19 Lasers & semiconductors

Content

19.1 Basic principles of lasers
19.2 Energy bands, conductors and insulators
19.3 Semiconductors
19.4 Depletion region of a p-n junction

Learning Outcomes

Candidates should be able to:

(a) recall and use the terms spontaneous emission, stimulated emission and population inversion in related situations.

(b) explain the action of a laser in terms of population inversion and stimulated emission. (Details of the structure and operation of a laser are not required.)

(c) describe the formation of energy bands in a solid, with reference to conduction electrons and holes.

(d) distinguish between conduction band and valence band.

(e) use band theory to account for the electrical properties of metals, insulators and intrinsic semiconductors, with reference to conduction electrons and holes.

(f) analyse qualitatively how n- and p-type doping change the conduction properties of semiconductors.

(g) discuss qualitatively the origin of the depletion region at a p-n junction and use this to explain how a p-n junction can act as a rectifier.

9646_2014

19.1 Basic principles of lasers

Explain

- In 1916, **Albert Einstein** laid the foundation for the invention of the **laser** (*Light Amplification by Stimulated Emission of Radiation*) and its predecessor, the maser (*Microwave Amplification by Stimulated Emission of Radiation*), in a ground-breaking re-derivation of **Max Planck's law of radiation** based on the concepts of *spontaneous* and *induced emission*. The theory was forgotten until after World War II.

- In 1953, **Townes** and graduate students **Gordon** and **Zeiger** produced the first maser, a device operating on similar principles to the laser, but producing microwave rather than optical radiation.

- The first application of *lasers* visible in the daily lives of the general population was the supermarket barcode scanner, introduced in 1974. The laserdisc player, introduced in 1978, was the first successful consumer product to include a laser, but the compact disc player was the first laser-equipped device to become truly common in consumers' homes, beginning in 1982, followed shortly by laser printers.

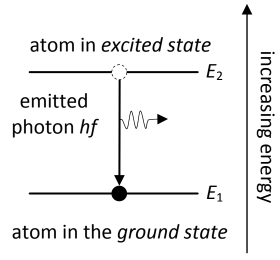
- Some of the other applications include medical (Bleed less Surgery, Laser healing, Survical treatment, Kidney stone treatment, Eye treatment, dentistry), Industrial (Cutting, Welding, Material heat treatment), Defense (Battle field, Anti-missile, Directed Energy Weapon (DEW), Electro Optic Counter Measures (EOCM)), Research tool (Spectroscopy, Laser ablation, Laser annealing, Laser scattering, Laser interferometers, LIDAR), Product development / Commercial (Laser Printers, Compact disc, Barcode scanners, Laser pointers, Holograms).

Absorption

- If light (photons) of frequency f pass through a group of atoms, there is a possibility of the light being absorbed by atoms which are in the *ground state*, which will cause them to be excited to the *higher energy state*. The probability of absorption is proportional to the radiation *intensity* of the light, and also to the number of atoms currently in the *ground state*.

Spontaneous emission

Spontaneous emission is the process by which an atom or molecule in an *excited state* drops to the *ground state*, resulting in the creation of a *photon*, without any external agent.



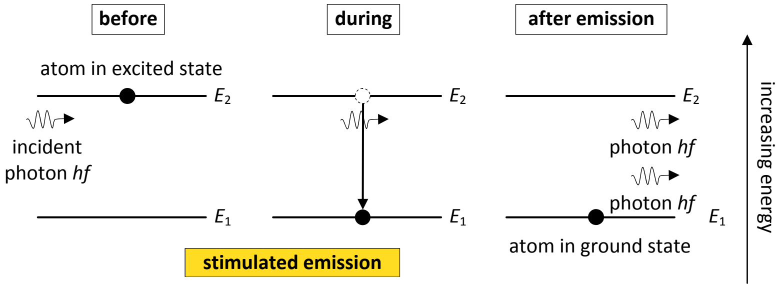
spontaneous emission

- *Spontaneous emission* of light or luminescence is a fundamental process that plays an essential role in many phenomena in nature and forms the basis of many applications, such as fluorescent tubes, television screens, plasma display panels, lasers and light emitting diodes.

- These spontaneously emitted photons are in many *energy quanta*, *phases*, *frequencies*, *polarisations* and *directions*, *i.e.*,
  - **polychromatic**: a range of wavelengths and frequencies
  - **incoherent**: a range of polarizations and directions

Stimulated emission

> **Stimulated emission** is the process by which, when perturbed by a photon, matter may lose energy resulting in the creation of another photon. The perturbing photon is not destroyed in the process (cf. absorption), and the second photon is created with the same phase, frequency, polarization and direction as the original.



- The *energy* of the *incident photon* $(E = hf)$ must match exactly the transition difference between the excited energy level and its *ground state* of the atom.

- For every one *incident photon* that enters the atom, two *photons* of the same *energy*, *phase*, *frequency*, *polarisation* and *direction* may be emitted.
  - **monochromatic**: one wavelength and one frequency only
  - **coherent**: same polarization and direction

Population inversion

- Under normal conditions, atoms are in *thermal equilibrium*, *i.e.*, they follow Boltzmann's statistics in that the number of atoms in higher energy states is smaller than the number in the lowest energy state – the *ground state*. The **population**, or number of atoms or molecules, is not favorable to *stimulated emission*. Any atom of higher energy states tend to spontaneously emit photon and decay rapidly to energy state that is similar to its surrounding low-energy atoms.

- In order for *stimulated emission* to take place, one needs to get more of the atoms in a non-equilibrium situation, whereby more atoms are "stuck" in an excited state than are currently in a lower energy state. To do this, one can shine the atoms with intense flashes of light, the process of optical absorption will excite the atoms from their ground state to the desired higher energy states. This process is called **pumping**, and in general does not always directly involve light absorption; other methods of exciting the laser medium, such as high voltage electrical discharge or chemical reactions may be used.

- This *pumping* process will excite an appreciable number of atoms into the excited energy states. To obtain *population inversion*, a necessary condition whereby *stimulated emission* can take place most often, at least half the population of atoms must be excited from the *ground state*. The laser medium is said to be strongly *pumped*.

- An *optical amplification* is now possible.

A **population inversion** occurs when a *population* (such as a group of atoms or molecules) exists in state with more members in an *excited state* than in *lower energy states*.

  - The concept is of fundamental importance in laser science because the production of an *inverted population* is a necessary step in the workings of a laser.
  - It is clear then, that to produce an **inverted population**, the system cannot be at *thermal equilibrium*.

Laser

A **LASER** (*Light Amplification by Stimulated Emission of Radiation*) is an optical source that emits photons in a *coherent* beam. Laser light is typically near-monochromatic, *i.e.* consisting of a single wavelength or hue, and emitted in a *narrow* beam.

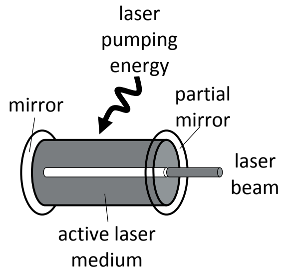
- This is in contrast to common light sources, such as the incandescent light bulb, which emit incoherent photons in almost all directions, usually over a wide spectrum of wavelengths.

- The verb "to lase" means "to produce coherent light" or possibly "to cut or otherwise treat with coherent light", and is a back-formation of the term *laser*.

Construction

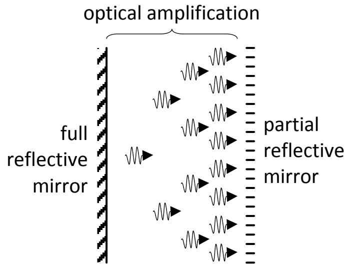
- A laser is composed of an active **laser medium** and a resonant **optical cavity**.

- The **gain medium** is a material of controlled purity, size, and shape, which uses a *quantum mechanical effect* called **stimulated emission** (discovered by *Einstein* while researching the *photoelectric effect*) to amplify the beam. For a laser to operate, the *gain medium* must be



principal components of a laser

"pumped" by an external energy source, such as electricity or light (from a classical source such as a *flash lamp*, or another laser). The pump energy is absorbed by the laser medium to produce **excited states** in the medium. When the number of particles in one excited state exceeds the number of particles in some lower state, **population inversion** is achieved. In this condition, an optical beam passing through the medium produces more *stimulated emissions* than *stimulated absorptions* so the beam is amplified. An excited laser medium can also function as an **optical amplifier**.



- The **optical amplification** process starts when a single incident photon passes an atom, parallel to the axis of the gain medium. It will produce 2 photons, which will stimulate other exciting atoms to emit 2 more photons and so on. The number of these emitted photons through stimulation grows *exponentially*. These photons are bounced back and forth by reflection with the mirrors producing even more stimulated photons (*amplification*).
- Photons that are <u>not</u> parallel to the axis of gain medium does not bounce back and forth but escape quickly through the sides of the oscillating medium <u>without</u> *amplification*.

- The light generated by *stimulated emission* is very similar to the input signal in terms of wavelength, frequency, phase, and polarization. This gives laser light its characteristic coherence, and allows it to maintain the uniform polarization and monochromaticity established by the optical cavity design.

- The **resonant cavity** (**cavity resonator**) contains a coherent beam of light between the two reflective mirrors so that each photon passes through the *gain medium* multiple times before being emitted from the output aperture or lost to diffraction or absorption. As light circulates through the cavity, passing through the *gain medium*, if the gain (amplification) in the medium is stronger than the resonator losses, the power of the circulating light can rise **exponentially**.

- Some of these photons are then released via gaps in the partial mirror as a narrow laser beam.

- The laser beam often has a very small divergence (highly collimated), but a perfectly collimated beam cannot be created, due to the effect of diffraction. Nonetheless, a laser beam will spread much less than a beam of incoherent light.

- Lasers are normally classified according to its gain medium (eg. Argon gas, ruby or helium) and the mode could be pulsating or continuous. The power can ranged from mW to MW.

Worked Example

Example 1

Describe how lasers can be used for:

(a) Medical purposes
(b) Music purposes
(c) Metal working

Solution:

(a) Since lasers can be controlled and can have such a small contact area, they are ideal for fine cutting and depth control in scalpels. Medical lasers can also be used to reattach retinas and can be used in conjunction with fiber optics to place the laser beam where it needs to be. Medical lasers can also be used to fuse together skin and stitch up incisions after surgery. **(ans)**

(b) One popular use of lasers is the reading of Compact Discs (CDs). CDs have a reflective aluminum layer that has very small pits put in the aluminum. The pits are translated into binary by the computer and then are used for information. **(ans)**

(c) Lasers allow better cuts on metals and the welding of dissimilar metals without the use of a flux. **(ans)**

19.2 Energy bands, conductors & insulators

Describe

Electrical conduction

- **Electrical conduction** is the movement of **electrically charged** particles through a **transmission medium**. The movement can form an **electric current** in response to an **electric field**. The underlying mechanism for this movement depends on the material.

- *Conduction* is well-described by **Ohm's Law**, which states that the *current* is proportional to the applied *electric field*. The ease with which current density (current per area) j appears in a material is measured by the **conductivity** σ, defined as:

  $$j = \sigma E$$

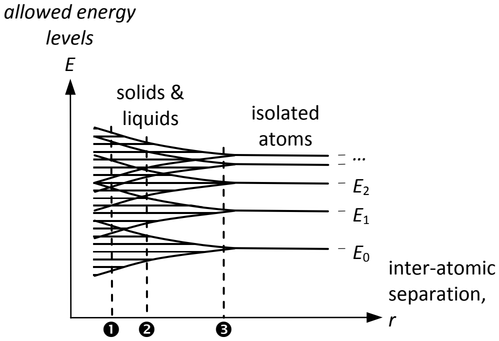
  or its reciprocal **resistivity** ρ:

  $$j = \frac{E}{\rho}$$

- In linear *anisotropic materials*, σ and ρ are tensors.

Energy bands

- In solid state physics, the **electronic band structure** (or simply *band structure*) of a solid is the series of "forbidden" and "allowed" energy bands that it contains. The *band structure* determines a material's electronic properties, optical properties, and a variety of other properties.



- The electrons of a single free-standing atom occupy **atomic orbitals**, which form a discrete set of **energy levels**. If several atoms are brought together into a molecule, their atomic orbitals split, producing a number of **molecular orbitals** proportional to the number of atoms. When a large number of atoms (of order 10^{20} or more) are brought together to form a solid, the number of orbitals becomes exceedingly large, and the difference in energy between them becomes very small. However, some intervals of energy contain no orbitals, no matter how many atoms are aggregated.

- These *energy levels* are so numerous as to be indistinct. First, the separation between energy levels in a solid is comparable with the energy that electrons constantly exchange with **phonons** (**atomic vibrations**). Second, it is comparable with the energy uncertainty due to the **Heisenberg uncertainty principle**, for reasonably long intervals of time.

- In crystalline solids, atoms interact with their neighbors, and the energy levels of the electrons in isolated atoms turn into **bands**. Whether a material conducts or not is determined by its **band structure**. Electrons, being **fermions**, follow the **Pauli exclusion principle**, meaning that two electrons cannot occupy the same state. Thus electrons in a solid fill up the energy bands up to a certain level, called the **Fermi energy**. *Bands* which are completely full of electrons cannot conduct electricity, because there is no state of nearby energy to which the electrons can jump. Materials in which all bands are full (*i.e.* the *Fermi energy* is between two bands) are **insulators**. In some cases, however, the *band theory* breaks down and materials that are predicted to be *conductors* by *band theory* turn out to be *insulators*. **Mott insulators** and **Charge transfer insulators** are two such classes of *insulators*.

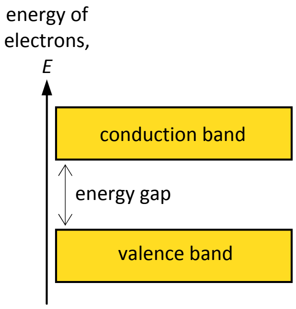
  ❶ – conductors — no energy gaps

  ❷ – semiconductors — small energy gaps

  ❸ – insulators — large energy gaps

Conduction band

- In *semiconductors* and *insulators*, the **conduction band** is the range of electron energy, higher than that of the *valence band*, sufficient to make the electrons free to accelerate under the influence of an applied *electric field* and thus constitute an *electric current*.



energy bands in solids

Valence band

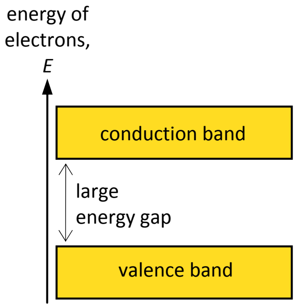
- In solid, the **valence band** is the highest range of electron energies where electrons are normally present at zero temperature. In *semiconductors* and *insulators*, there is a **band gap** above the *valence band*, followed by a *conduction band* above that. In *metals*, the *conduction band* is the *valence band*.

Conductors, insulators and semiconductors

- *In order to conduct a current, electrons must move and so must receive energy to do so. To receive energy, the electrons must be able to transit to an avaliable vacant higher energy level.*

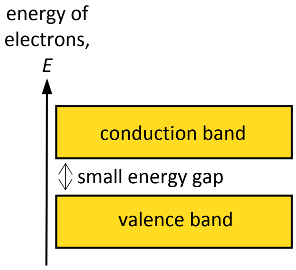
- Solids can be divided into three classes based upon their *band structure*:

❶ **Insulators** contain a **completely empty allowed band** directly above a **completely filled allowed band** (at absolute zero, when the *Fermi-Dirac distribution* is not smoothed). These bands are called the *conduction band* and *valence band*, respectively. The *Fermi level* falls almost exactly in the middle of the forbidden band between them. For reasons that are explained at *electrical conduction*, such a solid has very low *conductivity*.



energy bands in insulators

❷ **Semiconductors** are similar to insulators, but the *conduction* and *valence bands* are spaced closely enough together (≈ 1 eV) that, at room temperature, a nontrivial number of electrons is found in the *conduction band*. These materials have significant *conductivity* that is highly *temperature*-sensitive.

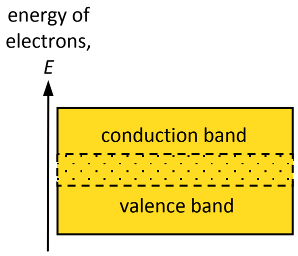


energy bands in semiconductors

- When ionizing radiation strikes a *semiconductor*, it will often excite an **electron** out of its energy level and consequently leave a **hole**. This process is known as the creation of an **electron–hole pair**. Ionizing radiation helps in conduction.
- At low temperatures, *semiconductors* behave like *insulators*. The valence electrons have no adjacent energy levels to transit into, and do not have sufficient energy to the energy gap.
- At high temperatures, *semiconductors* behave like *conductors*. The valence electrons have sufficient energy to cross over the small energy gap to assist in the conduction.
- The conductivity of *semiconductors* is very much lower than that of metallic *conductors* because the number of available conduction electrons is very small (≈ 10^{10} m^{-3}).

☺ *Semiconductors are materials whose electrical conductivities are higher than those of insulators but less than those of conductors.*

❸ **Conductors** contain a band that is partly empty and partly filled regardless of temperature. They have very high conductivity. *Conductors* such as metals, there is no distinction between the *valence band* and *conduction band*.



energy bands in conductors

- At low temperatures, *conductors* have maximum conduction electrons due to insignificant lattice vibrations.
- At high temperatures, *conductors* have lower conduction electrons due to significant lattice vibrations.
- The conductivity of *conductors* is good due to the large number of available conduction electrons (≈10^{28} m^{-3}). These electrons needs very little energy to transit to the conduction band for conduction.

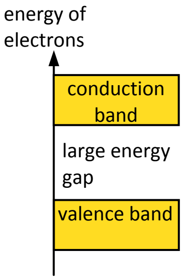
Worked Example

Example 1

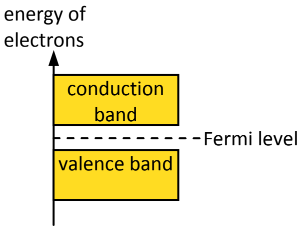
Compare the energy bands in insulators, semiconductors and conductors, illustrating with diagrams if needed.

Solution:

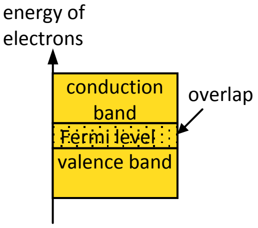
Most solid substances are *insulators*, and in terms of the band theory of solids, this implies that there is a large forbidden gap between the energies of the valence electrons and the energy at which the electrons can move freely through the material (the conduction band). Doping of insulators is insufficient to make them good conductors of electricity.



In *semiconductors*, the Fermi level is essentially halfway between the valence and conduction bands. Although no conduction occurs at 0 K, at higher temperatures, a finite number of electrons can reach the conduction band and provide some current. Doping has a greater effect in semiconductors and it results in extra energy levels.



In *conductors*, there is overlap of the valence band and the conduction band so that at least a fraction of the valence electrons can move through the material. (ans)



19.3 Semiconductors

Describe

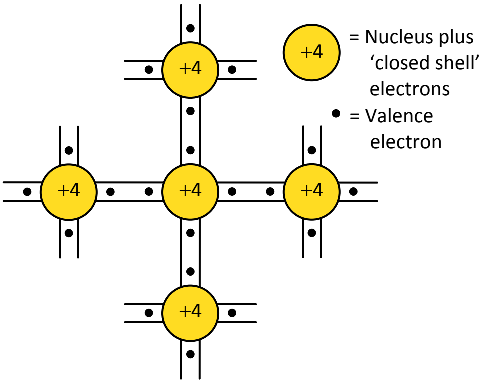
Doping of semiconductors

- One of the main reasons that *semiconductors* are useful in electronics is that their electronic properties can be greatly altered in a controllable way by adding small amounts of impurities. These impurities are called **dopants**.

- Heavily **doping** a *semiconductor* can increase its **conductivity** by a factor greater than a billion. In modern **integrated circuits**, for instance, heavily–doped **polycrystalline silicon** is often used as a replacement for metals.

Intrinsic and extrinsic semiconductors

- An **intrinsic semiconductor** is a *semiconductor* which is pure enough that the impurities in it do not appreciably affect its electrical behavior. In this case, all carriers are created by thermally or optically excited electrons from the full **valence band** into the empty **conduction band**. Thus equal numbers of electrons and holes are present in an *intrinsic semiconductor*. Electrons and holes flow in opposite directions in an electric field, though they contribute to current in the same direction since they are oppositely charged. Hole current and electron current are not necessarily equal in an *intrinsic semiconductor*, however, because electrons and holes have different **effective masses** (crystalline analogues to free inertial masses).



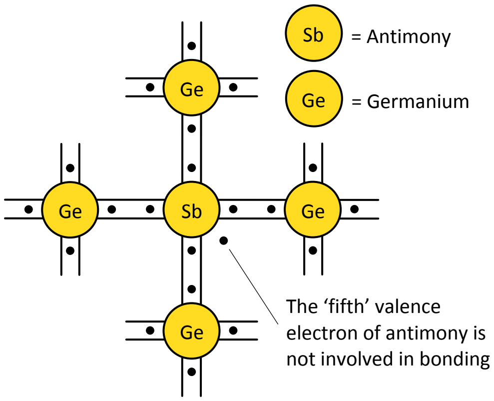
- The concentration of carriers in an *intrinsic semiconductor* is strongly dependent on the temperature. At low temperatures, the *valence band* is completely full, making the material an insulator. Increasing the temperature leads to an increase in the number of carriers and a corresponding increase in conductivity. This principle is used in **thermistors**. This behavior contrasts sharply with that of most metals, which tend to become less conductive at higher temperatures due to increased *phonon scattering*.

- An **extrinsic semiconductor** is a *semiconductor* that has been highly doped with impurities to modify the number and type of free charge carriers present.

- A *semiconductor* which is doped to such high levels that the dopant atoms are an appreciable fraction of the semiconductor atoms is called **degenerate**. A *degenerate semiconductor* acts more like a conductor than a semiconductor.

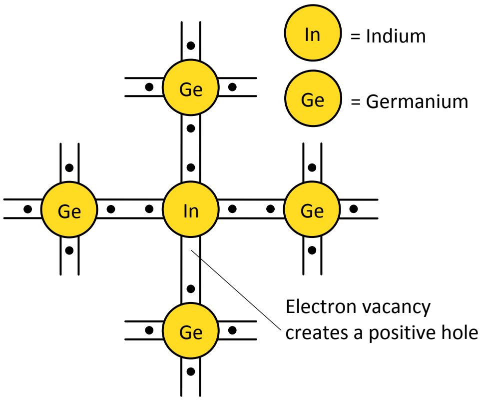
N–type doping

- The purpose of **n–type doping** is to produce an abundance of mobile or "carrier" **electrons** in the material. To help understand how *n-type doping* is accomplished, consider the case of silicon (Si) or germanium. Si atoms have four **valence electrons**, each of which is **covalently bonded** with one of four adjacent Si atoms. If an atom with five valence electrons, such as those from group 15 (a.k.a. group V) of the periodic table (e.g. phosphorus (P), arsenic (As), or antimony (Sb)), is incorporated into the crystal lattice in place of a Si atom, then that atom will have four covalent bonds and *one unbonded electron*. This extra electron is only weakly bound to the atom and can easily be excited into the *conduction band* (≈ 0.01 eV). At room temperatures, virtually all such electrons are excited into the *conduction band*. Since excitation of these weakly bound electrons does not result in the formation of a hole, the number of electrons in such a material far exceeds the number of thermally generated holes. In this case the *electrons* are the *majority carriers* ($\approx 10^{20}$ m^{-3}) and the *holes* are the *minority carriers*. Because the five-electron atoms have an extra electron to "donate", they are called **donor atoms**. Note that each movable electron within the *semiconductor* is never far from an immobile positive dopant ion, and the *n-doped* material normally has a net electric charge of zero.



P–type doping

- The purpose of **p–type doping** is to create an abundance of **holes**. In the case of silicon, a trivalent atom (such as boron) is substituted into the **crystal lattice**. The result is that one electron is missing from one of the four **covalent bonds** normal for the silicon lattice. Thus the dopant atom can accept an electron from a neighboring atoms' covalent bond to complete the fourth bond. Such dopants are called acceptors. The dopant atom accepts an electron, causing the loss of one bond from the neighboring atom and resulting in the formation of a "hole" (≈ 0.01 eV to release for conduction). Each hole is associated with a nearby negative-charged dopant ion, and the semiconductor remains electrically neutral as a whole. However, once each hole has wandered away into the lattice, one proton in the atom at the hole's location will be "exposed" and no longer cancelled by an electron. For this reason a hole behaves as a quantity of positive charge. When a sufficiently large number of acceptor atoms are added, the holes greatly outnumber the thermally-excited electrons. Thus, the *holes* are the *majority carriers* ($\approx 10^{20}\ \text{m}^{-3}$), while *electrons* are the *minority carriers* in *P-type* materials. Blue diamonds (Type IIb), which contain boron (B) impurities, are an example of a naturally occurring *P-type semiconductor*.



Worked Examples

Example 1

The semiconductors used in practical devices are made mainly from silicon or germanium that has been "*doped*" by the addition of a small amount of an impurity.

(a) Briefly describe what happens when a phosphorus atom is introduced into a silicon crystal.

(b) Explain what will happen if a boron atom is introduced into the silicon crystal instead.

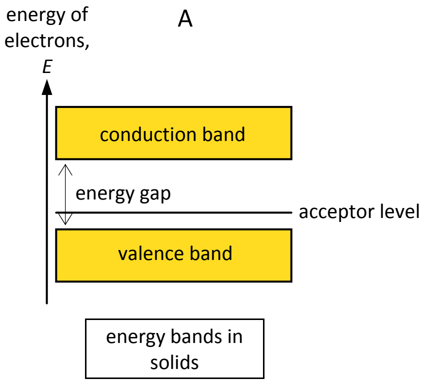
Solution:

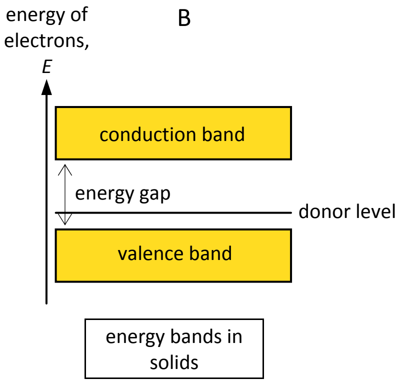
(a) Since phosphorus has five valence electrons while silicon has four, the phosphorous atom will shed one of the electrons from its outer shell in order to fit in with the silicon lattice. This electron is available to slide through the material, carrying current. Since the impurity adds some extra electrons that are free to move in the material, it is called n-type semiconductor because of the negative charge of the electrons. **(ans)**

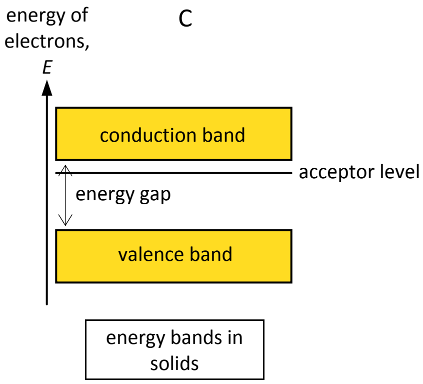
(b) Since boron has three valence electrons while silicon has four, the boron site will grab an electron out of the lattice to fit neatly into its atomic structure in the crystal, forming a current of effectively positive charge. Since the impurity removes some electrons, it creates "holes" in the crystal lattice: unoccupied places that could accommodate an electron. The holes are positively charged because an electron is missing from them. Holes can move through the crystal, since a nearby electron can jump into a hole, filling it up, but leaving a new hole nearby. Such materials are called p-type semiconductors, because it is positive charge that can “move” freely. **(ans)**

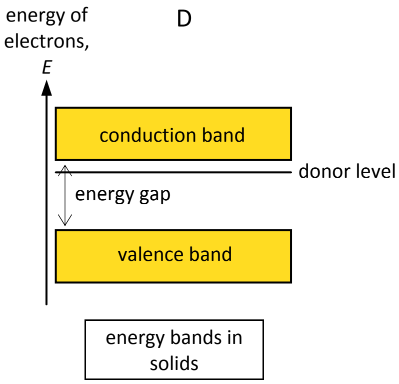
Example 2

Which diagram represents the energy levels in an n–type semiconductor material?









(D) (ans)

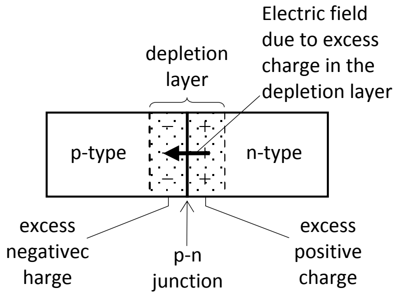
19.4 Depletion region of a p-n junction

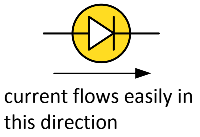
Discuss

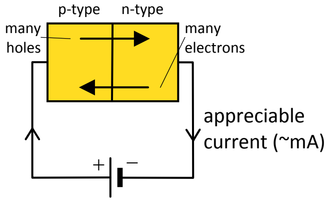
P–N junctions

- A **p-n junction** may be created by doping adjacent regions of a *semiconductor* with **p–type** and **n–type** dopants. If a **forward–** (**positive**) **biased** voltage is placed on the p–type side, the dominant positive carriers (holes) are pushed toward the junction. At the same time, the dominant negative carriers (electrons) in the n–type material are attracted toward the junction. Since there is an abundance of carriers at the junction, the junction behaves as a conductor, and the voltage placed across the junction produces a current. As the clouds of holes and electrons are forced to overlap, electrons fall into holes and become part of the population of immobile covalent bonds. However, if the bias polarity is reversed or **reverse–biased**, the holes and electrons are pulled away from the junction. Since only very few new electron/hole pairs are created at the junction, the existing mobile carriers are swept away to leave a **depletion region** (or *zone* or *layer*); a region of relatively non-conducting silicon. The reverse bias voltage will produce only a very low current across the junction. The p–n junction is the basis of an electronic device called a **diode** or **rectifier**, which allows electric charges to flow in only one direction. Similarly, a third semiconductor region can be doped n–type or p-type to form a three–terminal device, such as the **bipolar junction transistor** (which can be either p–n–p or n–p–n).

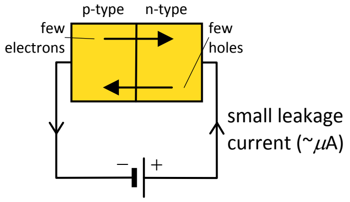
- When an *electron-hole pair* is created in the depletion region by ionizing radiation the two liberated charged particles will be swept out of the area. After the electron-hole pair is created in the depletion region, the hole will be swept towards the p-type region by the electric field, and the electron will be swept towards the n-type region by the electric field. The movement of these charge carriers constitutes a small **electrical current** which can be measured and analyzed.







forward-bias



reverse-bias

Worked Example

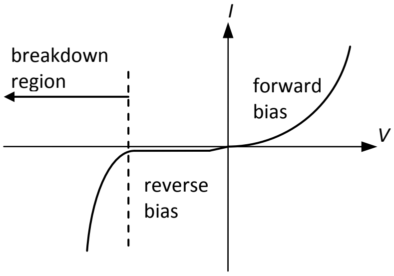
Example 1

(a) Describe the *IV* characteristics of a diode p-n junction.

(b) State two specialized versions of the diode.

Solution:

(a) The p-n junction is not an ideal rectifier diode having infinite resistance in the reverse direction and no resistance in the forward direction.



In the forward direction (forward biased) it can be seen that very little current flows until a certain voltage has been reached. This represents the work that is required to enable the charge carriers to cross the depletion layer. This voltage varies from one type of semiconductor to another.

From the diagram it can be seen that a small amount of current flows in the reverse direction (reverse biased). It has been exaggerated to show it on the diagram, and in normal circumstances it is very much smaller than the forward current. The reverse current results from minority carriers which are a very small number of electrons found in a P type region or holes in an N type region. The current through the device is very low for all reverse voltages up to a point where a process called reverse breakdown (*i.e.*, **Zener voltage**) occurs which causes the device to be damaged (*i.e.*, **Zener breakdown**) along with a large increase in current (*i.e.*, **avalanche breakdown**). **(ans)**

(b) Two versions are the light-emitting diode and the photodiode. **(ans)**

Worked Problem

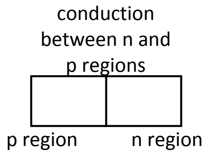
Example 1

The following diagrams A, B and C show the semiconductor diode p-n junctions under three situations.

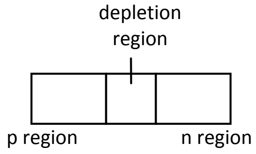
1. No bias applied
2. Forward bias applied
3. Reverse bias applied

Match the diagrams with each of the situations and briefly describe what happens in them.

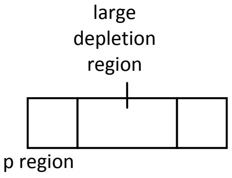
A:



B:



C:



Solution:

A p–n junction is formed from a piece of silicon by making one end p type and the other end n type. This means that both ends have different characteristics. One end has an excess of electrons whilst the other has an excess of holes.

Diagram B matches Situation 1 where no bias is applied. The two areas meet and the electrons fill the holes and there are no free holes or electrons. This means that there are no available charge carries in this region and it is known as the depletion region.

Diagram A matches Situation 2 where forward bias is applied. Even though the depletion region is very thin, current cannot flow in the normal way. It is dependent on the way in which the voltage is applied to the junction. If the voltage is applied such that the p region becomes positive and the N region becomes negative, holes are attracted towards the negative voltage and are assisted to jump across the depletion layer. Similarly electrons move towards the positive voltage and jump the depletion layer. Even though the holes and electrons are moving in opposite directions, they carry opposite charges and as a result they represent a current flow in the same direction.

Diagram C matches Situation 3 where reverse bias is applied. If the voltage is applied to the p-n junction in the opposite direction, no current flows. This is because the holes are attracted towards the negative potential that is applied to the p type region. Similarly the electrons are attracted towards the positive potential which is applied to the n type region. In other words the holes and electrons are attracted away from the junction itself and the depletion region increases in width. Accordingly no current flows. **(ans)**

Notes: