CORPORATE RESEARCH AND DEVELOPMENT • SCHENECTADY, NEW YORK

THEORY OF CONDUCTION IN ZnO VARISTORS

by

G.D. Mahan,* L.M. Levinson,
and H.R. Philipp
Signal Electronics Laboratory

Report No. 78CRD205 November 1978

TECHNICAL INFORMATION SERIES

CLASS 1

* Permanent Address: Physics Dept.,
Indiana University, Bloomington,
Indiana 47401

GENERAL ELECTRIC

General Electric Company
Corporate Research and Development
Schenectady, New York

TECHNICAL INFORMATION SERIES

AUTHOR Mahan, GD,* Levinson, LM, Philipp, HR	SUBJECT zinc oxide varistors	NO. 78CRD205
		DATE November 1978
TITLE Theory of Conduction in ZnO Varistors		GE CLASS 1
		NO. PAGES 17
ORIGINATING COMPONENT Signal Electronics Laboratory		CORPORATE RESEARCH AND DEVELOPMENT SCHENECTADY, N.Y.

SUMMARY

A theory is presented which quantitatively accounts for the important features of conduction in ZnO-based metal oxide varistors. This theory has no adjustable parameters. Using the known values of the ZnO band gap, donor concentration n_o and low voltage varistor leakage current activation energy, we predict a varistor breakdown voltage of ≈ 3.2 volts/grain boundary for $n_o = 10^{17}$ carriers cm^{-3} and T = 300°K. This compares well with measurements on a single grain-grain junction. The highly nonlinear varistor conduction derives from electron tunneling "triggered" by hole creation in the ZnO when the conduction band in the grain interior drops below the top of the valence band at the grain interface. The theory predicts coefficients of nonlinearity $\alpha = d\ln I/d\ln V$ as high as 50 or even 100.

*Permanent Address: Physics Dept., Indiana University, Bloomington, Indiana 47401.

KEY WORDS
zinc oxide varistor, tunneling, ceramic grain barrier

INFORMATION PREPARED FOR ______________________

Additional Hard Copies Available From — Corporate Research & Development Distribution, P.O. Box 43 Bldg. 5, Schenectady, N.Y., 12301

Microfiche Copies Available From — Technical Information Exchange, P.O. Box 43 Bldg. 5, Schenectady, N.Y., 12301

THEORY OF CONDUCTION IN ZnO VARISTORS

G. D. Mahan, L. M. Levinson, and H. R. Philipp

I. INTRODUCTION

Metal-oxide varistors are ZnO-based ceramic devices with highly nonlinear current-voltage characteristics. They are produced by sintering ZnO powder together with small amounts of other oxide additives, e.g., Bi_2O_3 and MnO_2. The resultant structure is comprised of semiconducting, n-type ZnO grains surrounded by insulating barriers at the ZnO grain boundaries. A review of the fabrication, microstructure, and electrical behavior of this material is given in Ref. 1.

ZnO varistors have proved to be useful in a variety of applications, particularly as high quality voltage surge suppressors. Consequently, active investigation has occurred in a number of laboratories to establish a mechanism for describing the extreme nonlinearity of this device.[1-16] While the experimental behavior of ZnO varistors is generally agreed upon,[1] the theories of varistor behavior have been unable to account for more than a limited portion of the measured data.

The observed current-voltage characteristic of varistors is often empirically described by the power law relation

$$I = kV^{\alpha}, \tag{1}$$

where the parameter α [$= d(\ln I)/d(\ln V)$] is a measure of the device nonlinearity. The parameter α varies with voltage. In the breakdown region, α can attain values well in excess of 50 and can exceed 100 under special conditions.

In addition to the explanation of the high observed values of α, a consistent theory of varistor conduction should explain the other features generally observed in these devices:

- The breakdown mechanism is substantially independent of temperature with a small negative temperature coefficient of breakdown voltage V_b at fixed-current[15]

$$\left[10^{-4} \lesssim \frac{1}{V_b}\left(\frac{dV_b}{dT}\right)_I \lesssim 10^{-3}\right]$$

- The varistor breakdown voltage is independent of carrier concentration in the ZnO grains.[17]
- Both increased temperature and increased ZnO carrier concentration increase the varistor leakage currents.[5, 15, 17, 18]
- The varistor breakdown voltage is about 3 to 3.5 volts per grain barrier[19] and is quite insensitive to the nature and amount of the various additives, the details of the firing procedure, and the processing steps used in the device fabrication.[1, 5] Indeed ZnO varistors without Bi_2O_3 have recently been reported,[20] and we have verified that these varistors (and other ZnO varistor systems without Bi_2O_3) have breakdown voltages of about 3.5 volts per grain boundary.
- The varistor capacitance decreases somewhat with applied DC bias[1, 12] and then sharply rises as the varistor enters the breakdown region.

Most previously proposed theories of varistor breakdown are either insufficiently nonlinear or have other discrepancies with the experimental data outlined above. Bernasconi _et al_.[11] and Philipp _et al_. (Ref. 15) have commented on a number of these theories. For completeness, a brief summary is presented here.

A. Avalanche Breakdown

Breakdown processes in semiconductor junctions often proceed via avalanche multiplication. The current-voltage characteristics associated with this process are extremely sharp,[21] and correspond in many cases to $\alpha > 1000$. Avalanche breakdown is characterized, however, by a positive temperature coefficient of breakdown voltage.[22] ZnO varistors invariably exhibit a negative temperature coefficient of breakdown voltage.

B. Space-Charge-Limited (SCL) Currents

The space-charge-limited process describes conduction within an insulator provided with ohmic contacts.[23] When applied to ZnO varistors[2, 11] the SCL theory describes conduction in an intergranular layer between adjacent ZnO grains. We believe, however, that the intergranular layer plays a relatively passive role in the varistor breakdown process. SCL descriptions of varistor behavior also neglect the injection problem at the grain-intergranular layer interface. Although large values of α can be achieved at the trap free limit voltage (V_{TFL}) this voltage is proportional to the trap density in the intergranular material. Hence V_{TFL} would be expected to be highly variable upon altering varistor composition or processing, in disagreement with experiment.

C. Thermal Activation Over a Double "Schottky" Barrier

The grain-grain interface is represented in this model by back-to-back depletion layer barriers with either a finite or zero thickness intergranular layer. For a general or exponential surface state density, maximum α values are limited to $\alpha \lesssim 25$ at room temperature.[11, 14] While it is understandable that device

Manuscript received September 13, 1978.

imperfection and inhomogeneity could reduce a theoretically predicted high value of α to a lower, experimentally observed value, it is clear that any theory predicting a lower α value than observed is faulty.

D. Simple Tunneling

The equations for Fowler-Nordheim tunneling[24] through a thin intergranular layer region lead to maximum values of $\alpha \sim 30$ at a current density of 10^{-3} ampere/cm^2. Although this theory accounts well for the general insensitivity of the varistor breakdown characteristic to varistor composition and processing, and gives the correct negative temperature coefficient of varistor breakdown voltage, the simple tunneling model is inadequate to explain α values in the 50 to 100 range.

More recently, Bernasconi et al.[13] and Knecht and Klein[25] have proposed a model based on back-to-back depletion layer barriers in which for $T \lesssim 400^\circ K$, conduction takes place by a hopping mechanism involving localized states within the barrier. The hopping process is described by an ad hoc parameter representing the number of hops made by an average electron upon crossing the barrier. An additional feature of this model is a tunneling process from deep traps (Ref. 13) or the valence band[25] of the ZnO into the conduction band at the low side of the barrier. This process serves to pin the ZnO conduction band edge. When the tunneling voltage is achieved, the depletion barrier height decreases rapidly with applied voltage, thereby giving rise to very high values of α.

A difficulty with this model is that it predicts the varistor breakdown voltage V_b to vary with ZnO donor density n_o. For example, Knecht and Klein[25] predict that V_b decreases by about 20% as n_o increases from 3×10^{16} to 3×10^{17} electrons/cm^3. Experimentally, however, V_b is essentially unaffected by variations in n_o as measured by infrared[8] and high frequency[26] techniques. Van Kemenade and Eijnthoven[19] have also observed a general insensitivity of varistor breakdown voltage to variations in processing and formulation. A second argument against this model is the need to justify the use of a single ad hoc hopping parameter to describe electron transport through ZnO depletion layers. These depletion layers should have different defect concentrations resulting from various fabrication procedures and varistor ingredients.

We shall outline a theory of varistor conduction which predicts high values of α and which is consistent with the experimental observations listed above (Ref. 27). The varistor breakdown mechanism is described in terms of electron tunneling through an abruptly thinned depletion layer at the ZnO grain boundary. The electrical barrier for conduction is modeled in terms of a double depletion layer barrier as shown in Fig. 1. The ZnO grains are also separated by an interface oxide layer which is assumed sufficiently thin so that it does not provide the dominant electrical barrier for the varistor current-voltage

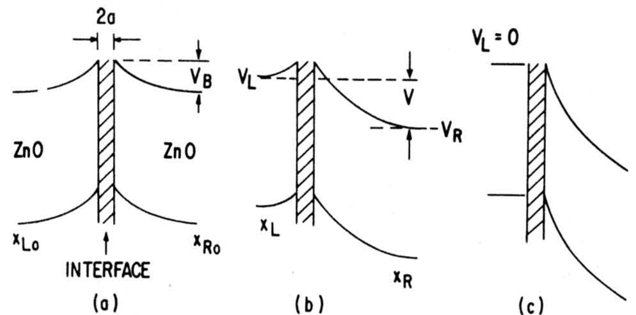


Fig. 1 (a) Double depletion layer model at zero applied voltage. Two ZnO grains are separated by an interface of a different oxide, thickness 2a. The barrier height is V_B. The depletion regions extend a distance $x_{Ro} - a = -a - x_{Lo}$ into the ZnO grains. (b) The application of a voltage V causes V_R to increase in magnitude and V_L to decrease. (c) If the interface charge is fixed, the back depletion layer vanishes ($V_L = 0$) at the critical voltage $V_c = 4V_B$.

behavior. We suggest that the varistor breakdown mechanism is controlled by hole creation in the forward bias side of the double Schottky barrier. Our calculations predict a breakdown voltage $V_b \approx 3.2$ eV in good agreement with observation. We emphasize that the holes in our model do not cause significant current flow by their recombination. Instead they affect the electron current by so altering the conduction band shape that electrons in the interface may tunnel out more easily to the adjacent ZnO grain. Thus the holes play a passive role in the current conduction. Nevertheless the theory readily predicts high values for the coefficient of nonlinearity α.

Our proposed theory of varistor behavior is presented in Section III. This theory evolved during a time when simpler theories were tried and found wanting. This was often the interplay between theory and experiment, whereby proposed theories were tested and found to disagree with new experimental information. Nevertheless, in Section II we present some of these simpler theories because they provide insight into the basis for the complete theory given in Section III.

II. PRELIMINARY MODELS

This section reviews our early theories and the reasons for their rejection. In this succession of theories, each becomes more complex to account for experimental observations. The final theory, presented in Section III, may be understood as the final step in this evolution. Other investigators were examining similar preliminary theories and rejecting them for similar reasons.[11]

Throughout this discussion, several dimensional symbols will be used. The first is a prefactor which enters into most of the expressions for current

$$J_o = n_o e v \sim 10^5 \text{ amperes/cm}^2 , \qquad (2)$$

where n_o is the number density of electrons in the ZnO conduction band ($n_o \sim 10^{17}cm^{-3}$), their thermal velocity is $v \sim 10^7$cm/sec, and e is the electron charge. These combine to give a maximum prefactor of about 10^5 amperes/cm^2. This prefactor is used in transport theories which describe the current through or over the interface barrier. For example, if the transport is by thermal excitation over the interface barrier, the current at very low voltages has the form

$$J = J_o \beta V e^{-\beta V_B}, \tag{3}$$

where V_B is the barrier height and $\beta = 1/kT$. Experimentally the activation energy is less than an electron volt, $V_B \sim 0.8$ eV.(4, 5) At room temperature we then obtain $\beta V_B \sim 32$ so that $J \approx 10^{-9}$ ampere/cm^2. In fact, this is the correct order of magnitude for the current density in the very low voltage region, where it is believed that transport is due to thermal activation.(1, 4-8, 10-14)

There is an important relationship which one soon appreciates in trying to construct models of varistor behavior. This is not really a theorem, but rather a rule of thumb relating the varistor nonlinearity to the magnitude of the predicted varistor currents. Many theories of varistor breakdown are models of tunneling through the interface barrier. The tunneling probability is usually evaluated by WKBJ theory, so that the current transport is given by

$$J \sim J_o e^{-W(V)}, \tag{4}$$

where $W = 2\int dx\, k(x)$ is the tunneling integral. Comparing Eq. (4) with Eq. (1) we obtain

$$\alpha = \frac{d \ln J}{d \ln V} = -\frac{d\, W(V)}{d \ln V}. \tag{5}$$

Since most WKBJ integrals W(V) are smooth functions of voltage, one typically finds $\frac{dW}{d\ln V} \sim W$, so that Eq. (4) becomes

$$J \sim J_o e^{-\alpha}, \tag{6}$$

where α is a smooth function of V.

The difficulty with this expression is that $\alpha \sim 50$, while $J_o \sim 10^5$ amperes/cm^2, so that the current density predicted by this formula is very small. Unfortunately, this is supposed to describe the breakdown region, where the current density is becoming very large, with J reaching the kiloampere range. Thus the rule of thumb, Eq. (6), relating the varistor nonlinearity α with the varistor current density J illustrates the difficulties inherent in any theory which has W a smooth function of voltage. This point was also made by Bernasconi, et al.(11) They correctly observe that most breakdown mechanisms fail to give a large enough α; alternatively if they give a large enough α, they give too small a current density.

The models we discuss have a sharp nonlinearity in the junction characteristics; something dramatic happens so that W is not a smooth function of voltage in the breakdown region. This feature is necessary for any tunneling model which describes varistor behavior characterized by a very sharp variation of current with voltage, i.e., $\alpha \gtrsim 50$.

A. Double Depletion Layer Model

We take as our starting point the double depletion layer barrier model description of the interface between adjacent ZnO grains. It is generally accepted (Refs. 1, 10-14) that such a model is appropriate in the low voltage region, i.e., before varistor breakdown is initiated. Difficulties with this model occur upon extending it to the highly nonlinear varistor breakdown region.

We consider two ZnO grains separated by an interface layer which is another oxide. The thickness and composition of this interface oxide are uncertain, in spite of intensive study.(1, 2, 12, 28) The configuration is shown in Fig. 1. The interface layer has a thickness 2a. In addition there is a depletion region in each ZnO semiconductor side, of width x_R-a and -a-x_L on the right and left sides, respectively. At V = 0, the depletion region has a thickness

$$-a - x_{Lo} = x_{Ro} - a = \left[\frac{\epsilon_o V_B}{2\pi n_o e^2}\right]^{1/2}, \tag{7}$$

where V_B is the barrier height.

By using constants appropriate for this system ($\epsilon_o \sim 10$, $n_o \sim 10^{17}$ electrons/cm^3, $V_B \sim 0.8$eV), one calculates that x_{Ro}-a $\sim$ 1000 Å. This number agrees with the junction thickness 2(x_{Ro}-a) $\sim$ 2000 Å deduced from capacitance measurements.(1, 2, 10-14) The actual interface thickness is quite variable in these polycrystalline ceramics. Some interface regions are quite thick; others appear quite thin. Our assumption is that transport is through the thinnest interface regions, which are probably much less than 30 Å thick.(12, 28) We shall thus assume $x_{Ro} \gg a$ and ignore thickness a, since it is unimportant in our theories.

An important aspect of the interface models is charge conservation. Since the depletion region contains a net amount of positive donor charge $n_o(x_R - x_L)$, then an equal amount of negative charge is somewhere in the interface region. ZnO, however, is believed to have no intrinsic surface states whatsoever in the band gap.(29) This implies that the negative charge resides in the interface oxide. We have assumed that the interface region is fairly thin. The interface region is assigned a charge per unit area σ. Although this has the dimensions of a surface charge, we mean it to be a bulk charge spread over a thin interface of thickness 2a. An estimate of the interface trap density needed to accommodate this is

$$n_T \sim \frac{2x_R\, n_o}{2a} \sim 10^{18} \text{ to } 10^{19} \text{ cm}^{-3}, \tag{8}$$

for an interface thickness between 100 Å and 10 Å. This trap density in the interface oxide appears reasonable, since it unquestionably has many bonding defects. The symbol σ_o will be used for the interface charge at zero applied voltage.

The first requirement of any model of varistor behavior is a formulation of how this interface charge density changes with applied voltage. Different models have been proposed. One preliminary model which we investigated thoroughly makes the assumption that it does not change. This model will now described.

Since the interface region has an overall charge neutrality, our assumption that the density of negative charge is fixed forces the amount of positive charge to be fixed. The amount of positive charge is just that in the two depletion regions

$$\sigma_o = n_o(x_R(V) - x_L(V)) = n_o\,(x_R + |x_L|). \tag{9}$$

Although both x_R and $|x_L|$ vary with voltage, their sum must be constant.

The other equation needed in this model is given by noting that the applied voltage per junction V is the difference of the barrier heights on the two sides of the junction, i.e.,

$$V = V_R - V_L = \frac{2\pi e^2 n_o}{e_o}\left(x_R^{\,2} - x_L^{\,2}\right). \tag{10}$$

Equations (9) and (10) are easily solved for x_R and x_L, with the result

$$\begin{aligned} V_R &= V_B(1 + V/4V_B)^2 \\ V_L &= V_B(1 - V/4V_B)^2 \\ x_L &= x_{Lo}(1 - V/4V_B) \\ x_R &= x_{Ro}(1 + V/4V_B)\ . \end{aligned} \tag{11}$$

The current flows by thermal excitation over the barrier, and is given by the relationship

$$J = J_o\left(e^{-\beta V_L} - e^{-\beta V_R}\right). \tag{12}$$

Our sign convention is that positive current and voltage mean electron flow to the right. Figure 1(b) shows the junction region when a positive voltage is applied. The right hand conduction band is lowered, so that V_R becomes larger. Also, the depletion region on the right is widened, so that x_R is increased. On the left side of the interface, the conduction band is raised and the depletion region is correspondingly narrowed.

According to Eq. (11), the left depletion region vanishes at a voltage $V_c = 4\ V_B$. We could thus interpret V_c as the critical voltage for varistor breakdown in this model. Others who have examined this model have assigned breakdown to a lower voltage, when the left depletion region is small but not zero. (Refs. 10, 11) Their assignment however does not lead to a theory possessing sufficient nonlinearity. We note also that for $V < V_c$ the junction current as given by Eq. (12) obeys the rule of thumb given by Eq. (6), i.e., if α is high, the current is low. However, if breakdown occurs when the flat band condition is reached, as shown in Fig. 1(c), then the current has a dramatic nonlinearity. For $V > V_c$, large amounts of current can flow with the junction offering no more resistance. Depending now on how one limits the current flow by putting some dissipation into this model, one may get values of α as high as one likes. The current-voltage characteristics are shown schematically in Fig. 2. Note the presence of a cusp in the curve at $V = V_c$. For $V < V_c$, the curve is essentially a plot of V_L in Eq. (11), and the cusp arises from the parabolic dependence of V_L on applied voltage.

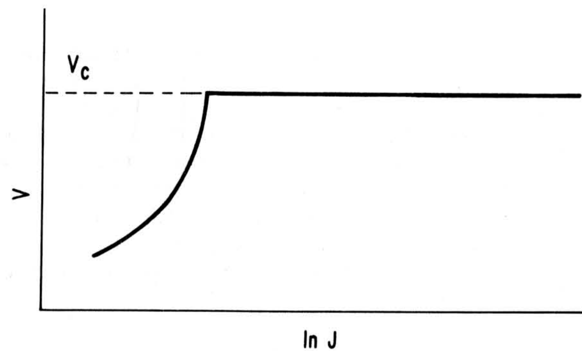


Fig. 2 Schematic of current-voltage characteristics for the model of Fig. 1. A sharp cusp in the lnJ versus V curves is predicted at the voltage V_c where $V_L = 0$ (flat band). The cusp is not observed experimentally.

This model has several attractive features. It predicts a critical voltage $V_c = 4V_B = 3.2$ eV which is in the correct range of experimental results.[18, 30] Another positive feature is that the breakdown mechanism has no explicit temperature dependence.

This model for varistor breakdown also has several features which disagree with experiment. First the experimental ln(I) vs. V plots[1] do not demonstrate the cusp behavior shown in Fig. 2. Although the cusp may be softened or smeared by averaging over junction thickness or other device characteristics, our attempts to do these averages show that the cusp shape remains. Its absence in experimental data is viewed as damaging evidence against the above model. A second experimental disagreement follows upon noting that the model predicts a strongly voltage-dependent activation energy E_A in the varistor prebreakdown region. From Eq. (11) we have

$$E_A = V_L = V_B(1 - V/V_c)^2\,. \tag{13}$$

This is a simple prediction which is easy to test. The test fails. The experimental activation energy is only slightly voltage dependent in the prebreakdown region.

The typical voltage dependence of the activation energy E_A for a 250 volt breakdown voltage varistor is shown in Fig. 3. The insensitivity of E_A to V can be regarded as conclusive evidence against this model. It also indicates that the interface charge density must vary with applied voltage, since that is the only assumption of the model. All of our other models incorporate a variation of charge density in the interface as the voltage is changed in the junction region.

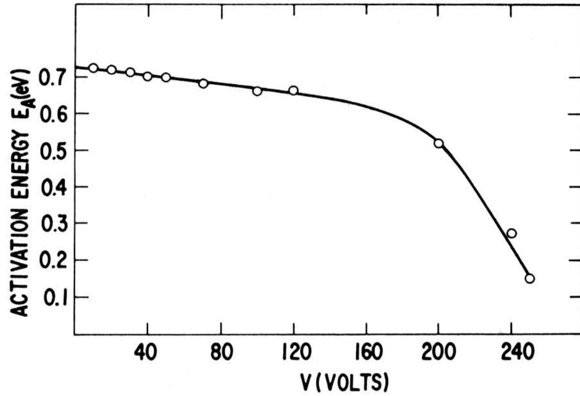


Fig. 3 Experimentally measured activation energy E_A as a function of applied voltage. This particular sample had $E_A = 0.73$ eV at $V = 0$. Note that E_A changes only slightly with voltage until the breakdown region, $V \approx 250$ volts.

The above model also has the feature that the capacitance is a constant in the prebreakdown region with

$$C_o = \frac{A\epsilon_o}{8\pi x_{Ro}} . \tag{14}$$

The experimental capacitance decreases slightly with voltage in the prebreakdown region. [1, 12, 14] and in the breakdown region the experimental capacitance rises dramatically. There is no way we have found to explain this within the model of constant interface charge. This is another argument against this model.

B. Two Step Transport

This section describes a model of prebreakdown behavior in which the restriction of fixed interface charge is relaxed. The immediate goal is to explain why the measured activation energy E_A for thermal excitation over the barrier is insensitive to voltage in the prebreakdown region. From the results of Section IIA, we must require that the negative charge density σ changes with voltage in the interface region. This may only happen if the electrons stop and spend time in the interface during their passage through it. Thus we are led to a model of two step transport. We assume that the electron transfer from ZnO to the interface, and then on to the next ZnO grain, are separate steps in the transport process. The theoretical procedure is to write a rate equation for each step. The interface charge σ is obtained by insisting that the current is conserved in the steady state; as much current flows into the interface as out. In this way, one obtains $\sigma(V)$, the steady state interface charge density as a function of applied voltage.

The four transport steps are defined in Fig. 4. On the right or left, current may flow either way between ZnO and the interface. The current flows from ZnO onto the interface are assumed to proceed by thermal activation. At low voltages (prebreakdown region), the depletion regions are so thick, that tunneling through them is very improbable. The activation energy is just the barrier height on the respective sides

$$J_{1L} = J_o\, e^{-\beta V_L} , \qquad J_{1R} = J_o\, e^{-\beta V_R} . \tag{15}$$

The current flow from the interface to the ZnO grains is assumed to be proportional to the charge density in the interface. It is also assumed that the flow is thermally activated. This is certainly valid at low values of V, but the assumption needs modification at higher values of V. This has been done, and the results are discussed below. For the present discussion, we examine the model using thermal activation for electron transport. In equilibrium with $V = 0$ the chemical potential is constant in the system, so that traps in the interface must be filled to the same chemical potential as in the ZnO grains. This means that the activation energy at $V = 0$ for transport from the interface must also be V_B, the zero voltage barrier height. Our last assumption is that the barrier height for the thermal excitation from the interface does not change when the interface charge density increases. This is equivalent to assuming that the trap density in the interface oxide is high, which seems reasonable. Although this assumption is ad hoc, it has the merit of simplicity. It then follows that the current flow from the interface oxide to the ZnO is given by

$$J_{2L} = J_{2R} = J_2 \sigma e^{-\beta V_B} , \tag{16}$$

where J_2 is a constant which is determined below.

The next step is to write the equation of continuity for current. In the steady state, with constant applied voltage, as many electrons flow into the interface as from it. This then implies

$$J_{1L} + J_{1R} = J_{2L} + J_{2R} = 2J_2 \sigma e^{-\beta V_B} . \tag{17}$$

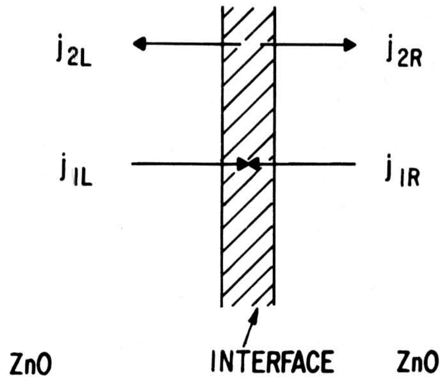


Fig. 4 Labeling scheme for the two step model of charge transport.

using Eq. (16). From Eqs. (17) and (15) we obtain the interface charge density

$$\sigma = \frac{J_o}{2J_2}\left[e^{-\beta(V_L - V_B)} + e^{-\beta(V_R - V_B)}\right]. \tag{18}$$

The value of J_2 is obtained from the zero voltage equilibrium condition where $V_L = V_R = V_B$ and $\sigma = \sigma_o$. It follows that

$$J_2 = J_o/\sigma_o . \tag{19}$$

Substituting Eq. (19) in Eq. (18) we obtain

$$S \equiv \sigma/\sigma_o = 1/2\left[e^{-\beta(V_L - V_B)} + e^{-\beta(V_L - V_B)}\right]. \tag{20}$$

In Eq. (20), $S = \sigma/\sigma_o$ is the normalized interface charge.

To solve for S, V_L, and V_R as functions of $V = V_R - V_L$, one more equation is required. This equation is obtained from charge conservation at the junction. Since the amount of interface charge must equal the amount of positive charge in the depletion region, we have

$$\sigma = n_o(x_R - x_L) . \tag{21}$$

Since the junction width is proportional to the square root of the junction voltage we obtain from Eq. (21)

$$S = \frac{1}{2\sqrt{V_B}}(\sqrt{V_R} + \sqrt{V_L}) . \tag{22}$$

These equations were solved numerically to obtain the dependence of S, V_R, and V_L on V. The numerical results may be represented, with a very high degree of accuracy, by the approximate analytical solution

$V > 0$:

$$V_R = V_B + V$$

$$V_L = V_B$$

$$S = \frac{1}{2}(1 + \sqrt{1 + V/V_B}) \tag{23}$$

$V < 0$

$$V_R = V_B$$

$$V_L = V_B - V$$

$$S = \frac{1}{2}(1 + \sqrt{1 - V/V_B}) .$$

The error in the analytical approximation is only of the order kT = 0.025 eV. One can also show that in this model the current flow to the right is exactly given by

$$J = J_{1L} - J_{2L} = J_o e^{-\beta V_B} S \tanh(\beta V/2) . \tag{24}$$

For V >> kT the hyberbolic tangent becomes unity, and the only voltage dependence of the current is through S(V). By using the approximate forms, we obtain the prebreakdown current behavior as

$$S = \frac{1}{2}J_o(1 + \sqrt{1 + |V|/V_B})\, e^{-\beta V_B} \tanh(\beta V/2) . \tag{25}$$

Thus in the prebreakdown region the current J is thermally activated, and the barrier height V_B is voltage independent.

In this model, the double depletion layer barrier behaves largely as a single metal-semiconductor Schottky barrier. The junction behavior becomes very one-sided, with the full voltage drop (within about kT) on one side. The left side of the junction hardly changes its barrier height or the thickness of the depletion region. This behavior is experimentally verified in Fig. 3.

It is interesting to derive the junction capacitance C in this model. From Eq. (23)

$$C = \frac{dQ}{dV} \approx \frac{C_o}{\sqrt{1 + V/V_B}} , \tag{26}$$

i.e., C decreases slowly with V. The experimental capacitance also shows a slow decrease with increasing voltage in the prebreakdown region.[1, 12] The theoretical dependence [Eq. (26)] is, however, too steep -- it falls about 50% by the breakdown voltage, while the experimental decrease is only around 20%.

The two step transport model provides a reasonably satisfactory account of the prebreakdown region. The main problem with it is theoretical. We have checked the assumption that the transport would all be thermally activated and found it untrue in one case. At large positive voltages, when the electron flow is to the right, the transport from the interface to the ZnO becomes dominated by tunneling rather than by thermal activation. This is because the tunneling barrier becomes very thin at large values of V. We have therefore modified the theory to take this into account. This is described in Section III, where we present our final theory of the varistor behavior.

III. VARISTOR THEORY

A. Prebreakdown Region

Our model for the prebreakdown region is an extension of that discussed in Section IIB. The interface has a charge density $S = \sigma/\sigma_o$ which varies with applied voltage. Transport is viewed as a two step process whereby electrons go from ZnO to the interface in one step, and on through the interface to the next ZnO grain in the second step. The interface charge S is determined by self-consistently solving several coupled equations which express the conservation of charge and current.

According to the definitions in Fig. 4, current flow is conserved if the four currents obey

$$1 = (J_{1L} + J_{1R}) / (J_{2R} + J_{2L}) . \quad (27)$$

The transport from the interface $J_{2L} + J_{2R}$ is assumed to be proportional to the amount of interface charge σ. When the electron flow is to the right, our calculations show that the right hand ZnO depletion layer becomes thin enough that tunneling through it is important. For right hand electron flow, we modify Eq. (16) for the transport from the interface to the ZnO, to give

$$J_{2R} = J_o S e^{-\beta V_B} \Lambda(V_R) , \quad (28)$$

$$\text{where} \quad \Lambda(V_R) = \frac{1}{u} \int_0^{VB} dE\, e^{\beta E} e^{-W(E, V_R)} , \quad (29)$$

$$u = \int_0^{V_B} dE\, e^{\beta E} e^{-W(E, V_B)} , \quad (30)$$

$$W(E, V_R) = (2/\hbar\omega_p)\left\{\sqrt{V_R E} - (V_R - E) \ln[(\sqrt{V_R} + \sqrt{E})/\sqrt{V_R - E}]\right\} , \quad (31)$$

and

$$\omega_p^2 = 4\pi n_o e^2/m\varepsilon_o .$$

The integral in $\Lambda(V_R)$ is over the interface energy from the top of the barrier (E = 0) down to the Fermi level ($E = V_B$) in the interface. The factor exp (-W) is the WKBJ tunneling probability through the barrier, and exp $[-\beta(V_B-E)]$ is the probability of a particle being thermally excited to this energy. The factor u is a normalization integral. At low voltages, tunneling is improbable, and transport is by thermal excitation; the main weight of the integrand comes from $E \sim 0$ and $\Lambda \sim 1$. At higher values of V, as the Schottky barrier becomes thinner, tunneling is more likely, and the electrons move through the barrier at larger values of E (E is the energy measured downward from the top of the interface). The factor Λ increases by about a factor of ten from zero voltage up to a critical voltage $V_c \approx 2.4$ eV (initiation of breakdown, see below). The derivation of Eq. (31) is given in Appendix A.

In the same way as we have generalized J_{2R} to take account of tunneling as well as thermal excitation, one could also take similar weighted averages of tunneling and thermal excitation for the other three transport steps. These currents, are however, dominated by thermal excitation, so we did not do so. The other three factors in Eq. (27) have the same definitions as in Section II, so that Eq. (27) may be rewritten as

$$S = \frac{e^{-\beta(V_L - V_B)} + e^{-\beta(V_R - V_B)}}{1 + \Lambda(V_R)} = e^{-\beta(V_L - V_B)} \frac{[1 + e^{\beta V}]}{1 + \Lambda(V_R)} . \quad (32)$$

In deriving the right hand equality, we used $V = V_R - V_L$. The third equation linking our three unknowns (S, V_R, and V_L) follows from charge conservation as in Eq. (22)

$$S = \frac{1}{2\sqrt{V_B}}\left[\sqrt{V_R} + \sqrt{V_L}\right] . \quad (33)$$

These three equations were solved self-consistently to obtain S, V_R, and V_L as a function of V. The solution is very similar to the prebreakdown model of Section II. For electrons flowing to the right ($V > 0$), we obtain, for example, the approximate solution

$$V_L = V_B$$

$$V_R = V_B + V \quad (34)$$

$$S = \frac{1}{2}\left[1 + \sqrt{1 + V/V_B}\right] .$$

This is identical to Eq. (23). These values are hardly changed, even when $\Lambda(v)$ increases by a factor of ten over the value of unity assumed in Section II. The reason for this is deduced upon examination of the right hand side of Eq. (32). Since S remains of order unity in the prebreakdown region, an increase in Λ must mean that

$$e^{-\beta(V_L - V_B)} \sim \Lambda . \quad (35)$$

Since $\beta = 40$ at room temperature, then $V_L - V_B$ need only change by 0.1 eV to satisfy this identity. So V_L changes only by about 0.1 eV in the prebreakdown region while V_R changes the remaining amount. Again the voltage drop is very one sided, with the back side of the junction hardly changing its activation energy V_L while V is changed. As we noted earlier, measurements on varistor samples show that the activation energy in the prebreakdown region is in fact insensitive to voltage (Fig. 3).

In order to calculate the current, we use the following expression for electron flow to the right

$$
\begin{aligned}
J &= J_{2R} - J_{1R} \\
&= J_o e^{-\beta V_B}\left[S\Lambda(V_R) - e^{-\beta(V_R - V_B)}\right], \qquad (36)
\end{aligned}
$$

from Eqs. (28) and (15). For $V >> kT \sim 0.03$ eV, the second term in brackets may be ignored. Then the current is given by the product $S\Lambda(V_R)$. In the prebreakdown region, S increases only by 50%, while $\Lambda(V_R)$ increases by a factor of 10. This is in agreement with actual varistor behavior, which shows a much larger increase in current than would be accounted for by the factor S alone.

While the factor $\Lambda(V_R)$ is quite important in describing an increase in the current as electrons begin to tunnel through the depletion layer barrier rather than being thermally activated over it, this increase in current seems to have little influence on other prebreakdown varistor properties, such as the interface charge density S or the activation energy as a function of voltage. Thus, the capacitance is still given by Eq. (26). It depends only on the charge densities, which are not altered by the increased tunneling.

This model agrees with the experimental properties of varistors in the prebreakdown regime. The capacitance decreases slowly with increasing voltage. The current is thermally activated, with a voltage independent activation energy. The current increases smoothly with voltage, with no unusual behavior until breakdown.

B. Breakdown Region

In our model the highly nonlinear breakdown behavior of varistors is caused by hole creation in the ZnO. This happens when the conduction band drops below the top of the valence band on the same side of the junction. The case of electron flow to the right is shown in Fig. 5. The chemical potential on the right is taken at the conduction band energy. It is assumed that this chemical potential is constant on the right of the interface so that holes are present near the interface. Holes first appear in the valence band of the right hand ZnO grain at a voltage given by

$$
V_c \approx E_G - V_B \approx 3.2 - 0.8 = 2.4 \text{ eV}, \qquad (37)
$$

where E_g is the ZnO band gap.[31] Here V_c should not be identified with the so called varistor breakdown voltage V_b. Experimentally V_b is defined as the voltage at which the device current density $J = 1$ ma/cm^2. Varistor breakdown is initiated well before this current density is achieved. Values of V_b (calculated numerically) are given below.

Referred to the valence band edge, the hole chemical potential is

$$
\mu_H = V_R - E_G \; . \qquad (38)
$$

We assume that the holes live a long time, i.e., that electron hole recombination is slow. The pronounced

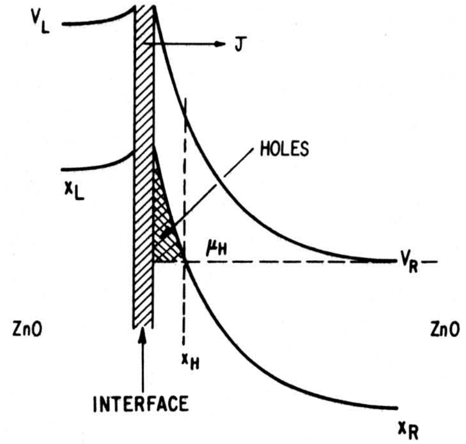


Fig. 5 The hole model. The holes are created on the forward side of the junction whenever the hole chemical potential is below the valence band edge. The electron tunneling from the interface is indicated by the arrow labelled J.

current nonlinearity, corresponding to the initiation of varistor breakdown is not due to the hole recombination. Rather it is caused by the drastic change the holes make in the conduction band shape U(x) near the interface. The electrons still tunnel from the interface near its energetic top, as shown by the arrow in Fig. 5. This tunneling becomes much easier because the tunneling barrier becomes very thin, i.e., the holes cause thinning of the potential barrier to electrons tunneling from the interface.

The change in potential shape is calculated from Poissons equation

$$
\frac{d^2}{dx^2} U = \frac{4\pi e^2}{\epsilon_o}\left[n_H(x) + n_o\right], \qquad (39)
$$

where $n_H(x)$ is the hole charge density. It is calculated in the Thomas-Fermi approximation as

$$
n_H(x) = \frac{1}{3\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\left[U - \mu_H\right]^{3/2}. \qquad (40)
$$

It is convenient to define

$$
f(x) = U(x) - \mu_H \qquad (41)
$$

so that Eq. (39) becomes

$$
\frac{d^2}{dx^2} f = \frac{4\pi e^2}{\epsilon_o}\left[\frac{1}{3\pi^2}\left(\frac{2mf}{\hbar^2}\right)^{3/2} + n_o\right]. \qquad (42)
$$

Multiplying by df/dx and integrating gives

$$
\frac{\epsilon_o}{8\pi e^2 n_o}\left(\frac{df}{dx}\right)^2 = b f^{5/2} + f + C, \qquad (43)
$$

where b is given by

$$b^{-1} = \frac{15}{2}\pi^2 n_o \left(\frac{\hbar^2}{2m}\right)^{3/2} \tag{44}$$

and C is a constant of integration with

$$C = V_R - \mu_H = E_G . \tag{45}$$

Details of the derivation of C are given in Appendix B. Note that the constant b is quite large, so that the term $bf^{5/2}$ is the largest on the right whenever f is not near zero.

From Eq. (43) the shape of the potential barrier is reduced to a quadrature

$$(x_H - x)\left[\frac{8\pi e^2 n_o}{\epsilon_o}\right]^{1/2} = \int_0^{f(x)} dy \left[E_G + y + by^{5/2}\right]^{-1/2} \tag{46}$$

where $y = (\gamma n_o)^{1/2}(x_R - x)$ and $\gamma = 2\pi e^2/\epsilon_o$ are introduced in Appendix A. The constant x_H is the maximum extent of the injected holes, as shown in Fig. 5. Since the integral on the right is absolutely convergent, even when $f(x) \to \infty$, there is a maximum absolute value of x_H. This equation determines x_H, since all other quantities are known.

The injected holes will result in a large increase in the positive charge density in the depletion regions. Because of charge neutrality, there must be an equal amount of electron charge in the interface. Thus the model implies the interface must charge up to cancel the charge of the holes. The total amount of charge in the hole region is

$$n_T = \int_o^{x_H} dx \left[n_H(x) + n_o\right] \tag{47}$$

which from Poissons equation gives

$$n_T = \frac{\epsilon_o}{4\pi e^2}\left[\left(\frac{df}{dx}\right)_{x_H} - \left(\frac{df}{dx}\right)_o\right] . \tag{48}$$

We can evaluate Eq. (48) using Eqs. (43) through (45), and from this we obtain a simple expression for the electron charge in the interface during breakdown

$$S = \frac{1}{2\sqrt{V_B}}\left[\sqrt{V_L} + \left(V_R + b\mu_H^{5/2}\right)\right]^{1/2} . \tag{49}$$

This expression is similar to the prebreakdown results [Eq. (22)]. In the breakdown region, the last factor in the square root totally dominates, so that an adequate approximation is

$$S \simeq \frac{1}{2}\left(\frac{b}{V_B}\right)^{1/2} \mu_H^{5/4} = \frac{1}{2}\left(\frac{b}{V_B}\right)^{1/2} (V_R - E_G)^{5/4} . \tag{50}$$

which is basically just the hole charge.

Another equation for S is obtained, as before, from the two step transport model. The current from the hole recombination is assumed small, and can be neglected, so the two step transport model has exactly the same mathematical form as in the prebreakdown model, i.e.,

$$S = e^{-\beta(V_L - V_B)}\left(\frac{1 + e^{-\beta V}}{1 + \Lambda(V_R)}\right) . \tag{51}$$

The main difference is that now $\Lambda(V_R)$ becomes very large in the breakdown region. This occurs because the holes make the tunneling barrier very thin. The same equation [Eq. (29)] is used to evaluate $\Lambda(V_R)$, except the WKBJ tunneling integral is different. One must account for the thin barrier due to holes. There are two possible situations. For $E < \mu_H$, the electron tunnels through a barrier which is totally determined by the shape deduced from Eq. (46) (see Appendix C)

$$W(E, V_R) = \frac{2}{\hbar\omega_p}\int_o^E dy \frac{\sqrt{E-y}}{\sqrt{V_R - y + b(\mu_H - y)^{5/2}}} \cdot \tag{52}$$

In the case $E > \mu_H$, the electron tunneling trajectory is partly in the potential region determined by the holes (W_1), and partly in the region where there is a parabolic conduction band shape for $x > x_H$, (W_2) i.e.,

$W = W_1 + W_2$, where

$$W_1 = \frac{2}{\hbar\omega_p}\int_o^{\mu_H} dy \frac{\sqrt{E - \mu_H + y}}{\sqrt{E_G + y + by^{5/2}}} \text{ and} \tag{53}$$

$$W_2 = \frac{2}{\hbar\omega_p}\left\{\sqrt{E_G(E-\mu_H)} - (V_R - E)\ln\left(\frac{\sqrt{E_G} + \sqrt{E - \mu_H}}{\sqrt{V_R - E}}\right)\right\} . \tag{54}$$

Eqs. (28) through (31), (49), and (51) determine the self-consistent results for the V-dependence of S, V_R, and V_L. These were solved numerically to obtain the correct description of breakdown characteristics.

In Fig. 6 we plot the theoretical dependence of the left junction voltage V_L as a function of applied voltage V for a grain electron density of 10^{17} cm^{-3} and a barrier height of V_B = 0.73 eV. The latter value was chosen to provide a comparison with the experimental activation energy in Fig. 3. These data have been replotted in Fig. 6 to provide a comparison between theory and experiment. The experimental curve is on a multigrain device, and its voltage scale has been set by making the two curves agree at the mid-height point. The theoretical curve for V_L is

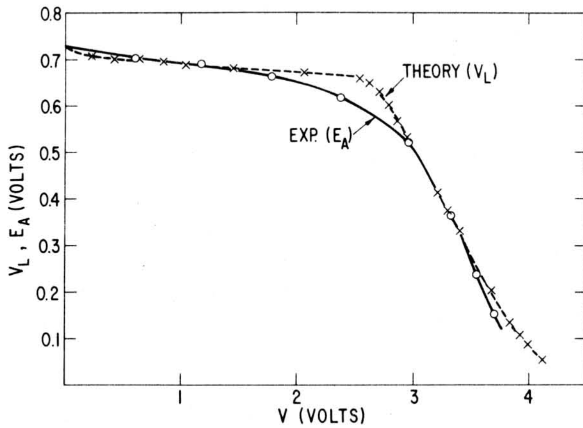


Fig. 6 The theoretical dependence of the left junction voltage V_L as a function of applied voltage V taking V_B = 0.73 eV, $n_o = 10^{17}$ cm^{-3}, T = 300°K. The measured activation energy E_A, normalized to a single junction (see text), is also plotted.

for a single grain interface. It shows a reasonable agreement with the experimental result for the multigrain data. The activation energy has only a slight voltage dependence at small values of applied voltage V, and decreases dramatically in the breakdown region. We regard this agreement between theory and experiment as an important confirmation of our theory.

The numerical solution of the coupled equations [Eqs. (28) through (31), (49), and (51)] is perhaps somewhat opaque from a physical understanding point of view. However one can deduce most of the results by some simple arguments.

An approximate solution is obtained from the following algorithm. One selects a value of V_R larger than E_G. The interface charge is given by Eq. (50). The computer is then used to calculate $\Lambda(V_R)$, since the WKBJ integrals are too complicated to permit analytical approximation. V_L is thereupon obtained from Eq. (51), i.e.,

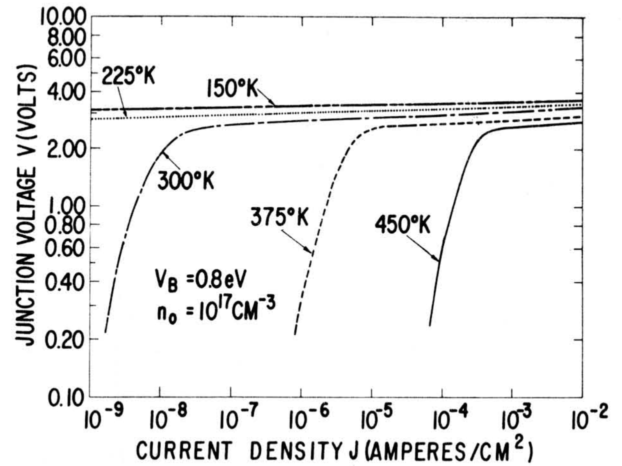
$$V_L \approx V_B - kT \ln (S + S\Lambda) \ . \quad (55)$$

Finally one deduces $V = V_R - V_L$.

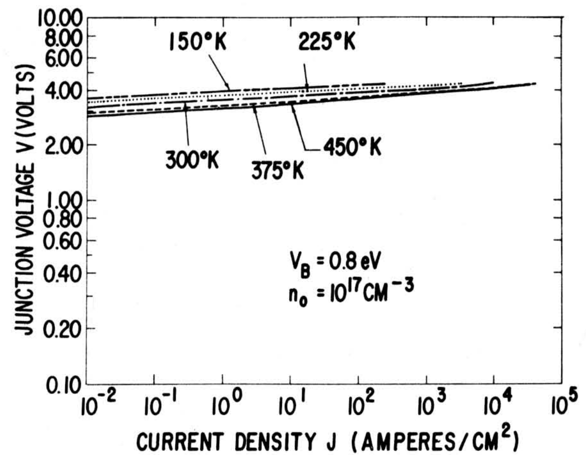
The total current to the right is still given by Eq. (36). The second term in the bracket is negligible, so that one has

$$J = J_o S e^{-\beta V_B} \Lambda \ . \quad (56)$$

As shown in Figs. 7 and 9 the current rises sharply with V in the varistor breakdown region. This increase is basically a result of the large increase in Λ. The increase in S is nearly a linear function of V and does not affect J to a marked extent. Calculated values of α are given in Figs. 8 and 10. Depending on values of n_o and T, the exponent α can be very large.



(a) low current densities



(b) high current densities

Fig. 7 Log-log plot of the predicted variation of current density J with applied voltage V for a single varistor grain junction. The curves are calculated for the temperatures indicated. We have set the ZnO grain donor density $n_o = 10^{17}$ cm^{-3} and the junction barrier height V_B = 0.8 eV.

We note that the current J in Eq. (56) is not thermally activated in the breakdown region. When $\Lambda(V_R)$ is dominated by tunneling we have $\Lambda(V_R) \sim \exp(\beta V_B)$ since the major contribution to the integral in Eq. (29) is at $E \sim V_B$. Thus the quantity $\Lambda \exp(-\beta V_B)$ is temperature insensitive as is of course verified by numerical evaluation (Figs. 7 and 9).

We may also calculate the varistor capacitance by evaluating the total electrostatic energy $\mathcal{E}$ and finding its second derivation with respect to V.

$$\mathcal{E} = \frac{\varepsilon_o A}{8\pi} \int dx \left[\frac{dU}{dx}\right]^2 , \quad (57)$$

$$C = \frac{d^2}{dV^2} \mathcal{E} \ . \quad (58)$$

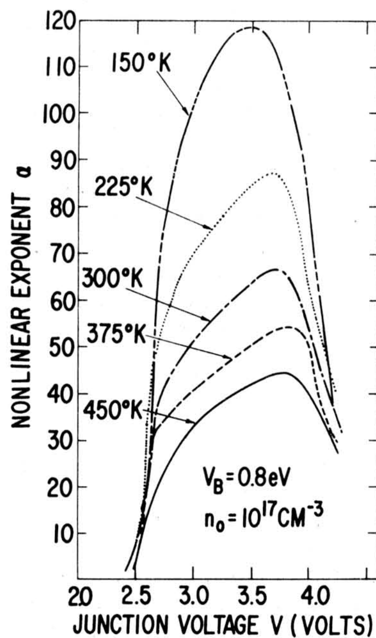
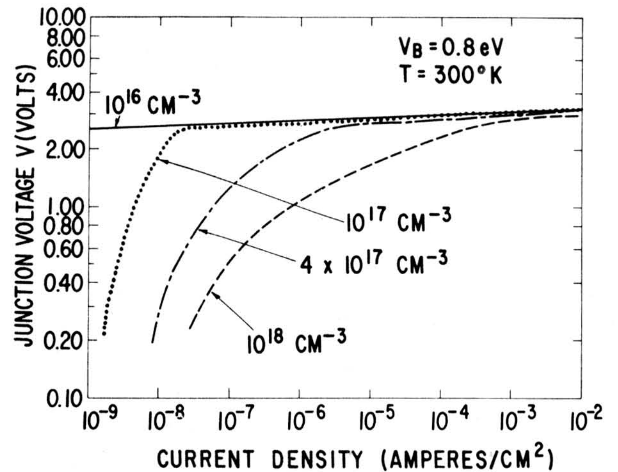


Fig. 8 Variation of the varistor coefficient of nonlinearity $\alpha = d\ln I/d\ln V$ for the curves of Fig. 7. The abscissa is the voltage V applied to the junction.

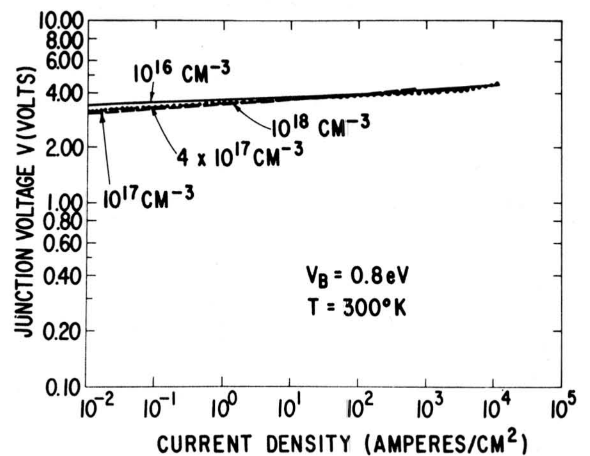
For example, in the breakdown region, the total energy density is

$$\mathcal{E} = \frac{A\epsilon_o}{8\pi}\left[\frac{8\pi e^2 n_o}{\epsilon_o}\right]^{1/2}\left\{\frac{2}{3}V_L^{3/2} + \frac{2}{3}(V_R - \mu_H)^{3/2} + \int_o^{\mu_H} dy\left[E_G + y + b\, y^{5/2}\right]^{1/2}\right\} . \qquad (59)$$

The three terms in braces are from the left depletion layer, the right depletion layer, and from the injected hole region, respectively. We have evaluated Eq. (59) and its second derivative numerically. The result is shown in Fig. 11 for $n_o = 10^{17}$ cm^{-3} and T = 300°K. The capacitance drops by about 40% with increasing bias voltage until V = 2.5 eV. Figure 11 shows that the theoretical capacitance then rises quite steeply. This is a prediction of the theory which was confirmed by our experiments; this is discussed below. The differential capacitance evaluated from Eq. (59) also shows a narrow (~0.1 eV) region of negative values at 2.5 eV before the steep rise. This is not observed experimentally, and is not expected because of considerations regarding the response time of the breakdown. The mechanism of hole creation, which is discussed below, is quite slow in the region of V ~ 2.5 eV, so that the holes are not created fast enough in this region to be measured. Thus Eq. (59) describes a system in equilibrium, which may take too long to measure. The response time of the system in this voltage range is discussed further below.



(a) low current



(b) high current

Fig. 9 Log-log plot of the predicted variation of current density J with voltage V for a single varistor grain junction. The curves are calculated for the ZnO grain donor densities indicated. We have set T = 300°K and the junction barrier height V_B = 0.8 eV.

The increase in capacitance is easily explained. Capacitance is charge divided by distance. In the breakdown region, the hole creation provides a large amount of charge, and the separation between it and the electron charge in the interface is small. Therefore the charge is large and the distance small. This results in a very large capacitance.

Before turning to a comparison of the model with experimental data, it is worthwhile to consider some details of possible mechanisms for creating the holes. The response time of a varistor is less than 10^{-9} sec (Ref. 1) and therefore holes must be created on this time scale. It is unlikely that holes are made by direct tunneling of electrons across the depletion region from valence band to conduction band. Using a two

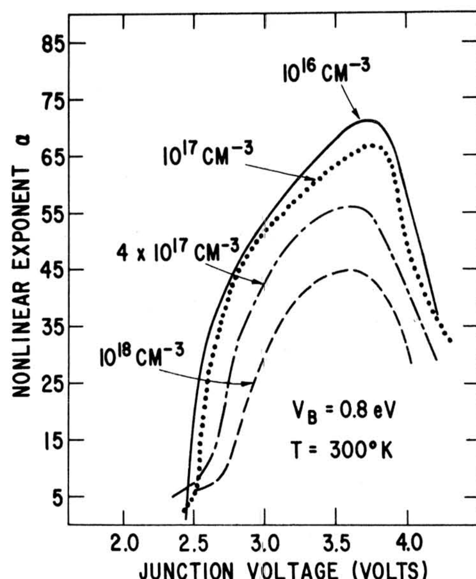


Fig. 10 Variation of the varistor coefficient of nonlinearity $\alpha = d\ln I/d\ln V$ for the curves of Fig. 9. The abscissa is the voltage V applied to the junction.

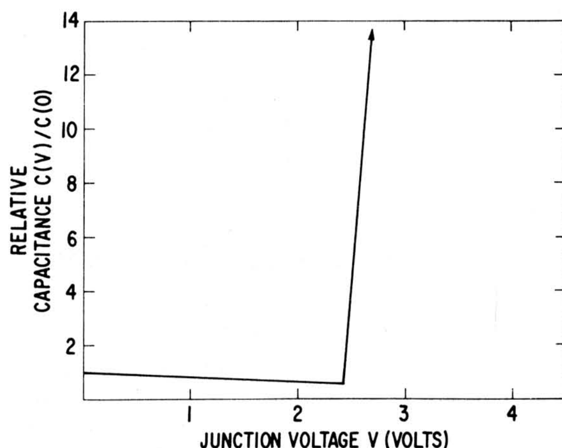


Fig. 11 Calculated variation of the varistor junction capacitance with applied DC bias voltage V. A steep rise in the capacitance is predicted to occur concurrent with the initiation of the hole region.

band model[32, 33] the probability of such an event is $\sim\exp(-W)$ where $W = \frac{4}{3} E_G/\hbar\omega_p \sim 400$ for $n_o = 10^{17}$ cm^{-3} It appears more plausible that holes are created by interband excitation as the energy loss step of the tunneling electrons. When electrons tunnel from the interface to the conduction band, some will have sufficient kinetic energy to create an electron-hole pair by interband excitation. For $V_R > E_G + V_B = 4.0$ eV all of the tunneling electrons have sufficient energy to allow this energy loss process. For intermediate voltages such that $E_G < V_R < E_G + V_B$, only thermally activated electrons have sufficient kinetic energy. The efficiency of interband excitation of electron-hole pairs was reported by Ludwig and Kingsley.[34] The quantum efficiency of pair production and subsequent luminescence by photon excitation was not reported for ZnO, because the results were found to be sensitive to surface preparations. The same authors also measured the production of electron-hole pairs in ZnO from low energy electron beam excitation and found a luminescence efficiency smaller than that due to photon excitation. In summary, a finite fraction (perhaps 10 to 30%) of low energy electrons will make an electron-hole pair. This provides an adequate rate of hole production to explain the buildup of the hole charge necessary for the present theory. A less certain feature is the mechanism by which the holes decay. Several mechanisms are possible and are under study.

The onset of hole production starts at $V_R = E_G$. Electrons which tunnel elastically from the interface states have a final kinetic energy of $E_f = E_G - V_B$ which is insufficient to excite interband transitions. But those electrons which are thermally excited over the barrier then enter the ZnO grain with kinetic energy $E_f = E_G$ which is at the threshold for being able to excite holes by interband transitions. For higher applied voltages, where $V_R > E_G$, some of the thermally excited electrons have $E_f > E_G$ and thus can excite interband excitations and make holes. This is rather a slow process for bias voltages which have $V_R \simeq E_G$, which is why the capacitance should not show its negative spike in this region.

Supporting evidence for this explanation of the hole creation process can also be found from the so-called "overshoot" phenomenon. As previously mentioned, in typical varistors breakdown occurs in less than a nanosecond.[1] However, when a fast constant current step I is applied to a varistor, a time dependent voltage response, V(t), is observed.[35, 36] This effect, commonly referred to as a voltage overshoot, is most marked in the first few nanoseconds of the applied pulse. The effect becomes negligible on a microsecond time scale. We interpret this time dependent phenomenon as arising from the time associated with hole creation.

IV. DISCUSSION

Any theory of conduction in metal oxide varistors must be consistent with the experimental observations listed in the Introduction.[1-5] We shall compare the predictions of the model of Section III with measured data, bearing in mind that the model is derived for a single grain-grain junction whereas many (but not all - cf. Fig. 13) measurements on ZnO varistors evaluate average properties, i.e., the experimental results are a complex average of numerous parallel-series paths traversing many junctions in the device under study.

The model of Section III contains only one unknown parameter, namely the depletion layer barrier height V_B. This parameter is easily evaluated from the activation energy of the very low voltage linear varistor

leakage characteristic. All other constants used are either known ZnO material parameters or universal physical constants. Nevertheless we predict the varistor breakdown voltage V_b (by definition the voltage at a current of 1 ma/cm^2) to be 3.2 volts per junction at T ~ 300°K, $n_o \sim 10^{17}$ electrons/cm^{-3} (Fig. 7a). In addition, wide variations in temperature T or ZnO donor density, n_o, produce little change in V_b (Figs. 7 and 9).

To compare our theoretical predictions with experiment requires measurements on a single ZnO varistor grain junction. Such data were obtained by carefully evaporating a pair of opposing Al electrodes having 10 μ separation on the surface of a varistor disc. The varistor grain size was about 25 μ and the electrodes lay on either side of a single ZnO grain junction. A photomicrograph of the experimental configuration is given in Fig. 12.

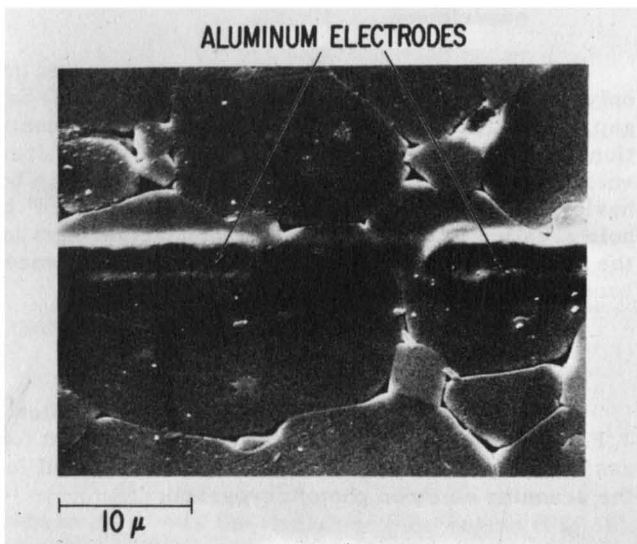


Fig. 12 Photomicrograph of evaporated Al electrode configuration on the surface of a ZnO varistor. The electrodes have ohmic contact to two ZnO grains separated by a single grain barrier. The measured current-voltage curve is given in Fig. 13.

In Fig. 13 we plot the room temperature current-voltage characteristic measured experimentally on a single ZnO varistor junction of the type shown in Fig. 12. Assuming that the area of grain-grain contact ≈10 μ x 10 μ, a current of 10^{-9} ampere corresponds to a current density of 10^{-3} ampere/cm^2. Thus from Fig. 13, we obtain the experimental value of $V_b \approx 3.3$ volts per grain junction. For the sample of Fig. 13, $n_o = 10^{17}$ electrons/cm^3, [8] and we have also plotted the appropriate calculated curve. The agreement between our model and the experimental single junction measurements is remarkably good except at very high current densities where the experimental voltage rises above the theoretical curve due to the series resistance of the ZnO grains. [1] This effect has not been included in our model.

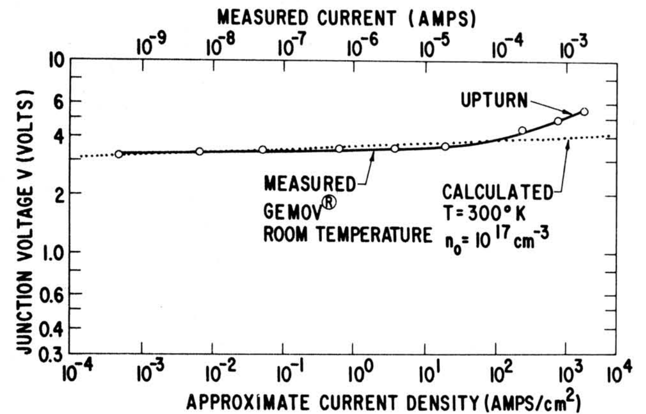


Fig. 13 Experimental and calculated current-voltage characteristic of a single ZnO varistor grain junction. The observed decrease in nonlinearity at high current densities is due to the series resistance of the ZnO grains. (Ref. 1) This has not been included in our model.

We have also repeated the single grain measurements on a variety of other varistor types and compositions (some without bismuth) and have found breakdown voltages close to that of Fig. 13. Van Kemenade and Eijnthoven[19] have made similar single junction studies. They find $V_b \approx 3.5$ volts at 10^{-6} ampere which we presume corresponds to a current density of about 1 ampere/cm^2. From Fig. 7b we predict $V_b = 3.6$ volts at J = 1 ampere/cm^2. Again the agreement is remarkably good and both the model and the experimental data exhibit similar values of V_b for wide variations in the varistor composition and processing.

We may also compare the predicted and experimentally measured temperature dependence of the varistor breakdown voltage, $1/V_b\,(dV_b/dT)_I$. This quantity is usually measured at a current density of 10^{-3} ampere/cm^2. From Fig. 7b the predicted value is $1/V_b\,(dV_b/dT) = 8 \times 10^{-4}\,{}^{\circ}K^{-1}$, between 150°K and 300°K in good agreement with experiment. [4, 5, 15]

In Fig. 8 we give the predicted values of the nonlinear experiment α corresponding to the curves of Fig. 7. Values of α well in excess of 50 are found with the higher values of α at the lower temperatures. This is confirmed experimentally. [15] Note also that the model predicts that the peak in α moves to lower currents at lower temperatures. This is also corroborated experimentally. [15]

In Figs. 9a and 9b we present plots of current density J vs. applied junction voltage for the indicated values of donor density in the ZnO grains. The point to be emphasized is that increased ZnO carrier concentration n_o leads to increased varistor leakage but does not appreciably affect the varistor breakdown. This is to be contrasted with the model of Bernasconi, et al. [11, 13] and Knecht and Klein[25] which predicts a significant dependence of V_b upon n_o.

We have also measured the variation of varistor capacitance with applied DC bias. Figure 14 gives C(V)/C(0) for a varistor with $V_b \approx 230$ volts. The measured capacitance drops by about 30% until the varistor breakdown region is reached, whereupon a sharp increase in C is observed. Values of C at higher voltages could not be obtained due to the presence of large DC currents in the device. It is interesting to note that the measured behavior is quite similar to that predicted in Fig. 11. To our knowledge, no other model of varistor behavior predicts this sharp rise in capacitance in the breakdown region.

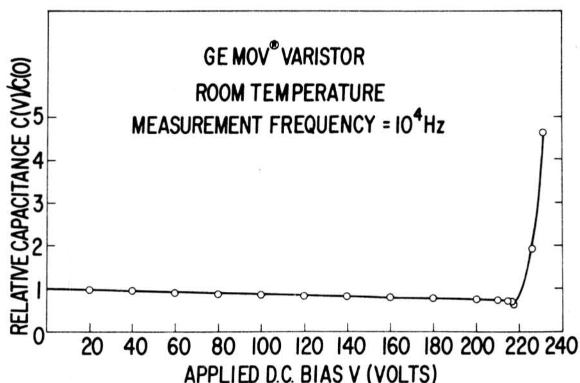


Fig. 14 Experimentally measured variation of the varistor capacitance with applied DC bias voltage. Note the steep rise for voltages in the varistor breakdown region.

V. SUMMARY

A theory is presented which quantitatively accounts for the important features of conduction in ZnO-based metal oxide varistors. It predicts:

- Coefficients of nonlinearity α as high as 50 or even 100
- Breakdown voltage insensitive to temperature and electron concentration in the ZnO grains
- Varistor leakage currents which increase with temperature and electron concentration in the ZnO grains
- Breakdown voltage at 1 ma/cm^2 of 3 to 3.5 volts/grain which is insensitive to the metal oxide additives and the firing and processing procedures
- Capacitance which initially decreases with increasing applied DC bias and then sharply rises as the breakdown voltage is approached
- A prebreakdown activation energy E_A having a voltage dependence in good agreement with experiment

The theory has no adjustable parameters and uses only known material parameters such as the ZnO band gap, measurable extrinsic ZnO grain donor concentration and varistor low voltage leakage current activation energy. The highly nonlinear varistor conduction behavior results from electron tunneling "triggered" by hole creation in the ZnO when the conduction band in the grain interior drops below the top of the valence band at the grain interface.

ACKNOWLEDGEMENTS

The authors wish to express their appreciation to J. P. Cocoma, L. J. Petrucco, and P. L. Sargent for assistance with the experiments and to M. E. Gill for the scanning electron photomicrograph.

These appendices provide details of the calculations of Section III.

A. PREBREAKDOWN

Here the potential energy U(x) for the band bending is $U(x) = n_o \gamma (x_R - x)^2$, where x_R is chosen to give $V_R = n_o \gamma x_R^2$ and $\gamma = 2\pi e^2/\epsilon_o$. The WKBJ integral for tunneling from the interface at an energy E below the top of the interface, as in Fig. 5, is

$$W = 2\int k(x)\,dx = 2\left(\frac{2m}{\hbar^2}\right)^{1/2} \int dx\,[\gamma n_o (x_R - x)^2 + E - V_R]^{1/2} . \quad (A1)$$

By changing integration variables to $y = (\gamma n_o)^{1/2}(x_R - x)$, the integral may be expressed as

$$W = \frac{4}{\hbar\omega_p} \int_{\sqrt{V_R - E}}^{\sqrt{V_R}} dy \sqrt{y^2 + E - V_R} \quad (A2)$$

where $\omega_p^2 = 4\pi n_o e^2/\epsilon_o m$.

This is a standard integral which gives Eq. (31).

B. HOLE CREATION

The potential U(x) has a non-parabolic shape in the region of the holes. The potential is

$$U(x) = E_G + f(x) \qquad 0 \le x \le x_H$$
$$U(x) = \gamma n_o (x_R - x)^2 \qquad x_H \le x \le x_R \quad (B1)$$

where x_H marks the end of the hole region (Fig. 5).

At $x = x_H$, $f = 0$, i.e.,

$$U(x_H) = E_G . \quad (B2)$$

The value of x_H is found by solving Eq. (43).

At $x = 0$, $U(x) = V_R$, i.e.,

$$f(0) = V_R - E_G . \quad (B3)$$

At the end of the hole region, U(x) is continuous and has a continuous derivative. From Eq. (B1),

$$U(x_H) = E_G = \gamma n_o (x_R - x_H)^2 , \quad (B4)$$

$$\left(\frac{df}{dx}\right)_{x_H} = 2\gamma n_o (x_R - x_H) , \quad (B5)$$

$$= 2(\gamma n_o E_G)^{1/2} , \quad (B6)$$

using Eq. (B4). From Eq. (B6) we obtain the constant of integration C in Eq. (45). This follows since at x_H, where $f = 0$, we have from Eq. (43)

$$\left(\frac{df}{dx}\right)_{x_H}^2 = 4\gamma n_o C , \quad (B7)$$

so that $C = E_G$. The value of x_H is found from Eq. (46) at $x = 0$, where $f(0) = V_R - E_G$. Thus

$$x_H = \frac{1}{2\sqrt{\gamma n_o}} \int_0^{V_R - E_G} dy\,[E_G + y + by^{5/2}]^{-1/2} . \quad (B8)$$

We then have from Eq. (B5)

$$x_R = x_H + (E_G/\gamma n_o)^{1/2} . \quad (B9)$$

This determines X_H and X_R, so the potential is completely specified.

C. TUNNELING BARRIERS WITH HOLE SPACE CHARGE

The WKBJ tunneling integral for a particle of energy E depends upon whether the tunneling barrier is confined to the hole region. This happens when $E < V_R - E_G$. The tunneling integral is

$$W = 2\left(\frac{2m}{\hbar^2}\right)^{1/2} \int dx \sqrt{U(x) + E - V_R} \quad (C1)$$

where

$$U(x) = E_G + f(x) . \quad (C2)$$

We evaluate W by changing the variable of integration to f(x) and using Eq. (43) for (df/dx). Thus

$$W = 2\left(\frac{2m}{\hbar^2}\right)^{1/2} \int \frac{df}{\left(\frac{df}{dx}\right)} \sqrt{f + E - V_R + E_G} \quad (C3)$$

and substituting Eq. (43) gives

$$W = \frac{2}{\hbar\omega_p} \int_{V_R - E_G - E}^{V_R - E_G} df \left[\frac{f + E - V_R + E_G}{E_G + f + bf^{5/2}}\right]^{1/2} . \quad (C4)$$

Equation (C4) becomes Eq. (52) with the change of variables $f = V_R - E_G - y$.

The other possibility is $E > V_R - E_G$. Here the tunneling is partly through the hole region, and partly in the donor depletion region. In the region of holes,

the WKBJ integral is similar to Eq. (C4), except now the integration limits extend to the edge of the region at $f = 0$, i.e.,

$$W_1 = \frac{2}{\hbar\omega_p} \int_0^{V_R - E_G} df \left[\frac{f + E - V_R + E_G}{E_G + f + bf^{5/2}} \right]^{1/2} . \qquad \text{(C5)}$$

Equation (C5) is identical to Eq. (53) since $\mu_H = V_R - E_G$. The other contribution is the part of the tunneling trajectory through the donor depletion region

$$W_2 = 2\left(\frac{2m}{\hbar^2}\right)^{1/2} \int_{x_H}^{x_E} dx \left[\gamma n_o (x_R - x)^2 + E - V_R \right]^{1/2} . \qquad \text{(C6)}$$

This is evaluated by changing variables to $y = (x_R - x)\sqrt{\gamma n_o}$, to give

$$W_2 = \frac{2}{\hbar\omega_p} \left\{ \sqrt{E_G (E + E_G - V_R)} - (V_R - E) \ln \left(\frac{\sqrt{E_G} + \sqrt{E + E_G - V_R}}{\sqrt{V_R - E}} \right) \right\} \qquad \text{(C7)}$$

which is Eq. (54).

REFERENCES

1. L. M. Levinson and H. R. Philipp, IEEE Trans. PHP-13, 338 (1977).
2. M. Matsuoka, Jpn. J. Appl. Phys. 10, 763 (1971).
3. W. B. Morris, J. Am. Ceram. Soc. 56, 360 (1973).
4. L. M. Levinson and H. R. Philipp, Appl. Phys. Lett. 24, 75 (1974).
5. L. M. Levinson and H. R. Philipp, J. Appl. Phys. 46, 1332 (1975).
6. L. M. Levinson and H. R. Philipp, J. Appl. Phys. 47, 1117 (1976).
7. H. R. Philipp and L. M. Levinson, J. Appl. Phys. 46, 3206 (1975).
8. H. R. Philipp and L. M. Levinson, J. Appl. Phys. 47, 1112 (1976).
9. J. Wong, J. Appl. Phys. 46, 653 (1975).
10. J. D. Levine, Crit. Rev. Solid State Sci. 5, 597 (1975).
11. J. Bernasconi, J. P. Klein, B. Knecht, and S. Strassler, J. Electron Mater. 5, 473 (1976).
12. W. G. Morris, J. Vac. Sci. Technol. 13, 926 (1976).
13. J. Bernasconi, S. Strassler, B. Knecht, H. P. Klein, and A. Menth, Solid State Commun. 21, 867 (1977).
14. P. R. Emtage, J. Appl. Phys. 48, 4372 (1977).
15. H. R. Philipp and L. M. Levinson, J. Appl. Phys. 48, 1621 (1977).
16. L. F. Lou, Extended Abstracts, 153rd Electrochemistry Society Meeting, Vol. 78-1, 308 (1978).
17. L. M. Levinson and H. R. Philipp, unpublished.
18. H. R. Philipp and L. M. Levinson, J. Appl. Phys. 47, 3177 (1976).
19. J. T. C. Van Kemenade and R. K. Eijnthoven, Ber. Dtsch. Keram. Ges. (in press).
20. K. Mukae, K. Tsuda, and I. Nagasawa, Jpn. J. Appl. Phys. 16, 1361 (1977).
21. A. Goetzberger, B. McDonald, R. Haitz, and R. Scarlett, J. Appl. Phys. 34, 1591 (1963); C. Lee, R. Logan, R. Batdorf, J. Kleimack, and W. Wiegmann, Phys. Rev. 134, A761 (1964).
22. K. G. McKay, Phys. Rev. 94, 877 (1954); G. A. Baraff, Phys. Rev. 128, 2507 (1962).
23. S. M. Sze, Physics of Semiconductor Devices (Wiley-Interscience, New York, 1969); R. H. Tredgold, Space Charge Conduction in Solids (Elsevier, New York, 1966); J. G. Simmons, J. Phys. D4, 613 (1971).
24. J. G. Simmons, J. Appl. Phys. 34, 1793 (1963).
25. B. Knecht and H. P. Klein, Ber. Dtsch. Keram. Ges. (in press).
26. L. M. Levinson and H. R. Philipp, J. Appl. Phys. 47, 3116 (1976).
27. For a preliminary account of this theory, see G. D. Mahan, L. M. Levinson and H. R. Philipp, Appl. Phys. Lett. (1978).
28. D. R. Clarke, J Appl. Phys. 48, 4372 (1977).
29. D. Eger, Y. Goldstein, and A. Many, RCA Review 36, 598 (1975).
30. It should be emphasized that evaluations of the breakdown voltage per grain barrier are best taken from single grain measurements.[18] Values deduced from measurements of average grain size and macroscopic breakdown field are generally too low. This arises since breakdown in this highly nonlinear device occurs along the path having the fewest grains.
31. D. G. Thomas, J. Phys. Chem. Solids 15, 86 (1960).
32. J. W. Conley and G. D. Mahan, Phys. Rev. 161, 681 (1967).
33. P. N. Butcher, J. A. Hulburt and K. F. Hulme, J. Phys. Chem. Solids 21, 320 (1961).
34. G. W. Ludwig and J. D. Kingsley, J. Electrochem. Soc. 117, 348, 353 (1970).
35. E. C. Sakshaug, J. S. Kresge, and S. A. Miske, IEEE Trans., Power Apparatus Syst. PAS-96, 647 (1977).
36. Transient Voltage Suppression Manual (D. C. Kay, ed.), General Electric Co., Semiconductor Products Dept., Syracuse, N. Y. (1976).

H78CRD205
CAMPBELL VC 6065102090
GENERAL ELECTRIC CO.
TB DEPT 6-183 EP
SYRACUSE NY 13221

G.D. Mahan
L.M. Levinson
H.R. Philipp

THEORY OF CONDUCTION
IN ZnO VARISTORS

Report No. 78CRD205
November 1978

GENERAL ELECTRIC COMPANY
CORPORATE RESEARCH AND DEVELOPMENT
P.O. BOX 8, SCHENECTADY, N.Y. 12301