler, *ibid.* (to be published).

[15]This suggests that two forms of $n_1$ concentration can exist within the trapping band—a higher energy less localized semiconducting state and a nonconducting bound state such as the exciton.

[16]G. Ingliss and F. Williams, Phys. Rev. Lett. 25, 1275 (1970).

[17]D. Adler and L. Flora, in *Amorphous and Liquid Semiconductors*, edited by J. Stuke and W. Brenig (Taylor and Francis, London, 1974), Vol. 2, p. 1407. See also K. E. Petersen and D. Adler, in *Amorphous and Liquid Semiconductors*, edited by W. E. Spear (Centre for Industrial Consultancy and Liaison, Edinburgh, 1977), pp. 695 and 707.