Chapter 21
Bipolar Devices

21.1 Introduction

The mathematical model of semiconductor devices, derived in Chap. 19, is applied here to the description of the fundamental bipolar device, the *p–n* junction. The term *bipolar* indicates that both electrons and holes contribute to the current. The analysis is carried out using the simple example of a one-dimensional abrupt junction in steady state, with the hypotheses of non-degeneracy and complete ionization, that lend themselves to an analytical treatment. The equilibrium condition is considered first, and the solution of Poisson's equation is tackled, showing that the structure can be partitioned into space-charge and quasi-neutral regions. Then, the Shockley theory is illustrated, leading to the derivation of the ideal $I(V)$ characteristic. The semiconductor model is then applied to illustrating two features of the reverse-bias condition, namely, the depletion capacitance and the avalanche due to impact ionization. The complements justify the simplification of considering only the diffusive transport for the minority carriers in a quasi-neutral region, and provide the derivation of the Shockley boundary conditions. Finally, the expression of the depletion capacitance is worked out for the case of an arbitrary charge-density profile.

21.2 *P–N* Junction in Equilibrium

A very simple, yet fundamental, semiconductor device is the *p–n junction*, whose one-dimensional version is sketched in Fig. 21.1. The device is fabricated by thermally diffusing (Chap. 23), or ion implanting *p*-type dopant atoms into an *n*-type substrate, or vice versa. As a consequence, the diffused or implanted profile is not spatially uniform. The substrate profile may in turn result from a similar process, so that in general it is not uniform either. The locus of points where the ionized dopant concentrations are equal to each other, $N_D^+ = N_A^-$, is a surface called *metallurgical*

M. Rudan, *Physics of Semiconductor Devices,*
DOI 10.1007/978-1-4939-1151-6_21

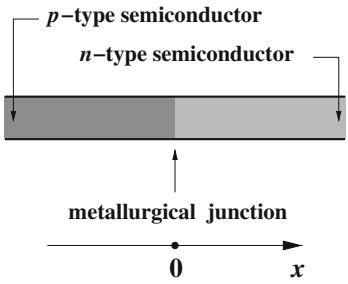

Fig. 21.1 Schematic example of a one-dimensional *p–n* junction

junction.[1] The theory of the *p–n* junction is carried out with reference to a simplified structure, where the device is one dimensional and aligned with the x axis; in this case the metallurgical junction is a plane normal to x and, as shown in Fig. 21.1, its position is made to coincide with the reference's origin. Also, the non-uniform dopant concentrations $N_A(x)$ and $N_D(x)$ are approximated by piecewise-constant functions, $N_A = \text{const}$ for $x < 0$ and $N_D = \text{const}$ for $x > 0$. The device obtained from this approximation is called *abrupt p–n junction*. Considering the actual form of the dopant distribution, the approximation is not realistic; however, the much simpler model based on it is still able to capture the essential features of the device characteristics. Moreover, the model assumes that the conditions of non-degeneracy and complete ionization hold; this assumption makes the analytical approach possible.

Within an integrated circuit the *p–n* junction is supplemented with contacts that connect it to the rest of the circuit. Such contacts are typically made of metals, metal silicides, or heavily-doped polycrystalline semiconductors; as a consequence, two more junctions are present: the first one is between the contact and the *p*-doped semiconductor, the other one between the contact and the *n*-doped semiconductor. It is implied that the contacts are made of the same material; if it is not so, more junctions must be considered as shown below.

21.2.1 Built-In Potential

A qualitative description of the device in the equilibrium condition starts from the assumption that the extension of the *p*-doped and *n*-doped regions along the x axis is large, so that, far away from the junction, the semiconductor can be considered as uniformly doped of the *p* or *n* type, respectively. This fixes the boundary conditions for the electron and hole concentrations:[2] in fact, remembering that in the non-degeneracy and complete-ionization conditions the equilibrium concentrations in a uniform semiconductor are given by (18.42) for the *p* type and by (18.30) for the *n* type, one finds

[1] The metallurgical junction is often indicated with the same term used for the whole device, namely, *p–n junction* or simply *junction*.

[2] The use of asymptotic conditions is not applicable to shallow junctions like, e.g., those used for the fabrication of solar cells. In this case, the theory is slightly more involved.

$$p_{p0} = p(-\infty) \simeq N_A, \qquad n_{p0} = n(-\infty) \simeq \frac{n_i^2}{N_A}, \tag{21.1}$$

$$n_{n0} = n(+\infty) \simeq N_D, \qquad p_{n0} = p(+\infty) \simeq \frac{n_i^2}{N_D}. \tag{21.2}$$

The above concentrations are also called *asymptotic concentrations*; the last approximations are derived from the assumption $N_A, N_D \gg n_i$ which, as outlined in Sect. 18.4.1, has a vast range of validity. The distance between the conduction-band edge and the Fermi level is found from $n = N_C \exp[-(E_C - E_F)/(k_B T)]$ (compare with (18.28)); combining with (21.1) and (21.2) yields

$$\frac{n_i^2}{N_A} \simeq N_C \exp\left[-\frac{E_C(-\infty) - E_F}{k_B T}\right], \quad N_D \simeq N_C \exp\left[-\frac{E_C(+\infty) - E_F}{k_B T}\right], \tag{21.3}$$

whence

$$E_C(-\infty) - E_C(+\infty) = k_B T \log\left(\frac{N_A N_D}{k_B T}\right). \tag{21.4}$$

An identical expression is found for $E_V(-\infty) - E_V(+\infty)$. These findings show that E_C, E_V are functions of position; their explicit form is determined below. Alternatively, one may use (18.60) and (18.61), to find

$$\psi_0 = k_B T \log\left(\frac{N_A N_D}{k_B T}\right), \qquad \psi_0 = \varphi(+\infty) - \varphi(-\infty), \tag{21.5}$$

where ψ_0 is called *built-in potential*.[3] One notes that so far the values of the constants $n^{(0)}$, $p^{(0)}$ in (18.60) and (18.61) have been left unspecified; remembering that $n^{(0)}$ is the value of n in the position(s) where $\varphi = 0$, and the same for $p^{(0)}$, the numerical values sought are determined by specifying the zero point of φ. Here such a point is fixed by letting $\varphi(+\infty) = 0$ whence, using (21.3) and (21.4), one finds

$$p^{(0)} = p(+\infty) = p_{n0}, \qquad n^{(0)} = n(+\infty) = n_{n0}. \tag{21.6}$$

The expressions of the carrier concentrations in terms of the band energies E_C, E_V must be coherent with those expressed in terms of φ. In fact, the relation between $E_C(x)$ and $\varphi(x)$ is found by combining $n = n^{(0)} \exp[q\,\varphi/(k_B T)]$ with $n = N_C \exp[(E_F - E_C)/(k_B T)]$ and using $n^{(0)} = N_C \exp\{[E_F - E_C(+\infty)]/(k_B T)\}$; a similar procedure is applied to E_V, to eventually find

$$E_C(x) = E_C(+\infty) - q\,\varphi(x), \qquad E_V(x) = E_V(+\infty) - q\,\varphi(x). \tag{21.7}$$

[3] The same quantity is also called *barrier potential* and is sometimes indicated with ψ_B.

Letting $N(x) = -N_A$ for $x < 0$ and $N(x) = +N_D$ for $x > 0$, the electric potential is found by solving the Poisson equation

$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}x^2} = \frac{q}{\varepsilon_{\mathrm{sc}}}\left[n_{n0}\,\exp\left(\frac{q\varphi}{k_B\,T}\right) - p_{n0}\,\exp\left(\frac{-q\varphi}{k_B\,T}\right) - N(x)\right] \tag{21.8}$$

with boundary conditions $\varphi(-\infty) = -\psi_0$ and $\varphi(+\infty) = 0$. One notes that within each half domain the charge density in (21.8) has the form $\rho = \rho(\varphi)$, which makes the theory of Sect. 19.5.8 applicable. Therefore, it is convenient to separately solve (21.8) in each half space, and apply suitable matching conditions at $x = 0$ afterwards. When the regional-solution method is used, the boundary conditions must be modified with respect to $\varphi(-\infty) = -\psi_0$ and $\varphi(+\infty) = 0$; the new conditions are shown below. In the n-doped region the charge density reads $\rho = q(p - n + N_D)$; when $x \to +\infty$ the latter becomes $0 = p_{n0} - n_{n0} - N_D$: in fact, as at large distances from the origin the material behaves like a uniformly-doped semiconductor, local charge neutrality is fulfilled at infinity. Using the dimensionless potential $u = q\,\varphi/(k_B\,T)$, and indicating the derivatives with primes, gives the equation the form

$$u'' = \frac{1}{L_D^2}\,A_D(u), \quad A_D = \exp(u) - 1 + \frac{p_{n0}}{n_{n0}}\,[1 - \exp(-u)], \quad L_D^2 = \frac{\varepsilon_{\mathrm{sc}}\,k_B\,T}{q^2\,n_{n0}}, \tag{21.9}$$

with L_D the *Debye length for the electrons*. The normalized charge density A_D vanishes for $u = 0$, and is positive (negative) when u is positive (negative). Note that, by letting $x \to +\infty$ in $n = n_i\,\exp[(\varphi - \varphi_F)/(k_B\,T)]$, $p = n_i\,\exp[(\varphi_F - \varphi)/(k_B\,T)]$, and using the normalized Fermi potential $u_F = q\,\varphi_F/(k_B\,T)$, one finds $p_{n0}/n_{n0} = (n_i/N_D)^2 = \exp(2\,u_F)$, where $u_F < 0$ on account of the fact that here an n-doped region is considered. Following the method illustrated in Sect. 19.5.8 transforms the left hand side of (21.8) into $u''\,u' = (1/2)\,[(u')^2]'$. This term is then integrated from $x = 0$ to $x = +\infty$; the result is simplified by observing that the region far from the junction is substantially uniform, whence the electric potential is constant there. As u' is proportional to the electric field, it follows that $u'(+\infty) = 0$: this is the second boundary condition to be added to $u(+\infty) = 0$. In conclusion, the integration of (21.9) from $x \geq 0$ to $+\infty$ yields

$$\left(u'\right)^2 = \frac{2}{L_D^2}\,B_D(u), \quad B_D = \exp(u) - 1 - u + \frac{n_i^2}{N_D^2}\,[u + \exp(-u) - 1]. \tag{21.10}$$

It is easily found that B_D is non negative; in fact, from (21.10) one derives $B_D = 0$ for $u = 0$, and at the same time $\mathrm{d}B_D/\mathrm{d}u = A_D(u)$ is positive for $u > 0$, negative for $u < 0$; as a consequence, B_D grows from zero in either direction when u departs from the origin. Letting $F_D^2 = B_D$ one finds $|u'| = \sqrt{2}\,F_D/L_D$; as the condition $u = 0$ holds only asymptotically, the modulus of u' always grows as u departs from the origin, showing that u is monotonic. Considering that u must fulfill the other boundary condition $u(-\infty) = -q\,\psi_0/(k_B\,T) < 0$, one concludes that u is a

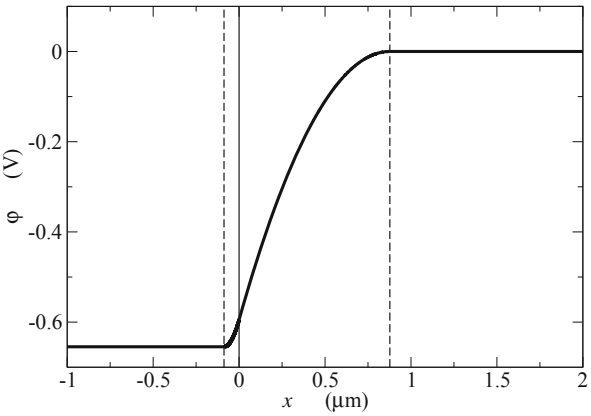

Fig. 21.2 Solution of the one-dimensional Poisson equation 21.8 in an abrupt *p–n* junction at equilibrium, with $N_A = 10^{16}$ cm^{-3}, $N_D = 10^{15}$ cm^{-3}. The *continuous vertical line* marks the position of the metallurgical junction, the *dashed vertical lines* mark the edges of the space-charge region

monotonically-growing function, whence, choosing the positive sign and separating the variables, one finds

$$\frac{\mathrm{d}u}{F_D(u)} = \frac{\sqrt{2}}{L_D}\,\mathrm{d}x \tag{21.11}$$

for $x \geq 0$. The above must be tackled numerically because it has no analytical solution.[4] Once the normalized potential is found from (21.11), its value in the origin, $u(x = 0)$, along with that of the derivative $u'(x = 0) = \sqrt{2}\,F_D[u(x = 0)]/L_D$, provide the boundary conditions for the solution in the *p*-doped region.

The solution for $x < 0$ follows the same pattern, where the asymptotic neutrality condition reads $n_{n0}\exp(-\chi_0) - p_{n0}\exp(\chi_0) + N_A = 0$, with $\chi_0 = q\,\psi_0/(k_B\,T)$. Letting $v = u + \chi_0$, $n_{p0}/p_{p0} = (n_i/N_A)^2 = \exp(-2\,u_F) \ll 1$, and using (21.5) provides

$$v'' = \frac{1}{L_A^2}\,A_A(v), \quad A_A = \frac{n_i^2}{N_A^2}\,[\exp(v) - 1] + 1 - \exp(-v), \quad L_A^2 = \frac{\varepsilon_{\mathrm{sc}}\,k_B\,T}{q^2\,p_{p0}}, \tag{21.12}$$

where L_A is the *Debye length for the holes*. The rest of the procedure is similar to that used in the *n*-doped region.

21.2.2 Space-Charge and Quasi-Neutral Regions

The form of the electric potential φ is shown in Fig. 21.2 for a *p–n* junction at equilibrium with $N_A = 10^{16}$ cm^{-3}, $N_D = 10^{15}$ cm^{-3}. The form of the bands is

[4] The numerical procedure is outlined in the note of Sect. 22.2.1.

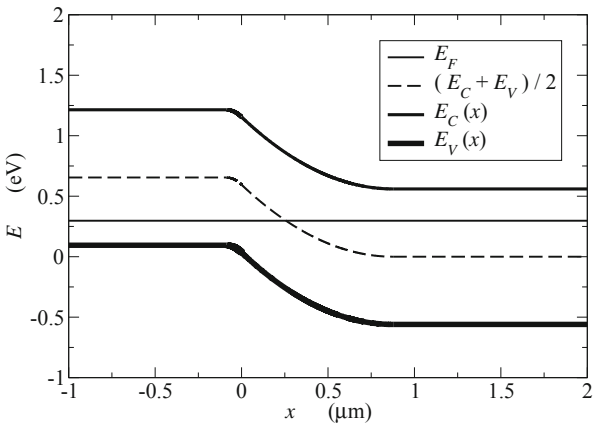

Fig. 21.3 Form of the bands for the same device as in Fig. 21.2

shown in Fig. 21.3 for the same device. It is interesting to note that the device can be thought of as made of three regions: in the intermediate region, whose boundaries are marked by dashed vertical lines in Fig. 21.2, the electric potential has a non-negligible curvature, this showing that the charge density is large. The region is called *space-charge region*, and contains the metallurgical junction, marked by the continuous vertical line. In the two regions on the sides of the space-charge region, the electric potential is nearly constant,[5] whence the electric field $-\mathrm{d}\varphi/\mathrm{d}x$ is negligibly small. As a consequence, the charge density $\rho = -\varepsilon_{\mathrm{sc}}\,\mathrm{d}^2\varphi/\mathrm{d}x^2$ is negligible as well; for this reason, the two regions under consideration are called *quasi-neutral regions*.[6]

The transition from the space-charge region and one or the other quasi-neutral region is sharp. Thanks to this, it is possible to identify the width l of the space-charge region and correlate it with other parameters of the device; such a correlation is worked out in Sect. 21.4 in a specific operating regime. For convenience, the width of the space-charge region is expressed as $l = l_p + l_n$, where l_p is the extension of the space-charge region on the p side of the metallurgical junction, and l_n the analogue on the n side. As shown in Fig. 21.2, it is $l_n > l_p$; as explained in Sect. 21.4, this is due to the global charge neutrality and to the fact that $N_D < N_A$.

If the equilibrium carrier concentrations corresponding to the electric potential of Fig. 21.2 are drawn in a logarithmic scale, the curves look similar to that of Fig. 21.2, apart from scaling factors and from the inversion due to the negative sign in $p = p_{n0}\,\exp[-q\,\varphi/(k_B\,T)]$. This is shown in Fig. 21.4. A more realistic representation, shown in Fig. 21.5, uses a linear scale. The hole concentration p ranges from $p_{p0} \simeq N_A = 10^{16}\ \mathrm{cm}^{-3}$ in the p-type quasi-neutral region to $p_{n0} \simeq n_i^2/N_D = 10^5\ \mathrm{cm}^{-3}$ in

[5] The electric potential can not be exactly constant, because the solution of (21.8) is an analytical function; as a consequence, if φ were constant in a finite interval, it would be constant everywhere.

[6] The inverse reasoning would not be correct: in fact, $\rho = 0$ may yield $\mathrm{d}\varphi/\mathrm{d}x = \mathrm{const} \neq 0$, which makes φ a linear function of x; deducing $\varphi = \mathrm{const}$ from $\rho = 0$ is correct only if the additional condition of spatial uniformity holds.

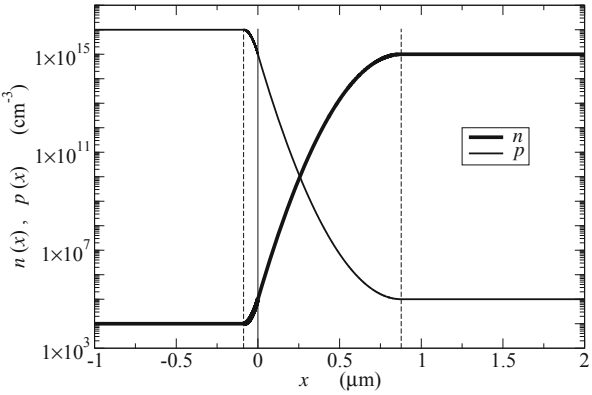

Fig. 21.4 Electron and hole concentrations in a one-dimensional, abrupt *p–n* junction at equilibrium, with $N_A = 10^{16}$ cm^{-3}, $N_D = 10^{15}$ cm^{-3}. The figure is drawn in a logarithmic scale. The *continuous vertical line* marks the position of the metallurgical junction, the *dashed vertical lines* mark the edges of the space-charge region

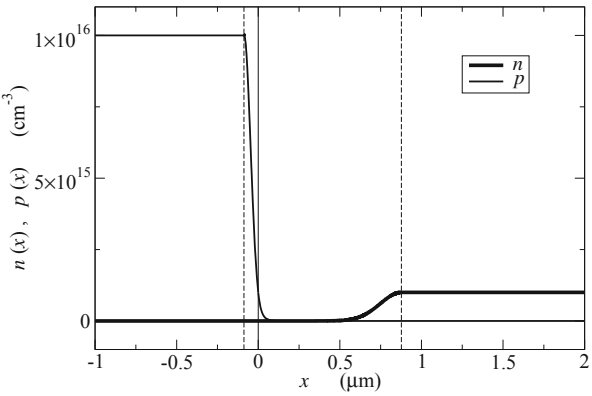

Fig. 21.5 The same concentrations as in Fig. 21.4, drawn in a linear scale

the *n*-type quasi-neutral region; similarly, the electron concentration n ranges from $n_{p0} \simeq n_i^2/N_A = 10^4$ cm^{-3} in the *p*-type quasi-neutral region to $n_{n0} \simeq N_D = 10^{15}$ cm^{-3} in the *n*-type quasi-neutral region. This shows that the two concentrations vary by several orders of magnitude over the space-charge region, whose length is about 1 μm; for this reason, if these gradients existed alone, they would make holes (electrons) to diffuse in the positive (negative) direction of the x axis. Such diffusions in fact do not occur, because in the equilibrium condition they are balanced by the electric field. The latter is negative: in fact, the electric potential increases with x, so that the force associated to the electric field E is negative (positive) for holes (electrons); the equations for the current densities of the semiconductor device model in (19.129) and (19.130) yield in this case

$$-q\mu_p\, pE = -q D_p \frac{\mathrm{d}p}{\mathrm{d}x} > 0, \qquad -q\mu_n\, n\, E = +q D_n \frac{\mathrm{d}n}{\mathrm{d}x} > 0, \tag{21.13}$$

whence $J_p = J_n = 0$. The description of the *p–n* junction in the equilibrium condition is completed by adding the contacts; as indicated above, this amounts to introducing two more junctions. The contacts are made of materials different from

the semiconductor of which the *p–n* junction is made, hence the atomic structure of the contact's material must adapt to that of the semiconductor when the contact is deposited on it; for this reason, the structure of the contact-semiconductor junction must be described on a case-by-case basis. From the qualitative standpoint, one can use the analogy with the *p–n* junction to deduce the existence of a built-in potential Φ_{mp} between the contact and the *p*-type semiconductor, and of another built-in potential Φ_{mn} between the contact and the *n*-type semiconductor. The built-in potentials are influenced by the dopant concentration of the semiconductor, namely, $\Phi_{mp} = \Phi_{mp}(N_A)$ and $\Phi_{mn} = \Phi_{mn}(N_D)$. Assume that the contacts are made of the same material; if they are short-circuited, a closed loop is formed where, from Kirchhoff's voltage law, the built-in potentials fulfill the relation

$$\psi_0 + \Phi_{mn} - \Phi_{mp} = 0. \tag{21.14}$$

This situation is schematically illustrated in Fig. 21.6, where it is assumed that the material of the contacts is the same.[7] In the figure, the built-in potentials at the contacts are represented by discontinuities; in the practical cases, in fact, to prevent the contact-semiconductor junction from behaving like a rectifying device, a heavy dose of dopant is preliminarily introduced into the semiconductor region onto which the contact is to be deposited. For this reason, the spatial extension where Φ_{mp} or Φ_{mn} occurs is negligibly small with respect to the typical scale length of the device. Regardless of this, another important outcome of the fabrication process mentioned above is that the concentration of carriers available in the materials forming a contact is very large; for this reason, as mentioned in Sect. 19.5.6, a contact is able to supply the amount of charge necessary to keep the equilibrium and charge-neutrality conditions in the semiconductor layer adjacent to it. This also implies that, within some limits to be specified later, in a non-equilibrium condition the built-in potential is practically the same as in equilibrium. In circuit theory, a contact whose built-in potential is independent of the current that crosses it is called *ideal Ohmic contact*; this condition is equivalent to that of a vanishing differential resistivity of the contact.

21.3 Shockley Theory of the *P–N* Junction

The analytical derivation of the current-voltage characteristic of the *p–n* junction is based on the hypothesis that the device is not too far from equilibrium, so that the weak-injection condition (20.35) is fulfilled in the quasi-neutral regions. Within this limit, the approximation that the contacts are ideal is acceptable. A non-equilibrium condition is obtained by applying, e.g., a bias voltage V between the contacts; if V is such that the electric potential at the contact of the *p* region is higher than that of the *n* region, the condition is called *forward bias*; when it is lower, the condition is called *reverse bias*. Fig. 21.7 shows the symbol of the *p–n* junction used in circuit

[7] If it is not so, one must add to the left hand side of (21.14) the barrier between the two materials.

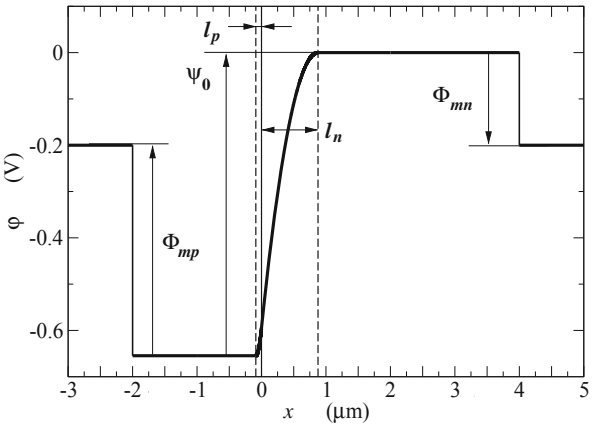

Fig. 21.6 Electric potential for the same device as in Fig. 21.2, including the built-in potentials of the contacts

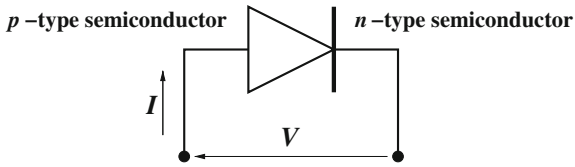

Fig. 21.7 Symbol and typical I, V reference for the p–n junction

theory, along with the standard references for the applied voltage V and current I. In a one-dimensional case the current is given by $I = A_e J$, with A_e the device cross-sectional area and J the total current density.

Numerical solutions of the semiconductor-device model show that in the weak-injection condition the partitioning of the device into space-charge and quasi-neutral regions still holds; this implies that, when a bias voltage is applied between the contacts, such a voltage adds algebraically to the built-in potential. In fact, the discontinuities at the contacts are the same as in the equilibrium condition due to the contacts' ideality, and the electric potential in the quasi-neutral regions is nearly constant; the extension of the space-charge region, instead, changes due to the application of the external voltage, $l_p = l_p(V)$, $l_n = l_n(V)$. When a forward bias is applied, l_p and l_n slightly decrease, as qualitatively shown in Fig. 21.8 (the drawing is not in the same scale as Fig. 21.6, and is meant only to show the change in l; the solution in the space-charge region is omitted). The same applies to Fig. 21.9, that refers to the reverse bias and shows that in this case l_p and l_n increase. Using (21.14), the application of Kirchhoff's voltage law to either case yields, for the voltage drop ψ across the space-charge region, the expression

$$\psi + \Phi_{mn} + V - \Phi_{mp} = 0, \qquad \psi + V - \psi_0 = 0. \tag{21.15}$$

In reverse bias ($V < 0$) it is always $\psi > \psi_0 > 0$; in forward bias ($V > 0$) a sufficiently large value of V in (21.15) could make ψ to become negative. However, when V becomes large the weak-injection condition does not hold any more, and

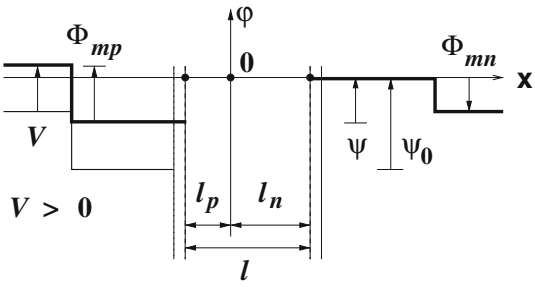

Fig. 21.8 Schematic description of the change in the extension l of the space-charge region in a forward-biased p–n junction ($V > 0$). The *thin lines* refer to the equilibrium case. The drawing is not in the same scale as Fig. 21.6

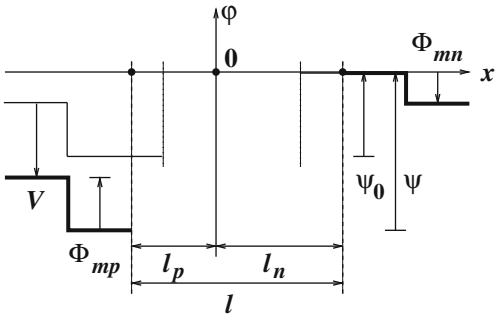

Fig. 21.9 Schematic description of the change in the extension l of the space-charge region in a reverse-biased p–n junction ($V < 0$). The *thin lines* refer to the equilibrium case. The drawing is not in the same scale as Fig. 21.6

(21.15) does not apply; in conclusion, the range of forward biases to be considered here is such that the condition $\psi_0 > \psi > 0$ is always fulfilled.

When a forward bias is applied, due to $\psi < \psi_0$ the electric field within the space-charge region decreases with respect to the equilibrium case;[8] thus, the drift term in the drift-diffusion equations of (19.129) and (19.130) becomes weaker. The diffusion term, in contrast, becomes slightly stronger, because the values of the electron concentrations in the quasi-neutral regions are fixed by the asymptotic conditions, and the width of the space-charge region slightly decreases with respect to the equilibrium case. In conclusion, the diffusion term prevails and the current-density equations yield

$$-q\,D_p\,\frac{\mathrm{d}p}{\mathrm{d}x} > -q\,\mu_p\,p\,E > 0, \qquad q\,D_n\,\frac{\mathrm{d}n}{\mathrm{d}x} > -q\,\mu_n\,n\,E > 0, \tag{21.16}$$

so that $\mathbf{J}_p \cdot \mathbf{i} = qp\,v_p > 0$ and $\mathbf{J}_n \cdot \mathbf{i} = -qn\,v_n > 0$. The total current density $(\mathbf{J}_p + \mathbf{J}_n) \cdot \mathbf{i}$ is positive as well.

When a reverse bias is applied, due to $\psi > \psi_0$ the voltage drop across the space-charge region increases with respect to the equilibrium case. The region's width increases as well; however, the increase in l is relatively weak, whence the

[8] The width of the space-charge region decreases as well (Fig. 21.8); such a decrease, however, is small, and does not compensate for the decrease in the potential drop.

electric field within the space-charge region increases and the drift term in the drift-diffusion equations of (19.129) and (19.130) becomes stronger. The diffusion term, in contrast, becomes weaker, because the values of the electron concentrations in the quasi-neutral regions are fixed by the asymptotic conditions. In conclusion, the drift term prevails and the current-density equations yield

$$-q\mu_p\, pE > -qD_p \frac{\mathrm{d}p}{\mathrm{d}x} > 0, \qquad -q\mu_n\, nE > qD_n \frac{\mathrm{d}n}{\mathrm{d}x} > 0, \tag{21.17}$$

so that $\mathbf{J}_p \cdot \mathbf{i} = qp\, v_p < 0$, $\mathbf{J}_n \cdot \mathbf{i} = -qn\, v_n < 0$. The total current density is negative as well.

The $I(V)$ relation of the *p–n* junction is worked out here in the one-dimensional and steady-state case, this leading to the *Shockley equations* [97, 98]. The steady-state form $\mathrm{div}\mathbf{J} = 0$ of the continuity equation (4.23) reduces in one dimension to $\mathrm{d}J/\mathrm{d}x = 0$, whence

$$J = J_p(x) + J_n(x) = \text{const.} \tag{21.18}$$

The hole and electron current densities depend on position and fulfill the steady-state continuity equations of (19.129) and (19.130); in the latter, only the net thermal recombination term is considered, to find

$$\frac{\mathrm{d}J_p}{\mathrm{d}x} = -qU_{\mathrm{SRH}}, \qquad \frac{\mathrm{d}J_n}{\mathrm{d}x} = qU_{\mathrm{SRH}}. \tag{21.19}$$

Once J_p, J_n are determined from (21.19), they are specified at a suitable position and added up. As shown below, such positions are the boundaries $-l_p$ and l_n between the space-charge and quasi-neutral regions, for instance, $J = J_p(-l_p) + J_n(-l_p)$. Observing that $-l_p$ is the boundary of the *p*-type region, $J_p(-l_p)$ is a majority-carrier current density, whereas $J_n(-l_p)$ is a minority-carrier current density. The opposite happens if the other boundary is chosen, to yield $J = J_p(l_n) + J_n(l_n)$. To proceed, it is convenient to seek for an expression of J where both current densities refer to minority carriers; this is achieved by integrating (21.19) over the space-charge region, to define the *recombination current density*

$$J_U = \int_{-l_p}^{l_n} qU_{\mathrm{SRH}}\ \mathrm{d}x = J_p(-l_p) - J_p(l_n) = J_n(l_n) - J_n(-l_p). \tag{21.20}$$

Combining (21.20) with the expression of the total current density at, e.g., l_n provides

$$J = J_p(l_n) + J_n(l_n) = J_p(l_n) + J_n(-l_p) + J_U, \tag{21.21}$$

which has the desired form.

21.3.1 Derivation of the $I(V)$ Characteristic

Remembering the simplified form (20.38) or (20.40) of the net thermal-recombination term in the weak-injection condition, and using $p^{\mathrm{eq}} = p_{n0}$ in the n-type region and $n^{\mathrm{eq}} = n_{p0}$ in the p-type region, transforms (21.19) into, respectively,

$$\frac{\mathrm{d}J_p}{\mathrm{d}x} = q\frac{p - p_{n0}}{\tau_p}, \qquad x > l_n, \tag{21.22}$$

$$\frac{\mathrm{d}J_n}{\mathrm{d}x} = q\frac{n - n_{p0}}{\tau_n}, \qquad x < -l_p. \tag{21.23}$$

In this way, the hole- and electron-continuity equations are decoupled from each other; also, they are to be solved over disjoint intervals. To the current-continuity equations one associates the corresponding drift-diffusion equation taken from (19.129) or (19.130); it follows that in each quasi-neutral region the drift-diffusion equation is that of the minority carriers. It can be shown that in a quasi-neutral region, when the weak-injection condition holds, the diffusion term of the minority carries dominates over the drift term (the details are worked out in Sect. 21.6.1). Thus, for $x < -l_p$ (p-type region), $J_n = q\,\mu_n\,n\,E + q\,D_n\,\mathrm{d}p/\mathrm{d}x \simeq q\,D_n\,\mathrm{d}p/\mathrm{d}x$. Inserting the latter into (21.23) yields $D_n\,\mathrm{d}^2n/\mathrm{d}x^2 = (n - n_{p0})/\tau_n$. In this derivation the diffusion coefficient $D_n = k_B\,T\,\mu_n/q$ is not subjected to the derivative; in fact, the two parameters that influence bulk mobility, T and N_A (Sect. 20.5), are independent of position. The equation then reads,

$$\frac{\mathrm{d}^2(n - n_{p0})}{\mathrm{d}x^2} = \frac{n - n_{p0}}{L_n}, \qquad L_n = \sqrt{\tau_n\,D_n}, \tag{21.24}$$

with L_n the *diffusion length* of the minority carriers in the p-type region. It is a second-order, linear equation in the unknown n; it is decoupled from the rest of the semiconductor-device model: in fact, the simplified form of the net-recombination term contains only the electron concentration, and the neglect of the drift term eliminates the coupling with the Poisson equation. The boundary conditions must be fixed at $x \to -\infty$ and $x = -l_p$; the former is $n(-\infty) = n_{p0}$, whereas the latter needs a more elaborate derivation, given in Sect. 21.6.2, whose outcome (also called *Shockley's boundary condition*) is $n(-l_p) = n_{p0}\,\exp[q\,V/(k_B\,T)]$. The general solution of (21.24) is

$$n = n_{p0} + A_n\,\exp(x/L_n) + B_n\,\exp(-x/L_n), \tag{21.25}$$

whence the asymptotic boundary condition yields $B_n = 0$. The other boundary condition provides $n(-l_p) = n_{p0} + A_n^-$, with $A_n^- = A_n\,\exp(-l_p/L_n)$. The electron current density is then found from

$$J_n = q\,D_n\,\frac{\mathrm{d}n}{\mathrm{d}x} = q\,\frac{D_n}{L_n}\,A_n\,\exp(x/L_n) = J_n(-l_p)\,\exp[(x + l_p)/L_n], \tag{21.26}$$

where, using $n(-l_p) = n_{p0} + A_n^-$ and the boundary condition at $x = -l_p$,

$$J_n(-l_p) = \frac{q D_n A_n^-}{L_n} = \frac{q D_n n_{p0}}{L_n} F, \qquad F(V) = \exp[qV/(k_B T)] - 1. \quad (21.27)$$

In the same manner one finds

$$J_p(l_n) = \frac{q D_p p_{n0}}{L_p} F, \qquad L_p = \sqrt{\tau_p D_p}, \quad (21.28)$$

where F is the same as in (21.27), and L_p is the diffusion length of the minority carriers in the n-type region. Inserting (21.27) and (21.28) into (21.21) yields the total current density,

$$J = J_p(l_n) + J_n(-l_p) + J_U = q\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right) F + J_U. \quad (21.29)$$

Multiplying (21.29) by the cross-sectional area A_e and defining the *saturation current density*

$$J_s = q\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right) = q n_i^2 \left(\frac{\sqrt{D_p/\tau_p}}{N_D} + \frac{\sqrt{D_n/\tau_n}}{N_A}\right), \quad (21.30)$$

yields the expression of the $I(V)$ characteristic of the p–n junction:

$$I = I_s \left[\exp\left(\frac{qV}{k_B T}\right) - 1\right] + I_U, \qquad I_s = A_e J_s, \quad I_U = A_e J_U. \quad (21.31)$$

The characteristic fulfills the equilibrium condition $I(0) = 0$; in fact, at equilibrium it is $U_{\mathrm{SRH}} = 0$. When $qV/k_B T \gg 1$, the exponential term in (21.31) prevails over the other terms and the characteristic becomes $I \simeq I_s \exp[qV/(k_B T)]$, namely, the well-known exponential form of the forward-bias case. Finally, when $qV/k_B T \ll -1$, the current becomes $I \simeq -I_s + I_U$. As the order of magnitude of I_s may be similar to that of I_U, it is necessary to calculate the latter explicitly. The analysis is made easier by the observation that the electric field (which in the reverse-bias condition prevails over diffusion) drains the holes from the space-charge region to the p-type quasi-neutral region; similarly, the electrons of the space-charge region are drained towards the n-type quasi-neutral region. As a consequence, the carrier concentrations in the space-charge region are negligible, and the full-depletion condition (20.34) applies there; using the non-degenerate expression one finds, for the reverse-bias current,

$$I \simeq -I_s + I_U \simeq -I_s + A_e \int_{-l_p}^{l_n} q \frac{-n_i}{\tau_g} \, \mathrm{d}x = -I_s - q A_e \frac{n_i}{\tau_g} \, l(V) < 0, \quad (21.32)$$

where the expression (21.20) of the recombination current density has been used, and $l = l(V)$ is the width of the space-charge region. As shown in Sect. 21.4, in

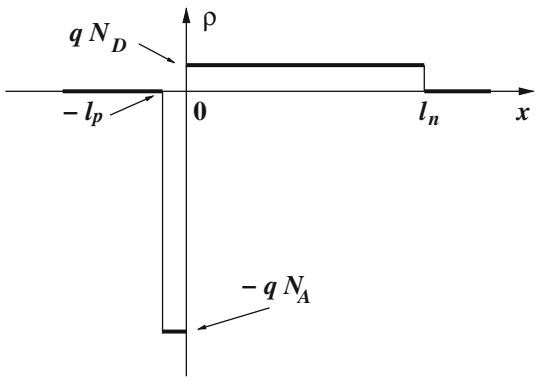

Fig. 21.10 Charge density in a reverse-biased p–n junction using the ASCE approximation, in arbitrary units. The ratio N_A/N_D is the same as in Fig. 21.6

the reverse-bias condition and for an abrupt junction it is $l \propto \sqrt{\psi_0 + |V|}$; as a consequence, I_U increases with $|V|$.

The approximations that have been introduced to derive the $I(V)$ characteristic are many; in fact, (21.31) is referred to as the *ideal characteristic*. However, it captures quite well the general behavior of the device as long as the applied voltage is within the limit of the weak-injection approximation. When the forward bias exceeds such a limit, the drift term is not negligible anymore and the effect of the electric field in the quasi-neutral regions must be accounted for. Considering in turn the reverse-bias condition, at large values of $|V|$ the electric field in the space-charge region becomes sufficiently strong to induce impact ionization (Sect. 20.3) and, possibly, the junction breakdown due to avalanche (Sect. 21.5).

The dependence of the $I(V)$ characteristic on temperature is due, besides that of $q\,V/(k_B\,T)$, to the coefficients of J_s and J_U. To this purpose, the second form of (21.30) is more useful because it shows explicitly the term n_i^2, whose temperature dependence is exponential (18.14); in fact, the temperature dependence of D_p/τ_p, D_n/τ_n is much weaker. The same considerations apply to J_U, whose main dependence on temperature is due to factor n_i. The dependence on n_i of the reverse current (at constant temperature) has been used in the considerations about the parasitic currents in integrated circuits made in Sect. 18.7.

21.4 Depletion Capacitance of the Abrupt *P–N* Junction

It has been anticipated in Sect. 21.3.1 that, when a reverse bias is applied to the junction, the full-depletion condition holds in the space-charge region; as a consequence, the charge density ρ in the latter is essentially due to the dopant atoms. In the abrupt junction considered here, the dopants' concentration is piecewise constant; a simplified description of the charge density in the situation in hand is obtained from the *abrupt space-charge edge* (ASCE) approximation, which describes ρ with the form sketched in Fig. 21.10. In other terms, the approximation consists in replacing

with a discontinuity the smooth change of ρ at $x = -l_p$, $x = 0$, and $x = l_n$; thus, the space charge is given by

$$\rho = -qN_A, \qquad -l_p < x < 0, \tag{21.33}$$

$$\rho = qN_D, \qquad 0 < x < l_n, \tag{21.34}$$

and $\rho = 0$ elsewhere. The electric field and the electric potential are continuous because there are no charge layers or double layers; letting $E_0 = E(0)$, from $\mathrm{d}E/\mathrm{d}x = \rho/\varepsilon_{sc}$ and (21.33), (21.34) one draws

$$\frac{E_0 - E(-l_p)}{l_p} = -\frac{qN_A}{\varepsilon_{sc}}, \qquad \frac{E(l_n) - E_0}{l_n} = \frac{qN_D}{\varepsilon_{sc}}, \tag{21.35}$$

the first of which holds for $-l_p \le x \le 0$, the second one for $0 \le x \le l_n$. In the quasi-neutral regions the field is negligible; in the order of approximation used here one lets $E = 0$ in such regions, whence $E(-l_p) = E(l_n) = 0$ and, from (21.35),

$$E_0 = -\frac{qN_D}{\varepsilon_{sc}}\, l_n = -\frac{qN_A}{\varepsilon_{sc}}\, l_p < 0, \qquad N_D\, l_n = N_A\, l_p. \tag{21.36}$$

In conclusion, the electric field is a piecewise-linear function whose form is shown in Fig. 21.11. Due to $\mathrm{d}\varphi/\mathrm{d}x = -E$, the integral of $-E$ over the space-charge region equals the potential drop ψ:

$$\psi = \varphi(l_n) - \varphi(-l_p) = -\int_{-l_p}^{+l_n} E\ \mathrm{d}x = -\frac{1}{2}\, E_0\, (l_n + l_p). \tag{21.37}$$

Inserting into (21.37) one or the other form of E_0 from (21.36), one obtains two equivalent expressions for $l_p + l_n$:

$$l_n + l_p = l_n + \frac{N_D}{N_A}\, l_n = N_D\, l_n \left(\frac{1}{N_D} + \frac{1}{N_A}\right) = N_A\, l_p \left(\frac{1}{N_D} + \frac{1}{N_A}\right). \tag{21.38}$$

Then, combining (21.38) with (21.37) one finds

$$\psi = \frac{q}{2\,\varepsilon_{sc}} \left(\frac{1}{N_D} + \frac{1}{N_A}\right) (N_D l_n)^2 = \frac{q}{2\,\varepsilon_{sc}} \left(\frac{1}{N_D} + \frac{1}{N_A}\right) \left(N_A l_p\right)^2, \tag{21.39}$$

whence

$$l_n = \frac{1}{N_D} \left(\frac{2\,\varepsilon_{sc}\,\psi/q}{1/N_D + 1/N_A}\right)^{1/2}, \qquad l_p = \frac{1}{N_A} \left(\frac{2\,\varepsilon_{sc}\,\psi/q}{1/N_D + 1/N_A}\right)^{1/2}, \tag{21.40}$$

and

$$l = l_n + l_p = \left[\frac{2\,\varepsilon_{sc}}{q} \left(\frac{1}{N_D} + \frac{1}{N_A}\right) \psi\right]^{1/2}, \qquad \psi = \psi_0 - V. \tag{21.41}$$

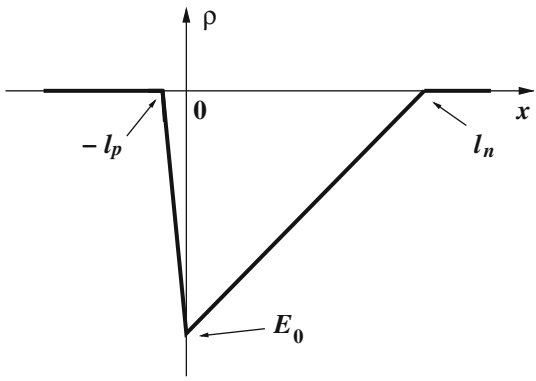

Fig. 21.11 Electric field consistent with the charge density of Fig. 21.10, in arbitrary units

Multiplying by qA_e both sides of the second relation in (21.36) yields $qN_D\,A_e\,l_n = qN_A\,A_e\,l_p$, that represents the global charge conservation in the device. Such a conservation is implied by the assumption that $E = 0$ in the quasi-neutral regions, as is found by integrating $\mathrm{div}\mathbf{D} = \rho$ over the space-charge region. In the charge-conservation relation the widths l_p, l_n depend on V through (21.40); it follows that in the reverse-bias condition the device can be assimilated to a non-linear capacitor where the charge per unit area of the two oppositely-charged sides is, respectively, $Q_p = -qN_A\,l_p$ and $Q_n = qN_D\,l_n$. The differential capacitance per unit area is defined as $C = \mathrm{d}Q_p/\mathrm{d}V = -\mathrm{d}Q_n/\mathrm{d}V$; from the definition,[9] two equivalent expressions follow,

$$C = -qN_A\,\frac{\mathrm{d}l_p}{\mathrm{d}V} = q\,\frac{\mathrm{d}(N_A\,l_p)}{\mathrm{d}\psi}, \qquad C = -qN_D\,\frac{\mathrm{d}l_n}{\mathrm{d}V} = q\,\frac{\mathrm{d}(N_D\,l_n)}{\mathrm{d}\psi}. \tag{21.42}$$

Using (21.40), the differential capacitance per unit area of the abrupt *p–n* junction is found to be

$$C = \left[\frac{q\varepsilon_{\mathrm{sc}}/(2\,\psi)}{1/N_D + 1/N_A}\right]^{1/2} = \frac{[(q\varepsilon_{\mathrm{sc}}/2)/(1/N_D + 1/N_A)]^{1/2}}{[\psi_0\,(1 - V/\psi_0)]^{1/2}}, \tag{21.43}$$

which is given the more compact form[10]

$$C = C_0\left(1 - \frac{V}{\psi_0}\right)^{-1/2}, \qquad C_0 = C(V = 0) = \left[\frac{q\varepsilon_{\mathrm{sc}}/(2\,\psi_0)}{1/N_D + 1/N_A}\right]^{1/2}. \tag{21.44}$$

[9] Definition $C = \mathrm{d}Q_p/\mathrm{d}V$ is coherent with the choice of the reference in Fig. 21.7. The units of C are $[C] = \mathrm{F\ cm^{-2}}$. Compare with the calculation of the MOS capacitance in Sect. 22.3.

[10] It is worth reminding that the result holds only in the reverse-bias condition. In the forward-bias condition the injection of carriers from the quasi-neutral regions into the space-charge region prevents one from neglecting the contribution of the carrier concentrations to the charge density and makes the use of (21.44) erroneous.

Combining (21.44) with (21.41) one derives the interesting relation

$$\frac{1}{C^2} = \frac{2}{q\varepsilon_{sc}} \left(\frac{1}{N_D} + \frac{1}{N_A} \right) \psi = \frac{l^2}{\varepsilon_{sc}^2}, \qquad C = \frac{\varepsilon_{sc}}{l}, \tag{21.45}$$

namely, the standard expression for the capacitance per unit area of the parallel-plate capacitor. Such an expression is not limited to the case where the dopant concentration is piecewise constant; as shown in Sect. 21.6.3, it applies in fact to all cases.

From the standpoint of circuit design, the capacitance associated to a *p–n* junction is a parasitic effect that hampers the circuit's speed. However, the effect is also exploited to manufacture voltage-controlled capacitors, called *variable capacitors* or *varactors*. In these devices, that are operated in reverse bias, the geometry is designed to maximize the capacitance; they are used, e.g., in voltage-controlled oscillators, parametric amplifiers, and frequency modulation.[11] The bias range of these devices is such that the reverse current is negligibly small; if the modulus of the reverse bias is made to increase, and is eventually brought outside this range, carrier multiplication due to impact ionization (Sect. 20.3) takes place; this, as shown in Sect. 21.5, leads to a strong increase of the reverse current.

21.5 Avalanche Due to Impact Ionization

The situation where impact ionization dominates over the other generation-recombination mechanisms has been illustrated in Sect. 20.3.1, showing that in the steady-state case the continuity equations for electrons and holes reduce to (20.45). Such equations are applicable, for instance, to the space-charge region of a reverse-biased *p–n* junction; when the value of $|V|$ becomes large, the increase in the number of carriers due to impact ionization may give rise to an avalanche phenomenon that eventually leads to the junction's *avalanche breakdown*, namely, a strong increase in the current due to carrier multiplication.[12] The absolute value V_B of the voltage at which the phenomenon occurs is called *breakdown voltage*. To illustrate avalanche, the one-dimensional case is considered, so that (20.45) become

$$\frac{\mathrm{d}J_n}{\mathrm{d}x} = k_n J_n + k_p J_p, \qquad \frac{\mathrm{d}J_p}{\mathrm{d}x} = -k_n J_n - k_p J_p, \tag{21.46}$$

[11] Varactors are also manufactured using technologies other than the bipolar one; e.g., with MOS capacitors or metal-semiconductor junctions.

[12] If the breakdown is accompanied by current crowding, the junction may be destroyed due to excessive heating. Special *p–n* junctions, called *avalanche diodes*, are designed to have breakdown uniformly spread over the surface of the metallurgical junction, to avoid current crowding. Such devices are able to indefinitely sustain the breakdown condition; they are used as voltage reference and for protecting electronic circuits against excessively-high voltages.

where the impact-ionization coefficients k_n, k_p depend on x through the electric field E and are determined experimentally.[13] To ease the notation the boundaries of the space-charge region are indicated with a, b; also, considering the reference's orientation (Fig. 21.7), it is $E, J_n, J_p < 0$, where the electric field is significant only for $a \le x \le b$. As in the one-dimensional and steady-state case it is $J = J_n(x) + J_p(x) = \text{const}$, eliminating J_p from the first equation in (21.46) yields

$$\frac{\mathrm{d}J_n}{\mathrm{d}x} = k_n\, J_n + k_p\,(J - J_n) = (k_n - k_p)\, J_n + k_p\, J, \tag{21.47}$$

namely, a first-order equation in J_n containing the yet undetermined parameter J. The equation is recast as

$$\frac{\mathrm{d}J_n}{\mathrm{d}x} - \frac{\mathrm{d}m}{\mathrm{d}x}\, J_n = k_p\, J = k_n\, J - \frac{\mathrm{d}m}{\mathrm{d}x}\, J, \qquad m = \int_a^x (k_n - k_p)\, \mathrm{d}x' \tag{21.48}$$

where $m(a) = 0$, $\mathrm{d}m/\mathrm{d}x = k_n - k_p$. Multiplying by the integrating factor $\exp(-m)$ and dividing by J transforms (21.48) into

$$\frac{1}{J}\,\frac{\mathrm{d}}{\mathrm{d}x}\left[J_n\, \exp(-m)\right] = k_n\, \exp(-m) - \frac{\mathrm{d}m}{\mathrm{d}x}\, \exp(-m), \tag{21.49}$$

where $-(\mathrm{d}m/\mathrm{d}x)\, \exp(-m) = \mathrm{d}\exp(-m)/\mathrm{d}x$. Integrating (21.49) from a to b and using $m(a) = 0$ yields

$$\frac{J_n(b)}{J}\, \exp[-m(b)] - \frac{J_n(a)}{J} = Y_n + \exp[-m(b)] - 1, \tag{21.50}$$

where the *electron-ionization integral* is defined as

$$Y_n = \int_a^b k_n\, \exp(-m)\, \mathrm{d}x. \tag{21.51}$$

The above result is somewhat simplified by observing that, due to impact ionization, the concentration of electrons in the conduction band increases from a to b, whereas that of holes increases from b to a; the concept is rendered in Fig. 21.12: consider a portion $x_1 < x < x_2$ of the space-charge region, where it is assumed for simplicity that the electric potential is linear. As the electric field is oriented to the left, electrons (indicated by the black dots) are accelerated to the right, holes (the white dots) are accelerated to the left. The vertical lines indicate the exchange of energies involved in impact-ionization events initiated by electrons or holes. An electron transited from the valence to the conduction band is accelerated by the field and may acquire a kinetic energy sufficient for initiating an impact-ionization event itself; the same applies to holes. As a consequence, the number of conduction-band electrons is

[13] An example of model for k_n, k_p is that proposed by Chynoweth [14]: $k_n = k_{ns}\, \exp(-|E_{cn}/E|^{\beta_n})$, $k_p = k_{ps}\, \exp(-|E_{cp}/E|^{\beta_p})$, where the parameters depend on temperature [83, 111].

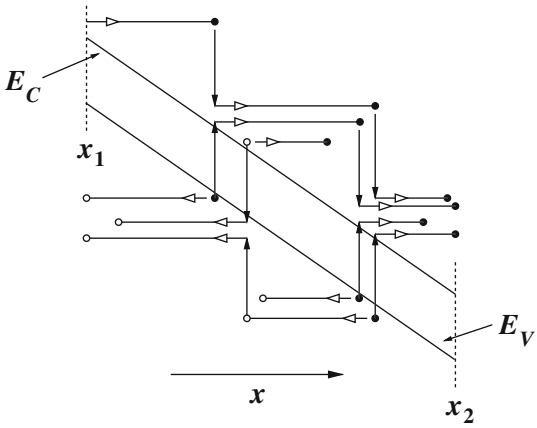

Fig. 21.12 Schematic description of the avalanche phenomenon. The details are given in the text

multiplied from left to right, namely, the number of those exiting at x_2 is larger than the number entering at x_1; similarly, the number of valence-band holes is multiplied from right to left. Due to the multiplication mechanism, taking $x = b$ by way of example, the major contribution to $J = J_n(b) + J_p(b)$ is given by $J_n(b)$. Then, letting $J \simeq J_n(b)$ in (21.50) and canceling out some terms provides

$$1 - \frac{1}{M_n} = Y_n, \qquad M_n = \frac{J_n(b)}{J_n(a)} \geq 1, \tag{21.52}$$

where M_n is the *electron-multiplication factor*. The latter is a measure of the impact ionization's level. As long as $Y_n < 1$, corresponding to a finite value of M_n, the avalanche condition does not occur; when $Y_n \to 1$, then $M_n \to \infty$: in this case, the injection of a negligibly-small number of electrons at a produces a large electron current density at b. The operating conditions where M_n is large must be avoided because an excessive current may damage the device. Note that in the design stage of the device one first calculates Y_n from (21.51), then obtains M_n from (21.52). Thus, it may well happen that $Y_n > 1$, corresponding to $M_n < 0$; this outcome is not physically sound, and simply indicates that the parameters used in the calculation of the ionization integral (21.51) are not consistent.[14]

The analysis of the impact-ionization condition can also be carried out starting with the elimination of J_n from the second equation in (21.46). The equation corresponding to (21.48) reads

$$\frac{\mathrm{d}J_p}{\mathrm{d}x} = -k_p\, J_p - k_n\, (J - J_p) = \frac{\mathrm{d}m}{\mathrm{d}x}\, J_p - k_n\, J\,; \tag{21.53}$$

in turn, the equation corresponding to (21.50) is

$$\frac{J_p(b)}{J}\, \exp\left[-m(b)\right] - \frac{J_p(a)}{J} = -\frac{Y_p}{\exp\left[m(b)\right]} + \exp\left[-m(b)\right] - 1, \tag{21.54}$$

[14] This happens, for instance, if a value of $|V|$ larger than the breakdown voltage is used in (21.51).

with the *hole-ionization integral* given by

$$Y_p = \int_a^b k_p \, \exp\left[m(b) - m\right] \, \mathrm{d}x. \tag{21.55}$$

Letting $J \simeq J_p(a)$ in (21.54) yields

$$1 - \frac{1}{M_p} = Y_p, \qquad M_p = \frac{J_p(a)}{J_p(b)} \geq 1, \tag{21.56}$$

where M_p is the *hole-multiplication factor*. Using the definition (21.48) of m, the relation between the ionization integrals is found to be

$$Y_n = \exp\left[-m(b)\right] Y_p + 1 - \exp\left[-m(b)\right]. \tag{21.57}$$

The above shows that $Y_p = 1$ corresponds to $Y_n = 1$, namely, the avalanche condition $Y_p = 1$ for the holes coincides with that of the electrons, as should be.

21.6 Complements

21.6.1 Weak-Injection Limit of the Drift-Diffusion Equations

In the calculation of the $I(V)$ characteristic of the p–n junction carried out in Sect. 21.3.1 it has been stated that in a quasi-neutral region, when the weak-injection condition holds, the diffusion term of the minority carries dominates over the drift term. To better discuss this issue, the case of a p-doped region is considered, so that the majority-carrier concentration is $c^{\mathrm{eq}} = p^{\mathrm{eq}} = p_{p0}$ and (20.35) become $|p - p_{p0}| \ll p_{p0}$, $|n - n_{p0}| \ll p_{p0}$. The latter may be recast as

$$|p - p_{p0}| \leq \alpha \, p_{p0}, \qquad |n - n_{p0}| \leq \alpha \, p_{p0}, \tag{21.58}$$

with $\alpha \ll 1$. Indicating with p_m, p_M the minimum and maximum values of p imposed by (21.58), one finds $p_M - p_{p0} = \alpha \, p_{p0}$, $p_{p0} - p_m = \alpha \, p_{p0}$, whence

$$p_M = (1 + \alpha) \; p_{p0}, \qquad p_m = (1 - \alpha) \; p_{p0}. \tag{21.59}$$

Similarly, for the minority-carrier concentration one finds $n_M - n_{p0} = \alpha \, p_{p0}$, $n_{p0} - n_m = \alpha \, p_{p0}$, whence

$$n_M = n_{p0} + \alpha \, p_{p0}, \qquad n_m = n_{p0} - \alpha \, p_{p0}. \tag{21.60}$$

The maximum absolute variation of p turns out to be:

$$p_M - p_m = 2 \, \alpha \, p_{p0}. \tag{21.61}$$

Instead, the maximum variation of n must be treated with some care. In fact, using the non-degenerate case one finds

$$n_M = \frac{n_i^2}{p_{p0}} + \alpha\, p_{p0} = p_{p0}\left(\frac{n_i^2}{p_{p0}^2} + \alpha\right), \tag{21.62}$$

$$n_m = \frac{n_i^2}{p_{p0}} - \alpha\, p_{p0} = p_{p0}\left(\frac{n_i^2}{p_{p0}^2} - \alpha\right). \tag{21.63}$$

Even for a relatively low dopant concentration, say, $N_A \simeq p_{p0} = 10^{16}\ \mathrm{cm}^{-3}$, at room temperature one has $n_i^2/p_{p0}^2 \simeq 10^{-12}$, which is much smaller than the reasonable values of α. It follows $n_M \simeq \alpha\, p_{p0}$, $n_m \simeq 0$, where the limit of n_m must be chosen as such because n is positive definite. In conclusion, the maximum relative variations of p and n with respect to the equilibrium values are given by

$$\frac{p_M - p_m}{p_{p0}} = 2\,\alpha, \qquad \frac{n_M - n_m}{n_{p0}} \simeq \alpha\,\frac{p_{p0}}{n_{p0}} = \alpha\,\frac{p_{p0}^2}{n_i^2} \gg 2\,\alpha. \tag{21.64}$$

By way of example, one may let $\alpha = 10^{-3}$, still with $N_A = 10^{16}\ \mathrm{cm}^{-3}$. While the maximum relative variation of p is 2×10^{-3}, that of n is 10^9; it follows that the constraint imposed onto the derivative is strong in the case of p, much weaker for n. Within the same example, the maximum absolute variation is $2 \times 10^{13}\ \mathrm{cm}^{-3}$ for p and $10^{13}\ \mathrm{cm}^{-3}$ for n, in both cases much smaller than the majority-carrier concentration ($10^{16}\ \mathrm{cm}^{-3}$). The conclusion is that in a quasi-neutral region, under the weak-injection conditions, the diffusive transport prevails for the minority carriers, whereas the transport of the majority carriers is dominated by drift. With reference to the p-doped region considered here, one has $J_p \simeq q\,\mu_p\, p\, E$ and $J_n \simeq q\, D_n\, \mathrm{grad} n$, respectively.

21.6.2 Shockley's Boundary Conditions

The derivation of the analytical model of the p–n junction's $I(V)$ characteristic, worked out in Sect. 21.3.1, requires the boundary conditions for the minority-carrier concentrations at the boundaries of the space-charge region; specifically, one needs to determine $n(-l_p)$ and $p(l_n)$. The derivation is based on calculating approximate expressions for the ratios $n(-l_p)/n(l_n)$, $p(l_n)/p(-l_p)$, where the denominators are majority-carrier concentrations that are in turn approximated with $n(l_n) \simeq n_{n0} \simeq N_D$ and $p(-l_p) \simeq p_{p0} \simeq N_A$.

To proceed, one considers the electron drift-diffusion equation $J_n = q\,\mu_n\, n\, E + q\, D_n\, \mathrm{d}n/\mathrm{d}x$, and observes that in the space-charge region the drift and diffusion terms have opposite signs; also, their moduli are much larger than that of the current density. In fact, the latter is small due to the weak-injection condition,

whereas the terms at the right hand side of the equation are large because the electric potential and the electron concentration have non-negligible variations over the space-charge region. It follows that the moduli of the drift and diffusion terms are comparable to each other: $-q\mu_n n E \simeq q D_n \mathrm{d}n/\mathrm{d}x \gg |J_n|$ and, similarly, $-q\mu_p p E \simeq -q D_p \mathrm{d}p/\mathrm{d}x \gg |J_p|$ for holes. Now, the approximation is introduced, that consists in neglecting J_n and J_p; this yields equilibrium-like expressions for the concentrations, $n \simeq n^{(0)} \exp[q\varphi/(k_B T)]$, $p \simeq n^{(0)} \exp[-q\varphi/(k_B T)]$, which are used to calculate the ratios sought:

$$\frac{n(-l_p)}{n(l_n)} \simeq \exp\left[\frac{q(V-\psi_0)}{k_B T}\right] = \frac{n_i^2}{N_A N_D} \exp\left(\frac{qV}{k_B T}\right), \tag{21.65}$$

$$\frac{p(l_n)}{p(-l_p)} \simeq \exp\left[\frac{q(V-\psi_0)}{k_B T}\right] = \frac{n_i^2}{N_A N_D} \exp\left(\frac{qV}{k_B T}\right). \tag{21.66}$$

The last form of (21.65), (21.66) is obtained from the definition (21.5) of the built-in potential. Using $n(l_n) \simeq n_{n0} \simeq N_D$ and $p(-l_p) \simeq p_{p0} \simeq N_A$ along with $n_{p0} = n_i^2/N_A$ and $p_{n0} = n_i^2/N_D$ (compare with (21.1,21.2)), finally yields the Shockley boundary conditions

$$n(-l_p) \simeq n_{p0} \exp\left(\frac{qV}{k_B T}\right), \qquad p(l_n) \simeq p_{n0} \exp\left(\frac{qV}{k_B T}\right). \tag{21.67}$$

21.6.3 Depletion Capacitance—Arbitrary Doping Profile

The expression of the depletion capacitance worked out in Sect. 21.4 for an abrupt p–n junction is extended here to an arbitrary doping profile, still in one dimension. Let $a < x < b$ be the region where the charge density ρ differs from zero, and assume for the electric potential that $\varphi = \varphi(a) = \text{const}$ for $x < a$ and $\varphi = \varphi(b) = \text{const}$ for $x > b$. Also, it is assumed that there are no single layers or double layers of charge, whence the electric field E and φ are continuous. The constancy of φ in the outside regions implies the global charge neutrality, as is found by integrating $\varepsilon_{\mathrm{sc}} \mathrm{d}E/\mathrm{d}x = \rho$ from a to b and using the continuity of E:

$$\int_a^b \rho \, \mathrm{d}x = 0. \tag{21.68}$$

Thanks to (21.68) one finds, for any x,

$$\int_a^x \rho \, \mathrm{d}x + \int_x^b \rho \, \mathrm{d}x = 0 \qquad Q = -\int_a^x \rho \, \mathrm{d}x = \int_x^b \rho \, \mathrm{d}x, \tag{21.69}$$

which provides the definition of the charge per unit area Q. The definition holds also if x is outside the interval $[a, b]$; in this case, however, one finds $Q = 0$. In the

following it is assumed that x is internal to the space-charge region. The solution of the Poisson equation is taken from Prob. 4.2; using $E(a) = 0$ and the global charge-neutrality condition after letting $\psi = \varphi(b) - \varphi(a)$, yields

$$\varepsilon_{\rm sc}\,\psi = \int_a^b x\rho\ \mathrm{d}x. \tag{21.70}$$

If the voltage drop changes by a small amount, $\psi \leftarrow \psi + \mathrm{d}\psi$, the space-charge boundaries are modified, $a \leftarrow a + \mathrm{d}a$, $b \leftarrow b + \mathrm{d}b$, whence Q changes as well:[15]

$$\mathrm{d}Q = \int_b^{b+\mathrm{d}b} \rho\ \mathrm{d}x = \rho(b)\,\mathrm{d}b = \int_a^{a+\mathrm{d}a} \rho\ \mathrm{d}x = \rho(a)\,\mathrm{d}a. \tag{21.71}$$

On the other hand, from (21.70) it follows

$$\varepsilon_{\rm sc}\ \mathrm{d}\psi = \int_{a+\mathrm{d}a}^{b+\mathrm{d}b} x\ \rho\ \mathrm{d}x - \int_a^b x\rho\ \mathrm{d}x = b\,\rho(b)\ \mathrm{d}b - a\,\rho(a)\ \mathrm{d}a = (b-a)\ \mathrm{d}Q. \tag{21.72}$$

Thus, the capacitance per unit area of the space-charge region is

$$C = \frac{\mathrm{d}Q}{\mathrm{d}\psi} = \frac{\varepsilon_{\rm sc}}{b-a}, \tag{21.73}$$

which is the expected generalization of (21.45). Note that the absence of charge layers makes the variation $\mathrm{d}Q$ to depend on the variations in a and b only. As a consequence it is $a = a(\psi)$, $b = b(\psi)$, whence $C = C(\psi)$. If ψ and $\rho(x)$ are prescribed, the values of a, b are determined by the system of Eqs. (21.68) and (21.70).

It is interesting to note that if the charge density has a power form, $\rho \sim x^n$, then ψ depends on the $(n+2)$th power of $b - a$. Consider by way of example a *diffused junction*, namely, a junction obtained, e.g., by diffusing a dopant of the p type into a uniform n-type substrate. Expanding ρ to first order around the metallurgical junction, and using the full-depletion and ASCE approximations, yields $\rho \simeq k\,x$ for $-l/2 < x < l/2$ and $\rho = 0$ elsewhere. Using (21.70) and (21.72) then yields

$$\varepsilon_{\rm sc}\,\psi = \frac{1}{12}\,k\,l^3, \qquad C = C_0\left(1 - \frac{V}{\psi_0}\right)^{-1/3}, \qquad C_0 = \left(\frac{k\,\varepsilon_{\rm sc}^2}{12\,\psi_0}\right)^{1/3}. \tag{21.74}$$

The general expression (21.72) of the capacitance per unit area of the space-charge region finds a useful application in a measuring technique for the doping profile (Sect. 25.5).

[15] After the change in the boundaries' positions, x in (21.69) is still internal to the space-charge region.

21.6.4 Order of Magnitude of Junction's Parameters

Still considering an abrupt, p–n silicon junction with $N_A = 10^{16}$ cm^{-3}, $N_D = 10^{15}$ cm^{-3}, the built-in potential at room temperature is

$$\psi_0 = \frac{k_B\,T}{q}\,\log\left(\frac{N_A\,N_D}{n_i^2}\right) \simeq 0.65\ \text{V} \tag{21.75}$$

(compare with (21.5)). The carrier mobilities have been estimated in Sect. 19.6.6; in fact, hole mobility is smaller than electron mobility and, as outlined in Sect. 20.5.3, the mobility degradation due to impurity scattering is expected to vary from one side of the junction to the other because the doping concentrations are different. The experimental minority-carrier mobilities for the doping concentrations and temperature considered here are $\mu_n \simeq 1000$ cm^2 V^{-1} s^{-1} in the p region and $\mu_p \simeq 500$ cm^2 V^{-1} s^{-1} in the n region [103, Sect. 1.5], whence

$$D_n = \frac{k_B\,T}{q}\,\mu_n \simeq 26\ \text{cm}^2\,\text{s}^{-1}, \qquad D_p = \frac{k_B\,T}{q}\,\mu_p \simeq 13\ \text{cm}^2\,\text{s}^{-1}. \tag{21.76}$$

The experimental values of the minority-carrier lifetimes are $\tau_n \simeq 5 \times 10^{-5}$ s and $\tau_p \simeq 2 \times 10^{-5}$ s. The corresponding diffusion lengths (21.24), (21.28) are

$$L_n = \sqrt{\tau_n\,D_n} \simeq 360\ \mu\text{m}, \qquad L_p = \sqrt{\tau_p\,D_p} \simeq 160\ \mu\text{m}. \tag{21.77}$$

The above values provide for the saturation current density (21.30)

$$J_s = q\left(\frac{D_p\,p_{n0}}{L_p} + \frac{D_n\,n_{p0}}{L_n}\right) \simeq 14\ \text{pA cm}^{-2}. \tag{21.78}$$

From (21.41), the width of the depletion region at zero bias is found to be

$$l(V=0) = l_n + l_p = \left[\frac{2\,\varepsilon_{\text{sc}}}{q}\left(\frac{1}{N_D} + \frac{1}{N_A}\right)\psi_0\right]^{1/2} \simeq 1\ \mu\text{m}, \tag{21.79}$$

with $l_n/l_p = N_A/N_D = 10$. The permittivity of silicon $\varepsilon_{\text{sc}} = 11.7 \times \varepsilon_0$ has been used, with $\varepsilon_0 \simeq 8.854 \times 10^{-14}$ F cm^{-1} the vacuum permittivity. Finally, the value of the differential capacitance per unit area at zero bias (21.44) is

$$C_0 = \left[\frac{q\,\varepsilon_{\text{sc}}/(2\,\psi_0)}{1/N_D + 1/N_A}\right]^{1/2} = \frac{\varepsilon_{\text{sc}}}{l(V=0)} \simeq 11\ \text{nF cm}^{-2}. \tag{21.80}$$

Problems

21.1 Evaluate the built-in potential at room temperature in an abrupt p–n junction with $N_A = 10^{16}$ cm^{-3} and $N_D = 10^{15}$ cm^{-3}.

21.2 Show that avalanche due to impact ionization is possible only if both coefficients k_n and k_p are different from zero.