CHAPTER 7

Photodetectors

The optical detectors discussed in this chapter may be classified as either *thermal* or *photon* devices. In thermal detectors, the absorption of light raises the temperature of the device and this in turn results in changes in some temperature-dependent parameter (e.g. electrical conductivity). As a consequence, the output of thermal detectors is usually proportional to the amount of energy absorbed per unit time by the detector and, provided the absorption efficiency is the same at all wavelengths, is independent of the wavelength of the light. In photon detectors, on the other hand, the absorption process results directly in some specific quantum event (such as the photoelectric emission of electrons from a surface) which is then 'counted' by the detection system. Thus the output of photon detectors is governed by the rate of absorption of light quanta and not directly on their energy. Furthermore, all the photon processes considered here require a certain minimum photon energy to initiate them. Since the energy of a single photon is given by $E = h\nu = hc/\lambda_0$ (see eq. 1.2), photon detectors have a long wavelength 'cut-off', that is a maximum wavelength beyond which they do not operate.

A problem encountered with photon detectors operated in the infrared is that the photon energies involved become comparable with the average thermal energies ($\approx kT$) of atoms in the detector itself. A relatively large number of quantum events may then be generated by thermal excitation rather than by light absorption and will thus constitute a source of noise. The obvious way to reduce this noise signal is to reduce the temperature of the detector; indeed most photon detectors operating above a wavelength of 3 μm or so must be cooled to liquid nitrogen temperatures (77 K) or below.

7.2.1 Thermoelectric detectors

Thermoelectric detectors use the principle of the thermocouple (i.e. the Seebeck effect) whereby the heating of one junction between two dissimilar metals relative to the other causes a current to flow round the circuit which is proportional to the temperature difference between the junctions. In thermoelectric detectors one junction is used to sense the temperature rise of the receiving element whilst the other is maintained at ambient temperature, as shown in Fig. 7.3. A rather more sensitive detector may be made by connecting several thermocouples together in series; the device is then known as a *thermopile*. Efficient operation calls for materials with large electrical conductivities (to minimize Joule heating effects) and also small thermal conductivities (to minimize heat conduction losses). These two requirements are usually incompatible, and a compromise has to be reached. Although metals are most often used for the junction materials, certain heavily doped semiconductors can offer improved sensitivity, but these are generally less robust and give rise to constructional problems. The usefulness of thermoelectric detectors lies in their simplicity and their rugged construction.

7.2.2 The bolometer

In the bolometer the incident radiation heats a fine wire or metallic strip causing a change

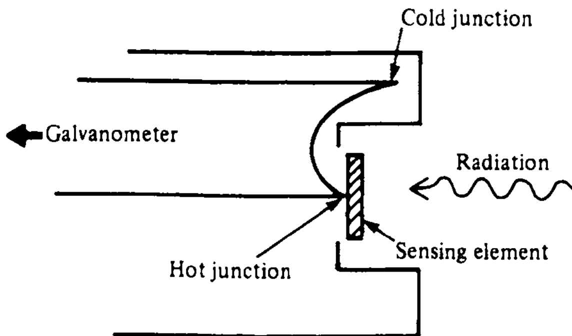


FIG. 7.3 Thermoelectric detector: the temperature change of the sensing element induced by the absorption of radiation is detected using a thermocouple, one junction of which is attached to the sensing element, and the other shielded from the radiation.

in its electrical resistance. This may be detected in several ways: for example, the element may be inserted into one arm of a Wheatstone bridge (Fig. 7.4), or in place of the photoconductor in the circuit of Fig. 7.17. Care must be taken to ensure that any currents flowing through the element are sufficiently small not to raise its temperature by a significant amount. The main parameter of interest in assessing the performance is the temperature coefficient of resistance α, which is given by

$$\alpha = \frac{1}{\rho}\frac{d\rho}{dT}$$

where ρ is the resistivity of the material and T the temperature. The resistivity of metals increases with increasing temperature, and hence for these α will be a positive quantity. Platinum and nickel are the most commonly used, and both have α values of about 0.005 K^{-1}. Greater sensitivity may be achieved by using semiconducting elements, which are sometimes called *thermistors*. These consist of oxides of manganese, cobalt or nickel and have α values of about −0.06 K^{-1} (for these materials, α is dependent on temperature). The negative sign arises because of the characteristic decrease in resistivity with increasing temperature of semiconductors above a certain temperature (see eq. 2.24 and the discussion following eq. 2.36).

Carbon resistance bolometers cooled to liquid helium temperature (4.2 K) have proved successful in far-IR astronomy where very sensitive detectors are required. Use can also be made of superconducting materials; in these the resistivity drops suddenly to zero below a particular temperature (the 'critical temperature'). In operation the temperature of the element is held just below this value so that the absorption of even a small amount of radiation will cause a very rapid increase in the resistivity. The devices exhibit exceptional sensitivity but

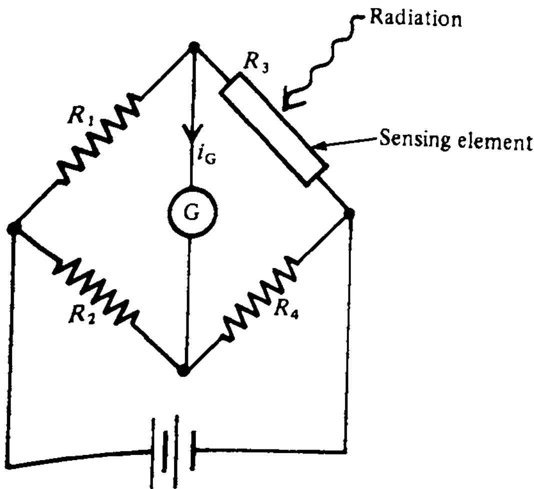


FIG. 7.4 Wheatstone bridge circuit incorporating a bolometer radiation sensing element. When the resistance values are such that $R_1/R_2 = R_3/R_4$ then the current i_G through the galvanometer is zero. If, however, the sensing element resistance changes slightly then a current will flow which is proportional to the resistance change.

require that the ambient temperature be very accurately controlled. The latest designs use high temperature superconducting materials such as yttrium barium copper oxide (YBCO).

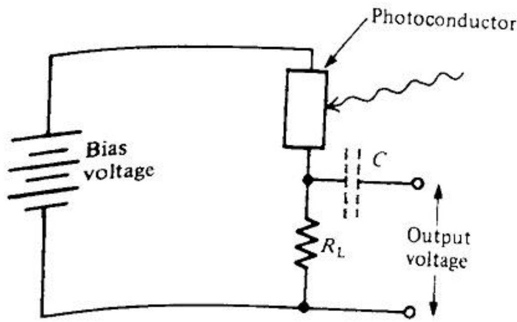


FIG. 7.17 Photoconductor bias circuit. The photoconductor is placed in a series circuit comprising a voltage source, a load resistor R_L and the photoconductor itself. Changes in the resistance of the photoconductor cause changes in the voltage appearing across R_L. If only the a.c. component of this voltage is required, then a blocking capacitor C may be placed as shown.

7.2.4 Pyroelectric detectors

Pyroelectric detectors are a more recent development, and

they can be made with very rapid response times and are very robust. The incident radiation is absorbed in a ferroelectric material which has molecules

with a permanent electrical dipole moment. Below a critical temperature, the *Curie temperature* T_c, the dipoles are partially aligned along a particular crystallographic axis giving rise to a net electrical polarization of the crystal as a whole. When the material is heated, the increased thermal agitation of the dipoles decreases the net polarization, which eventually becomes zero above T_c, as shown in Fig. 7.6.

The most sensitive material in use is triglycine sulfate (TGS), but this has an inconveniently low Curie temperature of only 49°C and more commonly used materials are ceramic based, such as lead zirconate, which have Curie temperatures of several hundred degrees centigrade.

The detector consists of a thin slab of ferroelectric material cut such that the spontaneous polarization direction is normal to the large area faces. Transparent electrodes are evaporated onto these faces and connected together with a load resistor (which can have values as high as $10^{11}\ \Omega$) as shown in Fig. 7.7(a). A temperature change of the ferroelectric causes the spontaneous polarization to vary and hence also the amount of captive surface charge on the faces. Changes in the surface charge induce corresponding changes in the charge on the electrodes, thus causing a current to flow through the load resistor. This in turn results in a changing voltage signal appearing across the load resistor. Radiation of constant irradiance will not cause any change in the charge stored on the electrodes and will consequently not give rise to an output signal. The frequency response of the pyroelectric detector is considered in detail in Problem 7.6. At low frequencies, the output rises from zero to reach a plateau when $f > 1/2\pi\tau_H$, where τ_H is the thermal time constant given by eq. (7.5).

At higher frequencies, the electrode capacitance C acts as a signal shunt across the load resistor R_L and the output voltage falls to $1/\sqrt{2}$ of its maximum value at a cut-off frequency f_c given by

$$f_c = \frac{1}{2\pi R_L C} \tag{7.7}$$

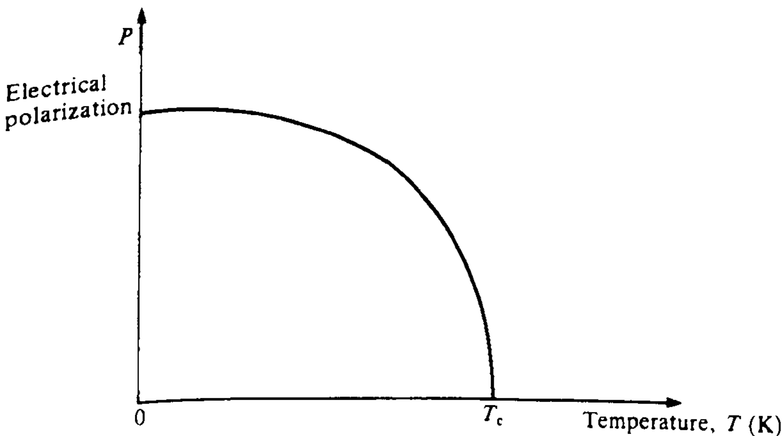


FIG. 7.6 Spontaneous electrical polarization versus temperature for a ferroelectric material (schematic diagram). The polarization falls to zero at the Curie temperature T_c.

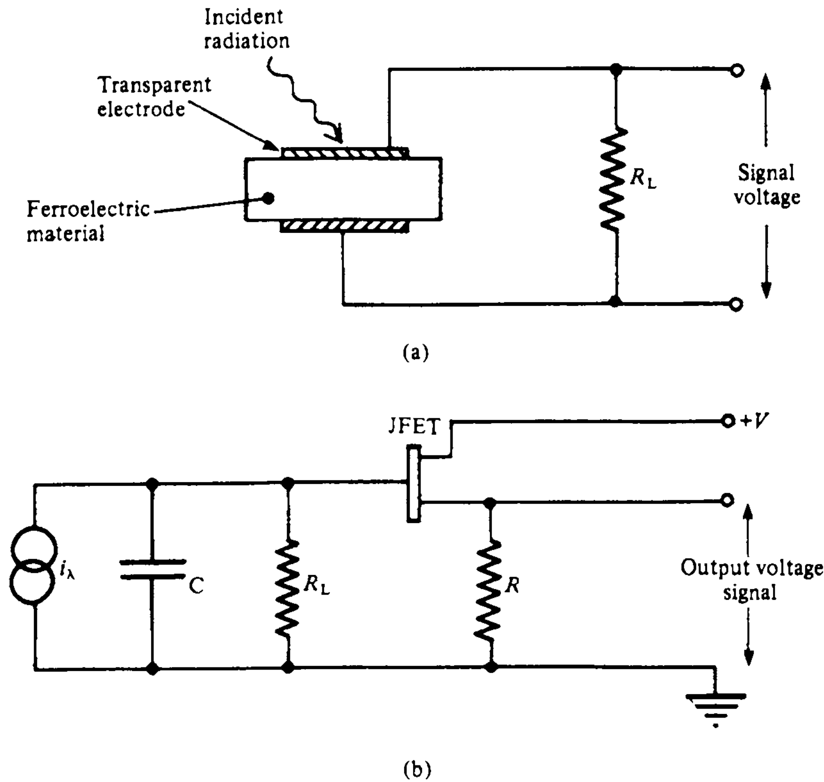


FIG. 7.7 (a) Pyroelectric detector. A slab of ferroelectric material is sandwiched between two electrodes (one being transparent). The electrodes are connected by a load resistor R_L. Radiation absorbed within the ferroelectric material causes it to change its polarization. The induced charge on the electrodes changes and current flows through R_L causing a voltage signal to appear across R_L. (b) Equivalent circuit and typical impedance matching circuitry for a pyroelectric detector. The varying amounts of charge stored on the electrodes are equivalent to a current generator feeding into the electrode capacitance C. The load resistor R_L is in parallel with C. Since R_L is usually very high (about 10^9 Ω or more), an impedance matching circuit is often employed to reduce the signal source impedance. A typical circuit using a JFET is shown here; the output impedance in this case is then R (≈ 1 kΩ).

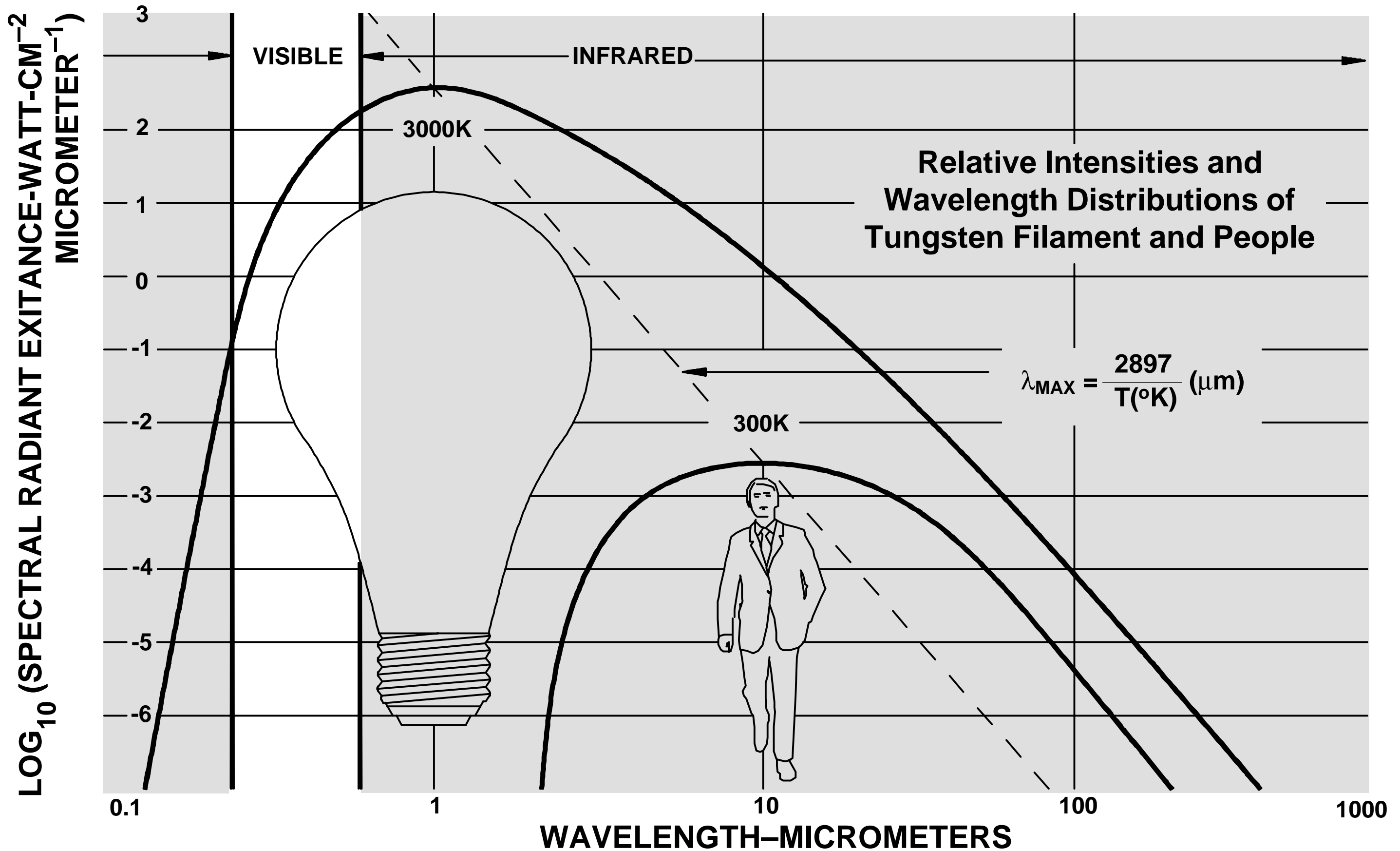
As the voltage output is proportional to R_L (again see Problem 7.6) there is a trade-off between sensitivity and frequency response. Typically, a detector with a frequency bandwidth of 1 Hz at an operating frequency of 100 Hz can detect radiation powers of about 10^{-8} W.

Because of the comparatively large values of the load resistor encountered in pyroelectric detectors, an impedance matching circuit is usually built into the detector. A source follower circuit using a JFET is commonly used as shown in Fig. 7.7(b).

Pyroelectric detectors can be made with response times in the nanosecond region and with a wavelength response extending out to 100 μm. They have proved very useful as low cost, robust IR detectors in such uses as fire detection and intruder alarms.

ELTECdata # 100

Introduction To Infrared Pyroelectric Detectors



Use Infrared: It's Already There

Pyroelectric detectors make mid-range infrared affordable. You use what is already there — 100% natural and harmless.

You can use the invisible glow of objects and people to detect, count, monitor, locate, activate, conserve, protect, or warn. It is passive technology.

Beyond Photodiodes

Visible light goes from 0.4 to 0.7 micrometers on the wavelength spectrum. Beyond that is infrared. Photodiodes are inexpensive and practical even to 1 micrometer. But, 1 micrometer corresponds to the "wavelength of maximum energy" of a blackbody at about 2,900 Kelvin (4,700^{o}F) which is the temperature of an incandescent bulb's white-hot filament.

To use the infrared emitted from ourselves or objects that we can touch, wavelengths well beyond 1 micrometer and especially those around 10 micrometers must be detected.

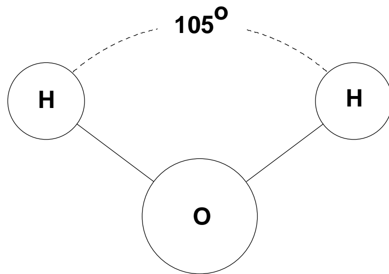
Pyroelectrics Are Practical

Detection of mid-range infrared is not new. Thermistors and thermopiles (thin-film thermocouples) have long been available. Although these components are relatively inexpensive, the circuitry required to make them work is not. Moreover, both thermistors and thermopiles are generally found wanting in terms of signal strength and speed of response.

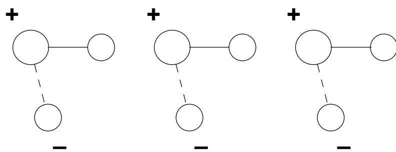
Pyroelectrics are today's practical choice for broad-band IR detection. Pyroelectrics offer technical advantages in signal strength, speed of response and in minimizing interconnecting circuitry. And, as has happened with other components, use of more sophisticated production techniques pioneered at ELTEC INSTRUMENTS, INC., has increased the availability of lithium tantalate pyroelectric detectors while lowering cost.

The Pyroelectric Effect: The Material

If a material has an internal electrical symmetry, it's neutral. If it's unsymmetrical — like water — it has a permanent electric dipole. Most unsymmetrical materials in bulk have a zero dipole effect because of a random or self-cancelling arrangement.



Water is unsymmetrical



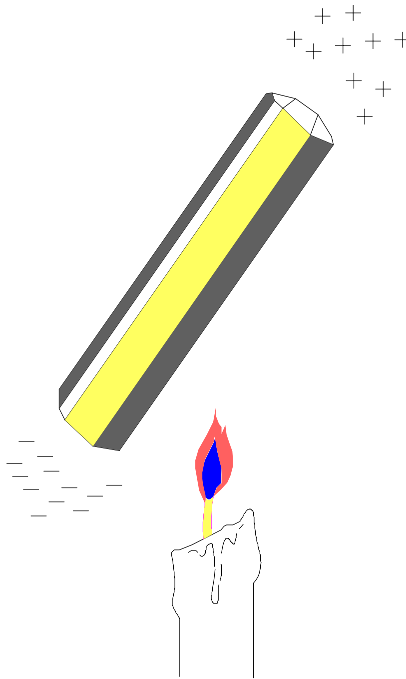
Dipoles acting in unison

There are some unsymmetrical materials which maintain a net dipole orientation even in bulk. Heating such a material (within limits) doesn't randomize the dipoles, but rotates them in unison and thus maintains a polarization. Since this occurs in the absence of an external electric field, it is called spontaneous polarization.

Dipoles will act in unison to an upper temperature point called the Curie point. Lithium tantalate is a practical pyroelectric material because it has a Curie point of 610° C. Also, lithium tantalate is a very responsive synthetic crystal with an established, long-term stability.

The Pyroelectric Effect: Simplified

The Greeks discovered the pyroelectric effect 23 centuries ago. They observed that when tourmaline was placed in hot ashes, it first attracted and then repelled them (charge generation ... attraction by induction ... contact/reversal ... repulsion of like charge). Hence "pyro", for fire, plus electric !

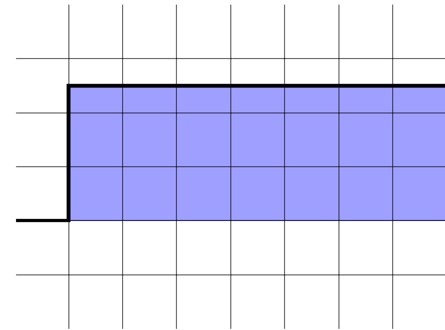


Heated tourmaline develops electric charges

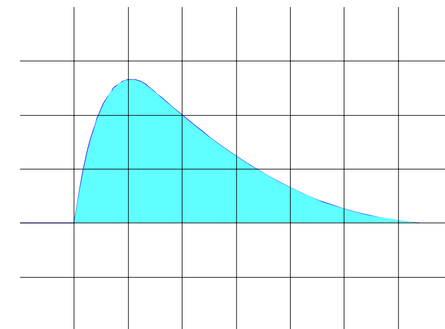
Pyroelectric isn't thermo-electric. In a thermoelectric device, like a thermocouple, a steady voltage is produced when two junctions of dissimilar metals are held at steady but different temperatures. In a pyroelectric device, a change in temperature creates a change in polarization. "Electrical polarization" is just another way of saying "electrical charge". The charge is collected by electrodes on the crystal. So the "open circuit" Voltage = (Q, charge) / (C, crystal capacitance).

In THEORY, if the crystal were levitated in a vacuum in a perfectly reflective Dewar, at infinite impedance (and some other conditions), and a thermal step function applied, the voltage would follow the step function.

In REALITY, the crystal has a thermal time constant, so it will quickly thermalize to its environment (return to ambient) after a step input. This releases the strain on the crystal lattice and the crystal "reabsorbs" the electrons as the lattice returns to its neutral state. Thus, "step function in," "voltage pulse out."

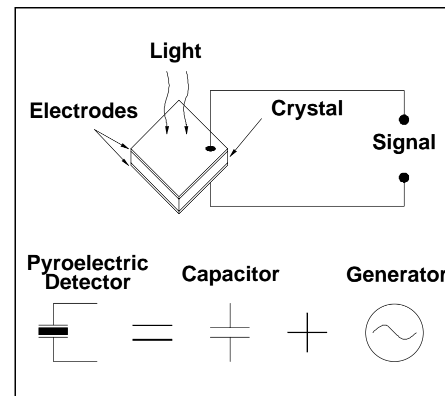


Infrared input to detector (step)



Response of pyroelectric detector

The nature of the pyroelectric detector makes it both fast and useful. Since every object is emitting infrared light, every object is a transmitter. And since the infrared detector responds to infrared, it is a receiver. An intruder entering a room is like an invisible light being turned on; the detector responds to the change in infrared light, generating a useful signal. A glass object (transparent in the visible and near-infrared) may pass right through a light beam undetected, but its infrared emissions will identify it every time. In short, wherever there's a change in infrared light, there's a potential pyroelectric application.



The Pyroelectric Detector

A thin wafer of lithium tantalate has electrodes deposited on both faces. The electrodes gather the charge which is unable to leak through because the material is such a good dielectric (insulator). In its simplest form, the pyroelectric detector is both a capacitor and a charge generator (in response to infrared light striking a face, being absorbed as heat, creating change in polarization). And all this at room temperature, without the need for cooling or electrical biasing.

Electrical Considerations

Think of a pyroelectric detector as a tiny flat-plate Active Capacitor. Typical capacitance is about 30 picofarads. Insulation resistance is 5×10^{12} Ohms. So, except in laser applications, the extreme source impedance makes use of the crystal by itself impractical. PRACTICAL pyroelectrics contain either a JFET source follower (Voltage Mode) or a transimpedance amplifier (Current Mode).

For a rough idea of the signal you might get using a pyroelectric detector, use the following formula:

V responsivity = (current responsivity)(effective impedance)

$$I \cdot (R/\sqrt{1+(2\pi fRC)^2})$$

for I, use 0.5 to 1 microamp per watt

for R, use either your load resistor value or feedback resistor value

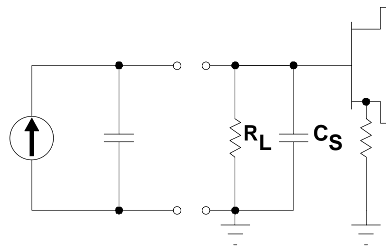
for C, use detector capacitance for Voltage Mode (typically 30 pF) or use stray feedback capacitance for Current Mode (typically 0.03 pF)

The formula is very useful to get an approximation of voltage responsivity at different frequencies (modulation rates).

Both of these amplification schemes have positive and negative features. The voltage mode circuit will generally yield the best signal to noise ratio and it can operate at a very small supply voltage and current. However, it does not have a large output responsivity and output response will be frequency dependent (unless a low value R_L is used). The current mode offers a substantial increase in output signal. It can have a "flatter" frequency response and that response can be set independent of the crystal. Unfortunately, the noise characteristics of the operational amplifier limit the signal-to-noise ratio and the operating voltage and current requirements are greater.

NOTE: Although the voltage (Field Effect Transistor) or current (Op Amp) circuits can be added externally to the basic detector package, it is accomplished with the addition of stray capacitance, susceptibility to EMI, testing problems, expense and possibly a compromise in reliability. To circumvent these problems, detectors are offered with the FET and appropriate load resistor or op amp and appropriate feedback resistor in the detector package.

Detector connected with source follower



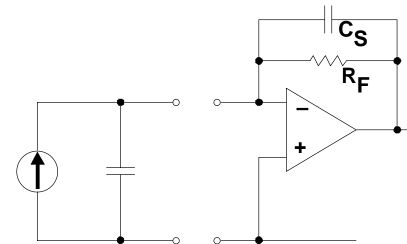
The voltage follower is basically a FET connected as a source follower.

In this configuration, the voltage output will be:

$R_V = R_i \cdot Z_{eff} \cdot A_o$

where
- R_V = voltage response in V/W
- R_i = current responsivity
- Z_{eff} = lumped impedance of crystal, R_L, and stray capacitance at the input
- A_o = follower gain (approx. 0.8)

Detector connected with a current to voltage converter



The current to voltage converter can be an operational amplifier connected as shown.

In this configuration the voltage output will be:

$R_V = R_i \cdot Z_F$

where Z_F = lumped impedance of feedback loop including R_F and C_S, stray feedback loop resistance and capacitance

7.3.3 Photomultipliers

In the photomultiplier, the photoelectrons are accelerated towards a series of electrodes (called *dynodes*) which are maintained at successively higher potentials with respect to the cathode. On striking a dynode surface, each electron causes the emission of several secondary electrons which in turn are accelerated towards the next dynode and continue the multiplication process. Thus, if on average δ secondary electrons are emitted at each dynode surface for each incident electron and if there are N dynodes overall, then the total current amplification factor between the cathode and anode is given by

$$G = \delta^N \tag{7.10}$$

Considerable amplification is possible: if we take, for example, $\delta = 5$ and $N = 9$, we obtain a gain of 2×10^6.

Four of the most common photomultiplier dynode configurations are illustrated in Fig. 7.12. Three of them (venetian blind, box and grid, and linear focused) are used in 'end-on' tubes. These have a semitransparent cathode evaporated onto the inside surface of one end of the tube envelope. The photoelectrons are emitted from the opposite side of the cathode layer to that of the incident radiation. Obviously, in this arrangement the thickness of the photocathode is very critical. If it is too thick, few photons will penetrate to the electron-emitting side, whilst if it is too thin few photons will be absorbed.

In the venetian blind type, electrons strike a set of obliquely placed dynode slats at each dynode stage; the electrons are attracted to the next set of slats by means of the interdynode potential applied between a thin wire grid placed in front of the slats. This arrangement is compact, relatively inexpensive to manufacture and is very suitable for large area cathodes. The box and grid type (Fig. 7.12b) is somewhat similar in performance. In both of these, very little attempt is made to focus the electrons, which is in contrast to the linear focused and circular cage focused types (Figs 7.12c and d), where some degree of electron focusing is obtained by careful shaping and positioning of the dynodes.

The focused types have somewhat higher electron collection efficiencies and a much better response to high signal modulation frequencies (we discuss frequency response later in this section). The circular cage focused type is very compact and usually used in conjunction with a side window geometry. In this, the photocathode material is deposited on a metal substrate within the glass envelope and the photoelectrons are emitted from the same side of the cathode as that struck by the incident radiation.

The dynode potentials are usually provided by means of the circuit shown in Fig. 7.13. Care must be taken to ensure that the voltage between the cathode and the first dynode is large enough to maintain proportionality between cathode current and cathode illumination. Usually a voltage value is recommended for a particular tube, and in some circumstances (e.g. when examining fast pulses) it may be preferable to use a Zener diode in place of the fixed resistor R_k in Fig. 7.13 to keep the voltage at this value.

The intermediate stages usually operate satisfactorily over quite a wide voltage range provided the voltage is distributed uniformly. To maintain this uniformity, the current flowing down the dynode chain must be considerably larger (say 100 times) than the anode current.

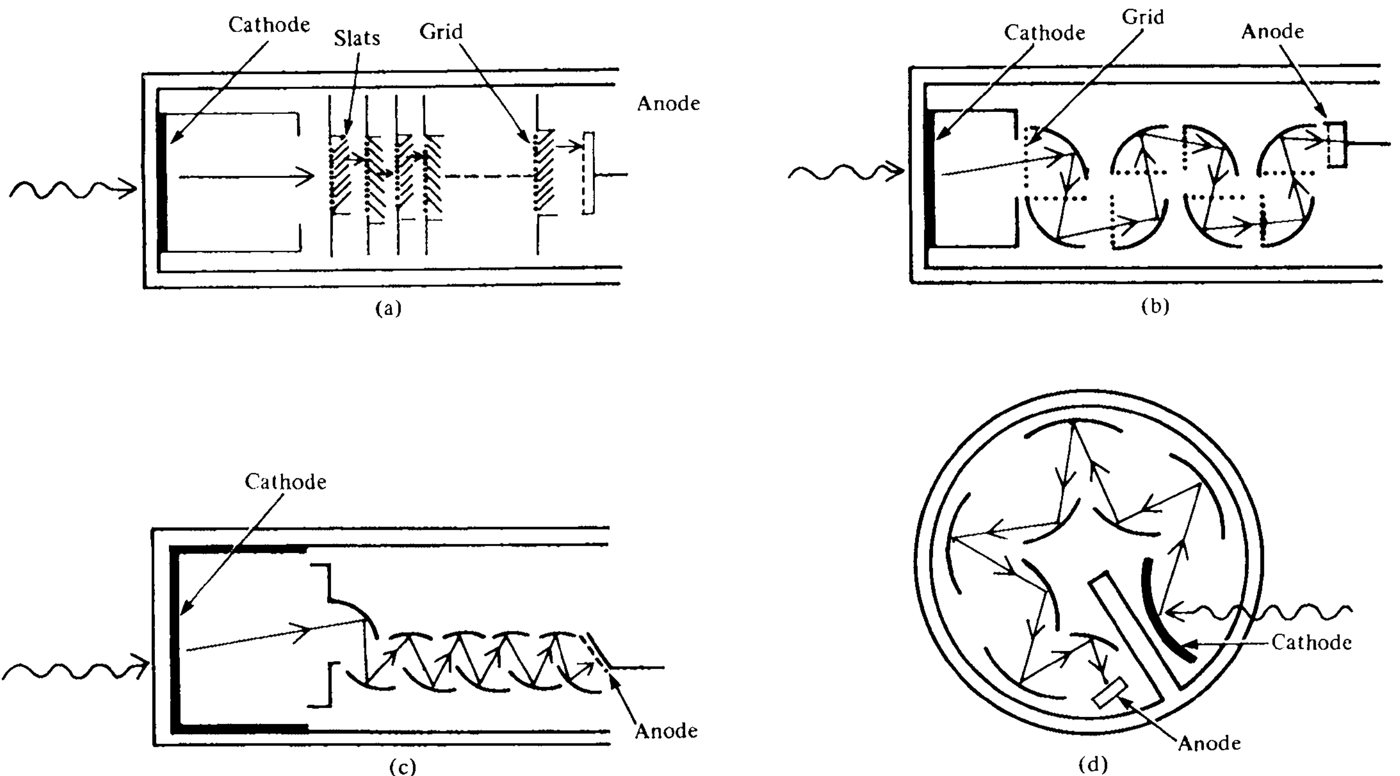


FIG. 7.12 Dynode structures of four common types of photomultiplier: (a) venetian blind, (b) box and grid, (c) linear focused and (d) circular cage focused. Typical trajectories of an electron through the systems are also shown.

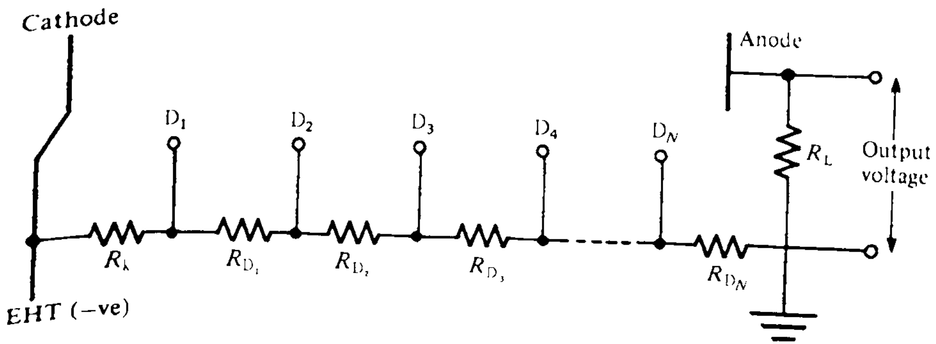


FIG. 7.13 Dynode biasing circuit using a linear resistor chain. The EHT voltage supply is applied across a resistor chain (R_k, R_{D1}, R_{D2}, etc.) which acts as a potential divider and maintains the dynodes (D_1, D_2, D_3, etc.) at increasingly higher positive potentials relative to the cathode. When an amplified signal current pulse arrives at the anode, it flows through the anode load resistor R_L causing a voltage to appear across it.

If high anode currents are likely, then the last few stages may also be biased using Zener diodes.

The photomultiplier responds to light input by delivering charge to the anode. This charge may be allowed to flow through a resistor R_L or to charge a capacitor: the corresponding voltage signal then provides a measure of the input optical signal. If individual pulses need to be examined (as, for example, when using photon counting techniques, which we discuss in section 7.3.9), then it is important to ensure that the response time of the external circuitry is less than that of the pulse rise time. This usually implies a low value for the load resistor.

Traditionally photomultipliers have been relatively bulky devices with photocathode diameters of 25 mm or more. Recently much smaller devices have become available contained in small metal cans with photocathode diameters of 8 mm or so.

SEMICONDUCTOR-BASED PHOTODETECTORS

Introduction

Optical detectors have been used for almost two centuries to measure optical power. Astronomer Sir William Herschel was first to use a thermometer to detect (and hence discover!) the infrared portion of the spectrum in April 1800 when he was studying the spectral output of the sun [1]. Since that time, the science of optical detection has advanced to the point where performance is limited only by quantum effects. Four basic parameters are listed to characterize the performance of an optical detector. These are:

1. *Responsivity*: How much signal is obtained per unit radiated power? This signal is generally specified in units of amps/watt (A/W) or in volts/watt (V/W).
2. *Spectral response*: How does the responsivity vary with wavelength?
3. *Detectivity*: What is the minimum detectable power required to achieve a signal-to-noise ratio of unity? The signal-to-noise (S/N) ratio indicates the relative size of the signal and noise currents (or voltages). As the S/N ratio increases, the quality of the received signal increases (e.g., the number of errors in transmission reduce).
4. *Time response*: How fast does the detector respond to a change in signal?

Responsivity and Quantum Efficiency

Consider a semiconductor photodetector which absorbs photons with energies at or above E_g, the bandgap energy of the semiconductor. Suppose the incident optical power is given by P_{in}, and it is assumed that all the incident photons enter the semiconductor. Suppose the photocurrent generated as a result of this incident optical power is given by I_p, then the relationship between P_{in} and I_p is

$$I_p = RP_{in}, \tag{4.168}$$

where R is the responsivity of the photodetector in units of AW^{-1}.

The quantum efficiency of the detector η may be defined as ratio of the number of hole–electron pairs generated to the number of incident photons, and is given by

$$\eta = \frac{\dfrac{I_p}{q}}{\dfrac{P_{in}}{h\nu}} = \frac{h\nu}{q}R. \tag{4.169}$$

Thus, R may be written as

$$R = \frac{q\eta}{h\nu} = \frac{\eta\lambda q}{hc}. \tag{4.170}$$

The responsivity can be degraded if the detector is presented with an excessively large optical power. This condition, which is called detector saturation, limits the detector's **linear dynamic range**, which is the range over which it responds linearly with the incident optical power.

Suppose the thickness of the semiconductor is w, and the absorption coefficient is α Np m^{-1}, then the transmitted optical power escaping from the semiconductor is given by

$$P_{\text{tr}} = P_{\text{in}} e^{-\alpha w}. \tag{4.171}$$

Thus, the absorbed power is given by

$$P_{\text{abs}} = P_{\text{in}} - P_{\text{tr}} = P_{\text{in}}(1 - e^{-\alpha w}). \tag{4.172}$$

Since every absorbed photon creates one hole–electron pair, the quantum efficiency is given by

$$\eta = \frac{P_{\text{abs}}}{P_{\text{in}}} = (1 - e^{-\alpha w}), \tag{4.173}$$

which assumes that all the incident photons enter the semiconductor with no reflection.

Indeed, not all incident photons produce electron-hole pairs because not all incident photons are absorbed. Some photons simply fail to be absorbed because of the probabilistic nature of the absorption process. Others may be reflected at the surface of the detector, thereby reducing the quantum efficiency further. Furthermore, some electron-hole pairs produced near the surface of the detector quickly recombine because of the abundance of recombination centers there and are therefore unable to contribute to the detector current. Finally, if the light is not properly focused onto the active area of the detector, some photons will be lost.

Therefore, more precisely the quantum efficiency can be written as:

$$\eta = (1 - \mathscr{R})\zeta[1 - \exp(-\alpha w)],$$

where $\mathscr{R}$ is the optical power reflectance at the surface, ζ the fraction of electron–hole pairs that contribute successfully to the detector current, α the absorption coefficient of the material (cm^{-1}) discussed in Sec. 15.2B, and w the photodetector depth. Equation (17.1-1) is a product of three factors:

- The first factor $(1 - \mathscr{R})$ represents the effect of reflection at the surface of the device. Reflection can be reduced by the use of antireflection coatings.
- The second factor ζ is the fraction of electron–hole pairs that successfully avoid recombination at the material surface and contribute to the useful photocurrent. Surface recombination can be reduced by careful material growth.
- The third factor, $\int_0^w e^{-\alpha x}\,dx / \int_0^\infty e^{-\alpha x}\,dx = [1 - \exp(-\alpha w)]$, represents the fraction of the photon flux absorbed in the bulk of the material. The device should have a sufficiently large value of w to maximize this factor.

The effects of absorption on quantum efficiency are illustrated in Fig. 17.1-1.

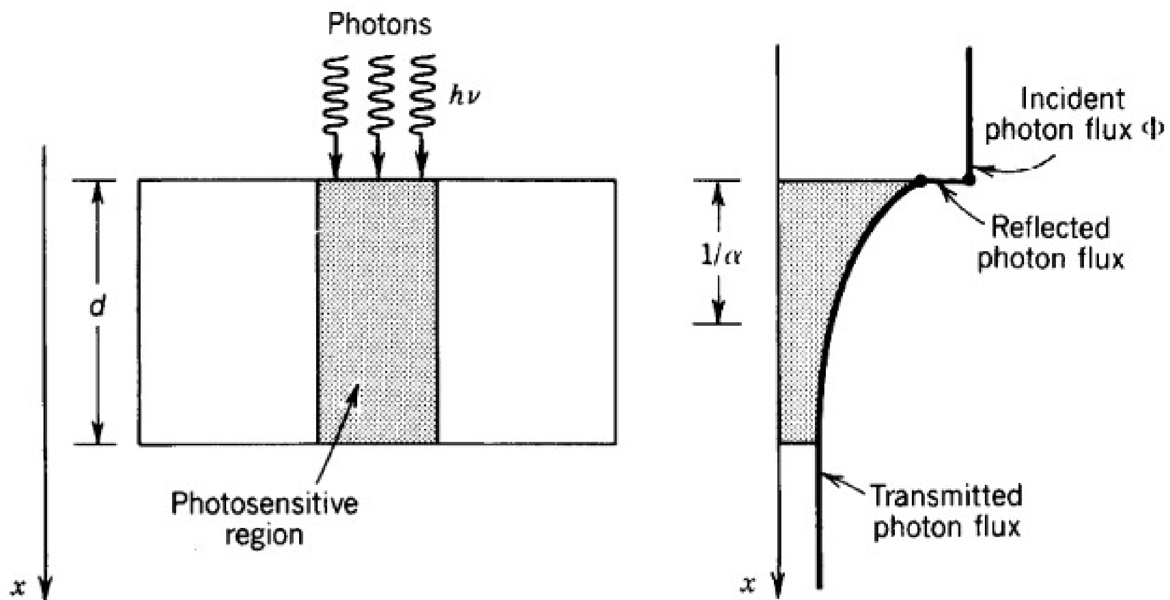


Figure 17.1-1 Effect of absorption on the quantum efficiency η.

Dependence of η on Wavelength

The quantum efficiency η is a function of wavelength, principally because the absorption coefficient α depends on wavelength. For photodetector materials of interest, η is large within a spectral window that is determined by the characteristics of the material. For sufficiently large λ_o, η becomes small because absorption cannot occur when $\lambda_o \geq \lambda_g = hc_o/E_g$ (the photon energy is then insufficient to overcome the bandgap). The bandgap wavelength λ_g is the **long-wavelength limit** of the semiconductor material.

For sufficiently small values of λ_o, η also decreases, because most photons are then absorbed near the surface of the device (e.g., for $\alpha = 10^4\ \text{cm}^{-1}$, most of the light is absorbed within a distance $1/\alpha = 1\ \mu\text{m}$). The recombination lifetime is quite short near the surface, so that the photocarriers recombine before being collected.

Responsivity and quantum efficiency are linked by the formula (4.170), which can be rewritten as

$$\Re = \frac{\eta e}{h\nu} = \eta \frac{\lambda_o}{1.24}.$$

An appreciation for the order of magnitude of the responsivity is gained by setting $\eta = 1$ in (17.1-3), whereupon $\Re = 1$ A/W, i.e., 1 nW $\rightarrow$ 1 nA, at $\lambda_o = 1.24\ \mu\text{m}$. The linear increase of the responsivity with wavelength, for a given fixed value of η, is illustrated in Fig. 17.1-2. $\Re$ is also seen to increase linearly with η if λ_o is fixed. For thermal detectors $\Re$ is independent of λ_o because they respond directly to optical power rather than to the photon flux.

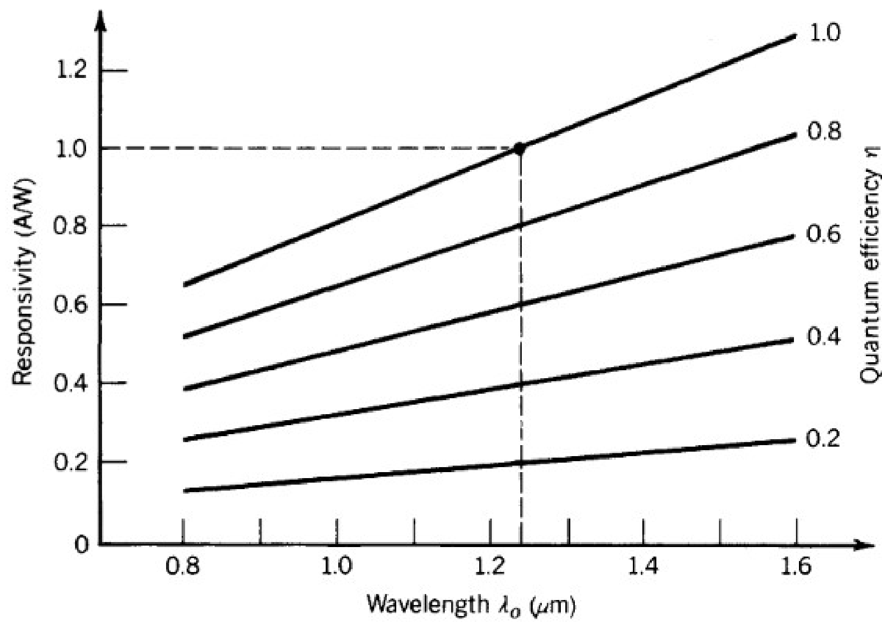


Figure 17.1-2 Responsivity $\Re$ (A/W) versus wavelength λ_o with the quantum efficiency η as a parameter. $\Re = 1$ A/W at $\lambda_o = 1.24\ \mu$m when $\eta = 1$.

Devices with Gain

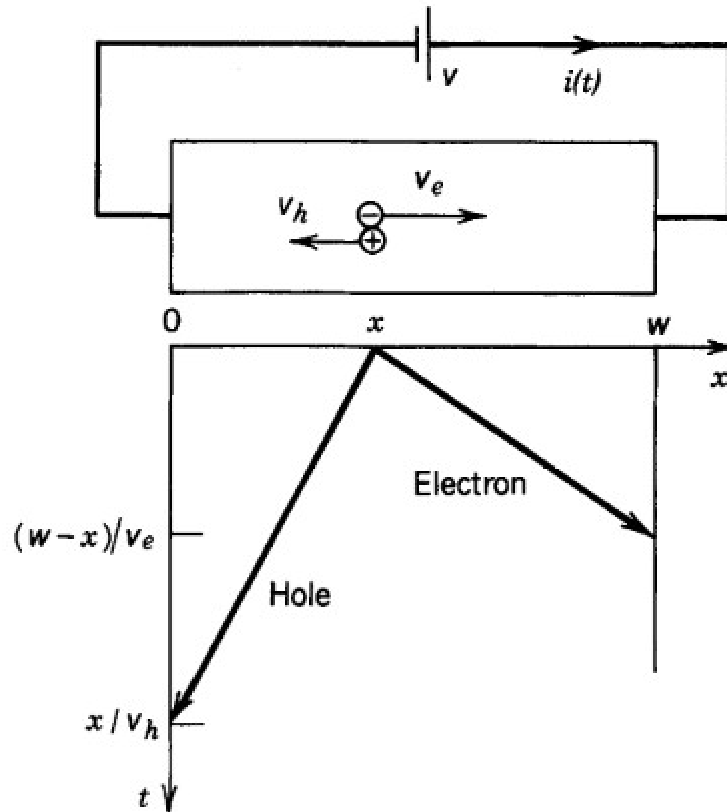
The formulas presented above are predicated on the assumption that each carrier produces a charge e in the detector circuit. However, many devices produce a charge q in the circuit that differs from e. Such devices are said to exhibit gain. The gain G is the average number of circuit electrons generated per photocarrier pair. G should be distinguished from η, which is the probability that an incident photon produces a detectable photocarrier pair. The gain, which is defined as

$$G = \frac{q}{e}, \tag{17.1-4}$$

can be either greater than or less than unity, as will be seen subsequently.

Response Time

One might be inclined to argue that the charge generated in an external circuit should be $2e$ when a photon generates an electron–hole pair in a photodetector material, since there are two charge carriers. In fact, the charge generated is e, as we will show below. Furthermore, the charge delivered to the external circuit by carrier motion in the photodetector material is not provided instantaneously but rather occupies an extended time. It is as if the motion of the charged carriers in the material draws charge slowly from the wire on one side of the device and pushes it slowly into the wire at the other side so that each charge passing through the external circuit is spread out in time. This phenomenon is known as **transit-time spread.** It is an important limiting factor for the speed of operation of all semiconductor photodetectors.

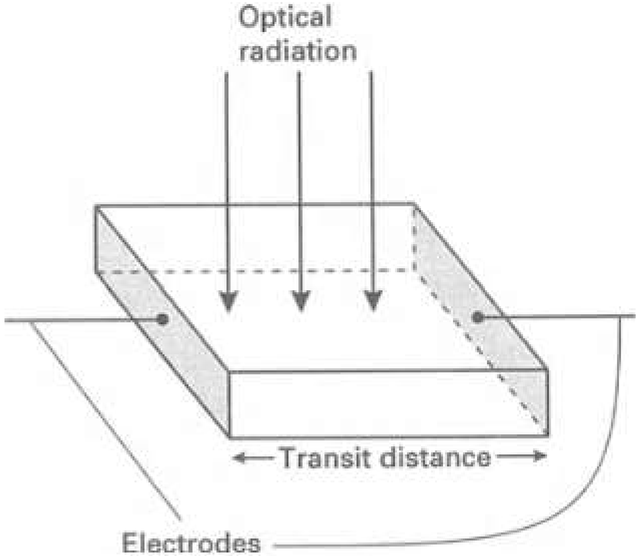


Another response-time limit of semiconductor detectors is the RC time constant formed by the resistance R and capacitance C of the photodetector and its circuitry.

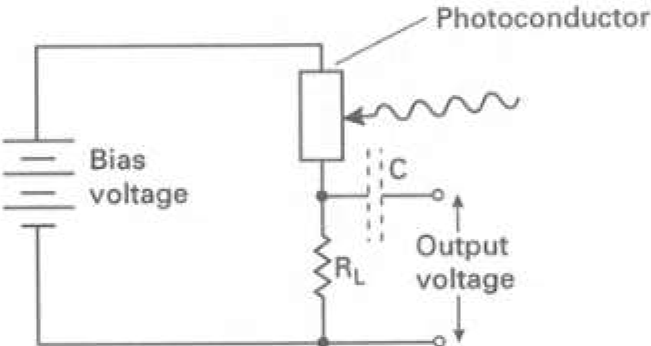
As a final point, we mention that photodetectors of a given material and structure often exhibit a fixed gain–bandwidth product. Increasing the gain results in a decrease of the bandwidth, and vice versa. This trade-off between sensitivity and frequency response is associated with the time required for the gain process to take place.

PHOTOCONDUCTIVE DETECTORS (PHOTOCONDUCTORS)

We know that the conductivity of a semiconductor depends upon the number of charge carriers in the conduction band. A photon, when absorbed by a semiconductor atom, is capable of raising an electron from the valence band to the conduction band (provided that its energy, $h\nu$, is greater than the band-gap energy) and thus of increasing the conductivity. In doing so, an electron hole pair is created; this will eventually be annihilated, since it represents a pair of charges in excess of the thermal equilibrium value. However, while the extra charge exists, it can be used to contribute to an electric current, whose value will depend upon the light power incident upon the semiconductor. Thus, a small voltage across the photoconductive material layer will give rise to the required, measurable current in an external circuit (Fig. 7.25).



(a) Slab of photoconductor



(b) Photoconductor detection circuit

The physical basis is shown in Figure 13.4, where a photon with sufficient energy can lift an electron from the valence band to the conduction band, creating two free-charge carriers which alter the conductivity of the material. Photoconductive detectors usually employ semiconductor materials with a bandgap suited to the wavelength of the light that is being detected.

Their use in optical communication systems, which presently use near-infrared light, is less common than *pin* diodes, but photoconductive detectors have seen renewed interest in recent years because they can have better responsivities than most other detectors.

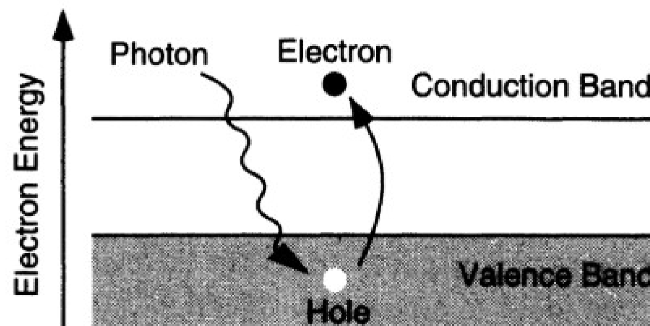


Figure 13.4 Optical absorption of a photon leads to the creation of a free electron and a free hole in an intrinsic semiconductor.

In thermal equilibrium, there are always some free carriers in a semiconductor. These holes and electrons contribute directly to the conductivity, σ, of the material

$$\sigma = nq\mu_e + pq\mu_h \tag{13.3}$$

where n and p are the concentration of free electrons and holes, respectively, and μ_e and μ_h are the mobilities of the electrons and holes, respectively, in the material. In semiconductors at room temperature, the mobilities are typically in the range of 10^2 to 10^4 $cm^2/V \cdot sec$, with the electron mobility usually being about a factor of 10 greater than the hole mobility.

The number of thermal carriers depends on the doping and on the generation rate of the material. The generation rate depends on the temperature, the number and type of lattice defects, and the size of the bandgap. The microscopic population of carriers is dynamic, even though the overall population may appear to be static. Electrons are constantly being excited to the conduction band via thermal excitation, where they exist for a short time before recombining with a hole. The equilibrium concentration of carriers determines the dark conductivity of the material.

From this simple arrangement it is straightforward to derive some simple, but important, relationships.

First, for the photon to yield an electron-hole pair its energy must satisfy $h\nu > E_g$, where E_g is the band-gap energy of the material. If, however, ν is too high, then all the photons will be absorbed in a thin surface layer, and the charge pairs will not be collected efficiently. Thus there is a frequency 'responsivity' spectrum for each type of photodetector which must be matched to the spectrum of the light which is to be detected.

Secondly, suppose that we are seeking to detect a light power of P watts at an optical frequency ν. This means that $P/h\nu$ photons are arriving every second. Suppose now that a fraction, η, of these produce electron hole pairs. Then there are $\eta P/h\nu$ charge carriers of each sign produced every second. If all are collected, and noting that each charge carrier only travels, on average, a distance of half the depletion width, and thus contributes, effectively, half the flow rate, the observed electric current is:

$$I = \frac{e\eta P}{h\nu} \tag{7.17}$$

Thus the current is proportional to the optical power. This means that the electrical power is proportional to the square of the optical power. It is important, then, when specifying the signal-to-noise ratio (SNR) for a detection process, to be sure about whether the ratio is stated in terms of electrical or optical power. (This is a fairly common source of confusion in the specification of detector noise performance.)

Another very important feature is speed of response. How quickly can the above processes respond to changes in input light level? Again some simple ideas establish the basic principles. Consider the electron created in the electron-hole creation process. Each electron will continue to contribute to the current flow until it is annihilated. If the average time for recombination of an electron-hole pair in the semiconductor is τ and the transit time across the layer is t_{tr}, then the average number of times the electron is used in the current, clearly, is τ/t_{tr}. Hence the current due to the light flux is

$$i_p = \frac{e\eta P}{h\nu}\frac{\tau}{t_{tr}} \tag{7.18}$$

However, the minimum time within which the photoconductor can respond to a change in light level is τ, so the bandwidth of the measurement is given by

$$\Delta f = \frac{1}{2\pi\tau} \tag{7.19}$$

We can define a responsivity for the device as the current produced for a given optical power. From (7.18) this is seen to be:

$$\mathscr{R} = \frac{i_p}{P} = \frac{e\eta\tau}{h\nu t_{tr}} \tag{7.20}$$

The factor τ/t_{tr} is generally higher than 1 (it can reach several hundreds) and is proven that it coincides with the *gain G* of the photoconductors. Therefore, the responsivity for a photoconductors can be written as:

$$\mathscr{R} = \frac{e\eta}{h\nu} G = \frac{e\lambda}{hc} \eta G \, ,$$

where $G = \tau/t_{tr}$. The value for τ, the recombination time, will depend upon the material: the transit time, instead, depends on the distance between the electrodes and will increase with them.

We will discuss about the gain with more details in the section related to the UV photoconductors.

There are two significant disadvantages possessed by photoconductive devices when compared with other types of photodetector. The first is that, with the absolute values which must be inserted in equation (7.18), high sensitivity (values of $\tau/t_{tr} \geq 10^3$, say) can only be achieved for response times ~ 50 ms. This is much too slow a response for most optoelectronic applications. The second is that the electron/hole recombination process is a random one, and thus one which introduces noise. This 'recombination noise', as it is called, is troublesome in many applications. An advantage, however, of these devices is that they can be used at long wavelengths, up to $\sim 20\,\mu$m. Since the band gap must be quite small for such low energy photons there is a problem with thermally created electron hole pairs which give rise to thermal noise. Hence such detectors usually must be cooled to, at least, liquid nitrogen temperatures (77 K).

Another advantage of these devices is that they are very cheap. For this reason they were often used in domestic cameras as CdS or CdSe cells.

Other materials which are commonly used are:

PbS	$1 - 3\,\mu$m wavelength
InS	$3 - 7\,\mu$m wavelength
HgCdTe	$5 - 14\,\mu$m wavelength.

The latter will be extensively studied in the section dedicated to the IR detectors.

Application Circuits

Photoconductors are usually used for detection of infrared radiation. When a bias is applied to the photoconductor in the absence of radiation, a current referred to as the dark current is generated. When light is incident on the photoconductor, its resistance decreases and the current flowing through it increases. Photosignal is the increase in the current caused by radiation. Generally this photosignal is much smaller (of the order of few parts per thousand) than the dark current. Extracting this small signal from the dark current is the primary task of the front-end circuit.

Figure 10.5a and b depict some simple circuits using photoconductors. Using photoconductors in these configurations however reduces the responsivity of the conductor as the relative change in the circuit resistance is smaller because of the load resistance R. The choice of R and R_{sen} also affects the output voltage from the circuit. For Figure 10.5a, the higher the value of R, the higher the output voltage but the relative responsivity is poorer. Similarly, in the case of circuit of Figure 10.5b, the higher the value of R, the lower the output voltage but the relative responsivity is better.

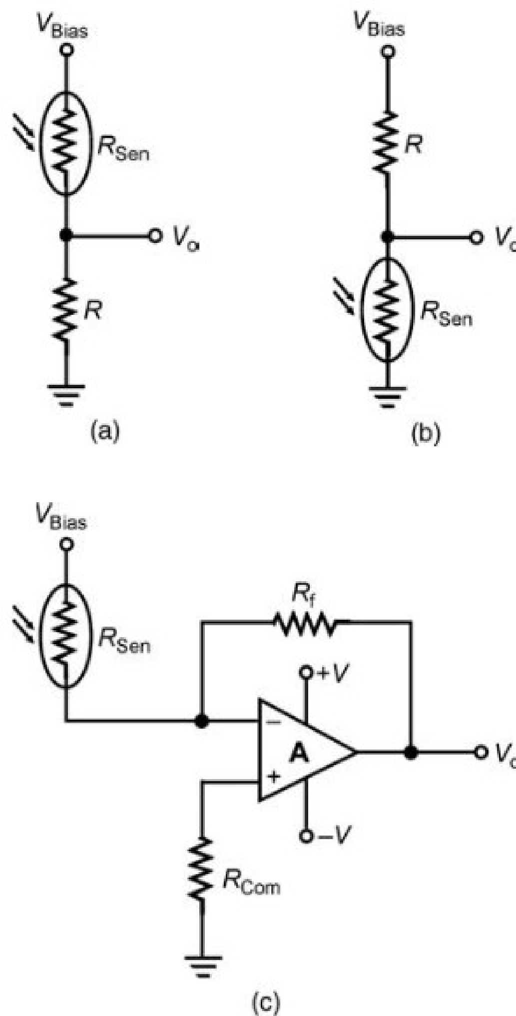


Figure 10.5 (a, b) Simplest application circuits using photoconductors; and (c) application circuit of photoconductor using opamp in the transimpedance mode.

To overcome these problems, photoconductors are used in conjunction with amplifiers to obtain both better responsivity and a high output voltage. There are two possible circuit configurations: voltage-mode amplifiers and current-mode or transimpedance amplifiers. The basic transimpedance amplifier is shown in Figure 10.5c. The non-inverting input of the operational amplifier (opamp) is connected to ground through resistance R_{com} to minimize the DC offset voltage.

The output voltage V_O is given by Equation 10.19:

$$V_O = -(R_f/R_{sen}) \times V_{bias} \tag{10.19}$$

The gain of the transimpedance amplifier should be set such that the amplifier does not saturate at the maximum expected radiation intensity. Also, if the bias voltage of the photoconductor is more than the maximum rated input voltage of the opamp, then a Zener diode should be connected between the inverting input of opamp and the ground terminal. Theoretically, the signal voltage can be obtained by subtracting the output voltage under dark conditions from the voltage signal in Equation 10.19, given by Equation 10.20:

$$V_O = -[(R_f/R_{sen}) - (R_f/R_{dark})] \times V_{bias} \tag{10.20}$$

where R_{dark} is the resistance value of the photoconductor in the absence of radiation.

Practically however, this is not a feasible solution as the dark resistance of the photoconductor is a strong function of temperature and even a slight increase in temperature decreases the value of dark resistance by a large amount and *vice versa*. The sensor temperature therefore has to be controlled to the

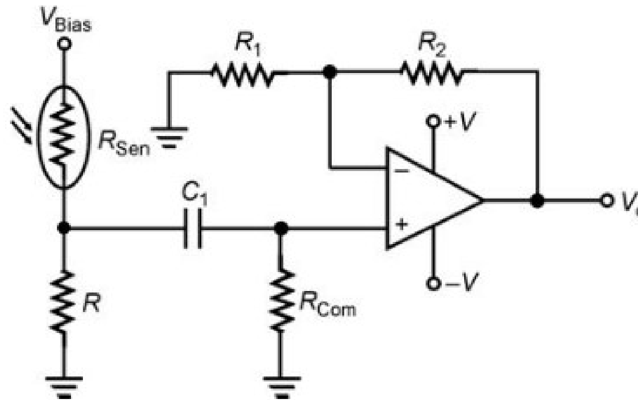


Figure 10.6 Application circuit of photoconductor using voltage mode amplifier with AC coupling.

order of 0.01 °C or better, which is often not feasible. The most common method used to extract the signal is to modulate the incident radiation at a specific frequency, either by placing a mechanical chopper in front of the sensor or by electrically modulating the radiation source. The signal generated due to radiation is now an AC signal while the dark current is a DC signal. The AC signal can be separated from the DC background signal using an AC coupled amplifier. A voltage-mode amplifier using AC coupling is shown in Figure 10.6.

Light Depending Resistor (LDR)

Photoresistors are the most typical application of the phenomenon of photoconductivity. An electromagnetic radiation of a suitable wavelength that penetrates inside a semiconductor crystal can give it the energy to allow some electrons to pass from the valence band to the conduction band. As a result, the electrical resistance of the semiconductor decreases, thanks to the increased density of free electrons. This phenomenon depends on:

- the semiconductor used;

- the concentration of doping impurities in the semiconductor;

- the wavelength of the incident radiation.

This physical phenomenon is exploited to fabricate *photoresistors*, or *Light Depending Resistors* (LDRs). For their realization, generally, the material used is a thin film of CdS (cadmium sulphide) which has the maximum sensitivity in the red range. It is deposited on an insulating support with two metal electrodes on it. To have a high sensitivity, the LDR must be able to capture a large number of photons, therefore in general has a sufficiently large surface. However, it should be kept in mind that the electron-hole pairs created by the light radiation tend to recombine, bringing the electrons back into the valence band. In order to reduce the recombination, it is then necessary to make them travel a distance as short as possible, so that there is not enough time for recombination. To obtain a high photoconductive surface and at the same time a short path for the charge carriers, a comb-shaped metal electrode structure is created. This is a good trade-off for the two problems described (see Fig. 8.12a).

Fig. 8.12 - LDR: a) Device picture; b) Circuit symbol

The relationship between the resistance and the amount of incident radiation is as follows:

$$R = A \cdot L^{-\alpha}$$

where R is the resistance value generated by lighting, L is the amount of the incident radiation measured in lux, A and α are constant. Note that this component is also used for the realization of *crepuscular switches* (devices that allow one to turn on one or more lights when the sun goes down).

2.9 The quantum well

There has been considerable interest recently in devices based on semiconductor structures in the form of a very thin slab, in which the thickness is less than about 10 nm, and very much less than the length and breadth of the slab. Quantum well structures may, for example, be formed by sandwiching a region of gallium arsenide between two gallium aluminium arsenide regions; the gallium arsenide is then bounded by two heterojunctions. If electrons are confined within such a structure then the energy levels they occupy can be obtained from eq. (2.13), we suppose that the slab thickness is L_z, and that the other dimensions are L_x and L_y, so that

$$E(n_1,n_2,n_3) = \frac{h^2}{8m_e^*}\left(\frac{n_1^2}{L_x^2} + \frac{n_2^2}{L_y^2} + \frac{n_3^2}{L_z^2}\right) \tag{2.64}$$

where n_1, n_2, n_3 are positive integers greater than zero. There is of course a corresponding expression for holes in the valence band.

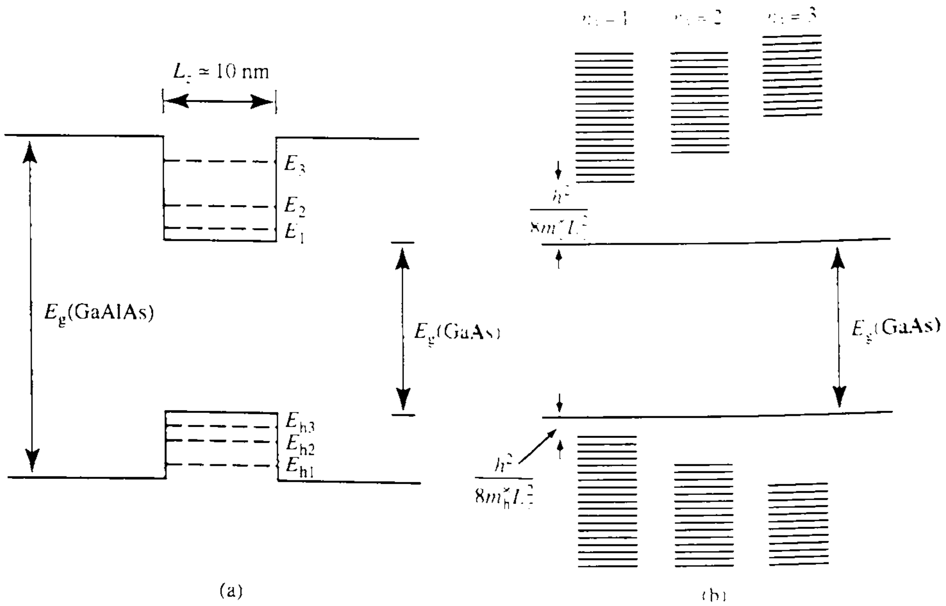


FIG. 2.32 Energy levels within a quantum well structure of GaAs sandwiched between two GaAlAs regions: (a) shows the discrete electron and hole energy levels for $n_3 = 1, 2, 3$ while (b) shows the sub-bands corresponding to these values of n_3 (they are separated for clarity, but in reality they are superimposed on one another).

Since L_z is so small compared with L_x and L_y, the energy levels will form groups, or sub-bands, of closely spaced levels determined by the various integral values of n_1 and n_2, for *each* value of n_3 as shown in Fig. 2.32. The density of states distribution in this case is no longer the smooth parabolic curve depicted in Fig. 2.14(a), but rather it has the 'staircase' structure shown in Fig. 2.33. There are a number of consequences of this situation. Firstly we see that the lowest energy states, corresponding to $n_1 = 1$, are situated at an energy approximately $h^2/8m_e^* L_z^2$ above the bottom of the conduction band, and correspondingly the uppermost states in the valence band are about $h^2/8m_h^* L_z^2$ below the top of the band. Thus the effective energy gap between the bands has increased by an amount

$$\Delta E_g = \frac{h^2}{8L_z^2}\left(\frac{1}{m_e^*} + \frac{1}{m_n^*}\right)$$

If we take $L_z = 10$ nm, then as Example 2.8 shows, ΔE_g is of the order of 0.1 eV in GaAs.

7.3.5.4 *Multiple quantum well detectors*

The introduction of techniques for fabricating multiple quantum well (MQW) structures (see section 2.9) has led to the development of some very interesting detectors based on them which operate in the far-IR region. Within a quantum well an optically induced transition can take place between the sub-bands provided that the transition is 'vertical' on the $E-k$ diagram (see section 4.6.1.1 and Fig. 4.13). That is, in the notation of section 2.9, a transition can take place between states of differing n_3 only if they have the same values of n_1 and n_2 (Fig. 7.24a). If we take an MQW structure and apply a field then the energy bands become 'tilted' and it is then possible for an electron in an excited state to tunnel through the potential barriers (Fig. 7.24b) and hence give rise to a current flow. A slightly different device can be made by using a well so narrow that only one sub-band is present within the well. Direct absorption can now take place between this state and the continuum states which lie above the top of the quantum well (Fig. 7.24c).

The sorts of transitions described here cover the wavelength range between about 5 µm and 20 µm (see Example 7.5). Although this range is already covered by the mercury cadmium telluride photoconductor there are considerable processing problems associated with making devices from this material.

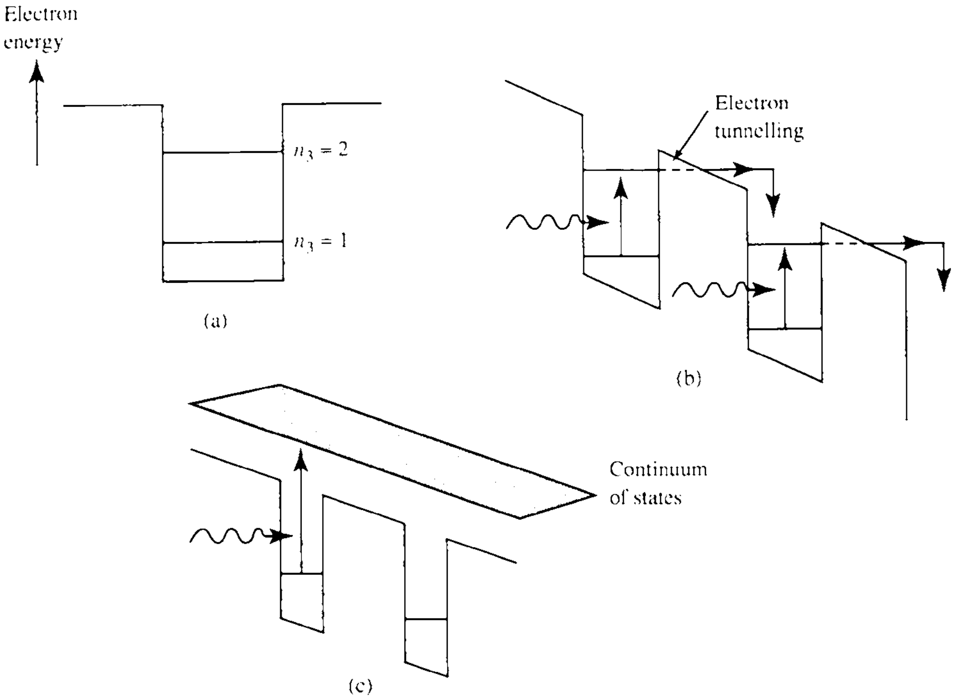


FIG. 7.24 (a) Electron states in a quantum well (no applied field); the dimensions are such that only two levels are present (quantum number n_3 = 1and 2). With an applied field we may observe photoconductivity due to (b) transitions between n_1 and n_2 followed by interwell tunnelling or (c) transitions to the electron continuum of states.

EXAMPLE 7.5 Detection wavelengths in quantum well detectors

According to eq. (2.64) the energy difference between the $n_3 = 1$ and $n_3 = 2$ states in a quantum well (n_1 and n_2 remaining constant) is given by

$$\Delta E = \frac{h^2}{8m_e^*}\left(\frac{2^2}{L_z^2} - \frac{1^2}{L_z^2}\right)$$

Here we have replaced m by m_e^* to allow for the fact that the well is within a semiconductor material and is not simply in a vacuum. Assuming a well 10 nm wide ($=L_z$) of GaAs (where $m_e^* = 0.068m$) we then obtain $\Delta E = 0.165$ eV. The corresponding optical wavelength is then given by $hc/\Delta E$, that is 7.5 μm.

7.3.6 Junction detectors

7.3.6.1 The p–n junction detector

When a p–n junction is formed in a semiconductor material, a region depleted of mobile

charge carriers is created which has a high internal electric field across it (see the discussion in section 2.8.1). If an electron–hole pair is generated by photon absorption within this region the electric field separates the electron and hole as shown in Fig. 7.25.

We may detect the charge separation in three distinct ways. First, if the device is left on open circuit (or with a very high resistance between the external contacts) an externally measurable potential will appear between the p and n regions; this is the *photovoltaic* mode of operation. Secondly, in the *photoamperic* mode a very low external resistance is connected between the external contacts and a photogenerated current flows through it. Finally the most usual way to operate the device is in the *photoconductive* mode where a reverse bias is applied across the junction and the resulting current flow through an external load resistor measured. The load resistor in this case need not be as small as in the photoamperic mode.

The junction will also respond to electron–hole pairs generated away from the depletion region provided they are able to diffuse to the edge of the depletion region before recombination takes place. From the discussion in Chapter 2, it is evident that only carriers generated within a minority carrier diffusion length or so of the edge of the depletion region are likely to be able to do this; nevertheless, this does increase the sensitive volume of the device.

In operation, we may represent the photodiode by a constant current generator (the current flow i_λ being generated by light absorption) with an ideal diode across it (to simulate the effect of the p–n junction), as shown in Fig. 7.26. The internal characteristics of the cell may be better modelled by the introduction of a shunt resistor (R_{sh}), a shunt capacitor (C_d) and a series resistor (R_s). If we assume a fraction η of the incident radiation is absorbed within the cell, then we may write

$$i_\lambda = \frac{\eta I_0 A e \lambda_0}{hc} \tag{7.25}$$

where I_0 is the light irradiance falling on a cell of area A.

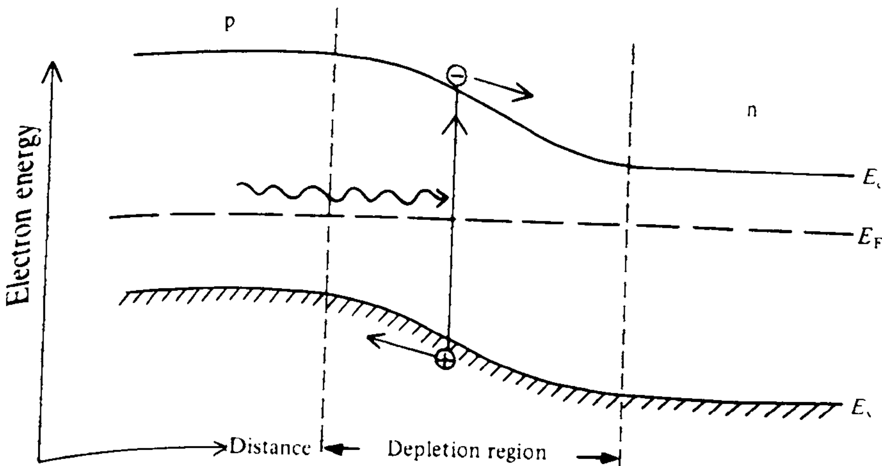


FIG. 7.25 Electron energy level diagram illustrating the generation and subsequent separation of an electron–hole pair by photon absorption within the depletion region of a p–n junction.

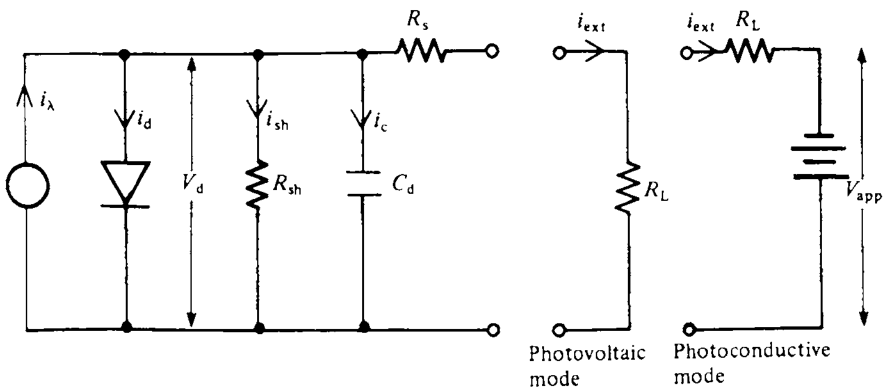


FIG. 7.26 Photodiode equivalent circuit. In operation, the photodiode can be represented by a photogenerated current source i_λ feeding into an ideal diode. The internal cell characteristics are better modelled by the introduction of a shunt resistor (R_{sh}), a shunt capacitor (C_d) and a series resistor (R_s). In the photovoltaic mode, an external high value resistor (R_L) is connected across the output and the voltage across this measured. In the photoconductive mode, an external bias (V_{app}) is applied in conjunction with a series load resistor (R_L). The current flowing through R_L is monitored by measuring the voltage across it.

In the photoconductive mode, a relatively large reverse bias (≈10 V or more) is usually applied across the diode (Fig. 7.26). Since the diode current saturates at i_0 for relatively small values of reverse bias (i.e. a few tenths of a volt), eq. (7.26) can be written as

$$i_\lambda = i_0 + i_{sh} + i_{ext}$$

As in the photovoltaic mode we may assume that both i_0 and i_{sh} are much less than i_λ and hence we may write $i_{ext} = i_\lambda$.

Substituting from eq. (7.25) we then have

$$i_{ext} = \frac{\eta I_0 A e \lambda_0}{hc} \tag{7.30}$$

With an external load resistor of R_L, the output voltage will then be

$$V_{ext} = \frac{\eta I_0 A e \lambda_0}{hc} R_L \tag{7.30a}$$

Hence, in the photoconductive mode the external current flowing is directly proportional to the incident light irradiance.

Finally, in the photoamperic mode, the situation is somewhat similar to that in the photoconductive mode: provided the external resistance is sufficiently small, almost all of the photogenerated current will flow through it and hence the output current and voltage will be the same as eqs (7.30) and (7.30a) respectively. However, if the external voltage ($=R_L \times i_{ext}$) becomes too large, the internal diode may become sufficiently forward biased to divert an appreciable amount of current from the external circuit; this will lead to a non-linear response.

The photoconductive mode thus offers the advantage of an inherently linear response over a relatively wide dynamic range. It also offers a more rapid response than the photovoltaic mode. The main drawback is the presence of a dark current ($i_0 + i_{sh}$) which, as in the photomultiplier, gives rise to shot noise (section 7.3.3.2) and limits the ultimate sensitivity of the device. All modes of operation are subject to generation noise (section 7.3.5.1), but recombination noise is absent since the charge carriers are separated in the depletion region before they can recombine.

A typical structure for a p–n diode junction detector based on silicon is shown in Fig. 7.27. Contact to the semiconductor material is made via a metal–n^+ (or $-p^+$) junction; this is found to be the most convenient way of providing an ohmic contact (see section 2.8.6). If we assume an abrupt junction together with the external bias, V, being much larger than the internal junction potential, V_0, and also that $N_a \gg N_d$, then the depletion layer widths can be written as (see eqs 2.60, 2.59 and 2.55)

$$x_n = \left(\frac{2\varepsilon_0\varepsilon_r V}{eN_d}\right)^{1/2} \quad \text{and} \quad x_p = \left(\frac{2\varepsilon_0\varepsilon_r V N_d}{eN_a^2}\right)^{1/2} \tag{7.31}$$

Since we have a p^+–n structure it follows that $x_n \gg x_p$; Fig. 7.28 illustrates the resulting electric field variation within the depletion regions. For efficient detection the electron–hole pairs should be generated within the depletion region. At short wavelengths, where the absorption

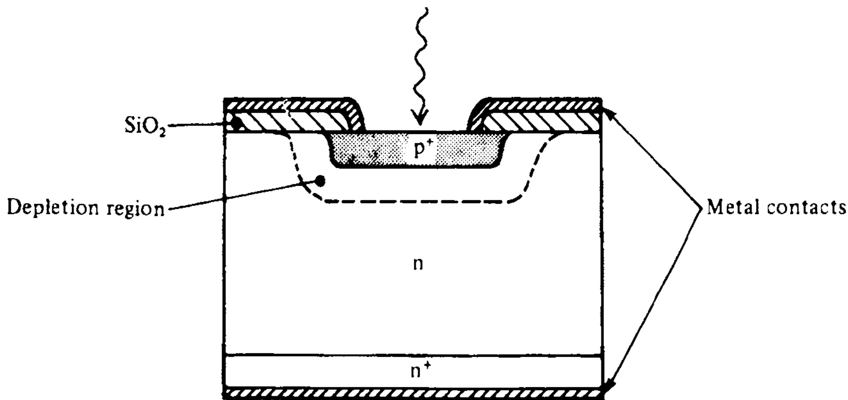


FIG. 7.27 Typical silicon photodiode structure for photoconductive operation. A junction is formed between heavily doped p-type material (p^+) and fairly lightly doped n-type material so that the depletion region extends well into the n material. The p^+ layer is made fairly thin. Metallic contacts can be made directly to the p^+ material but to obtain an ohmic contact to the n material an intermediate n^+ layer must be formed.

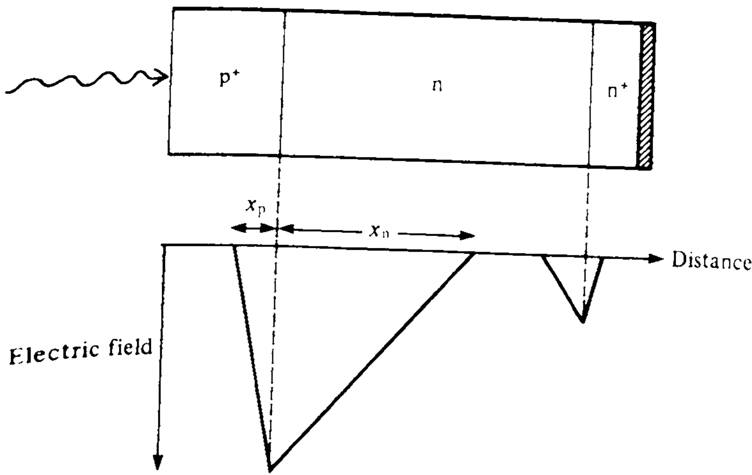


FIG. 7.28 Electric field distribution within the p^+–n junction diode shown in Fig. 7.27, assuming an abrupt diode structure.

coefficient is relatively high (Fig. 7.18), they will be generated close to the surface. Consequently, to achieve a good short wavelength response the p^+ region should be made as thin as possible. Conversely, at the upper wavelength range of the detector the absorption coefficient will be relatively small and a wide depletion region is required for high detection efficiency. This implies that large values of the reverse bias voltage are needed (see Example 7.6) which may approach or exceed the diode breakdown voltage. Detection efficiency may also be improved by providing an antireflection coating to the front surface of the detectors consisting of a $\lambda/4$ thick coating of SiO_2.

7.3.6.2 *The p–i–n photodiode*

A structure that results in a good long wavelength response with only relatively modest bias levels is the so-called p–i–n (or PIN) structure, illustrated in Fig. 7.29. Here the intrinsic (i)

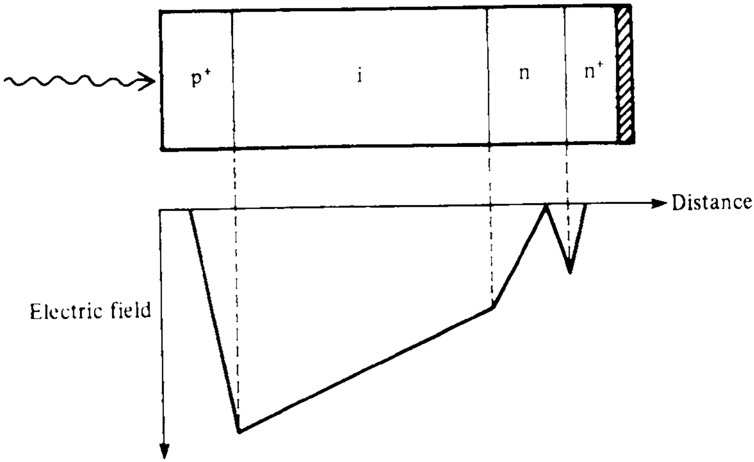


FIG. 7.29 Electric field distribution within a p–n structure.

region has a high resistivity (i.e. low values of N_a and N_d) so that only a few volts of reverse bias are needed to cause the depletion region to extend all the way through to the n region and thus provide a large sensitive volume. In practice, the bias is maintained at a considerably higher voltage than the minimum value and the intrinsic region then remains fully depleted of carriers even at high light levels. The depletion region width in a p–i–n structure is then practically independent of applied voltage and thus much better delineated than in a p–n structure where the depletion region width will vary appreciably with applied voltage. For this reason most simple photodiode structures are of the p–i–n rather than the p–n type.

For efficient detection of photons we require that as many as possible are absorbed within the intrinsic region (assuming a p–i–n structure).

Silicon p–i–n photodiodes can achieve quantum efficiencies of up to 80% in the wavelength range 0.8–0.9 μm. A typical spectral response of a p–i–n silicon photodiode is shown in Fig. 7.30.

The problem of low detector efficiencies at wavelengths close to the bandgap wavelength can also be addressed by using detectors that are illuminated from the side (i.e. parallel to the junction) although these are not commonly available.

7.3.6.3 *Photodiode materials*

For the detection of radiation in the visible and near-IR region one of the most popular

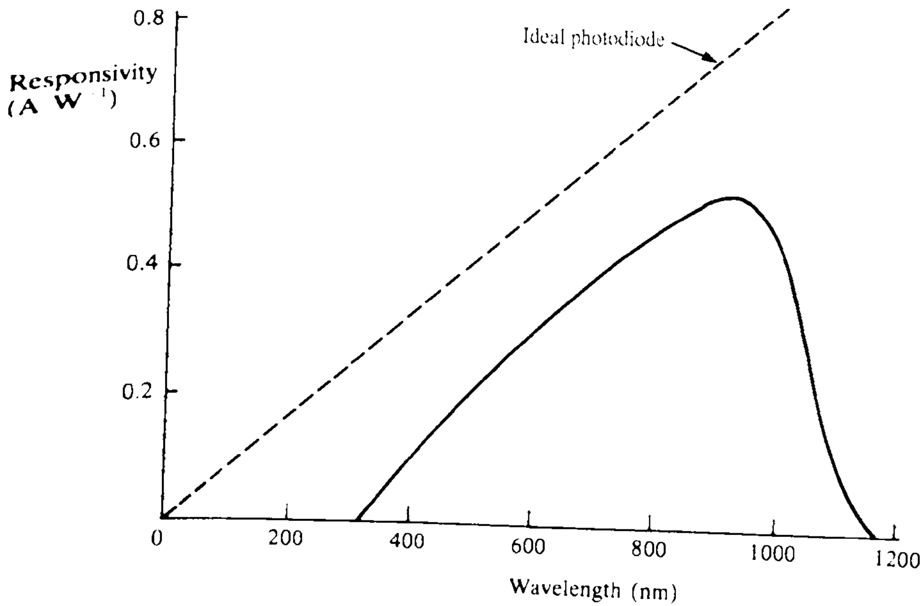


FIG. 7.30 Typical current responsivity of a silicon photodiode. Also shown is the responsivity of an ideal photodiode with unit quantum efficiency.

photodiode materials is silicon. This has an energy gap of 1.14 eV and, as mentioned above, can achieve quantum efficiencies of up to 80% between 0.8 μm and 0.9 μm. An examination of Fig. 7.18 shows that at these wavelengths the absorption coefficient is about 10^5 m^{-1}. Thus to obtain efficiencies as high as 80% we require an intrinsic layer thickness of about 20 μm and also a surface p$^+$ layer thickness that is substantially less than α^{-1}, or 10 μm.

As we shall see in Chapter 9, the development of optical fiber communication systems has led to the demand for detectors operating at wavelengths of 1.3 μm and 1.55 μm which have high sensitivities and exceptionally wide bandwidths. Photodiodes based on germanium are available (the bandgap wavelength for Ge is 1.88 μm) but they suffer from rather low sensitivities and large reverse bias leakage currents, this latter being a general problem in small bandgap semiconductors. Much more successful have been materials that can be grown on InP substrates such as the ternary compound $In_xGa_{1-x}As$ and the quaternary compound $Ga_xIn_{1-x}As_yP_{1-y}$. The materials must be lattice matched to InP which, in the case of $In_xGa_{1-x}As$, takes place when $x = 0.53$. The bandgap is then 0.74 eV which corresponds to a bandgap wavelength of 1.68 μm. Unfortunately when homojunctions are made from such narrow bandgap materials they display relatively low breakdown voltages and large reverse bias leakage currents. In fact by taking eq. (2.51a) and using the relationships $p_n \times n_n = n_i^2 = n_p \times p_p$ together with eq. (2.36), it can be shown that $J_0 \propto \exp(-E_g/kT)$. Consequently instead of homojunctions, it is usual to form heterojunctions with a wider bandgap material. An example of such a structure is shown in Fig. 7.31, where an i layer of $In_{0.53}Ga_{0.47}As$ is sandwiched between a p$^+$ layer of InP and an n layer of InP. Because the radiation has to pass through a layer of InP which has a bandgap wavelength of 0.92 μm,

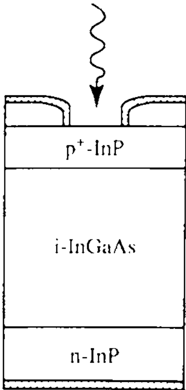


FIG. 7.31 Basic structure of a p–i–n heterojunction InGaAs detector.

then no radiation below this wavelength will be detected. A further advantage of this structure is that since no carriers of interest will be generated in either the surface p^+ layer or the lower n layer there will be no diffusion of carriers to the junction from outside the depletion region, which, as we shall see in the next section, can worsen the response time of the detector.

From Fig. 7.18 we can see that the absorption coefficient for $In_{0.53}Ga_{0.47}As$ close to the bandgap wavelength is higher than in the case of silicon. This is essentially because $In_{0.53}Ga_{0.47}As$ is a direct bandgap semiconductor whereas silicon is an indirect bandgap semiconductor. As a consequence the intrinsic layer thickness can be much smaller than in the case of silicon. At 1.5 μm, for example, the absorption coefficient of $In_{0.53}Ga_{0.47}As$ is about 10^6 m^{-1} and thus to achieve the condition $w_i = 2\alpha^{-1}$ we require an intrinsic layer thickness of the order of only 2 μm.

Phototransistors

A phototransistor is essentially fabricated as a normal BJT. The difference is that the base region can be exposed to incident radiation. Therefore, as in a photodiode, the package is provided with a transparent lens which allows the base to be illuminated. Normally, the base is left floating but in some circuits it can also be connected externally. The operation of the phototransistor is similar to that of the photodiode, but the resulting current in this case is much higher than that of a photodiode, as it experiences the multiplicative effect typical of the transistor. In fact, by biasing the transistor so that the CB junction is inversely polarized, the output current flowing from the collector to the emitter (in an *npn* device) will be given by the base current i_b (plus the reverse current of thermal origin due to the reverse bias of the CB junction) multiplied by the gain of the transistor β. If we consider that i_b is a photocurrent generated in the base region, this means that the output emitter current i_e is about β times larger than that of a photodiode. Fig. 8.11 shows the circuit symbol of an *npn* phototransistor, inserted in a typical operating circuit. The light absorbed by the base generates an emitter current which flows on the load resistance R_L and produces a voltage drop (V_{out}) on it.

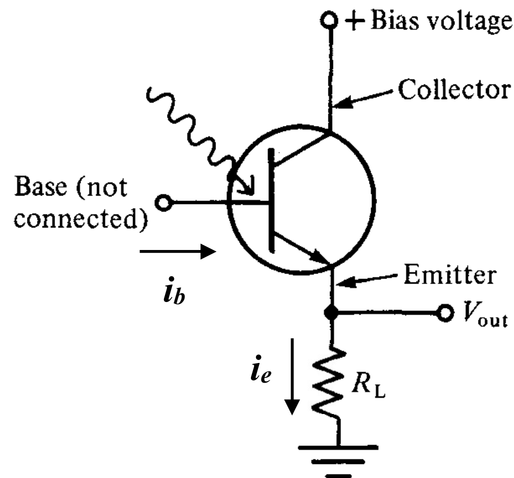


Fig. 8.11 - Circuit symbol of an *npn* phototransistor and its external connections

The sensitivity of the device can be adjusted by connecting a variable resistance between base and mass, but usually the base is left floating to maximize light sensitivity. The trade-off for a better sensitivity is a slower speed. Phototransistors are more sensitive than photodiodes, but their response times are significantly slower, due to the time constant associated with the capacitance and resistance of the base region. The typical output current of a photodiode is some μA while the saturation-interdiction switching times are of the order of some ns. In a phototransistor, on the other hand, the output current is a few mA but the switching times are of the order of μs.

7.3.6.6 *Avalanche photodiodes*

As explained above, most fast photodiodes are designed for use with a 50 Ω load impedance and the voltage output ($= i_\lambda R_L$) often requires considerable amplification. Useful internal amplification of the photocurrent is achieved in the avalanche photodiode (APD). In this device, a basic p–n structure is operated under very high reverse bias. Carriers traversing the depletion region therefore gain sufficient energy to enable further carriers to be excited across the energy gap by impact excitation. The process is illustrated in Fig. 7.36.

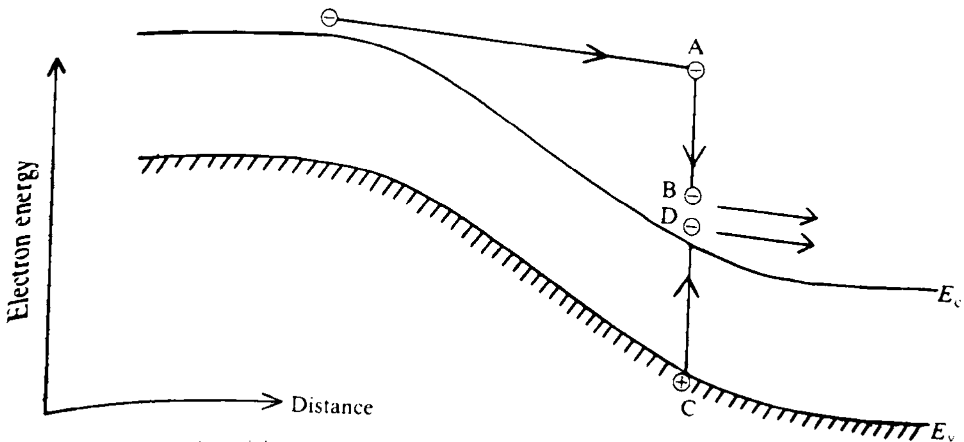


FIG. 7.36 Illustration of the principle of operation of an avalanche photodiode. An electron having reached the point A has sufficient energy above the conduction band bottom to enable it to excite an electron from the valence band into the conduction band (C → D). In so doing it falls from A to B.

An electron having reached point A on the diagram has enough energy above the conduction band bottom such that it can collide with an electron from the valence band and raise it to the conduction band (C $\longrightarrow$ D) (the minimum energy required to initiate this process was discussed in section 4.3 when dealing with cathodoluminescence). This generates a new electron–hole pair; in so doing, the electron will of course lose an equivalent amount of energy and move from A to B. The newly generated carriers may both subsequently generate further electron–hole pairs by the same process. The probability that an ionizing event takes place is governed by the carrier densities and also the *ionization coefficients*. The latter vary rapidly with electric field $\mathscr{E}$ according to a relation of the form $\exp(-A/\mathscr{E})$ where A is a constant. In addition it should be noted that the magnitudes of the ionization coefficients can differ depending on the type of carrier. In silicon, for example, the ionization coefficient for electrons is considerably greater than that for holes, whilst in germanium the two are almost identical.

Current gains in excess of 100 are readily achievable. However, as shown in Fig. 7.37, the current gain is very sensitive to the value of the bias voltage, and if the bias voltage is made too large there is the danger of creating a self-sustaining avalanche current that flows in the absence of any photoexcitation, which sets an upper limit on the voltage that may be used. In fact non-uniformities within the device can cause small regions of premature breakdown known as *microplasmas*. Obviously the device requires a very stable voltage supply and, in addition, since the avalanche process itself depends on temperature, temperature stabilization is required. Another problem that can occur is that of excessive leakage currents at the junction edges. In silicon APDs this latter problem can be addressed by the use of a *guard ring* as shown in Fig. 7.38. This also serves to restrict the avalanche region to the central illuminated part of the cell and thus helps to reduce premature breakdown. Unfortunately the guard ring also increases the capacitance of the device thus restricting the high frequency performance.

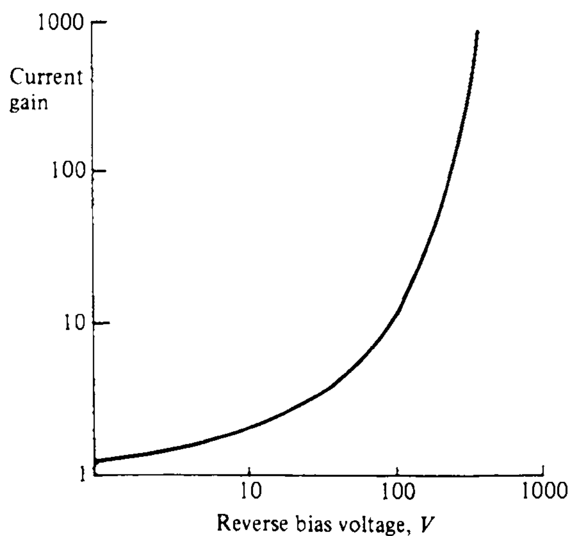


FIG. 7.37 Typical variation of current gain with reverse bias voltage for an avalanche photodiode.

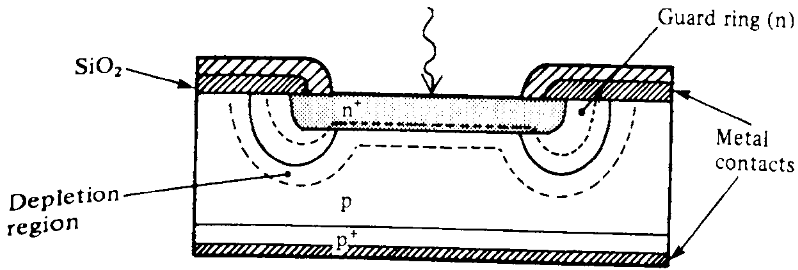


FIG. 7.38 Silicon avalanche photodetector with guard-ring structure. The guard ring is a region of comparatively low doping, and hence the depletion region extends an appreciable distance into it. Thus, in the vicinity of the guard ring the total depletion layer thickness is greater, and hence the maximum electric field strengths are less, than in the central region.

5.4 CHARGE-COUPLED DEVICES (CCD)

A schematic view of a CCD is shown[5] in Fig. 18. The basic device consists of a closely spaced array of MOS capacitors on a continuous insulator (oxide) layer that covers the semiconductor substrate. A CCD can perform a wide range of electronic functions, including image sensing and signal processing. The operating principle of the CCD involves the charge storage and transfer actions controlled by the gate electrodes.

Figure 18*a* shows a CCD to which sufficiently large positive bias pulses have been applied to all the electrodes to produce surface depletion. A slightly higher bias has been applied to the central electrode so that the center MOS structure is under greater depletion and a potential well is formed there; i.e., the potential distribution is shaped like a well because of the larger depletion-layer width under the central electrode. If minority carriers (electrons) are introduced, they will be collected in the potential well. If the potential of the right-hand electrode is increased to exceed that of the central electrode, we obtain the potential distribution shown in Fig. 18*b*. In this case, the minority carriers will be transferred from the central electrode to the right-hand electrode. Subsequently, the potential on the electrodes can be readjusted so that the quiescent storage site is located at the right-hand electrode. By continuing this process, we can transfer the carriers successively along a linear array.

CCD Shift Register

Figure 19 shows more details about the basic principle of charge transfer in a three-phase, *n*-channel CCD array. The electrodes are connected to the ϕ_1, ϕ_2, and ϕ_3 clock lines. Figure 19*b* shows the clock waveforms and Fig. 19*c* illustrates the corresponding potential wells and charge distributions.

At $t = t_1$, clock line ϕ_1 is at a high voltage and ϕ_2 and ϕ_3 are at low voltages. The potential wells under ϕ_1 will be deeper than the others. We assume that there is a signal charge at the first ϕ_1 electrode. At $t = t_2$, both ϕ_1 and ϕ_2 have high bias as charge starts to transfer. At $t = t_3$, the voltage at ϕ_1 is returning to the low value while ϕ_2 electrodes are still held at high voltage. The electrons stored under ϕ_1 are being emptied in this period. At $t = t_4$, the charge transfer is complete and the original charge packet is now stored under the first ϕ_2 electrode. This process will be repeated and the charge packet continues to shift to the right. CCDs can be operated with two, three, or four phases, with different design structures. Multiple electrode structures and clocking schemes have been proposed and implemented.[6]

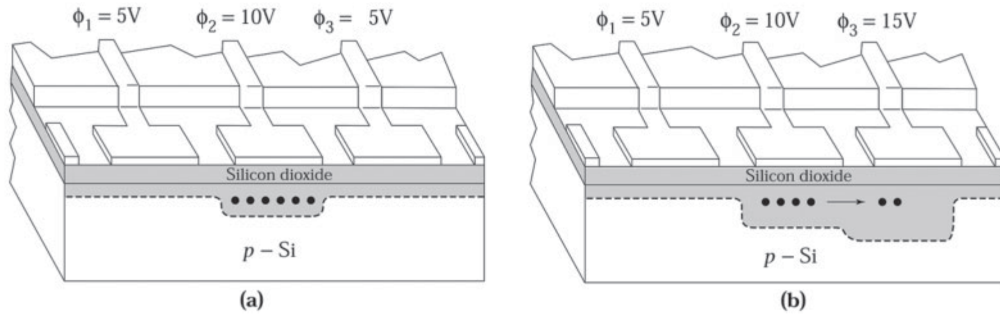


Fig. 18 Cross section of a three-phase charge-coupled device.[5] (*a*) High voltage on ϕ_2. (*b*) ϕ_3 pulsed to a higher voltage for charge transfer.

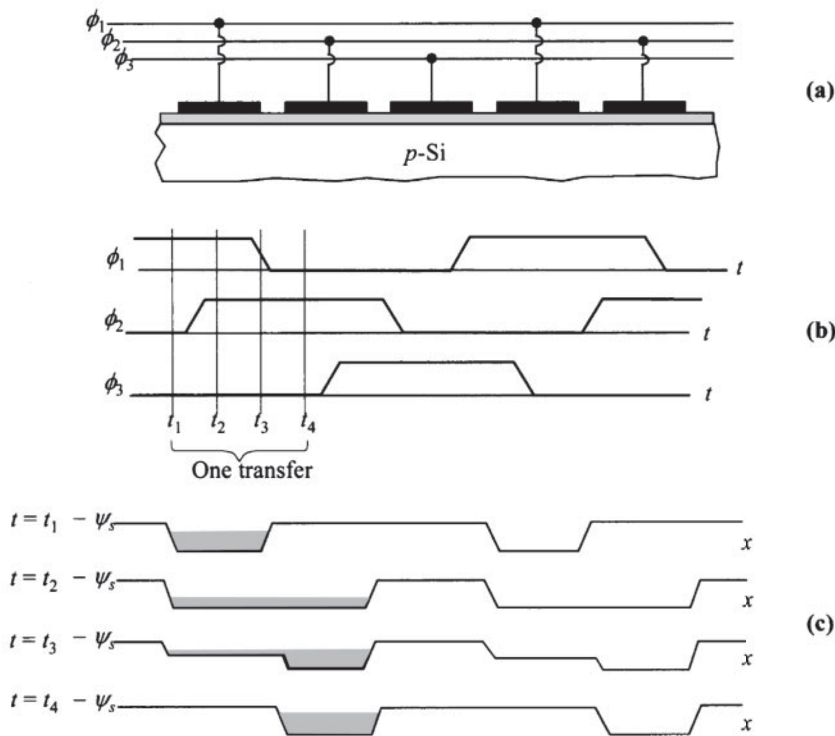


Fig. 19 Illustration of CCD charge transfer. (*a*) Application of three-phase gate bias. (*b*) Clock waveforms. (*c*) Surface potential (and charge) vs. distance at different times.[6]

CCD Image Sensor

For analog and memory devices, the charge packets are introduced by injection from a *p-n* junction in the vicinity of the CCD. For optical imaging applications, the charge packets are formed as a result of electron-hole pair generation caused by incident light.

When CCD used in imaging array systems such as a camera or video recorder, CCD image sensors must be spaced c1osely to one another in a chain and function as shift registers to transport the signals. The structure of the surface-channel CCD image sensor is similar to that of the CCD shift register, with the exception that the gates are semitransparent to let light pass through. Common materials for the gates are metal, polysilicon, and silicide.

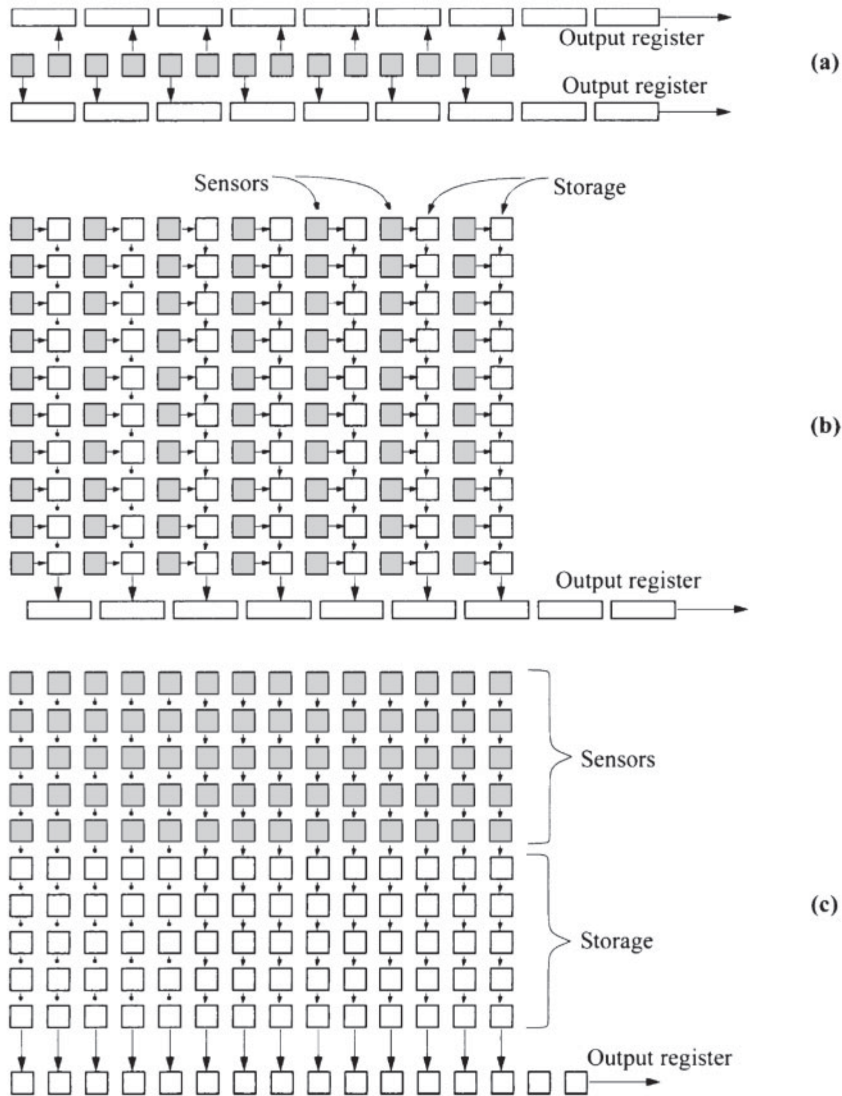


Fig. 20 **Schematic layouts showing the readout mechanisms of (*a*) line imager with dual output registers, and area imagers with (*b*) interline transfer and (*c*) frame transfer. Gray pixels represent CCDs as photodetectors. The output register is usually clocked at higher frequency than the internal transfer.**

Alternatively, the CCD can be illuminated from the back of the substrate to avoid light absorption by the gate. In this configuration, the semiconductor has to be thinned so that most of the light can be absorbed within the depletion region at the top surface.

Because the CCDs can also be used as a shift registers, there is great benefit to using CCDs as photodetectors in an imaging-array system since the signals can be brought out sequentially to a single node, without complicated x-y addressing to each pixel. The photogenerated carriers are integrated during light exposure, and the signal is stored in the form of a charge packet, to be transported and detected later. The detection mode of the integrated charge over a long period of time enables detection of weaker signals. In addition, the CCDs have the advantages of low dark current, low noise, low-voltage operation, good linearity, and good dynamic range. The structure is simple, compact, stable, and robust, and is compatible with MOS technology. These factors contribute to high yield, which makes the CCDs desirable in consumer products.

Different readout mechanisms for the line imager and the area imagers are shown in Fig. 20.[4] A line imager with dual output registers has improved readout speed (Fig. 20*a*). Most common area imagers use either interline-transfer (Fig. 20*b*) or frame-transfer (Fig. 20*c*) readout architecture. In the former case, signals are transferred to the neighboring pixels and are subsequently passed along to the output register chain while the light-sensitive pixels start to collect a charge for the next data. In the frame-transfer scheme, signals are shifted to a storage area away from the sensing area. The advantage of this over the interline transfer is a more efficient light-sensing area, but there is more image smear since CCDs continue to receive light as signal charges are passed through them. For both interline transfer and frame transfer, all columns advance their charge signals to the horizontal output register simultaneously, and the output register carries these signals out at a much higher clocking rate.