Theory of Semiconductor Devices
(반도체 소자 이론)

Lecture 19

Young Min Song
Assistant Professor
School of Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology
http://www.gist-foel.net
ymsong@gist.ac.kr, ymsong81@gmail.com
A207, ☎2655

Contents

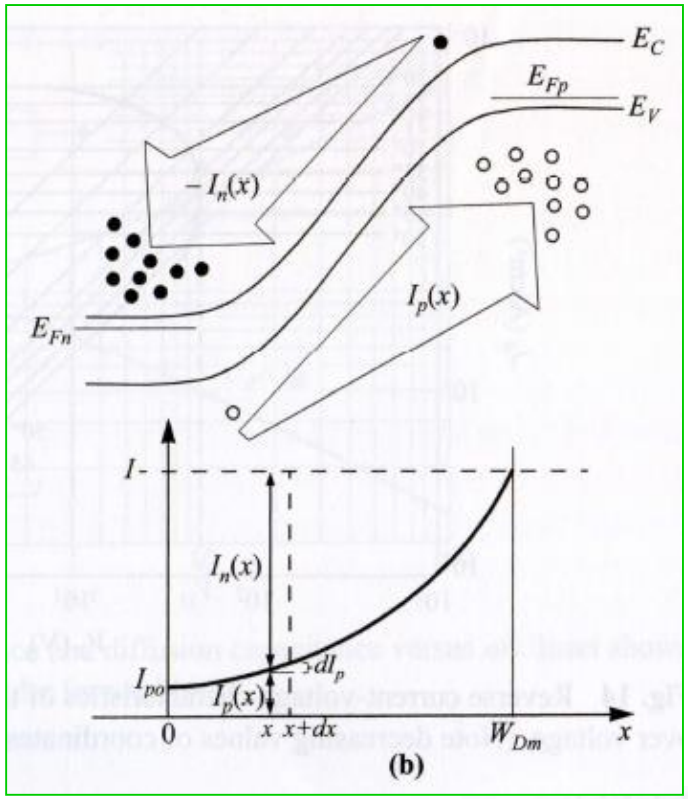
p-n Junction

Avalanche Multiplication

- ***Avalanche multiplication***, or impact ionization, is the most-important mechanism in junction breakdown.
- The avalanche breakdown voltage imposes an ***upper limit*** on the reverse bias for most diodes, on the collector voltage of bipolar transistors, and on the drain voltages of MESFETs and MOSFETs.
- In addition, the impact ionization mechanism can be used to generate microwave power, as in IMPATT devices, and to amplify optical signals, as in avalanche photodetectors.

- We first derive the basic ionization integral which determines the breakdown condition.
-Assume that a current I_{po} is incident at the left-hand side of the depletion region with width W_{Dm} (Fig. 15b).
- If the electric field in the depletion region is high enough that electron-hole pairs are generated by the impact ionization process, the hole current I_p will increase with distance through the depletion region and reach a value $M_p I_{po}$ at $x = W_{Dm}$.
- Similarly, the electron current I_n will increase from $I_n(W_{Dm}) = 0$ to $I_n(0) = 1 - I_{po}$,where the total current I ($= I_p + I_n$) is ***constant*** at steady state.

Avalanche Multiplication



The incremental hole current is equal to the number of electron-hole pairs generated per second in the distance dx,

$$dI_p = I_p \alpha_p dx + I_n \alpha_n dx \tag{93}$$

or

$$\frac{dI_p}{dx} - (\alpha_p - \alpha_n) I_p = \alpha_n I. \tag{94}$$

α_n and α_p are the electron and hole ionization rates.

- The solution of Eq. 94 with the boundary condition of $I = I(W_{Dm}) = M_p I_{po}$ is given by

$$I_p(x) = I\left\{\int_0^x \alpha_n \exp\left[-\int_0^x (\alpha_p - \alpha_n)dx'\right]dx + \frac{1}{M_p}\right\} \Big/ \exp\left[-\int_0^x (\alpha_p - \alpha_n)dx'\right] \tag{95}$$

where M_p is the multiplication factor of holes and is defined as

$$M_p \equiv \frac{I_p(W_{Dm})}{I_p(0)} \equiv \frac{I}{I_{po}}. \tag{96}$$

Avalanche Multiplication

- With a relationship

$$\int_0^{W_{Dm}} (\alpha_p - \alpha_n) \exp\left[-\int_0^x (\alpha_p - \alpha_n) dx'\right] dx = -\exp\left[-\int_0^x (\alpha_p - \alpha_n) dx'\right]\Bigg|_0^{W_{Dm}} = -\exp\left(\left[-\int_0^{W_{Dm}} (\alpha_p - \alpha_n) dx'\right] + 1\right), \quad (97)$$

Equation 95 can be evaluated at $x = W_{Dm}$ and be rewritten as

$$1 - \frac{1}{M_p} = \int_0^{W_{Dm}} \alpha_p \exp\left[-\int_0^x (\alpha_p - \alpha_n) dx'\right] dx. \quad (98)$$

- Note that M_p is a function of α_n in addition to α_p.
- The ***avalanche breakdown voltage*** is defined as the voltage where M_p approaches ***infinity***.

Hence the breakdown condition is given by the ionization integral

$$\int_0^{W_{Dm}} \alpha_p \exp\left[-\int_0^x (\alpha_p - \alpha_n) dx'\right] dx = 1. \quad (99a)$$

- If the avalanche process is initiated by electrons instead of holes, the ionization integral is given by

$$\int_0^{W_{Dm}} \alpha_n \exp\left[-\int_x^{W_{Dm}} (\alpha_n - \alpha_p) dx'\right] dx = 1. \quad (99b)$$

Avalanche Multiplication

- Equations 99a and 99b are equivalent; that is, the breakdown condition depends *only* on what is happening within the depletion region and *not* on the carriers (or primary current) that initiate the avalanche process. The situation does not change when a mixed primary current initiates the breakdown, so either Eq. 99a or Eg. 99b gives the breakdown condition.
- For semiconductors with equal ionization rates ($\alpha_n = \alpha_p = \alpha$) such as GaP, Eq. 99a or 99b reduces to the simple expression

$$\int_0^{W_{Dm}} \alpha dx = 1 . \qquad (100)$$

- From the breakdown conditions described above and the field dependence of the ionization rates, the breakdown voltage, maximum electric field, and depletion-layer width ***can*** *be calculated*.
- As discussed previously, the electric field and potential in the depletion layer are determined from the solutions of the Poisson equation. Depletion layer boundaries that satisfy Eq. 99a or 99b can be obtained numerically using an iteration method.
- With known boundaries we obtain the breakdown voltage

$$V_{BD} = \frac{\mathscr{E}_m W_{Dm}}{2} = \frac{\varepsilon_s \mathscr{E}_m^2}{2qN} \qquad (101)$$

for one-sided abrupt junctions, and

Avalanche Multiplication

$$V_{BD} = \frac{2\mathscr{E}_m W_{Dm}}{3} = \frac{4\mathscr{E}_m^{3/2}}{3}\left(\frac{2\varepsilon_s}{qa}\right)^{1/2} \qquad (102)$$

for linearly graded junctions, where N is the ionized background impurity concentration of the lightly doped side, a the impurity gradient, and $\mathscr{E}_m$ the maximum field.

-**Figure 16a** shows the *calculated* breakdown voltage as a function of N for abrupt junctions in Si, (1 0 0)-oriented GaAs, and GaP. The experimental results are generally in good agreement with the calculated values.

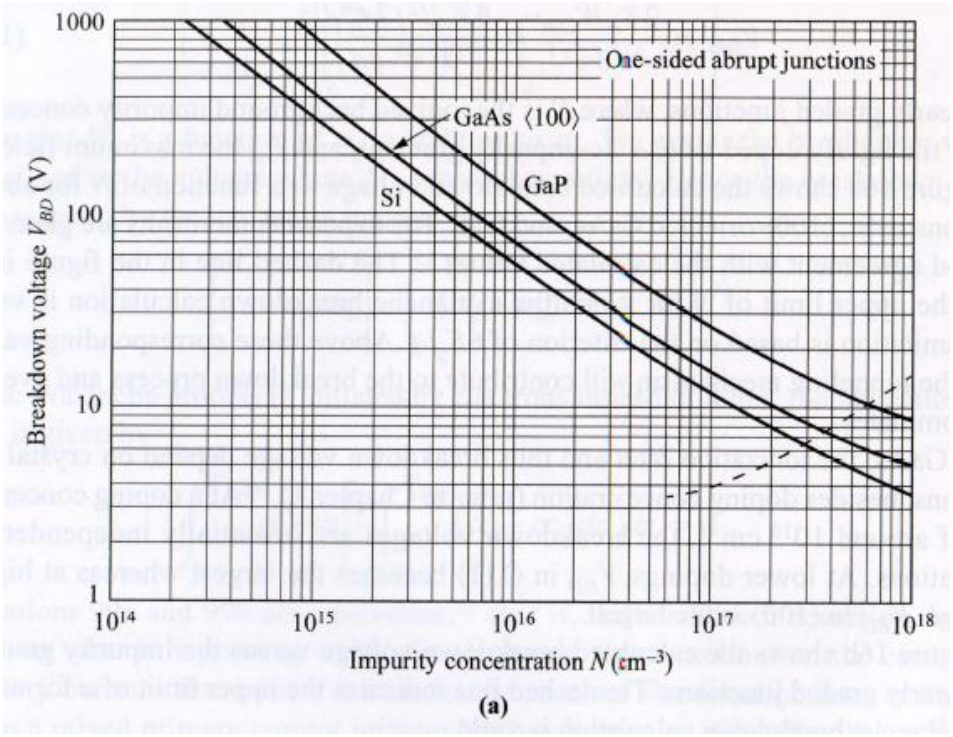
- **Figure 16b** shows the calculated breakdown voltage versus the impurity gradient for linearly graded junctions. The dashed line indicates the upper limit of a for which the avalanche breakdown calculation is valid.

- The calculated values of the maximum field $\mathscr{E}_m$ and the depletion-layer width at breakdown for the three semiconductors above are shown in **Fig. 17a** for the abrupt junctions, and in **Fig. 17b** for the linearly graded junctions. For the Si abrupt junctions, the maximum field at breakdown can be expressed as

$$\mathscr{E}_m = \frac{4\times10^5}{1-(1/3)\log_{10}(N/10^{16}\ \mathrm{cm^{-3}})} \quad \mathrm{V/cm} \qquad (103)$$

where N is in $\mathrm{cm^{-3}}$.

Avalanche Multiplication



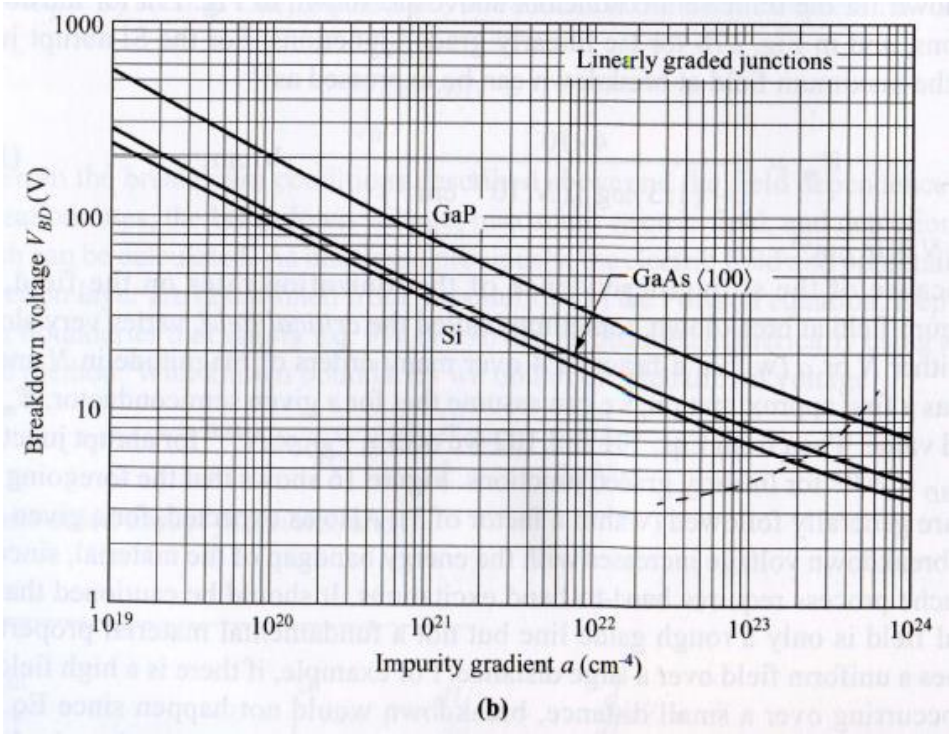
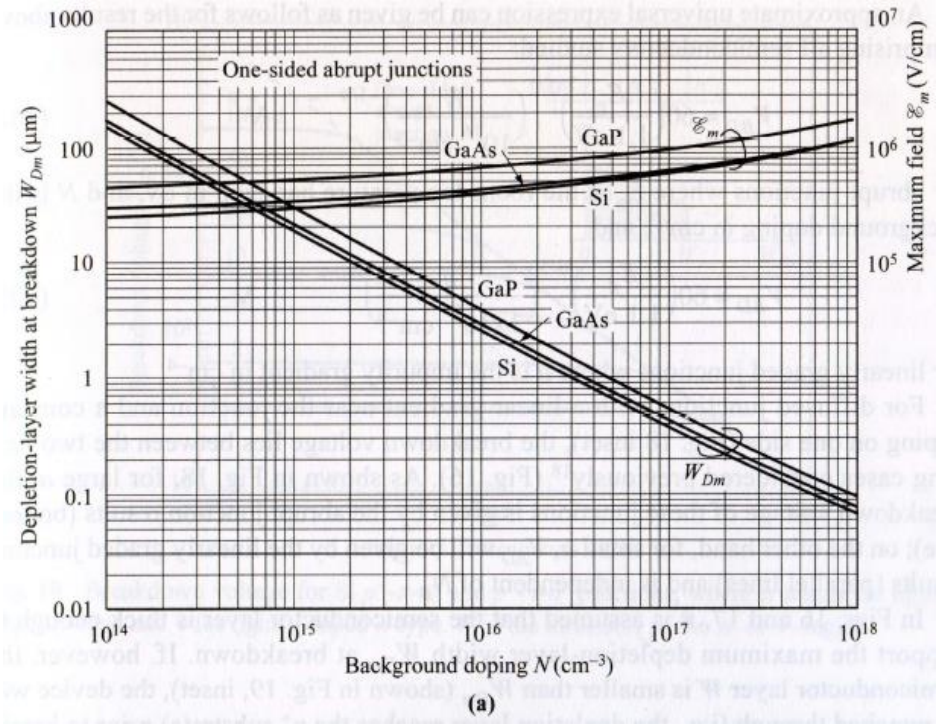


Fig. 16 Avalanche breakdown voltage in Si, ⟨100⟩-oriented GaAs, and GaP, for (a) one-sided abrupt junctions (vs. impurity concentration) and (b) linearly graded junctions (vs. impurity gradient). The dashed lines indicate the maximum doping or doping gradient beyond which tunneling will dominate the breakdown characteristics.

Avalanche Multiplication



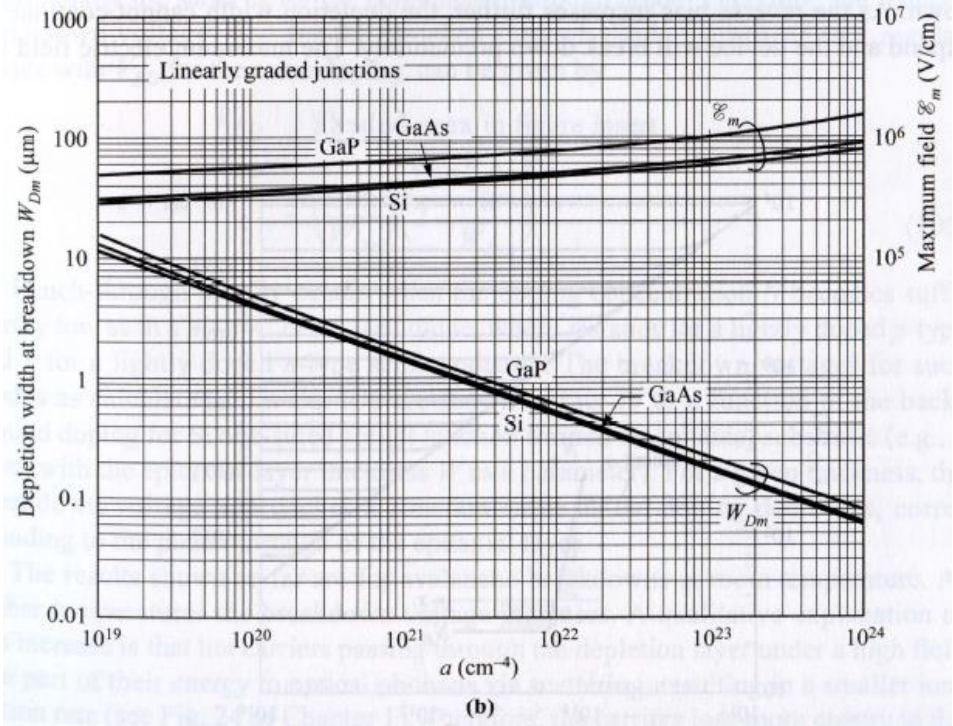


Fig. 17 Depletion-layer width and maximum field at ***breakdown*** in Si, ⟨100⟩-oriented GaAs, and GaP for (a) one-sided abrupt junctions and (b) linearly graded JunctIons.

Avalanche Multiplication

- An approximate *universal* expression can be given as follows for the results above comprising all semiconductors studied:

$$V_{BD} \approx 60\left(\frac{E_g}{1.1\ \text{eV}}\right)^{3/2}\left(\frac{N}{10^{16}\ \text{cm}^{-3}}\right)^{-3/4} \quad \text{V} \qquad (104)$$

for abrupt junctions where E_g is the room-temperature bandgap in eV, and N is the background doping in cm-3; and

$$V_{BD} \approx 60\left(\frac{E_g}{1.1\ \text{eV}}\right)^{6/5}\left(\frac{a}{3\times10^{20}\ \text{cm}^{-4}}\right)^{-2/5} \quad \text{V} \qquad (105)$$

for linearly graded junctions where a is the impurity gradient in cm^{-4}.

- For diffused junctions with a linear gradient near the junction and a constant doping on one side (Fig. 18 inset), the breakdown voltage lies between the two limiting cases considered previously.

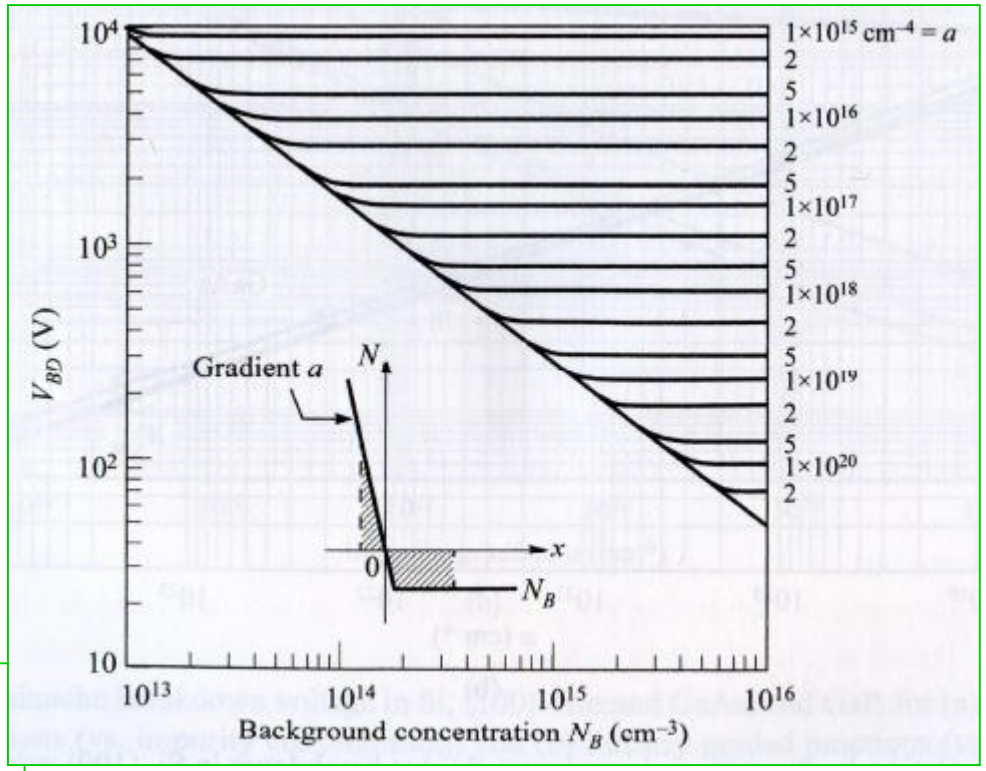


Fig. 18 Breakdown voltage for Si diffused junctions at 300 K.

Contents

p-n Junction

Heterojunctions

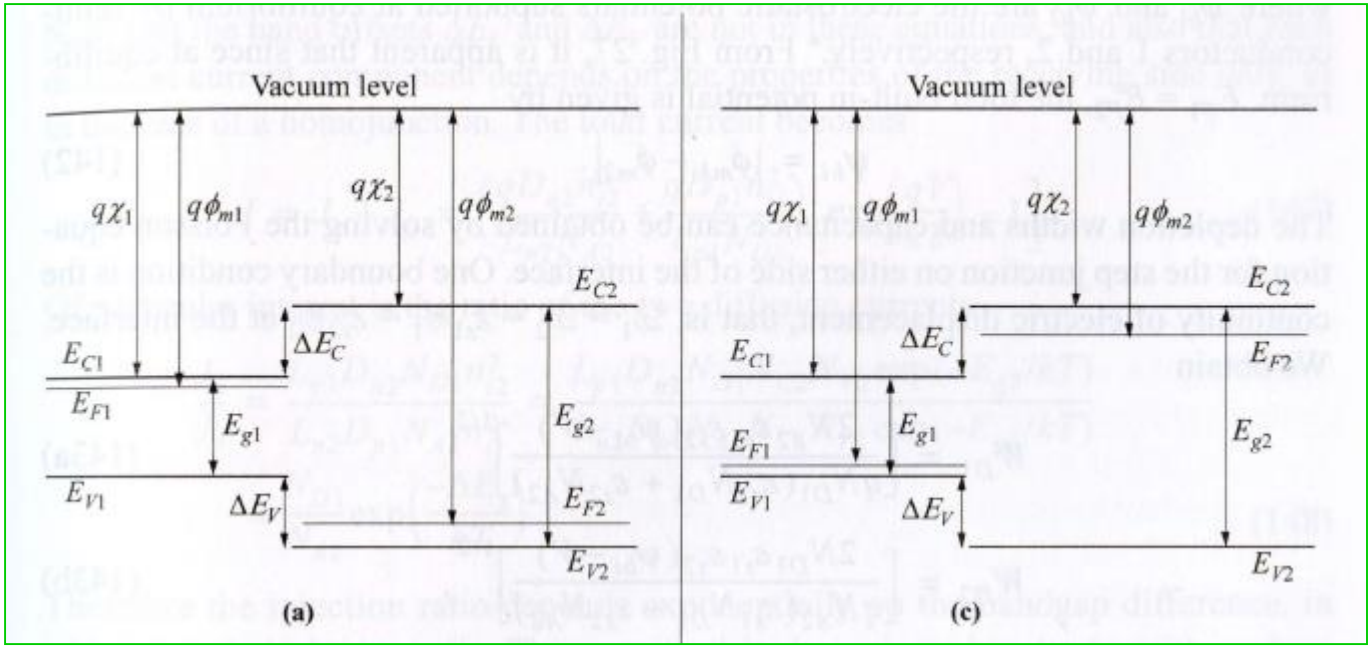
- When the two semiconductors have the same type of conductivity, the junction is called an ***isotype*** heterojunction.
- When the conductivity types differ, the junction is called an ***anisotype*** heterojunction which is a much more useful and common structure than its counterpart.

- In 1951, Shockley proposed the *abrupt* heterojunction to be used as an efficient emitter-base injector in a ***bipolar*** transistor.
- In the same year, Gubanov published a *theoretical* paper on heterojunctions.
- Kroemer later analyzed a similar, although graded, heterojunction as a wide-bandgap emitter.
- Since then, heterojunctions have been extensively studied, and many important applications have been made, among them the room-temperature injection laser, light-emitting diode (LED), photodetector, and solar cell, to name a few.

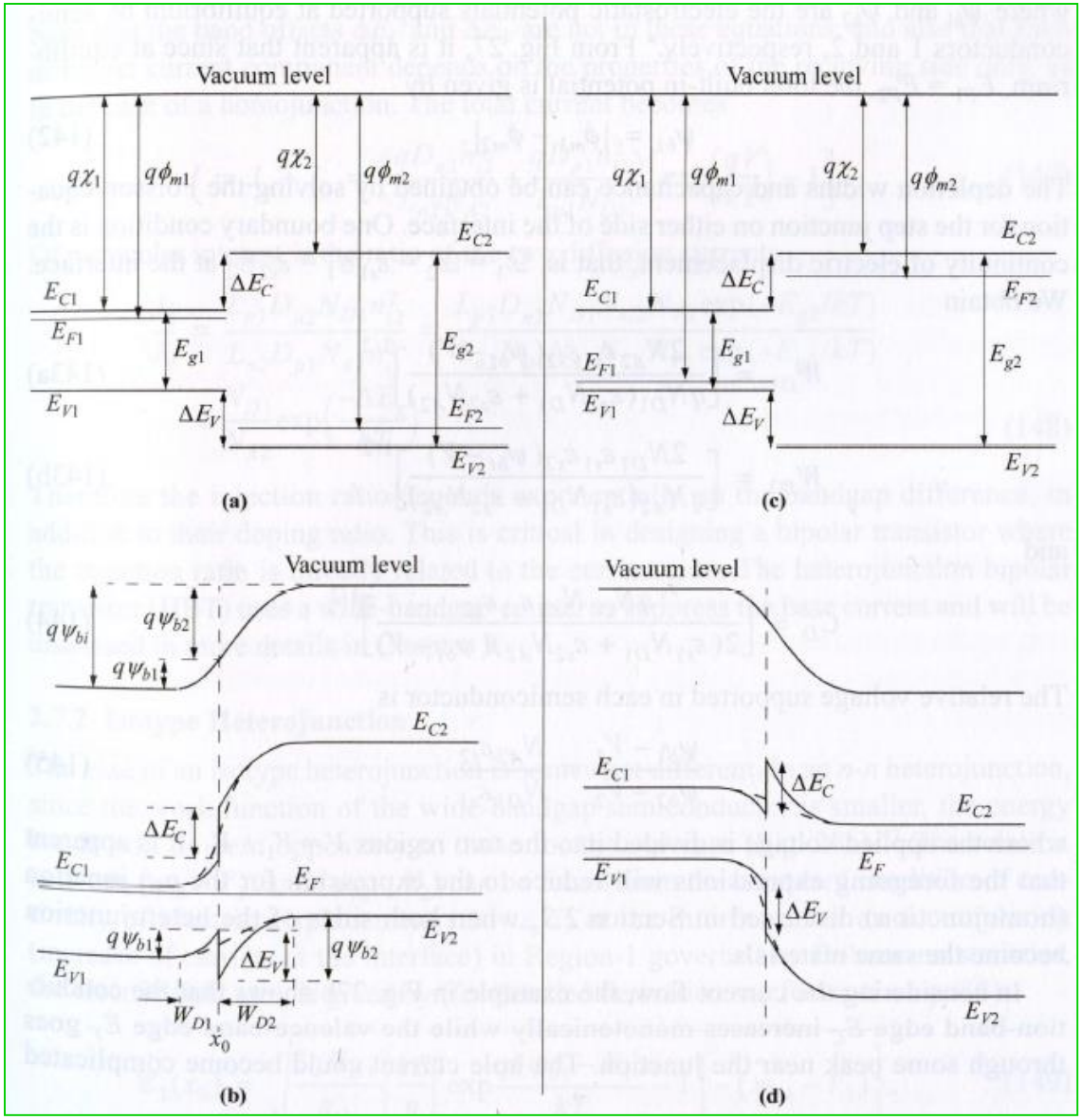
- In many of these applications, by forming *periodic* heterojunctions with layer thickness of the order of 10 nm, we utilize the interesting properties of quantum wells and superlattices.

Anisotype Heterojunctions

- The energy-band model of an *idealized* anisotype abrupt heterojunction without interface traps was proposed by Anderson based on the previous work of Shockley.
-We consider this model next, since it can adequately explain *most* transport processes, and only slight modification of the model is needed to account for *nonideal* cases such as interface traps.
- **Figures 27**a and c show the energy-band diagrams of two isolated semiconductors of opposite types. The two semiconductors are assumed to have different band gaps E_g, different permittivities ε_s , different work functions ϕ_m , and different electron affinities χ.



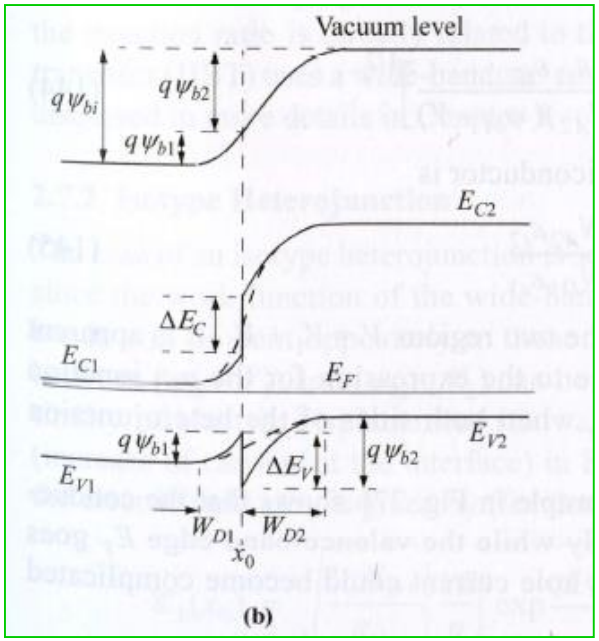
Anisotype Heterojunctions



-Work function and electron affinity are ***defined*** as the energy required to remove an electron from the Fermi level E_F and from the bottom of the conduction band E_C respectively, to a position just outside the material (vacuum level).

Fig.27 Energy-band diagrams for (a) two isolated semiconductors of opposite types and different E_g (of which the smaller bandgap is *n*-type) and (b) their idealized anisotype heterojunction *at* ***thermal equilibrium***. In (c) and (d), the smaller bandgap is *p*-type. In (b) and (d), the dashed lines across the junctions represent graded composition.

Anisotype Heterojunctions



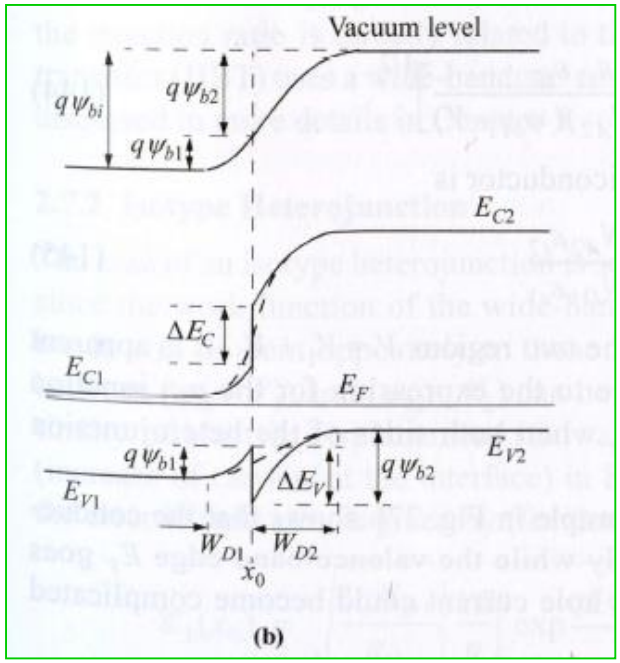
- The difference in energy of the conduction-band edges in the two semiconductors is represented by ΔE_C and that in the valence-band edges by ΔE_V .

- When a junction is formed between these semiconductors, the energy-band profile *at equilibrium* is shown in Fig. 27b for an *n-p* anisotype heterojunction where, in this example, the narrow-bandgap material is n-type.

-Since the ① Fermi level ***must*** coincide on both sides in equilibrium and ② the vacuum level is everywhere parallel to the band edges and is continuous, ③ the discontinuity in the conduction-band edges (ΔE_C) and valence-band edges (ΔE_V) is invariant with doping in those cases where E_g and χ are not functions of doping (i.e., nondegenerate semiconductors).
- The total built-in potential ψ_{bi} is equal to the sum of the partial built-in voltages ($\psi_{b1} + \psi_{b2}$), where ψ_{b1} and ψ_{b2} are the electrostatic potentials supported at equilibrium by semiconductors 1 and 2, respectively. From Fig. 27, it is apparent that since at equilibrium, $E_{F1} = E_{F2}$, the ***total built-in potential*** is given by

$$\psi_{bi} = |\phi_{m1} - \phi_{m2}| . \qquad (142)$$

Anisotype Heterojunctions



- The depletion widths and capacitance can be obtained by solving the Poisson equation for the step junction on either side of the interface. One boundary condition is the ***continuity*** of electric displacement, that is, $\mathscr{D}_1 = \mathscr{D}_2 = \varepsilon_1 \mathscr{E}_1 = \varepsilon_2 \mathscr{E}_2$ at the interface.
- We obtain

$$W_{D1} = \left[\frac{2N_{A2}\varepsilon_{s1}\varepsilon_{s2}(\psi_{bi} - V)}{qN_{D1}(\varepsilon_{s1}N_{D1} + \varepsilon_{s2}N_{A2})}\right]^{1/2}, \quad (143a)$$

$$W_{D2} = \left[\frac{2N_{D1}\varepsilon_{s1}\varepsilon_{s2}(\psi_{bi} - V)}{qN_{A2}(\varepsilon_{s1}N_{D1} + \varepsilon_{s2}N_{A2})}\right]^{1/2}, \quad (143b)$$

$$W_{Dp} = \sqrt{\frac{2\varepsilon_s \psi_{bi}}{q}\frac{N_D}{N_A(N_A + N_D)}}$$

$$W_{Dn} = \sqrt{\frac{2\varepsilon_s \psi_{bi}}{q}\frac{N_A}{N_D(N_A + N_D)}}$$

and

$$C_D = \left[\frac{qN_{D1}N_{A2}\varepsilon_{s1}\varepsilon_{s2}}{2(\varepsilon_{s1}N_{D1} + \varepsilon_{s2}N_{A2})(\psi_{bi} - V)}\right]^{1/2}. \quad (144)$$

The relative voltage supported in each semiconductor is

$$\frac{\psi_{b1} - V_1}{\psi_{b2} - V_2} = \frac{N_{A2}\varepsilon_{s2}}{N_{D1}\varepsilon_{s1}} \quad (145)$$

where the applied voltage is *divided* into the two regions $V = V_1 + V_2$.

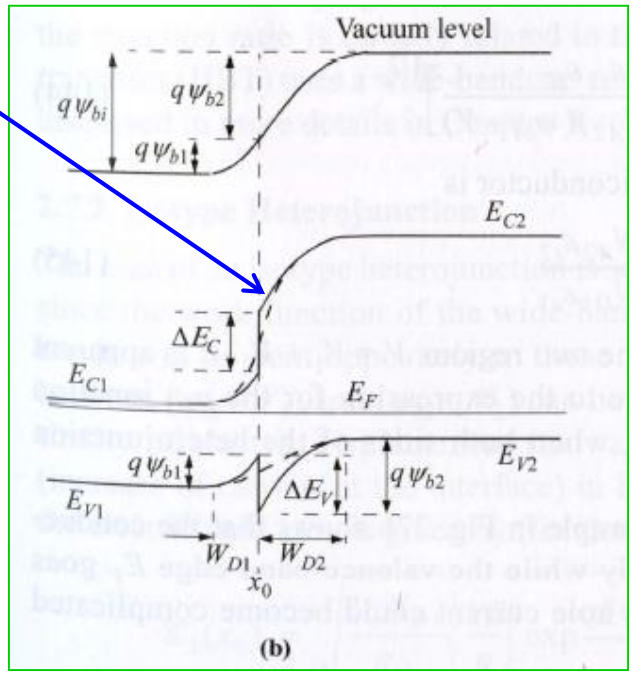
Anisotype Heterojunctions

- In considering the current flow, the example in Fig. 27b shows that the conduction-band edge E_C increases *monotonically* while the valence-band edge E_V goes through some ***peak*** near the junction.
- The hole current could become *complicated* because of the added barrier which might present a *bottle-neck* in thermionic emission, in series with diffusion. The analysis can be greatly ***simplified*** by *assuming* a ***graded junction*** where ΔE_C and ΔE_V become *smooth* transitions inside the depletion region. With this assumption, the diffusion currents are similar to a regular *p-n* junction but with the appropriate parameters in place. The electron and hole diffusion currents are:

$$J_n = \frac{qD_{n2}n_{i2}^2}{L_{n2}N_{A2}}\left[\exp\left(\frac{qV}{kT}\right) - 1\right], \quad (146a)$$

$$J_p = \frac{qD_{p1}n_{i1}^2}{L_{p1}N_{D1}}\left[\exp\left(\frac{qV}{kT}\right) - 1\right]. \quad (146b)$$

- Note that the band offsets ΔE_C and ΔE_V are *not* in these equations, and also that each diffusion current component depends on the properties of the *receiving side only*, as in the case of a homojunction.



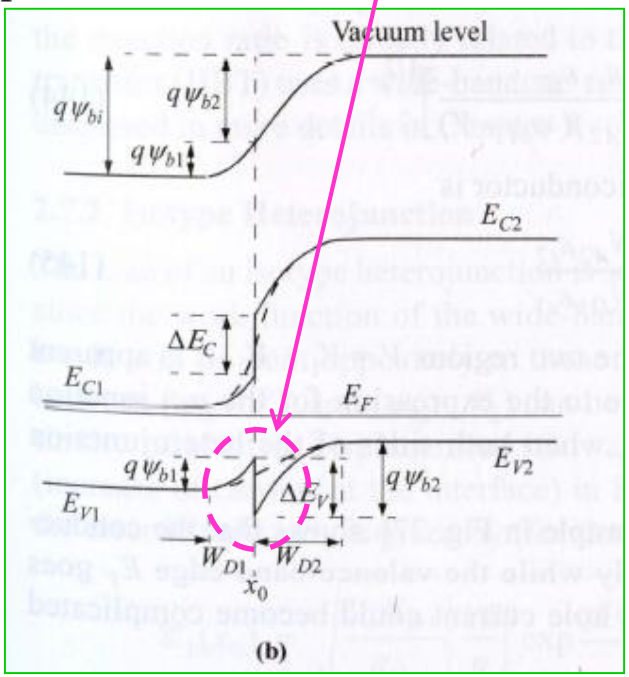
Anisotype Heterojunctions

- In considering the current flow, the example in Fig. 27b shows that the conduction-band edge E_C increases *monotonically* while the valence-band edge E_V goes through some ***peak*** near the junction.
- The hole current could become *complicated* because of the added barrier which might present a *bottle-neck* in thermionic emission, in series with diffusion. The analysis can be greatly ***simplified*** by *assuming* a ***graded junction*** where ΔE_C and ΔE_V become *smooth* transitions inside the depletion region. With this assumption, the diffusion currents are similar to a regular *p-n* junction but with the appropriate parameters in place. The electron and hole diffusion currents are:

$$J_n = \frac{qD_{n2}n_{i2}^2}{L_{n2}N_{A2}}\left[\exp\left(\frac{qV}{kT}\right) - 1\right], \qquad (146a)$$

$$J_p = \frac{qD_{p1}n_{i1}^2}{L_{p1}N_{D1}}\left[\exp\left(\frac{qV}{kT}\right) - 1\right]. \qquad (146b)$$

- Note that the band offsets ΔE_C and ΔE_V are *not* in these equations, and also that each diffusion current component depends on the properties of the *receiving side only*, as in the case of a homojunction.



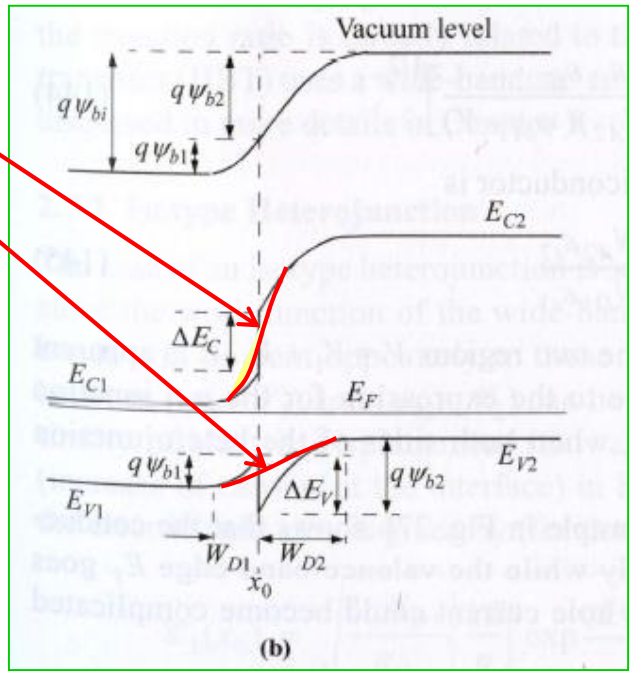
Anisotype Heterojunctions

- In considering the current flow, the example in Fig. 27b shows that the conduction-band edge E_C increases *monotonically* while the valence-band edge E_V goes through some ***peak*** near the junction.
- The hole current could become *complicated* because of the added barrier which might present a *bottle-neck* in thermionic emission, in series with diffusion. The analysis can be greatly ***simplified*** by *assuming* a ***graded junction*** where ΔE_C and ΔE_V become *smooth* transitions inside the depletion region. With this assumption, the diffusion currents are similar to a regular *p-n* junction but with the appropriate parameters in place. The electron and hole diffusion currents are:

$$J_n = \frac{qD_{n2}n_{i2}^2}{L_{n2}N_{A2}}\left[\exp\left(\frac{qV}{kT}\right) - 1\right], \quad (146a)$$

$$J_p = \frac{qD_{p1}n_{i1}^2}{L_{p1}N_{D1}}\left[\exp\left(\frac{qV}{kT}\right) - 1\right]. \quad (146b)$$

- Note that the band offsets ΔE_C and ΔE_V are *not* in these equations, and also that each diffusion current component depends on the properties of the *receiving side only*, as in the case of a homojunction.



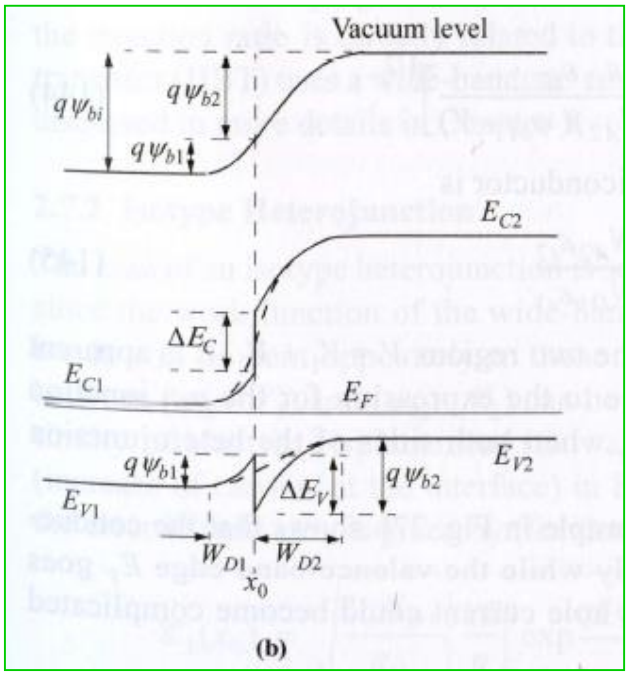
Anisotype Heterojunctions

- In considering the current flow, the example in Fig. 27b shows that the conduction-band edge E_C increases *monotonically* while the valence-band edge E_V goes through some ***peak*** near the junction.
- The hole current could become *complicated* because of the added barrier which might present a *bottle-neck* in thermionic emission, in series with diffusion. The analysis can be greatly ***simplified*** by *assuming* a ***graded junction*** where ΔE_C and ΔE_V become *smooth* transitions inside the depletion region. With this assumption, the diffusion currents are similar to a regular *p-n* junction but with the appropriate parameters in place. The electron and hole diffusion currents are:

$$J_n = \frac{qD_{n2}n_{i2}^2}{L_{n2}N_{A2}}\left[\exp\left(\frac{qV}{kT}\right)-1\right] \quad \boxed{\frac{qD_n n_{po}}{L_n}\left[\exp\left(\frac{qV}{kT}\right)-1\right]} \quad (146a)$$

$$J_p = \frac{qD_{p1}n_{i1}^2}{L_{p1}N_{D1}}\left[\exp\left(\frac{qV}{kT}\right)-1\right] \quad \boxed{\frac{qD_p p_{no}}{L_p}\left[\exp\left(\frac{qV}{kT}\right)-1\right]} \quad (146b)$$

- Note that the band offsets ΔE_C and ΔE_V are *not* in these equations, and also that each diffusion current component depends on the properties of the *receiving side only*, as in the case of a homojunction.



Anisotype Heterojunctions

- The total current becomes

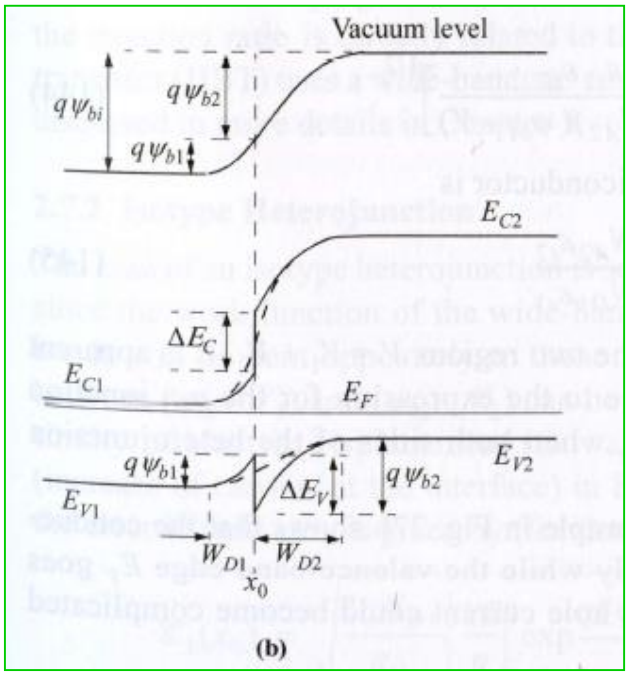
$$J = J_n + J_p = \left(\frac{qD_{n2}n_{i2}^2}{L_{n2}N_{A2}} + \frac{qD_{p1}n_{i1}^2}{L_{p1}N_{D1}}\right)\left[\exp\left(\frac{qV}{kT}\right) - 1\right]. \quad (147)$$

- Of particular interest is the ***ratio*** of the two diffusion currents.

$$\frac{J_n}{J_p} = \frac{L_{p1}D_{n2}N_{D1}n_{i2}^2}{L_{n2}D_{p1}N_{A2}n_{i1}^2} = \frac{L_{p1}D_{n2}N_{D1}N_{C2}N_{V2}\exp(-E_{g2}/kT)}{L_{n2}D_{p1}N_{A2}N_{C1}N_{V1}\exp(-E_{g1}/kT)}$$
$$\approx \frac{N_{D1}}{N_{A2}}\exp\left(\frac{-\Delta E_g}{kT}\right). \quad (148)$$

- Therefore the injection ratio depends ***exponentially*** on the bandgap difference, in addition to their doping ratio.
-This is critical in *designing* a bipolar transistor where the injection ratio is directly related to the current gain.

- The heterojunction bipolar transistor (HBT) uses a wide-bandgap emitter to suppress the base current and will be discussed in more details in Chapter 5.



Isotype Heterojunctions

- The case of an ***isotype*** heterojunction is somewhat different.
- In an *n-n* heterojunction, since the work function of the wide-bandgap semiconductor is smaller, the energy bands will be bent ***oppositely*** to those for the *n-p* case (Fig. 28a).
- The relation between (ψ_{b1} - V_1) and (ψ_{b2} - V_2) can be found from the boundary condition of continuity of electric displacement ($\mathscr{D}=\varepsilon_s \mathscr{E}$) at the interface.

- For an accumulation (increase of carriers at the interface) in Region-l governed by Boltzmann statistics, the electric field at x_0 is given by

$$\mathscr{E}_1(x_0) = \sqrt{\frac{2qN_{D1}}{\varepsilon_{s1}}\left\{\frac{kT}{q}\left[\exp\frac{q(\psi_{b1}-V_1)}{kT}-1\right]-(\psi_{b1}-V_1)\right\}}. \qquad (149)$$

- The electric field at the interface for a depletion in Region-2 is given by

$$\mathscr{E}_2(x_0) = \sqrt{\frac{2qN_{D2}(\psi_{b2}-V_2)}{\varepsilon_{s2}}}. \qquad (150)$$

- If the ratio $\varepsilon_{s1}N_{D1} / \varepsilon_{s2}N_{D2}$ is of the order of unity and ψ_{bi} ($\psi_{b1} + \psi_{b2}$) $\gg kT/q$, we obtain

$$\exp\left[\frac{q(\psi_{b1}-V_1)}{kT}\right] \approx \frac{q}{kT}(\psi_{bi}-V) \qquad (151)$$

where V is the total applied voltage and is equal to ($V_1 + V_2$) .

Isotype Heterojunctions

- For the carrier transport, because of the potential barrier as shown in Fig. 28**a**, the conduction mechanism is governed by thermionic emission of *majority* carriers, **electrons** in this case

$$J = \frac{q\psi_{bi}A^*T}{k}\left(1-\frac{V}{\psi_{bi}}\right)\exp\left(\frac{-q\psi_{b1}}{kT}\right)\exp\left(\frac{-q\phi_b}{kT}\right)\left[\exp\left(\frac{qV}{kT}\right)-1\right] = J_0\left[\exp\left(\frac{qV}{kT}\right)-1\right] \qquad (154)$$

- The ***reverse*** current *never* saturates but increases *linearly* with voltage at large – V.
- In the ***forward*** direction, the dependence of J on V can be approximated by an ***exponential*** function $J \propto \exp(qV/\eta kT)$.

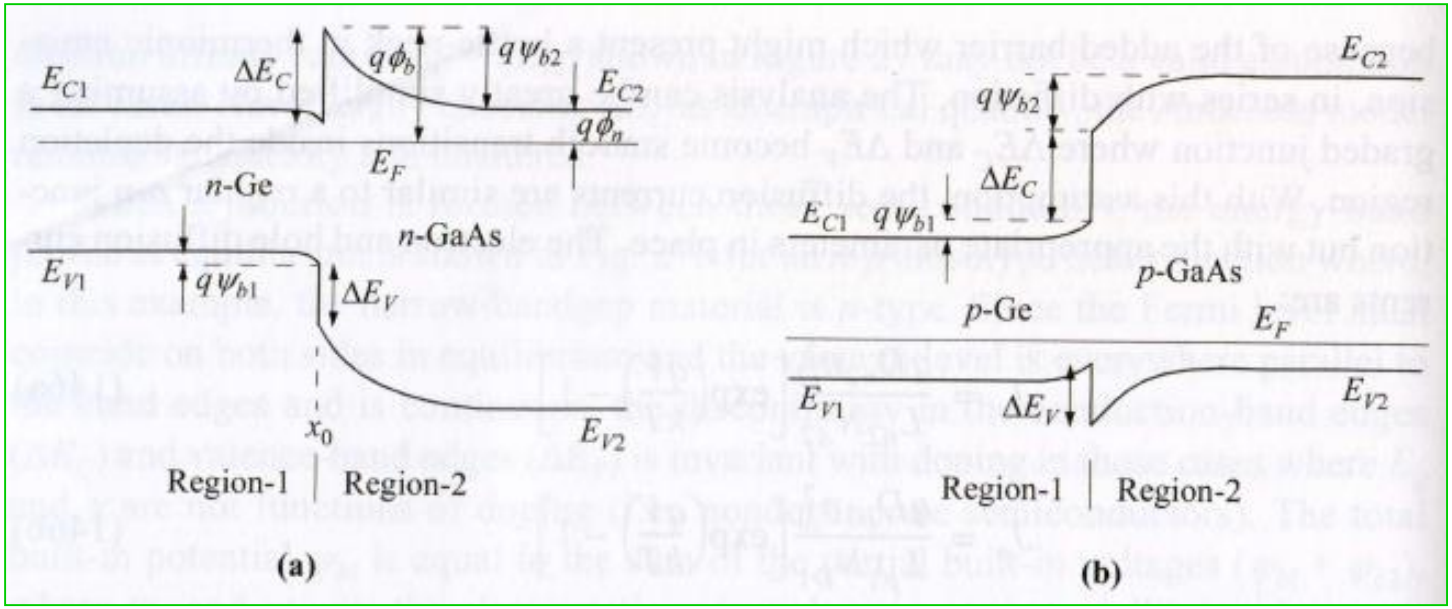


Fig.28 Energy-band diagrams for ideal *(a)* *n-n* and (b) *p-p* isotype heterojunctions.