Elements of Photonics, Volume II: For Fiber and Integrated Optics. Keigo Iizuka

ISBNs: 0-471-40815-8 (Hardback); 0-471-22137-6 (Electronic)

APPENDIX A

PIN PHOTODIODE ON AN ATOMIC SCALE

A.1 PIN PHOTODIODE

A PIN photodiode consists of p-type, intrinsic type, and n-type semiconductors. The "I"of PIN stands for intrinsic, and we will begin with an explanation of the intrinsic-type semiconductor. The simplest intrinsic-type semiconductor is made out of highly purified group IV elements such as silicon (Si) or germanium (Ge), which have four valence electrons (four electrons in the outermost orbit). Such elements form a crystalline structure by acquiring an additional four electrons from neighboring atoms, making the total number of electrons in the outer orbit eight, and thus completely filling the allowed eight electron states in the outer valence shell. The most stable shell configuration is that of the completely filled shell. A stable shell means that there are no free electrons at 0 K, and the conductivity is zero at 0 K. The bond picture for the intrinsic semiconductor is shown in Fig. A.1a.

The above situation is very often represented by energy levels. Free electrons are at a higher energy level than those in orbit. The energy levels of the free electrons are in the conduction band and those of the orbital electrons are in the valence band. The lowest energy level of the conduction band is denoted by E_c, and the uppermost energy level of the valence band is denoted by E_v. The energy gap between conduction and valence bands is denoted by the symbol E_g, and an energy greater than or equal to E_g is needed for an electron in orbit to be freed. For an intrinsic semiconductor at 0 K, all electrons are in the valence band, and none are in the conduction band. The Fermi function $f(E)$ gives the probability that a given energy state E is occupied by an electron. The *Fermi energy level*, which marks the border between filled and unfilled states, lies midway between the valence band and the conduction band, as shown in Fig. A.1b. There are no actual electron states right at this particular Fermi level, as it lies in a forbidden region. Rather, the Fermi level serves as a convenient border marker between filled and unfilled states.

Figure A.1c illustrates what happens as the temperature is raised in the intrinsic semiconductor. The increased thermal energy means that some electrons will gain enough energy to leave orbit and become free electrons or carrier electrons. These free

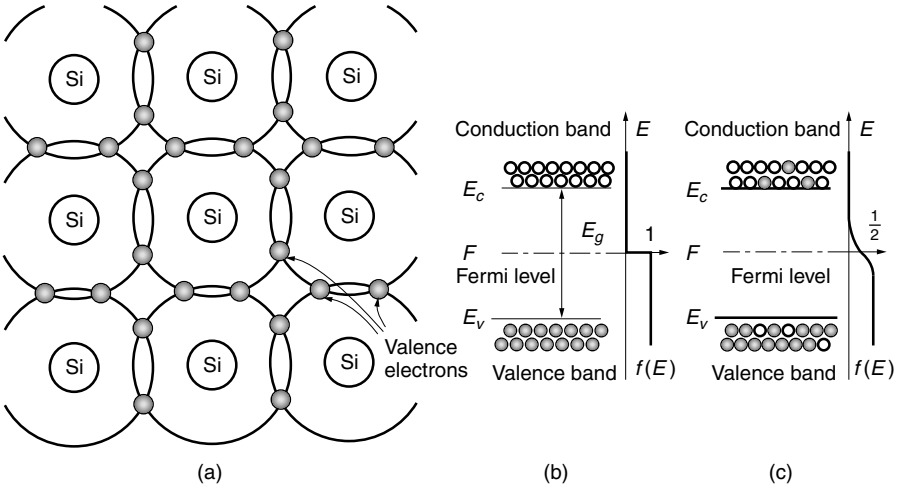

Figure A.1 Silicon intrinsic semiconductor. (a) Bond picture. (b) Energy levels at $T = 0$ K. (c) Energy levels at $T > 0$ K.

electrons are represented by the filled circles in the conduction band in Fig. A.1c. For each electron that leaves orbit to become a free electron, a vacant valence state is left behind, as represented by the open circles in the valence band. The Fermi level is still midway between the valence and conduction levels, and in a broader sense, the Fermi level is defined as that level for which the probability of the state being filled equals the probability of the state being empty. In an intrinsic silicon semiconductor at room temperature, the concentration of free electrons is of the order of 10^{10} cm^{-3} and is much lower than the 10^{22} cm^{-3} concentration characteristic of a metal conductor.

The atomic structure of the n-type layer will be explained next. The n-type layer is made by substituting some of the group IV atoms of the intrinsic semiconductor by donors that are group V atoms like N, P, As, Sb, Bi, or Pa and have five valence electrons. Let us take Si as the intrinsic semiconductor and As as the donor and refer to Fig. A.2a. Wherever the Si atom is replaced by the donor, the total number of available valence electrons becomes nine. At lower temperatures, the As atom uses eight electrons to complete the outermost shell, and the ninth electron is loosely bound in the Coulomb field of the nucleus of the As atom. As the temperature is raised, it only takes a little bit of energy for the ninth electron to break loose and become a free electron. The dissociation of the ninth electron from the As atomic site needs much less energy than that of the other electrons.

In terms of the energy band picture in Fig. A.2b, the energy levels of the ninth electrons of the donor atoms are just slightly below the conduction band, as only a small amount of energy is needed for these electrons to be freed. As a consequence, the Fermi level for the n-type semiconductor is closer to the conduction band, as shown in Fig. A.2b. As the concentration of donor atoms is increased, the Fermi level moves closer and closer to the conduction band, and for very high concentrations of donor atoms, the Fermi level is actually within the conduction band. For a given concentration of donor atoms, raising the temperature doesn't change the Fermi level, but it does cause more electrons to move to the conduction band, and as they do so, vacant states are left behind as shown in Fig. A.2c.

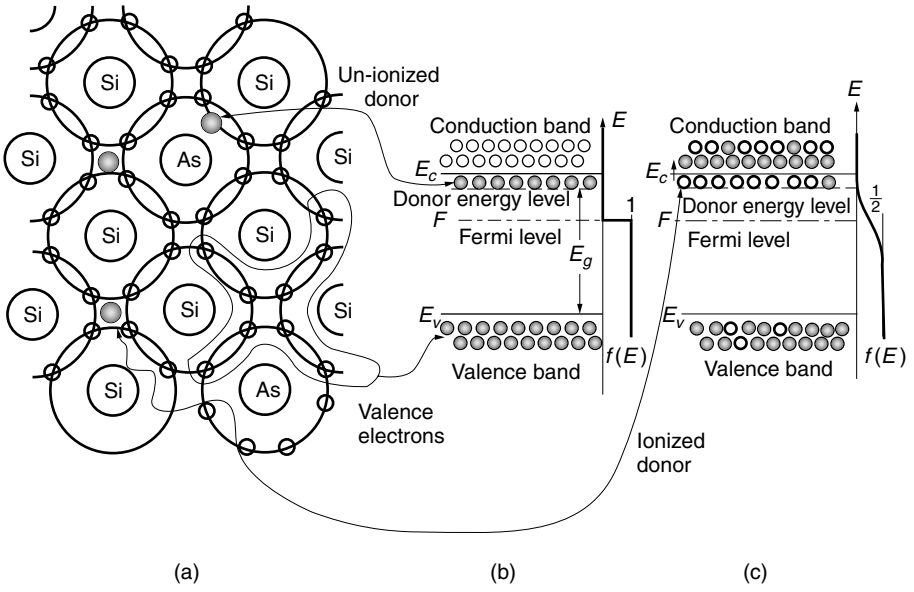

Figure A.2 n-Type semiconductor. (a) Bond picture. (b) Energy levels at $T = 0$ K. (c) Energy levels at $T > 0$ K.

Finally, the p-type semiconductor will be explained. The p-type semiconductor is made by substituting group III atoms for some of the group IV atoms in the intrinsic semiconductor. The group III atoms have three valence electrons and are called acceptor atoms. Examples are B, Al, Ga, In, Tl, or Ac. For the sake of illustration, let us take a specific p-type semiconductor consisting of Si with Ga as the acceptor, as shown in Fig. A.3a. Wherever the Si atom is replaced by the acceptor, the total number of available valence electrons is only seven, which is one short of forming a complete shell. In its eagerness to form a complete shell, the acceptor site may "accept" an electron from a nearby Si atom. If this happens, Ga will have a complete shell, but the Si atom that gave up the electron will now be incomplete. This process continues. The Si atom with the missing electron will capture an electron from another Si atom, thereby completing its shell but leaving an incomplete shell elsewhere. The movement of the missing site is electrically equivalent to the movement of a positive charge. This fictitious mobile charge is called a *hole*. Another term that is used to describe the hole is positive carrier.

It is important to realize that n-type and p-type semiconductors are electrically neutral overall. For the n-type semiconductor, for every electron that is freed from the As site to produce a negative carrier, an immobile positive charge is left behind. Likewise, for the p-type semiconductor, for every positive carrier (hole), there is a corresponding immobile negative charge at the Ga site.

The energy bands for the p-type semiconductor at $T = 0$ are shown in Fig. A.3b. The "holes"of Si are represented by the open circles in the valence band, while the electrons are represented by the filled circles. The unfilled eighth electron state of the acceptor atom is represented by an energy level just slightly above the silicon valence band, as it only takes a little bit of energy for an electron to leave the valence band and fill a state at the acceptor energy level. As a consequence, the Fermi level for

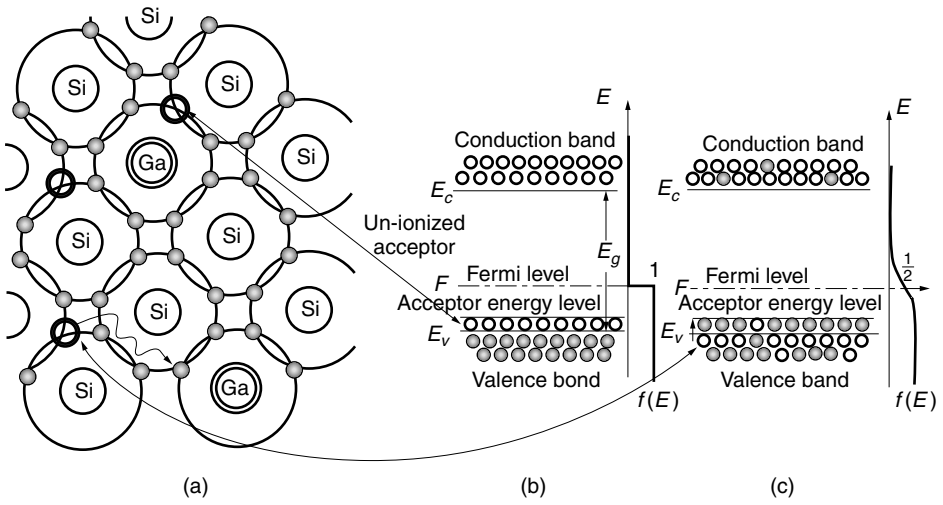

Figure A.3 p-Type semiconductor. (a) Bond picture. (b) Energy levels at $T = 0$ K. (c) Energy levels at $T > 0$ K.

the p-type semiconductor is closer to the valence band, and for high concentrations of acceptor atoms, the Fermi level is within the valence band. As the temperature is raised, some valence electrons will gain enough energy to become free electrons, leaving holes behind, as is the case for Fig. A.3c. Thus, at room temperature, a p-type semiconductor can have some free electrons. A point to keep in mind is that the density of holes will be much greater in a p-type semiconductor than the density of free electrons.

So far, only the simplest forms of semiconductors have been discussed. Before starting the topic of p-n junctions, it is worth mentioning that an intrinsic semiconductor can also be made by using compounds of group III and group V atoms in equal portions. Examples are gallium arsenide (GaAs) and indium phosphide (InP). Using group II atoms to replace some of the group III atoms in the III–V compound produces a p-type semiconductor. Likewise, using group VI atoms to replace some of the group V atoms in the III–V compound results in an n-type semiconductor.

So far, n- and p-type semiconductors have been considered separately, as indicated in Fig. A.4a. They will now be joined together as indicated in Fig. A.4b to form a p-n junction. The electron carrier density in the n-type layer is about 10^5 times larger than the p-type layer. The free electrons in the n-type layer start to diffuse into the p-type layer. This diffusion, however, does not continue forever. If it did, the result would be a perpetual current generator!

The neighborhood of the As atom was originally electrically neutral before the connection. As the electrons start leaving the site of As due to diffusion, the neighborhood of the As atom effectively becomes electrically positive. This positive charge $\oplus$ is, however, immobile because the As atom is held firm in the crystal. Immobile charges are encircled in Fig. A.4 and free charges are not. In order to emphasize this lack of mobility, the pattern of $\oplus$ and $\ominus$ in Fig. A.4a is preserved in Fig. A.4b.

The center of the junction region is normally an intrinsic region because when a free electron meets a hole (or empty site of the orbit), there is a good chance that

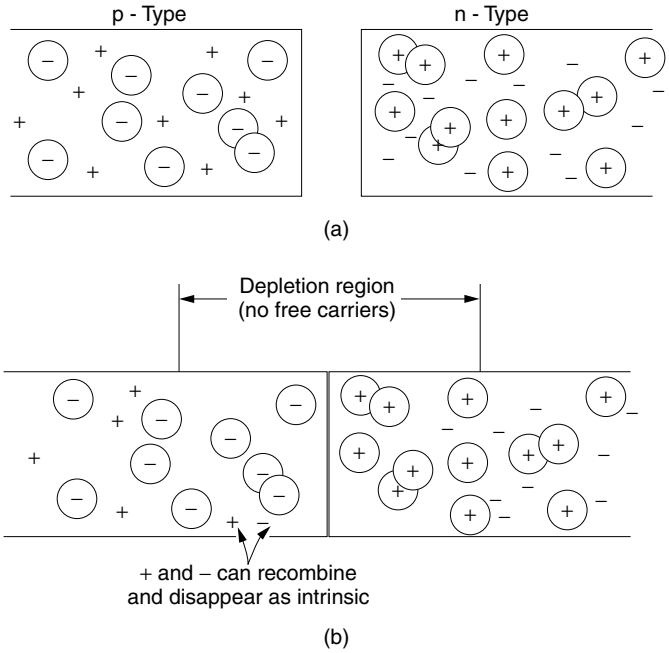

Figure A.4 Flow of charges after junction is made. (a) Before junction is made. (b) After junction is made.

the free electron will get back into orbit. One way of describing this process is to say that the free electron and hole have recombined, and the result of the recombination is a filled valence state. In analogy with particle physics, an alternate description is to view the process as the *annihilation* of a free electron–hole pair. The reverse process is called *pair production*.

It should be mentioned that recombination is most likely to take place between the positive and negative carriers. Recombination of a free electron with an immobile $\oplus$ charge is far less common. For instance, $\oplus$ of As has already disposed of the ninth electron, and because it is reasonably content with its completed external shell of eight electrons, it is almost indifferent to gaining back the ninth electron. On the other hand, the positive carrier means that a Si atom somewhere has an incomplete shell of seven valence electrons and this atom is very eager to grab a free electron to complete its outershell to a full complement of eight.

Getting back to the diffusion in the junction region, as the diffusion of free electrons progresses, the amount of accumulated immobile charges $\oplus$ increases. The Coulomb force between the immobile positive charge $\oplus$ and the negatively charged electrons slows down the diffusion. At equilibrium, there is no diffusion and the n-type layer is positively charged. With precisely the same reasoning but with opposite polarity, the equilibrium distribution of the holes is established. Holes diffuse into the n-layer, but at equilibrium, the diffusion is halted by the accumulation of negative $\ominus$ immobile charges left behind in the p-type layer. Thus, the p-type layer is negatively charged. The potential difference caused by the immobile charges is the contact potential of the junction.

A.2 I–V CHARACTERISTICS

The movement of electrons across the p-n junction will be calculated using energy bands. The contact potential V_0 causes a downward translation of the energy levels of the n-type layer by eV_0. In equilibrium, the Fermi level is the same throughout the p and n regions as indicated in Fig. A.5a. The case of lightly doped n and p regions at room temperature is considered, although the analysis is applicable to high dopant concentrations as well. Without an external bias voltage, the lower edges of the conduction bands of the p and n layers are higher than the Fermi level by A and B electron volts, respectively, as indicated in Fig. A.5a, and

$$A = eV_0 + B \tag{A.1}$$

With a forward bias V (positive potential) applied to the p-type layer and negative potential to the n-type layer, the whole energy band structure of the n-type layer translates upward by eV as indicated in Fig. A.5b. In order to cut down on the number of notations, the spacing between the levels rather than the levels themselves will be used here. When the term electron density is used without qualification, the density of free electrons is implied.

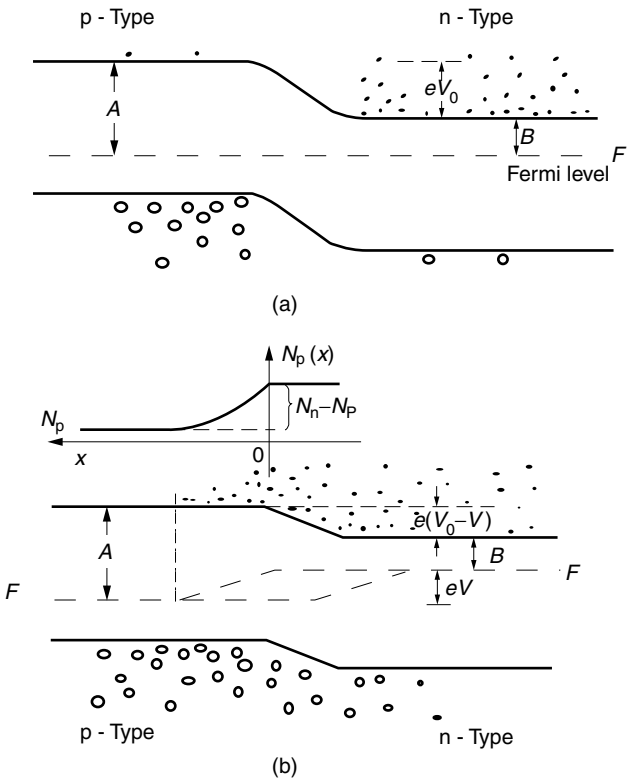

Figure A.5 Energy band structure of a p-n junction. (a) Energy levels without bias. (b) Energy levels with bias.

In the present explanation, simplicity will override a rigorous derivation of the expression and the Boltzmann distribution function rather than the Fermi distribution function will be used, even though this assumption is true only when the levels considered are away from the Fermi level. According to the Boltzmann distribution, the electron density at an energy level higher than the Fermi level by B electron volts is simply expressed by

$$N'_{\mathrm{n}} = N_0 e^{-B/kT} \tag{A.2}$$

where N_0 is the effective electron density extrapolated to the Fermi level. N'_{n} is the electron density at the bottom of the conduction band of the n-type layer.

Now, the electron density when a forward bias is applied is calculated. The electron density in the n layer at the level corresponding to the bottom of the conduction band of the p-type layer is, from Fig. A.5b,

$$N_{\mathrm{n}} = \left(N_0 e^{-B/kT}\right) e^{-e(V_0 - V)/kT} \tag{A.3}$$

Using Eq. (A.1) in (A.3),

$$N_{\mathrm{n}} = N_0 e^{-A/kT} e^{eV/kT} \tag{A.4}$$

The electron density N_{p} at the bottom of the conduction band of the p-type layer can be calculated in a similar manner and

$$N_{\mathrm{p}} = N_0 e^{-A/kT} \tag{A.5}$$

Note that Eq. (A.5) is identical with the first factor of Eq. (A.4), and

$$\begin{aligned} N_{\mathrm{p}} &< N_{\mathrm{n}} \quad (V > 0, \text{ forward bias}) \\ N_{\mathrm{p}} &< N_{\mathrm{n}} \quad (V < 0, \text{ backward bias}) \end{aligned} \tag{A.6}$$

A step of the electron density $(N_{\mathrm{n}} - N_{\mathrm{p}})$ is created across the junction. With a forward bias, electrons in the n-type layer diffuse into the p-type layer passing through the intrinsic layer without much loss of electrons. Inside the p-type layer, however, there is a sea of holes to annihilate the electrons (recombine with electrons). The electron density decays as the electrons go deeper into the p layer. Without a supply of electrons from an external source, the prescribed difference $(N_{\mathrm{n}} - N_{\mathrm{p}})$ could not be maintained because of the annihilation inside the p layer.

The amount of necessary supply current is calculated from the diffusion equation. The spatial distribution of the electrons inside the p-type layer is

$$N_{\mathrm{p}}(x) = N_{\mathrm{p}}(0) + (N_{\mathrm{n}} - N_{\mathrm{p}}) e^{-x/L_{\mathrm{n}}} \tag{A.7}$$

where x is the distance away from the junction. The origin is taken at the beginning of the p-type layer and not the center of the p-n junction. The electric current due to the gradient of the electron density is

$$J_e(x) = -eD_{\mathrm{n}} \frac{d}{dx} N_{\mathrm{p}}(x) \tag{A.8}$$

where D_n is the diffusion constant for the electrons. Insertion of Eq. (A.7) into (A.8) with $x = 0$ gives the current at the junction as

$$J_e(0) = \frac{eD_n}{L_n}(N_n - N_p) \tag{A.9}$$

By inserting Eqs. (A.4) and (A.5), one finally obtains the current for the p-n junction:

$$J_e = I_{so}\left(e^{eV/kT} - 1\right) \tag{A.10}$$

with

$$I_{so} = \frac{eD_n}{L_n}N_p + \frac{eD_p}{L_p}N_n \tag{A.11}$$

The second term in Eq. (A.11) accounts for the contribution of the holes. As seen from Eq. (A.10), I_{so} is the current when the p-n junction is deeply back-biased. It is called the saturated back-bias current.

The expressions so far are for the case without incident light. When light whose quantum hv is larger than the bandgap is incident on the junction, there is a chance

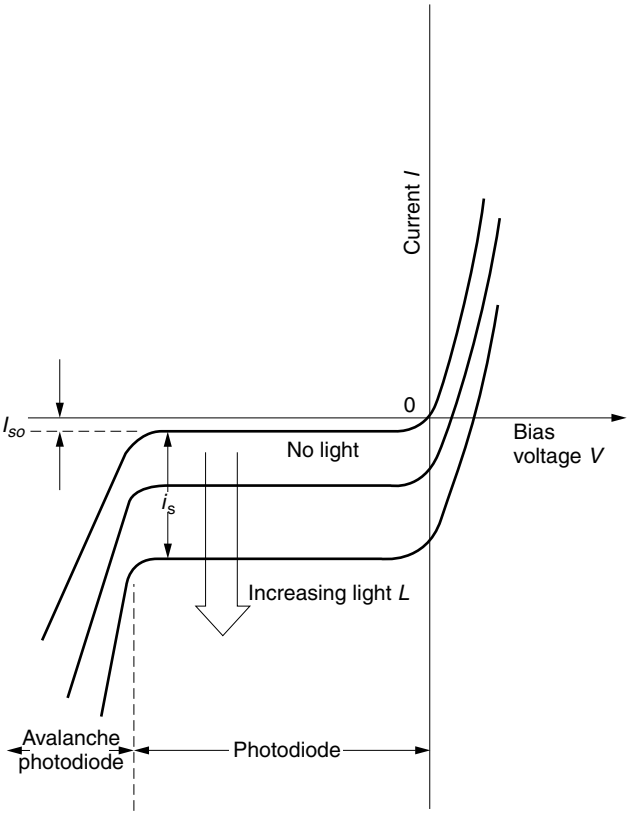

Figure A.6 Current–voltage characteristic of a p-n junction.

that an electron in the valence band will absorb the energy of the photon and move into the conduction band, leaving a hole behind. This process is called *photon-induced pair production*. The current due to the incident photons or photocurrent has to be added to Eq. (A.10). The characteristic curve including the photocurrent is shown in Fig. A.6. Increasing the number of incident photons causes a downward translation of the I–V curve. In the negative bias region, also referred to as the back-biased region, the n layer is positive and the p layer is negative. The negative bias region is used as a photodetector because the dark current (current in absence of the incident light) is at its minimum, having reached the saturated back-bias current. The output current is then primarily due to the photocurrent. In this region, the incremental impedance of the p-n junction is practically independent of the back-bias voltage.

The region of extreme negative bias is that of the avalanche effect. With a large bias voltage, the accelerated electrons start creating additional electron–hole pairs by colliding with the orbital electrons. The photocurrent can be multiplied up to a thousand times. This is the avalanche photodiode (APD).