Fischler, J. Phys. Soc. Japan Suppl. 21, 230 (1966).
[33]G. D. Watkins and J. W. Corbett, Phys. Rev. 134, A1359 (1964).
[34]E. L. Elkin and G. D. Watkins, Phys. Rev. 174, 881 (1968).
[35]G. D. Watkins, in *Radiation Damage in Semiconductors* (Dunod, Paris, 1965), p. 97.
[36]G. D. Watkins, in *Radiation Effects in Semiconductors*, edited by F. L. Vook (Plenum, New York, 1968), p. 67.
[37]G. D. Watkins (unpublished).

PHYSICAL REVIEW B VOLUME 5, NUMBER 2 15 JANUARY 1972

Nonequilibrium Carrier Phenomena in *n*-Type InSb

Charles L. Dick, Jr. and Betsy Ancker-Johnson
Boeing Scientific Research Laboratories and University of Washington, Seattle, Washington* 98105
(Received 2 June 1971)

Four nonequilibrium carrier phenomena, namely, the hot-electron effect, injection, impact ionization, and bulk negative differential conductivity (BNDC), produce nonlinear conduction in *n*-InSb in a strongly interrelated manner. These interrelationships are shown in detail. The hot-electron effect causes more decrease in average conduction under most transport conditions than injection causes increase and thus injection is not so easily observed. As the electric field strength is increased into the impact ionization range which begins at $\simeq 225$ V/cm, the generation rate progressively exhibits an e^E, an $e^{1/E}$, and then an e^{1/E^2} dependence. The distribution of the carriers in energy space shifts from a peaked distribution for 225 V/cm $\lesssim E \lesssim 425$ V/cm to an isotropic distribution for $E \gtrsim 500$ V/cm. While the carriers are in a peaked distribution, the polar-optical-phonon scattering changes from small-angle scattering to large-angle as E is increased. The origin of type-*S* BNDC is explained, as is its temporal and spatial relationship to type-*N* BNDC. The *S*-BNDC results from the presence of copious excess carriers that are generated by impact ionization within the propagating high-field domain (HFD) associated with the Gunn effect and the *N*-BNDC. A low-electric-field-strength region is responsible for the *S*-BNDC, a region that propagates in the wake of the HFD associated with type *N*. A HFD in InSb makes a single transit as a result of impact ionization within the domain. The *N*-BNDC in *n*-InSb decays in typically <10 nsec, the transit time for a HFD, whereas the *S*-BNDC persists for several hundred nanoseconds. Radiative recombination measurements show conclusively that the *S*-BNDC is caused by excess carrier generation as proposed. The decay of these excess carriers is found to be characterized by two times. The faster, ~27 nsec, persists for typically ~150 nsec, a much shorter duration than the *S*-BNDC decay time, and is not understood at this time. However, the longer excess-carrier decay time prevails well beyond the *S*-BNDC disappearance, a phenomenon that is explainable by the hot-electron effect.

I. INTRODUCTION

The hot-electron effect,[1] injection,[2] impact ionization,[3–5] and bulk negative differential conductivity (BNDC)[6–8] are known nonequilibrium conduction phenomena associated with semiconductors, in particular with *n*-InSb. Nonequilibrium conduction refers here to any condition of nonlinear (or non-Ohmic) conduction that results when a voltage is applied to a semiconductor sample. The conduction may become nonlinear because carriers in excess of the number in thermal equilibrium with the lattice are introduced by the applied voltage. Nonlinearity may also occur because the transport properties of the charge carriers are modified by the applied field. A combination of these possibilities may also exist. Excess carriers are introduced into a semiconductor under the influence of an applied voltage by injection and/or impact ionization. The hot-electron effect influences the transport properties of electrons because of the attendant mobility reduction. BNDC can be the result of a change in the mobility of some charge carriers, or the result of the introduction of excess carriers, or a complicated combination of both of these possibilities.

These four nonequilibrium phenomena are strongly interrelated in *n*-InSb, a fact complicating their study, particularly because these phenomena occur on extremely short time scales. The short time scales require rather sophisticated diagnostic techniques. The measurements are further complicated by the fact that the lattice is held at a low temperature, 77 K, unless stated otherwise. This temperature is chosen because it nearly maximizes both carrier mobilities in InSb and is conveniently attained.

This work shows the interrelationship of these four nonequilibrium phenomena in *n*-InSb and gives new information about each, including a new effect: a temporal and spatial relationship between type-*N* and type-*S* BNDC.[8]

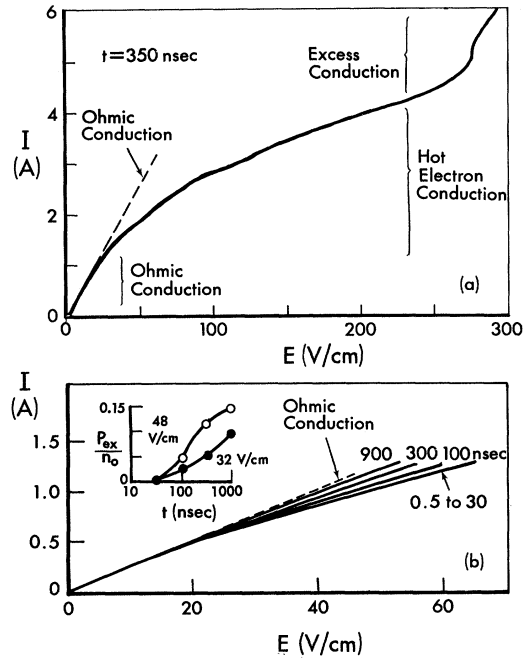

FIG. 1. (a) Conduction characteristic illustrating Ohmic, sub-Ohmic, and excess-carrier conduction, recorded 350 nsec after the application of the applied voltage. Data refer to sample No. S20-4C-15n. (b) Conduction characteristics, parametric in time, illustrating the hot-electron effect and injection. The inset shows the relative amount of injection, relative to $n_0 = 1.7 \times 10^{14}$ cm^{-3}, as a function of field strength and time. Data refer to sample No. S35-1F-11n.

II. NON-OHMIC CONDUCTION CAUSED BY HOT ELECTRONS AND INJECTED HOLES

With the application of an increasing voltage to a (single-crystal) sample of n-InSb, electron heating occurs as the average energy of the electrons is raised above kT_L; T_L is the lattice temperature.[1,9] Electron heating, or the hot-electron effect, causes a reduced electron mobility μ_n because electron scattering from polar-optical phonons increases as the average electron energy is raised. Sub-Ohmic conductivity results from the μ_n reduction, a phenomenon that is readily observable by simple conduction measurements for electric field strengths in the range 20 V/cm $\lesssim E \lesssim$ 250 V/cm, Fig. 1(a). For fields exceeding 250 V/cm, carrier multiplication dominates conduction, masking further heating of the electrons (carrier multiplication is discussed in Sec. III). Ohmic, hot-electron, and excess-carrier conduction are illustrated in Fig. 1(a).

Another phenomenon also occurs in the same E range (20–250 V/cm), namely, the injection of nonequilibrium holes, an effect which remained unobserved in n-InSb until recently[2] because of the dominance of the hot-electron effect. The influence of hole injection on the average (over the sample length[10]) conduction is generally small because the holes carry such a small percentage of the current ($\mu_n/\mu_p \approx 25$–100). In contrast, the effect on the conductivity in a narrow, e.g., 1-mm, region of the sample near the anode (where the holes are injected) can be quite large because a considerable number of holes can accumulate in such a region of the sample.[2] The uniform filling with injected carriers of a long, e.g., 1-cm, sample does not occur in n-InSb as it does so readily in p-InSb.[11] It is prevented by the majority-carrier drift velocity being so much larger than the minority-carrier drift velocity. This disparity in velocities causes the injected plasma front to reverse its direction and move back to the anode.[12]

It is possible, however, in the case of a relatively short sample, e.g., 2.5 mm, to detect an increase in conduction, as a function of time, caused by hole injection.[2] A description of the method used to measure conduction characteristics as a function of time has been published[3] and recent improvements are given elsewhere.[13] Figure 1(b) contains conduction characteristics taken to demonstrate hole injection. For 0.5 nsec $\lesssim t \lesssim$ 30 nsec, no time dependence is seen in the conduction, but for 30 $\lesssim t \lesssim$ 900 nsec, the effects of hole injection are readily apparent.

The sub-Ohmic conduction resulting from electron heating occurs on such short time scales that it cannot be resolved with presently available techniques: A simple calculation of collision time τ_c, assuming optical-phonon collisions only, yields[14] $\tau_c \approx 2 \times 10^{-12}$ sec; this result agrees well with more sophisticated calculations.[15] On the other hand, injected-carrier transit times are $\sim 10^{-7}$ sec/cm, and thus injected-carrier conduction, enhanced over the hot-electron conduction, is easily measured with standard sampling methods.

The ratio of the total current to the Ohmic current may be expressed as

$$\frac{I_T}{I_\Omega} = \left[1 + \frac{p_{ex}}{n_0}\left(\frac{\mu_n + \mu_p}{\mu_n}\right)\right] \approx 1 + \frac{p_{ex}}{n_0} \quad ; \qquad (1)$$

thus the ratio of injected carriers p_{ex} to background carriers n_0 can be found as a function of E and t. The inset in Fig. 1(b) demonstrates the relative amount of injection as a function of E and t calculated from the data using Eq. (1). For example, with an applied voltage of 12 V, which corresponds to $E = 48$ V/cm (upper curve in the inset), 2.6×10^{13} holes/cm^3 and an approximately equal number of electrons are injected by 900 nsec.

The hot-electron effect and injection are not apparent in conduction characteristics for $t \gtrsim 3$ nsec when $E \gtrsim 250$ V/cm, independent of sample length. The reason for this, impact ionization, is the subject of Sec. III.

III. IMPACT IONIZATION

Impact ionization in n-InSb causes copious carrier generation when E exceeds ~250 V/cm. Knowledge of the generation-rate dependence on E is crucial to the understanding of the BNDC phase of conduction, which is discussed in Sec. IV below. This section treats the correlation between the theoretically predicted and experimentally determined generation rates in order to ascertain the correct generation-rate formulation. An outgrowth of this study is an improved estimate of the electron distribution in energy space.

A. Results of Existing Theories

As E is increased beyond ~250 V/cm in an n-InSb sample at 77 K, a few electrons can gain enough energy to create an electron-hole pair by an impact-ionizing collision before being "cooled" in a collision that creates an optical phonon. As E is increased further, more and more electrons gain enough energy to participate in impact ionization. The questions of how the accelerated electrons are distributed in energy space and through what type of collisions they lose their energy are the subjects of many theoretical papers.

A number of early works concentrated specifically on impact ionization in p-n junctions in Si and Ge.[16–18] Later works considered covalent semiconductors with parabolic band structures.[19,20] Chuenkov[21] discussed ionic semiconductors with both parabolic and hyperbolic band structures; the band structure of InSb is hyperbolic. The principal differences among the theories is the assumed form of the energy distribution function of the electrons. The results of all these works can be expressed in two general generation-rate equations:

For "low" E,

$$g(E)=v_{\text{drift}}\ f_1(E)e^{-C_L/E}; \tag{2}$$

for "high" E,

$$g(E)=v_{\text{drift}}f_2(E)\,e^{-C_H/E^2} \tag{3}$$

The functions f_1 and f_2 depend so weakly on E that they are usually neglected. For the "low"-E case, first treated by Shockley,[17] the distribution function of the carriers is peaked in the direction of the applied field, resulting in a drift of carriers in energy space. The few carriers that have sufficient energy to impact ionize are contained in the high-energy tail of the distribution. In the "high"-E case, a nearly symmetric or Maxwellian distribution function is assumed (first by Wolff[16]), which results in essentially all the carriers gaining the same amount of energy in a given time period. Both Eqs. (2) and (3) assume an isotropic scattering probability for optical-phonon collisions, and no electron-electron collisions.

Baraff[19] combined the results of Eqs. (2) and (3) into a spherically symmetric distribution with a field-directed high-energy tail and obtained numerical solutions for $g(E)$ involving several adjustable parameters: ϵ_R, the optical-phonon energy; λ_R, the mean free path for optical-phonon collisions; and ϵ_i, the ionization energy. He also developed a criterion to determine which type of distribution function (peaked or symmetric) is applicable: For $eE\lambda_R \leq \epsilon_R$ the distribution function is peaked, and for $eE\lambda_R \gg \epsilon_R$ the distribution function is symmetric.

The constants C_L and C_H are defined in the following manner[17,18]:

$$C_L=\epsilon_i/e\lambda_R\ , \tag{4}$$

$$C_H=3\epsilon_i\,\epsilon_R/e^2\lambda_R^2\ , \tag{5}$$

so that values for λ_R and ϵ_i can be obtained by fitting the experimental results to the theory, if ϵ_R is known.

A somewhat different mathematical approach was taken by Dumke.[22] In addition to the assumptions of hyperbolic bands and a strongly peaked distribution function, he assumed essentially all small-angle (forward) scattering with polar optical phonons. His results yield a rather complicated equation for $g(E)$. One term involves diffusion in energy space and is weakly dependent on E for "low" fields. The more important term is exponentially dependent on E. For "low" E (what "low" means is discussed in Sec. III C below), the $g(E)$ dependence may be expressed as

$$g(E)\approx g_0 e^{E/\Theta}\ . \tag{6}$$

Unfortunately, Θ is not expressed as a function of the parameters contained in Eqs. (4) and (5). Thus, comparison of the constants in the exponents obtained from the two different approaches, isotropic and anisotropic scattering, is not directly possible.

The validity of the treatments expressed by Eqs. (2) and (3) has been shown experimentally in Si junctions[23,24] and bulk[25] Si for "low" E and in Si junctions[23] for "high" E. Two papers[4,5] contain experimental investigations of $g(E)$ in n-InSb and report conflicting results. Before comparing the present data with the results of these two papers, a discussion of the measurement of $g(E)$ is given in Sec. III B.

B. Measurement of Generation Rates

McGroddy and Nathan[4] suggested a novel technique for measuring $g(E)$, a method that is valid if all recombination processes may be neglected during the short duration of the measurement and if there is no generation by holes. The first assumption is shown to be valid in Sec. IV B 3 b below. The second is valid since electrons (injected), rather than holes, produce impact ionization even in p-

InSb.[26] By maintaining E constant for a time long enough to measure dI/dt, the generation rate as a function of E may be determined from

$$g(E)=\frac{1}{n_0}\frac{dn}{dt}=\frac{1}{I_0}\frac{dI}{dt}\bigg|_{t=0^+}, \tag{7}$$

where n_0 is the thermal equilibrium density of electrons, I_0 the Ohmic current, and I the total current. To satisfy the assumption of no recombination, dI/dt is measured in the earliest time period possible.

The earliest time at which measurements can be made is limited by the risetime, 10–90% ≈ 290 psec, of the applied voltage pulse.[3,13] To take advantage of this fast risetime and to reduce spurious signals, it is important to keep the sample holder as free from parasitic inductances and capacitances as possible. This is best accomplished by keeping the physical size of the sample holder small, using low-inductance and low-capacitance resistors in the circuitry, and making the connections to the sample in the form of coaxial cables insofar as possible. All of the samples studied were enclosed in a holder constructed of coaxial connectors containing the samples and two disclike 1-Ω (at 77 K) resistances. Each of these resistances consists of twelve 10-Ω 0.1-W carbon resistors soldered in a radial pattern.[13] To maintain a constant $E(t)$, one 1-Ω resistance is placed in parallel with the sample. The other of these resistances is in series on the ground side of the sample and is used to monitor the current through the sample. The uncertainty in the E measurements is estimated to be $\lesssim \pm 4\%$ and the uncertainty in the dI/dt values $\lesssim \pm 5\%$. Thus, $g(E)$ may be measured with an accuracy of $\sim \pm 9\%$.

Using the technique described by Eq. (7), $g(E)$ was measured in n-InSb at much higher fields than previously employed.[4,5] The next three figures present these data, plotted to exhibit the correlation, if any, with the theoretical behavior expressed by Eqs. (6), (2), and (3), respectively. The previous data[4,5] are included as appropriate, as well as Dumke's calculated $g(E)$ (in Fig. 2).

In Fig. 2 the functional dependence of all the experimental data is seen to agree well with Dumke's calculation, although the calculation refers to a zero temperature. The measured values of Θ in Eq. (6) range from 26–35 V/cm compared with the calculated[22] $\Theta = 25$ V/cm. The ranges in E of agreement between the data and the function $e^{E/\Theta}$ are

Present data:	$225 \leq E \leq 425$ V/cm;
McGroddy and Nathan[4]:	$225 \leq E \leq 330$ V/cm;
Ferry and Heinrich[5]:	$250 \leq E \leq 350$ V/cm;
Dumke's theory[22]:	$160 \leq E \leq 280$ V/cm;

Only the McGroddy-Nathan data extend to low enough E to indicate a departure from $e^{E/\Theta}$ behavior.

Figure 3 gives the range of agreement between the data and an $\exp(-C_L/E)$ dependence:

Present data:	$300 \leq E \leq 475$ V/cm;
McGroddy and Nathan:	$250 \leq E \leq 330$ V/cm;
Ferry and Heinrich:	$250 \leq E \leq 300$ V/cm.

From Eq. (4) and an estimate[26] of $\epsilon_i \approx 1.1\epsilon_g = 0.25$ eV, values of λ_R may be calculated. These values appear in Fig. 3.

Figure 4 shows that the present data at high fields best fit an e^{-C_H/E^2} dependence. The upper limit in E is given here by the maximum available applied voltage and not by the deviation from an e^{-C_H/E^2} dependence. Equation (5) yields the magnitude of λ_R, noted in Fig. 4, as obtained using the value of ϵ_i given above and a value[27] of $\epsilon_R = 0.0244$ eV. The so-obtained magnitude of λ_R is larger than its value deduced at lower fields, Fig. 3; this result is discussed in Sec. III C.

C. Discussion of Results

The data are presented in Figs. 2–4 in forms chosen to test the validity of the various theoretical assumptions, namely, (i) peaked distribution func-

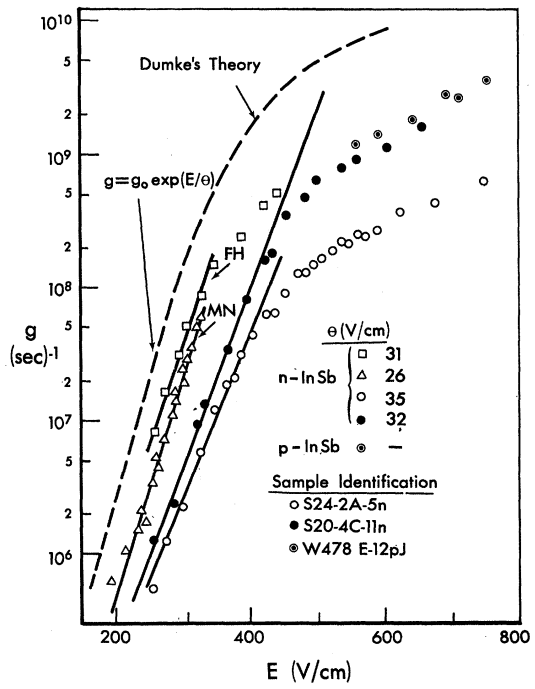

FIG. 2. Semilogarithmic plot of generation-rate data for comparison with Dumke's theory in the low-field approximation.

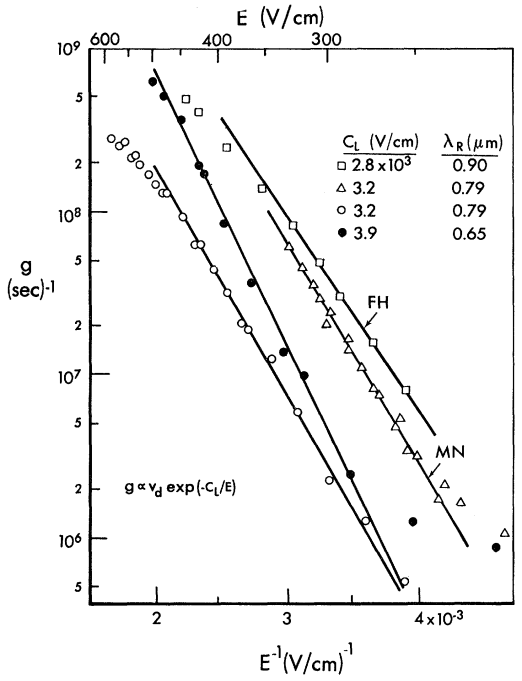

FIG. 3. Generation-rate data plotted for comparison with an $e^{-C_L/E}$ dependency.

tion, small-angle polar-optical scattering, and "low" fields, Eq. (6); (ii) peaked distribution function and isotropic polar-optical scattering, Eq. (2); (iii) symmetric distribution function and isotropic polar-optical scattering, Eq. (3).

McGroddy and Nathan[4] concluded that their data best fit an $e^{E/\Theta}$ dependence, while Ferry and Heinrich[5] concluded that their data best fit an $e^{-C_L/E}$ dependence for $E \lesssim 330$ V/cm and at higher fields tended toward an e^{-C_H/E^2} dependence. The present data, which cover the largest range in E, exhibit a best fit to all three dependences in overlapping field ranges:

$$\text{for } 225 \leq E \leq 425 \text{ V/cm}, \quad g \propto e^{E/\Theta};$$

$$\text{for } 300 \leq E \leq 475 \text{ V/cm}, \quad g \propto e^{-C_L/E};$$

$$\text{for } 425 \lesssim E \lesssim 700 \text{ V/cm}, \quad g \propto e^{-C_H/E^2};$$

where the upper limit is a circuit limitation and not the result of a fundamental bulk property.

Baraff's criterion[19] for the fields yielding a peaked distribution function is $E \leq \epsilon_R/e\lambda_R \approx 300$ V/cm for $\lambda_R = 0.8$ μm, Fig. 3. For $E \gg \epsilon_R/e\lambda_R$ or $\gg 300$ V/cm, the distribution function is symmetric. According to the present data the transition toward a Maxwellian (isotropic) distribution begins at ~425 V/cm, since in this range the data shift from exhibiting an $e^{1/E}$ dependence toward following an e^{1/E^2} dependence.

The best fit of the present data to the various $g(E)$ dependencies over the specified ranges in E implies that the distribution function is peaked for $E \lesssim 425$ V/cm and isotropic for $E \gtrsim 500$ V/cm. The significant decrease in the generation rate predicted by Dumke's[22] theory for high (>500 V/cm) E arises from the increased importance of the term involving diffusion in energy space. This results in the assumed distribution function becoming more isotropic, and so supports the conclusion that the distribution function is probably isotropic for $E \gtrsim 500$ V/cm.

The λ_R deduced from the high-E data, Fig. 4, are larger than the values deduced from the lower-E data, Fig. 3, as noted above. This result agrees qualitatively with a theoretical prediction[28] that says λ_R increases with energy ϵ:

$$\lambda_R \propto \frac{\epsilon^{1/2}}{\sinh^{-1}\epsilon^{1/2}} \propto \frac{\epsilon^{1/2}}{\ln(2\epsilon^{1/2})} \,. \tag{8}$$

The comparison between theory and observation at low fields is not sufficiently quantitative to determine if the assumption of small-angle polar-optical scattering made by Dumke[22] is closer to reality than the isotropic polar-optical scattering assumption made by the other authors. The present data, however, suggest that as the applied voltage is increased, first small-angle scattering is dominant in the impact ionization process, then

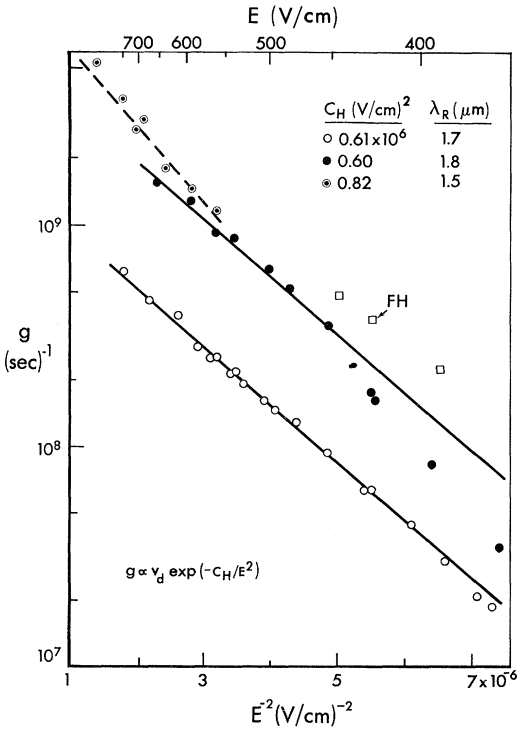

FIG. 4. Generation-rate data plotted for comparison with an e^{-C_H/E^2} dependency.

isotropic scattering is most important, and finally peaked distributions smooth into symmetric distributions at sufficiently high fields.

A transient instability appears in the conduction characteristics of n-InSb at high fields. This instability, BNDC, is discussed in Sec. IV.

IV. BNDC

Two forms of BNDC are exhibited by semiconductors. One, known as the type-N or voltage controlled BNDC, results in a decreased conductivity. This appears in I-E characteristics as a decrease in I and an increase in E; thus the curve resembles the letter "N." The other BNDC, known as the type-S or current controlled BNDC, produces an increase in conductivity. The associated I-E characteristic displays an increase in I and a decrease in E and so has the appearance of the letter "S".

A. Gunn Type-N BNDC

A special type of N-BNDC, the Gunn effect,[29] is observed in semiconductors having a particular band structure. The possible occurrence of this effect was proposed independently by Ridley and Watkins[30] and by Hilsum[31] for a class of semiconductors having two conduction bands characterized by different mobilities and spaced in energy such that the lower energy band has the higher mobility and the emission of optical phonons is the dominant mechanism for energy relaxation in both energy bands. They theorized that a sufficiently high E would "heat" the electrons in such a band structure enough to cause them to transfer to the higher-energy lower-mobility minima, thus causing a new mobility decrease or a negative differential mobility (NDM). Ridley[32] explored theoretically the tendency for a sample of such a class of semiconductors to divide into high-field domains and low-field regions, given a condition of NDM. Ridley and Watkins[30] additionally pointed out that the direct band-gap separation should be sufficiently large to prevent the generation of high mobility carriers by impact ionization, a condition which overrides the effect of the intervalley scattering.

In 1964, Gunn[29] reported some current instabilities (and voltage instabilities) in samples of GaAs and InP that are now known to originate from the Ridley-Watkins-Hilsum mechanism. The high-field domains (HFD), found to form in samples exhibiting NDM, can propagate in space from cathode to anode[29] or be essentially stationary adjacent to the anode.[33] The field may also either grow, remain fixed, or decrease in amplitude.[34,35] The front half of a HFD is a depletion region while the back half is an accumulation region.[36] Also, the ratio of the peak field E_{peak} within the HFD to the threshold field E_{th} necessary to form the HFD can be quite high (e.g., in n-GaAs, $E_{peak}/E_{th} \approx 35$).[35]

A transient example of the Gunn effect has been found[6] in n-InSb at atmospheric pressure and 77 K. An earlier attempt[37] to observe the Gunn effect in this semiconductor had failed; the failure was attributed to direct band-gap ionization dominating over intervalley scattering because the higher-energy minima were believed to be too high in energy to be populated before impact ionization dominated the transport process. Later measurements[7] support this explanation by showing that the separation between the central ($k = 0$) minimum and the next higher $\langle 111 \rangle$ minimum is ~ 0.45 eV, which is approximately twice the direct band-gap separation,[38] $E_g = 0.23$ eV. Smith *et al.*,[6] nevertheless, concluded that the Gunn effect does occur in n-InSb and attributed the earlier failure to observe the N-BNDC to the experimental conditions, namely relatively small voltages applied for times long enough to create large carrier densities by impact ionization and so obscure the effects of any intervalley scattering. They further suggested that the ability to produce the Gunn effect in n-InSb implies that the cross section for impact ionization is small enough to allow carrier heating into the higher minima. An earlier[4] estimate of the ratio of the cross section for ionization to the total collision cross section is 0.05, a result which Smith *et al.*[6] noted supports their suggestion. The possibility that the observed NDM is caused by electron transfer to higher-energy states within the nonparabolic central band by means of intravalley scattering was ruled out by invoking the results of pressure experiments[7] and recent theoretical calculations.[39]

The presence of an N-BNDC in n-InSb is demonstrated in Fig. 5 for $0.6 \text{ nsec} < t < 3.5$ nsec. The two curves for $t = 1.0$ and $t = 2.0$ nsec in Fig. 5(a) show the characteristic "N" shape of the Gunn effect. Figure 5(b) contains $I(t)$ and $E(t)$ for two applied voltages; V_1 is just below the threshold for N-BNDC and $V_2 = 1.02V_1$ is just above this threshold. For times up to ~ 3 nsec V_2 produces an E greater than V_1 does and simultaneously an I less than V_1 does, another demonstration of N-BNDC.

B. Type-S BNDC

In 1966, Ancker-Johnson[3] reported the existence of an S-BNDC in single-crystal n-InSb at 77 K. Figure 5 shows S-BNDC for $t > 3.5$ nsec and demonstrates the temporal link between the N-BNDC and S-BNDC observed in n-InSb.[8] The conduction curves in Fig. 5(a) for $t = 6$ and 10 nsec have the characteristic "S" shape associated with the increased conductivity of the S-BNDC. Figure 5(b) shows the transition time at which the enhanced conduction begins (crossover points in the E curves and in the I curves), namely 3.5 nsec. Once the S-BNDC has formed, the average conduction continues to increase with time until a nonequilibrium

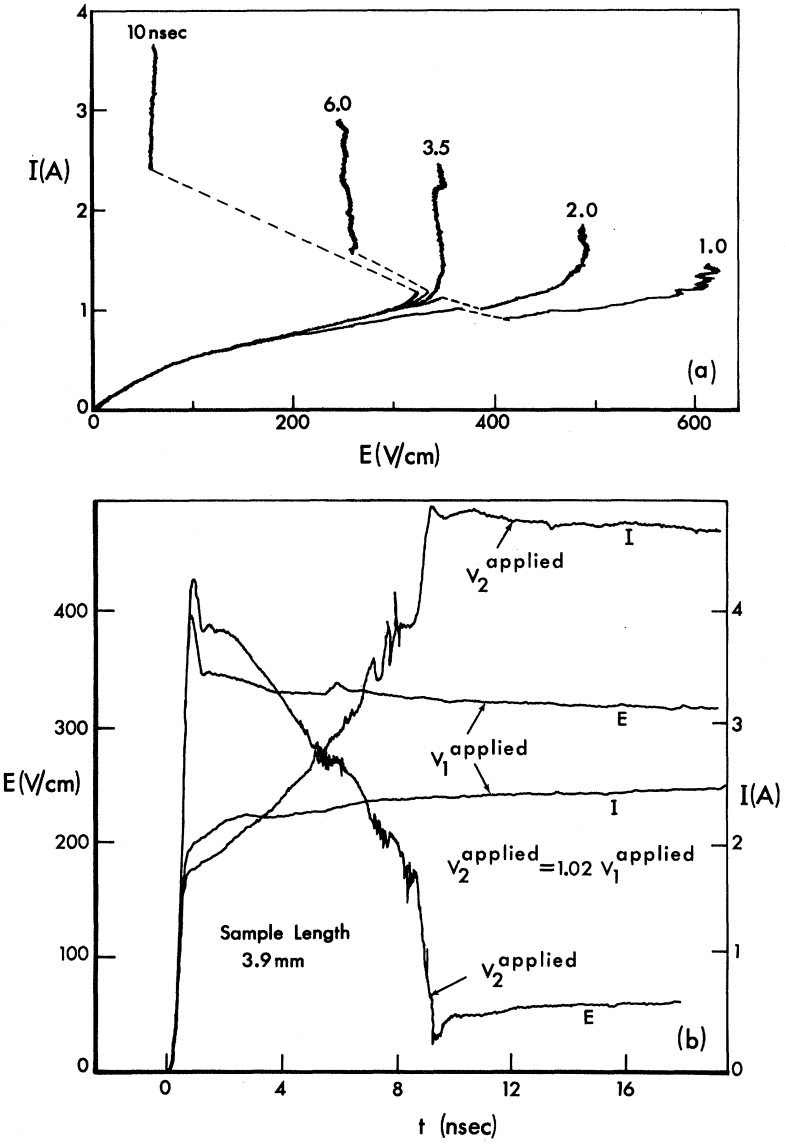

FIG. 5. (a) Measured conduction characteristics showing the temporal relationship between N-BNDC and S-BNDC. Data refer to sample No. S20-4G-22n. (b) Current and field strength (averaged over the sample length) as functions of time. The voltage V_1 is just below the threshold for N-BNDC and V_2 is just above this threshold. The large $-dE/dt$ and $+dI/dt$ at $t=9$ nsec are associated with the arrival of a HFD at the anode of the sample. Data refer to sample No. S20-4C-22n.

steady state is reached at $t \approx 9$ nsec, Fig. 5(b). The average conduction is seen in this figure to increase with time between 0.7 and 9 nsec for an applied voltage exceeding threshold since $E(t)$ decreases and $I(t)$ increases.

1. *Model for Origin of Type-S* BNDC

The S-BNDC present in n-InSb is created in the wake of a propagating HFD and the increased conductivity associated with the S-BNDC is caused by strong impact ionization occurring in the HFD, as briefly reported previously.[8]

Figure 6 qualitatively represents the phenomenon. Although parts of Fig. 6 are speculative, the curves are closely related to the experimental results recorded in Fig. 5 (and Fig. 7).

In Figs. 6(a) and 6(b), E and n, the electron concentration, are shown as functions of x, length down the sample current axis, for an applied voltage just above E_{th}. The time t_1 corresponds to the formation of the HFD; t_2 and t_3 are subsequent times. The conduction characteristics in Figs. 6(c)–(e) show E averaged over three different regions of the sample for the three different times used in all of Fig. 6. Figure 6(f) contains typical I-V curves, where V is the nonconstant (with time) applied voltage, and is similar to Fig. 5(a).

The formation of a nonequilibrium HFD near the cathode of a sample is depicted in Fig. 6 at $t=t_1$. At its formation the field averaged over the sample length increases and the current decreases as shown for $t=0.7$ nsec in Fig. 5(b). The peak field within a HFD, Fig. 6(a), is quite high ($E_{peak} \gtrsim 4900$ V/cm; this value is discussed later in conjunction

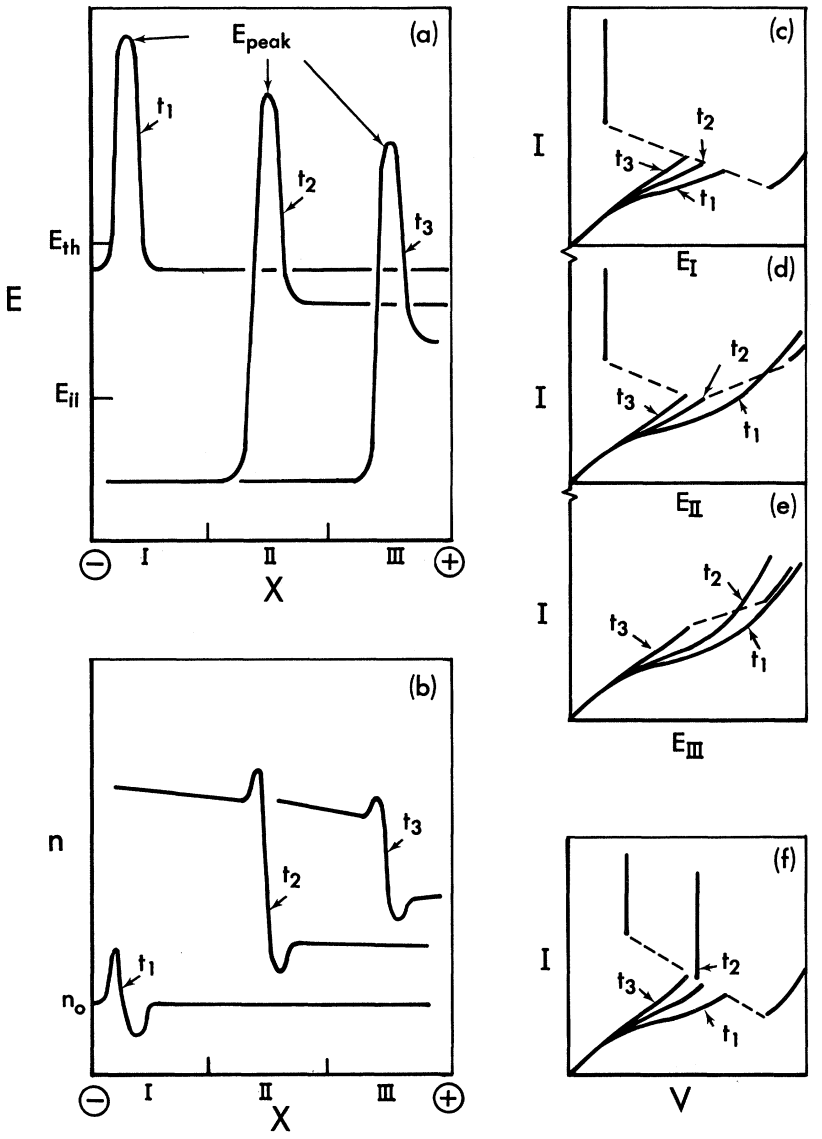

FIG. 6. Model for the occurrence of S-BNDC. (a) Electric field strength as a function of distance between cathode and anode during domain formation and transit. (b) Electron density as a function of this same distance. (c) Conduction characteristics with electric field averaged over region I (cathode third of the sample). (d) Conduction characteristics with electric field averaged over region II (middle of the sample). (e) Conduction characteristics with electric field averaged over region III (anode third of the sample). (f) Current-voltage (across the entire sample) characteristics as functions of time.

with Fig. 7), thus a strong avalanche commences in the HFD by the time t_1 elapses, since the threshold field for impact ionization, E_{ii}, is only ~250 V/cm. The current decreases at the formation of a HFD because the partially depleted region of the HFD, Fig. 6(b), is the point of lowest density in the sample, which together with the characteristic carrier velocity, determines the current. (Diffusion current is negligible on these nanosecond time scales.)

According to this model, the density of electrons and holes is increasing in front of the propagating HFD because the field there also exceeds E_{ii} (but by much less than E_{peak}); Figs. 6(a) and (b). Therefore, as the HFD propagates toward the anode, the current increases because of the increasing carrier density ahead of the HFD; Figs. 5(b) and 6(b). Simultaneously, E behind the HFD is decreasing, Fig. 6(a), as excess carriers generated by E_{peak} increase at a rate much faster than the current can increase. Hence, in contrast to the current, the averaged field decreases as the HFD propagates toward the anode. This E decrease is seen in Fig. 5(b) for 0.7 nsec $\leq t \leq$ 9 nsec, and in Fig. 6(a) for $t_1 \leq t \leq t_3$.

The effect of the decreasing field behind the HFD considerably influences the conduction characteristics averaged over different regions of the sample; Figs. 6(c)-(e). In region I, an N-BNDC characteristic occurs at t_1 and an S-BNDC characteristic at t_2 and t_3; Fig. 6(c). In region II, no BNDC occurs at t_1, but at t_2 an N-BNDC is seen and at t_3 an S-BNDC occurs; Fig. 6(d). At t_1, regions II and III are unaffected by the HFD located in region I, thus the conduction characteristics are simply monotonic;

at t_2, the HFD has not yet arrived at region III, so there is no indication of any BNDC in region III for t_1 and t_2; Figs. 6(d) and (e).

Upon arrival of the HFD at the anode, the high-field and associated depletion region disappear, causing the field to drop sharply and the current to increase simultaneously, as Fig. 5(b) for $t = 9$ nsec illustrates.

The study of a single type of BNDC, i.e., N-BNDC or S-BNDC, is aided by choosing a circuit condition that tends to maximize the change in current or voltage that is associated with the particular BNDC. However, in order to observe the transition from an N-BNDC to an S-BNDC a circuit condition that is neither constant voltage nor constant current is most useful. Such a circuit was used to collect all of the data pertaining to both the N-BNDC and S-BNDC in this work. Our model also assumes an applied voltage that decreases with time, as seen in Fig. 6(a): The area under the $E(x)$, i.e., the applied voltage, decreases with advancing time.

The roots of this model lie in some particular effects associated with BNDC in GaAs. In 1966 Heeks[35] reported observing a steady-state infrared radiation of band-gap frequency from high-resistivity n-type bulk GaAs samples exhibiting the Gunn effect. This radiation he attributed to impact ionization within the HFD associated with the Gunn effect. Also in 1966, Liu[40] reported observing a steady-state S-BNDC and infrared radiation from low-resistivity n-GaAs for fields just above the threshold for the Gunn effect. Through experimentation with various sample lengths and resistivities, Liu concluded that the S-BNDC observed at the Gunn threshold was associated with the length and the carrier concentration of the sample, but the actual nature of the association was not proposed. In 1967, Southgate[41] reported the results of detailed studies on excess-carrier generation and recombination processes in n-GaAs. He found that the amount of ionization generated by a HFD depended on the resistivity of the sample or, in other words, on the thermal equilibrium density of electrons, n_0. Because of this either a steady-state N-BNDC or a steady-state S-BNDC, but not both in one sample, could be generated simply by using samples of different n_0. A high-resistivity sample has a low ionization rate and, therefore, few excess carriers are produced by impact ionization within the HFD. Thus, the conduction characteristics are little modified and the sample can still exhibit N-BNDC. A low-resistivity sample has a high ionization rate and hence copious excess-carrier production by impact ionization, so it exhibits S-BNDC because there are sufficient excess carriers present to reduce the average conductivity.

The reason that both N-BNDC and S-BNDC can be seen in the same sample of n-InSb, although neither is a steady-state condition, is due in part to impact ionization outside the HFD, a process which is not present in n-GaAs. In n-InSb, the density in front of the HFD is continuously increasing with time [Fig. 6(b)] because of ionization there, while the field behind the HFD drops below E_{ll} because of the large number of excess carriers created by E_{peak} [Fig. 6(a)]. For early times, $t \lesssim 3$ nsec, the total excess-carrier density is still low enough that a net N-BNDC can be seen; Fig. 5(a). However, for $t > 3.5$ nsec, the excess-carrier density has increased so much that only an S-BNDC can be seen; Fig. 5(a) [compare $t \gtrsim t_2$ in Fig. 6(f)].

In n-GaAs, the steady-state condition of BNDC is characterized by successive nucleation and propagation of HFD, independent of the I-E characteristic shape, N or S. This is because impact ionization within the HFD is the only mechanism for excess-carrier generation in such experiments. Hence, a steady-state condition between generation and recombination can be reached that still allows formation of NDM.

Smith *et al.*[6] suggested that impact ionization throughout a sample would prevent formation of the Gunn effect in n-InSb, given a condition of NDM, however, they did not conclude that domain propagation can occur only once per applied power pulse. This "single-transit" property of the HFD is an inherent feature of the here-proposed model for the origin of the S-BNDC. The impact ionization in the HFD occurs at a much higher rate than it does outside the HFD because the field within the HFD is so much higher than the field outside it. Thus, even if the domain propagation time is reduced by making a very short sample (like the 600-μm sample Smith *et al.*[6] used), so that the effect of ionization ahead of the HFD is minimized, another domain would never be able to propagate in the wake of a previous HFD because of the ionization generated by the previous HFD.

Neither the N-BNDC nor the S-BNDC is a steady-state condition in n-InSb because of the single transit of the HFD. Since only one HFD may propagate over a given region of the sample, the Gunn effect lasts only as long as a HFD exists. Because the S-BNDC is produced by the ionization within the HFD, when these nonequilibrium carriers decay and no new HFD is formed, the S-BNDC decays. Then the usual I-E characteristic associated with impact ionization is observed: a curve that exhibits positive slope at all field strengths. The decay of the S-BNDC is discussed further in Sec. IV B 3.

2. *Experimental Evidence in Support of Proposed Model for Origin of* S-BDNC

In this section the results of four independent ex-

periments are presented to demonstrate the validity of the model just proposed.

a. *Conduction measurements*. Figure 5(a) shows a series of I-E characteristics generated in a manner previously described.[3] Two features of this method were significantly improved during this work and so are described here.

The necessity of measuring conduction characteristics on short, namely, subnanosecond time scales requires rather sophisticated triggering techniques. A coaxial pulse generator, fitted with a mercury-wetted relay, cannot be triggered externally because of the uncertainty in the closing of the relay. Instead, a trigger signal must be taken from the output of the pulser. The amplitude of this signal varies with the amplitude variation of the output pulse. If this signal is used to trigger a Tektronix sampling oscilloscope with a 3T77 time base, the time of triggering will change as the amplitude of the trigger signal changes, due to inherent characteristics of the oscilloscope. Since the dynamic range of amplitude needed from a pulse generator to generate data like that of Fig. 5(a) is typically 20:1, and such a range causes a shift in the time of triggering of ~1.6 nsec, a great deal of care must be taken to generate a fixed-amplitude constant-risetime trigger pulse. A description of a device designed to meet this requirement is found elsewhere.[13]

To eliminate undesirable ringing and to utilize

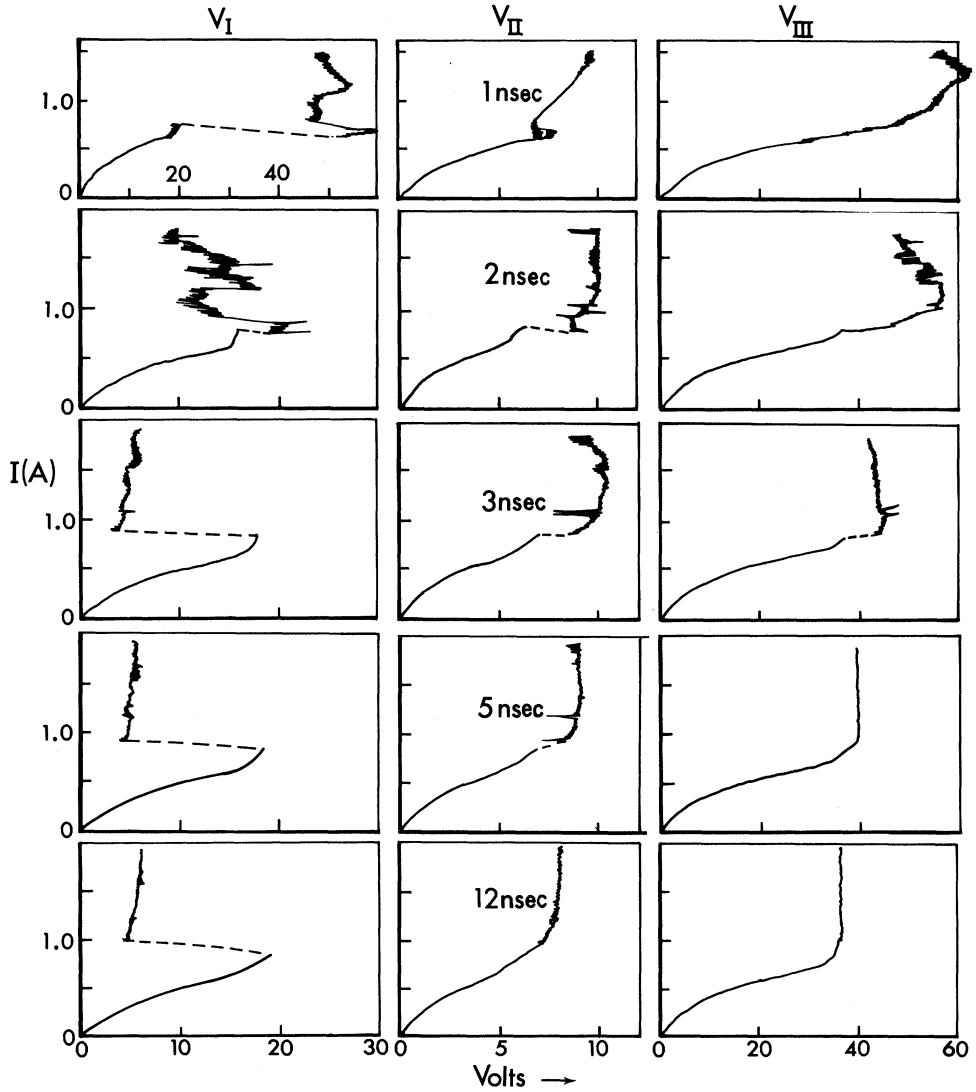

FIG. 7. Current as a function of voltage for three regions of a sample (shown in Fig. 8) parametric in time for $1 \leq t \leq 12$ nsec. The voltage scale for the upper left-hand curve ($t = 1$ nsec and region I) is on a compressed scale; all the other voltage scales are as marked on the bottom row of curves. Data refer to sample No. S20-4C-13n.

TABLE I. Properties, at 77 K, of the n-InSb single-crystal samples used to investigate BNDC.

Boule	Sample No.	Size (mm^3)	n_0 (10^5 cm^{-3})	μ_n (10^5 cm^2/Vsec)
S24-2A	6n	0.25×1.1×3.9	0.5	3
	9n	0.3×1.2×4.0	0.5	3
S35-1F	11n	0.4×0.4×2.5	1.7	7
	14n	0.7×0.7×6.0	1.7	7
	ga	0.3×0.7×4.2	1.7	7
S20-4C	12n	0.2×0.4×2.5	2.7	2
	13n	0.2×0.4×2.5	2.7	2
	22n	0.3×0.55×3.9	2.7	2

the subnanosecond capabilities of the sampling system, a 50-Ω coaxial stripline sample holder was developed for the BNDC measurements. Details are found elsewhere.[13] The overshoot in E seen in Fig. 5(b) for $t \lesssim 0.5$ nsec is an unavoidable consequence of the sample inductance plus the connecting leadwire inductance. Though small (on the order of 10 nH), these inductances combined with the typical sample resistance of 10 Ω have an influence on subnanosecond time scales.

The eight samples used to investigate BNDC were studied in a circuit that was intentionally neither constant voltage nor constant current, so that the transition from N-BNDC to S-BNDC could best be studied as discussed above. The physical properties of these samples are listed in Table I.

Figures 7–10 (see also Fig. 2 of Ref. 8) show effects on the conduction properties of the transition from type-N to type-S BNDC resulting from differences in HFD formation and propagation.

An example of domain propagation through most of a sample has been shown previously (Fig. 2 of Ref. 8). The collapse in transit of a HFD is demonstrated in Figs. 7 and 8, and has been seen[35] also in n-GaAs. Such an event is expected from the model, which has the HFD controlling the current because it controls the minimum carrier concentration that occurs in the sample. Thus, as stated above, as a domain propagates down the sample and encounters an ever increasing carrier concentration caused by ionization in its path, Fig. 6(b), the current increases. As the current increases, the additional voltage necessary to satisfy the increased potential drop ahead of the HFD is taken from the domain; Fig. 6(a). If the current increases sufficiently to require a major portion of the bias, the HFD cannot be sustained.

A large N-BNDC is seen, Fig. 7, in region I (0.8 mm of a 2.5-mm-long sample) at $t = 1$ nsec. By 3 nsec, an S-BNDC has been formed in region I. It slowly decays and disappears by 200 nsec as shown in Fig. 8, curve V_I. However, E_{peak} of this HFD is sufficiently suppressed as it propagates so that the domain does not last long enough to cause significant impact ionization in region II, much less in region III; thus an S-BNDC never appears in these two regions, as Figs. 7 and 8 show.

The magnitudes of the voltage across region I just before and just after the threshold for HFD formation, Fig. 8, allow E_{peak} to be estimated at an early time, 1.0 nsec. The estimate is made by finding the field intensities ahead E_A and behind E_B, the HFD, and by determining the lengths of these parts of the sample. The magnitude of the field where the HFD has not yet propagated is obtained from the conduction characteristics just below threshold. In Fig. 7 at $t = 1.0$ nsec in region I the voltage across the 0.8 mm is 20 V or $E_A = 250$ V/cm. The quantity E_B is found by noting that the voltage across region I is 5 V after the HFD has traversed it ($t \geq 3$ nsec), so $E_B = 63$ V/cm. The distance that the HFD has propagated by 1.0 nsec is determined as follows: The data of Fig. 5(b) indicate that a HFD is formed by 0.7 nsec (this is a typical value for the pulse risetime of 290 psec used

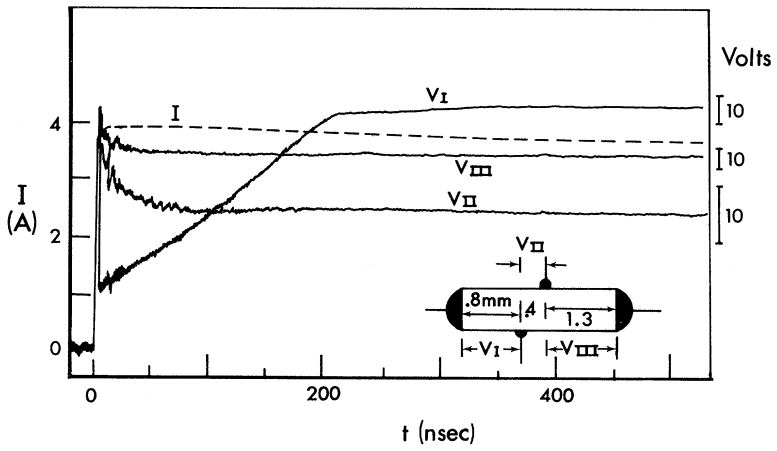

FIG. 8. Current and voltage as functions of time for the sample of Fig. 7. The decrease in V_I, associated with the S-BNDC in region I, is not seen in V_{II} and V_{III}, thus confirming the results shown in Fig. 7. The voltage scales for V_I, V_{II}, and V_{III} are given on the right-hand ordinate; the zero is the same as on the left-hand ordinate. Data refer to sample No. S20-4C-13n.

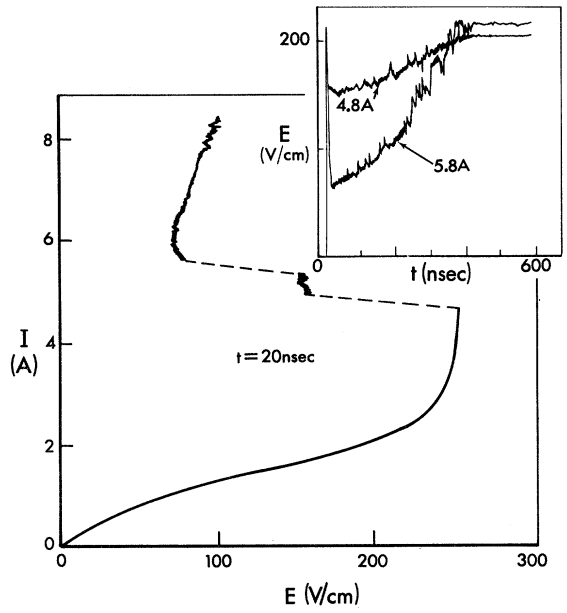

FIG. 9. Conduction characteristics illustrating a double S-BNDC at $t=20$ nsec. The inset shows $E(t)$ for two current levels, one associated with the S-BNDC at $I=4.8$ and the other associated with the S-BNDC at $I=5.8$ A. Data refer to sample No. S35-1F-11n.

in this work); therefore by 1.0 nsec the HFD has propagated 0.1 mm. The HFD width is $\lesssim 90\ \mu$m[6] (it is typically 50 μm in n-GaAs[36]). Since the field intensities outside the HFD are now known as well as the lengths over which they prevail, the voltage distribution just after threshold can be determined: $V_A = 25$ V/mm $\times [0.8 - (0.1 + 0.09)]$ mm $= 15$ V, and $V_B = 6.3$ V/mm $\times 0.1$ mm $= 0.6$ V. The voltage across the entire sample at 1.0 nsec is 60 V at the HFD threshold (Fig. 7, region I at 1 nsec), so $E_{peak} = (60 - 15.6)\ \text{V}/9\times10^{-3}$ cm $= 4900$ V/cm or more if the HFD width is smaller than 90 μm.

The data of Fig. 7 (and Fig. 2 of Ref. 8) preclude the possibility of the S-BNDC being caused by a current filament formation, a possible result of NDM suggested by Ridley.[32] This possibility is eliminated by the demonstrated propagating nature of the S-BNDC. Any current filament formation, according to Ridley's arguments, would have to be as instantaneous (<1 nsec) throughout the length of the sample, as the HFD formation is in a cross section of the length. Furthermore, Ridley's model could not possibly fit a case (Fig. 7) for the current being pinched into a filament over only part of the sample length.

The existence of multiple S-BNDC was briefly reported[42] in 1968, and an example of a double S-BNDC is shown in Fig. 9. Multiple S-BNDC are probably the result of the place of nucleation of a particular HFD being at any place that a large field strength exists, wherever it may occur in a sample (e.g., defects in a sample would cause initial nonuniformity in E). Thus, in a given sample, the conduction condition at which a particular HFD is nucleated depends on the amplitude of the applied voltage and its distribution (nonuniform) in the sample. For example, in the inset of Fig. 9, the magnitude of the decrease in E at 4.8 A for an applied V_1 is less than the magnitude of the decrease in E at 5.8 A for an applied V_2. This implies that fewer carriers are generated by the HFD associated with V_1 (HFD 1) than by the HFD associated with V_2 (HFD 2). The generation rate is probably approximately equal in the two domains because the E_{peak} probably does not vary much from domain to domain as is true in[35] n-GaAs and because the generation rate is relatively weakly dependent on E at such high fields; Fig. 4. Then the conclusion is that HFD 2 propagated over a greater distance in the sample than HFD 1 propagated, and so HFD 2 must have been nucleated closer to the cathode than HFD 1.

Smith *et al.*[6] suggested that the risetime of the leading edge of the applied voltage pulse should af-

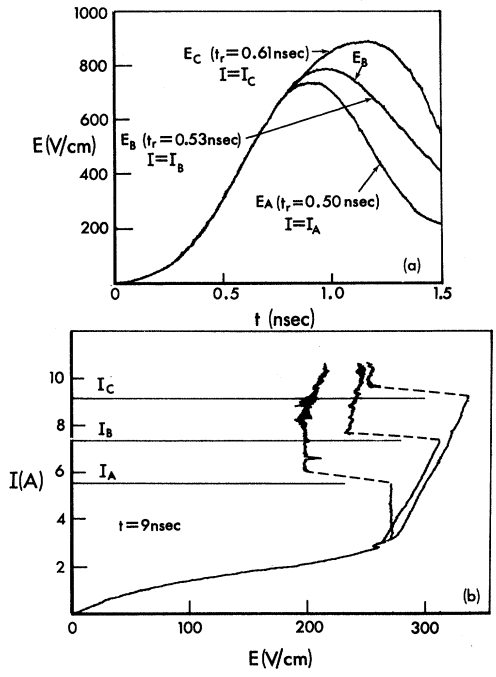

FIG. 10. (a) Electric field strength (averaged over 2.5 mm) as a function of time, for three different rise times and applied voltages. The magnitudes of the applied voltages are just below those producing a HFD; therefore, the maximum field for each rise time is very nearly the threshold field for that case. (b) Conduction characteristics at 9 nsec, demonstrating the fact that the threshold field strength for formation of an S-BNDC is highly dependent on the risetime of the applied voltage. Data refer to sample No. S35-1F-11n.

fect the formation of the N-BNDC; indeed as discussed above, too slow a risetime prevents its formation. According to the proposed model of S-BNDC formation, the risetime should similarly affect the S-BNDC occurrence. Figure 10 shows the results of a study designed to test this hypothesis.

Normally, the risetime of the applied voltage pulse produced by the type of pulser used in these experiments[3] is independent of the pulse amplitude, since the characteristics of the coaxial line which stores the energy used in forming the pulse are independent of the amount of energy stored (below dielectric breakdown). In order to control the risetime of the applied pulse and so generate the data in Fig. 10, a variable shunt capacitor (air dielectric, 3–7 pF) was placed in the pulsing circuit close to the sample. Increasing this shunt capacitance increases the risetime of the applied pulse; furthermore for a given capacitance, the resulting risetime is again independent of the pulse amplitude.

Figure 10(a) shows the average field strengths as functions of time for three different risetimes and applied voltages. The chosen applied voltages in each case are just below those that produce HFD formation (and subsequent S-BNDC formation). The resulting field maxima are, therefore, essentially equal to the E_{th} in each case. The shortest risetime ($t_r = 0.50$ nsec) and lowest applied voltage develop a field intensity of 720 V/cm at threshold. As Fig. 10(a) also shows, larger fields and longer times are required for threshold with lengthening risetimes. The influence of these risetime changes on the S-BNDC at 9 nsec is shown in Fig. 10(b). The S-BNDC occurs at increasingly higher currents and field intensities as the risetime is lengthened. This threshold dependence is a result of impact ionization occurring throughout the sample before HFD formation. As noted above, impact ionization is in competition with intervalley scattering, thus the longer risetimes necessitate higher fields to create NDM, a HFD, and subsequently S-BNDC. Somewhat longer risetimes than used to generate Fig. 10 result in so much carrier generation that intervalley transfer becomes negligible.

It is difficult to determine accurately the threshold field strength E_{th} necessary to create BNDC effects, because the risetime influence just discussed tends to make the observed E_{th} larger than it really is. Furthermore, slight nonuniformities in initial conductivity caused, for example, by random crystal defects, result in the measurement of too small an E_{th}. This erroneous measurement occurs because random defects are the sites of higher fields than the average and so sites for NDM initiation. Since these initiation regions are generally much smaller than the regions in which the E is measured, too small an E_{th} is observed.

The effects of these two sources of error tend to cancel so the average of many E_{th} observations is probably more accurate than the range in measurements indicates. The value obtained from the present work is $E_{th} \approx 500$ V/cm. The range in observed E_{th} is 250–770 V/cm for $t_r = 0.3$ nsec. Smith *et al.*[6] estimated $E_{th} \approx 600$ V/cm and Fawcett and Ruch[39] calculated $E_{th} \approx 525$ V/cm.

The ratio E_{peak}/E_{th} is thus $4900/500 \approx 10$. In n-GaAs this ratio has been found[35] to be ≥ 35. The low value in InSb is another indication of the competition between ionizing collisions and intervalley-scattering collisions.

b. *Radiative recombination measurements.* It was proposed in Sec. IV B 1 that the increase in conductivity associated with the S-BNDC results from copious excess-carrier generation by a HFD. The purpose of this section is to present additional data which demonstrate that excess-carrier generation is responsible for the S-type instability.

The presence of excess carriers in n-InSb at 77 K can be determined by detecting photons of bandgap wavelength ~5 μm that are produced by the direct recombination of electrons in the conduction band with holes in the valence band. Several authors have reported making such measurements.[43–45] The radiative recombination flux Φ is related to the carrier density in the following manner:

$$\Phi = \frac{dn}{dt} = Rnp \ , \tag{9}$$

where R is the radiative recombination rate, and n and p represent the electron- and hole-carrier densities, respectively.

The rapid changes in conduction associated with BNDC require a radiation detection system with a fast response time, thus a solid-state detection system was developed to measure 5-μm radiation that has the following specifications[46]: a 35-nsec risetime (27–73%), a peak signal-to-noise ratio of ~40:1, an average noise-equivalent power of $\sim 10^{-15}$ W, a spatial resetability of 10 μm, a spatial resolution of ~100 μm, and a temporal resolution of 10 nsec. The temporal resolution is limited by the gate width of the boxcar integrator. The sensitive area of the detector, 3 mm in diameter, was scaled down through a lens to a sampling area ~0.5 mm in diameter.

The spatial distribution of Φ over the surface of a sample of n-InSb at 77 K for a constant current (16 A) and time (170 nsec) is shown in Fig. 11. At this current the sample is stably biased just above the threshold for S-BNDC and at this sampling time Φ is maximized in this particular case. Both the cross scan $\Phi(Y)$, and length scan $\Phi(Z)$, show Φ to be relatively constant over the surface area. These data and the similar results obtained at higher currents (not shown) provide the absolute spatial reference for the measurements given below.

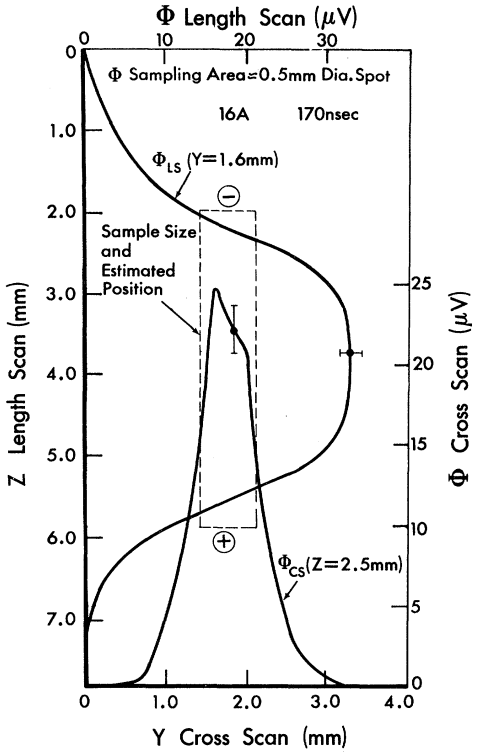

FIG. 11. Spatial distribution of the radiative recombination flux Φ for a sample biased just above the threshold for S-BNDC. Data refer to sample No. S35-1F-14.

If a HFD domain is responsible for the production of copious excess carriers, then the radiative recombination of these carriers is expected to produce a significant increase in Φ coincident with E reaching the threshold for the S-BNDC. Figure 12 shows this correlation.

The I-E characteristics in Fig. 12 demonstrate the presence of an S-BNDC for a threshold current of 13 A. The "ragged" appearance (compared with, e.g., the I-E curve in Fig. 9) of the 80-nsec I-E curve for 15 A$\lesssim I \lesssim$23 A is caused by the point of domain nucleation from pulse-to-pulse varying as the applied voltage is increased. Thus the amount of carrier generation, and hence the magnitude of the decrease in E, varies as I increases. This incidental feature is of no import in the radiation measurements. The conduction characteristics show no significant change for $t>270$ nsec, so signaling the establishment of the nonequilibrium steady-state impact ionization condition. The insets (a)–(c) in Fig. 12 show Φ as a function of current for sampling times correlated with the sampling times of the I-E characteristics. The precipitous increase in Φ at 13 A, the threshold for the S-BNDC, shows the marked increase in carriers associated with the formation of the S-BNDC.

Also apparent in Figs. 12(d) is a discontinuous increase in Φ at $t=520$ nsec, which is not accompanied by a corresponding decrease in E (the steady-state is established before this time as stated above). This increase in Φ is interpreted to mean that there are still excess carriers present above that number required to satisfy the circuit current requirements, yet these carriers do not affect E. These excess carriers do not affect E, to $t\gtrsim 300$ nsec, because when they were generated E decreased, resulting in an increase in conductivity just from the carrier cooling. As the excess carriers begin to decay, the conductivity begins to decrease, causing E to increase, which further reduces the conductivity by heating the carriers and contributes to the disappearance of the S-BNDC apparent in the I-E characteristics. The carrier heating can be a large effect. Glicksman and Hicenbothem[1] have shown that the electron mobility (μ_n) at 30 V/cm is three times larger than μ_n at 300 V/cm. Hence, the decay of the S-BNDC does not reflect the decay of the excess carriers which produced it. Excess-carrier decay is discussed further in Sec. IV B 3.

By $t=800$ nsec, Fig. 12(e), there is no detectable radiative recombination present that can be attributed to the carriers which produced the S-BNDC, although there is significant light flux caused by the radiative decay of excess carriers generated by ionization throughout the sample, as expected.[43–45]

The technique demonstrated in Fig. 10, used to cause E_{th} to increase by lengthening the risetime of

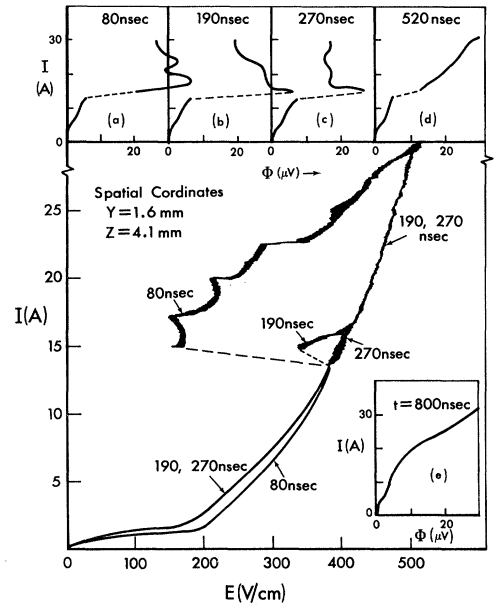

FIG. 12. Conduction characteristics parametric in time, illustrating the presence and decay of S-BNDC. The insets (a)–(e) show the associated radiative recombination flux Φ for the same times as used to sample the conduction characteristics. Data refer to sample No. S35-1F-ga.

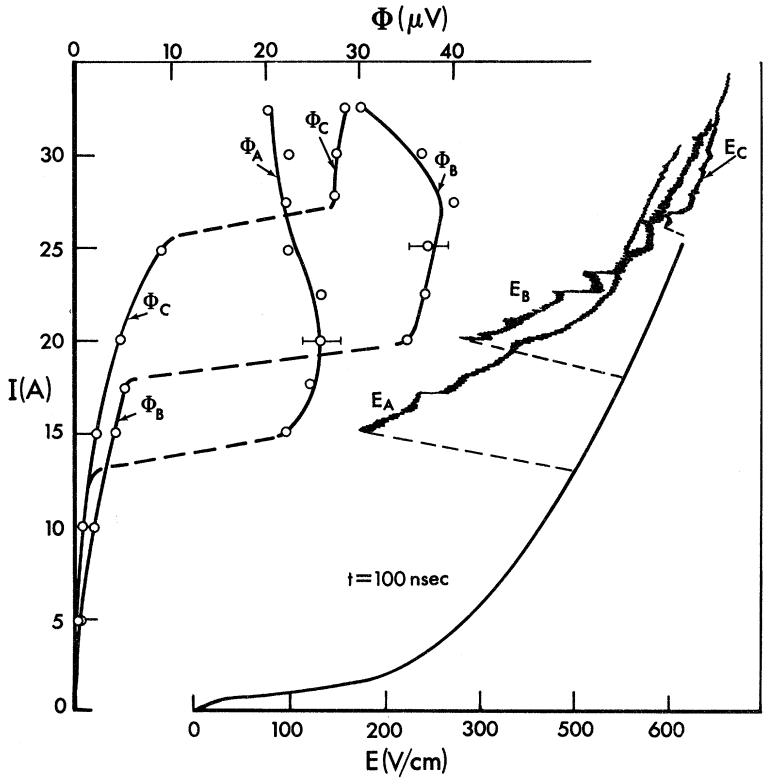

FIG. 13. Field intensity and radiative recombination flux as functions of current for $t = 100$ nsec. Conditions "*A*," "*B*," and "*C*" are associated with the risetime ($t_{rA} < t_{rB} < t_{rC}$) of the applied voltage, similar to the data of Fig. 10. Data refer to sample No. S35-1F-ga.

the applied pulse, enables another demonstration of the correlation between *S*-BNDC formation and simultaneous increase in Φ. In Fig. 13, Φ and E are shown as functions of current for a constant observation time, 100 nsec, and three different pulse risetimes. It is clear that a large flux of excess carriers attends the *S*-BNDC threshold.

The finite-domain propagation velocity, (3–4) $\times 10^7$ cm/sec, results in a time dependence of the generation of excess carriers by a HFD. This means that radiation generated at the cathode end of a sample is generated earlier in time, relative to the application of the bias, than radiation generated at the anode end. Hence, there is a delay in the arrival time at the detector of the half-amplitude point of Φ, measured as a function of increasing Z, as previously reported.[8]

c. *Modification of low-field conduction by the S-BNDC.* Another demonstration that carrier generation is responsible for the *S*-BNDC arises from a special low-field conduction-measurement technique. Southgate[41] used this type of measurement to determine excess-carrier lifetime in *n*-GaAs and various modifications of this technique have been used previously.[11,47]

Since the excess carriers generated by the HFD have a finite lifetime (see Fig. 12), it is possible to detect their presence in low-field conduction characteristics. First, a voltage (V_1) high enough to establish a condition of NDM must be applied to a sample for a long enough time (~ 10 nsec) to allow propagation of a HFD through the sample, thus producing a copious quantity of excess carriers throughout the sample. Then, if the applied voltage is quickly (≲ 1 nsec) reduced to V_2, a value well below the threshold for ionization, any super-Ohmic conduction detected during the application of V_2 can be attributed to the presence of excess carriers generated during the first, high-voltage, part of the applied voltage pulse. (Injection is a negligible effect for voltages of the magnitude of V_2.)

Figure 14(a) shows the presence of an *S*-BNDC during the high-amplitude ($V = V_1$) part of the pulse. After 10 nsec have elapsed, the amplitude of the applied pulse is reduced to $V_2 = \frac{1}{9} V_1$ by the use of a reflected signal. The details of how the applied pulse is generated are given elsewhere[13]; an example of the pulse shape is given in the inset of Fig. 14(a). Conduction characteristics generated at two specific times during the stepped-down portion of the applied pulse are shown in Fig. 14(b).

The fact that the maximum value of V_2 does not generate excess carriers by ionization is clear from Fig. 1(b); those data demonstrate that the conduction characteristics for a flat-topped pulse ($V_1 = V_2$) with $E \lesssim 50$ V/cm and for 30 nsec $\lesssim t \lesssim 900$ nsec exhibit only Ohmic and sub-Ohmic conduction (hole injection is negligible compared to the effects dis-

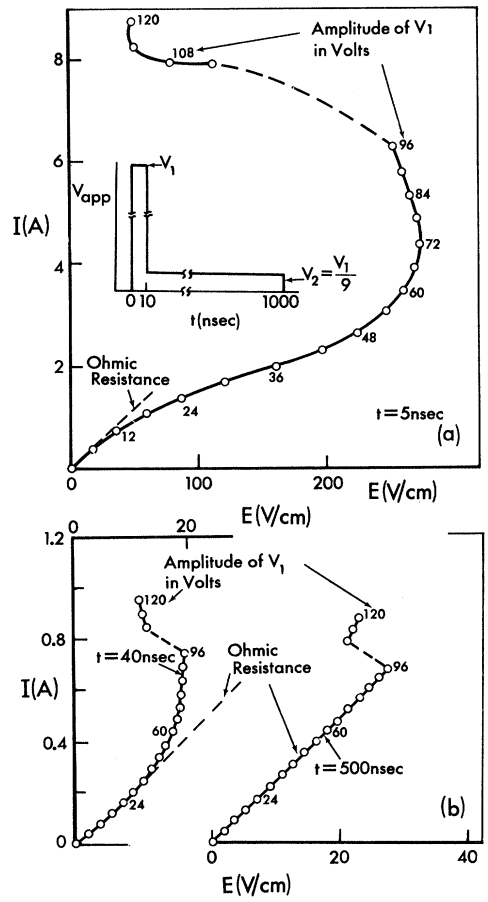

FIG. 14. (a) Conduction characteristic for $t=5$ nsec, generated during the high-voltage (V_1) part of the applied pulse, an example of which is shown in the inset. Successive data points correspond to 6-V increments in V_1 amplitude. Data refer to sample No. S35-1F-11n. (b) Conduction characteristics for $t=40$ and 500 nsec, generated during the low-voltage (V_2) part of the applied pulse. Successive data points correspond to 6-V increments in V_1 amplitude as they do in (a).

cussed in this section).

Successive data points in Figs. 14(a) and 14(b) correspond to 6-V increments in V_1; observing these increments allows correlation between conduction properties in the first part ($t<10$ nsec) and in the second part ($t>10$ nsec) of the applied pulse. In Fig. 14(a), the hot-electron effect is seen for $10 \lesssim V_1 \lesssim 42$ V. This effect is not seen in Fig. 14(b) because V_2 is too low to cause carrier heating, and the carriers that were heated for $t<10$ nsec have cooled off on a time scale of $\sim 10^{-12}$ sec. The beginning of carrier generation, seen in Fig. 14(a) for $42<V_1<96$ V, is also apparent in the $t=40$-nsec curve of Fig. 14(b) as nonlinear conduction, but is not apparent in the $t=500$-nsec curve. This lack of nonlinear conduction results because the carriers, generated during the first 10 nsec of applied power by ionization at a field less than required for BNDC but greater than that required for impact ionization, have almost completely decayed away by the time the 500-nsec curve was recorded. This technique of measuring excess-carrier lifetime is discussed in more detail in Sec. IV B 3.

The result of most importance in this section is as follows: When V_1 exceeds 96 V, an S-BNDC is established as illustrated in Fig. 14(a). The S-BNDC is seen also in the conduction curves of Fig. 14(b) and is precisely correlated with its occurrence during the high-voltage portion of the applied pulse. Since the S-conduction characteristics in Fig. 14(b) occur at $E<30$ V/cm, a magnitude far below E_{th}, these results establish that the occurrence of an S-BNDC requires excess-carrier generation.

d. *Hall-effect measurements.* A well-known technique used to determine carrier concentrations in semiconductors is the Hall effect. Charge carriers moving in orthogonal electric and magnetic fields are forced to a surface in the third direction, causing a potential to be developed known as the Hall voltage V_H. In n-InSb V_H is related to the net excess free-carrier density in the following manner[48]:

$$V_h \propto -I/n_f\,, \qquad \text{for } \mu B<1\,, \tag{10}$$

$$V_h \propto -I/(n_f-p_f)\,, \qquad \text{for } \mu B>1\,, \tag{11}$$

where n_f and p_f are, respectively, the free-carrier densities of electrons and holes, B is the magnetic field, I is the current, and μ is the carrier mobility. The analysis of Hall-effect data in n-InSb at high fields is complicated by the large mobility ratio $\mu_n/\mu_p \approx 30$–100 and the hot-electron effect. Also μ_n and, to a lesser extent, μ_p are functions of E and B. These equations are only valid if the stated conditions are fulfilled for both μ_n and μ_p.

V_H measurements in n-InSb during S-BNDC have been reported by Heinrich and Müller,[49] but were not interpreted in a manner consistent with the present model for the origin of the S-BNDC. One set of their measurements at $t=20$ nsec for $500 \lesssim B \lesssim 2000$ G shows a sharp decrease in V_H at a current coincident with the sharp E decrease in the conduction curve that attends the S-BNDC. For the two lowest values of B they used, 500 and 750 G, $\mu B<1$ so Eq. (10) is valid (for their applied fields, $\mu_n \approx 1.1\times10^5$ cm^2/V sec^2 and so $\mu_n B<1$ for $B \lesssim 900$ G). The sharp decrease in V_H (note that the current is held essentially constant) is now seen, therefore, to be the result of a sharp increase in n_f, as the S-BNDC is established.[50] Thus once again, the origin of the S-BNDC has been shown to be related to an increase in the excess-carrier density.

3. *Decay of Excess Carriers*

After a finite time the S-BNDC decays, leaving

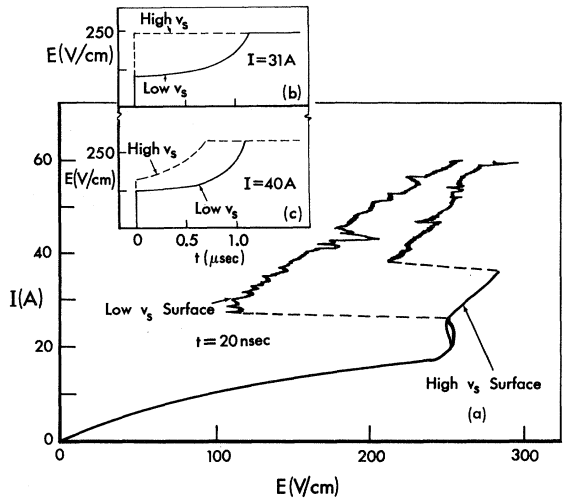

FIG. 15. (a) Conduction characteristics at $t = 20$ nsec for a sample prepared first with high-loss (high-v_s) surfaces and then with low-loss (low-v_s) surfaces. The insets (b) and (c) show the $E(t)$ associated with particular current and surface conditions. Data refer to sample No. WS85B-4n.

just conventional ionization in the nonequilibrium steady state (Fig. 12, I-E curve for $t = 270$ nsec). The ionization process which produced the S-BNDC is a transient effect since a HFD makes only one transit per applied-voltage pulse through the sample. As was pointed out in connection with Fig. 12, more than just the decay of the excess carriers produced by the HFD is involved in the decay of the S-BNDC. It is a complicated process involving alsc the hot-electron effect and generation outside the HFD. The modes by which the excess carriers decay can be radiative and/or nonradiative, through bulk traps and/or surface damage-layer traps. In

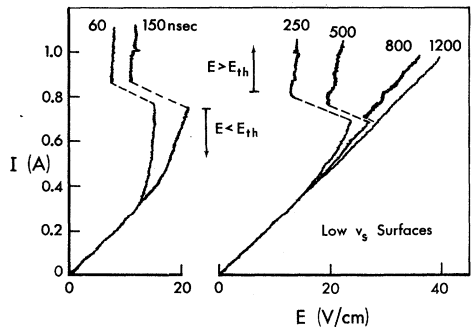

FIG. 16. Conduction characteristics parametric in time, generated by the stepped pulse shown in the inset of Fig. 14(a). These data show the decay of excess carriers generated by ionization throughout the sample ($E < E_{th}$) and excess carriers generated by a traveling HFD ($E \geq E_{th}$). The sample length is 2.5 mm. Data refer to sample No. S35-1F-11n.

TABLE II. Measured decay times of carriers produced by impact ionization throughout the sample ($E < E_{th}$) and during the transit of a high-field domain ($E \geq E_{th}$).

Sample	$E < E_{th}$ t_1(nsec)	$E < E_{th}$ t_2(nsec)	$E \geq E_{th}$ t_1(nsec)	$E \geq E_{th}$ t_2(nsec)
S24-2A-5n	26	160	17	130
S35-1F-11n	28	88	39	129
S20-4C-12n	13	120	39	145

this section are some measurements of the decay of excess carriers generated by impact ionization.

a. *Role of surfaces.* Surface properties play an important role in the decay of nonequilibrium carriers in InSb.[11] This is readily shown by observing nonequilibrium carrier decay in samples whose surfaces have either a relatively high recombination velocity v_s or a relatively low v_s.[51] Surfaces with a relatively high v_s are made by mechanically polishing the InSb with a 0.3-μm grit. This polishing technique creates a damage layer at the surface of the sample at least 0.4 μm deep[52] but probably[53] less than 3 μm. Then, v_s is reduced by etching ~10 μm of material from each of the mechanically polished surfaces with a lactic-acid etch.[54]

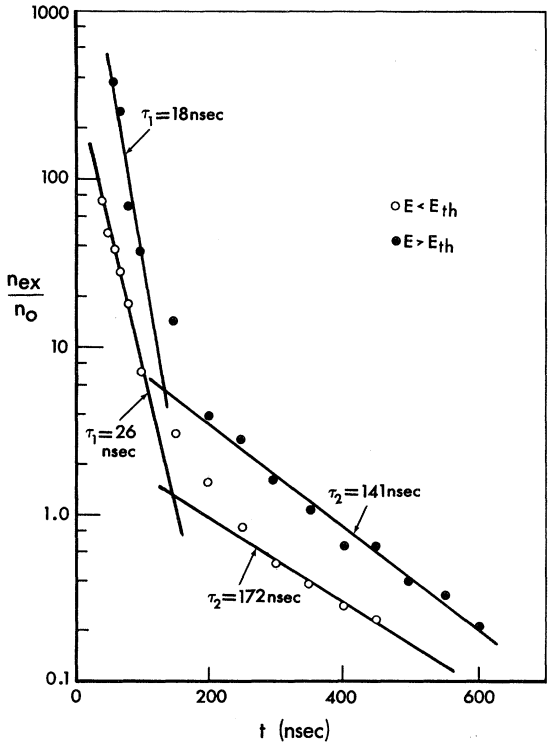

FIG. 17. Normalized excess-carrier densities as functions of time. From these data, two lifetimes τ_1 and τ_2 are apparent. Data refer to sample No. S24-2A-5n.

Figure 15 shows the effects of different types of surfaces on the same sample, in both the conduction characteristics and in E as a function of time. A current of 31 A in a sample with low-v_s surfaces produces an S-BNDC that persists for ~1.0 μsec; Fig. 15(b). At this current level, however, the same sample prepared with high-v_s surfaces exhibits no S-BNDC, Fig. 15(a), because the carrier losses are so high. When the voltage applied to the sample with high-v_s surfaces is increased sufficiently to trigger formation of a HFD in it, at 40A, the excess carriers created by the HFD decay in 0.7 μsec; Fig. 14(c).

b. *Lifetime measurements*. One approach to measuring the decay of excess carriers in a sample that experiences large conductivity changes employs the stepped pulse described in Fig. 14. As was pointed out in Sec. IV B 2 c, any carrier generation mechanism, such as injection and/or impact ionization, that occurs during the high-amplitude part ($t < 10$ nsec) of the applied pulse is detectable as super-Ohmic conduction in the low-amplitude part of the pulse, $t < 10$ nsec. The problem of carrier sweepout is reduced to a minimum by keeping the E low, always < 30 V/cm, so that the ambipolar drift[55] is < 750 μm during the decay measurements.

Figure 16 shows conduction characteristics generated using the stepped pulse. Two types of nonlinear conduction are seen: that due to excess carriers created during the application of V_1 [inset in Fig. 14(a)] by impact ionization throughout the sample at a field $E_{ii} < E < E_{th}$, and that due to excess carriers generated during the application of V_1 by impact ionization in a HFD, $E \geq E_{th}$. Both types of nonlinear conduction decay with increasing time until, at 1200 nsec, the conduction is simply linear.

The decay of excess carriers may be described as exponential, as shown in Fig. 17. The three samples for which such data were collected all exhibit two distinct lifetimes τ_1 and τ_2 for both $E < E_{th}$, as illustrated in Fig. 17. The values of the measured lifetimes are given in Table II. The shorter lifetime $\tau_1 \approx 27$ nsec persists for ~150 nsec. Carrier sweepout is negligible during this time (ambipolar drift ≈ 180 μm compared with a sample length of 2.5 mm), thus τ_1 is probably characteristic of the impact ionization decay, whereas τ_2 may be significantly affected by sweepout. Insufficient measurements have been made to conclude if τ_1 is significantly different for $E < E_{th}$ and $E \geq E_{th}$. Furthermore, the results presented here should be correlated with other types of nonequilibrium carrier-decay measurements in InSb before an interpretation of these results is attempted.

ACKNOWLEDGMENTS

We are grateful to J. O. Page for the use of his radiation recombination detection system and to both R. M. Lantz and H. R. T. Seibold for expert technical aid.

*Research on plasmas in semiconductors in the Department of Electrical Engineering is partially supported by the National Science Foundation.

[1]M. Glicksman and W. A. Hicenbothem, Jr., Phys. Rev. 129, 1572 (1963).

[2]B. Ancker-Johnson and C. L. Dick, Jr., Appl. Phys. Letters 15, 141 (1969).

[3]B. Ancker-Johnson, J. Phys. Soc. Japan Suppl. 21, 694 (1966); and 22, 1156 (1967).

[4]J. C. McGroddy and M. I. Nathan, J. Phys. Soc. Japan Suppl. 21, 437 (1966).

[5]D. K. Ferry and H. Heinrich, Phys. Rev. 169, 670 (1968).

[6]J. E. Smith, Jr., M. I. Nathan, J. C. McGroddy, S. A. Porowski, and W. Paul, Appl. Phys. Letters 15, 242 (1969).

[7]M. R. Lorenz, J. C. McGroddy, T. S. Plaskett, and S. Porowski, IBM J. Res. Develop. 13, 583 (1969).

[8]C. L. Dick, Jr., and B. Ancker-Johnson, Appl. Phys. Letters 18, 124 (1971).

[9]W. Shockley, Bell System Tech. J. 30, 990 (1951).

[10]Throughout this manuscript the symbol E refers to an averaged electric field strength, averaged over a portion of the sample or averaged over the entire length. The symbol $E(x)$ is used to denote the electric field strength at a particular point within a sample.

[11]B. Ancker-Johnson and W. P. Robbins, J. Appl. Phys. 42, 762 (1971).

[12]S. Tosima and K. Ando, J. Phys. Soc. Japan 23, 812 (1967).

[13]C. L. Dick, Jr. and B. Ancker-Johnson, Boeing Document No. D1-82-0997, 1970 (unpublished).

[14]$\tau = m_e^* \mu_e(E)/e \approx 2 \times 10^{-12}$ sec for $m_e^* = 0.015 m_0$, $\mu_e(E) = 2 \times 10^5$ cm^2/V sec, where m_e^* is the electron effective mass and $\mu_e(E)$ is the hot-electron mobility.

[15]V. A. Kovarskii, I. A. Chalkovskii, and E. P. Sinyavskii, Fiz. Tverd. Tela 6, 2131 (1964) [Sov. Phys. Solid State 6, 1679 (1965)].

[16]P. A. Wolff, Phys. Rev. 95, 1415 (1954).

[17]W. Shockley, Solid State Electron. 2, 35 (1961).

[18]J. L. Moll and N. Meyer, Solid State Electron. 3, 155 (1961); J. L. Moll and R. van Overstraeten, *ibid.* 6, 147 (1963); D. J. Bartelink, J. L. Moll, and N. I. Meyer, Phys. Rev. 130, 972 (1963).

[19]G. A. Baraff, Phys. Rev. 128, 2507 (1962).

[20]L. V. Keldysh, Zh. Eksperim. i Teor. Fiz. 48, 1692 (1965) [Sov. Phys. JETP 21, 1135 (1965)].

[21]V. A. Chuenkov, Fiz. Tverd. Tela 9, 48 (1967) [Sov. Phys. Solid State 9, 35 (1967)].

[22]W. P. Dumke, Phys. Rev. 167, 783 (1968).

[23]V. K. Aladinskii, Fiz. Tverd. Tela 7, 1813 (1965) [Sov. Phys. Solid State 7, 1460 (1965)].

[24]C. A. Lee, R. A. Logan, R. L. Batdorf, J. J. Kleimack, and W. Wiegmann, Phys. Rev. 134, A761 (1964).

[25]V. L. Dalal, Appl. Phys. Letters 15, 379 (1969).

[26]S. A. Porowski, W. Paul, J. C. McGroddy, M. I. Nathan, and J. E. Smith, Jr., Solid State Commun. 7, 905 (1969).

[27]D. N. Nasledov, Yu. G. Popov, Yu. S. Smetannikova,

Fiz. Tverd. Tela 6, 3728 (1964) [Sov. Phys. Solid State 6, 2989 (1965)].

[28]E. M. Conwell, *High Field Transport in Semiconductors*, Vol. 9 of *Solid State Physics*, edited by F. Seitz D. Turnbull, and H. Ehrenreich (Academic, New York, 1967), p. 157; see Eq. 3.6.23.

[29]J. B. Gunn, IBM J. Res. Develop. 8, 141 (1964).

[30]B. K. Ridley and T. B. Watkins, Proc. Phys. Soc. (London) 78, 293 (1961).

[31]C. Hilsum, Proc. IRE 50, 185 (1962).

[32]B. K. Ridley, Proc. Phys. Soc. (London) 82, 954 (1963).

[33]K. W. Böer and G. Döhler, Phys. Rev. 186, 793 (1969).

[34]P. N. Butcher, Rept. Progr. Phys. 30, 97 (1967).

[35]J. S. Heeks, IEEE Trans. Electron Devices ED13, 68 (1966).

[36]E. M. Conwell, Phys. Today 23 (No. 6), 35 (1970).

[37]A. G. Foyt and A. L. McWhorter, IEEE Trans. Electron Devices ED13, 79 (1966).

[38]A. Mooradian and H. Y. Fan, Phys. Rev. 148, 873 (1966).

[39]W. Fawcett and J. G. Ruch, Appl. Phys. Letters 15, 368 (1969).

[40]S. G. Liu, Appl. Phys. Letters 9, 79 (1966).

[41]P. D. Southgate, J. Appl. Phys. 38, 4589 (1967).

[42]C. L. Dick, Jr. and B. Ancker-Johnson, Bull. Am. Phys. Soc. 13, 1495 (1968).

[43]B. D. Osipov and A. N. Khvoshchev, Zh. Eksperim. i Teor. Fiz. 43, 1179 (1962) [Sov. Phys. JETP 16, 833 (1963)].

[44]J. Pehek and H. Levinstein, Phys. Rev. 140, A576 (1965).

[45]A. P. Shotov, S. P. Grishechkina, B. D. Kopylovskii, and R. A. Muminov, Fiz. Tverd. Tela 8, 1038 (1966) [Sov. Phys. Solid State 8, 865 (1966)].

[46]The radiation recombination detection system was developed by J. O. Page.

[47]B. Ancker-Johnson, in *Proceedings of the International Conference on Physics of Semiconductors* (Nauka, Moscow, 1968), p. 813.

[48]See, for example, R. A. Smith, *Semiconductors* (Cambridge U.P., Cambridge, England, 1959), p. 103.

[49]H. Heinrich and E. Müller, Solid State Commun. 6, 391 (1968).

[50]The Heinrich-Müller data also show a sharp decrease in V_H for $1 \leq B \leq 2$ kG; however, for those magnitudes of B, $\mu_n B > 1$ and $\mu_p B < 1$, so that the interpretation is no longer as straightforward as it is for $\mu_n B < 1$.

[51]B. Ancker-Johnson, Boeing Document No. D1-82-0908, 1969 (unpublished).

[52]E. N. Pugh and L. E. Samuels, J. Appl. Phys. 35, 1966 (1964).

[53]W. S. Chen (private communication).

[54]This etch is composed of 2.1:42.9:4.5 parts by volume of (i) distilled H_2O, (ii) 88.4% $CH_3CHOHCOOH$ (lactic acid), and (iii) 70% HNO_3 at 35–40 °C. Etch rate is ~ 0.5 μm/min.

[55]Calculated from $v_a t_m$, where v_a is the ambipolar drift velocity ($\mu_p \times E = 4 \times 10^3 \times 30$ cm/sec) and t_m is the maximum measurement time (600 nsec); see Fig. 17.

PHYSICAL REVIEW B VOLUME 5, NUMBER 2 15 JANUARY 1972

Zero-Bias-Conductance-Peak Anomaly of Ta-*I*-Al Tunnel Junctions at 0.3 K and 90 kG

Joel A. Appelbaum and L. Y. L. Shen
Bell Telephone Laboratories, Murray Hill, New Jersey 07974
(Received 10 August 1971)

We have studied the zero-bias–conductance-peak anomaly in a Ta-*I*-Al tunnel junction over a magnetic field range of 1 to 90 kG at a temperature of 0.3 K, and have made a detailed comparison between experiment and the *s-d* exchange theory of electron tunneling. Good agreement between theory and experiment was obtained for the line shape of the conductance peak and its zero-bias temperature dependence. The line shape of the conductance curves found experimentally at high magnetic fields agreed reasonably well with the theory if magnetic-field-induced lifetime broadening was included. There was, however, a serious discrepancy between theory and experiment for the zero-bias magnetoresistance.

INTRODUCTION

The occurrence of zero-bias–conductance-peak anomalies has been extensively studied in the past.[1–9] Three types of junctions exhibit this phenomenon. The first kind involves a transition metal as one electrode of the junction, such as Ta-*I*-Al and Nb-*I*-Al. The second kind consists of tunnel junctions like Sn-*I*-Sn or Al-*I*-Al with copper or transition-metal impurities as dopants in the insulating oxides. The third kind consists of metal-semiconductor junctions or semiconductor *p-n* junctions, whose conductance peaks originate from the impurities in the semiconductors. It is now generally assumed that these anomalies are due to the exchange scattering of tunneling electrons by magnetic impurities in or near the insulating barrier. Comparison between experiments and theory indicates qualitative as well as some quantitative agreements. A natural way to study this effect is to apply a strong magnetic field, which splits the spins into their Zeeman levels, thus modifying the conductance of the junction. In the meantime, the temperature of the junction is maintained as low as possible to re-