Optical Communication Devices

Optical sources 1: LASER

Contents

1 Introduction

- The optical source fundamental function is to convert electrical energy in the form of a current into optical energy (light) in an efficient manner which allows the light output to be effectively launched or coupled into the optical fiber.
- Three main types of optical light source are available. These are:
  1. wideband ‘continuous spectra’ sources (incandescent lamps)
  2. monochromatic incoherent sources (light-emitting diodes, LEDs)
  3. monochromatic coherent sources (lasers).
- The major requirements for an optical fiber emitter are:
  1. Compatible with launching light into an optical fiber (highly directional).
  2. Track the electrical input signal (linear).
  3. emit light at wavelengths where the fiber has low losses and low dispersion
  4. capable of simple signal modulation
  5. couple sufficient optical power to overcome attenuation in the fiber
  6. have a very narrow spectral bandwidth (linewidth)
  7. capable of maintaining a stable optical output
  8. the source is comparatively cheap and highly reliable

2 Basic Concepts

- The laser is a device which amplifies light, hence the derivation of the term LASER as an acronym for Light Amplification by Stimulated Emission of Radiation.
- The practical realization of the laser is as an optical oscillator.
- LASER provides an output of monochromatic, highly coherentradiation.

2.1 Absorption and Emission of Radiation

- The interaction of light with matter takes place in discrete packets of energy or quanta, called photons.
- Furthermore, the quantum theory suggests that atoms exist only in certain discrete energy states such that absorption and emission of light causes them to make a transition from one discrete energy state to another.

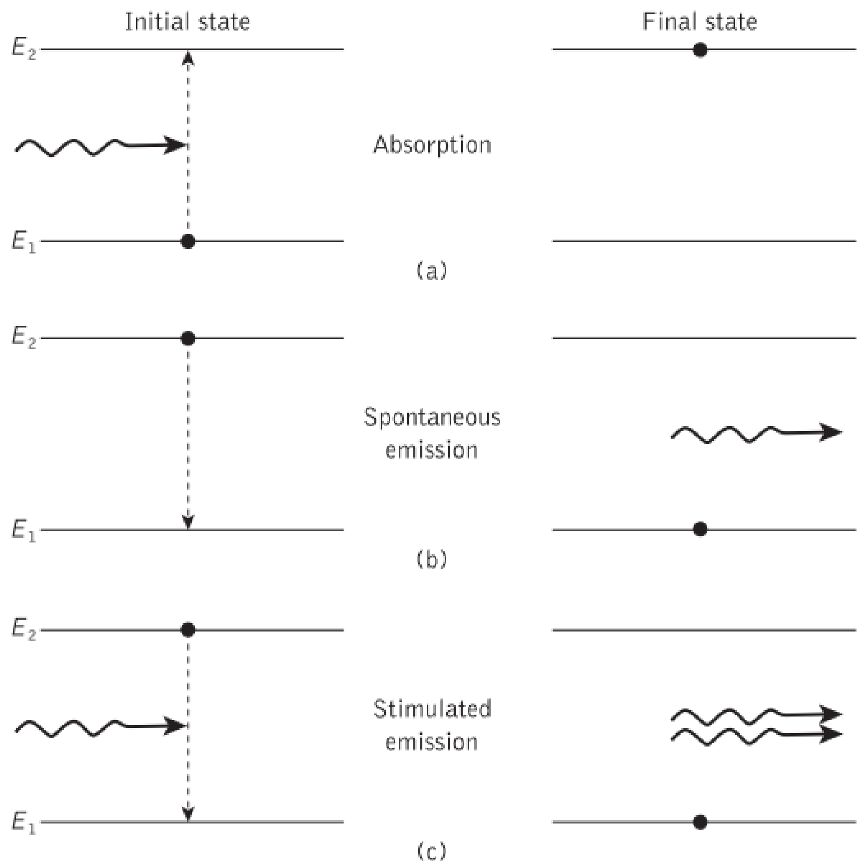

Figure 1: Energy state diagram showing: (a) absorption; (b) spontaneous emission; (c) stimulated emission. The black dot indicates the state of the atom before and after a transition takes place

- The frequency of the absorbed or emitted radiation f is related to the difference in energy E between the higher energy state E_2 and the lower energy state E_1 by:
$$E = E_2 - E_1 = hf \tag{1}$$
where $h = 6.62610^{-34}$ J.s is Planck's constant.
- Figure 1(a) illustrates a two energy state or level atomic system where an atom is initially in the lower energy state E_1.
  - When a photon with energy $E_2 - E_1$ is incident on the atom it may be excited into the higher energy state E2 through absorption of the photon.
  - This process is sometimes referred to as stimulated absorption.
- when the atom is initially in the higher energy state E_2 it can make a transition to the lower energy state E_1 providing the emission of a

photon at a frequency corresponding to Eq.1. This emission process can occur in two ways:

1. by spontaneous emission (random manner) (incoherent radiation) (LED). Figure 1(b).
2. by stimulated emission (a photon with $E = E_2 - E_1$ interacts with the atom in E_2 causing it to return to E_1 with the creation of a second photon with the same energy, freq, and polarization.) (coherent radiation) (LASER) Figure 1(c).

2.2 The Einstein Relations

- Einstein demonstrated that the rates of absorption, spontaneous emission and stimulated emission were related mathematically.
- Considering the atomic system to be in thermal equilibrium.

  (the rate of the upward transitions must equal the rate of the downward transitions.)

2.2.1 Absorbtion

The upward transition rate R_{12} (indicating an electron transition from level 1 to level 2) may be written as:

$$R_{12} = N_1 \rho_f B_{12} \tag{2}$$

where:

R_{12} is electron transition from E_1 to E_2.

N_1 is the density of atoms in energy level E_1.

ρ_f is the spectral density of the radiation energy at the transition frequency f.

B_{12} is the Einstein coefficient of absorption.

2.2.2 Sponteneous Emission

The sponteneous downward transition rate R_{21} (indicating an electron transition from level 2 to level 1) may be written as:

$$R_{21} = N_2 A_{21} \tag{3}$$

where:

R_{21} is electron transition from E_2 to E_1.

N_2 is the density of atoms in energy level E_2.

A_{21} is the Einstein coefficient of sponteneous emission.

NOTE: $A_{21} = 1/\tau_{21}$ where τ_{21} is the spontaneous lifetime.

2.2.3 Stimulated Emission

The stimulated downward transition rate R_{21} (indicating an electron transition from level 2 to level 1) may be written as:

$$R_{21} = N_2 \rho_f B_{21} \tag{4}$$

where:

R_{21} is electron transition from E_2 to E_1.

N_2 is the density of atoms in energy level E_2.

ρ_f is the spectral density of the radiation energy at the transition frequency f.

B_{21} is the Einstein coefficient of stumilated emission.

The total transition rate from level 2 to level 1 is given by:

$$R_{21} = N_2 A_{21} + N_2 \rho_f B_{21} \tag{5}$$

For a system in thermal equilibrium $R_{12} = R_{21}$, hence:

$$N_1 \rho_f B_{12} = N_2 A_{21} + N_2 \rho_f B_{21} \tag{6}$$

$$\rho_f = \frac{N_2 A_{21}}{N_1 B_{12} - N_2 B_{21}} \tag{7}$$

$$\rho_f = \frac{A_{21}/B_{12}}{(N_1 B_{12}/N_2 B_{21}) - 1} \tag{8}$$

The population of the two energy levels of such a system is described by Boltzmann statistics:

$$\frac{N_1}{N_2} = \frac{g_1 \exp\left(-E_1/KT\right)}{g_2 \exp\left(-E_2/KT\right)} = \frac{g_1}{g_2} \exp\left(E_2 - E_1/KT\right) = \frac{g_1}{g_2} \exp(hf/KT) \tag{9}$$

where:

N_1 and N_2 represent the density of atoms in energy levels E_1 and E_2.

g_1 and g_2 being the corresponding degeneracies of the levels.

(In many cases the atom has several sublevels of equal energy within an energy level which is then said to be degenerate)

K is Boltzmann's constant.

T is the absolute temperature.

Then by substituting Eq. 9 in Eq. 8:

$$\rho_f = \frac{A_{21}/B_{21}}{[(g_1 B_{12}/g_2 B_{21}) \exp(hf/KT)] - 1} \tag{10}$$

However, since the atomic system under consideration is in thermal equilibrium it produces a radiation density which is identical to black body radiation. Planck showed that the radiation spectral density for a black body is given by:

$$\rho_f = \frac{8\pi h f^3}{c^3}\left[\frac{1}{\exp(hf/KT)-1}\right] \tag{11}$$

Comparing Eq.11 with Eq.10 we obtain the Einstein relations:

$$B_{12} = \left(\frac{g_2}{g_1}\right) B_{21} \tag{12}$$

and:

$$\frac{A_{21}}{B_{21}} = \frac{8\pi h f^3}{c^3} \tag{13}$$

when the degeneracies of the two levels are equal ($g_1 = g_2$), then the probabilities of absorption and stimulated emission are equal. Then:

$$\frac{\texttt{Stimulated emission rate}}{\texttt{Spontaneous emission rate}} = \frac{B_{21}\rho_f}{A_{21}} = \frac{1}{\exp(hf/KT)-1} \tag{14}$$

Then, for systems in thermal equilibrium spontaneous emission is by far the dominant mechanism.

- NOTES:
  - From Eq. 9, as $E_2 > E_1$, then $\frac{N_1}{N_2} \uparrow\uparrow$ hence $N_1 >> N_2$
  - For stimulated emission to dominate over absorption and spontaneous emission in a two-level system, then we need $N_2 > N_1$ (population inversion)and radiation density $\uparrow\uparrow$ (optical feedback).

2.3 Population Inversion

- Under the conditions of thermal equilibrium, given by Eq. 9, the lower energy level E_1 of the two-level atomic system contains more atoms than the upper energy level E_2. This condition is illustrated in Figure **??**(a).
- To achieve optical amplification it is necessary to create a non-equilibrium distribution of atoms such that the population of the upper energy level is greater than that of the lower energy level (i.e. $N_2 > N_1$).
- This condition, which is known as population inversion, is illustrated in Figure 2(b).
- In order to achieve population inversion it is necessary to excite atoms into the upper energy level E_2 and hence obtain a non-equilibrium distribution. This process is achieved using an external energy source and is referred to as 'pumping'.

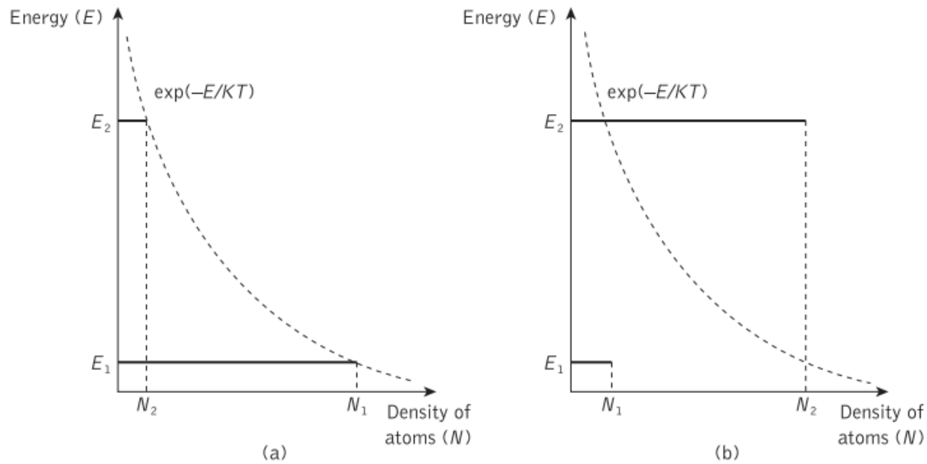

Figure 2: Populations in a two-energy-level system: (a) Boltzmann distribution for a system in thermal equilibrium; (b) a non-equilibrium distribution showing population inversion

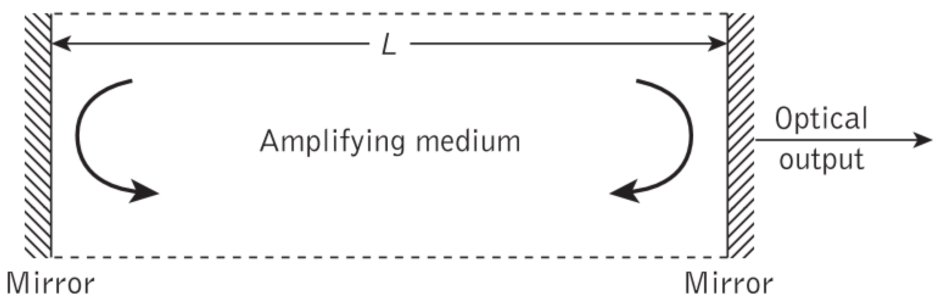

Figure 3: The basic laser structure incorporating plane mirrors

2.4 Optical Feedback

- Light amplification in the laser occurs when a photon colliding with an atom in the excited energy state causes the stimulated emission of a second photon and then both these photons release two more.
- Continuation of this process effectively creates avalanche multiplication, and when the electromagnetic waves associated with these photons are in phase, amplified coherent emission is obtained.
- This is accomplished by placing or forming mirrors (plane or curved) at either end of the amplifying medium, as illustrated in Figure 3.
- A stable output is obtained at saturation when the optical gain is exactly matched by the losses incurred in the amplifying medium.
- The major losses result from factors such as absorption and scattering

in the amplifying medium, absorption, scattering and diffraction at the mirrors and non-useful transmission through the mirrors.

- when sufficient population inversion exists in the amplifying medium the radiation builds up and becomes established as standing waves between the mirrors. These standing waves exist only at frequencies for which the distance between the mirrors is an integral number of half wavelengths.

  Thus when the optical spacing between the mirrors is L, the resonance condition along the axis of the cavity is given by:

$$L = \frac{\lambda q}{2n} \tag{15}$$

where:

λ is the emission wavelength.

n is the refractive index of the amplifying medium.

q is an integer.

then:

$$\lambda = \frac{2Ln}{q} \tag{16}$$

Alternatively, discrete emission frequencies f are defined by:

$$f = \frac{qc}{2Ln} \tag{17}$$

The different frequencies of oscillation within the laser cavity are determined by the various integer values of q and each constitutes a resonance or mode.

It may be observed that these modes are separated by a frequency interval δf where:

$$\delta f = \frac{c}{2nL} \tag{18}$$

The mode separation in terms of the free space wavelength, assuming $\delta f << f$:

$$\delta\lambda = \frac{-c}{f^2}\delta f \tag{19}$$

$$\delta\lambda = \frac{\lambda^2}{2nL} \tag{20}$$

A large number of modes may be generated within the laser cavity, the spectral output from the device is defined by the gain curve. Hence the laser emission will only include the longitudinal modes contained within the spectral width of the gain curve. This situation is illustrated in Figure 4.

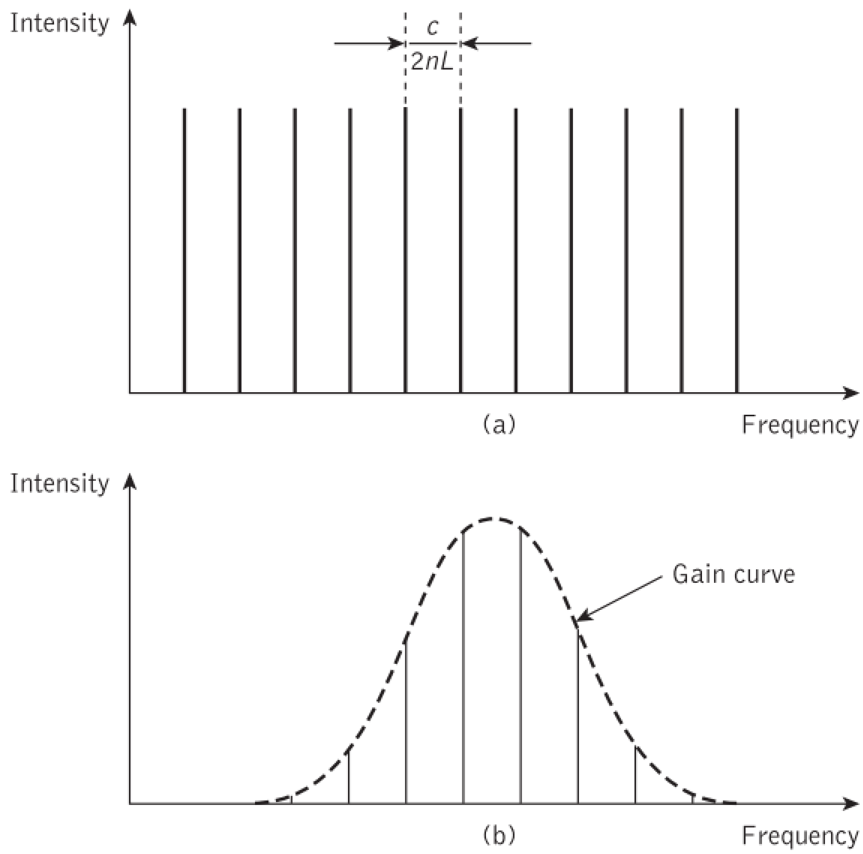

Figure 4: (a) The modes in the laser cavity. (b) The longitudinal modes in the laser output

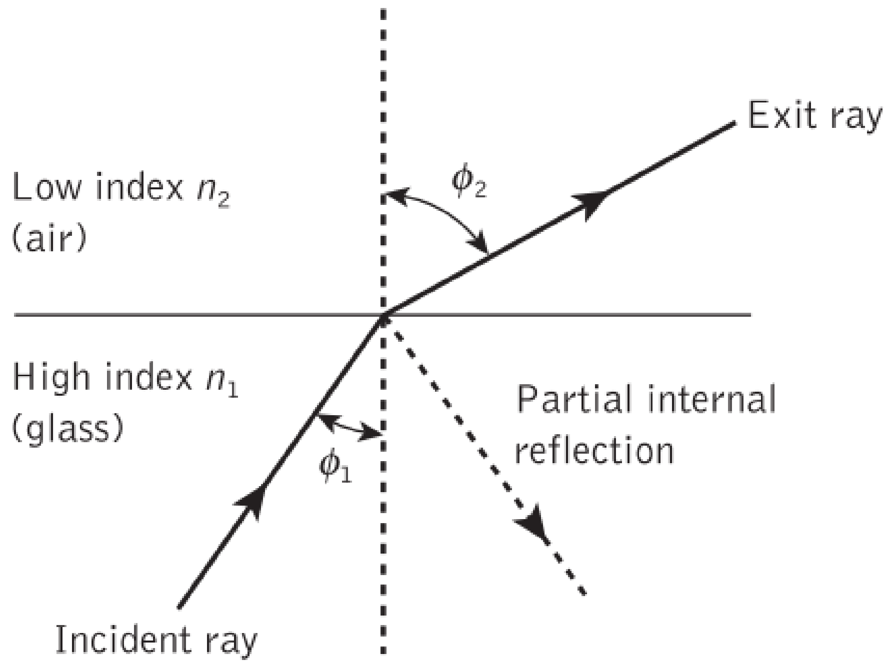

Figure 5: Light rays incident on a high to low refractive index interface

2.5 Reflectivity

For a light beam passes from one substance with refractive index n1 into another with refractive index n2, we define refraction coefficient as:

$$\Gamma = \frac{\texttt{reflected field}}{\texttt{incident field}} = \frac{n_1 - n_2}{n_1 + n_2} \tag{21}$$

therefore the reflectivity:

$$r = \Gamma^2 = \frac{P_{\texttt{reflected}}}{P_{\texttt{incident}}} = \frac{(n_1 - n_2)^2}{(n_1 + n_2)^2} \tag{22}$$

2.6 Threshold condition for laser oscillation

- A minimum or threshold gain within the amplifying medium must be attained such that laser oscillations are initiated and sustained.

$$\texttt{Fractional loss} = r_1 r_2 \exp(-2\bar{\alpha}L) \tag{23}$$

where: $\bar{\alpha}$ is loss coefficient per unit length cm^{-1}.

r_1 and r_2 are reflectivity of first and second mirror.

L is amplifying medium length.

The fractional round trip gain is given by:

$$\texttt{Fractional gain} = \exp(2\bar{g}L) \tag{24}$$

where:

$\bar{g}$ is the gain coefficient per unit length cm^{-1}.

Hence:

$$\exp(2\bar{g}L) \times r_1 r_2 \exp(-2\bar{\alpha}L) = 1 \tag{25}$$

$$r_1 r_2 \exp[2(\bar{g} - \bar{\alpha})L] = 1 \tag{26}$$

The threshold gain per unit length:

$$\bar{g} = \bar{\alpha} + \frac{1}{2L} \ln \frac{1}{r_1 r_2} \tag{27}$$

For laser action to be easily achieved it is clear that a high threshold gain per unit length is required in order to balance the losses from the cavity.

3 Optical Emission from Semiconductors

3.1 The p-n Junction

- A perfect semiconductor crystal containing no impurities or lattice defects is said to be intrinsic.

- The energy band structure of an intrinsic semiconductor is illustrated in Figure 6(a) which shows the valence and conduction bands separated by a bandgap E_g which varies for different semiconductor materials.

- At a temperature above absolute zero where thermal excitation raises some electrons from the valence band into the conduction band, leaving empty hole states in the valence band.

- For a semiconductor in thermal equilibrium the energy-level occupation is described by the Fermi–Dirac distribution function which describes energy-level occupation and the probability that an electron occupies a particular energy level.

$$P(E) = \frac{1}{1 + \exp(E - E_F)/KT} \tag{28}$$

  where:

  E_F is Fermi Level.

  (The Fermi level is only a mathematical parameter but it gives an indication of the distribution of carriers within the material).

  K is Boltzmann's constant.

- To create an extrinsic semiconductor the material is doped with impurity atoms which create either more free electrons (donor impurity) or holes (acceptor impurity). These two situations are shown in Figure 7.

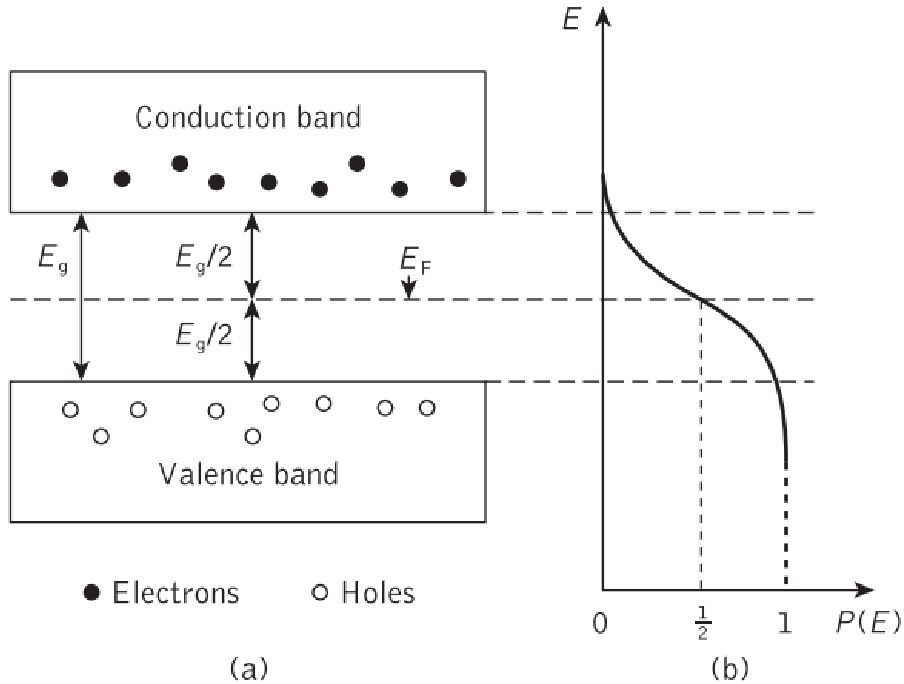

Figure 6: (a) The energy band structure of an intrinsic semiconductor at a temperature above absolute zero, showing an equal number of electrons and holes in the conduction band and the valence band respectively. (b) The Fermi–Dirac probability distribution corresponding to (a)

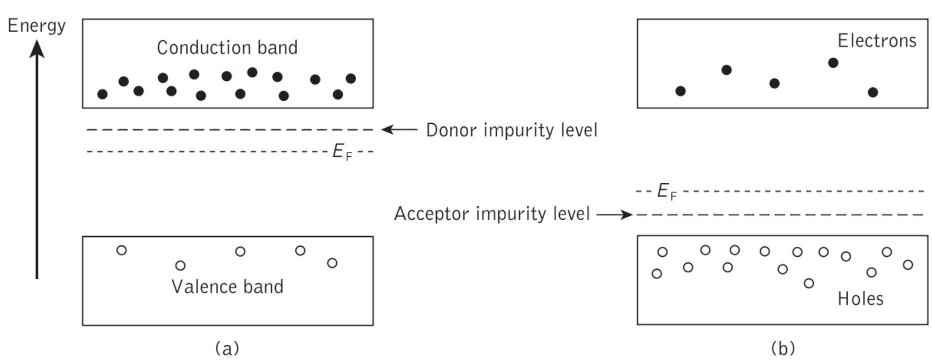

Figure 7: Energy band diagrams: (a) n-type semiconductor; (b) p-type semiconductor

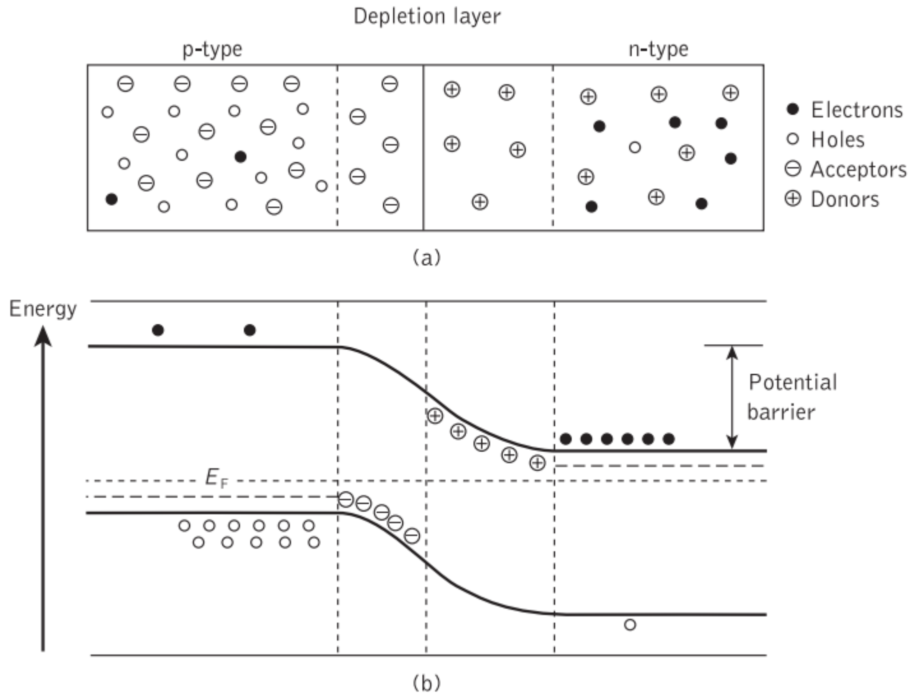

Figure 8: (a) The impurities and charge carriers at a p–n junction. (b) The energy band diagram corresponding to (a)

- The p–n junction diode is formed by creating adjoining p- and n-type semiconductor layers in a single crystal, as shown in Figure 8(a).
- A thin depletion region or layer is formed at the junction through carrier recombination which effectively leaves it free of mobile charge carriers.
- This establishes a potential barrier between the p- and n-type regions which restricts the interdiffusion of majority carriers from their respective regions, as illustrated in Figure 8(b).
- The width of the depletion region and thus the magnitude of the potential barrier is dependent upon the carrier concentrations (doping) in the p- and n-type regions and any external applied voltage.
- When an external positive voltage is applied to the p-type region with respect to the n-type, both the depletion region width and the resulting potential barrier are reduced and the diode is said to be forward biased.
- Electrons from the n-type region and holes from the p-type region can flow more readily across the junction into the opposite type region.
- This situation in suitable semiconductor materials allows carrier recombination with the emission of light.

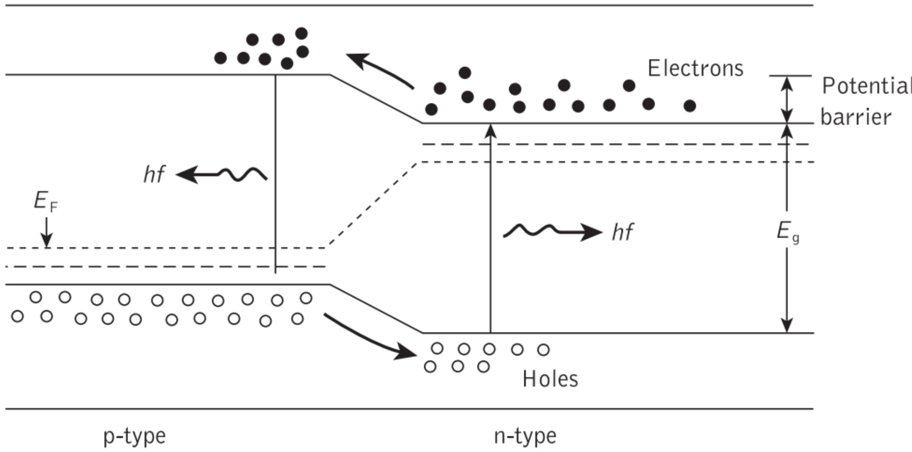

Figure 9: The p–n junction with forward bias giving spontaneous emission of photons

3.2 Spontaneous Emission

- The increased concentration of minority carriers in the opposite type region in the forward-biased p–n diode leads to the recombination of carriers across the bandgap. This process is shown in Figure 9 for direct bandgap.
- The energy released by this electron–hole recombination is approximately equal to the bandgap energy Eg.
- Excess carrier population is therefore decreased by recombination which may be radiative or nonradiative.
  - In nonradiative recombination the energy released is dissipated in the form of lattice vibrations and thus heat.
  - In radiative recombination the energy is released with the creation of a photon with a frequency following:

$$E_g = hf = \frac{hc}{\lambda} \tag{29}$$

$$\lambda = \frac{1.24}{E_g} \tag{30}$$

    where λ is written in μm and E_g in eV.

- This spontaneous emission of light from within the diode structure is known as electroluminesce.
- the amount of radiative, recombination and the emission area within the structure is dependent upon the semiconductor materials used and the fabrication of the device.

3.3 Carrier Recombination

3.3.1 Direct and Indirect Bandgap Semiconductors

In order to encourage electroluminescence it is necessary to select an appropriate semiconductor material.

1. Direct Bandgap

    - Direct bandgap semiconductors are the most useful materials for this purpose in which electrons and holes on either side of the forbidden energy gap have the same value of crystal momentum and thus direct recombination is possible.
    - It may be observed that the energy maximum of the valence band occurs at the same (or very nearly the same) value of electron crystal momentum* as the energy minimum of the conduction band.
    - Hence when electron–hole recombination occurs the momentum of the electron remains virtually constant and the energy released may be emitted as light.
    - This direct transition of an electron across the energy gap provides an efficient mechanism for photon emission and the minority carrier lifetime is short (10^{-8} to 10^{-10} s).
    - Example: GaAs

2. Indirect Bandgap

    - In indirect bandgap semiconductors, the maximum and minimum energies occur at different values of crystal momentum.
    - For electron–hole recombination to take place it is essential that the electron loses momentum such that it has a value of momentum corresponding to the maximum energy of the valence band.
    - The conservation of momentum requires the emission or absorption of a third particle, a phonon.
    - the recombination in indirect bandgap semiconductors is relatively slow (10^{-2} to 10^{-4} s). This is reflected by a much longer minority carrier lifetime.
    - Example: Si

The recombination coefficient is obtained from the measured absorption coefficient of the semiconductor, and for low injected minority carrier density relative to the majority carriers it is related approximately to the radiative minority carrier lifetime τ_r by:

$$\tau_r = [B_r(N + P)]^{-1} \tag{31}$$

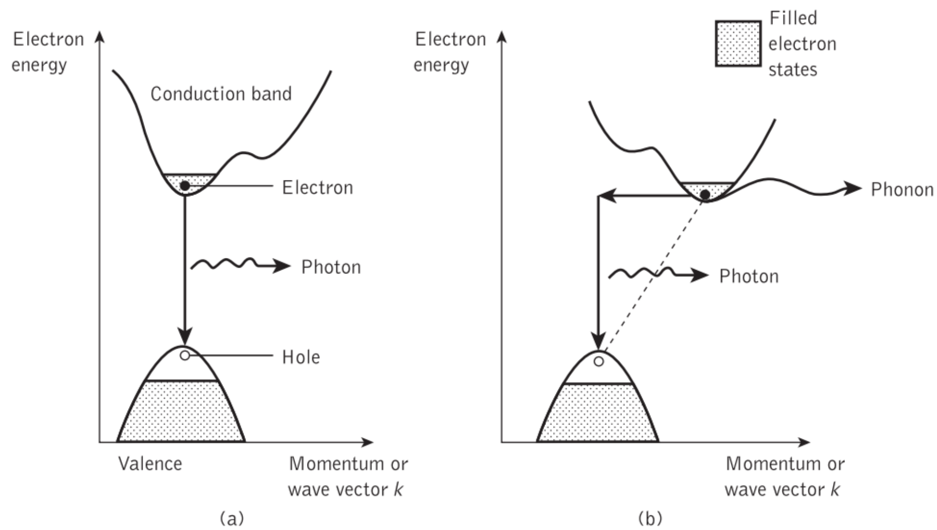

Figure 10: Energy–momentum diagrams showing the types of transition: (a) direct bandgap semiconductor; (b) indirect bandgap semiconductor

where:

N and P are the majority carrier concentrations in the n- and p-type regions.

B_r is the recombination coefficient.

Direct bandgap semiconductor devices in general have a much higher internal quantum efficiency. This is the ratio of the number of radiative recombinations (photons produced within the structure) to the number of injected carriers which is often expressed as a percentage.

3.3.2 Other Radiative Recombination Processes

- Only full bandgap transitions have been considered to give radiative recombination. However, energy levels may be introduced into the bandgap by impurities or lattice defects within the material structure which may greatly increase the electron–hole recombination (effectively reduce the carrier lifetime).

- The recombination process through such impurity or defect centers may be either radiative or nonradiative.

- Major radiative recombination processes at 300 K other than band-to-band transitions are shown in Figure **??**.

- Hence, an indirect bandgap semiconductor may be made into a more useful electroluminescent material by the addition of impurity centers which will effectively convert it into a direct bandgap material.

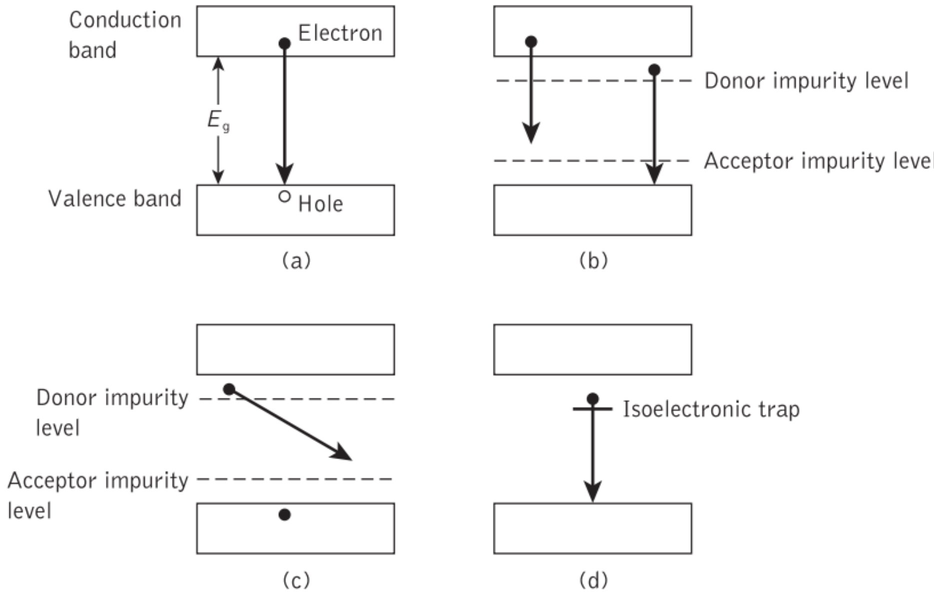

Figure 11: Major radiative recombination processes at 300 K: (a) conduction to valence band (band-to-band) transition: (b) conduction band to acceptor impurity, and donor impurity to valence band transition; (c) donor impurity to acceptor impurity transition; (d) recombination from an isoelectronic impurity to the valence band

3.4 Stimulated Emission and Lasing

- Carrier population inversion is achieved in an intrinsic (undoped) semiconductor by the injection of electrons into the conduction band of the material.

- Incident photons with energy E_g but less than the separation energy of the quasi-Fermi levels $E_q = E_{Fc} - E_{Fv}$ cannot be absorbed because the necessary conduction band states are occupied. However, these photons can induce a downward transition of an electron from the filled conduction band states into the empty valence band states, thus stimulating the emission of another photon.

- The basic condition for stimulated emission is therefore dependent on the quasi-Fermi level separation energy as well as the bandgap energy and may be defined as:

$$E_{Fc} - E_{Fv} > hf > E_g \tag{32}$$

- Population inversion may be obtained at a p–n junction by heavy doping (degenerative doping) of both the p- and n-type material.

  - Heavy p-type doping with acceptor impurities causes a lowering of the Fermi level or boundary between the filled and empty states into valence band

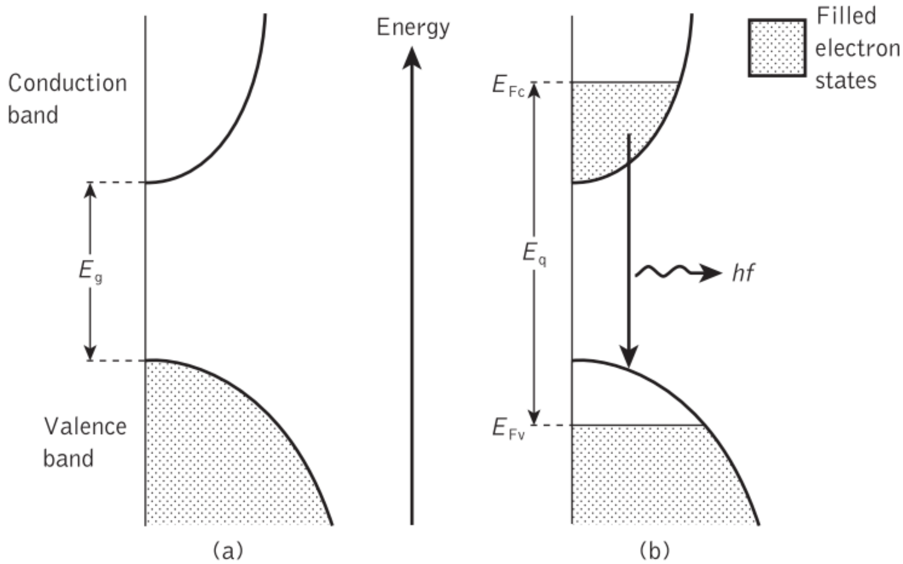

Figure 12: The filled electron states for an intrinsic direct bandgap semiconductor at absolute zero: (a) in equilibrium; (b) with high carrier injection

- Degenerative n-type doping causes the Fermi level to enter the conduction band of the material.

- In general, the degenerative doping distinguishes a p–n junction which provides stimulated emission from one which gives only spontaneous emission as in the case of the LED.

- Finally, it must be noted that high impurity concentration within a semiconductor causes differences in the energy bands in comparison with an intrinsic semiconductor. These differences are particularly apparent in the degeneratively doped p–n junctions used for semiconductor lasers.

3.4.1 Electron Density n and Photon Density ϕ Rate Equations

- The behavior of the semiconductor laser can be described by rate equations for electron and photon density in the active layer of the device.

  The two rate equations for electron density n, and photon density ϕ, are:

$$\frac{dn}{dt} = \frac{J}{ed} - \frac{n}{\tau_{sp}} - Cn\phi \tag{33}$$

  where:

  $\frac{J}{ed}$ is the increase in the electron concentration in the conduction band as the current flows into the junction diode.

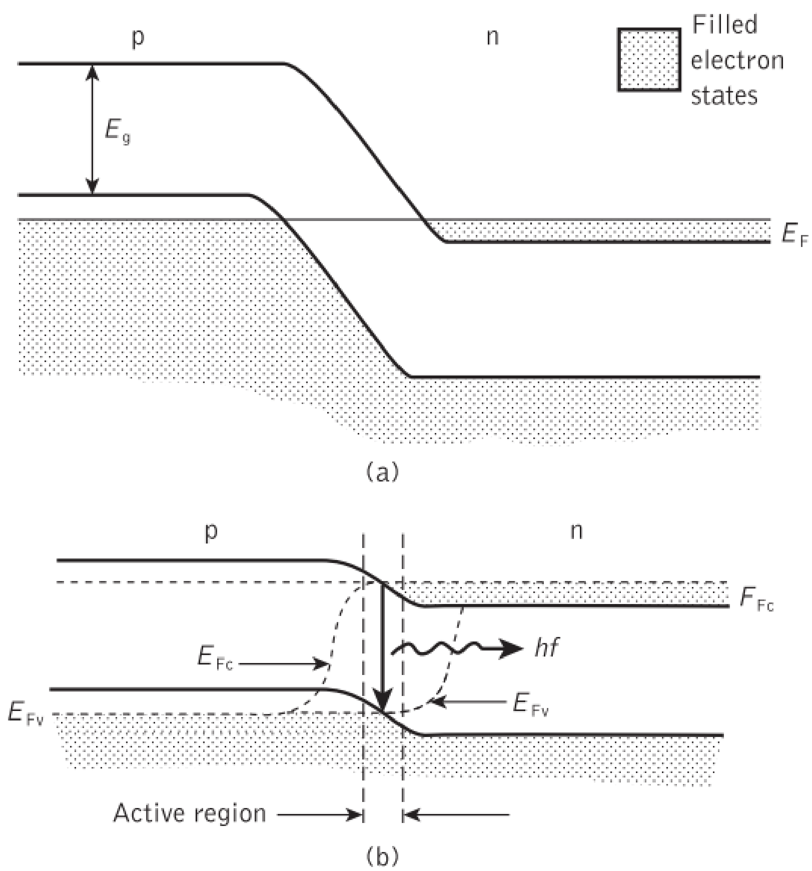

Figure 13: The degenerate p–n junction: (a) with no applied bias; (b) with strong forward bias such that the separation of the quasi-Fermi levels is higher than the electron–hole recombination energy hf in the narrow active region. Hence stimulated emission is obtained in this region

$\frac{n}{\tau_{sp}}$ is The electrons lost from the conduction band by spontaneous transitions.

$Cn\phi$ is The electrons lost from the conduction band by stimulated transitions.

J is the current density A/m^2

e is the charge on an electron

d is the thickness of the recombination region.

τ_{sp} is the spontaneous emission lifetime.

C is a coefficient which incorporates the B coefficients (Einistien coefficient).

ϕ is the photon density.

n is the electron density.

$$\frac{d\phi}{dt} = Cn\phi + \delta\frac{n}{\tau_{sp}} - \frac{\tau}{\phi_{ph}} \tag{34}$$

where:

$Cn\phi$ is the stimulated emission as a source of photons.

$\delta\frac{n}{\tau_{sp}}$ is the fraction of photons produced by spontaneous emission which combine to the energy in the lasing mode (often neglected as $\delta <<$)

$\frac{\tau}{\phi_{ph}}$ is the decay in the number of photons resulting from losses in the optical cavity.

δ is a small fractional value.

τ_{ph} is the photon lifetime.

τ_{sp} is the spontaneous emission lifetime.

C is a coefficient which incorporates the B coefficients (Einistien coefficient).

ϕ is the photon density.

n is the electron density.

- Although these rate equations may be used to study both the transient and steady-state behavior of the semiconductor laser, we are particularly concerned with the steady-state solutions.
- The steady state is characterized by:
  * The left hand side of Eqs 33 and 34 being equal to zero, when n and ϕ have nonzero values.
  * the fields in the optical cavity which are represented by ϕ must build up from small initial values, and hence $\frac{d\phi}{dt}$ must be positive when ϕ is small.

Therefore, setting δ equal to zero in Eq. 34, it is clear that for any value of ϕ, $\frac{d\phi}{dt}$ will only be positive when:

$$Cn - \frac{1}{\tau_{ph}} \geq 0 \tag{35}$$