The semiconductor diode as a rectifier, a light source, and a solar cell: A simple explanation

F. Herrmann[a)]
Abteilung für Didaktik der Physik, Universität Karlsruhe, D-76128 Karlsruhe, Germany

P. Würfel
Institut für Angewandte Physik, Universität Karlsruhe, D-76128 Karlsruhe, Germany

An explanation of the principles of a *pn* junction is proposed without recourse to the band model, the space charge of the junction, and the charge carrier depletion at the interface. The explanation assumes that the processes in a *pn* junction can be considered as a chemical reaction between electrons, holes, and photons and that an *n*-type material is a conductor for electrons and an insulator for holes, and a *p*-type material is a conductor for holes and an insulator for electrons. We give a simple and concise explanation of rectification, light emission, and current generation by *pn* junctions. © *2006 American Association of Physics Teachers.*
[DOI: 10.1119/1.2192786]

I. INTRODUCTION

In this paper I discuss three different applications of a *pn* junction. When used as a rectifier, a *pn* junction is called a semiconductor diode. As a light source, it is a light emitting diode or LED, and when used as a photoelectric energy converter, a *pn* junction is often called a solar cell. These devices are so ubiquitous that a physics student should know their working principles.

To obtain a satisfactory understanding of *pn* junctions, a comprehensive knowledge of solid state and semiconductor physics might seem necessary.[1–3] Many of the concepts involving *pn* junctions go back to Shockley.[4] We will see that a simple and correct explanation is possible if we are willing to adopt some ideas from chemistry, and we will find that the *pn* junction is a close relative of the electrochemical cell. Understanding the band model, the space charge of the junction, and the charge carrier depletion at the junction will not be necessary.

We limit our discussion to the explanation of the physical principles. The derivation of the diode's current-voltage characteristic and thermodynamic efficiency based on these principles is given in Ref. 5.

II. CHARGE CARRIERS IN SEMICONDUCTORS

A. Chemical reactions

A semiconducting material is essentially an insulator. Its electrons are either bound to the inner shells of the atoms or form covalent bonds with neighboring atoms and thus are fixed and not available for conduction. However, if photons or phonons are present, such a material can become a conductor. A photon or phonon of appropriate energy can free an electron from its bond so that the electrons become mobile and contribute to the conductivity. In such a process another mobile charge carrier is also created. The missing electron or hole can jump from one site to a neighboring site and move through the crystal like a positive charge carrier.

Because the properties of mobile electrons differ from those of bound electrons, it is convenient to imagine the mobile electrons as a different species of particles. They can be considered as the particles of a substance in the sense of chemistry. We will use the symbol *e* for the mobile electrons, and *h* for holes in the same way as chemists use the symbol H for hydrogen. We also need to consider light as a substance for which the corresponding particles are photons with the symbol γ.

The absorption of a photon, which frees an electron from its site and creates a free electron and a free hole, can be described as a chemical reaction:

$$\gamma \rightarrow \square \ \square . \tag{1}$$

The reaction can also proceed in the opposite direction:

$$\square \ \square \rightarrow \gamma . \tag{2}$$

In the case of chemical equilibrium with the surroundings, we have

$$\gamma \rightleftarrows e + h . \tag{3}$$

Here γ stands for those photons of the thermal background radiation (if no other radiation is present) that have sufficient energy to free an electron. Due to this radiation a semiconductor always contains a small number of electrons and holes. However, the conductivity that results from these charge carriers is so small that the material can still be considered an insulator.[6]

The conductivity of the semiconductor material can be increased by adding a small amount of donors or acceptors. Because an electron of a donor atom is only weakly bound, it can be set free by a phonon. Correspondingly, a hole can be freed from an acceptor atom. At room temperature the donors and acceptors are completely ionized. Consequently, in an *n*-doped material the number of additional electrons equals the number of donors, and in a *p*-doped material the number of additional holes equals the number of acceptors. For instance, silicon can be *n*-doped with arsenic and *p*-doped with boron.

B. The law of mass action

In chemical equilibrium, electron-hole-pair recombination is on average as frequent as the inverse process, that is, the creation of an electron-hole pair by a photon. The production rate of electron-hole pairs depends on the concentration of photons of sufficient energy in the background radiation. It does not depend on the concentration of donors and accep-

Table I. The electron and the hole concentrations in undoped, n-doped, and p-doped silicon in particles per cm^3. The product $c_e c_h$ is the same in every case.

	c_e	c_h	$c_e c_h$	
Undoped	10^{10}	10^{10}	10^{20}	Insulating
n-doped	10^{17}	10^{3}	10^{20}	Electron conductor
p-doped	10^{3}	10^{17}	10^{20}	Hole conductor

tors as long as their presence does not change the absorption properties of the semiconductor. The rate of electron-hole recombination (which is equal to the rate of photon production) depends on the concentration of electrons and holes. Because both an electron and a hole are involved in the reaction, the reaction rate is proportional to the product of the concentrations of the electrons and the holes. In steady state the rate of production must equal the rate of annihilation. The product of the two concentrations is therefore independent of the doping:

$$c_e c_h = \text{constant}. \tag{4}$$

For twice the electron concentration (due to doping), the hole concentration will be smaller by a factor of 2. Because the concentration of photons in the ambient radiation depends on the temperature, the same holds for the value of the product $c_e c_h$. Chemists recognize Eq. (4) as the law of mass action. For silicon at a temperature of 300 K, $c_e c_h \approx 10^{20}/cm^6$. (We have expressed the concentration in terms of particles per cm^3.)

If the silicon is not doped, the concentration of the electrons must equal the concentration of the holes. To obtain $10^{20}/cm^6$ for the product in Eq. (4), the concentration of each of the two substances (electrons and holes) must be $10^{10}/cm^3$ (see Table I). If the silicon is n-doped with 10^{17} arsenic atoms per cm^3, it contains 10^{17} free electrons per cm^3 and thus 10^3 holes. Correspondingly, if it is p-doped with 10^{17} boron atoms per cm^3, the electron concentration is 10^3 electrons per cm^3.

C. Substance-specific conductivities

To describe the electrical properties of a material, we usually consider the electrical conductivity, and not the nature of the charge carriers. However, there are phenomena that can be understood only if we specify which particles carry the electric current. Such a distinction is important in electrochemistry. A material is called an electrolyte if it is a conductor for ions and a non-conductor for electrons. Metallic conductors are conductors for electrons and (in general) non-conductors for ions. The workings of an electro-chemical cell is based on the differences of these substance-specific conductivities.

A semiconductor diode is also based on such differences. An n-doped material is a conductor for electrons and an insulator for holes, whereas a p-doped material is a hole conductor and an electron insulator (see Table I). Let us consider an n-doped material with appropriate (ohmic) contacts on the left and the right sides. If electrons are fed into the material from the left, electrons will come out on the right. (This behavior is analogous to that of a water pipe that is already filled with water. When water is pushed into the pipe from
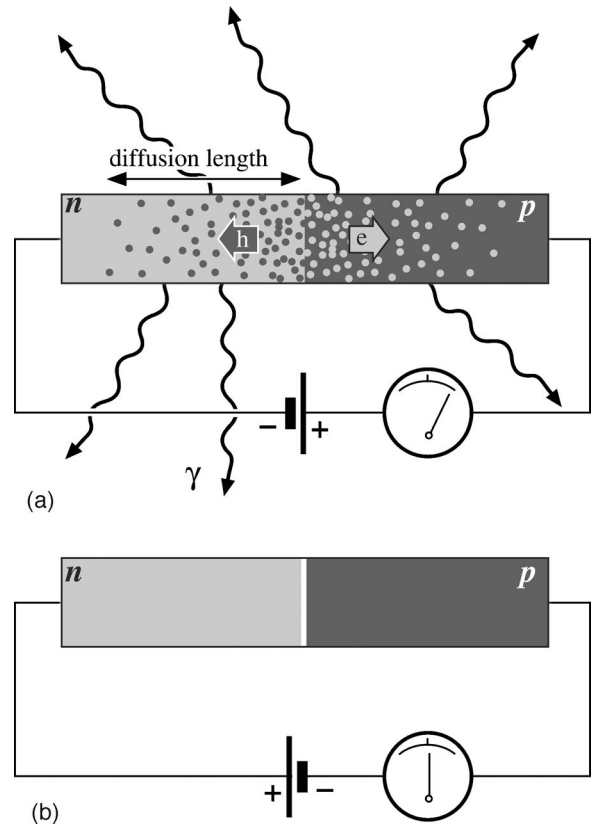

Fig. 1. (a) Forward polarity: Electrons and holes move toward the interface. Electrons penetrate into the p region and holes into the n region. Both recombine within a diffusion length, thereby emitting light. (b) Reverse polarity: To have an electric current, electrons and holes would have to move from the interface. Because they are generated by the few photons of the thermal background radiation (and by thermal phonons), the current is very small.

one side, water will immediately come out at the other side, although a particular portion of water may take a long time to cross the entire pipe.)

III. THE SEMICONDUCTOR DIODE AS A RECTIFIER AND AS A LIGHT SOURCE

A semiconductor diode consists of two adjacent regions of a semiconducting material, one being n-doped and the other p-doped. That is, one region is an electron conductor and the other is a hole conductor. Each region has an ohmic metallic contact. An ohmic contact to an n-conductor exchanges mainly electrons and an ohmic contact to a p-conductor exchanges mainly holes.

We first connect the diode to a battery in the low-resistance or forward direction [see Fig. 1(a)]. The charge carriers in the n and the p region move in opposite directions—the electrons to the right, the holes to the left. We consider what happens in the vicinity of the interface between the n and the p material. From the left, electrons cross the contact interface into the p region. As a consequence, the e concentration in the p region becomes larger than the equilibrium value, which means that the chemical equilibrium of the reaction in Eq. (3) is disturbed. Hence, we obtain a net reaction rate: light is produced at the expense of electrons and holes.

In the same way holes that cross the pn interface and pen-

F. Herrmann and P. Würfel   2

etrate into the *n* region perturb the chemical equilibrium on the left-hand side and thus also give rise to the production of light, which is more intense than the background radiation. Because the electrons and holes in a diode recombine before they reach the opposite contact, the electric current is limited by the reaction rate.

When the diode is connected in the reverse direction [see Fig. 1(b)], there is no electric current. To obtain a current, charge carriers would have to move away from the interface. Because there is no source for such carriers, no carriers can move away or more precisely, almost no carriers. As we argued in Sec. II A, the ambient radiation produces *e-h* pairs at a low rate. This small supply of charge carriers is responsible for the reverse current. The reverse current does not depend on the voltage because no more carriers can move away from the contact region than are produced by the photons of the background radiation (and by phonons).

Let us return to the forward direction. We now ask where reaction (2) occurs. One might think that the electrons that move into the *p* region react with the holes as soon as they cross the *pn* interface. Correspondingly, it might be expected that the holes which cross the interface from *p* to *n* recombine immediately with the electrons. In reality, the charge carriers do not behave in this way. The reason is the same as for many other chemical reactions.

A chemical reaction can proceed only if two conditions are fulfilled: The chemical potential of the reactants must be greater than that of the reaction products, and the reaction must not be inhibited. (The analogous conditions hold for an electric current: a voltage is needed and the electric resistance must not be too large.) A familiar example of a chemical reaction that normally is strongly inhibited is the burning of hydrogen:

$$2H_2 + O_2 \rightarrow 2H_2O. \tag{5}$$

Although the chemical potential (under normal conditions of pressure and temperature) of the reactants is substantially higher than that of the water, hydrogen and oxygen can be stored together in a container with almost no reactions taking place. The reason is that the reaction rate is strongly inhibited.

In reality, reaction (2) is not as strongly inhibited as reaction (5), but nevertheless, an electron can move a great distance into the *p* material before it reacts with a hole. We can say that for an electron in the *p* region or a hole in the *n* region, it is difficult to find a partner. This observation is crucial for understanding the solar cell. The average distance that an electron in the *p* material or a hole in the *n* material covers before reacting is called the diffusion length. For electrons in high quality silicon it may be as large as several hundred microns. The region within a diffusion length on both sides of the *pn* interface is the region where the processes of interest occur: The reactions of electrons and holes with photons that give rise to rectification, light emission, and current generation.

IV. THE SEMICONDUCTOR DIODE AS A SOLAR CELL

Before trying to understand the solar cell, we consider a simpler situation. What happens when we shine a small light beam on a *p*-doped semiconductor? The light will produce electron-hole-pairs at a rate that depends only on the light intensity and does not depend on anything that happens sub-
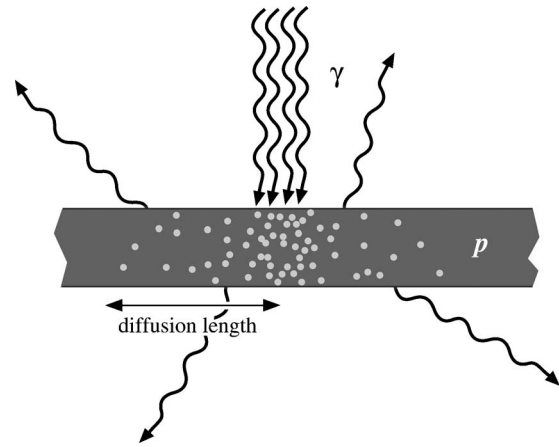

Fig. 2. Light incident on a *p* material. The electron concentration at the location of the light absorption is much higher than in the surrounding area.

sequently to the electrons and holes (see Fig. 2). Because the *p* material is a conductor for holes only, the electrons that are produced by the light accumulate and a concentration gradient develops. Due to their high concentration, the holes can easily accommodate in such a way that the material remains neutral at every point. The concentration (chemical potential) gradient of the electrons drives them away from where they were produced and they react with the holes according to Eq. (2). The average distance that an electron covers before reacting with a hole is the diffusion length.

The device of Fig. 2 can easily be transformed into a solar cell. Instead of shining the light anywhere in the middle of a *p*-conducting crystal, we direct it on a point within a diffusion length from the *pn* interface (see Fig. 3). Contrary to the case of Fig. 2, the electrons now can escape. Instead of recombining, the electrons reach the *pn* interface behind which they encounter a region of high electron conductivity. Thus, they flow toward the left. The holes find themselves in a region of high hole conductivity from the beginning and flow toward the right. They cannot move to the left because their movement is blocked by the *n*-type material.

We also could have directed the light to a point on the left of the *pn* interface. Essentially the same behavior would occur: The holes reach the hole conductor, which is less than a
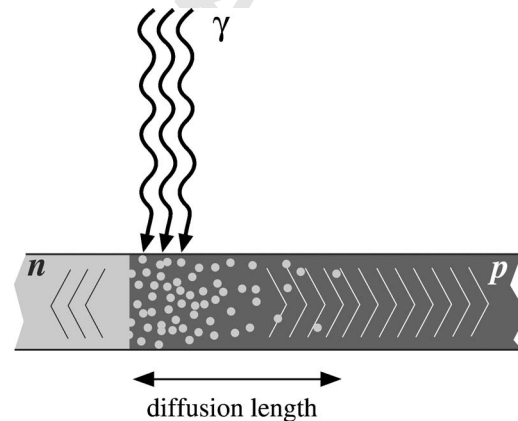

Fig. 3. Equal amounts of electrons and holes are produced in the *p* material. Because the production takes place within a diffusion length from the interface, the electrons can escape into the *n* material on the left.

diffusion length away, and thus can flow to the right, whereas the electrons are in an electron conducting material and can flow to the left. In reality, the light is not directed to a spot on either side of the *pn* interface, but to the entire volume within a diffusion length of both sides of the *pn* interface. In this way, we take the maximum advantage of the *pn* junction.

The driving force for the charge carrier current in a solar cell is the excess concentration of the *e-h* pairs, which is caused by the incident light. The newly produced carriers can flow only in opposite directions because the material is an electron conductor and a hole insulator on one side, and a hole conductor and an electron insulator on the other side of the *pn* interface.

V. DEMONSTRATION

Semiconductor diodes are used as rectifiers, light sources, and solar cells. Each of these devices is optimized according to the purpose for which it is made. However, these differences are not essential for understanding the underlying physical principles. In a classroom demonstration we show all three effects by using the same diode. An appropriate choice is a commercial red LED, which should not be too small. By first showing the rectifier effect, we see that the diode emits light when connected in the forward direction. Then we connect the diode directly, that is, without any power supply, to a voltmeter with a high internal resistance. If light falls on the diode, the meter will display a voltage of the order of 1 V. (Because the *pn* interface area is very small, the short-circuit current is difficult to measure in this experiment.)

VI. SUMMARY

In a semiconductor diode the current in one direction is larger than in the opposite direction because of the asymmetry of the reaction of electrons and holes with photons. In the forward direction electrons and holes recombine and produce photons. The photons can easily escape and therefore the current can be large. These photons make up the light emitted by a LED. In the reverse direction electrons and holes are produced by the few photons of the background radiation (and by phonons) and therefore the current is very small.

The carriers produced by the additional radiation in a solar cell can move only in opposite directions because the solar cell is an electron conductor on one side and a hole conductor on the other side of the *pn* interface. The driving force for the charge carriers results from the excess concentration of the *e-h* pairs, which is caused by the incident light.

[a] Electronic mail: friedrich.herrmann@physik.uni-karlsru

[1] P. A. Tipler and G. Mosca, *Physics for Scientists and Engineers* (Freeman, New York, 2004), pp. 1248–1250.

[2] C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1996), p. 573.

[3] R. Tilley, *Understanding Solids* (Wiley, New York, 2004), p. 463.

[4] W. Shockley, *Electrons and Holes in Semiconductors* (Van Nostrand, Toronto, 1950).

[5] P. Würfel, *Physics of Solar Cells* (Wiley-VCH, Weinheim, 2005).

[6] In addition to the reaction with photons in Eqs. (1)–(3) there is a similar reaction with phonons (non-radiative transitions). This contribution to the reaction increases the reaction rates in both directions but leaves the concentrations unaltered. Our treatment of the *pn* junction is therefore not restricted to the ideal case where all transitions are radiative.