

## Chapter 1

# Introductory Chapter

### 1.1 Elementary act of impact ionization

The basis of all ionization, breakdown and avalanche effects, without exception, is the elementary act of ionization (Fig. 1.1).

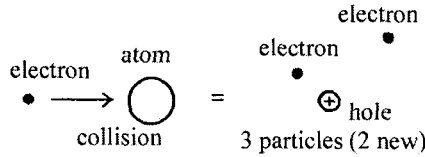


Fig. 1.1 The elementary act of impact ionization. Collision of an energetic electron (or hole) with an atom produces two new free carriers: an electron and a hole.

A free carrier (electron or hole) “impact” on the atom of a semiconductor. If the energy of the carrier is large enough, this carrier will “knock out” the electron from the valence shell of the atom. As a result, two new free carriers, an electron and a hole, appear. In other words, if an initial carrier has enough energy, it can initiate the transition of an electron from a valence band to a conduction band. The minimal energy necessary to carry out the act of impact ionization is called the *threshold energy*  $E_{th}$ . It is clear from the law of energy conservation that the threshold energy cannot be less than the energy gap of the semiconductor  $E_g$ .

The laws of energy conservation and momentum conservation must nevertheless be satisfied simultaneously in the process of an elementary act of ionization. As a result,  $E_{th} > E_g$ . In the case of the simplest dispersion law for electrons and holes (Fig. 1.2), the relation between the energy of the particles  $E$  and their wave vector  $\mathbf{k}$  is defined as  $E = \frac{\hbar^2 k^2}{2m^*}$  (an approximation for the isotropic effective mass  $m^*$ ).

In this case the threshold energy can be calculated fairly simply [1]. The ionization threshold for the electrons,  $E_{the}$  is

$$E_{the} = E_g \left( 1 + \frac{m_e^*}{m_e^* + m_h^*} \right) \quad (1.1)$$

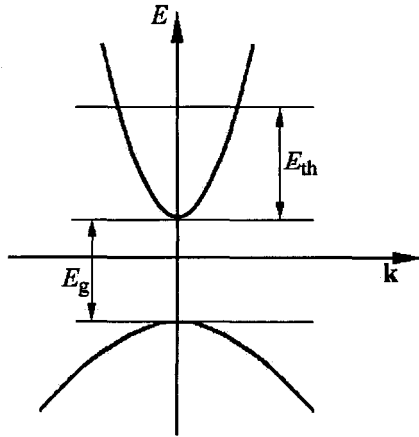


Fig. 1.2 The simplest “parabolic” dispersion law. At any energy, electron and hole can be characterized by isotropic effective mass  $m_e^*$  and  $m_h^*$ , respectively.

Analogously, for ionization initiated by holes, we obtain for the hole ionization threshold  $E_{thh}$

$$E_{thh} = E_g \left( 1 + \frac{m_h^*}{m_e^* + m_h^*} \right) \tag{1.2}$$

It is worth noting that when the effective masses of the electron and hole are equal ( $m_e^* = m_h^*$ ),  $E_{the} = E_{thh} = 3/2E_g$ . On the other hand,  $E_{the} + E_{thh} = 3E_g$  at any effective mass ratio  $m_e^*/m_h^*$ .

The band structure of real semiconductors at high electron or hole energies can never be described by this simple parabolic law, however. A schematic of a GaAs band structure [2] is shown in Fig. 1.3.

At low energy, practically all the electrons are located at the bottom of the central  $\Gamma$ -valley, and can be characterized by an isotropic effective mass  $m_e^* = 0.063m_0$  ( $m_0 = 0.911 \times 10^{-30}$  kg is the mass of a free electron at rest). At high electron energy, however, the electron effective mass even in the  $\Gamma$ -valley depends to an appreciable extent on the electron energy (nonparabolicity). In addition, as seen in Fig. 1.3, there are two side valleys in the conduction band: the  $L$ -valley, in the  $\langle 111 \rangle$  direction of the Brillouin Zone (with energy separation between the bottoms of the  $L$  - and  $\Gamma$  -valleys  $E_{\Gamma L} = 0.29$  eV), and the  $X$ - valley, in the  $\langle 100 \rangle$  direction of the Brillouin Zone (with energy separation between the bottoms of the  $X$ - and  $\Gamma$ - valleys  $E_{\Gamma X} = 0.48$  eV). In both the  $L$  - and  $X$  - valleys the surfaces of equal energy are ellipsoids with a high ratio of longitudinal to transverse effective mass. As for valence band, three bands, those of heavy holes and light holes together with the split-off band, must be taken into account in order to calculate the threshold energy  $E_{th}$  (for details, see Review [3]).

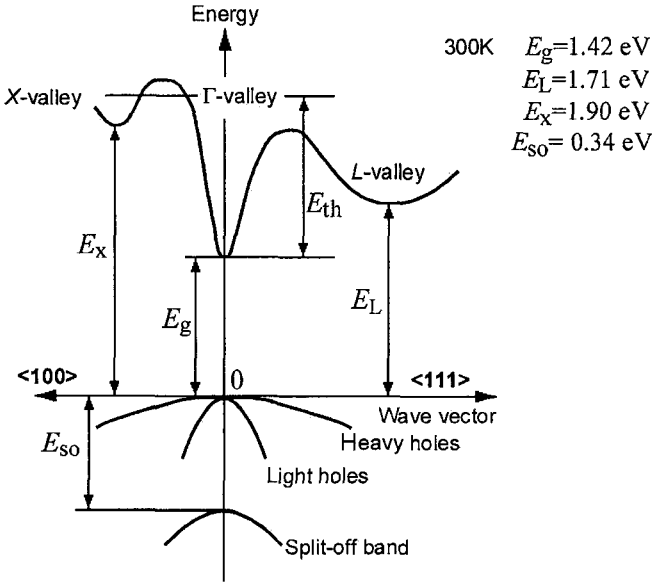


Fig. 1.3 Schematic of a GaAs band structure. The most important minima of the conduction band and maxima of the valence band are indicated.

As a consequence,  $E_{th}$  may be substantially larger than the energy gap, and the threshold energy will demonstrate considerable variation with crystallographic orientation. For GaAs, for example,  $E_{the} = 2.01$  eV for impact ionization by electrons propagating in the  $\langle 110 \rangle$  direction and  $E_{the} = 2.05$  eV for electrons propagating in the  $\langle 100 \rangle$  direction, while electrons moving in the  $\langle 111 \rangle$  direction do not cause impact ionization at all [4].

A schematic Si band structure is shown in Fig. 1.4 [2].

As seen in the figure, the existence of several valleys in the conduction and valence bands must be taken into account in order to calculate the threshold energy  $E_{th}$ . Nevertheless, the estimates show that the magnitude of  $E_{th}$  in the Si is close to the energy gap  $E_g$  (1.1 eV at room temperature) [5].

If the energy of an electron (or hole) is exactly equal to the threshold energy  $E_{th}$ , the cross-section of the impact ionization is zero. As the carrier energy increases, the probability of ionization  $\bar{p}$  increases approximately in the manner [6]:

$$\bar{p} \propto (E - E_{th})^2 \quad (1.3)$$

However, the number of very “energetic” carriers, with an energy  $E$  exceeding  $E_{th}$ , decreases exponentially as  $E$  increases. Thus the effective ionization energy lies very close to the threshold value.

In zero or very low electric fields (close to equilibrium), the role of impact ionization depends to a critical extent on the energy gap  $E_g$ . In relatively wide-gap

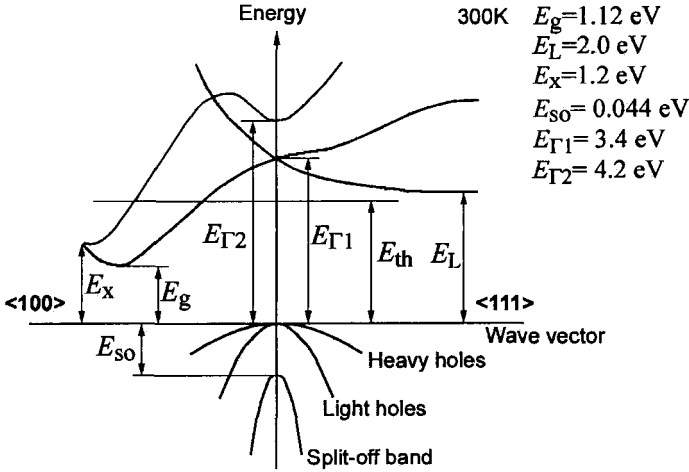


Fig. 1.4 Schematic of the Si band structure. The most important minima of the conduction band and maxima of the valence band are indicated.

semiconductors such as Si and GaAs the role of impact ionization is negligible.

Let us estimate the probability  $\bar{p}$  of finding an electron that is able to cause an act of impact ionization in Si at room temperature. The equilibrium thermal energy of a free electron at 300 K  $\bar{E} = \frac{3}{2}kT \approx 0.039\text{ eV}$  (here  $k = 8.617 \times 10^{-5}\text{ eV K}^{-1}$  is the Boltzmann constant). With a threshold energy  $E_{th} \approx E_g \approx 1.1\text{ eV}$ , this probability is equal to

$$\bar{p} \approx \exp(-E_{th}/\bar{E}) \approx 6 \times 10^{-13} \tag{1.4}$$

In a sample with an equilibrium concentration of, say,  $n_0 = 10^{18}\text{ cm}^{-3}$ , the concentration of electrons which are able to cause impact ionization is about  $6 \times 10^5\text{ cm}^{-3}$ . Given a modern field-effect transistor with characteristic dimensions of  $0.1\mu\text{m} \times 0.05\mu\text{m} \times 50\mu\text{m} = 2.5 \times 10^{-13}\text{ cm}^{-3}$  and an equilibrium concentration in the channel  $\sim 10^{18}\text{ cm}^{-3}$ , the probability of finding just a single electron which can cause the ionization act is about  $10^{-7}$ . In the largest device known in semiconductor electronics, a silicon power rectifier diode, with a characteristic operation area of approximately  $10\text{ cm}^2$ , a base of thickness about  $500\mu\text{m}$  and a characteristic electron equilibrium concentration in the base of about  $10^{13}\text{ cm}^{-3}$ , the total number of electrons in the base is  $5 \times 10^{12}$ . Even in this huge device there are on average only 2-3 electrons with this threshold energy.

However, as seen from Eq. (1.4), the probability  $\bar{p}$  increases exponentially with the mean energy of the carrier  $\bar{E}$ , so that in a strong electric field, when the mean energy  $\bar{E}$  is large enough, the effects of impact ionization becomes very important in semiconductors at any value of  $E_g$ .

In narrow-gap semiconductors with a small  $E_g$ , the probability  $\bar{p}$  can be large enough even at equilibrium, in the absence of an electric field. For example, in InSb ( $E_g = 0.17$  eV) at 300 K, taking  $E_{th} = E_g$ , we have  $\bar{p} \approx 10^{-2}$ . It is apparent that impact ionization processes are very important for such narrow-band semiconductors, even when at equilibrium.

## 1.2 Auger recombination

According to the principle of detailed balance, a state of equilibrium implies that the number of carriers that appear per unit of time due to a distinct generation process must be equal to the number that disappear due to the inverse recombination process. For band-to-band generation, for example, the inverse process is band-to-band recombination, in which the electron and hole recombine and the energy  $E_g$  is transformed into the energy of a photon or a number of phonons.

The inverse of impact ionization is Auger recombination (Fig. 1.5), which is similarly a “three-particle” process, in that an electron and hole recombine and the energy  $E$  that is released ( $E \geq E_{th}$ ) is transferred to a third particle, which can be either an electron or a hole. Two electrons and a hole are involved in the Auger recombination process in an  $n$ -type material ( $e$ - $e$ - $h$  process) and two holes and an electron in a  $p$ -type material ( $h$ - $h$ - $e$  process).

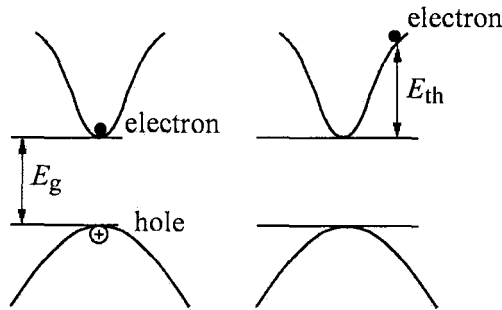


Fig. 1.5 Schematic representation of a three-particle electron-electron-hole ( $e$ - $e$ - $h$ ) Auger recombination process. An electron and a hole recombine, and the energy released  $E \geq E_{th}$  is transferred to the electron.

Decrease in the number of excess carriers (e.g. electrons) due to Auger recombination can be described by the expression [7]:

$$\frac{dn}{dt} \approx -(C_n n + C_p p) \cdot n \cdot p, \quad (1.5)$$

where  $C_n$  and  $C_p$  are the Auger coefficients for  $e$ - $e$ - $h$  and  $h$ - $h$ - $e$  processes, respec-

tively. If  $n = p$  (intrinsic semiconductor or the case of a high injection level):

$$\frac{dn}{dt} = -n^3(C_n + C_p) = -Cn^3 \quad (1.6)$$

For a low injection level in an  $n$ -type semiconductor ( $p \ll n_0 \approx N_d$ , where  $N_d$  is the donor doping level), we have

$$dp/dt = -C_n p n_0^2 = -C_n p N_d^2 \quad (1.7)$$

For a low injection level in a  $p$ -type semiconductor ( $n \ll p_0 \approx N_a$ , where  $N_a$  is the acceptor doping level), we obtain

$$dn/dt = -C_p n p_0^2 = -C_p n N_a^2 \quad (1.8)$$

Comparing the expressions for Auger recombination and conventional (linear) Shockley-Read recombination via traps,  $dp/dt \approx -p/\tau_p$  for an  $n$ -type semiconductor and  $dn/dt \approx -n/\tau_n$  for a  $p$ -type semiconductor, the recombination lifetimes associated with Auger recombination can be written in the form:

$$\tau_{nA} = \frac{1}{C_p N_a^2}; \quad \tau_{pA} = \frac{1}{C_n N_d^2} \quad (1.9)$$

In narrow band semiconductors it is Auger recombination that determines the maximum achievable lifetime in a pure material. In InSb, for example, the intrinsic concentration, i.e. the minimum possible concentration at room temperature, is  $n_i = p_i \approx 2 \times 10^{16} \text{ cm}^{-3}$  [2]. With  $C = 5 \times 10^{-26} \text{ cm}^6 \text{ s}^{-1}$  [2],  $\tau_{nA} = \tau_{pA} = 1/Cn_i^2 \approx 5 \times 10^{-8} \text{ s}$ . Hence, the lifetime in InSb at 300 K cannot exceed this value even at a zero concentration of recombination traps (Fig. 1.6).

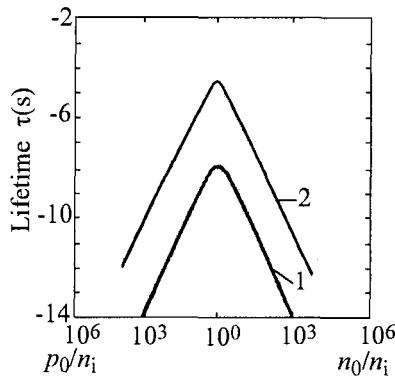


Fig. 1.6 Dependencies of lifetimes associated with Auger recombination on carrier concentration in InSb (Curve 1) and InAs (Curve 2) at 300K.

In InAs ( $E_g = 0.354$  eV) at 300 K,  $n_i = p_i \approx 10^{15}$  cm $^{-3}$  [2], and with  $C = 2.2 \times 10^{-27}$  cm $^6$  s $^{-1}$  [2],  $\tau_{nA} = \tau_{pA} = 1/Cn_i^2 \approx 4.5 \times 10^{-4}$  s. This value is rather large, and at room temperature the lifetime, even in pure material, is limited by either Shockley-Read or band-to-band recombination. The rate of Auger recombination nevertheless increases with temperature or doping in a manner that is directly proportional to the square of the doping (Fig. 1.6, Curve 2), and it is Auger recombination that limits the maximum achievable lifetime at elevated temperatures (large intrinsic concentration) and / or moderate doping levels.

In relatively wide-gap semiconductors such as Si ( $E_g = 1.12$  eV,  $n_i = p_i \approx 10^{10}$  cm $^{-3}$  at 300K [2]), the part played by Auger recombination is important either at a high doping level or in the event of a high carrier concentration caused by optical pumping or injection.

The dependence of the hole lifetime  $\tau_p$  in pure silicon on the concentration of shallow donors ( $n_0 = N_d$ ) is demonstrated in Fig. 1.7. At a low doping level,  $\tau_p$  falls into a very broad range from several milliseconds to nanoseconds, depending on the concentration of deep traps, and the lifetime is determined by the linear Shockley-Read recombination lifetime  $\tau_{SR}$ . Auger recombination comes into play when  $\tau_{SR} \sim \tau_A = 1/CN_d^2$  or  $N_d \geq \left(\frac{1}{C\tau_{SR}}\right)^{1/2}$ .

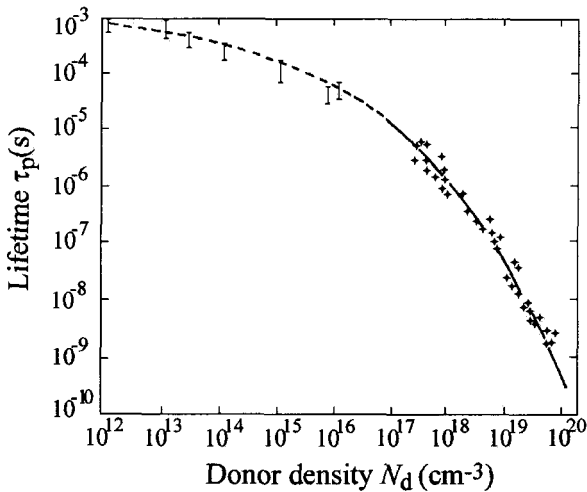


Fig. 1.7 Hole lifetime  $\tau_p$  as a function of shallow donor density in pure  $n$ -type Si. The fact that, at  $N_d \geq 5 \times 10^{17}$  cm $^{-3}$ ,  $\tau_p \sim 1/N_d^2$  indicates the decisive importance of Auger recombination at a high doping level.

At  $C = 1.1 \times 10^{-30}$  cm $^6$  s $^{-1}$  and  $\tau_{SR} \approx 10^{-3}$  s (Si of very high purity), the role of Auger recombination will be decisive at  $N_d \gg 3 \times 10^{16}$  cm $^{-3}$ , while at a fairly

typical  $\tau_{SR}$  of  $10^{-6}$  s Auger recombination predominates at  $N_d \gg 10^{18} \text{ cm}^{-3}$ .

### 1.3 Energy of electrons and holes as a function of electric field

In an electric field  $F$  the carriers take their energy from the field:

$$\frac{dE}{dt} = ev \cdot F, \quad (1.10)$$

where  $e$  is the electron charge and  $v$  is the electron velocity.

On the other hand, they give up their excess energy to the crystal lattice due to various types of scattering (acoustic phonons, optical phonons, piezoelectric scattering, impurity scattering, etc.):

$$\frac{dE}{dt} = -\frac{E - E_0}{\tau_e} \quad (1.11)$$

Here  $E_0 = \frac{3}{2}kT$  is the equilibrium energy of the carrier and  $\tau_e$  is the effective electron energy relaxation time ( $\tau_e$  is usually of the order of  $10^{-12} - 10^{-13}$  s). In a steady state,

$$evF = \frac{E - E_0}{\tau_e} \quad (1.12)$$

and

$$E = E_0 + evF \cdot \tau_e \quad (1.13)$$

In a low electric field, Ohm's law is satisfied, so that  $v = \mu F$ , where  $\mu$  is a low field mobility. Then

$$E = E_0 + e\mu F^2 \tau_e \quad (1.14)$$

and the excess energy  $\Delta E = E - E_0$  is proportional to  $F^2$ .

In strong electric fields the carrier drift velocity *saturates* in the majority of important cases and becomes almost independent of the electric field (see Fig. 1.8). In this case,  $v = v_s$ , and

$$\Delta E = E - E_0 = ev_s F \tau_e \quad (1.15)$$

The excess energy in this case is proportional to  $F$ :  $\Delta E \sim F$ .

Although  $\tau_e$  depends on the electric field in strong electric fields, Eqs. (1.14) and (1.15) give a qualitatively correct idea of the dependencies of carrier energy on the electric field. The field dependencies of the mean electron energy  $E_n$  (a) and mean hole energy  $E_p$  (b) in Si, calculated by the Monte-Carlo technique, are shown in Fig. 1.9. As seen,  $E_n$  and  $E_p$  are at their equilibrium value  $E_n = E_p \approx 0.039$  eV in low electric fields. The mean energy of the carriers then increases with  $F$ , so that in relatively low fields the dependencies  $E(F)$  come close to the law  $\Delta E \sim F^2$

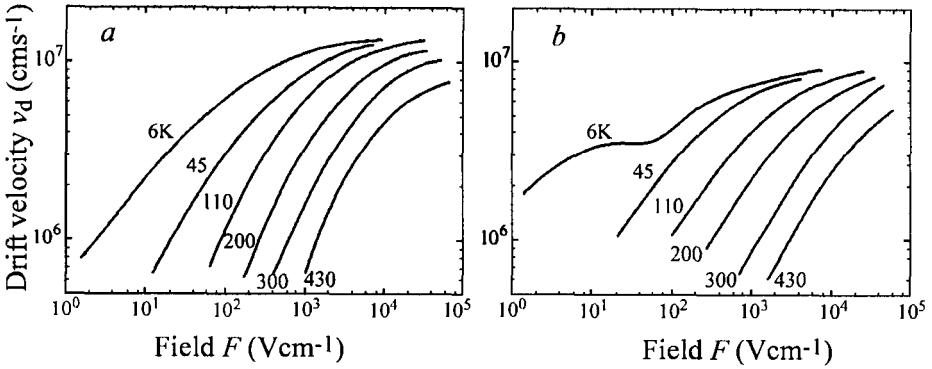


Fig. 1.8 Dependencies of (a) electron drift velocity and (b) hole drift velocity in Si on the electric field at different temperatures [8]. (With kind permission from Elsevier)

(see Eq. (1.14)) and in relatively strong fields ( $v = v_s$ ),  $\Delta E \sim F$  (Eq. (1.15)). The higher the doping level, the smaller is the low field mobility  $\mu$  and the higher value of  $F$  required to reach a given magnitude of  $E$ .

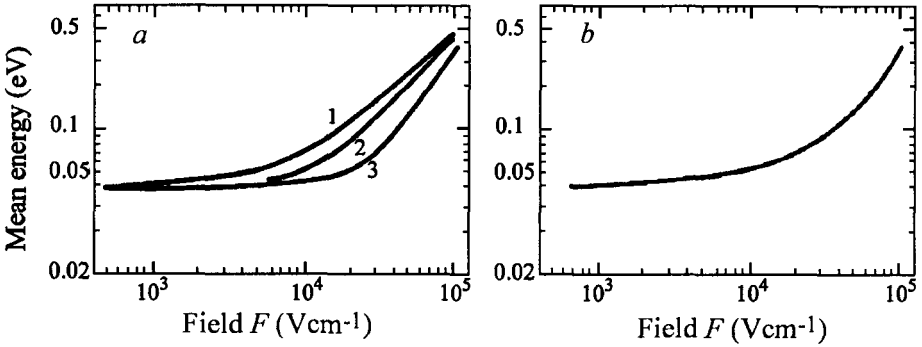


Fig. 1.9 Field dependencies of (a) mean electron energy and (b) mean hole energy in Si at 300 K, calculated by the Monte-Carlo technique. The dependencies for electrons are calculated for three doping levels  $N_d(\text{cm}^{-3})$ : 1 - 0 (high purity Si), 2 -  $4 \times 10^{18}$  and 3 -  $4 \times 10^{19}$ . The dependence  $E(F)$  in Fig. 1.9 b is calculated for pure  $p$ -Si [8]. (With kind permission from Elsevier)

As mentioned earlier, the probability  $\bar{p}$  of finding an electron that is able to cause an act of impact ionization increases vastly with an increase in the strength of the field. For Si, for example, the characteristic field for impact ionization  $F_i$  is about  $3 \times 10^5$  V/cm. As seen in Fig. 1.9 a, at a field  $F$  of  $2 \times 10^4$  V/cm, which is an order of magnitude less than  $F_i$ , the mean energy of electrons  $\bar{E}$  in pure Si (curve 1) is approximately 0.1 eV. Using Eq. 1.4 it is easy to estimate  $\bar{p}$  at  $E_{th} = 1.1$  eV and  $\bar{E} = 0.1$  eV:  $\bar{p} \approx 1.7 \times 10^{-5}$ , and by comparing this value with the result for

equilibrium conditions at room temperature (Eq. (1.4)) one can see that a 2.5-fold increase in  $\bar{E}$  will cause  $\bar{p}$  to increase by seven orders of magnitude.

### 1.4 Main approaches for describing ionization phenomena

There are 3 main approaches to the study of ionization phenomena:

- Approximation of the characteristic breakdown field  $F_i$ ,
- Monte-Carlo simulation,
- Approximation of ionization rates.

#### 1.4.1 Approximation of the characteristic breakdown field $F_i$

Approximation of the characteristic breakdown field  $F_i$  is the simplest way of describing impact ionization and breakdown phenomena. In the framework of this approach one assumes that avalanche breakdown occurs if the maximum field  $F_m$  at any point in the structure exceeds a value known as the characteristic breakdown electric field  $F_i$ . If  $F_m$  is less than  $F_i$ , there will be no impact ionization phenomena at all.

This approach is illustrated with an abrupt reverse biased  $p - n$  junction in Fig. 1.10. The field  $F$  reaches its maximum at the boundary of the  $p$  and  $n$  regions and (in the simplest case of homogeneous doping) decreases linearly with distance from the junction. The slope  $F(x)$  is determined by the one-dimensional

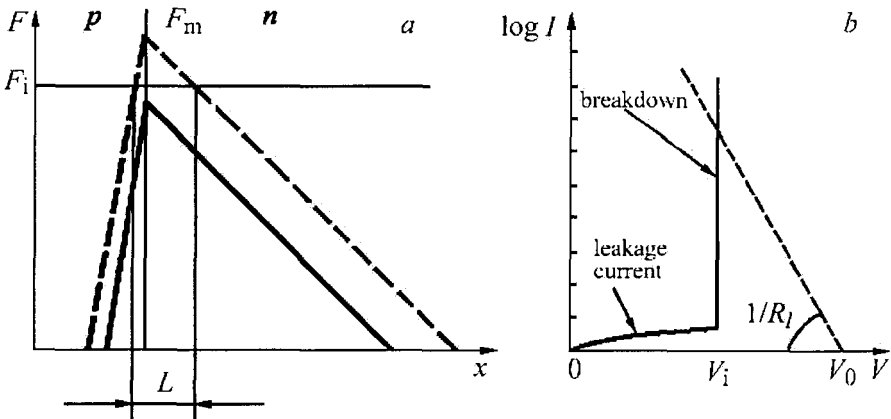


Fig. 1.10 Approximation of characteristic breakdown field  $F_i$ . If  $F_m < F_i$  and, accordingly, the bias applied to the structure,  $V$ , is less than breakdown voltage  $V_i$ , there are no ionization effects. The current flowing through reverse biased  $p - n$  junction is the leakage current. At  $F_m = F_i$  (and, respectively  $V = V_i$ ), the “mature” breakdown occurs, and current is controlled by external load resistance  $R_l$ .

Poisson equation  $dF/dx = eN/\epsilon\epsilon_0$ , where  $N = N_a$  in the  $p$  region and  $N = N_d$  in the  $n$  region,  $\epsilon$  is the dielectric constant of the semiconductor, and  $\epsilon_0$  is the permittivity of the vacuum. The width of the space-charge region  $W$  in each region is connected with the maximum field  $F_m$  and the doping level  $N$  by the obvious relation  $W = \epsilon\epsilon_0 F/eN$ . At  $F_m = F_i$ ,  $W_i = \epsilon\epsilon_0 F_i/eN$ , and the breakdown voltage drop  $V_i$  across each ( $p$  and  $n$ ) region is  $V_i = \frac{F_i W_i}{2} = \frac{\epsilon\epsilon_0 F_i^2}{2eN}$ . The breakdown voltage of this structure, calculated in the framework of this approximation, is  $V_i = \frac{\epsilon\epsilon_0 F_i^2}{2e} \left( \frac{1}{N_a} + \frac{1}{N_d} \right)$ .

If the applied bias  $V < V_i$ , and accordingly  $F_m < F_i$ , the leakage current will flow through the reverse biased junction. Mature breakdown occurs at  $V = V_i$ , and ( $F_m = F_i$ ), and (in the framework of this approximation) the current flowing through the structure is controlled only by the external load resistance  $R_l$  (Fig. 1.10b).

This is obviously a very rough approximation indeed, since the transient problems cannot be even formulated in the framework of this approach, for example. Moreover, it is clear that at  $V > V_i$ , the maximum field  $F_m$  exceeds  $F_i$  value and ionization effects occur not at a single point but across the whole region  $L$ , where  $F_m > F_i$  (Fig. 1.10a). In reality, the current-voltage characteristic at breakdown does not take the form of a vertical straight line with zero differential resistance  $dV/dI = 0$  (Fig. 1.10b), etc.

Nevertheless, this approximation is not infrequently used, e.g. to estimate the magnitude of the breakdown voltage at  $p-n$  junctions and in Schottky diodes with arbitrary doping distribution across the base, in field-effect and bipolar transistors (FETs and BJTs), in thyristors and in many other semiconductor devices. On the other hand, it can be used for qualitative analysis in very complicated situations connected with breakdown phenomena (see Chapters 4 and 5).

The characteristic breakdown electric field  $F_i$  in all semiconductor materials depends on the doping level and the temperature. Its dependence on the doping level  $N$  for Si is shown in Figure 1.11, where  $F_i$  is seen to increase monotonically from  $2 \times 10^5$  V/cm to  $6 \times 10^5$  V/cm with a rise in doping from  $\sim 10^{14}$  cm $^{-3}$  to  $\sim 10^{17}$  cm $^{-3}$ . Analogous dependencies are common in other semiconductors. In GaAs  $F_i$  increases from  $\sim 3 \times 10^5$  to  $8 \times 10^5$  V/cm in the same doping concentration range, while in SiC,  $F_i$  falls over a range between  $2 \times 10^6$  and  $6 \times 10^6$  V/cm.

Such type of  $F_i(N)$  and  $V_i(N)$  dependencies are also characteristic of other semiconductors (Fig. 1.12).

As a rule, both the breakdown voltage  $V_i$  and the breakdown field  $F_i$  increase with temperature (Fig. 1.13), so that the dependencies shown in the figure are typical of most semiconductors and semiconductor structures.  $V_i$  increases monotonically with temperature, so that the smaller the doping level, the stronger is the temperature dependence of  $V_i$ .

The characteristics of the dependencies can be explained by simple qualitative

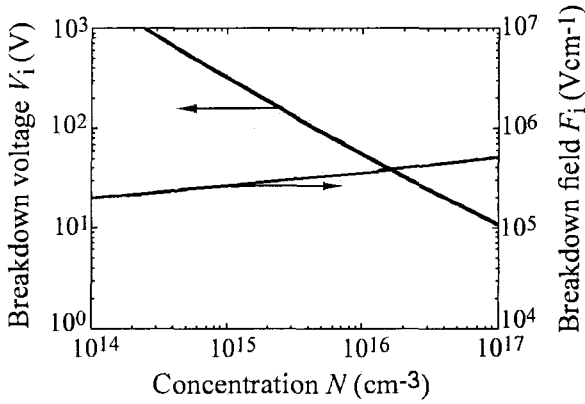


Fig. 1.11 Breakdown field  $F_i$  and breakdown voltage  $V_i$  of an abrupt and highly asymmetrical ( $N_a \gg N_d$ ) Si  $p-n$  junction as a function of doping level.

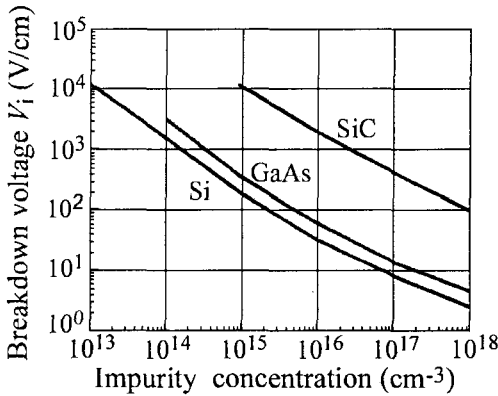


Fig. 1.12 Breakdown voltage as a function of doping level for abrupt  $p^+-n$  junctions fabricated on the basis of the Si, GaAs and SiC.

considerations. As the impact ionization process is defined by the energy of the carrier, gaining from the electric field between scattering collisions, the probability of impact ionization decreases as scattering events become more frequent. Thus, since the frequency of phonon scattering increases with temperature, it becomes more difficult for an electron (hole) to take a large amount of energy from the electric field. (This can be described formally as a decrease in  $\tau_e$  in Eqs. (1.13–1.15) as temperature increases.) As a result, the breakdown field and breakdown voltage increase with temperature.

The lower the doping level, the larger is the relative contribution of phonon scattering to the total scattering processes. That is why the temperature dependence

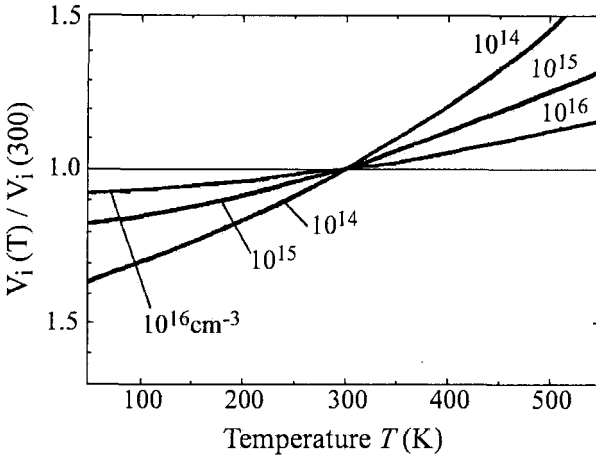


Fig. 1.13 Normalized breakdown voltage versus temperature for an abrupt Si  $p$ - $n$  junction at different doping levels [9].

of  $V_i$  becomes greater as the doping level decreases.

There are exceptions to this rule, however, as voltage breakdown  $V_i$  can *decrease* with increasing temperature in semiconductors with a high concentration of deep levels, due to thermal ionization of the traps. As we will see later, such a situation is very dangerous from the point of view of possible thermal instabilities.

#### 1.4.2 Monte-Carlo simulation

The Monte-Carlo technique is a very powerful numerical method that allows us to simulate any transport phenomena in semiconductors, including ionization and breakdown effects. It is based on the approach suggested in Ref. [10]. The idea is to simulate carrier motion in a  $\mathbf{k}$ -space (and generally speaking  $\mathbf{x}$ -space) under the action of an electric field and scattering processes. By observing the motion of a single electron (hole) in a  $\mathbf{k}$ -space for a sufficiently long time, we obtain a distribution function  $f(\mathbf{k})$ . All the transport parameters, such as the drift velocity  $v(F)$ , diffusion coefficient  $D(F)$ , etc., can then easily be calculated.

Between the scattering events electron (hole) moves in the electric field, and the change in the carrier wave vector  $\mathbf{k}$  is determined by electric field  $F$ :

$$\vec{k}(t) = \vec{k}_0 + \frac{e\vec{F}}{\hbar}t \quad (1.16)$$

A scattering event is defined as occurring at an instant  $t_1$  determined by a computer-generated random number  $r_1$ . Another random number  $r_2$  defines which scattering process occurs: acoustic scattering, polar optical scattering, impurity scattering,

etc. The next random number (or numbers) will be taken to define the parameters of the electron state after scattering, and so on . . . The probabilities of the scattering events should be known from microscopic theory or from experimental data. When considering ionization phenomena one must take into account the probabilities attached to the elementary acts of impact ionization [1; 11; 12; 13]. Details of the Monte-Carlo algorithm can be found in many books and handbooks (see, for example [14]).

Monte-Carlo technique has become a standard numerical method nowadays, and is a conventional attribute of many commercial simulators (ATLAS, DESSIS, MEDICI etc.). The accuracy of its calculations in the present instance is limited by only the accuracy of our knowledge of the band structure and scattering rates.

This technique is rarely used to calculate the operating regimes of devices, however, because it usually takes up too much computer time. It is used as a rule to check the principal problems and to calculate ionization rates, and it has also been successfully used to simulate extremely small semiconductor devices when all other techniques have failed due to the large space inhomogeneities and very high space derivatives that are characteristic of small devices.

### 1.4.3 *Approximation of ionization rates*

The approximation of ionization rates is the “workhorse” of the theory of ionization phenomena. It is a very productive and effective compromise between the “oversimplified” approach of the effective breakdown field  $F_i$  and the rigorous but rather complicate Monte-Carlo simulation procedure.

In the framework of this approach one assumes that impact ionization is characterized by ionization rates of  $\alpha_i$  for electrons and  $\beta_i$  for holes, which are defined as *probabilities of impact ionization per unit length*.

For example, if in a given electric field  $F$  an electron moves an average distance of  $l_i = 10^{-3}$  cm between two acts of impact ionization, then  $\alpha_i$  is equal to  $10^3$  cm $^{-1}$ . If  $l_i$  is equal to  $10^{-5}$  cm, then  $\alpha_i = 10^5$  cm $^{-1}$ , and so on . . . Ionization rates are assumed to be instant functions of the electric field  $F$ :  $\alpha_i(F)$  and  $\beta_i(F)$  (the *local model*). This assumption has obvious limitations, however.

Let us assume that at  $t = 0$  the field  $F$  increases instantly from  $F = 0$  to a high value  $F_0$  (Fig. 1.14). It takes some time for an electron (or hole) to acquire the threshold energy  $E_{th}$  which is necessary to produce an elementary act of impact ionization. Roughly speaking, this time will be equal to the energy relaxation time  $\tau_e$  ( $\sim 10^{-12}$ – $10^{-13}$  s in high electric fields). Hence, when considering processes with characteristic times of some picoseconds or less (i.e. frequencies of some hundreds of Gigahertz and higher), we must remember that the local model may not be valid.

A similar situation emerges if an electric field  $F$  changes very sharply in a space (Fig. 1.15). It is clear that if a field changes notably along the mean free path  $l_0$ , it will be impossible to say which value of  $F$  should we use to calculate  $\alpha_i(F)$  or  $\beta_i(F)$ .

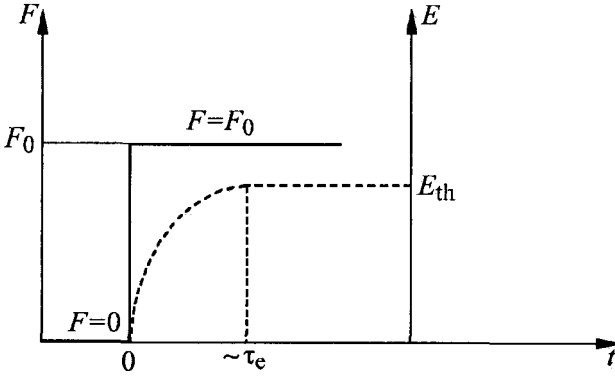


Fig. 1.14 If field  $F$  increases instantly from  $F = 0$  to  $F = F_0$ , it will take the time  $t \sim \tau_e \sim (10^{-12} - 10^{-13} \text{ s})$  for an electron or hole to acquire the appropriate energy  $E$ . The solid line represents the time dependence of  $F$  and the dashed line the qualitative time dependence of  $E$ .

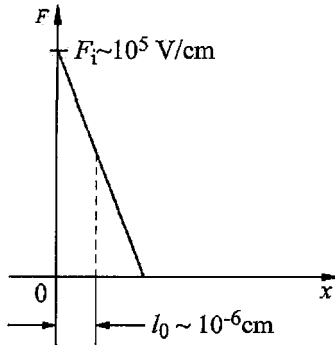


Fig. 1.15 If an electric field  $F$  changes very sharply ( $dF/dx \geq 10^{11} \text{ V/cm}^2$ ), local models (including in part the ionization rates approach) will not be valid.

Taking a characteristic mean free path of  $l_0 \sim 10^{-6} \text{ cm}$  and a characteristic  $F_i$  of  $10^5 \text{ V/cm}$ , one can estimate a characteristic magnitude for  $dF/dx$  of about  $10^{11} \text{ V/cm}^2$ . Such large values of  $dF/dx$  are realized either in extremely small semiconductor structures with characteristic sizes of about  $100\text{--}1000 \text{ \AA}$  and less or in the case of very high doping levels  $N \geq 10^{18} \text{ cm}^{-3}$ . In these cases Monte-Carlo simulation should be used to describe the ionization processes correctly.

The approximation of ionization rates is nevertheless the most popular and most efficient tool for studying ionization and breakdown phenomena in its region of applicability:  $f \leq 400 - 500 \text{ GHz}$ ,  $L \geq 0.1 \text{ }\mu\text{m}$ , and  $N \leq 10^{18} \text{ cm}^{-3}$ .

Even in very strong electric fields it is the case as a rule that only a small portion

of the electrons (or holes) have an energy which exceeds the characteristic critical energy  $E_0$  ( $E_0 \sim E_{th}$ ). On average the carrier energy is much smaller, and it is limited by optical phonon scattering with an energy of  $E_{ph} = \hbar\omega_0 \ll E_0$ . The values concerned are  $\hbar\omega_0 = 0.063$  eV for Si, for example,  $\hbar\omega_0 = 0.035$  eV for GaAs and  $\hbar\omega_0 \approx 0.1$  eV for SiC.

In order to achieve an energy of  $E_0$ , an electron has to move without collision for a distance

$$l_F = \frac{E_0}{eF} \quad (1.17)$$

The probability of such an event is

$$p \approx \exp(-l_F/l_0) \approx \exp\left(-\frac{E_0}{eFl_0}\right) \quad (1.18)$$

(where  $l_0$  is the mean free path).

Hence, the expressions for  $\alpha_i(F)$  and  $\beta_i(F)$  take the forms

$$\alpha_i = \alpha_0 \exp[-F_{n0}/F] \quad (1.19)$$

and

$$\beta_i = \beta_0 \exp[-F_{p0}/F] \quad (1.20)$$

where  $F_0 = E_0/el_0$ .

To calculate  $\alpha_0$  and  $\beta_0$  and define the exponential parts of these expressions more precisely, it is necessary to decide on the form of the energy distribution function for electrons (holes) [11; 12; 13] (see Review [15])

The experimental  $\alpha(F)$  and  $\beta(F)$  dependencies for the most important semiconductor materials are usually described by the following empirical equations:

$$\begin{aligned} \alpha_i &= \alpha_0 \exp[-F_{n0}/F]^{m_n} \\ \beta_i &= \beta_0 \exp[-F_{p0}/F]^{m_p}. \end{aligned} \quad (1.21)$$

In Si, for example,  $m_n = m_p = 1$ , and

$$\alpha_i = 3.318 \cdot 10^5 \exp[-1.174 \cdot 10^6/F] \quad (\text{cm}^{-1})$$

To describe the  $\beta(F)$  dependence correctly over a wide range of  $F$  values, it is necessary to use a "two-piece" approximation:

If  $2 \cdot 10^5 \text{ V/cm} < F < 5.3 \cdot 10^5 \text{ V/cm}$ ,

$$\beta_i = 2 \cdot 10^6 \exp[-1.97 \cdot 10^6/F] (\text{cm}^{-1}),$$

while for  $F > 5.3 \cdot 10^5 \text{ V/cm}$ ,

$$\beta_i = 5.6 \cdot 10^6 \exp[-1.32 \cdot 10^6/F] (\text{cm}^{-1}),$$

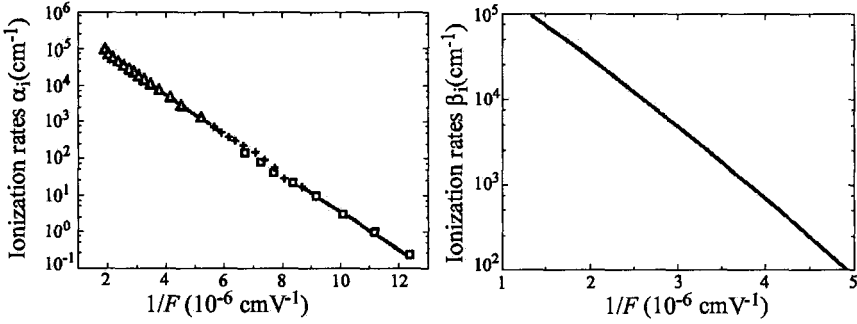


Fig. 1.16 Dependencies of the ionization rates of electrons,  $\alpha_i$ , and holes,  $\beta_i$ , in Si on  $1/F$  at 300K [16; 17]. (With kind permission from Elsevier)

(see Figure 1.16).

The  $\alpha_i(F)$  and  $\beta_i(F)$  dependencies for many semiconductor materials can be found in Refs. [2; 18; 19].

It is worth noting that if the electric field  $F$  is relatively small ( $F \ll F_0$ ),  $\alpha_i$  and  $\beta_i$  will be very strongly dependent on the field strength, while if  $F \sim F_0$ ,  $\alpha_i$  and  $\beta_i$  will show a fairly weak dependence on  $F$ .

In very strong fields,  $F \gg F_0$ ,  $\alpha_i$  and  $\beta_i$  tend towards their limiting values of  $\alpha_0$  and  $\beta_0$ , respectively, which fall within the range  $\sim 10^4$  to  $\sim 10^6 \text{ cm}^{-1}$  for different semiconductors. These limiting values correspond to a situation in which the distance between two elementary acts of impact ionization  $l = 1/\alpha_0$  is equal to the mean free path  $l_0$ , i.e. in which electrons (holes) ionize at *every scattering act*.

Let us consider the fluxes of electrons and holes passing through a region of a semiconductor (Figure 1.17). While travelling a distance  $dx$ , each electron will create an average of  $(\alpha_i dx)$  electron-hole pairs. The increase in the electron current *due to electron multiplication* will thus be

$$\left. \frac{dj_n}{dx} \right|_n = \alpha_i j_n dx \quad (1.22)$$

In addition, the electron current density will increase *due to hole multiplication*:

$$\left. \frac{dj_n}{dx} \right|_p = \beta_i j_p dx \quad (1.23)$$

Hence

$$\frac{dj_n}{dx} = \alpha_i j_n + \beta_i j_p \quad (1.24)$$

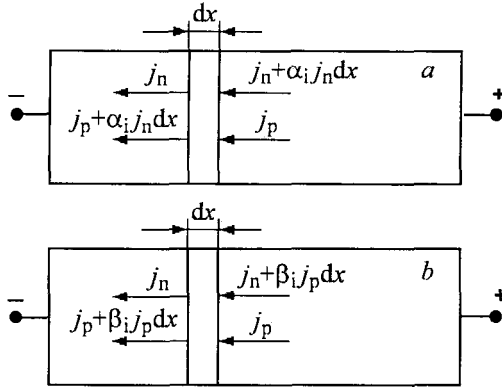


Fig. 1.17 Augmentation of the electron and hole current densities  $j_n$  and  $j_p$  caused by impact ionization by electrons (a) and holes (b).

Analogously,

$$\frac{dj_p}{dx} = -\beta_i j_p - \alpha_i j_n \tag{1.25}$$

and

$$\frac{dj_p}{dx} = -\frac{dj_n}{dx} \tag{1.26}$$

The total current density

$$j = j_n + j_p, \tag{1.27}$$

where  $j_n = env_n(F) + eD_n \frac{dn}{dx}$ ;  $j_p = epv_p(F) - eD_p \frac{dp}{dx}$ .

As we will see, by solving simultaneously the set of equations (1.25)–(1.27) and the Poisson equation

$$\frac{dF}{dx} = \frac{e}{\epsilon \epsilon_0} (N_d - N_A + p - n), \tag{1.28}$$

with appropriate boundary conditions allow us to describe (in a one-dimensional approximation) the steady state electron and hole distributions under conditions of avalanche multiplication and breakdown.

The transient characteristics can be described by a set of partial differential

equations:

$$\begin{aligned}
 \frac{\partial n}{\partial t} &= \frac{1}{e} \frac{\partial j_n}{\partial x} + \frac{1}{e} (\alpha_i j_n + \beta_i j_p) \\
 \frac{\partial p}{\partial t} &= -\frac{1}{e} \frac{\partial j_p}{\partial x} + \frac{1}{e} (\alpha_i j_n + \beta_i j_p) \\
 j_n &= env_n(F) + eD_n \frac{dn}{dx} \\
 j_p &= epv_p(F) - eD_p \frac{dp}{dx} \\
 \frac{dF}{dx} &= \frac{e}{\varepsilon\varepsilon_0} (N_d - N_A + p - n)
 \end{aligned} \tag{1.29}$$

with appropriate boundary and initial conditions.