Optical Communication

Notes Part C: Sources and Detectors

7. Optical Sources

A serious limitation of optical communication is that tunable, narrow-spectrum sources are very difficult to achieve, compared to radio-frequency sources. The most straightforward way to generate light is to simply heat a body, which causes it to radiate over a broad spectral range, with a peak determined by the temperature. However, the spectrum is generally too broad for efficient communication, and not easy to modulate or tune. This is also more-or-less true for light generated by combustion.

While some systems will provide narrow spectral output due to particular electronic transitions, such as gas discharge tubes (neon lights etc), most are impractical for communications due to complexity (size/cost/reliability/efficiency). The most practical is the band-gap transition in semiconductors, which can also be easily modulated (in intensity).

When an electron falls from the conduction band to the valence band of a semiconductor, it needs to emit the band-gap energy E_g, in the form of light or heat. If we pass current through a forward-biased p-n junction, we effectively pump holes and electrons in from opposite sides, and they recombine. If most of these recombinations are radiative (photon emitting) rather than non-radiative (phonon emitting), we have a reasonably efficient light source - the light emitting diode (LED). The wavelength of this light is determined by $h\nu = E_g$. We can use the convenient relation:

$$\lambda \text{ emitted} \cong \frac{1.24}{E_g} \ \mu\text{m eV}$$

We can characterise the radiative and non-radiative processes by statistical lifetimes τ_{rr} and τ_{nr}, the mean time for a hole-electron pairs to decay by each process. The internal efficiency of the LED (the fraction of hole-electron pairs which generate photons) is then given by

$$\eta_{int} = \frac{1}{1 + \tau_{rr} / \tau_{nr}}$$

There is also a finite spectral width to the radiation due to the width of filled valence hole and conduction electron states on each side of the bandgap. This width is mainly determined by temperature, and is usually in the range ΔE = 1.5 to 3.5 kT. For example, taking ΔE = 2kT, we have, for the following λ values: (at room temperature)

λ	$\Delta\lambda$
850 nm	30nm
1300 nm	70 nm
1550 nm	100nm

In each case the fractional spread in wavelengths is between 1% and 10% - this is <u>many</u> orders of magnitude higher than that associated with a typical signal bandwidth: for example, for a 1MHz signal at λ = 1300 nm, the minimum spectral width would be about 6 x 10^{-6} nm!

The most important factor determining η_{int} is whether the semiconductor is direct or indirect bandgap. In indirect bandgap materials, the minimum energy transition from the conduction to the valence band is associated with a significant change in momentum. For a given energy, photons have negligible momentum compared to phonons, and so a single photon cannot simultaneously satisfy conservation of energy and momentum for such a transition. Therefore, non-radiative recombination dominates in indirect bandgap materials such as Si and Ge, and consequently these are not useful as sources.

Once the photons are generated, they must escape from the semiconductor to be of use. Four factors contribute to this external quantum efficiency. Take for example a small emitting strip below a planar semiconductor/air boundary. The four factors limiting the escape of photons into the air are:

i) half the light goes downward;
ii) much of the light arrives at an angle > the critical angle $\theta_c = \sin^{-1} n_a/n_s$, and thus is totally internally reflected;
iii) photons at angles < θ_c undergo some (Fresnel) reflection anyway.
iv) some light is reabsorbed during its passage through the semiconductor.

(An example of the relevant calculation is found in problem 4.)

The response speed of the LED is primarily determined by the recombination lifetime. At low injection currents, this is approximately inversely proportional to dopant concentration, but excessive doping reduces the internal quantum efficiency. In GaAs a practical limit on acceptor concentration is ~ $N_a = 10^{24}$ m^{-3}, which gives a radiative lifetime $\tau_{rr} \cong 10$ n_{sec}, and consequently a modulation bandwidth of about 100 MHz.

External efficiency can be improved by "blooming" the semiconductor surface, ie. by using an antireflection coating, which ideally is $\lambda/4$ thick and has an index which is the geometric mean of the semiconductor and external medium indices. This reduces the Fresnel reflections; total internal reflection can be reduced by "lensing" the surface, or by some other surface shape modification. This is used in high efficiency LEDs for lighting. However, for fibre launching, surface shape modification does not necessarily help.

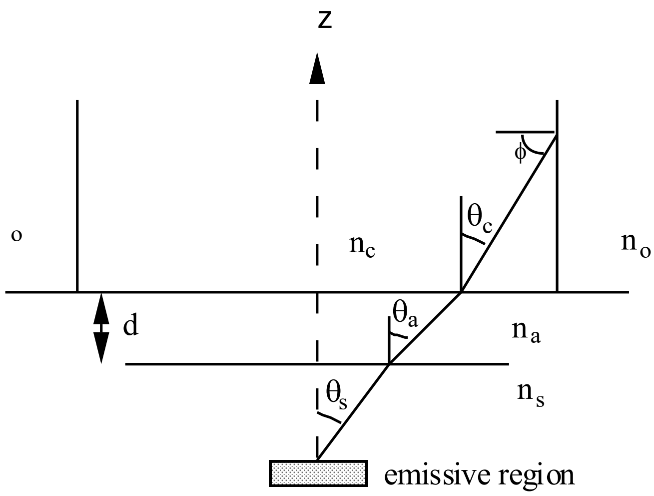


Figure 7.2 Geometry of light launching into a step-index fibre.

For launching into fibres, the rays must enter the core at an angle that will subsequently be guided. If we consider launching from air, then we have:
$n_a \sin\theta_a = n_c \sin\theta_c$

We need $\phi > \sin^{-1} n_o/n_c$ for guiding.
$\sin \phi = \cos\theta_c > n_o/n_c \qquad \therefore \; 1 - \sin^2\theta_c > (n_o/n_c)^2$

$$\therefore \; \sin^2\theta_a < \left(\frac{n_c}{n_a}\right)^2 (1-(n_o/n_c)^2) \qquad = \frac{n_c^2 - n_o^2}{n_a^2}$$

For air, $n_a = 1$, giving $\sin\theta_a < (n_c^2 - n_o^2)^{½}$

We call this quantity $(n_c^2 - n_o^2)^{½}$ the <u>numerical aperture</u> of the fibre; like the N.A. of a microscope or camera lens, it is a measurement of its light gathering ability. For $n_c - n_o \ll 1$, which is usually the case:

$$\text{N.A.} \cong (2n_o\Delta n)^{1/2}$$

the N.A. will have a very low value, and will cause a substantially more severe restriction on the useful emission directions if an LED is used for launching then total internal reflection at the semiconductor surface.

A much improved source is obtained by combining the p-n junction with a laser cavity to form a semiconductor laser diode. Lasers operate by **l**ight **a**mplification by **s**timulated **e**mission of **r**adiation. An electron hole pair will eventually recombine according to the statistical lifetimes as described above; in the case of radiative recombination this is called spontaneous emission. However, if a photon within the correct energy range passes close enough to interact with the pair, it can stimulate the recombination to take place immediately. The original photon is not absorbed or altered by this process, and the new photon resulting from the stimulated recombination has exactly the same direction, phase, energy and polarisation as the original.

We do not need to do anything special for stimulated emission to occur. The relative number of stimulated and spontaneous events simply depends on the local photon density. However, in order to increase this density so that stimulated emission dominates, we introduce a resonant cavity. This simply consists of two parallel mirrors between which we place the active material, in this case the p-n junction. Photons travelling perpendicular to the mirrors reflect back and forth in the cavity, generating more and more photons in this same direction. Eventually the density of photons becomes high enough that most injected hole-electron pairs will generate a photon in this direction.

By making one (or both) of the mirrors partially transmitting, we produce an output, The advantage for fibre launching is plain; since the output photons all travel in the same direction, they can easily be launched within the N.A. of the fibre. In practice there is some spread in the output direction because of diffraction, since the cavity cross section is finite. In fact, since the laser diode actually consists of a waveguide, rather than a bulk medium, between mirrored end facets, the output lies entirely within the fibre N.A. as long as the N.A. of the laser diode is lower. (We also require that the laser core be smaller than the fibre core to maximise coupling.)

If we examine the output power as a function of input current, we find that initially the output rises slowly until a *threshold current* is reached. This is the point at which the injected carrier density is sufficient to provide net round-trip gain, i.e. enough stimulated emission to compensate for both re-absorption and loss at the mirrors. After this point the output power rises sharply, and the slope of this line is known as the *slope efficiency*. It gives the differential optical power per unit input current, and is given by :

$$\frac{d\Phi}{dI} = \frac{h\nu}{e}\eta$$

i.e. the photon energy over the unit charge, times the quantum efficiency η. This latter quantity is simply the fraction of generated photons that leave the cavity (the external efficiency η_e), times the internal efficiency η_i: the fraction of injected carriers resulting in useful photons. As the photons make the round trip in the cavity, they are subject to absorption and scattering loss, to which we assign an absorption coefficient α_s, such that $\Phi(z) = \Phi(0)\exp(-\alpha_s z)$. For a cavity length L, we have a total reduced power $\Phi(2L) = \Phi(0)\exp(-2\alpha_s L)$. However, we also have a reduced power due to the mirrors; if these have reflectivity R_1 and R_2, then $\Phi(2L) = \Phi(0) R_1 R_2 \exp(-2\alpha_s L)$. To compare more easily the two effects, we introduce an equivalent absorption coefficient for the mirrors α_m, such that $R_1R_2 = \exp(-2\alpha_m L)$, giving:

$$\alpha_m = \frac{1}{2L}\ln\left(\frac{1}{R_1R_2}\right)$$

We can now obtain an expression for the quantum efficiency:

$$\eta = \eta_i \frac{\alpha_m}{\alpha_m + \alpha_s} = \eta_i \frac{\ln\left(1/R_1R_2\right)}{\ln\left(1/R_1R_2\right) + 2\alpha_s L}$$

Taking an example of $\alpha_s = 1$ mm^{-1}, L = 400 μm, $\eta_i = 0.8$, $R_1 = 1$ and $R_2 = 0.33$, we obtain $\eta = 0.465$, giving a slope efficiency at $\lambda = 1.3$ μm of 0.44 W/A.

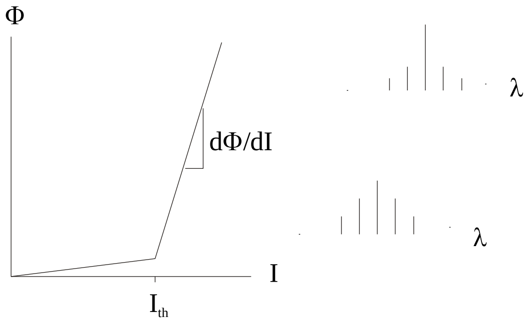


Figure 7.3 A typical laser diode output characteristic, showing the threshold current I_{th}, slope efficiency $d\Phi/dI$, and output spectra for low and high intensities.

The laser diode also offers much increased modulation speed compared to the LED. In the latter case, response is limited by the recombination lifetime, as we have seen. For the laser, however, when operating above threshold, stimulated emission can much more quickly alter the photon flux in response to changes in the input current. In order to take full advantage of this enhanced speed, modulation schemes should maintain the lasing condition, so that, for example, digital on-off keying will usually use an 'off' level just above threshold. Detailed analysis of the temporal response of the laser diode is beyond our scope, but we can note that speeds in excess of 1 Gbit/s can be obtained.

The existence of a lasing threshold has another important practical implication; if the input signal is analogue, careful biasing is needed to keep the output in its linear region. This is further complicated by the temperature variation of the threshold current, and the ease of damaging the output facets if excessive output powers are reached. For this reason, laser diodes will normally have active temperature stabilisation, e.g. using a Peltier effect device. For analogue applications, particularly cable television distribution, minimal threshold current values are sought.

Below threshold, the photons generated have the same spectral distribution as for an LED, as the device is effectively operating as one. But as lasing begins to be established, an additional spectral limitation applies: the round-trip phase change must be a whole multiple of cycles, so that successive passes add constructively. This is equivalent to the oscillations on a guitar string consisting of integer numbers of half wavelengths fitting on the string's length. Thus, within the output envelope there are a series of spectral lines, which we call longitudinal modes. Their spacing is determined by the condition above, i.e. $L = m\lambda_o/2n'$, (where λ_o is the free-space wavelength, n' the effective index of the laser waveguide, and m an integer), and so decreases with increasing cavity length. As the photon flux in the cavity grows with increasing current, the stronger central modes consume proportionally more carriers, and so they grow more quickly than the outlying ones, making the overall spectrum narrower.

The width of the individual spectral lines is mainly determined by the dependence of refractive index on carrier concentration, which makes the effective index vary and thus the relation between length and wavelength that determines the longitudinal mode frequencies. Associated with this effect is the fact that modulation by varying the injected current also causes frequency modulation, or *chirp*. This can be avoided by using external modulators, from which high modulation speeds can also be obtained.

As the dispersion is directly proportional to laser linewidth (with single mode fibre), narrow output spectra are very important. An important way to obtain these is by the use of single (longitudinal) mode lasers. These are based on replacing the facet mirrors with gratings: corrugations in the waveguide surface that have a only reflect particular wavelengths. When a light beam encounters a grating, it can be scattered (diffracted) in such directions that the wave vector **k** of the scattered light is equal to that of the incident light plus or minus that of the grating (1st order) or some multiple of it (higher order diffraction). Consider a grating having a period Λ, etched on the laser waveguide. To get reflection of a mode having propagation constant β, through first order diffraction, we require $2\pi/\Lambda = 2\beta = 4\pi\, n'/\lambda_o$, so that the scattered signal has the opposite propagation constant $-\beta$. Thus we choose $\Lambda = \lambda_o/2n'$. Two variants of such lasers exist: distributed feedback (DFB) lasers, where the grating runs along the whole waveguide, and distributed Bragg reflector (DBR) lasers, where the central gain section is separated from two grating reflectors. In such devices, linewidths below 100 MHz (sometimes far below!) are routinely achieved.

8. Photodetection

While the recombination of a hole-electron pairs across a semiconductor band gap is a suitable light generation mechanism for optical communications, the reverse process is highly suitable for detection, namely the absorption of a photon by the excitation of a hole-electron pair.

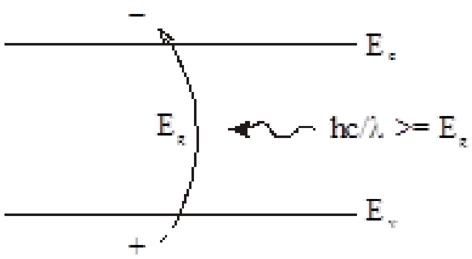


Figure 8.1 Inter-band excitation by a photon.

For this process to happen, we simply need $h\nu > E_g$. We do <u>not</u> need a direct bandgap material, so silicon is highly suitable. However, for wavelengths longer than E_g/hc, semiconductors are quite transparent; for Si, this cut-off is at about 1.1 μm, which unfortunately makes it unusable for the two main fibre communications wavelengths.

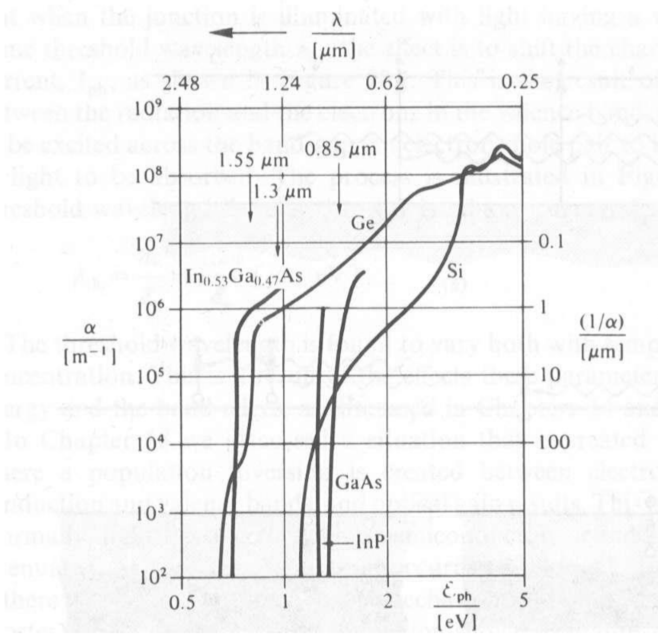


Figure 8.2 Absorption coefficient vs photon energy for various semiconductors (from Gowar).

The simplest way to detect light is to place a slab of semiconductor into a circuit and shine the light on it : as the number of electron-hole pairs goes up, the conductivity goes up, and so does the current in the external circuit. For a steady light flux Φ, the number of e-h pairs increases until their rate of recombination equals the rate of generation. Therefore a small τ gives a high response speed but low sensitivity while a high τ gives the reverse.

Unfortunately, due to the presence of thermally generated carriers, and thus noise, the basic photoconductor described above is unsuitable for communications. A much better solution is to use a reverse-biased p-n junction. Here there is a depletion layer at the junction where the carriers have been pulled away by the external field, so there is virtually no current in the external circuit. If an e-h pair is generated by absorption of a photon in this depletion layer, a current will flow in the circuit as these photo-carriers are pulled to opposite sides of the junction, at which time they will recombine and the current will stop.

The speed of the device is now mainly determined by the carrier transit across the depletion layer, so to make this fast, we need high field intensities throughout this layer, although not so high that dielectric breakdown occurs. Values between 10^6 and 10^7 V/m are sufficient generally to achieve the maximum (saturation) drift velocity without breakdown.

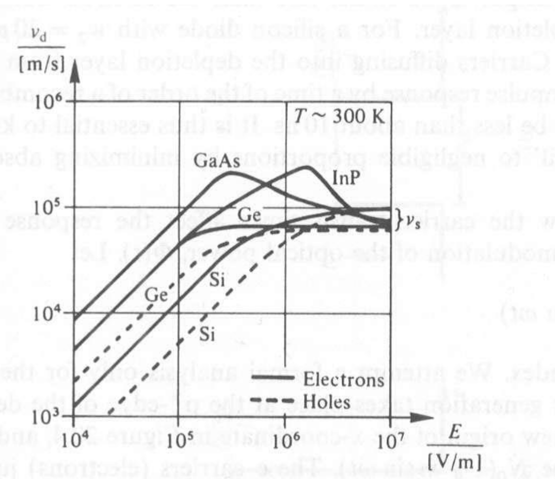
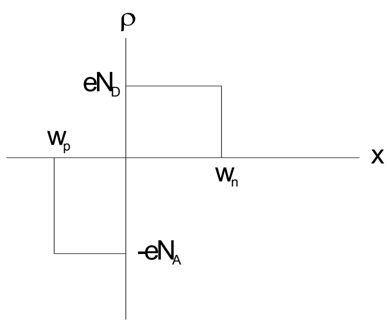


Figure 8.3 Variation of drift velocity with electric field strength (from Gowar).

To analyse the fields in the p-n photodetector, we assume that the carriers are completely depleted from a finite depletion width in each material, and that the charge concentration is therefore given by the dopant level in each. Thus we have :



ρ = charge density

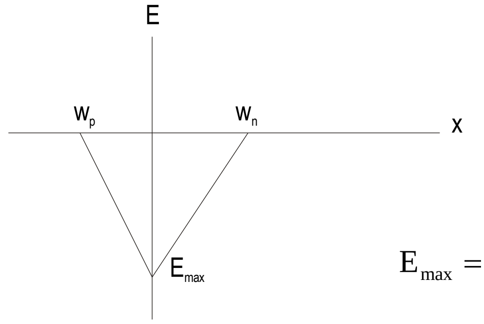
N_A, N_D are acceptor and donor dopant concentrations

Figure 8.4 Charge density distribution in reverse-biased p-n junction.

Charge conservation gives $Q_p = - Q_n$, or:

$$w_p N_A = w_n N_D$$

The E field is found by $dE/dx = \rho/\varepsilon$, and since E is zero outside the depletion layer, where the material is conducting, the field distribution will be as shown below.



$$E_{max} = -\frac{e}{\varepsilon} w_p N_A$$

Figure 8.5 Electric field distribution in reverse-biased p-n junction.

The voltage is found by $dV/dx = -E$. Absolute values of V have no significance, but the total voltage drop will be given by the area under the E curve:

$$\Delta V = E_{max} \frac{(w_p + w_n)}{2}$$

From this we can determine w_p & w_n for any reverse bias potential.

In practice, we need E to be, as much as possible, high throughout the depletion layer. The way to accomplish this is by the p-i-n structure, where a weakly doped (N_D^-) n layer is placed between the highly doped p & n materials - this is called the intrinsic layer. When this is depleted, the field drops slowly across it, and then suddenly falls to zero in the n^+ region. This allows the E field to be much more uniform, and also makes the depletion layer width, and thus the junction capacitance, much less dependent on the applied voltage.

As the wavelength decreases, the responsivity (amps of photocurrent per watt of optical power) also decreases, since the increasing photon energy doesn't increase I_{ph}. Photodetectors also have a quantum efficiency, as not all the photons are absorbed in the depletion region. Some are reflected from the surface (again, blooming can greatly reduce this), some are absorbed in the p-layer and some carry on through to the n^+ layer.

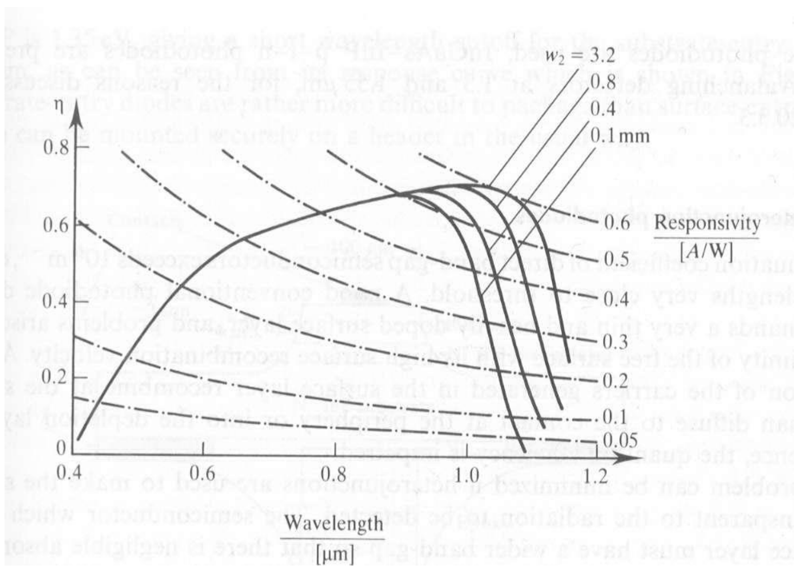


Figure 8.6 Quantum efficiency and responsivity vs. wavelength for p-i-n photodiodes of various i-layer thicknesses

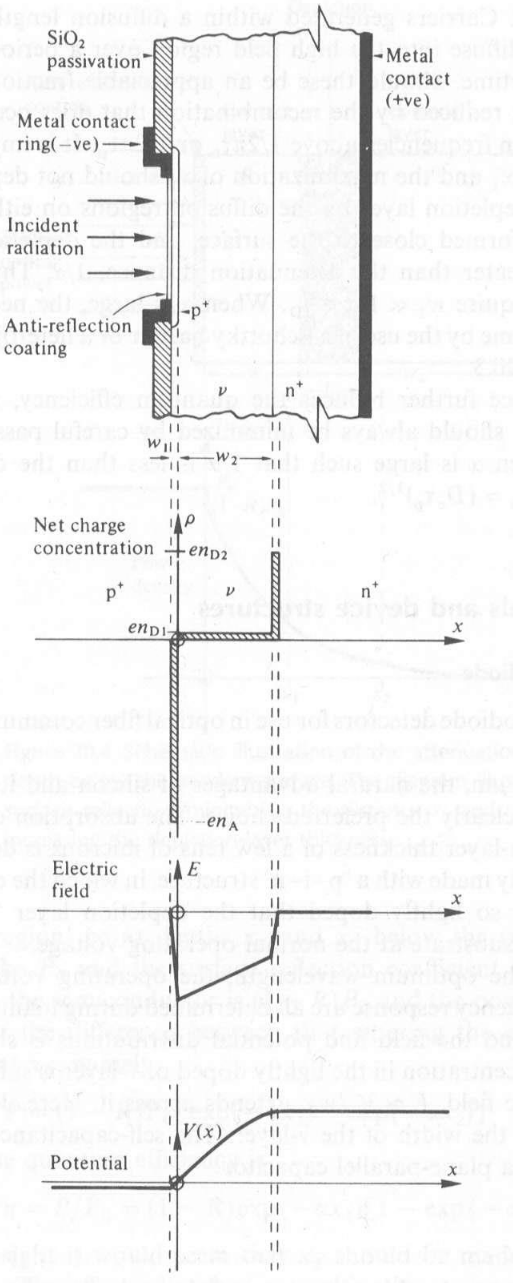


Figure 8.7 Structure, charge, field and voltage distributions in a p-i-n photoconductor (from Gowar)

The absorption coefficient α will be constant throughout the material (not very dependent on doping), so that:

$$\eta = (1-R)\left(e^{-\alpha X_1} - e^{-\alpha X_2}\right)$$

$$R = \left(\frac{n_s - n_o}{n_s + n_o}\right)^2$$

where X_1 and X_2 are the distances from the surface to the p-i and i-n boundary respectively.

One interesting thing about the response of the photodiode is that there is not a power-to-power relationship between the optical and electrical signals: rather, each photon produces a single carrier, and therefore the photocurrent generated is proportional to the number of photons per second. Since the photon flux is Φ and the photon energy is hν, this rate is Φ/hν, and we get:

$$\frac{I_{ph}}{\Phi} = \eta \frac{e\lambda}{hc} \equiv R, \text{ the responsivity}$$

The number of e-h pairs per photon can be greater than one if the field is high enough that the accelerated carriers ionise further carriers by collision. A device using this effect is called an avalanche photodiode, or APD. The photocurrent is multiplied by a factor M, but the shot noise is increased by a further factor F, so this is only useful when the thermal and amplifier noise are much greater than shot noise, i.e. for weak signals.

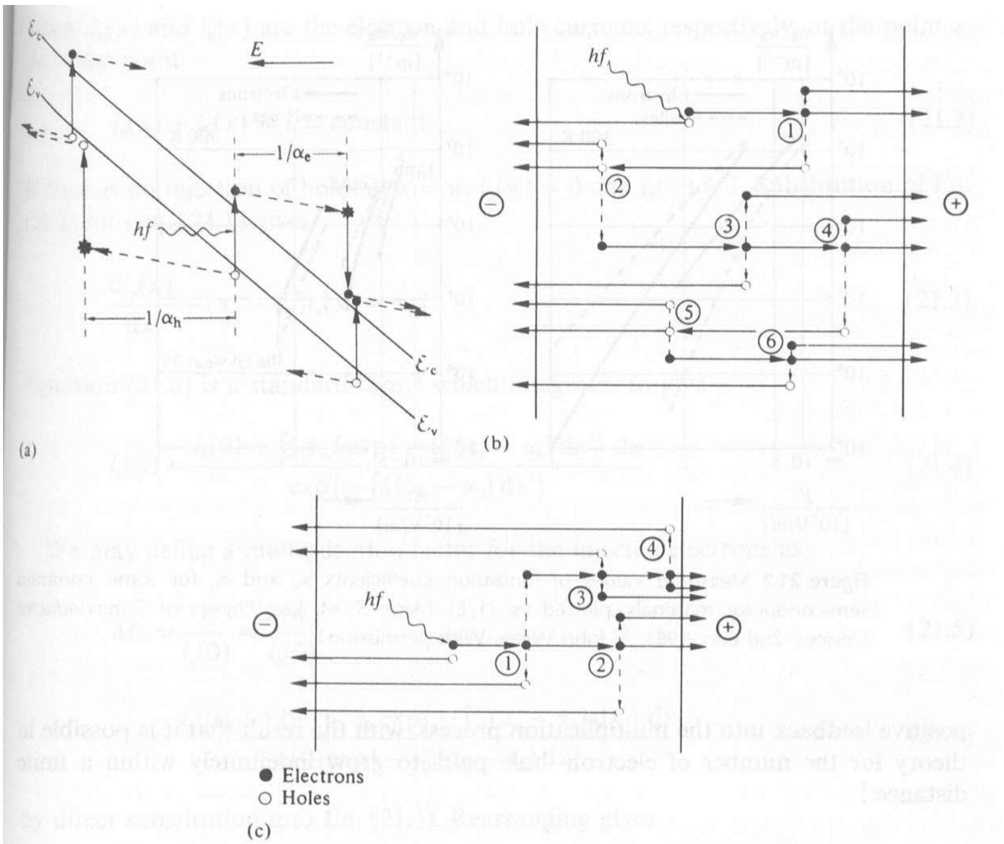
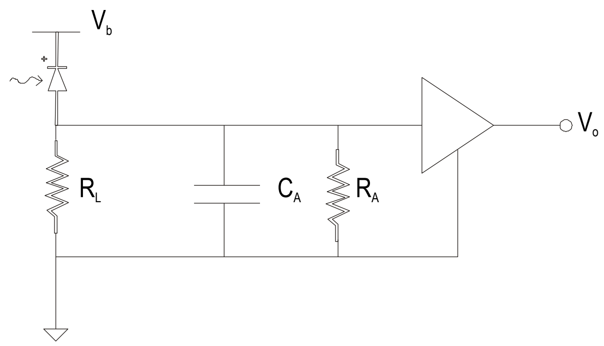


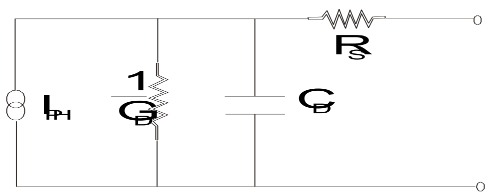
Figure 8.8 Schematic of the avalanche process.

9. The Receiver Amplifier

To be of use the photodiode must be connected to an amplifying circuit which conditions the signal for further use. The simplest method is to place the reverse biased photodiode in series with a load resistor which is then used as the input to a voltage amplifier:



The amplifier will have its own input resistance and capacitance R_A & C_A. For the photodiode itself, we can write an equivalent circuit:



Here I_{ph} is an ideal current source, G_D is the effective conductance due to the slope of the I-V characteristic and is almost negligible, C_D is the junction capacitance, and R_s is the series resistance of the bulk semiconductor and contacts. C_D may be < 1 $_p$F, while R_s may be a few Ω.

To analyse the performance of this circuit we need to know what the noise sources are. There are three main ones: shot noise, thermal noise, and amplifier noise. These sources tend to generate an amount of noise power per unit frequency, so the total noise power is proportional to the signal bandwidth. Since we are interested in the SNR of the photo<u>current</u> , since this is proportional to the optical power, we will talk about equivalent noise current sources I^{*}, but the simplest expressions are for noise current squared per unit frequency. We call this noise square spectral density: $(I^*)^2$

where the noise current will be : $I_{noise} = ((I^*)^2 \Delta f)^{1/2}$

Shot noise is due to the fact that photon detection is a quantised process, so there is a statistical variation in photon arrival times. This noise has a power proportional to the optical signal, and is in that sense unusual. This is because the photocurrent is proportional to the statistical average number of photons arriving, and the variation (uncertainty) in this number is approximately equal to its square root. Thus:

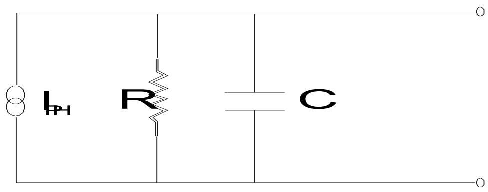
$(I_{sh}^*)^2 = 2eI_{ph}$

Thermal noise is due to thermal motion of conductors in the resistors of the circuit, and is given by:

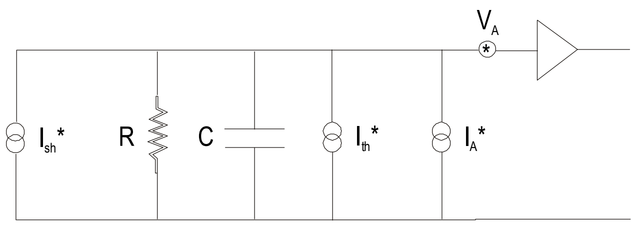
$(I_{Th}^*)^2 = 4kT/R$

The amplifier will itself have intrinsic noise which we can characterise by an equivalent noise input current and an equivalent noise input voltage.

If we ignore the photodiode series resistance, we can lump the other R & C components together with those of the amplifier, giving:



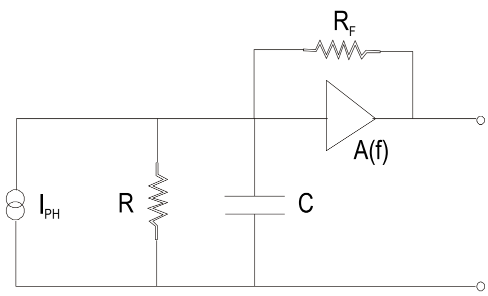
where $C = C_D + C_A$, $1/R = G_D + 1/R_A$. The equivalent noise circuit will be:



We need to include R & C because they will determine the equivalent current due to V_A.

$$\text{SNR} = \frac{I_{ph}}{[(V_{A^*})^2(\frac{1}{R^2} + \frac{4\pi^2}{3}\Delta f^2 C^2) + 2eI_{ph} + \frac{4kT}{R} + (I_{A^*})^2]^{1/2}\,\Delta\text{f}^{1/2}}$$

The voltage amplifier has the serious disadvantage that the gain is frequency dependent due to C, so equalisation is needed on the amplified signal. This problem is eliminated in the transimpedance amplifier, which is much more commonly used in practice.



here the capacitance is effectively shorted out, as the photocurrent will flow through the feedback resistor R_F.

The SNR needed in a receiver depends on the application. For a digital signal, the important criterion is the bit error rate (B.E.R.), which is the probability that a bit will be incorrectly interpreted. For many applications, a BER of 10^{-9} is considered appropriate (one mistake per 10^9 bits), and for this a SNR of 12 is required on the photocurrent.

Part C: Summary

Optical Sources

Photon energy, relation to band-gap. Energy-wavelength relationship.
Recombination of electrons and holes in a p-n junction; indirect and direct band-gaps.
Radiative and non-radiative recombination; quantum efficiency.
External quantum efficiency: Re-absorption, total internal reflection, Fresnel reflection.
Response speed of LED's.
Launching into fibre; significance of numerical aperture.
The semiconductor laser diode: basic operating principles.
Laser diode spectra, efficiency, speed.

Photodetection

Absorption by hole-electron pair creation (bandgap transition).
Absorption spectra of semiconductors, mobility and saturation drift velocities.
Photoconductive detectors.
Depletion regions in reverse biased diodes: charge density, field, and potential distributions.
Responsivity and quantum efficiency.
The p-i-n photodiode structure: advantages and properties.
Avalanche photodiodes.

The Receiver Amplifier

Voltage amplifier and transimpedance amplifier configurations.
Equivalent circuit of photodiode.
Sources of noise: thermal, shot and amplifier noise.
Noise equivalent circuits, SNR.

Part C: Problems

1) An optical source produces pulses whose intensity is Gaussian in time, with a spectral width resulting purely from the Fourier Transform of the time variation of the field due to modulation. Derive and plot an expression relating the transmitted pulse width giving minimum received pulse width with the fibre length, for a wavelength of 850 nm and a dispersion coefficient of D = 100 ps/(nm km). [This is hard!]

2) A p-i-n photodetector has an intrinsic layer width and doping level of $w_i = 10\mu m$ and $N_D^- = 10^{19}\ m^{-3}$, and a relative permittivity $\varepsilon_r = 12$.

   a) If the p and n layer doping levels, N_A and N_D^+, are equal, what must they be to ensure that when the bias voltage is high enough to maintain $E > 10^6$ V/m everywhere in the intrinsic layer, the total depletion layer width is no more that 10% greater than w_i ?

   b) What voltage needs to be applied to achieve the condition in (a)?

   c) How much variation will there then be in E in the intrinsic layer?

3) A silicon p-i-n photodetector has $w_i = 20\mu m$ $\varepsilon_r = 12$, $\alpha = 8 \times 10^4\ m^{-1}$, and is detecting a $\lambda = 0.85\mu m$ signal.

   a) Determine the quantum efficiency, neglecting absorption in the p region.

   b) Find the incident optical power needed to make the shot noise and the thermal noise equal, assuming a voltage amplifier is used with a combined input resistance of 10 kΩ.

   c) For such a signal level, what is the maximum bandwidth to maintain SNR = 12, ignoring the amplifier noise terms?

4) An LED has a Lambertian emitting region which is 5μm below the surface of a semiconductor having $\varepsilon_r = 12$, $\alpha = 2 \times 10^4\ m^{-1}$

   a) Find the external quantum efficiency of this diode, due to all 4 loss terms. For Fresnel reflection, use the normal incidence approximation, and for absorption, you may use the first order approximation for the exponential $e^{-x} = 1 - x$.

   b) How much is this increased if the diode is capped with a glass hemisphere, n = 1.5? Consider reflections at the two interfaces, but not multiple reflections.