CHAPTER I

A REVIEW OF THE BREAKDOWN MECHANISMS
1.1 Introduction

Most of the electrically insulating materials can sustain a fairly high voltage without harm to themselves. If the voltage applied to the materials is progressively increased, the leakage or capacitive current also increased accordingly. This rise becomes faster as one approaches the voltage 'Vm' (Fig. 1.1.) when a point 'A' is reached at which the slope \( \frac{dI}{dV} \) of the curve tends to become infinity. This condition may be called the on-set of breakdown of a material. It may be defined as that value of the electric intensity which is just sufficient to cause the material to breakdown or lose its insulating properties temporarily or permanently.

However, if the current through the solid material is allowed to increase above the point 'A' there is a drop in the voltage across it. This drop 'B' is explained by a decrease in the insulation resistance. A strongly conductive breakdown channel is formed during the process. The maximum voltage 'Vm' applied to the material at the moment of breakdown is known as the breakdown voltage. This breakdown is similar to the well known phenomenon of breakdown in gases and liquids. However, there is one important difference, viz., in solids when the breakdown voltage is removed, a trace of the earlier breakdown in the form of a punctured, fused
Fig. 1.1 - Current through insulator versus voltage.

F.L. = Fermi level; C = Conduction band; V = Valence band

Fig. 1.2 - The electron energies are plotted against distance through the materials.
hole or a track of an irregular shape is left over. These tracks (holes) provide the channels for future breakdown. No such previous history of breakdown is left over in case of gases or liquids, because of high mobility of ionized particles, provided the duration of breakdown and power associated is very low.

The studies of breakdown in materials could be broadly classified into three divisions, viz.,

(a) materials are in bulk form e.g. crystal plates, pellet or slab whose thickness is of the order of mm.,

(b) materials are taken in the form of thick films whose thickness is of the order of a few microns, and

(c) materials are evaporated to deposit in the form of thin films under vacuum technique. The thickness of the films is taken of the order of a few hundred Angstrom unit.

The studies of Whitehead,² Von Hippel,³ Frohlich ⁴ showed that the breakdown in bulk materials is generally of destructive type, resulting in burning or fusing of the specimen which cannot be used again, whereas Riehl, Baessler, Hunklinger, Spannring and Vaubel⁵ (1959) developed a technique on non-destructive breakdown in film materials. A large number of investigating experiments can be carried out on one sample, as
the follow up current does not destroy the evidence of a single breakdown event. Dean also observed in 1910 that the breakdown was initiated at certain region called spots. Klein (1969), Klein and his coworkers (1965) have investigated the mechanism of the breakdown process on self healing specimen. They have suggested about the formation of a conducting channel in the insulator followed by the discharge of electrostatic energy causing destruction mainly by evaporation.

1.2 Survey of the earlier concepts of electric breakdown

The earlier observations of electric breakdown of dielectric solid materials showed that it was dependent on the nature of -

(1) electrodes (geometry and materials),
(2) ambient medium,
(3) temperature,
(4) thickness, and
(5) nature and form of applied voltage (uniform, non-uniform, impulse, a.c. and d.c. voltages).

The dissipation of heat, accompanying the breakdown caused indefiniteness in the determination of breakdown strength of a material, therefore, it was thought that the non-thermal breakdown would be homogeneous. Semenoff and Walther (1928) proposed that the proportional variation of electric strength with inverse square root of thickness was due to electric stress
at the edges of electrodes. When due care was taken to avoid
discharges in the ambient medium between the electrodes to the
dielectric and using pulsed voltages to avoid thermal effect
then it was found that the electric strength was independent of
thickness, nature of electrodes and form of applied voltages.
This gave rise to the concept of intrinsic electric strength,
dependent only on the physical state of the dielectric material
but independent of external conditions. The medium in contact
with the dielectric should have greater conductivity and electric
strength than that of the material. Von Hippel \(^{10}\) (1931) used
nitrogen at high pressure as the ambient medium.

Rogowski \(^{11}\) calculated the electric field necessary to
cause distortion in the lattice structure of a simple cubic ionic
crystal, where ions are held in equilibrium by inverse square
Coulomb forces and Born repulsion forces. These values were
generally greater by a factor of 100.

Horowitz \(^{12}\) used the principle that breakdown begins
at microscopic cracks or scratches in a crystal and thus
calculated the electric field at which the crack becomes unstable
( prevented from spreading by surface energy similar to surface
tension ).

Suggestions were also made about the presence of
mechanical stress arising from the electric field, finally leading
to electric breakdown. This was successfully applied to mechanically weak or soft dielectrics. Rogowski 13 and Whitehead 14 made a suggestion that, if the electric field causes a relative displacement of ions in a polar structure, or a molecular distortion in the dielectric, the thermal motion super-imposed might give an effective displacement sufficient to lead to disintegration as in the phenomenon of melting. Ganger 15 derived an expression for the electric field required to cause critical displacement leading to breakdown. The values thus obtained were too high. Attempts were also made by Vorobiev 16 and Boning 17 to correlate chemical composition and the ion theory with electric breakdown. All these theories were elementary in nature, merely, trying to explain some observed facts.

Rogowski extended the well-known phenomenon of ionization by collision in gases to the problem of breakdown in solids. It yielded a field dependent conductivity which made the current unstable at a certain field strength. Whitehead and Nethercot 18 considered the self induced ionization by ions, leading to instability due to same factor which was suggested by Joffe 19 to be the ability of the dielectric to sustain a critical current density. Joffe supposed a characteristic ionic binding energy and a fractional dissipation of energy by a fast moving ion along its path. On this basis he obtained an expression for
the breakdown voltage, whose variation with thickness and temperature, as predicted by theory, was not in accordance with the experimental observations.

Zener,\textsuperscript{20} using the equation derived by Fowler and Nordheim for the escape of electrons from a potential barrier obtained an expression for the rate of escape of electrons under high field $10^6$ to $10^3$ V/cm. Although this theory may explain pre-breakdown currents but does not explain the mechanism of breakdown. Austen\textsuperscript{21} used the idea of ion escape by an activation process to explain the increased conductivity at very high fields.

1.3 Band theory model for solids

The breakdown phenomenon and other properties of the solids could be understood in terms of the well known band theory model. The solids are made up of regular and periodic arrangement of a pattern in three dimensions. Many physical properties of solids like the resistivity, conductivity, mechanical stress etc. could be explained on the basis of periodicity of the lattice field of the crystals. The motion of electrons under periodic potential field inside a crystal has been discussed by Bloch who came to the conclusions that :

(i) the energy spectrum of the electrons consists of a number of allowed energy bands separated by forbidden regions, and

(ii) the width of the allowed energy bands increases with increase of energy values.
The quantum mechanical solution give rise to a set of Brillouin zone with energy gap known as forbidden zone or band. The electron remaining in one zone or band cannot go over to another zone by a continuous application of force i.e. by trying to gradually increase its energy. The two consecutive Brillouin zones are separated by a forbidden gap. Unless the electron in first zone absorbs an energy equal to that of the forbidden gap in a single encounter, the electron cannot go over to the next Brillouin zone, however long the duration of the applied force may be. In fact the width of the forbidden gap is going to decide whether a solid is a conductor or an insulator as shown in Fig. 1.2. This illustrates a picture of a dielectric in which a series of allowed electronic energy bands, are completely occupied by electrons up to a certain level, and empty thereafter. The second band i.e. the conduction band does not give rise to any conductivity since it contains no electrons. The valence band (1st band) also does not conduct, since there are no unoccupied states into which an electron can be accelerated by an applied field. Such a material would be perfect crystalline insulator at absolute zero of temperature.

When the temperature is different from zero, there will be a few electrons in the second band and a few unoccupied places in the first band, therefore, some conduction will be possible. As the temperature rises, the number of electrons giving rise to
Fig. 13 Schematic energy level diagram for an amorphous semiconductor.

Fig. 14 The horizontal lines represent localized centres and the circles are electrons in them.
conduction increases and hence the conductivity tends to increase. The excitation of thermal vibration of the lattice, however, will tend to decrease it. At low temperature the transfer of electrons into the conduction band must predominate and hence the rapid rise of conductivity is observed with temperature for insulating materials. At higher temperature the effect of thermal vibration of the lattice may predominate over the normal transfer process of electrons. The temperature coefficient of resistance may change sign, the effect being observed for semiconductors.

In general, only the conduction and valence bands are important in pure strain free crystals at low temperature. However, the admixture of foreign ions, mechanical strain, and defects of the crystals will provide isolated electronic energy levels called electron traps in the forbidden band. The number of such levels in this band depends upon the degree of the imperfections of the crystals. The relative importance of isolated levels increases with rise in temperature. Frohlich 22(1947 a) proposed a model Fig.1.3 that can be used to calculate the temperature dependence of trap controlled band conduction. The picture representing the localized levels below the 'C' levels (conduction levels), is assumed to be of two types: the shallow traps i.e. 'S' levels, which extend over a range of energy $\Delta V$ below the non-localized levels, and deep traps i.e. 'D' levels, which lie below the 'C' levels by an amount $2W \gg \Delta V$. The electrons in C, S, and D levels are assumed to be sufficiently numerous to establish thermal equilibrium
amongst themselves at some temperature $T_e$, which may be different from the lattice temperature $T$. The validity of the inequalities required for above case is,

$$k_0 T_e \ll \Delta V \ll W \quad (1.1)$$

For impurity conduction in semiconductors, there is an impurity band conduction model which has been reviewed by Mott and Twose $^{23}$ (1961), Mott $^{24}$ (1967), and Klinger $^{25}$ (1968). This field of conduction in amorphous semiconductors has been reviewed by Cohen $^{26}$ (1971). The basic idea is that electrons occupying isolated donor levels (or holes occupying acceptor levels) may contribute to the electrical conductivity in a manner other than the usual process of thermal excitation followed by band conduction. The electron occupying the isolated donor is described by a localized wave function, which will overlap by a small but finite amount with neighbouring donors. Charge transport can occur by tunnelling to an adjacent vacant donor level and this has been called impurity band conduction. The character of the electronic band scheme leading to impurity conduction is exhibited in Fig. 1.4. In an intrinsic semiconductor, charge carriers are produced by thermal excitation of electrons from the valence band, across the energy gap $E_g$. In an extrinsic semiconductor the presence of certain impurities (donors) may provide electrons in levels that are close enough to the conduction band for the electrons to be excited to the conduction band at relatively low temperature.
Alternatively, other impurities (acceptors) may provide vacant levels, close to the valence band, into which electrons from it may be raised.

The vacant sites in valence band behave like free electrons with a positive charge and are known as 'positive holes'. Semiconductors are called n type when the conduction of electricity is primarily due to electrons and p type when conduction is mainly due to positive holes.

The characteristic features of impurity conduction are as follows.

(1) The centres at which the electron wave function can be localized are randomly distributed, even though their average density may be uniform.

(2) There must be an adjacent vacant centre to which an electron can move.

If all of the localized levels for electrons and holes are due to charged centres (donors and acceptors), then impurity conduction can take place only if there is compensation i.e. the presence of some minority centres. The compensation is based on the notion that trapping centres for electrons and holes are necessarily charged, so that an n type material containing no minority centres can have no vacant trapping centres for electrons.
This may be true for crystalline semiconductors but for dielectrics trapping may occur at centres that are uncharged. This may account for the behaviour of amorphous dielectrics. The above models may help to investigate the high field electrical conduction as well as the breakdown mechanism to be considered (vide supra).

1.4 High field conduction

While the above mentioned picture may help in considering the low field electrical conductivity exhibited by semiconductors as well as insulators, it may also help in explaining the electrical breakdown. It is a continuing process from low field conductivity to the high field conductivity which finally leads to breakdown. The problem of high field conduction could be considered as a combination of (i) electrode effects, (ii) bulk effects, and (iii) both the electrode and bulk effects combined.

1.4.1 Fowler-Nordheim electrode emission

The electron emission from the electrode into the dielectric may be explained on the basis of Fowler-Nordheim field emission mechanism, using a potential barrier model as modified for the dielectric. Good and Muller (1956) had
discussed this problem quantum mechanically and obtained an expression for field emission at temperatures other than zero.

Emtage 29 (1957) has investigated the effect of interaction between longitudinal optical phonons in a polar dielectric and electrons tunnelling into it. The results show that even in zero-temperature approximation a low field Schottky-type (vide supra) characteristic is continuous with a high field Fowler-Nordheim characteristic.

The presence of traps in the insulator may have a marked effect on the emission current. A theory to account for enhanced emission due to the presence of empty trapping centres, has been given by Penley 30 (1962) and by Gadzuk 31 (1970 a,b). The transmission coefficient for an electron wave, incident on the potential barrier for the insulator having impurity (foreign material) traps, is very much enhanced due to resonance tunnelling.

(1.6) Tunnelling through thin insulating films

If an insulating layer is so thin that an electron may tunnel directly from cathode to anode without at any time occupying the conduction band of the dielectric. The film thickness for which this will occur depends on both the work function and the applied voltage. Generally a thickness of 30 A or less involve, direct tunnelling between the electrodes for typical
experimental conditions. Frenkel \(^{32}(1930)\) explained electron tunnelling through a thin insulating film by using a simple rectangular potential barrier model. Sommerfeld and Beth\(^{33}(1933)\), Holm \(^{34}(1951)\), and Simmons \(^{35}(1963\ a,b\ ;1964\ )\), Statton \(^{36}(1962)\) have modified the earlier picture of Frenkel.

(1.c) Schottky emission

Schottky emission is essentially thermionic emission from a metal electrode into the conduction band of a dielectric taking into consideration the image force correction which lowers the potential barrier by a small amount. More recently Antula \(^{37}(1971)\) has found Schottky emission at even room temperature in films \(\approx 100\ \AA\) thickness, both pure \(\text{Al}_2\text{O}_3\) and doped with positive metallic ions.

(2) High field ionic conduction (Bulk effects)

A conduction process may be regarded as a bulk process involving the field dependence of the conductivity. It is assumed that the flow of charge carriers under uniform field is determined only by the properties of the dielectric. In this connection it is pointed out that a process is not regarded as a bulk process if the electrode emission changes the charge density within the dielectric. Hanscomb, Kao Calderwood, O'Dwyer, and
Emtage (1966) found that the conductivity of NaCl and KCl was appreciably non-ohmic even for field less than 1 MV cm$^{-1}$ at room temperature. The field dependent ionic mobility was therefore rejected as a possible explanation of the non-ohmic behaviour of the NaCl and KCl. However, Ieda, Kosaki and Sugiyama (1970) have investigated that the field dependence of the mobility is too small to account for the observation of non-ohmic conduction.

Just as for ionic conduction, the effect of high field on the bulk electronic conductivity of insulators may be either to change the mean mobility per carrier or to increase the carrier density. The field dependence of the mean mobility in semiconductors has been extensively reviewed by Conwell (1967) and applied particularly to n type germanium. The increase of conductivity may occur due to charge carriers multiplication, which is due to ionization either from traps or from the valence band. The ionization from the valence band is regarded as a prelude to electrical breakdown.

(3) Combined electrode and bulk effects

The current flow through dielectric between electrodes at high field was investigated by Lampert and Mark (1970). The current-injecting properties of the electrodes play the role of ohmic and blocking contacts in the process. The
electric field in the insulator is assumed to move the injected charge with a velocity determined by its mobility. At the ohmic contact additional carriers are added by the injection of electrons. The general current-voltage characteristics are predicted by Lampert's theory.

Blocking contacts are specified by conditions in which the total injected current of the carriers is zero. However, there may be a non-zero steady state current, if there is collision ionization within the insulator. The electron current is then ejected at the anode and hole current at the cathode. The double ejection was investigated by O'Dwyer \(42(1968)\) who has further developed the breakdown mechanism.

1.5 Thermal breakdown

When an electric field is applied to a solid dielectric there are always some currents which generate heat. All cases of solid dielectric breakdown ultimately involve thermal processes. The type of breakdown that can be adequately described in terms of the thermal properties of the dielectric and the pre-breakdown electrical conductivity may be termed as thermal breakdown.

A dielectric to which voltage is applied liberates heat, the temperature of the dielectric rises and the dielectric losses are therefore increased still more. The progress is intensified
until the dielectric is heated so much that it gets damaged (fused, burnt, pierced with cracks etc. depending on the nature of a given material and the conditions in which it is placed). The breakdown may occur at a specific temperature even at low applied field at which it would never develop at a lower temperature in an undamaged condition of the material. It is not necessary that the entire volume of the dielectric be heated for a breakdown to take place, it is enough to heat some portion of a dielectric in which Joule's heating is intensified in view of its inhomogeneity. Though the mean temperature at the entire volume of the dielectric may differ slightly from the original temperature.

In thermal breakdown the main processes are the Joule heat generated by the current flow and the conduction of this heat away to the surroundings. If thermal conduction is the only significant heat-loss process, the relation describing the lattice energy balance is given by

\[
\frac{dT}{dt} = - \text{div} (k \text{ grad } T) = -F^2
\]

(1.2)

where \( Cv \) is the specific heat per unit volume, \( T \) and \( k \) are electrical and thermal conductivity respectively and given by equations:

\[
\tau = \tau(F, T)
\]

(1.3)

\[
k = k(T)
\]

(1.4)
F the field, \( \frac{dT}{dt} \) is the time derivative, and \( \nabla T \) the space gradient of the temperature.

The minimum thermal critical field \( F_m \) corresponds to the condition when the applied field causes breakdown after a very long time (almost a steady state for the lattice processes). Thus \( \frac{dT}{dt} \) becomes zero, and hence, the thermal breakdown is given by:

\[
- \nabla (k \nabla T) = -F^2
\]

(1.5)

The second case occurs when the field is applied as a short pulse; it may then be a satisfactory approximation to ignore the heat conduction term in Equ. (1.2) and obtain the equation

\[
C_v \frac{dT}{dt} = -F^2
\]

(1.6)

A third approach can be regarded as a form of integrated approximation to Equ. (1.5). For a thin slab or film it is assumed that the energy dissipation results in a constant temperature in the dielectric, which is different from the temperature of the surroundings. The rate of energy loss to the surroundings may be written as:

\[
\Lambda (T - T_0) = IV
\]

(1.7)

where \( I \) is the current, \( V \) the applied voltage, and \( \Lambda \) a constant for the particular test set-up.
Thermal breakdown was observed in thick single crystal plates of NaCl by Inge, Semenoff and Walther (1925). Using a.c. voltages of slowly increasing amplitude, they found evidence of thermal breakdown in NaCl at a temperature greater than about 220°C. Essentially the same result was found by Whitehead and Nethercot (1935) (cf. Whitehead, 1951) and by Klein and Gafni (1966) (cf. Klein, 1969). Later on, the work is extended for thermal breakdown in thin films with field dependent conductivity by them. Hanscomb (1969) also investigated the thermal breakdown of NaCl and KCl at temperatures between 200°C and 320°C using flat-topped pulsed voltage. The relationship between critical field strength and time to breakdown was compared with theoretical results which gave satisfactory agreement with experimental results.

The thermal breakdown has been found to be effected by the following factors.

(1) high temperature,
(2) size and shape of the sample,
(3) thermal properties of the electrodes and their shape, and
(4) time of application of the field of breakdown strength being larger for voltage pulses of short duration.

1.6 Breakdown mechanism

Besides the thermal breakdown, the other type of breakdown, which is not connected in any way with Joule's heating
and associated with the pre-breakdown current is designated as purely electrical breakdown. The earlier theories of breakdown (Frohlich) had introduced the concept of intrinsic breakdown which was thought to be a physical constant of the material. Subsequent work showed that it was controlled by avalanche process which was affected by thickness and other parameters.

In order to account for electrical breakdown it is necessary to inquire how the electrons can escape from the filled zone to the conduction zone. If the forbidden gap is small, thermally excited transitions can occur. The thermal electrons will be extremely small in the case of dielectrics, although they are a characteristic feature of semiconductors which may become conductors at elevated temperature. The difficulty of thermal transitions is minimised by the existence of the intermediate imperfection levels as shown in Fig. 1.3. A series of stepping stones like S and D energy levels are available between the conduction and filled zones. These intermediate levels are rarer in more perfect crystals. They will be ineffective below a certain temperature, in facilitating escape from the filled zone. Electrons may be excited from lower levels by the absorption of radiation or when electrons or other particles with sufficient energy approach so closely as to make an inelastic collision. Zener described the escape of bound
electrons in a strong field on the Fowler-Nordheim mechanism. Though it is unsatisfactory for the theory of breakdown, it provides an additional mechanism for the appearance of free electrons.

In a dielectric there are always some free electrons, either by the effect of the distribution of thermal energy and intermediate imperfection energy levels at temperature above absolute zero, or by the effect of a strong electric field, or by absorption of radiation as in photo-conductivity. There may also be a number of loosely bound electrons which require only fairly small amount of energy to make transitions. If the applied field strength is low, the electrons and neutral molecules will be more or less in thermal equilibrium and the electrons would drift or diffuse in the direction of the field with a constant velocity, without gaining other than a negligible amount of energy from the field. When the field becomes stronger, the electrons are able to abstract greater amount of energy from the field. These electrons are likely to expel electrons from the lower zone of energy levels viz., ground state electrons. The electrons have a longer mean free path such that they have a fairly good chance of acquiring sufficient energy from the field for further ionization.

In analogy with Townsend theory of electrons in gases and
Fig. 1.5a - Sodium and Chlorine atoms are arranged in the unit cell.

Fig. 1.5b - Potential contours in principal directions.
ionization by collision, in which nonexciting collisions between electrons and neutral molecules corresponding to gases are replaced by interaction between electrons and lattice oscillations which is analogous to scattering or collision. Inelastic collision between electrons and molecules is replaced by the excitation of a ground state electron to a conduction or intermediate level by the close approach of a fast electron. The concept of mean free path and mean time between collisions or relaxation time can also be applied to solids. However the necessary applied field for the above mechanism will be much more in the case of solids. In a solid the electron moves as a wave though the strong and complicated fields of the massive ions of the structure so that it tends to move along certain preferred direction.

Von Hippel hypothesis

Von Hippel⁴⁶ in his theory pointed that the potential variation is the steepest along an edge (1 0 0) where alternate positive and negative ions are most closely spaced as shown in Fig. 1.5 a,b. The variation is less along the body diagonal (1 1 1) where the ions, still of alternate sign, are spaced \( \frac{1}{3} \) times the edge spacing, and the variations are much less along a face diagonal (1 1 0) where the ions are of the same sign and spaced at \( \frac{1}{2} \) times the (1 0 0) spacing. Hence Von Hippel
concluded the face diagonal to be a preferred direction and the body diagonal a secondary direction.

Von Hippel and his coworkers have examined the discharge paths with care in a number of experiments. According to them, the essential mechanism is directed by the creation of electron avalanches which are formed by impact ionization of the lattice atoms by electrons, accelerated by the applied field. Before the electrons can be accelerated they must gain more energy from the field than they lose in exciting or interacting with longitudinal polar modes of the lattice vibration. This interaction can be thought of as a barrier which is greatest when the electron has energy in the thermal range. Once an avalanche of sufficient size is formed, the part of material becomes molten due to the local high temperature involved. The avalanche trail forms a conducting path of plasma. The potential applied to the material is carried by the plasma formed along the avalanche channel. Thus the applied voltage is brought to the tip of the path which rapidly propagates into the dielectric medium. The fields surrounding the path are in-homogeneous so that the electrons will be accelerated towards or away from the path in all directions. The path will be oriented if the electrons experience less friction in some direction than in other. The ionization will occur first in the favourable direction and will thus extend the plasma path in preferred
directions. The above ideas have been treated analytically by Seeger and Teller and by Callen. The breakdown conditions were summarized by Von Hippel as follows.

(1) electrons move in the field direction and gain energy proportional to the field strength by direct acceleration,

(2) energy gain must exceed the loss for electrons of all energies.

Frohlich hypothesis

Frohlich assumed that the electrons do not always move in the field direction but are scattered by the lattice. They are thus gradually accelerated after scattering just as in ordinary conductivity. His theory also considers the rate of change of electronic energy on the average over many collisions. This rate must be positive for at least some of the electrons for breakdown. He considered all the circumstances in which a group of fast electrons having a mean given energy would tend to have a net gain or loss of energy in their average motion. If they gain energy, breakdown will occur provided an adequate supply of electrons with the appropriate energy could be visualized. The breakdown condition can be summarized as follows:

(1) electrons frequently change direction, the net energy gain is proportional to the square of the field strength as in conductivity.
(2) net energy gain must exceed loss for some energies.

O'Dwyer hypothesis

O'Dwyer $^{50}(1967,1969a)$ has suggested that the space charges (both electrons and holes) caused by the build-up of an electron avalanche should result in a non-uniform field strength distribution in the material. The assumption of an uniform field (as made in the previous theories) has been replaced by the consideration of hole current as well as the electron current in the material as postulated by O'Dwyer$^{42}(1968)$. When the space charge immediately in front of the cathode (due to relatively immobile holes) reaches such proportion that the electron current injected from the cathode is enhanced, the breakdown of the material occurs.

1.7 The concept of critical field strength

A purely electrical breakdown phenomenon is one which is not connected in any way with Joule heating effect of the pre-breakdown current. The instability develops very fast but it is not of thermal origin. All calculations of the critical breakdown field strength depend upon the nature of processes that take place in the breakdown phenomenon. Mainly three kinds of breakdown processes can be visualized which may form the basis
of the following three distinct theories of critical field strengths:-

(1) Theories of intrinsic critical field strengths.

(2) Avalanche critical field strength theories.

(3) Theories of space-charge-enhanced critical fields.

As already mentioned (page 11) in the case of high field conduction, the bulk effect theory, the electrode effect theory, and combination of the two may be used to explain the breakdown phenomenon. The calculations of intrinsic critical field strengths are based on the bulk effects, while the avalanche and space-charge theories depend on the electrode and the combination of bulk and electrode effects.

To explain the intrinsic critical field strengths, an uniform field is assumed to be acting on the dielectric of unspecified extent. The influence of electrodes is completely ignored. The theories of intrinsic breakdown yield values of the field strength for which collision with the lattice resulting in ionization in the dielectric becomes an important process between the two electrodes. There is a sudden growth in the number of conduction electrons over and above the steady value. On the other hand, avalanche critical field theories consider the conduction electron multiplication as a process that gradually
reaches intolerable proportions, as the field strength rises. A single electron starting at the cathode can cause an avalanche of electrons of sufficient size to destroy the dielectric. The theories of space-charge involve an electrode ejection processes rather than electron avalanches in the breakdown. Breakdown occurs when the space-charge immediately in front of the electrodes reaches such proportions that the electron current is ejected at the anode and hole current at the cathode. However, these currents are sufficient to destroy the material.

(1.a) Theories based on the single-electron approximation

The calculations of the intrinsic critical field strengths are based on the conduction electron energy balance equation (1.9). If the electric field $F$ is applied to the dielectric and causes the current flow of density $j$, then the rate of electron energy gain from the field is given by:

$$ A = jF $$  

(1.8)

The rate $A$ depends on $F$, the lattice temperature $T$ and average energy $E$ of the electrons. If some mechanism exists whereby conduction electrons can transfer energy to the lattice, the rate of energy transfer (loss) is given by $B$ which depends on $T$ and $E$. The condition for energy balance in the field is given by:

$$ A(F, T, E) = B(T, E) $$  

(1.9)
The following assumptions may be made for the development of this theory.

(i) As it is a purely electrical process, the thermal effect is completely neglected. The slight temperature rise of the lattice due to Joule heating effect is ignored.

(ii) The density of conduction electrons are so low that only the applied field and interaction with the lattice determine the motion of the electron.

(iii) A (the gain from the applied field) and B (the loss to the lattice) are capable of a meaningful interpretation only if the energy change per inelastic collision is much smaller than the average electron energy $E$. Thus $E \gg \hbar \omega$ is a strict requirement (where the lattice frequency $\omega(W)$ is a function of the wave vector $W$).

(iv) A drifting electron loses momentum in a more complicated manner at the equilibrium state, Equ. (1.9). The interaction between a slow electron and the vibrational modes of a polar lattice is too strong to be treated by perturbation theory. This problem of solid state was solved by Frohlich et al.\textsuperscript{51} (1950). The essential physical idea is that the polarization of the lattice caused by the electron, acts back on the electron itself and reduces its energy. As the electron moves through the polar lattice, it carries with it this polarization field. The electron together with the accompanying
polarization field, may behave as a quasi-particle, which is called a "polaron". At equilibrium state, the rate of energy loss \( B \) has been qualitatively confirmed by a numerical calculation of Thornber and Feynman\(^{52}(1970)\) for Frohlich\(^{53}(1954)\) polaron model.

The rate of energy gain per unit volume of a dielectric material is given by:

\[
A = \frac{e^2 F^2 \tau}{m^*} \quad (E, T) \quad (1.10)
\]

where \( \tau \) is the relaxation time for an electron of energy \( E \), \( m^* \) rigid-band effective mass of the electron under the periodic fields of the lattice when the ions are fixed at their equilibrium positions.

The rate of energy loss to the lattice, mainly determined by interaction with the long wavelength phonons, is given by:

\[
B (E, T) = \frac{\hbar \omega \log \tau}{2 \tau_0 (E)} \quad (1.11)
\]

where \( \tau_0 (E) \) is value of \( \tau \) for electrons having an upper limit of energy \( E_0 \) given by:

\[
E_0 = \frac{\hbar \omega^2}{8 m^*} \quad (1.12)
\]
and
\[ \gamma' = \frac{\sqrt{2} E}{a m^{1/2}} \tag{1.13} \]

where \( a \) is anion-cation separation distance.

The Frohlich polaron model contains two important characteristic constants, viz., \( \gamma \) and \( \omega \). The first constant \( \gamma \) is given by:
\[ \gamma = \frac{2 m^* \omega}{\hbar} \tag{1.14} \]

The electronic polarization of the ions due to the motion of a slow moving electron, may also contribute to an effective mass \( m^* \). The length \( \gamma^{-1} \) appears in the theory as a measure of the size of the polaron and hence is of dimensions of length. Taking typical values for an electron in the conduction band of an alkali halide; \( m^* \approx m \) and \( \omega \approx 3 \times 10^{13} \) per sec., one gets \( \gamma^{-1} \approx 10 \text{ Å} \), which justifies large polaron theory.

The second important constant of the theory is the dimensionless quantity:
\[ \chi = e^2 \left( \frac{m^*}{2 \frac{\hbar^2}{3 m^*}} \right)^{1/2} \left( \frac{\hbar}{\gamma} \right) \tag{1.15} \]

It is a measure of the strength of the interaction. For an electron in the conduction band of an alkali halide, \( \chi \approx 5 \), which is a case of intermediate coupling. If low and high frequency
dielectric constants are \( \epsilon_s \) and \( \epsilon_0 \), then the effective dielectric constant is given by:

\[
\frac{1}{\epsilon_s} - \frac{1}{\epsilon_0} = \frac{1}{\epsilon^*}
\]

(1.16)

(v) The lattice frequency \( \omega (W) \) is related with the Reststrahlen angular frequency \( \omega_L \) for long wavelength phonons as follows:

\[
\omega = \omega_L \left( \frac{\epsilon_0}{\epsilon_s} \right)^{1/2}
\]

(1.17)

The development of a critical field strength criterion from the energy-balance equation (1.9) amounts to an identification of an energy \( E \) for which this energy balance holds good. The high-energy criterion as formulated by Frohlich ⁴ (1937) corresponds to the situation when indefinitely large multiplication of electrons in the conduction levels will destroy the dielectric. Let 'I' be the ionization energy corresponding to an electron transition from the valence band to the conduction band, the field applied under such condition may be termed as the critical field \( F_C \), then the critical field strength is that for which collision ionization from the valence band cannot be balanced by the inverse process of recombination (electron hole recombination). If the applied field exceeds \( F_C \), the high energy electron resulting from a recombination collision is not able to completely lose its
Fig. 1.6 - The average rate of energy gain $A(F,T,E)$ from an applied field for various field strengths and the average rate of energy loss to the lattice $B(T,E)$. 
energy to the lattice and may cause further ionization. The above condition gives the relation:

\[ A(\mathcal{F}_c, T, I) = B(I, T) \] (1.18)

Thus the critical situation considers three possible values of the field strength \( \mathcal{F}_1 \), \( \mathcal{F}_2 \) and \( \mathcal{F}_c \) such that \( \mathcal{F}_1 < \mathcal{F}_c < \mathcal{F}_2 \) as shown in Fig. 1.5. For given values of the field \( \mathcal{F}_1 \), \( \mathcal{F}_c \) and \( \mathcal{F}_2 \) the corresponding equilibrium values of electron energy are \( E_1 \), \( I \), \( E_2 \), for which the energy gained from the field is all lost to the structure. The equilibrium energy decreases as the field increases. An electron having less than the equilibrium energy, will lose more energy than it gains. It will slow down and eventually be trapped. An electron with energy greater than the equilibrium energy will increase in energy, for \( A > B \) and will therefore cause ionization. Thus self increasing ionization is only possible by the presence of electrons with energy more than the equilibrium energy. The probability of the appearance of an electron with energy more than \( 'I' \) is very remote because an electron would lose its energy in inelastic collision before \( I \) was reached. Thus \( I \) is the highest value of equilibrium energy which any electron has a chance of securing. Therefore the field \( \mathcal{F}_c \), for which \( I \) is the equilibrium energy, is the lowest field in which self increasing ionization is reasonably possible and is therefore the
minimum breakdown field strength. This can be obtained from equation (1.18) in terms of other crystal constants; and is given by:

\[ F_c = \frac{2 \eta^2}{2 \pi \alpha \beta} \left( \frac{\gamma \exp(\frac{\gamma}{\alpha \beta})^{-1}}{\gamma \exp(\frac{\gamma}{\alpha \beta})^{-1}} \right)^{1/2} \left[ 1 + \frac{2}{\gamma \exp(\frac{\gamma}{\alpha \beta})^{-1}} \right]^{-2/2} \]

(1.19)

where

\[ \eta^2 = \frac{4 \pi \alpha \beta}{\gamma \exp(\frac{\gamma}{\alpha \beta})^{-1}} \left( \frac{\varepsilon}{\varepsilon_0} \right)^{1/2} \]

(1.20)

A criterion for intrinsic breakdown proposed by Von Hippel (1935) and developed by Callen (1949) identifies the energy \( E \) of the Eq. (1.9) with \( E' \) of the Fig. 1.6. This energy corresponds to the maximum of Breakdown-Energy curve. The high energy critical field strength criterion may be stated as:

\[ A(F_c, T, E') = B(E', T) \]

(1.21)

This corresponds to that critical field which will accelerate all conduction electron against the retarding field of the lattice. Hence the critical intrinsic field strength calculated from the energy balance Eq. (1.21) may give:

\[ F_c = \frac{e m}{2 \pi \alpha \beta} \left\{ \frac{\varepsilon_0 \varepsilon_s (\frac{\gamma}{\alpha \beta})^{1/2}}{\gamma \exp(\frac{\gamma}{\alpha \beta})^{-1}} \right\}^{1/2} \left[ 1 + \frac{2}{\gamma \exp(\frac{\gamma}{\alpha \beta})^{-1}} \right]^{-2/2} \]

(1.22)
where

\[ h \omega_c = \frac{\frac{e^2}{\epsilon^*}}{\hbar} \left( \frac{e}{\hbar} \right)^{1/2} \]  

(1.23)

So far, the average distribution of electrons in the conduction band has not been taken into account. If Maxwellian distribution is considered, the electron temperature \( T_e \) will be an appropriate measure of the average energy of the electrons. Hence energy-balance equation becomes:

\[ A( F, T, T_e ) = B( T, T_e ) \]  

(1.24)

Frohlich and Paranjape and Stratton (1961) have shown that the critical field strength derived from Eqn. (1.24) for ionic crystals is given by

\[ F_c = G \left( \frac{T}{\Theta} \right) F_0 \]  

(1.25)

where

\[ F_0 = \frac{m e^2 \omega_c}{\hbar} \left( \frac{\epsilon^*}{\epsilon_s} \right)^{1/2} \frac{1}{\Theta} \]  

(1.26)

* \( \epsilon^*/\epsilon_s \) is effective dielectric constant.

and \( G \) is a function of \( T/\Theta \) alone. The temperature \( \Theta \) is defined by

\[ \Theta = \frac{K}{K} \left( \frac{\epsilon_s}{\epsilon_0} \right)^{1/2} \]  

(1.27)

where \( K \) is the electron wave vector.

The function \( G \) is of the order of unity, and increases slowly with increasing temperature. The above theory of intrinsic
breakdown gives a temperature dependence rather than an absolute value for the critical field strength.

(2) Avalanche critical field strength theories

As already mentioned (page 25), after the steady state is reached, there is a sudden increase in the number of conduction electrons as soon as the critical field is applied. This multiplication is called avalanche process. The mechanism of initiation of the avalanche could be considered in two ways:

(i) The first type is that in which the avalanche of electrons arises by field emission from the valence to conduction band. This has been called field emission breakdown as considered by Zener\textsuperscript{20}(1934).

(ii) The second manner of initiation of avalanche is by collision ionization process, so well known in gaseous discharge phenomenon.

(2.a) Field-emission critical field strengths

The field emission critical field strength introduced by Zener\textsuperscript{20}(1934) was based on the calculation of the probability per unit time that an electron would escape from the valence to the conduction band. Franz\textsuperscript{56}(1939) who obtained an expression for the transition probability by taking into consideration the idea of impulse-thermal breakdown and the interaction between the emitted electron and valence band hole. Finally the critical
breakdown field comes out to be:

\[
F_c = \frac{4 \times 10^7 I^{3/2}}{\log \left(10^{20} t_c^2\right)}
\]

(1.28)

where \( I \) is in eV and \( t_c \) in \( \mu s \). This equation gives approximate (trial) values of the critical field over a wide ranging voltage application time \( t_c \).

(2.b) The forty-generations critical field strength

This theory comes under the second type of initiation process, viz., avalanche initiation. It is based on a single electron collision ionization theory, originally developed by Frohlich\(^{37}\)(1940) and Seitz\(^{62}\)(1949) and formulated in a somewhat more general way by Stratton\(^{55}\)(1961).

One electron, initially at the cathode receives sufficient energy from the applied field to ionize a bound electron. If both of these electrons now receive the same energy from the field, they will each cause further ionization. Eventually \( 2^i \) free electrons will be formed if there are \( i \) generations between cathode and anode. The avalanche will lead to breakdown if \( i \) is sufficiently large. Seitz\(^{58}\)(1949) has argued that by assuming critical field strength \( \approx 10^5 \text{V/cm} \), mobility of the order of \( 1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1} \) and diffusion coefficient of the order of \( 1 \text{ cm}^2 \text{ s}^{-1} \), obtained the result that to cause disruption an avalanche of
$10^{12}$ electrons or 40 generations would be needed. Thus,

$$i = 40 = \epsilon_c \cdot d$$  \hspace{1cm} (1.29)

where $d$ is the inter-electrode distance and $\epsilon_c$ the collision ionization rate per unit length.

Shockley\(^{59}\) (1961) has considered the probability $P(1)$ that a conduction electron has suffered no collision with the lattice over a path length $l$ and thereby obtained an expression:

$$P(1) = \text{Exp} \left( -\frac{I}{eF\langle \lambda \rangle} \right)$$

where $\langle \lambda \rangle$ is a suitable average free path,

and $l I = \frac{I}{eF}$ is the collision free path length required to produce ionization.

Thus $\epsilon = \epsilon_n \text{Exp} \left( -\frac{H}{F} \right)$  \hspace{1cm} (1.30)

where $H = \frac{I}{e\langle \lambda \rangle}$, a constant, and $\epsilon_n$ is a quantity having the dimension of an inverse length. With the use of Equ. (1.29) and (1.30) the critical field strength

$$\frac{H}{F_c} = \log \left( \frac{\epsilon_n \cdot l}{\epsilon_c} \right)$$  \hspace{1cm} (1.31)

Fitting of the experimental data ($F_c$ and $d$) to Equ. (1.31) yields the values for the ionization parameters as calculated in Chapter III.
(2.C) Avalanche - enhanced Cathode emission

Forlani and Minnaja (1964) proposed a combination of the two processes mentioned earlier (2.a and 2.b) in which both the field-emission and the avalanche multiplication were combined together rather than taking account of the avalanche multiplication resulting from only one electron starting from the cathode. The current injection into the conduction levels from the cathode is given by Fowler-Nordheim equation

\[ j_{cath} = j_0 \exp \left( - \frac{4 \sqrt{2} m^* F^{3/2}}{3 \hbar e F} \right) \]  \hspace{1cm} (1.32)

where \( j_0 \) is a field-dependent pre-exponential quantity. The collision ionization rate per unit length is given by:

\[ \lambda = \frac{e F}{I} \]  \hspace{1cm} (1.33)

therefore the current arriving at the anode as a result of this collision-ionization multiplication will be given by:

\[ j_{an} = j_{cath} \exp \left( \frac{e F}{I} d \right) \]  \hspace{1cm} (1.34)

\[ = j_0 \exp \left( - \frac{4 \sqrt{2} m^* F^{3/2}}{3 \hbar e F} + \frac{e F d}{I} \right) \]  \hspace{1cm} (1.35)
If somewhat arbitrary criterion is now adopted that zero exponent in Equ. (1.35) corresponds to critical conditions; then the critical field strength is given by:

\[
F_c = \frac{4 \sqrt{2} m^* \varphi^{3/2} I}{3 \hbar e^2 d}
\]  

(1.36)

Inserting values of reasonable order of magnitude in Equ.(1.36) the field \( F_c \) becomes,

\[
F_c \approx \frac{3 \times 10^4}{d^{1/2}} \text{ V cm}^{-1}
\]  

(1.37)

with \( d \) measured in cm, \( \varphi \) is the work function from the Fermi level in the metal to conduction level of the dielectric. Experimental results consistent with Equ. (1.37) have been found for various inorganic dielectrics by Budenstein and Hayes\(^{51}(1967)\), Smith and Budenstein\(^{62}(1959)\), and Budenstein, Hayes, Smith and Smith \(^{63}(1969)\).

(3) Space-charge-enhanced critical fields

The collision ionization with subsequent separation of mobile electrons and mobile holes may produce a distortion of the field in a dielectric and also in an ionic conductor where charge may accumulate adjacent to an electrode. Both these effects may modify the observed mean breakdown field strength.
(3.a) Space charge produced by collision ionization

To get a relationship between critical field strength and thickness of the dielectric, it is supposed that the mobility of the conduction electrons is much greater than the mobility of the holes. Further the current is assumed to be carried by electrons alone, hence

$$\frac{n \mu_n}{p \mu_p} \gg 1 \quad (1.38)$$

($$\mu_n$$ and $$\mu_p$$ are the mobility of electrons and holes respectively). The electrons continually create new holes by collision ionization. As such over a short time interval, the continuity equation for electron current is then:

$$n e \mu_n F = n \left[ 1 + \chi(F) \right] e \int_n (F + dF) \quad (1.39)$$

where $$\chi(F)$$ is the collision ionization function. The Equ. (1.39) is expanded and with the help of Equ. (1.30) it reduces to

$$dx = - \text{Exp} \frac{-H}{\chi_n F} dF \quad (1.40)$$

Introducing a dimensionless distance $$X = \chi_n x \quad (1.41)$$
and a dimensionless field strength \( \tilde{F} = \frac{F}{H} \).

The Eqn. (1.40) can be integrated from \( F_{\text{cath}} \) to \( F_c \) to give,

\[
X = - \int_{F_{\text{cath}}}^{F_c} \frac{\text{Exp} \left( \frac{1}{F} \right)}{F} \, dF
\]  

(1.42)

Thus \( X \) can be expressed formally in terms of the exponential integral \( E_1(z) \) to yield:

\[
X = E_1 \left( \frac{1}{F} \right) - E_1 \left( \frac{1}{F_{\text{cath}}} \right)
\]  

(1.43)

For given values of the dielectric thickness, this equation predicts a very sharply rising value of \( F_{\text{cath}} \) when the mean field strength \( \tilde{F} \) passes a certain critical value \( F_c \).

Various experimental investigations have supported the idea that breakdown is preceded with the occurrence of a space charge. The distortion of fields in alkali halides was investigated by Von Hippel, Gross, Jelatis, and Geller \( ^{64}(1953) \) and Geller \( ^{65}(1956) \) for KBr and by Williams \( ^{65}(1964) \) for NaCl.

(3.b) Space charge produced by ionic migration

Alger and Von Hippel \( ^{67}(1949) \) suggested that breakdown may be influenced by ionic space charges as was the
case for KBr above 200°C. Watson and Heyes (1970) have interpreted breakdown results on NaCl at 300°C and 350°C as caused in part by ionic migration. If positive ions are transported to the cathode at a rate that exceeds their discharge, then a positive space charge will form in front of the cathode, which will enhance the field in this region. If a total space charge $q$ is distributed uniformly over a distance $x$ in front of the cathode, Poisson's equation gives

$$F_{\text{cath}} - F = \frac{4\pi q}{\epsilon s} \quad (1.44)$$

where $F$ is the field strength in the neutral body of the dielectric. The mean field strength will be given by:

$$\langle F \rangle = F + (F_{\text{cath}} - F) \frac{x}{2d} \quad (1.45)$$

and combination of Eqs. (1.44) and (1.45) gives:

$$F_{\text{cath}} = \langle F \rangle + \frac{4\pi q}{\epsilon s} \left(1 - \frac{x}{2d}\right) \quad (1.46)$$

The rate of accumulation of charge at the cathode is assumed to be:

$$\frac{d q}{d t} = \gamma (u - F) \quad (1.47)$$
where $\eta$ is the fraction of charge transported that is not discharged at the electrode. If breakdown occurs within a short time of application of the impulse, the total space charge is small and $\langle F \rangle$ does not differ greatly from $F$. By solving Equ. (1.47):

$$q = \eta \langle F \rangle t$$  \hspace{1cm} (1.48)

is obtained. From Equs. (1.46) and (1.48), a third equation may be written as:

$$F_{cath} = \langle F \rangle \left\{ 1 + \frac{4\pi \eta \langle F \rangle t}{\langle F \rangle} \left( 1 - \frac{x}{2d} \right) \right\}$$  \hspace{1cm} (1.49)

If breakdown occurs when $F_{cath}$ exceeds some assigned critical field strength $F_C$, the relation between the mean critical field $\langle F_c \rangle$ and the time is given by

$$\frac{F_C}{\langle F_c \rangle} = \left\{ 1 + \frac{4\pi \eta \langle F_c \rangle t}{\langle F_c \rangle} \left( 1 - \frac{x}{2d} \right) \right\}$$  \hspace{1cm} (1.50)

Watson and Heyes (1970) have compared their experimental results with Equ. (1.50). The assigned critical field $F_C$ can be found from the zero time intercept. The value of $\eta$ and $\frac{x}{2d}$ are not separately obtainable from this simple analysis.
REFERENCES


11. Rogowski. (1927), Arch. fur. El. 18, 123.


16 Vorobiev. (1940), J. Tech. Physics, USSR 3, 73.
17 Boning. (1936), Arch. fur. El. 20, 88.
35 Simmons, J.G. (1963 a), J. Appl. Phys. 34, 1793;
(1963 b), J. Appl. Phys. 34, 2581;


38 Hanscomb, J.R., Kao, K.C., Calderwood, J.H., O'Dwyer, J.J.

of the Conference on Electrical Insulation and Dielectric
Phenomena, P. 17.

40 Conwell, E.M. (1967), High field transport in semiconductors

41 Lampert, M.A. and Mark, P. (1970), Current injection in


44 Klein, N. and Gañni, H. (1966), IEEE Trans. electron devices,
ED 13, 281.


47 Von Hippel, A., Gross, E.P., Jelatis, J.G. and Geller, M.
(1953), Phys. Rev. 91, 558.


Frohlich, H., Pelzer, H. and Zienau, S. (1950), Phil. Mag. 41, 221.


