Session 6摇Nonequilibrium Excess Carriers in Semiconductor

Excess electrons in the conduction band and excess holes in the valence band may exist in addition to the thermal-equilibrium concentrations if an external excitation is applied to the semiconductor. Excess electrons and excess holes do not move independently of each other.[1] They diffuse, drift, and recombine with the same effective diffusion coefficient, drift mobility, and lifetime. This phenomenon is called ***ambipolar transport***. The behavior of ***excess carriers*** is fundamental to the operation of semiconductor devices.

6.1摇Recombination

The thermodynamic ***nonequilibrium excess charges*** can be present in the semiconductor. They can be created by ***carrier injection*** through contacts, an electron beam or the absorption of light with wavelength smaller than the bandgap.[2] After the external excitation is turned off, the semiconductor will return to the equilibrium state. The relaxation of carriers into energetically lower states (and energy release) is called ***recombination***. The term stems from the electron recombining with the hole created after absorption of a photon.[3] However, there are other recombination mechanisms.

6.1.1摇Band to Band Recombination

The ***band-band recombination*** is the relaxation from an electron in the conduction band into the valence band (the empty state there is the hole). In a direct semiconductor, electrons can make an optical transition between the bottom of the conduction band to the top of the valence band. In an indirect semiconductor, this process is only possible with the assistance of a phonon and is thus much less probable.

Fig. 6.1(a) shows the processes of the ***spontaneous recombination*** of an electron of energy E_e and a hole of energy E_h. A similar consideration is made for the ***absorption*** process [Fig. 6.1(b)]. An electron is transferred upon light absorption from a valence-band state (occupied) to a conduction-band state that must be empty. The process is proportional to the light intensity.

In ***stimulated emission***, an incoming photon triggers' the transition of an electron in the conduction band into an empty state in the valence band. The emitted photon is in phase with the initial photon [Fig. 6.1(c)].[4]

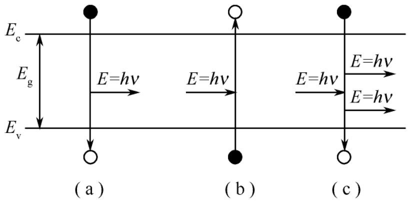


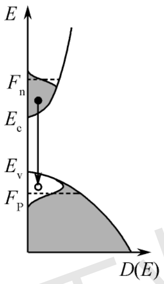
Fig. 6.1摇Processes of band band recombination: (a) spontaneous emission, (b) absorption and (c) stimulated emission. A full (empty) circle represents an occupied (unoccupied) electron state.

6.1.2摇Free-Exciton Recombination

The observation of free-excitons is limited for semiconductors with a small exciton binding energies (such as in GaAs) to low temperatures.[5] However, for large exciton binding energy, recombination from free-excitons is observed even at room temperature, as shown in Fig. 6. 2 for ZnO.

6. 1. 3摇Auger Recombination

In competition with the radiative, bimolecular recombination is the ***Auger recombination*** (Fig. 6. 3).[6] In the Auger process, the energy that is released during the recombination of an electron and hole is not emitted with a photon but, instead, transferred to a third particle. This can be an electron [eeh, Fig. 6. 3(a)] or a hole [hhe, Fig. 6. 3(b)]. The energy is eventually transferred nonradiatively from the hot third carrier via phonon emission to the lattice.[7] The probability for such process is $\propto n^2p$ if two electrons are involved and $\propto np^2$ if two holes are involved. The Auger process is a three-particle process and becomes likely for high carrier density, either through doping, in the presence of many excess carriers, or in semiconductors with small bandgap.[8] Auger recombination is the inverse of the ***impact ionization*** .



摇

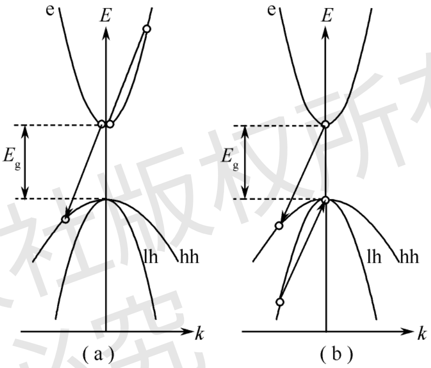


Fig. 6. 2摇Charge-carrier distribution during inversion, necessary for lasing. Shadedareas are populated with electrons. A stimulated transition between an electron and a hole is indicated.

Fig. 6. 3摇Schematic representation of Auger recombination. An electron recombines with a hole and transfers the energy to (a) another electron in the conduction band, (b) another electron in the valence band.

6. 1. 4摇Band-Impurity Recombination

Another recombination process is the capture of carriers by impurities. This process is in competition with all other recombination processes, e. g. the radiative recombination and the Auger mechanism.[9] The ***band-impurity recombination*** is the inverse process to the carrier release from impurities. It is particularly important at low carrier densities, for high dopant concentration and in indirect semiconductors since for these the bimolecular recombination is slow. The theory of capture on and recombination involving impurities is called Shockley-Read-Hall (SRH) kinetics.[10] An example of-band impurity recombination is shown in Fig. 6. 4

6. 1. 5摇Surface Recombination

A surface is typically a source of recombination, e. g. by midgap levels induced by the break of crystal symmetry. The ***surface recombination*** velocity for GaAs is shown in Fig. 6. 5. For InP, if the surface Fermi level is pinned close to midgap, the surface recombination velocity increases from ~ 5×10^3

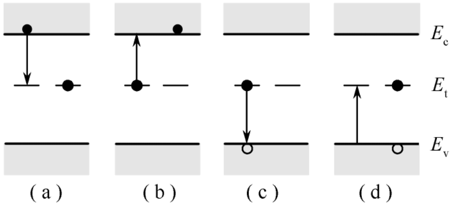


Fig. 6.4摇Band-to-impurity processes at an impurity with one level (left: initial, right: final state in each part): (a) electron capture (from conduction band), (b) electron emission (into conduction band), (c) hole capture (from valence band), (d) hole emission (into valence band). The arrows indicate the transition of the electron.

cm/s for a doping level of $n \sim 3\times10^{15}$ cm^{-3} to $\sim 10^6$ cm/s for a doping level of $n \sim 3\times10^{18}$ cm^{-3}. For Si, the surface recombination rate depends on the treatment of the surface and lies in the range between 10 ~ 10^4 cm/s . The Si-SiO_2 interface can exhibit $S \leqslant 0.5$ cm/s.

6.2摇Minority Carrier Lifetime

Consider what happens when an n-type semiconductor is uniformly illuminated with appropriate wavelength light to photogenerate electron-hole pairs (EHPs). We will now define thermal equilibrium majority and minority carrier concentrations in a extrinsic semiconductor. In general, the subscript "n" or "p" is used to denote the type of semiconductor, and "o" to refer to thermal equilibrium in the dark.[11]

In an n-type semiconductor, electrons are the ***majority carriers*** and holes are the ***minority carriers***.

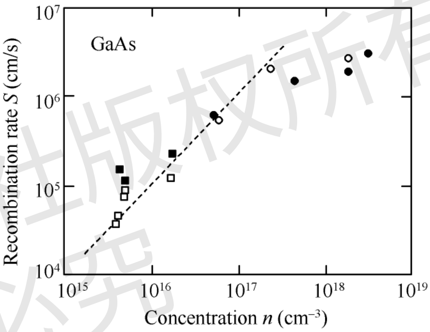


Fig. 6.5摇Surface recombination velocity for GaAs as a function of n-type doping concentration. Different experimental points correspond to different surface treatment methods. Dashed line is a guide to the eye.

n_{n0} is defined as the majority carrier concentration (electron concentration in an n-type semiconductor) in thermal equilibrium in the dark. These electrons constituting the majority carriers, are thermally ionized from the donors.

p_{n0} is defined as the minority carrier concentration (hole concentration in an n-type semiconductor) in thermal equilibrium in the dark.

When we illuminate the semiconductor, we create excess EHPs by photogeneration. Suppose that the electron and hole concentrations at any instant are denoted by n_n and p_n, which are defined as the instantaneous majority (electron) and minority (hole) concentrations, respectively. At any instant and at any location in the semiconductors, we define the departure from the equilibrium by excess concentrations as follows:[12]

Δn_n is the excess electron concentration: $\Delta n_n = n_n - n_{n0}$

Δp_n is the excess hole concentration: $\Delta p_n = p_n - p_{n0}$

Fig. 6. 6 shows a pictorial view of what is happening inside an n-type semiconductor when light is switched on at a certain time and then later switched off again. Obviously when the light is switched off, the condition $p_n = \Delta p_n$ (state B in Fig. 6. 6) must eventually revert back the dark case (state A) where $p_n = p_{n0}$. In other words, the excess minority carriers Δp_n and excess majority carriers Δn_n must be removed. This removal occurs by recombination. Excess hole recombine with the electrons available and disappear. This, however, takes time because the electrons and holes have to find each other. In order to describe the rate of recombination, we introduce a temporal quantity, denoted by τ and called the ***minority carrier lifetime*** (mean recombination time), which is defined as follows: τ is the average time a hole exists from its generation to its recombination, that is, the mean time the hole is free before recombining with an electron.[13] An alternative and equivalent definition is that $1/\tau$ is the average probability per unit time that a hole will recombine with an electron. We must remember that the recombination process occurs through recombination centers, so the recombination time τ will depend on the concentration of these centers and their effectiveness in capturing the minority carriers.[14] Once a minority carrier has been captured by a recombination center, there are many majority carriers available to recombine with it, so τ in an indirect process is independent of the majority carrier concentration. This is the reason for defining the recombination time as a minority carrier lifetime.

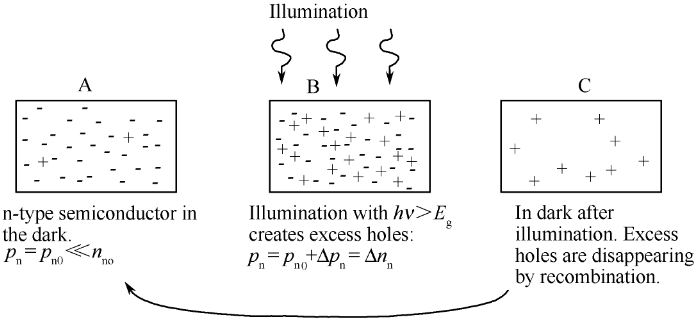


Fig. 6. 6摇illumination of an n-type semiconductor results in excess electron and hole concentrations. After the illumination, the recombination process restores equilibrium; the excess electrons and holes simply recombine.

We should note that the recombination time τ depends on the semiconductor material, impurities, crystal defects, temperature, and so forth, and there is no typical value to quote. It can be anywhere from nanoseconds to seconds. Later it will be shown that certain applications require a short τ, as in fast switching of pn junctions, whereas others require a long τ, for example, persistent luminescence.[15]

6. 3摇Ambipolar Transport

The generation and recombination rates of excess carriers are important parameters, but how the excess carriers behave with time and in space in the presence of electric fields and density gradients is of equal importance. As mentioned in the preview section, the excess electrons and holes do not move independently of each other, but they diffuse and drift with the same effective diffusion coefficientand with the same effective mobility. This phenomenon is called ambipolar transport.

If a pulse of excess electrons and a pulse of excess holes are created at a particular point in a semiconductor with an applied electric field, the excess holes and electrons will tend to drift in opposite directions. However, because the electrons and holes are charged particles, any separation will induce an internal electric field between the two sets of particles. This internal electric field will create a force attracting the electrons and holes back toward each other. This effect is shown in Fig. 6.7.

This E-field may be written as

$$E = E_{app} + E_{int} \qquad (6.1)$$

where E_{app} is the applied electric field and E_{int} is the induced internal electric field. Since the internal E-field creates a force attracting the electrons and hole, this E-field will hold the pulses of excess electrons and excess holes together.

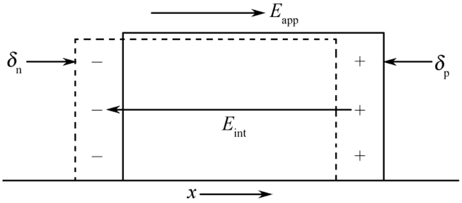


Fig. 6.7摇The creation of an internal electric field as excess electrons and holes tend to separate.

Reading Materials

Why do the deep levels act as effective centres of recombination?

Let us imagine an electron and hole wandering in the crystal. In order to meet, recombine and disappear, it is necessary that they should be close to each other in the vicinity of one and the same atom of the crystal lattice.[16] Such a situation is, in general, possible but seldom occurs.

Let us now assume that there is an impurity centre in the crystal whose activation energy ΔE is great. Should an electron appear in the vicinity of this centre, it is sure to be trapped by the impurity centre.[17]

The centre will keep the electron trapped until a hole appears in the vicinity. As soon as that happens the electron and hole recombine. The "killer" has committed this task of annihilating the electron-hole pair, and is ready to start all over again.[18]

Sometimes it is important for electrons and holes to perish in the device as soon as possible. It is often quite essential for the fast switching of semiconductor devices. Then impurities creating effective recombination centres should be incorporated into the material. Sometimes, on the contrary, electrons and holes must live long. In this case, the semiconductor is to be thoroughly purified.

Words and Expressions

constitute *vt.*	制定（法律），建立（政府），组成，任命
annihilate *vt.*	消灭

Glossary of Important Term

ambipolar transport	双极输运
excess carriers	过剩载流子
nonequilibrium excess charges	非平衡过剩载流子
generation	产生
recombination	复合
carrier injection	载流子注入
band-band recombination	带间复合（直接复合）
spontaneous recombination	自发复合
absorption	吸收
stimulated emission	受激发射
Auger recombination	俄歇复合
impact ionization	碰撞离化
band-impurity recombination	能带—杂质能级复合（SRH 复合）
surface recombination	表面复合
majority carriers	多子，多数载流子
minority carriers	少子，少数载流子
minority carrier lifetime	少子寿命

Notes

1. Excess electrons and excess holes do not move independently of each other.

提示: independently of …表示“与……无关”。each other 表示“相互”。

2. They can be created by ***carrier injection*** through contacts, an electron beam or the absorption of light with wavelength smaller than the bandgap.

提示: by 引导的介宾短语作方式状语,包括三个并列的宾语。

3. The term stems from the electron recombining with the hole created after absorption of a photon.

提示: stem from 表示“出自、起源于”;electron recombining with the hole 是动名词作主句的宾语;created after absorption of a photon 是过去分词作定语,修饰 hole。

4. The emitted photon is in phase with the initial photon [Fig. 6.1(c)].

提示: in phase 表示“同相地,协调地”。

5. The observation of free-excitons is limited for semiconductors with a small exciton binding energies (such as in GaAs) to low temperatures.

提示: be limited to…表示“被限制于……”,句子在 limited 和 to 之间插入了一个介词短语(for …);如果直译为“对自由激子的观察只能限制在低温下”,显得生硬,翻译时可做适当调整。

6. In competition with the radiative, bimolecular recombination is the ***Auger recombination*** (Fig. 6.3).

提示: 句中的词汇意义应与物理过程相结合。radiative 指复合中伴有发射过程;bimolecular 指复合过程中有双粒子参与;句子采用了倒装的形式,真正的主语为 the Auger recombination, In competition with the radiative, bimolecular recombination 为表语;In competition with…表示“与……存在竞争关系”,the radiative, bimolecular recombination 指前文提到的伴随发射过程的双粒子复合,在翻译时可以根据物理意义做出调整。

7. The energy is eventually transferred nonradiatively from the hot third carrier via phonon emission to the lattice.

提示: the hot third carrier 的理解应结合上下文,上文提到“在俄歇复合的过程中,电子与空穴复合放出的能量并不以发射一个光子的形式释放,而是将能量转移给第三个粒子”,因此 the hot third carrier 指上文提到的接受了能量的第三个粒子(具有较高能量的热载流子);from the hot third carrier、via phonon emission 和 to the lattice 这三个介词短语都作状语,表示能量传递是“从何处”、“通过什么方式”和“到哪里”。

8. The Auger process is a three-particle process and becomes likely for high carrier density, either through doping, in the presence of many excess carriers, or in semiconductors with small bandgap.

提示: 句中 is 和 becomes 是并列的系动词,句中的 likely 是形容词,作 becomes 的表语,表示“很可能的”。either…or…引导的状语指出俄歇复合容易发生的两种情况。in the presence of many excess carriers 是插入语,表示掺杂可能导致的一种容易引起俄歇复合的结果。

9. This process is in competition with all other recombination processes, e. g. the radiative recombination and the Auger mechanism.

提示: in competition with 可表示几种复合机制存在竞争的关系。

10. The theory of capture on and recombination involving impurities is called Shockley-Read-Hall (SRH) kinetics.

提示: of 引导的介宾短语作主语 theory 的定语,这一介宾短语的宾语包括并列的两部分:capture on impurities（杂质能级上的俘获）和 recombination involving impurities（与杂质相关的复合），由于两者中的介宾短语（on impurities 和 involving impurities）共用宾语impurities，所以前者将其省略。

11. In general, the subscript “n” or “p” is used to denote the type of semiconductor, and “o” to refer to thermal equilibrium in the dark.

提示：subscript 指参数中的下标。

12. At any instant and at any location in the semiconductors, we define the departure from the equilibrium by excess concentrations as follows:

提示：define 为谓语，departure 为宾语。

13. τ is the average time a hole exists from its generation to its recombination, that is, the mean time the hole is free before recombining with an electron.

提示：that is 的意思为“换句话说、更确切地说”。

14. We must remember that the recombination process occurs through recombination centers, so the recombination time τ will depend on the concentration of these centers and their effectiveness in capturing the minority carriers.

提示：so 引导从句，表示一定的因果关系。

15. Later it will be shown that certain applications require a short τ, as in fast switching of pn junctions, whereas others require a long τ, for example, persistent luminescence.

提示：it 是先行主语，that 引导的从句才是真正主语，因为真正主语过长，所以采用这种形式；that 引导的从句包含两个部分，用 whereas 连接表示转折的意思；as in fast switching of pn junctions 和 for example, persistent luminescence 是插入语，分别对两种情形给出了例子；由于句式复杂，插入语太多，可以在翻译时适当调整。

16. In order to meet, recombine and disappear, it is necessary that they should be close to each other in the vicinity of one and the same atom of the crystal lattice.

提示：meet、recombine、disappear 是并列的关系，都是动词不定式，后两者的前面省略了 to；it 是先行主语，用法同上一个注释；one and the same atom 强调必须是在同一原子附近出现。

17. Should an electron appear in the vicinity of this centre, it is sure to be trapped by the impurity centre.

提示：should 前置，表示一种强调语气。

18. The “killer” has committed this task of annihilating the electron-hole pair, and is ready to start all over again.

提示：原文作者为了更为生动地说明深能级杂质的作用，采用了拟人的手法，将深能级杂质视为一个“杀手”。在翻译成中文时，不妨灵活地处理。

Exercises

1. Translate the reading material into Chinese.

2. Answer the following questions in English.

(1) Describe the concept of excess generation and recombination.

(2) Why are the electron generation rate and recombination rate equal in the thermal equilibrium?

(3) Describe the concept of an excess carrier lifetime.

Session 7摇The pn Junction（Ⅰ）

7.1摇Introduction

Most ***semiconductor devices*** contain at least one junction between p-type and n-type semiconductor regions. Semiconductor device characteristics and operation are intimately connected to these pn junctions, so considerable attention is devoted initially to this basic device.[1]摇The ***pn junction diode*** itself provides characteristics that are used in ***rectifiers*** and ***switching circuits***. In addition, the analysis of the pn junction device establishes some basic terminology and concepts that are used in the discussion of other semiconductor devices. The fundamental analysis techniques used for the pn junction will also be applied to other devices. Understanding the physics of the pn junction is, therefore, an important step in the study of semiconductor devices.

The pn junction must not contain more than a small number of imperfection.[2]摇In practice this means either that the device has been made from a slice cut from a large single crystal, parts of which have been transformed by diffusing or ion-implanting doping atoms from the surface, or that new material has been grown epitaxially to extend a crystal substrate and to allow the including of a pn junction.[3]

7.2摇Basic Structure of the pn Junction

Fig. 7.1(a) schematically shows the pn junction. It is important to realize that the entire semiconductor is a single-crystal material in which one region is doped with acceptor impurity atoms to form the p region and the adjacent region is doped with donor atoms to form the n region.[4]摇The interface separating the n and p regions is referred to as the ***metallurgical junction.***

The impurity doping concentrations in the p and n regions are shown in Fig. 7.1(b). For simplicity, we will consider a ***step junction*** in which the doping concentration is uniform in each region and there is an abrupt change in doping at the metallurgical junction. Initially, at the metallurgical junction, there is a very large density gradient in both the electron and hole concentrations. Majority carrier electrons in the n region will begin diffusing into the p region and majority carrier holes in the p region will begin diffusing into the n region. If we assume there are no external connections to the semiconductor, then this diffusion process cannot continue indefinitely. As electrons diffuse from the n region, positively charged donor atoms are left behind. Similarly, as holes diffuse from the p region, they uncover negatively charged acceptor atoms. The net positive and negative charges in the n and p regions induce an electric field in the region near the metallurgical junction, in the direction from the positive to the negative charge, or from the n to the p region.[5]

The net positively and negatively charged regions are shown in Fig. 7.2. These two regions are referred to as the ***space charge region***. Essentially all electrons and holes are swept out of the space charge region by the electric field. Since the space charge region is depleted of any mobile charge, this region is also referred to as the ***depletion region***: these two terms will be used interchangeably. Density gradients still exist in the majority carrier concentrations at each edge of the space charge region. We can think of a density gradient as producing a ‘diffusion force’ that acts on the majority carriers. These diffusion forces, acting on the electrons and holes at the edges of the space charge region, are shown in the figure. The electric field in the space charge region produces another force on the electrons and holes which is in the opposite direction to the diffusion force for each type of particle.^{6} In thermal equilibrium, the diffusion force and the E-field force exactly balance each other.

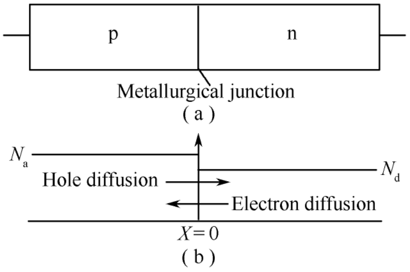


Fig. 7. 1摇(a) Simplified geometry of a pn junction; (b) doping profile of an ideal uniformly doped pn junction.

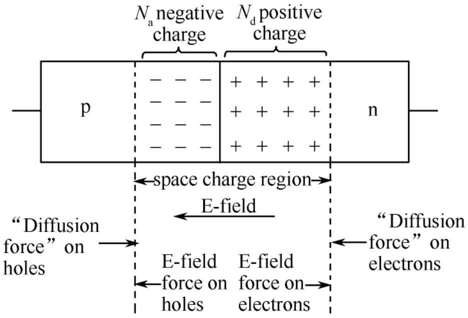


Fig. 7. 2摇The space charge region, the electric field, and the forces acting on the charged carriers.

7. 3摇Energy Bands for a pn Junction

The electron-energy diagram conveys a great deal of information about a semiconductor device, and it is well worthwhile learning how to construct these diagrams. It is not a good idea to try to memorize the diagram for every device-the simple stages of construction are what must be mastered.^{7}

(1) Start by putting the Fermi level on paper for one of the layers of semiconductor-any one will do.^{8}

(2) Build the band round this Fermi level. The conduction band is close to the Fermi level for n-type material, but the valence band is close in p-type material.^{9}

(3) Draw the other Fermi levels at the right height on the diagram, allowing for applied voltages. The more positive of two layers is nearer the bottom of the page.^{10}

(4) Complete these bands, keeping the gap between conduction and valence bands constant.

(5) Join up the conduction band from each layer to the next, using S-shaped double curves, and do the same for the valence band.^{11}

(6) Fill in details such as free carriers, doping ions, and applied voltages. Remember that doping ions are present in depletion layers, but that large numbers of free carriers are not.^{12}

A pn junction and its associated energy band is shown in Fig. 7. 3.

7. 4摇Ideal Current-Voltage Relationship

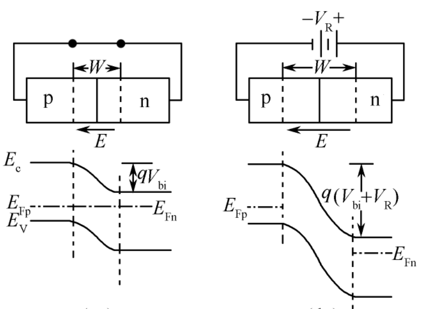
1. Assumptions

The ideal current-voltage relationship of a pn junction is derived on the basis of the following four assumptions.

(1) The abrupt depletion layer approximation applies. The space charge regions have abrupt boundaries and the semiconductor is neutral outside of the depletion region.

(2) The Maxwell-Boltzmann approximation applies to carrier statistics.

(3) The concept of *low injection* applies.



(a) (b) (c)

Fig. 7. 3摇 (a) zero bias, (b) reverse bias, and (c) forward bias.

(4) The individual electron and hole currents are constant throughout the depletion region.

2. Ideal-diode equation

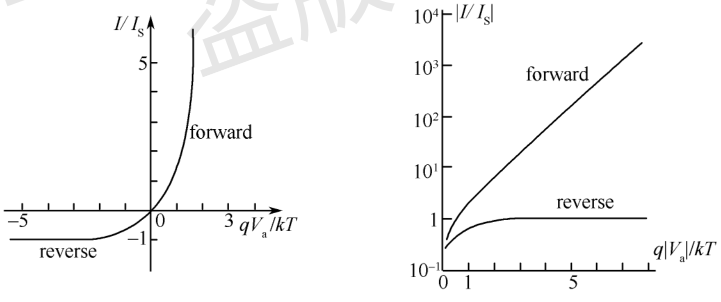
By solving the ***continuity equation*** for minority carriers in the quasi-neutral regions near a pn junction based on the above assumptions, it is possible to obtain ***current-voltage characteristics*** in the steady state for several simple cases with important practical applications. 13摇 These solutions lead to the following celebrated ***Shockley equation***, or ***ideal diode law***.

$$I_D = I_S[\exp(qV_a/kT)] - 1 \tag{7.1}$$

where I_D is the diode current and I_S the diode reverse ***saturation current***.

3. The current-voltage characteristic

The ideal *I-V* characteristic predicted by the above equation is illustrated in Figs. 7.4(a) and (b) in the linear and semilog plots respectively. In these universal curves I and V are plotted as multiples of I_S and kT/q respectively. 14摇 In the forward direction for $V_a > 3kT/q$, the rate of current rise is constant [Fig. 7.4 (b)]; at 300 K for every decade change of current, the voltage changes by 59.5 mV ($= 2.3\ kT/q$). 15摇 Note that the reverse current saturates for reverse voltages of more than about $-3(kT/q)$. 16



(a) linear scale (b) semilog scale

Fig. 7. 4摇 The current-voltage characteristic on two different scales.

7.5摇 Characteristics of a Practical Diode

The Shockley equation adequately predicts the current-voltage characteristics of germanium p-n junctions at low current densities. For Si and GaAs p-n junctions, however, the ideal equation can only give qualitative agreement. The departures from the ideal are mainly due to the following factors:

(1) The generation and recombination of carriers within the narrow depletion region

Based on the quantitative analysis, the current-voltage relation for a real pn junction is accurately modeled by an equation having two terms:

$$I = I_S[\exp(qV_a/kT) - 1] + I_{r0}[\exp(qV_a/2kT) - 1] \qquad (7.2)$$

The first term, which is normally dominant in forward bias, is exactly as given in Eq. (7.1). The second term, due to carrier generation and recombination in the depletion region, has an extra factor of 2 in the denominator of the exponent. Hence it varies roughly as the square root of the first term. Thus at small forward currents, and in reverse bias (i. e. $V_a < 0$), the second term becomes an important factor [Fig. 7.5, curve (a)]. The pre-factor I_{r0} is proportional to the volume in which this generation and recombination occurs, i. e. the volume of the depletion regions. Hence I_{r0} increases with the increase in ***depletion-layer width***, which occurs in ***reverse bias***, and the reverse current is thus prevented from saturating in a real diode [Fig. 7.5, curve (e)].[17]

(2) The ***high-injection*** condition at high current densities (under the forward-bias condition)

Because of the high-injection the current then becomes roughly proportional to exp$(qV_a/2kT)$, as shown in Fig. 7.5, curve (c).

(3) Voltage drops across the ***series resistance***

At high forward currents the diode current is limited more by the small series resistance R_s in the bulk semiconductor. As a result, the voltage across a junction carrying a high current is smaller by an amount IR_s than the total applied voltage as shown in Fig. 7.5, curve (d).[18]摇 Thus Eq. (7.3) is a more accurate representation at high currents:

$$I = I_S\{\exp[q(V_a - IR_s)/kT] - 1\} \qquad (7.3)$$

(4) The ***breakdown*** mechanism

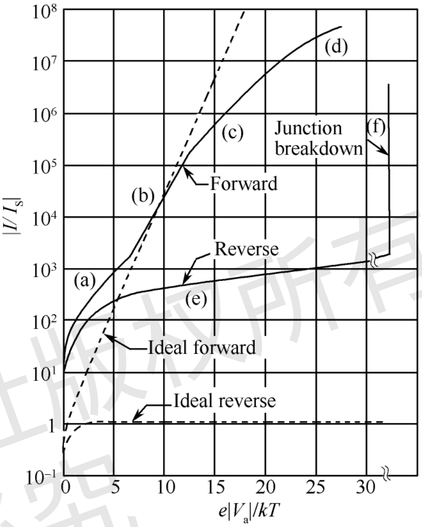


Fig. 7.5摇Current-voltage characteristics of a practical Si diode.

Beyond a few volts of reverse bias, the diode current rises dramatically as shown in Fig. 7.5, curve(f), a phenomenon termed breakdown discussed in next session.

Reading Materials

So-called ‘forward on voltage’

In the forward direction, the current rises very rapidly but smoothly as shown in Fig. 7.4(a), and there is no real ‘***forward on voltage***’ as sometimes assumed in books on circuit theory.[19]摇 However, as Fig. 7.6 shows, in strong ***forward bias***, the diode voltage varies rather little over a wide current range.[20]

For example, a diode designed to carry a current I at, say, 0.6V (i. e. $eV_a/kT = 24$) will carry a current of only 0.001 I at 0.427 V ($eV_a/kT = 17.1$).[21]摇 For practical purposes the diode is ‘off’ below this voltage.

The difference in the ‘on’ voltages of Ge and Si diodes also needs explanation. Note the ‘on’ current for a

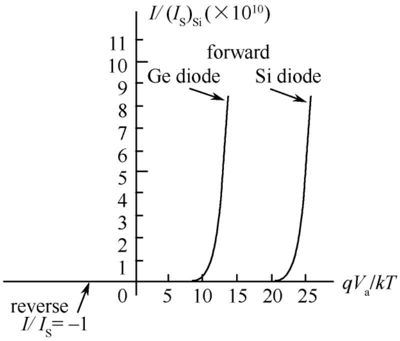


Fig. 7.6摇The current-voltage characteristic.

Ge diode with the same applied voltage as an Si diode（e. g. 0.7V）is the same multiple of its reverse saturation current I_S. Since I_S is proportional to the minority carrier density, which in turn depends on n_i^2, the reverse saturation current of a Ge diode-and hence the forward current- is about 10^8 times that of an Si diode with the same doping levels.[22]摇Thus a Ge diode ‘turns on’ at a lower voltage-about 0.25V is typical. Note that, to make a Si diode turn on at this voltage would need either a doping level about 10^8 times smaller, or an area 10^8 times bigger: neither is a practical proposition.[23]

Words and Expressions

intimately摇*adv*. 密切地
terminology摇*n*. 术语
imperfection摇*n*. 不完整性,缺陷
indefinitely摇*adv*. 无限期地
interchangeably摇*adv*. 可交换地
respectively摇*adv*. 分别地
dramatically摇*adv*. 引入注目地,明显地
semilog 摇*n*. 半对数
decade摇*n*. 十(进制,位)
denominator摇*n*. 分母
exponent 摇*n*. 指数
square root 摇*n*. 平方根

proposition 摇*n*. 提议
characteristics摇*n*. 特性
operation 摇*n*. 工作
celebrated 摇*adj*. 著名的
profile 摇*n*. 剖面
quantitative analysis 摇定量分析
techniques 摇*n*. 技术
to allow for 摇考虑
to be referred to… as… 摇将……称为……
be depleted of … 摇耗尽……
be worthwhile摇 ~ ing 摇值得……
to prevent … from ~ ing 摇阻止……(做……)

Glossary of Important Term

semiconductor device摇半导体器件
pn junction摇pn 结
diode摇二极管
rectifier摇整流器
switching circuit摇开关电路
metallurgical junction摇冶金结
step junction摇突变结
space charge region摇空间电荷区
depletion region摇耗尽层
depletion-layer width摇耗尽层宽度
reverse bias摇反向偏置

forward bias摇正向偏置
low injection摇小注入
high-injection摇大注入
continuity equation 摇连续性方程
current-voltage characteristics摇伏安特性
Shockley equation摇肖克莱方程
ideal diode law摇理想二极管定律
saturation current摇饱和电流
series resistance 摇串联电阻
breakdown摇击穿
forward on voltage摇正向导通电压

Notes

1. Semiconductor device characteristics and operation are intimately connected to these pn junctions, so considerable attention is devoted initially to this basic device.

提示:be connected to …(与……相关联)以及 be devoted to…(用于……)这两个词组都有一个副词修饰,但是副词的位置可以放在两种不同位置。

2. The pn junction must not contain more than a small number of imperfection.

提示:句子为否定形式,表示强调。翻译时用肯定语气表达原义,汉语表达比较简洁。

3. In practice this means either that the device has been made from a slice cut from a large single crystal, parts of which have been transformed by diffusing or ion-implanting doping atoms from the

surface, or that new material has been grown epitaxially to extend a crystal substrate and to allow the including of a pn junction.

提示：either that 和 or that 引导两个并列的宾语从句。其中第一个宾语从句中 parts of which 引导一个定语从句，结合专业内容分析，这个定语从句修饰 slice，而不是修饰直接位于定语从句前面的 single crystal。第二个宾语从句中 and 连接两个动词不定式 to extend 和 to allow，作状语。

4. It is important to realize that the entire semiconductor is a single-crystal material in which one region is doped with acceptor impurity atoms to form the p region and the adjacent region is doped with donor atoms to form the n region.

提示：in which 引导定语从句修饰 a single-crystal material。定语从句中又包括由 and 连接的两个并列句。and 可以翻译为 “而”，不一定机械地翻译为 “和”。

5. The net positive and negative charges in the n and p regions induce an electric field in the region near the metallurgical junction, in the direction from the positive to the negative charge, or from the n to the p region.

提示：基于物理含义考虑，翻译时将 “冶金结附近” 放在全句开始；in the n and p regions 修饰 The net positive and negative charges，翻译时根据物理含义明确翻译为 “n 区一侧的净正电荷” 和 “p 区一侧的净负电荷”；句子中 or 表示对前面内容的另一种表达，不是从两种中选择一种，因此翻译为 “或者说”，也可以翻译为 “即”，不要简单地翻译为 “或”。

6. The electric field in the space charge region produces another force on the electrons and holes which is in the opposite direction to the diffusion force for each type of particle.

提示：force on…中的介词 on 表示 “对……的作用力”；which 引导的定语从句修饰前面的名词 force，因为 force 还带有定语，因此定语从句位置没有紧跟在被修饰词 force 的后面。

7. It is not a good idea to try to memorize the diagram for every device-the simple stages of construction are what must be mastered.

提示：memorize 含有 “死记” 的意思；what must be mastered 是名词从句作表语。

8. Start by putting the Fermi level on paper for one of the layers of semiconductor-any one will do.

提示：any one will do 表示任何人都能做。因为对于突变结，同一个区域中费米能级为常数，绘制时只要绘制一条水平线，而且开始时对这条水平线位置没有制约，因此对于只要绘一条水平线的要求 “任何人都能做”。

9. The conduction band is close to the Fermi level for n-type material, but the valence band is close in p-type material.

提示：按照含义，第二个 close 的后面应该有 to the Fermi level，因为前面已经出现过，而且结构相同，因此可以省略。

10. Draw the other Fermi levels at the right height on the diagram, allowing for applied voltages. The more positive of two layers is nearer the bottom of the page.

提示：at the right height 表示 “正确的高度”；allowing for applied voltages 进一步解释说明要根据外加电压确定正确的高度，因为按照物理含义，两个区域之间费米能级之差就等于外加电压乘以电子电荷；后一句是比较级，进一步说明两个费米能级之间谁的位置更高；page 指绘制能带图的页面，翻译时可以直接翻成图的底部。

11. Join up the conduction band from each layer to the next, using S-shaped double curves, and do the same for the valence band.

提示：由图 7.3 可见，不管是什么偏置状态，p 区和 n 区界面附近能带图的导带底之间呈现 S

形，句子用 using S-shaped double curves 一方面点明了 S 形，而且更加形象地说明 S 形包括了两个方向转弯，即其中的 curve 表示转弯，double curves 表示“双弯”，不能翻译为双曲线，也不能翻译为两条曲线。

12. Remember that doping ions are present in depletion layers, but that large numbers of free carriers are not.

提示：注意本句中 Remember 与第 7 句中 memorize 含义的区别。Remember 是记住，memorize 带有“死记”的意思；这是祈使句，带有两个宾语从句，这两个宾语从句用 but 连接，表示转折。

13. By solving the ***continuity equation*** for minority carriers in the quasi-neutral regions near a pn junction based on the above assumptions, it is possible to obtain ***current-voltage characteristics*** in the steady state for several simple cases with important practical applications.

提示：动名词短语 solving the ***continuity equation*** 是 By 的介词宾语，其后面又带有多个修饰短语；主句中 it 是先行主语，代表从 to obtain…到句尾的动词不定式短语；with important practical applications修饰 several simple cases，可翻译为“几种具有重要实用价值的简单情况”。

14. In these universal curves I and V are plotted as multiples of I_S and kT/q respectively.

提示：these universal curves 表示“通用曲线”；注意，如果对照英语直译主句，会存在汉语表达的内容不清晰、可读性差的问题。应该在正确理解主句表示的内容后，采用意译的方式，表达的含义将清晰、明确。

15. In the forward direction for $V_a > 3kT/q$, the rate of current rise is constant [Fig. 7.4 (b)]; at 300 K for every decade change of current, the voltage changes by 59.5mV (= 2.3 kT/q).

提示：every decade change of current 表示电流变化一个数量级。

16. Note that the reverse current saturates for reverse voltages of more than about $-3(kT/e)$.

提示：由于负数的绝对值越大，表示负数代数值越小，因此该句不能翻译为大于 $-3(kT/q)$，而要翻译为比 $-3(kT/q)$ 更负。

17. Hence I_{r0} increases with the increase in ***depletion-layer width***, which occurs in ***reverse bias***, and the reverse current is thus prevented from saturating in a real diode [Fig. 7.5, curve (e)]。

提示：该句包含由 and 连接的两个句子；which occurs in ***reverse bias*** 是非限制性定语从句，修饰前面句子，起补充说明作用；to be prevented from 是固定搭配，表示“不……”；注意括号内容[Fig. 7.5, curve (e)] 紧跟在 diode 后面，但是括号内容并不是补充说明 diode 而是补充说明这个句子，因此注意中文翻译时括号的位置。

18. As a result, the voltage across a junction carrying a high current is smaller by an amount IR_s than the total applied voltage as shown in Fig. 7.5, curve (d).

提示：is smaller by an amount IR_s than…表示“比……小 IR_s”；结合课文的含义，以及图 7.5 表示的内容，curve (d)应该翻译为“(d) 段曲线”，而不要翻译为曲线(d)。

19. In the forward direction, the current rises very rapidly but smoothly as shown in Fig. 7.4(a), and there is no real ‘forward on voltage’ as sometimes assumed in books on circuit theory.

提示：句子的 and 具有转折的含义，因此将 and there is no 翻译为“并不存在”；books on …表示“关于……方面的书”；本句中 assume 表示“呈现，表现”的意思，不要总是翻译为“假设”。

20. However, as Fig. 7.6 shows, in strong ***forward bias***, the diode voltage varies rather little over a wide current range.

提示：in strong ***forward bias*** 表示正偏电压较大，其中的 strong 不要翻译为“强”；rather 是副词，修饰 little，表示相当小。

21. For example, a diode designed to carry a current I at, say , 0.6 V（i. e. $eV/kT = 24$）will carry a current of only 0.001 I at 0.427 V（$qV/kT = 17.1$）.

提示：这是个简单句，主语是 a diode，谓语是 will carry，其中主语后面带有过去分词 designed 引导的限制性定语；句子的 say 是插入语，假设一种情况，可以翻译为“例如”，并且具有以该假设情况为例说明某个结论的作用；i. e. 是拉丁语，代表 that is，可翻译为“也就是”或者“即”；本句采用“意译”的方式可以很简洁地将全句的含义表达清楚。

22. Since I_S is proportional to the minority carrier density, which in turn depends on n_i^2, the reverse saturation current of a Ge diode-and hence the forward current-is about 10^8 times that of an Si diode with the same doping levels.

提示：which in turn depends on n_i^2 是非限制性定语从句，修饰 the minority carrier density；and hence the forward current 引导的句子中谓语与 the reverse saturation current of a Ge diode 的谓语相同，因此省略了。

23. Note that, to make an Si diode turn on at this voltage would need either a doping level about 10^8 times smaller, or an area 10^8 times bigger: neither is a practical proposition.

提示：这是由 Note 引导的祈使句，后面是宾语从句；宾语从句中谓语 need 后面是由 either …, or…引导的两个宾语；代词 neither 作主语引导的简单句对前面内容起补充说明的作用，其中 neither 的含义是 not either，用于对前面 either …, or…内容的否定；proposition的含义是 suggestion。

Exercises

1. Translate the reading material into Chinese.

2. Draw the energy band diagram of a zero-biased, forward-biased and reverse-biased pn junction.

3. Describe why and how the space charge region is formed, and what happens to the parameters of the space charge region when a reverse bias voltage is applied.

4. What is the difference between the I-V characteristics of a practical pn junction diode and the ideal pn junction?

5. Summary the main factors causing the departures of a practical pn junction diode from the ideal pn junction.

6. How to understand the so-called ‘forward on voltage’ of a pn junction?

Session 8摇The pn Junction (Ⅱ)

8.1摇Breakdown in pn Junction

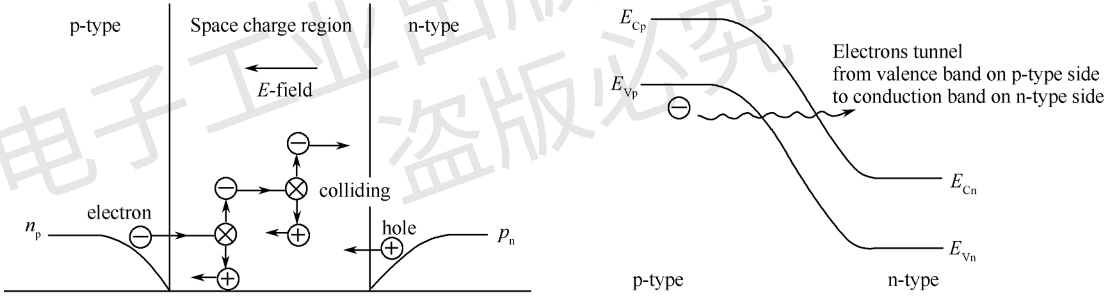
1. Breakdown

Breakdown in pn junction is the rapid increase of current as the reverse bias voltage is increased beyond some safe limit. Breakdown may result in the destruction of a device but this is not inevitable.[1] Damage is caused by local melting, and if the current is held to a small value by an external resistance, or is spread over a large area by careful design and manufacture, then the temperature reached are not high enough for Silicon or SiO_2 to melt, nor for impurities to diffuse out of their intended positions.[2]

There are two distinct breakdown mechanisms-***avalanche breakdown*** and ***Zener or tunnel breakdown*** .

2. Avalanche breakdown

The avalanche breakdown process occurs when electrons and/or holes, moving across the space charge region, acquire sufficient energy from the electric field to create electron-hole pairs by ***colliding***with atomic electrons within the depletion region. The newly created electrons and holes move in opposite directions due to the electric field and thereby add to the existing reverse-bias current. In addition, the newly generated electrons and/or holes may acquire sufficient energy to ***ionize*** other atoms, leading to the avalanche process.[3] This avalanche process is schematically shown in Fig. 8.1(a). For most pn junctions, the predominant breakdown mechanism will be the avalanche effect.



(a) (b)

Fig. 8.1摇(a) avalanche breakdown and (b) Zener breakdown mechanism.

3. Zener breakdown

Zener breakdown occurs in highly doped pn junctions through a ***tunneling mechanism*** . In a highly doped junction, the conduction and valence bands on opposite sides of the junction are sufficiently close during reverse bias that electrons may tunnel directly from the valence band on the p side into the conduction band on the n side, as shown in Fig. 8.1(b).

8.2摇Small-Signal Diffusion Resistance of the pn Junction

We have been considering the dc characteristics of the pn junction diode.[4] When semiconductor devices with pn junctions are used in ***linear amplifier circuits*** , for example, ***sinusoidal signals*** are superimposed on the dc currents and voltages, so that the small-signal characteristics of the pn junction become important.[5]

Assume that the diode is forward-biased with a dc voltage V_o producing a dc diode current I_{DQ}. If we now superimposes small, low-frequency sinusoidal voltage as shown in Fig. 8.2, then a small sinusoidal current will be produced, superimposed on the dc current. The ratio of sinusoidal current to sinusoidal voltage is called the ***incremental conductance*** . In the limit of a very small sinusoidal current and voltage, the small-signal incremental conductance is just the slope of the dc current-voltage curve, or

$$g_d = \left.\frac{dI_D}{dV_a}\right|_{V_a = V_o} \tag{8.1}$$

If we assume that the diode is biased sufficiently far in the forward-bias region, then the incremental conductance becomes

$$g_d = \left.\frac{dI_D}{dV_a}\right|_{V_a = V_o} = \left(\frac{q}{kT}\right) I_S \exp\left(\frac{qV_o}{kT}\right) \approx \frac{I_{DQ}}{V_t} \tag{8.2}$$

where $V_t = kT/q$, and is defined as the ***thermal voltage*** .

The small-signal incremental resistance is then the reciprocal function, or

$$r_d = \frac{V_t}{I_{DQ}} \tag{8.3}$$

where I_{DQ} is the dc quiescent diode current.

The incremental resistance decreases as the bias current increases, and is inversely proportional to the slope of the *I-V* characteristic as shown in Fig. 8.2. The incremental resistance is also known as the ***diffusion resistance*** .

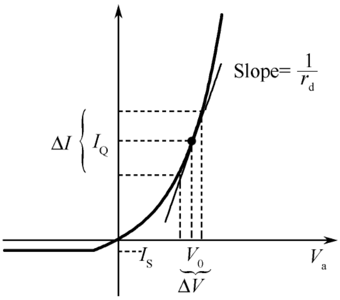


Fig. 8.2摇Curve showing the concept of the small-signal diffusion resistance.

8.3摇Junction Capacitance

Since we have a separation of positive and negative charges in the depletion region, a capacitance is associated with the pn junction. Fig. 8.3 shows the charge densities in the depletion region for applied reverse-bias voltages of V_R and $V_R + dV_R$. An increase in the reverse-bias voltage V_R will uncover additional positive charges in the n region and additional negative charges in the p region. The junction capacitance is defined as $C_j = dQ/dV_R$ and the following expression can be derived:

$$C_j = C_{j0}/(1 - V_a/V_{bi})^{1/2} \tag{8.4}$$

where C_{j0} is the capacitance at zero applied voltage.

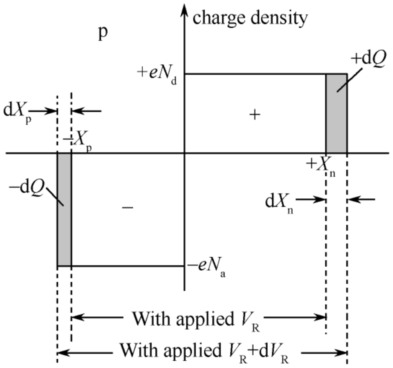


Fig. 8.3摇Illustrating the variation of depletion layer charge and width with junction voltage.

It is seen that the junction capacitance decreases as the bias becomes more negative.[6] When $V_{bi} \ll |V_a|$, C_j decreases roughly as the inverse square root of the reverse voltage.[7]

The junction capacitance C_j can also be expressed in terms of the depletion layer thickness, and we find that we can write

$$C_j = A\varepsilon_S / W(V_a) \tag{8.5}$$

So the junction capacitance is identical with that of an ordinary capacitor of the same size, shape, and permittivity as the depletion layer.[8] Keep in mind that the depletion layer width is a function of the reverse bias voltage so that the junction capacitance is also a function of the reverse bias voltage applied to the pn junction. This fact can be put to good use in a variety of ways, since it represents a voltage-controlled capacitance.[9] A diode specifically designed for such an application is called a ***varactor diode***

8.4摇Diffusion or Storage Capacitance

In forward bias, the excess charge stored in the neutral, or diffusion region, of the diode leads to a delay whenever an attempt is made to change the voltage across the junction.[10] Because this region is electrically neutral, there is not only an excess of minority carriers (already shown in Fig. 8.4) but also an equal excess of majority carriers, so that their charges balance and the region is neutral. All of these carriers must be re-adjusted in number when, for example, the external circuit causes an alternating voltage to appear across the diode. This results in a flow of charge in and out of the diode which is modeled by the ***diffusion capacitance*** sometimes called the ***storage capacitance.*** This small-signal capacitance can be calculated from the change ΔQ of the excess minority carrier charge Q, stored in the neutral diffusion regions of the diode, which accompanies a small change ΔV in the applied voltage.[11] The ΔQ charge is alternately being charged and discharged through the junction as the voltage across the junction changes.[12]

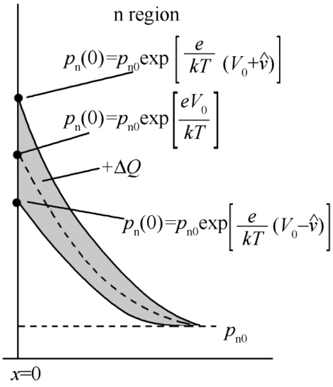


Fig. 8.4摇Minority carrier concentration changes with changing forward-bias voltage.

Define C_d as the dQ/dV, and the following result can be derived

$$C_d = \tau_t(qI_{DQ}/kT) = \tau_t g_d \tag{8.6}$$

where τ_t is called effective ***transit time*** .

The small-signal diffusion capacitance C_d is directly proportional to the current through the junction, becoming larger than the depletion layer capacitance for all reasonable forward biases.[13] C_d is always associated with the diode differential resistance r_d, so the forward-biased diode is inevitably lossy, and can not be used as a way of making a good capacitor.

From the foregoing we see that the relative significance of charge storage in the space-charge region (as represented by C_j) and charge storage in the quasi-neutral regions depends strongly on the junction voltage.[14] Under reverse bias, storage in the quasi-neutral regions is negligible and the storage represented by the ***junction capacitance*** dominates. Under forward bias, although C_j increases (because W_d decreases), the exponential factor in the formula for C_d generally makes diffusion capacitance and its associated charge storage dominant.

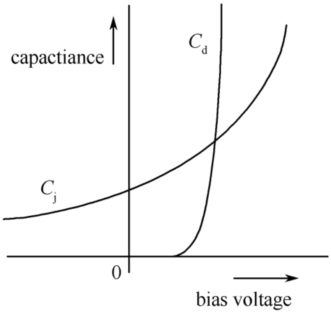


Fig. 8.5摇Voltage dependence of the small-signal junction capacitance C_j and the small signal diffusion capacitance C_d of a junction diode.

Fig. 8.5 illustrates the voltage dependence of the C_d and C_j, showing that with a forward bias of more than a few hundred millivolts the diffusion capacitance normally dominates, and C_j is often negligible.[15] The converse applies for reverse bias voltages.

8.5摇Diode Transients

The pn junction is typically used as an ***electrical switch***. A voltage or current ***pulse*** is applied to change the operating state between forward bias (***'on' state***) and reverse bias (***'off' state***). Since a pn junction has capacitance associated with it, it would be expected that some time would be required to make the transition from off to on (***turn-on time***) and from on to off (***turn-off time***). These ***transients*** are discussed qualitatively.

Of primary interest in circuit applications is the speed of the pn junction diode in switching states. Usually the turn-off time is much larger than the turn-on time, and the total turn-off time is the sum of ***storage time*** t_S and ***fall time*** t_2.

8.6摇Circuit Models for Junction Diodes

Two ways of modeling a diode for circuit analysis are given in this section: a 'small-signal' model suitable for hand calculations, and a more comprehensive, large-signal model as used in computer ***simulations*** of circuit behavior.

When the diode is used in a circuit in which small alternating signal voltages and currents are superimposed on static (d. c.) values, the diode can be replaced for the purposes of calculating signal voltages and currents by the small-signal ***equivalent circuit*** shown in Fig. 8.6.

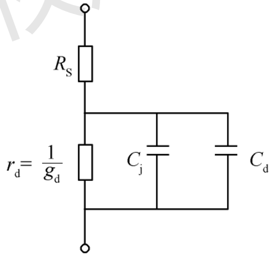


Fig. 8.6摇Small-signal equivalent circuit of a diode.

Let the static voltage across the junction be V, and treat the ***a. c. signal*** as an ***infinitesimal voltage*** dV. The total current is similarly expressed as the sum of I and dI. The equivalent a. c. conductance g_d is the ratio dI/dV, which is obtained by differentiating the I-V relation and given by Eq. (8.2) The inverse of g_d is the resistance of the equivalent resistor r_d in Fig. 8.6, whose value according to Eq. (8.2) depends on the d. c. quiescent diode current. The series resistance R_S is also included in Fig. 8.6. It is often negligibly small compared to r_d.

In reverse bias, leakage effects become important in determining the equivalent resistance, and the equivalent capacitance has a much lower impedance than the resistor. The two capacitors in Fig. 8.6 represent the two capacitive effects. C_d called the diffusion capacitance, or storage capacitance, is a current-dependent capacitance which models the storage of injected carriers in the neutral, or diffusion, region of the diode.[16] C_j is the depletion-layer, or junction capacitance given by either Eq. (8.4) or (8.5). Both capacitances depend on the applied voltage or current in such a way that in forward bias $C_d \gg C_j$, while in reverse bias $C_j \gg C_d$ as seen in Fig. 8.5.

An alternative way of modeling a diode is to use the full I-V relation. This approach is used in the standard computer simulations, of which the best known is probably SPICE (Simulation Program with Integrated Circuit Emphasis), which is widely used for simulating the behavior of ***integrated circuits***.[17]

Reading Materials

Breakdown voltage V_B depends on the doping density of a pn junction, in particular on the doping of the more lightly doped side of the junction. Fig 8.7 shows a plot of breakdown voltage for Si diodes of different doping. It indicates the general value of the quantities, but does not account for the grading of the junction, the doping of the more heavily doped side, nor the distinction between a ***planar*** and a ***spherical junction***. [18]

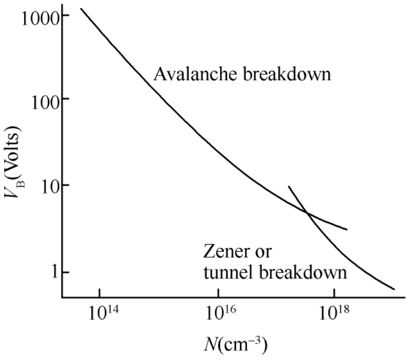


Fig. 8.7摇Breakdown voltage V_B versus doping density N for a Si junction diode.

The mechanism of breakdown for p-n junctions with breakdown voltages less than about $4E_q/q$ is due to the tunneling effect. For junctions with breakdown voltages in excess of $6E_q/q$, the mechanism is caused by avalanche multiplication. At voltages between $(4\sim6)E_q/q$, the breakdown is due to a mixture of both avalanche and tunneling.

Although not large, the temperature variation of the two types of breakdown is of opposite sign. For breakdown voltages in the range of about 5 ~ 6 *V* for a Silicon diode, both avalanche- and tunnel-breakdown can occur simultaneously so that the net temperature variation is very slight. This characteristic is useful for establishing a ***voltage reference*** in some integrated circuits.

Words and Expressions

collide摇*v.* 碰撞
predominant摇*adj.* 主要的
sinusoidal摇*adj.* 正弦的
to superimposed on摇叠加在……上
accompany摇*v.* 伴随,伴有
dominant摇*adj.* 占主要地位的
associated with摇*adj.* 与……有关
reciprocal摇*adj.* 倒数的
to be inversely proportional to…摇与……成反比
to be directly proportional to…摇与……成正比
inverse square root摇负二分之一次方
differentiate摇*v.* 求导
Keep in mind that…摇记住……

Glossary of Important Term

breakdown摇击穿
avalanche breakdown摇雪崩击穿
ionize摇离化
Zener breakdown摇齐纳击穿
tunneling mechanism摇隧穿机理
thermal voltage摇热电压
DC quiescent diode current摇二极管静态直流电流
varactor diode摇变容二极管
junction capacitance摇结电容
permittivity摇介电常数
diffusion capacitance摇扩散电容
storage capacitance摇储存电容
transit time摇渡越时间
a. c. signal摇交流信号
sinusoidal signals摇正弦信号
infinitesimal voltage摇无穷小电压
incremental conductance摇增量电导
diffusion resistance摇扩散电阻
linear amplifier circuits摇线性放大电路
electrical switch摇电开关
pulse摇脉冲
'on' state摇导通状态
'off' state摇断开状态
turn-on time摇导通时间
turn-off time摇断开时间
transient摇瞬态

storage time摇储存时间
fall time摇下降时间
simulation摇模拟,仿真
equivalent circuit摇等效电路
integrated circuit摇集成电路
planar junction摇平面结
spherical junction摇球面结
voltage reference摇电压基准

Notes

1. Breakdown may result in the destruction of a device but this is not inevitable.

提示:result in 是固定搭配,表示“导致”;is not inevitable 谓语是否定形式,表语 inevitable 表示“不可避免的”,也具有否定的含义,因此这是一种否定加否定的结构。与直接表示“可以避免的”的相比,采用这种否定加否定的方式,具有“强调”的作用。

2. Damage is caused by local melting, and if the current is held to a small value by an external resistance, or is spread over a large area by careful design and manufacture, then the temperature reached are not high enough for Silicon or SiO_2 to melt, nor for impurities to diffuse out of their intended positions.

提示:第一个 and 连接两个并列句。其中 and 后面是一个包括条件状语从句的复合句;if 引导的条件状语从句中 or 表示“或者”,连接由 is held 和 is spread 两个被动语气组成的并列谓语;enough应该放在被修饰的词的后面,与动词不定式组成固定搭配,还可以用 for 引导动词不定式的逻辑主语。本句中 enough 修饰 high,与 to melt 组成固定搭配,to melt 的逻辑主语是 for 后面的 Silicon or SiO_2;nor 后面连接的是又一个带逻辑主语的动词不定式,与前一个动词不定式并列。由于谓语为否定形式,因此这里用 nor 表示否定,而不是用 or。

3. In addition, the newly generated electrons and/or holes may acquire sufficient energy to ionize other atoms, leading to the avalanche process.

提示:and/or 直接翻译为“和/或”,表示“两者同时”或者“其中之一”;leading to 是分词短语作状语,其逻辑主语是主句中的主语。

4. We have been considering the dc characteristics of the pn junction diode.

提示:该句采用的是现在完成进行时。

5. When semiconductor devices with pn junctions are used in ***linear amplifier circuits*** , for example, ***sinusoidal signals*** are superimposed on the dc currents and voltages, so that the small-signal characteristics of the pn junction become important.

提示: When…引导时间状语从句,其中插入 for example,表示该状语从句列举的是一种实例情况。英语中经常直接采用一个例子来说明问题,而在该例之前并没有给出一般情况。翻译时可以添加说明。

6. It is seen that the junction capacitance decreases as the bias becomes more negative.

提示:more negative 表示负数绝对值增大,其代数值减小,因此翻译为“更负”才严谨,而不能翻译为“增大”。

7. When $V_{bi} \ll |V_a|$, C_j decreases roughly as the inverse square root of the reverse voltage.

提示:roughly 表示“约为”、“近似为”,不要机械地翻译为“粗略”;square root of…表示“平方根”、“二分之一次方”,前面加 inverse 表示平方根的倒数,也可翻译为“负二分之一次方”。

8. So the junction capacitance is identical with that of an ordinary capacitor of the same size, shape, and ***permittivity*** as the depletion layer.

提示:that 代表前面出现的 capacitance(电容);第一个 of 后面的 an ordinary capacitor(普通的电

容器)作 that 的定语;注意 capacitance 表示“电容”,capacitor 表示“电容器”;第二个 of 后面到句尾作 capacitor 的定语。

9. This fact can be put to good use in a variety of ways, since it represents a voltage-controlled capacitance.

提示:put to good use 是固定搭配,表示“充分利用”;voltage-controlled 是一种带逻辑主语的过去分词结构,作定语,修饰受到控制的对象 capacitance,破折号前面 voltage 表示动作的实施者,可翻译为“受电压控制的电容”。

10. In forward bias, the excess charge stored in the neutral, or diffusion, region of the diode leads to a delay whenever an attempt is made to change the voltage across the junction.

提示:or 表示“即”,不要翻译为“或者”;注意根据上下文灵活翻译 whenever。

11. This small-signal capacitance can be calculated from the change ΔQ of the excess minority carrier charge Q, stored in the neutral diffusion regions of the diode, which accompanies a small change ΔV in the applied voltage.

提示:过去分词 stored 修饰 Q;which 引导的非限制性定语从句修饰 ΔQ。

12. The ΔQ charge is alternately being charged and discharged through the junction as the voltage across the junction changes.

提示:is alternately being charged and discharged 是现在进行时被动语态。

13. The small-signal diffusion capacitance C_d is directly proportional to the current through the junction, becoming larger than the depletion layer capacitance for all reasonable forward biases.

提示:由 becoming 组成的现在分词短语作状语,becoming 的逻辑主语是主句中的主语;is directly proportional to 表示与……成正比;注意根据上下文灵活翻译 reasonable 一词。

14. From the foregoing we see that the relative significance of charge storage in the space-charge region (as represented by C_j) and charge storage in the quasi-neutral regions depends strongly on the junction voltage.

提示: From the foregoing 表示“从前面的分析中”;主语用 we 并不是一定要说明行为者,翻译时不一定必须表示为“我们……”;significance 表示大小,不要机械翻译为“意义”、“重要性”。

15. Fig. 8.5 illustrates the voltage dependence, showing that with a forward bias of more than a few hundred millivolts the diffusion capacitance normally dominates, and C_j is often negligible. The converse applies for reverse bias voltages.

提示:现在分词 showing 及其后面的宾语从句作状语;showing 后面是由 and 连接的两个宾语从句,此处 and 可以翻译为“而”;后面一句中 The converse applies 表示与前面一句表示的情况相反。

16. C_d called the diffusion capacitance, or storage capacitance is a current-dependent capacitance which models the storage of injected carriers in the neutral, or diffusion, region of the diode.

提示:主句主语是 C_d,谓语是 is a current-dependent capacitance,表示“与电流大小有关的电容”;过去分词短语 called…作定语,翻译时不一定按照定语修饰关系翻译;which…到句末是定语从句,修饰 capacitance;句子两处 or 的作用都是表示 or 连接的前后两个词是同一个术语的两种不同称谓,注意翻译时采用的不同汉语名词术语。

17. This approach is used in the standard computer simulations, of which the best known is probably SPICE (Simulation Program with Integrated Circuit Emphasis), which is widely used for simulating the behavior of ***integrated circuits*** .

提示:standard 表示“典型的”、“通用的”,不要机械地翻译为“标准的”;of which 引导的非限制

性定语从句修饰 simulations；句末 which…引导的非限制性定语从句修饰 SPICE；SPICE 是一个模拟软件的名称，翻译时不一定要用一个中文名称，只需像英文那样在 SPICE 后面给出其对应的全称。

18. It indicates the general value of the quantities, but does not account for the grading of the junction, the doping of the more heavily doped side, nor the distinction between a planar and a spherical junction.

提示：注意 general、account for、grading 和 nor 这 4 个词的翻译。

Exercises

1. Translate the reading material into Chinese.
2. How to avoid the destruction of a device resulted from the breakdown.
3. Compare the differences between the avalanche-breakdown and tunnel-breakdown.
4. Why docs the breakdown voltage of a pn junction decrease as the doping concentration increases?
5. Why does a junction capacitance exist in a biased pn junction?
6. Explain the physical mechanism of diffusion capacitance.
7. Compare the differences between the Voltage dependences of the small-signal junction capacitance C_j and diffusion capacitance C_d of a junction diode.
8. Explain what is diffusion resistance and why it is dependent on the dc quiescent diode current?
9. Why would some time be required to make the transition of a diode from off to on (turn-on time) and from on to off (turn-off time)?
10. When and why must the small-signal equivalent circuit be used?

Session 9摇Metal-Semiconductor Contacts

Most of the electronic devices that make up an integrated circuit are connected by means of metal-semiconductor contacts. Moreover, all integrated circuits communicate with the rest of an electrical system via metal-semiconductor contacts. As we will see, the properties of these contacts can vary considerably, and it is necessary to consider several factors in order to understand them.

It is well known that the quality of metal semiconductor contacts plays an important role in the performance of various semiconductor devices and integrated circuits. For example, good ohmic contacts are essential for achieving excellent performance of a semiconductor device, while Schottky (i. e. , rectifying) contacts can be used for a wide variety of device applications. In addition to different device and circuit applications, Schottky contacts can also be used as test vehicles for investigating the physical and electrical properties of a semiconductor material and its surfaces.[1] For example, a Schottky diode can be used to study bulk defects and interface properties of a metal semiconductor system. Therefore, it is essential to obtain a better understanding of the fundamental physical and electrical properties of the metal semiconductor systems so that technologies for preparing good ohmic and Schottky contacts can be developed for a wide variety of device applications.[2]

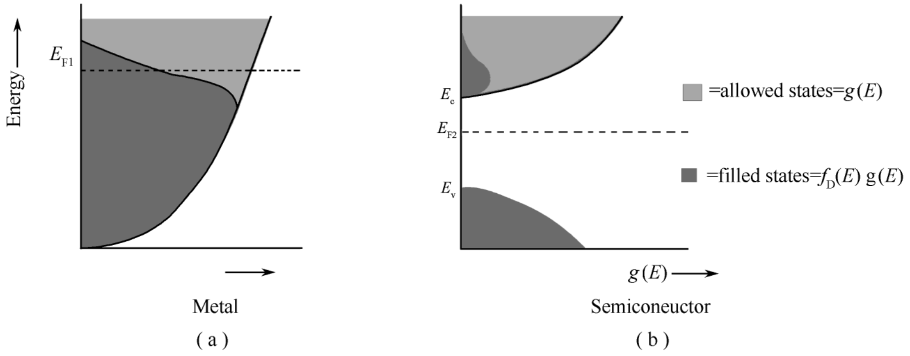
Two types of metal semiconductor contacts are commonly used in the fabrication of semiconductor devices and integrated circuits. They are the ***Schottky*** and ***ohmic contacts***. A Schottky barrier contact exhibits an asymmetrical current voltage (*I-V*) characteristic when the polarity of a bias voltage applied to the metal semiconductor contacts is changed. The ohmic contact, on the other hand, shows a linear *I-V* characteristic regardless of the polarity of the external bias voltage. A good ohmic contact is referred to the case in which the voltage drop across a metal semiconductor contact is negligible compared to that of the bulk semiconductor material.[3]摇

9.1摇Schottky Contacts

One of the first practical semiconductor devices used in the early 1900s was the metal-semiconductor diode. This diode, also called a point contact diode, was made by touching a metallic whisker to an exposed semiconductor surface. Now, the Schottky barrier diode is actually a variation of the point-contact diode in which the metal semiconductor junction is a surface rather than a point contact.[4] In fact, a large contact area between the metal and the semiconductor in a Schottky barrier diode provides some advantages over the point-contact diode. Lower forward resistance and lower noise generation are the most important advantages of the Schottky barrier diode. The applications of a Schottky barrier diode are similar to those of the point-contact diode. The low noise level generated by Schottky diodes makes them especially suitable for uses in microwave receivers, detectors, and mixers. The Schottky barrier diode is sometimes called the hot electron or hot carrier diode because the electrons flowing from the semiconductor to the metal have a higher energy level than electrons in the metal. The effect is the same as it would be if the metals were heated to a higher temperature than normal.[5]

9.1.1摇Schottky Contacts in Equilibrium

The most distinctive characteristic of the electronic energy states of the metal and the semiconductor is the relative positions of the Fermi levels within the densities of allowed states $g(E)$. In the metal, the Fermi level is immersed within a continuum of allowed states, while in a semiconductor, under usual circumstances, the density of states is negligible at the Fermi level. Plots of $g(E)$ versus energy for idealized metals and semiconductors are shown in Figs. 9.1(a) and 9.1(b).

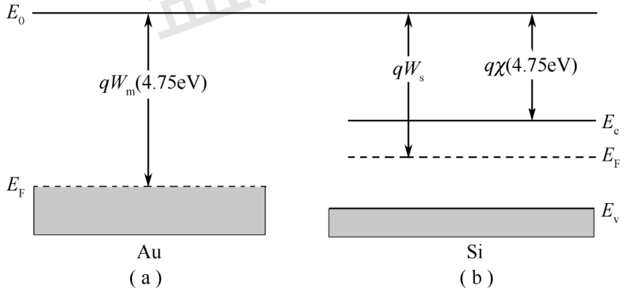


Metal
(a)

Semiconeuctor
(b)

Fig. 9. 1摇(a) Allowed electronic-energy states $g(E)$ for an ideal metal. The states indicated by cross-hatching are occupied. Note the Fermi level E_{F1} immersed in the continuum of allowed states. (b) Allowed electronic-energy states $g(E)$ for a semiconductor. The Fermi level E_{F2} is at an intermediate energy between that of the conduction-band edge and that of the valence-band edge.

Fig. 9.2 shows the energy levels for the metal gold and the semiconductor silicon. The difference between vacuum level E_0 and metal Fermi level E_F is called the ***work function***, usually given the symbol qW in energy units and often listed as W in volts for particular materials. The work function for various metals is shown in Fig. 9. 3. The minima of the work function exist for group-I elements. Since the electron density in the metal conduction band is very high, the position of the metal Fermi level does not change considerably when charge is exchanged between the metal and the semiconductor.



Au
(a)

Si
(b)

Fig. 9.2摇Pertinent energy levels for the metal gold and the semiconductor silicon. Only the work function is given for the metal, whereas the semiconductor is described by the work function qW_s, the electron affinity $q\chi_s$, and theband gap $(E_c - E_v)$.

In the case of the semiconductor, the difference between E_0 and E_F is a function of the dopant concentration of the semiconductor, because E_F changes position within the gap separating E_v and E_c as the doping is varied. The difference between the vacuum level and the conduction-band edge is, however, a constant of the material. This quantity is called the ***electron affinity***, and is conventionally denoted by $q\chi$ in energy units. Tables of χ in volts exist for many materials. (The symbol χ is a Greek cap-

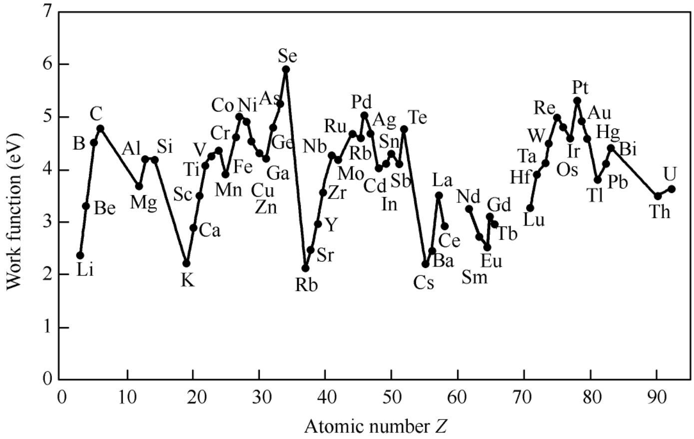


Fig. 9.3摇Work function of various metals.

ital letter *chi*.)

The choice of E_0 as a common energy reference makes clear that if W_m is less than W_s and the materials do not interact, an electron in the metal has, on average, a total energy that is higher than the average total energy of an electron in the semiconductor. On the contrary, if W_m is greater than W_s, the average total energy of an electron in the semiconductor is higher than it is in the metal. For the sake of discussion, we consider the latter case where $W_m > W_s$. When an intimate contact is established, the disparity in the average energies can be expected to cause the transfer of electrons from the semiconductor into the metal.

To construct a proper band diagram for the metal and the semiconductor in thermal equilibrium, we need to note two additional facts. The first is that the vacuum level E_0 must be drawn as a continuous curve. This is because E_0 represents the energy of a "just-free" electron and thus must be a continuous, single-valued function in space. Second, we note that electron affinity is a property associated with the crystal lattice like the forbidden-gap energy. Hence, it is a constant in a given material. Considering these three factors: constancy of E_F, continuity of E_0, and constancy of χ in the semiconductor, we can sketch the general shape of the band diagram for the metal-semiconductor system. The sketch is given in Fig. 9.4(a) for an n-type semiconductor for which $W_m > W_s$.

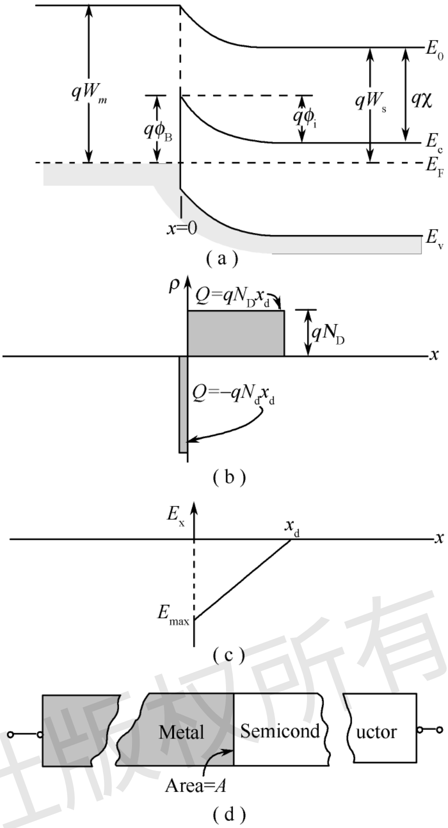
Fig. 9.4(a) indicates that electrons at the band edges (E_c and E_v) in the vicinity of the junction in the semiconductor are at higher energies than are those in more remote regions.[6] This is a consequence of the transfer of negative charge from the semiconductor into the metal. Because of the charge exchange, there is a field at the junction and a net increase in the potential energies of electrons within the band structure of the semiconductor. The free-electron population is thus depleted near the junction, as indicated by the increased separation between E_c and E_F at the surface compared to that in the bulk.

Before considering the electrical properties of the metal-semiconductor junction we note that our development thus far has relied on the important idealization that the basic band structures of the two materials are unchanged near their surfaces.[7]

The charge and field diagrams for an ideal metal-semiconductor junction are sketched in Figs. 9.4(b) and 9.4(c). To the extent that the metal is a perfect conductor, the charge transferred to it from the semiconductor exists in a plane at the metal surface. In the idealized n-type semiconductor, positive charge can consist either of ionized donors or of free holes while electrons make up the negative charge. We have made several assumptions about the semiconductor charge in drawing Figs. 9.4(b) and 9.4(c). First, the free-hole population is assumed to be everywhere so small that it need not be considered; second, the electron density is much less than the donor density from the interface to a plane at $x = x_d$. Beyond x_d, the donor density N_D is taken to be equal to n. These assumptions make up what is usually called the ***depletion approximation.*** Although they are not precisely true, they are generally sufficiently valid to permit the development of very useful relationships.

9.1.2摇Schottky Contacts Under Applied Bias

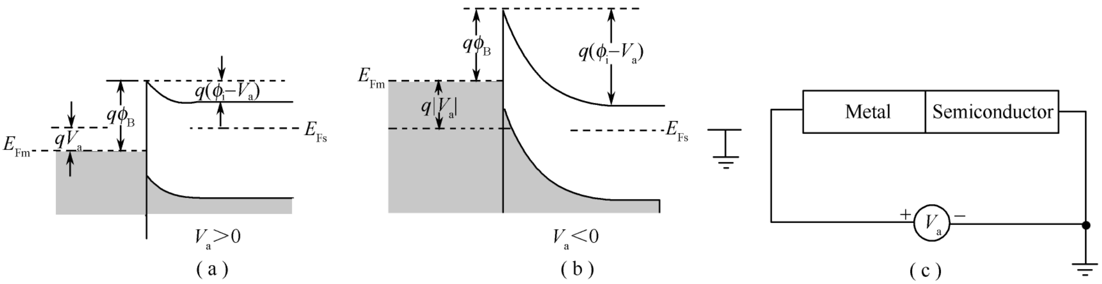
Up to this point, we have been considering thermal equilibrium conditions at the metal-semiconductor junction. Now we add an applied voltage and consider the resulting nonequilibrium condition. We saw in Fig. 9.4(a) that there is an abrupt step in allowed electron energies at the metal-semiconductor interface. This step makes it more difficult to cause a net transfer of free electrons from the metal into the semiconductor than it is to obtain a net flow of electrons in the opposite direction. There is a barrier of $q\phi_B$ electron volts between electrons at the Fermi level in the metal and the conduction band states in the semiconductor near its surface [Fig. 9.4(a)]. To first order this barrier height is independent of bias. Referring to Fig. 9.4(c), we see that the voltage drop across the near delta function of space charge in the metal (equivalent to the area between the $\boldsymbol{E}$-field curve and the axis) is effectively zero in equilibrium; that is, no voltage drop can be sustained across the metal.



摇Fig. 9.4摇(a) Idealized equilibrium band diagram 摇(energy versus distance) for a metalsemiconductor 摇rectifying contact (Schottky barrier). The physical 摇 junction is at $x = 0$. (b) Charge at an idealized 摇 metal-semiconductor junction. The negative charge 摇 is approximately a delta function at the metal 摇surface. The positive charge consists entirely of 摇ionized donors (here assumed constant in space) in 摇the depletion approximation. (c) Field at an 摇idealized metal-semiconductor junction. (d) the 摇 cross section of the Schottky contact.

The total voltage drop across the space-charge region ϕ_i occurs within the semiconductor, as can be seen in Fig. 9.4(a). An applied voltage is similarly dropped entirely within the semiconductor and alters the equilibrium-band diagram [Fig. 9.4(a)] by changing the total curvature of the bands, modifying the potential drop from ϕ_i.[8] Thus, electrons in the bulk of the semiconductor at the conduction-band edge are impeded from transferring to the metal by a barrier that can be changed readily from its equilibrium value $q\phi_i$ by an applied bias. The barrier is reduced when the metal is biased positively with respect to the semiconductor, and it is increased when the metal is more negative.

Energy-band diagrams for two cases of bias are shown in Fig. 9.5(a) and 9.5(b). Because these diagrams correspond to nonequilibrium conditions, they are not drawn with a single Fermi level. The Fermi energy in the region from which electrons flow is higher than is the Fermi energy in the region into which electrons flow.[9] Currents, of course, move in the direction opposite to the electron flow. To investigate bias effects on the barrier, we consider the semiconductor to be grounded and take forward bias to correspond to the metal electrode being made positive. The applied voltage is called $V_a >$ and the bias polarity is indicated in Fig. 9.5(c). Under reverse bias the metal is negatively biased ($V_a < 0$).



(a) (b) (c)

Fig. 9.5摇Idealized band diagrams (energy versus distance) at a metal-semiconductor junction (a) under applied forward bias ($V_a > 0$) and (b) under applied reverse bias ($V_a < 0$). The semiconductor is taken as the reference (voltage ground) as shown in (c). The vacuum levels for the two cases are not shown.

The basic dependence of current on voltage in a Schottky-barrier diode can be deduced from qualitative arguments. These arguments provide fundamental insight into the nature of the equilibrium behavior of the metal-semiconductor system.

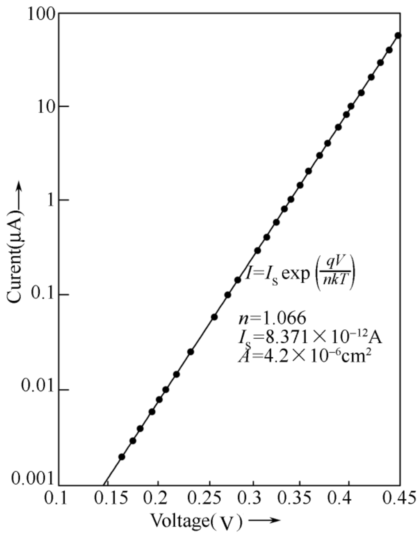
The band diagram at thermal equilibrium shown in Fig. 9.4 is the starting point for the derivation. At equilibrium, the rate at which electrons cross over the barrier into the semiconductor from the metal is balanced by the rate at which electrons cross the barrier into the metal from the semiconductor. From the discussion of diffusion in previous lessons, we know that free carriers in crystals are constantly in motion because of their thermal energies. For example, this fact was used to show that a density n_0 of free carriers in thermal motion can be considered to cause a current density equal to $qn_0 V_{th}/4$ in an arbitrary direction. At thermal equilibrium, of course, this current density is balanced by an equal and opposite flow, and there is zero net current. Applying this concept to the boundary plane of the band diagram in Fig. 9.4, we see that there is a tendency of electrons to flow from the semiconductor into the metal and an opposing balanced flux of electrons from the metal into the semiconductor. These currents are proportional to the density of electrons at the boundary. When a bias V_a is applied to the junction as in Fig. 9.5, the potential drop within the semiconductor is changed, and we can expect the flux of electrons from the semiconductor toward the metal to be modified. We can, therefore, approximate the current-voltage characteristic by

$$J_x = J'_S \left[\exp\left(\frac{qV_a}{nkT} \right) - 1 \right] \tag{9.1}$$

where J'_S is independent of voltage and n is taken to be a parameter having a value that is usually found experimentally to be between 1.02 and 1.15. Experimental measurements for a forward-biased, aluminum-silicon Schottky barrier are shown in Fig. 9.6. The good fit between the measured data in Fig. 9.6 and Equation (9.1) with $n = 1.07$ is typical.

9.2摇Ohmic Contacts

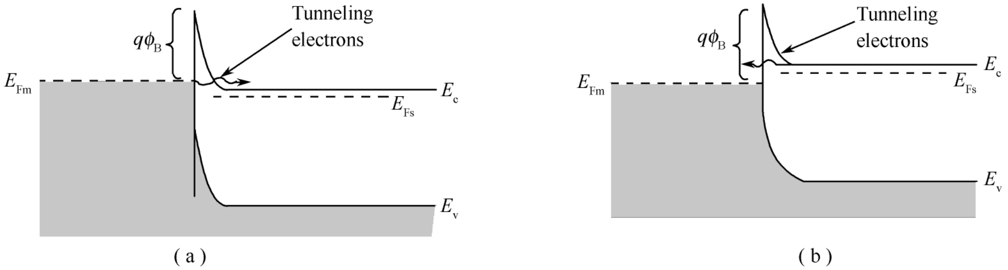
In our discussion of metal-semiconductor contacts, we have thus far considered cases in which the semiconductor near the metal has a lower majority-carrier density than the bulk and in which there is a barrier to electron transfers from the metal.[10] In such cases any applied voltage is dropped mainly across the junction region, and currents are contact limited. The inverse case, in which the contact itself offers negligible resistance to current flow when compared to the bulk, defines an ohmic contact.[11] Although this definition of an ohmic contact may sound awkward, it emphasizes one essential aspect: when voltage is applied across a device with ohmic contacts, the voltage dropped across the ohmic contacts is negligible compared to voltage drops elsewhere in the device. Thus, no power is dissipated in the contacts, and the ohmic contact can be described as being at thermal equilibrium even when currents are flowing. An important and useful consequence of this property is that all free-carrier densities at an ohmic contact are unchanged by current flow; the densities remain at their thermal-equilibrium values.



摇Fig. 9.6摇Measured values of current (plotted on a logarithmic scale) versus voltage for an aluminum-silicon Schottky barrier. Values for I_S and n are obtained from an empirical fit of the data to 摇摇摇摇摇摇Equation (9.1).

The metal-semiconductor contacts that we have considered in the previous section can, for example, be made ohmic if the effect of the barrier on carrier flow can be made negligible. In practice this is accomplished by heavily doping the semiconductor so that the barrier width x_d is very small. The space-charge region, therefore, narrows as N_D increases. When the barrier width approaches a few nanometers, a new transport phenomenon, *tunneling* through the barrier, can take place.

Fig. 9.7(a) is a schematic illustration of the tunneling process through a very thin Schottky barrier. When the barrier is of the order of nanometers and the metal is biased negatively with respect to the semiconductor, electrons in the metal need not be energetic enough to surmount the barrier to enter the semiconductor.[12] Instead, they can tunnel through the barrier into the conduction-band states in the semiconductor. Likewise, when the semiconductor is biased negatively with respect to the metal, electrons from the semiconductor can tunnel into electronic states in the metal [Fig. 9.7(b)]. Many electrons are



(a)

(b)

Fig. 9.7摇Metal-semiconductor barrier with a thin space-charge region through which electrons can tunnel. (a) Tunneling from metal to semiconductor. (b) Tunneling from semiconductor to metal.

available to take part in these processes and currents rise very rapidly as bias is applied. Hence, a metal-semiconductor contact at which tunneling is possible has a very small resistance. It is virtually always an ohmic contact. Ohmic contacts are frequently made in this way in practice. To assure a very thin barrier, the semiconductor is often doped until it is degenerate (i. e., until the Fermi level enters either the valence or the conduction band). In modem devices the conductivity of the semiconducting regions is higher than in previous device generations. Consequently, ohmic contacts must have even lower resistance so that no appreciable voltage drop occurs across them.

Another method of obtaining an ohmic contact is to cause the majority carriers to be more numerous near the contact than they are in the bulk of the semiconductor. An ohmic contact of this type results if the semiconductor surface is not depleted when it comes into equilibrium with the metal, but rather has an enhanced majority-carrier concentration. Using the ideal Schottky theory, we see that this condition occurs in a metal-semiconductor junction between a metal and an n-type semiconductor with a larger work function than that of the metal.[13] In this case, electrons are transferred to the surface of the semiconductor and the metal is left with a skin of positive charge. For an ohmic contact to a p-type semiconductor, the relative sizes of the work functions in the two regions need to be reversed to achieve a net positive charge in the semiconductor and, thereby, an enhanced hole density near the contact.

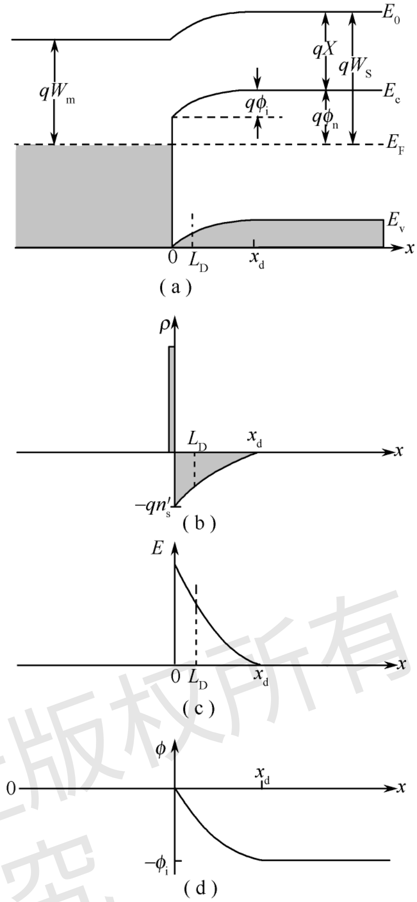


Fig. 9.8摇(a) Idealized equilibrium energy diagram for a Schottky ohmic contact between a metal and an n-type semiconductor. (b) Charge at an ideal Schottky ohmic contact. A delta function of positive charge at the metal surface couples to a distributed excess-electron density $n'(x)$ in the semiconductor. (c) Field, and (d) potential at an idealized Schottky ohmic contact. The Debye length L_D is a characteristic measure of the extent of the charge and field.

The relevant energy diagram is sketched in Fig. 9.8(a), and the corresponding charge and field diagrams are given in Fig. 9.8(b) and 9.8(c). There is a qualitative similarity between these figures and Figs. 9.4(a) to 9.4(c) that referred to a rectifying contact; the important distinction between the two situations is that the semiconductor charge consists of free electrons in the case of the ohmic contact, but it is fixed (on positive donor sites) in the barrier case. The charge distribution, field, and potential for such a contact can be calculated using techniques similar to those employed for the rectifying contact.

Reading Materials

Barrier Height and Capacitance

The barrier height can be determined by the capacitance measurement. When a small ac voltage is superimposed upon a dc bias, incremental charges of one sign are induced on the metal surface and charges of the opposite sign in the semiconductor. Fig. 9.9 shows some typical results where $1/C^2$ is plotted against the applied voltage.

To obtain the barrier height of semiconductor which contain both shallow-level and deep-level impurities, we need to measure the C-V curves at two different temperatures at multiple frequencies.

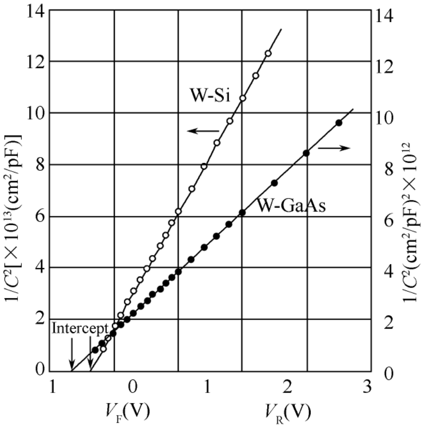


Fig. 9.9摇$1/C^2$ versus applied voltage for W-Si and W-GaAs diodes.

Drawing the Energy Bands

Start with the metal, draw the Fermi level for the semiconductor in the right place allowing for any external voltage, then draw the energy bands for the bulk semiconductor to fit round the Fermi level. Next locate the bottom of the conduction band at the surface by measuring up by W_m from the metal Fermi level, measuring down by W_s and correcting for the gap between conduction band and Fermi level in the semiconductor. This last stage can be more tidily expressed by saying that, at the surface, the bottom of the conduction band is $W_m - \chi_s$ above the metal Fermi level, where χ_s is the electron affinity of the semiconductor. Finally complete the bent bands in the semiconductor. Their thickness is once again governed by charges in the depletion layer through Poisson's equation. Notice that since the potential in the semiconductor is now no longer uniform, the carrier density will also vary. The local difference between the band edges and the Fermi level must be used to find hole and electron densities.

Words and Expressions

intimate摇*adj*. 亲密的
disparity摇*n*. 不一致, 不同, 不等
infinitesimal摇*adj*. 无穷小的, 极小的, 摇无限小的
vicinity摇*n*. 邻近, 附近, 接近
remote摇*adj*. 遥远的, 偏僻的, 细微的
curvature摇*n*. 弯曲, 曲率
impede摇*v*. 阻止

Glossary of Important Term

Schottky contacts摇肖特基接触
ohmic contacts摇欧姆接触
work function摇功函数
electron affinity摇亲和势
depletion approximation摇耗尽近似
tunneling摇隧穿

Notes

1. In addition to different device and circuit applications, Schottky contacts can also be used as test vehicles for investigating the physical and electrical properties of a semiconductor material and its surfaces.

提示:in addition to 的意思是"除……之外";vehicle 原指"交通工具、媒介",这里表示手段、方式。

2. Therefore, it is essential to obtain a better understanding of the fundamental physical and

electrical properties of the metal-semiconductor systems so that technologies for preparing good ohmic and Schottky contacts can be developed for a wide variety of device applications.

提示：it is essential to obtain…中，it 是先行主语，to 引导的不定式是真正的主语，由于真正主语太长，所以采用这种形式；so that 引导状语从句，可以翻译为“为了……有必要……”。

3. A good ohmic contact is referred to the case in which the voltage drop across a metal-semiconductor contact is negligible compared to that of the bulk semiconductor material.

提示：that 指代的是 voltage drop，in which 引导定语从句，修饰 case。

4. Now, the Schottky barrier diode is actually a variation of the point-contact diode in which the metal-semiconductor junction is a surface rather than a point contact.

提示：in which 引导的定语从句修饰的究竟是什么，这一点容易混淆，根据物理含义，应该是修饰前面的 Schottky barrier diode 而不是 point-contact diode。

5. The effect is the same as it would be if the metals were heated to a higher temperature than normal.

提示：as it would be if the metals were heated to a higher temperature than normal 是虚拟语气；结合上下文，the effect 可以翻译为“这时电子的能量”。

6. Fig. 9.4(a) indicates that electrons at the band edges (E_c and E_v) in the vicinity of the junction in the semiconductor are at higher energies than are those in more remote regions.

提示：该句中的比较从句采用了倒装形式，并用 those 代替了 electrons；at the band edges，in the vicinity 和 in the semiconductor 这三个介词短语依次指出了 electrons、band edges 和 junction 所在的位置。

7. Before considering the electrical properties of the metal-semiconductor junction we note that our development thus far has relied on the important idealization that the basic band structures of the two materials are unchanged near their surfaces.

提示：thus far 指“迄今，到目前为止”。development 原意为“进展”，这里可翻译为“分析”。

8. An applied voltage is similarly dropped entirely within the semiconductor and alters the equilibrium-band diagram [Fig. 9.4(a)] by changing the total curvature of the bands, modifying the potential drop from ϕ_i.

提示：modifying the potential drop from ϕ_i 作为伴随状态，表示能带图发生变化带来的结果。

9. The Fermi energy in the region from which electrons flow is higher than is the Fermi energy in the region into which electrons flow.

提示：本句直译为“流出电子区域的费米能级比流入电子区域的费米能级高”，非常不通顺，所以采用意译。

10. In our discussion of metal-semiconductor contacts, we have thus far considered cases in which the semiconductor near the metal has a lower majority-carrier density than the bulk and in which there is a barrier to electron transfers from the metal.

提示：本句的主句采用现在完成时，表示“一直”的意思。thus far 表示“迄今”；in which引导的从句修饰 bulk。

11. The inverse case, in which the contact itself offers negligible resistance to current flow when compared to the bulk, defines an ohmic contact.

提示：本句的主句是 The inverse case defines an ohmic contact，in which 引导的从句修饰 case。

12. When the barrier is of the order of nanometers and the metal is biased negatively with respect to

the semiconductor, electrons in the metal need not be energetic enough to surmount the barrier to enter the semiconductor.

提示:with respect to…表示“相对于……而言”;of the order…表示“……量级”;to surmount the barrier 和 to enter the semiconductor 这两个不定式表示目的。

13. Using the ideal Schottky theory, we see that this condition occurs in a metal-semiconductor junction between a metal and an n-type semiconductor with a larger work function than that of the metal.

提示:with a larger work function than that of the metal 作为定语,修饰 an n-type semiconductor。we see 不是句子要表达的重点,可以不翻译。

Exercises

1. Translate the first paragraph in the reading material into Chinese.

2. Answer the following questions in English.

(1) Describe the charge flow in a forward-biased Schottky barrier diode.

(2) Describe what is meant by an ohmic contact.

Session 10摇Heterojunctions

In the discussion of pn junctions in the previous lessons, we assumed that the semiconductor material was homogeneous throughout the entire structure. This type of junction is called a ***homojunction*** . When two different semiconductor materials are used to form a junction, the junction is called a semiconductor ***heterojunction*** . 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.

A growing number of modern devices are based on semiconductor heterojunctions. Modern bipolar transistors employ a p-n heterojunction in order to improve the emitter injection efficiency, while in HFET technology a heterojunction is used to form a high mobility channel.

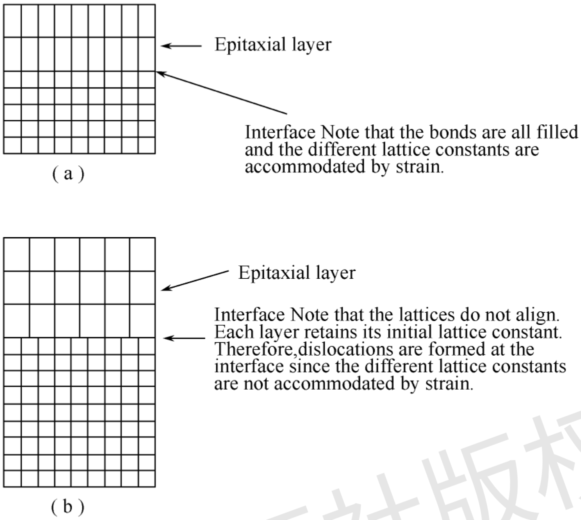
The heterojunction diodes offer a wide variety of important applications for laser diodes, light-emitting diodes (LEDs), photodetectors, solar cells, junction field-effect transistors (JFETs), modulation-doped field-effect transistors (MODFETs or HEMTs), heterojunction bipolar transistors (HBTs), quantum cascade lasers, quantum well infrared photodetectors (QWIPs), quantum dot lasers, and quantum dot infrared photodetectors. With recent advances in MOCVD and MBE epitaxial growth techniques for Ⅲ-Ⅴ compound semiconductors and SiGe/Si systems, it is now possible to grow extremely high-quality Ⅲ-Ⅴ heterojunction structures with layer thickness of 100Å or less for quantum dots, superlattices, and multiquantum-well (MQW) device applications.[1]

10.1摇Strain and Stress at Heterointerfaces

The simplest description of a bulk crystalline semiconductor is that it exhibits perfect or nearly perfect translational symmetry. In other words, suitable translations of the basic unit cell of a crystal restore the crystal back into itself.[2] Implicit in this definition is the assumption that the atoms within the crystal are regularly spaced throughout the entire bulk sample. This assumption is generally true for bulk materials. However, two important exceptions can arise. The first is that a bulk crystal can include impurities and dislocations such that the perfect periodicity of the material is disrupted locally.[3] The crystal can still retain its overall highly ordered structure, yet contain local regions in which perfect periodicity is disrupted by impurities or dislocations. These impurities and dislocations can significantly affect the properties of the material. The second situation arises in multilayered structures. Using exacting crystal growth procedures, heterostructures can be grown with atomic layer precision. A very thin layer of material can be grown on top of or sandwiched between layers grown with a different type of semiconductor material, even materials in which the lattice constant is different.[4]

When a thin layer of material is grown either on or between layers of a different semiconductor that has a significantly different lattice constant, the thin, epitaxial layer will adopt the lattice constant of the neighboring layers provided that the lattice mismatch is less than about 10%.[5] As can be seen from Fig. 10.1(a), when the thin, epitaxial layer adopts the lattice constant of the surrounding layers, it becomes strained, i.e., it is either compressed or expanded from its usual bulk crystal shape. There exists a maximum thickness of the thin layer below which the lattice mismatch can be accommodated through strain.

For layer thickness above the critical thickness, the lattice mismatch cannot be accommodated through strain, dislocations are produced and the strain relaxes as is seen in Fig. 10.1(b). The strain within the layer is homogeneous. The strained layer can be in either compressive or tensile strain. If the lattice constant of the strained layer is less than that of the surrounding layers the system is in tension. Conversely, if the lattice constant of the strained layer is greater than that of the surrounding layers, the strained layer is in compression.



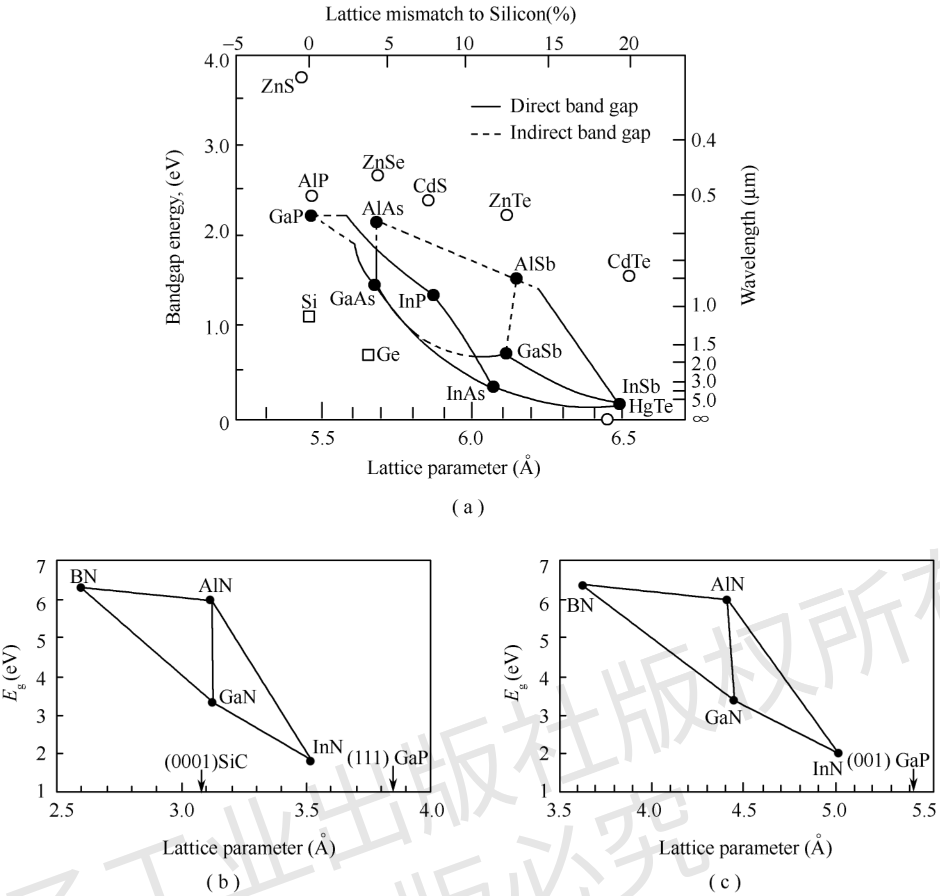
摇摇摇摇摇摇Fig. 10.1摇(a) Thin epitaxial layer strained to accommodate the various latticeconstants of the underlying semiconductor layer and (b) a thicker epitaxial layer thathas relaxed. In part (b) the epitaxial layer is thicker than the critical thickness anddislocations appear at the interface.

10.2摇Heterojunction Materials

Since the two materials used to form a heterojunction will have different energy bandgaps, the energy band will have a discontinuity at the junction interface. We have an abrupt junction in which the semiconductor changes abruptly from a narrow-bandgap material to a wide-bandgap material. On the other hand, if we have a GaAs-$Al_xGa_{1-x}As$ system, for example, the value of x may continuously vary on distance of several nanometers to form a graded heterojunction. Changing the value of x in the $Al_xGa_{1-x}As$ system allows us to engineer, or design, the bandgap energy.

Heterojunctions are generally formed from materials that can be grown upon each other epitaxially with low defect densities. Useful heterojunction systems are therefore comprised of materials that are relatively closely lattice matched.[6] Fig. 10.2 shows bandgaps versus lattice constants of common (a) cubic Ⅲ-Ⅴ and Si-based materials, and (b and c) GaN and related materials. The AlGaAs-GaAs system possesses only a small lattice mismatch over the entire composition range from GaAs to AlAs and thus was one of the first heterojunction systems to be developed and exploited in device structures.

For a given strain, determined by the lattice mismatch, a maximum thickness exists, called the critical thickness h_c, for the strained layer to be completely coherent with the substrate.[7] The critical thickness has a minimum under conditions of thermodynamic equilibrium, but can be enhanced by the epitaxial process. For thicknesses greater than the critical thickness, dislocations are created to reduce the ener-



(a)

(b)

(c)

Fig. 10.2摇Lattice constant versus bandgap for common semiconductor alloys: (a) cubic Ⅲ-Ⅴ and Si-based alloys; (b) GaN and related materials (wurtzite); (c) GaN and related materials (cubic).

gy of the system. These dislocations can significantly degrade device performance and reliability.

The velocity-field characteristics for relevant heterojunction materials are shown in Fig. 10.3. These characteristics determine the usefulness of a material in a given application, for example as the channel for a HEMT or as a collector in an HBT. Generally speaking, the higher the frequency, or speed, required for an application, the higher the desired velocity.[8] Thus the InGaAs alloys have shown excellent performance for millimeter-wave applications. A trade-off, however, in materials properties often exists. For example, the mobility of carriers in semiconductors tends to be greater for small-bandgap materials, as shown in Fig. 10.4. However, the critical electric field for breakdown tends to be greater for wider-bandgap materials. Thus, numerous material parameters must be considered for an application. A high velocity of electrons is desirable for the collector of an HBT, but the collector must also be able to support a high electric field resulting from a large output voltage. Thus InP is often used for an HBT collector for microwave or millimeter-wave power amplifierapplications.

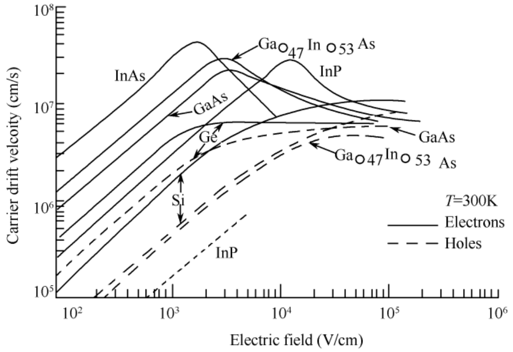


Fig. 10.3摇Velocity field characteristics of common semiconductors.

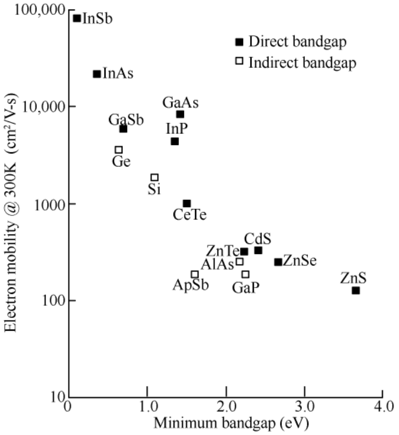


Fig. 10.4摇300K electron mobility versus bandgap for common semiconductors.

10.3摇Energy-Band Diagrams

In the formation of a heterojunction with a narrow-bandgap material and a wide bandgap material, the alignment of the bandgap energies is important in determining the characteristics of the junction. Fig. 10.5 shows three possible situations. In Fig. 10.5(a) we see the case when the forbidden bandgap of the wide-gap material completely overlaps the bandgap of the narrow-gap material. This case, called ***straddling***, applies to most heterojunctions. We will consider only this case here. The other possibilities are called ***staggered*** and ***broken gap*** and are shown in Figs. 10.5(b) and 10.5(c).

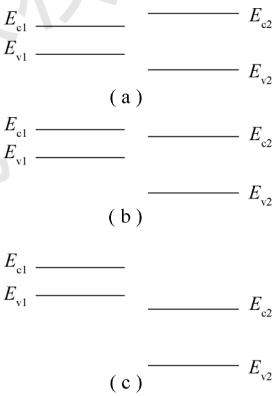


Fig. 10.5摇Relation between narrow-bandgap and wide-bandgap energies: (a) straddling, (b) staggered, and (c) broken gap.

Fig. 10.6 shows the energy-band diagrams of isolated n-type and P-type materials, with the vacuum level used as a reference. The electron affinity of the wide-bandgap material is less than that of the narrow-bandgap material. The difference between the two conduction band energies is denoted by ΔE_c, and the difference between the two valence band energies is denoted by ΔE_v. From Fig. 10.6 we can see that

$$\Delta E_c = e(\chi_n - \chi_P) \qquad (10.1)$$

$$\Delta E_c + \Delta E_v = E_{gP} - E_{gn} = \Delta E_g \qquad (10.2)$$

In the ideal abrupt heterojunction using nondegenerately doped semiconductors, the vacuum level is parallel to both conduction bands and valence bands. If the vacuum level is continuous, then the same ΔE_c and ΔE_v discontinuities will exist at the heterojunction interface. This ideal situation is known as the ***electron affinity rule.*** There is still some uncertainty about the applicability of this rule, but it provides a good starting point for the discussion of heterojunctions.

Fig. 10.7 shows a general ideal nP heterojunction in thermal equilibrium. In order for the Fermi levels in the two materials to become aligned, electrons from the narrow-gap n region and holes from the wide-gap P region must flow across the junction. As in the case of a homojunction, this flow of charge creates a space charge region in the vicinity of the metallurgical junction. The space charge width into the-type region is denoted by x_n, and the space charge width into the P-type region is denoted by x_P. The discontinuities in the conduction and valence bands and the change in the vacuum level are shown in the figure.

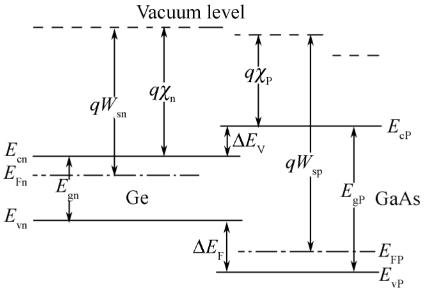
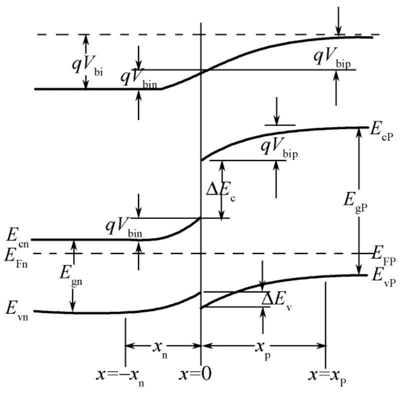


Fig. 10.6摇 Energy-band diagrams of a narrow-bandgep and a wide-bandgap material before contact.



摇摇

Fig. 10.7摇Ideal energy-band diagram of an nP heterojunction in thermal equil ibrium.

Reading Materials

Two-Dimensional Electron Gas

Fig. 10.8 shows the energy-band diagram of an nN GaAs-AlGaAs heterojunction in thermal equilibrium. The AlGaAs can be moderately to heavily doped n type, while the GaAs can be more lightly doped or even intrinsic. As mentioned previously, to achieve thermal equilibrium, electrons from the wide-bandgap AlGaAs flow into the GaAs, forming an accumulation layer of electrons in the potential well adjacent to the interface.[9] One basic quantum mechanical result that we have found previously is that the energy of an electron contained in a potential well is quantized. The phrase two-dimensional electron gas refers to the condition in which the electrons have quantized energy levels in one spatial direction (perpendicular to the interface), but are free to move in the other two spatial directions.

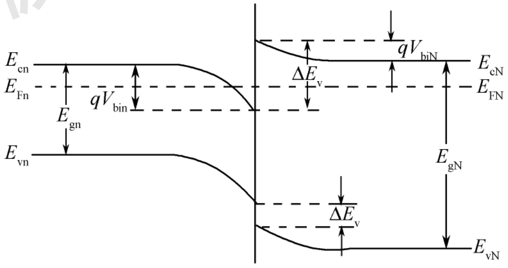
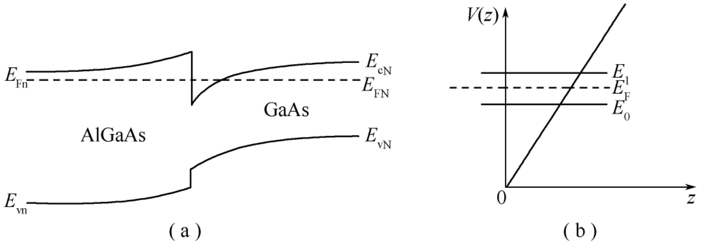


Fig. 10.8摇Ideal energy-hand diagram of an nN heterojunctiun in thermal equilibrium.

The potential function near the interface can be approximated by a triangular potential well. Fig. 10.9(a) shows the conduction band edges near the abrupt junction interface and Fig. 10.9(b) shows the approximation of the triangular potential well. The quantized energy levels are shown in Fig. 10.9 (b). Higher energy levels are usually not considered.



（a）　（b）

Fig. 10.9摇（a）Conduction-band edge at N-AlGaAs, n-GaAs eterojunction;（b）triangular wellapproximation with discrete electron energiesl.

Fig. 10.10摇Electron density in triangular potential well.

The qualitative distribution of electrons in the potential well is shown in Fig. 10.10. A current parallel to the interface will be a function of this electron concentration and of the electron mobility. Since the GaAs can be lightly doped or intrinsic, the two-dimensional electron gas is in a region of low impurity doping so that impurity scattering effects are minimized. The electron mobility will be much larger than if the electrons were in the same region as the ionized donors.[10]

Words and Expressions

egularly摇*adv*. 有规律地，有规则地，整齐地，匀称地	elastically摇*adv*. 有弹性地，伸缩自如地
epitaxial摇*adj*. (晶体)取向附生的，外延的	pseudomorphic摇*adj*. 伪形的，假的
coherent摇*adj*. 粘在一起的，一致的，连贯的	metallurgical摇*adj*. 冶金学的

Glossary of Important Term

homojunction摇同质结	straddling摇跨骑
heterojunction摇异质结	staggered摇交错
isotype heterojunction摇同型异质结	broken gap摇错层
anisotype heterojunction摇反型异质结	electron affinity rule摇电子亲和准则

Notes

1. With recent advances in MOCVD and MBE epitaxial growth techniques for Ⅲ-Ⅴ compound semiconductors and SiGe/Si systems, it is now possible to grow extremely high-quality Ⅲ-Ⅴ heterojunction structures with layer thickness of 100 Å or less for quantum dots, superlattices, and multiquantum-well (MQW) device applications.

提示: 在翻译时，extremely high-quality Ⅲ-Ⅴ heterojunction 和 layer thickness of 100 Å or less 都可以作为 structures 的定语。

2. In other words, suitable translations of the basic unit cell of a crystal restore the crystal back into itself.

提示: In other words 表示“换句话说”。

3. The first is that a bulk crystal can include impurities and dislocations such that the perfect periodicity of the material is disrupted locally.

提示: such that 引导表示结果的从句，可翻译为“这样”。结合上文，the first 后面省略了 exception，在翻译中应给出。

4. A very thin layer of material can be grown on top of or sandwiched between layers grown with a different type of semiconductor material, even materials in which the lattice constant is different.

提示：grown on top of 和 sandwiched between 有一个共同的宾语 layers，grown with a different type of semiconductor material 作 layers 的定语。

5. When a thin layer of material is grown either on or between layers of a different semiconductor that has a significantly different lattice constant, the thin, epitaxial layer will adopt the lattice constant of the neighboring layers provided that the lattice mismatch is less than about 10%.

提示：provided 表示“假如”。either…or…表示“或者……或者”的意思，而所引导的介词 on 和 between 具有相同的宾语 layer，所以 on 后面省略了宾语。

6. Useful heterojunction systems are therefore comprised of materials that are relatively closely lattice matched.

提示：comprised of 表示“由……组成”。relatively 修饰 closely，closely 修饰 matched。

7. For a given strain, determined by the lattice mismatch, a maximum thickness exists, called the critical thickness h_c, for the strained layer to be completely coherent with the substrate.

提示：本句的主句为 a maximum thickness exists；从句 called the critical thickness h_c 修饰 maximum thickness；for the strained layer to be completely coherent with the substrate 是表示目的的状语；For a given strain 是条件状语；determined by the lattice mismatch 是过去分词作定语，修饰 strain。

8. Generally speaking, the higher the frequency, or speed, required for an application, the higher the desired velocity.

提示：Generally speaking 表示“一般而言”。the higher…the higher…两个形容词的比较级联用，表示“越……越……”。

9. As mentioned previously, to achieve thermal equilibrium, electrons from the wide-bandgap AlGaAs flow into the GaAs, forming an accumulation layer of electrons in the potential well adjacent to the interface.

提示：该句中的主语为 electrons，谓语为 flow，from the wide-bandgap AlGaAs 表示流动的方向。forming an accumulation layer of electrons in the potential well adjacent to the interface 中，现在分词引导短语，说明电子流动的结果。

10. The electron mobility will be much larger than if the electrons were in the same region as the ionized donors.

提示：该句的被比较物是 electron mobility，为了使句子简化，在 than 后面省略了比较物，而是直接跟了一个采用虚拟语气的条件从句，指出如果在该区域存在离化施主，那么由于离化杂质散射，迁移率将低得多。

Exercises

1. Translate the first paragraph in the reading material into Chinese.

2. Answer the following questions in English.

(1) Explain what is meant by a two-dimensional electron gas.

(2) summarize the advantages of heterojunction.