

CHAPTER TWENTY TWO

Detection of Optical Radiation

Detection of Optical Radiation

22.1 Introduction

In this chapter we will review some of the fundamentals of the optical detection process. This will include a discussion of the randomly fluctuating signals, or noise that appear at the output of any detector. We will then examine some of the practical characteristics of various types of optical detector. The chapter will conclude with a discussion of the limiting detection sensitivities of important detectors used in various ways.

Photon detectors operate by absorbing the photons coming from a source and using the absorbed energy to produce a change in the electrical characteristics of their active element(s). This can occur in many ways. In a photomultiplier or vacuum photodiode the incoming photons are absorbed in a photoemittive surface and through the photoelectric effect free electrons are produced. These electrons can be accelerated and detected as an electrical current. In a semiconductor photodiode or photovoltaic detector, absorption of a photon at a p-n or p-i-n junction creates an electron-hole pair. The electron and hole separate because of the energy barrier at the junction | each carrier moves to the region where it can reduce its potential energy, as shown in Fig. (22.1).

Thermal radiation detectors use the heating effect produced by absorbed photons to change some characteristics of the detector element. In a bolometer, the heating of carriers changes their mobility and the resistivity of the detector element. In a thermopile the heating effect is used to generate a voltage through the thermoelectric (Seebeck) effect.

Fig. (22.1).

Pyroelectric detectors utilize the change in surface charge that results when certain crystals (ones that can possess an internal electric dipole moment) are heated.

22.2 Noise

22.2.1 Shot Noise

The ability of a photodetector to detect an incoming light signal is limited by the intrinsic fluctuations, or noise, both of the incoming light itself and of the background electrical current fluctuations generated by the detector. In the photon description of a monochromatic light beam an incoming beam of intensity $\langle I \rangle$ has an average photon flux associated with it of $\langle N \rangle$ photons/m² where

$$\langle N \rangle = \frac{\langle I \rangle}{h\nu} \quad (22:1)$$

If the rate of arrival of photons at the detector is examined closely it will be observed to fluctuate about this average value. Strictly speaking, all we can ever really observe is the rate of appearance of photo-electrons or carriers produced by the disappearance of photons at the active surface of the detector. We assume that these events are directly related by a quantum efficiency factor η , where

$$\eta = \frac{\text{number of carriers produced}}{\text{number of absorbed photons}} \quad (22:2)$$

The fluctuations in photon flux, N , give rise to a fluctuation in the rate of production of photo-produced carriers. These fluctuations appear as a randomly varying current superimposed on the average current. The

average current from the photodetector is

$$\bar{i} = \frac{e}{h\nu} \langle I \rangle A; \quad (22:3)$$

which can be written as

$$\bar{i} = R \langle I \rangle A; \quad (22:4)$$

where A is the detector area, and $R = e/h\nu$ is called the responsivity, which has units of A/watt. The fluctuations in current resulting from the fluctuating appearance of photo-produced carriers is called shot, or photon noise. The frequency spectrum of these current fluctuations can be calculated by considering the frequency components in the current produced by the elemental current of a single photo-produced carrier. If we take one of these elemental current pulses to be of a Gaussian shape we can write

$$i(t) = \frac{e}{\sqrt{2\pi} \tau} e^{-t^2/2\tau^2}; \quad (22:5)$$

where the shape of the pulse is normalized so that

$$\int_{-\infty}^{\infty} i(t) dt = e; \quad (22:6)$$

The charge on an electron is e .

The frequency spectrum of this current is

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} i(t) e^{i\omega t} dt; \quad (22:7)$$

which gives

$$F(\omega) = \frac{e}{\sqrt{2\pi}} \exp\left[-\frac{\omega^2 \tau^2}{2}\right]; \quad (22:8)$$

For sufficiently low frequencies, for example such that $\omega^2 \tau^2 \ll 2$, the exponential factor is very close to unity, and the Fourier transform is flat. Simply stated, this result says that randomly occurring narrow pulses contribute broadband "white" noise.

In terms of its Fourier transform the current can be written as

$$i(t) = \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega; \quad (22:9)$$

If this current flows into a resistance R , then the average power dissipated over time T , where T is greater than the pulse duration, is

$$P = \frac{1}{T} \int_{-T/2}^{T/2} \frac{i^2(t)}{R} dt = \frac{1}{RT} \int_{-\infty}^{\infty} i(t) F(\omega) e^{i\omega t} d\omega; \quad (22:10)$$

which can be rearranged as

$$P = \frac{1}{RT} \int_{-\infty}^{\infty} F(\omega) \int_{-\infty}^{\infty} i(t) e^{i\omega t} dt d\omega; \quad (22:11)$$

The limits on the inner integral have been set to ± 1 without significant error since the pulse is localized within the time interval T . Since $i(t)$ is real, it follows from Eq. (22.7) that

$$F(\omega) = \frac{1}{2^{1/4}} \int_{-1}^{+1} i(t) e^{i\omega t} dt = F(i\omega); \quad (22:12)$$

Substitution in Eq. (22.9) gives

$$P = \frac{2^{1/4}}{RT} \int_{-1}^{+1} |jF(\omega)|^2 d\omega; \quad (22:13)$$

which can be written as

$$P = \frac{4^{1/4}}{RT} \int_0^1 |jF(\omega)|^2 d\omega; \quad (22:14)$$

The fraction of the power spectrum between ω and $\omega + d\omega$ is $P(\omega) d\omega$, where the power spectral density function is clearly

$$P(\omega) = \frac{4^{1/4} (jF(\omega))^2}{RT} \quad (22:15)$$

For a flow of $\bar{v} < N >$ charge carriers per second, which are assumed to be uncorrelated, the total power spectral density is $\bar{v} < N > P(\omega)$. The spectral energy density supplied per second is

$$U(\omega) = \bar{v} < N > P(\omega) T; \quad (22:16)$$

where $T = 1s$, which can be written as

$$U(\omega) = 4^{1/4} \frac{\bar{v} < N > |jF(\omega)|^2}{R}; \quad (22:17)$$

This energy can be considered to result from a noise current generator whose mean squared magnitude is

$$\langle i_N^2(\omega) \rangle = 4^{1/4} \bar{v} < N > |jF(\omega)|^2 d\omega \quad (22:18)$$

For frequencies where $F(\omega)$ is ≈ 1 at

$$\langle i_N^2(\omega) \rangle = \frac{\bar{v} < N > e^2}{4} d\omega; \quad (22:19)$$

Since $\bar{v} < N > = \bar{i} = e$ and $d\omega = 2\pi df$ we can write the shot noise spectrum in terms of conventional frequency as

$$\langle i_N^2(f) \rangle = 2e\bar{i}df; \quad (22:20)$$

Even if all other sources of noise in the detector and its following electronics can be eliminated, shot noise is inescapable. It is produced by the signal. The best signal to noise (power) ratio to be expected is the shot-noise limited (SNL) value

$$\frac{P_S}{N_{SNL}} = \frac{\bar{i}^2 R}{i_N^2 R} = \frac{(e \bar{v} < I > A = h\nu)^2}{2e^2 \bar{v} < I > A = h\nu} \quad (22:21)$$

giving

$$\frac{S}{N_{\text{SNL}}} = \frac{\langle I \rangle A}{2h\nu\Delta f} = \frac{P}{2h\nu\Delta f} \quad (22:22)$$

where $P = \langle I \rangle A$ is the total power (W) falling on the detector.

This signal-to-noise ratio is frequently called the limiting video or direct detection limit.* It is the best that can be done with a detector directly receiving only the incoming light it is desired to detect. For a detector with unity quantum efficiency the minimum detectable light signal ($S=N = 1$) is

$$P_{\text{min}} = 2h\nu\Delta f; \quad (22:23)$$

The sampling time for digital data, by Nyquist's theorem, is generally set so that data is sampled at twice the maximum frequency being observed, i.e., $\Delta T = 1/(2\Delta f)$. The minimum detectable signal therefore corresponds to one detected photon per sampling interval. In practice, additional sources of noise are present in a detection system involving a detector and other electronic devices. Principal among these are 1/f noise, Johnson noise (also called thermal, Nyquist or resistance noise), and in semiconductor detectors generation-recombination (gr noise).

22.2.2 Johnson Noise

At temperatures above absolute zero the thermal energy of the charge carriers in any resistor leads to fluctuations in local charge density. These fluctuating charges cause local voltage gradients that can drive a corresponding current into the rest of the circuit. The quantitative treatment of this thermal noise can be carried out in several different ways^[22:1]. Nyquist^[22:2] proposed a theorem that stated that the noise power generated by a circuit element did not depend specifically on the nature of the circuit element, but only on the temperature of the component and the frequency band being examined. He proved this result thermodynamically and derived a quantitative expression for the noise power by considering the energy flow between two resistors of equal value connected together by a cable with characteristic impedance R . The noise power can also be calculated by considering the voltages and currents associated with a RLC tuned circuit. However, we choose here to calculate the noise power by relating the current and voltage fluctuations to the available radiation energy at temperature T described by black body radiation. We do this by considering the circuit shown in Fig. (22.2) in

* Also called the photon-noise or quantum-limited signal-to-noise ratio.

Fig. (22.2).

which an antenna is connected to a resistance R . The antenna can be modelled as a voltage source with an associated series resistance R_r | called its radiation resistance. If the antenna is a wire of length ℓ , then the electric field of the blackbody radiation will induce a voltage in the antenna. For radio and microwave frequencies, we can use the Rayleigh-Jeans approximation to Planck's radiation formula to calculate this induced voltage. The energy density of the radiation interacting with the antenna within a frequency band of width Δf at frequency f is

$$\Delta u = \frac{8\pi^5 f^2}{15c^3} kT \Delta f \quad (22:24)$$

On average, because the black body radiation is unpolarized, only one-third of this energy actually interacts with a linear antenna, oriented (say) in the x -direction. The mean-squared electric-field in the antenna direction within the frequency band Δf is *

$$\langle E_x^2 \rangle = \frac{cZ_0 \Delta u}{3} \quad (22:25)$$

The mean-square voltage induced in the antenna within the band Δf is

$$\langle V_r^2 \rangle = \ell^2 \langle E_x^2 \rangle = \frac{8\pi^5 f^2}{15c^2} kT Z_0 \Delta f \quad (22:26)$$

This voltage induces a current i_a in the antenna, which can flow into an external short-circuit.

Since an ideal antenna cannot dissipate any energy it must re-radiate the energy it receives. We can model this re-radiation process by associ-

* For a plane linearly polarized wave the intensity within Δf can be written as $\langle I \rangle = \langle \frac{E_x^2}{2Z_0} \rangle = \frac{c \Delta u}{6}$. The factor of 6 comes from the three orthogonal propagation directions in space and the 2 orthogonal polarization directions associated with each of them.

ating a radiation resistance R_r with the antenna such that the apparent dissipation in R_r balances the received power, as shown in Fig. (22.2a). For this to be so

$$\frac{c \langle V_r^2 \rangle}{R_r} = i_r^2 R_r: \quad (22:27)$$

Now the power radiated by an antenna of length λ carrying an oscillating current of magnitude i_r is^[22:3;22:4]

$$W = \frac{1/4 Z_0 i_r^2 \lambda^2 f^2}{3c^2} \quad (22:28)$$

Equating W with $\frac{1}{2} i_r^2 R_r$ gives

$$R_r = \frac{2/4 f^2 \lambda^2 Z_0}{3c^2} \quad (22:29)$$

If the antenna is connected to a resistor R , as shown in Fig. (22.2b), then the oscillating current in the antenna will dissipate a power $V_r^2 R / (R_r + R)^2$ in the resistor R . If the entire circuit shown in Fig. (22.2b) is in thermal equilibrium, then for the resistor R to heat would violate the second law of thermodynamics. Consequently, the resistor R must drive power back into the antenna to balance the power it receives. Consequently, the mean-square voltage generated by R within the band $c f$ must satisfy

$$\frac{\langle V_N^2 \rangle}{R} = \frac{c \langle V_r^2 \rangle}{R_r} = \frac{8/4 f^2 \lambda^2 Z_0}{3c^2} kT c f \times \frac{3c^2}{2/4 f^2 \lambda^2 Z_0} = 4kT c f \quad (22:30)$$

which gives

$$\langle V_N^2 \rangle = 4kTR c f: \quad (22:31)$$

Alternatively, the noisy resistor can be treated as a resistor R with a parallel current source whose mean-square value within the band $c f$ is

$$\langle i_N^2 \rangle = \frac{4kT}{R} c f: \quad (22:32)$$

Thus, this source of noise can be reduced by cooling the oscillating component to a low temperature.

If a shot-noise-limited detector drives a circuit with an input resistance R the equivalent current involves both the shot-noise current source and the Johnson noise current source, as shown in Fig. (22.3).

22.2.3 Generation-Recombination Noise and 1/f Noise

These two types of noise are important in semiconductor detectors. 1=f noise, or current noise, has a power spectrum that depends inversely on

Fig. (22.3).

the frequency. It can be described by a noise current generator

$$\langle i_N^2 \rangle = \frac{K \langle i \rangle^\alpha c f}{f^\beta} \quad (22:33)$$

where $\langle i \rangle$ is the average current through the detector K is a constant, and typically $\alpha = 2$ and $\beta = 1$. The noise that depends inversely on the frequency, the $1=f$ noise, originates from many causes, such as the diffusion of charge carriers, the presence of impurity atoms and lattice defects in the material, and interaction of charge carriers with the surface of the semiconductor. To achieve high $S=N$ ratios low frequency detection should be avoided because of $1=f$ noise which may increase in magnitude down to frequencies of 10^4 Hz. The trend towards higher and higher data rates in optical communication systems moves system operation well away from $1=f$ noise, although at high and intermediate frequencies generation-recombination (gr) noise will still be a factor in determining $S=N$ ratio. Generation-recombination noise arises from statistical fluctuations in the number of carriers in the detector. In this sense it is closely related to photon noise, the fluctuation in the number of generated carriers, but the gr noise results from the secondary carrier density fluctuations arising from random electron-hole recombinations. This recombination process has a characteristic lifetime τ_0 , which can be very short, 1ns or less, so we expect this source of noise to contribute up to frequencies $\gg 1/\tau_0$. Indeed, this is the case, the gr noise spectrum is flat up to a frequency $\gg 1/\tau_0$ and can be shown to correspond to an equivalent noise power generator

$$\langle i_N^2 \rangle = \frac{4 \langle i \rangle^2 \tau_0}{\langle N \rangle (1 + 4\tau_0^2 f^2)} \quad (22:34)$$

where $\langle N \rangle$ is the average number of charge carriers. If the time it takes a charge carrier to travel through the detector into the external

Fig. (22.4).

current is i_d then we can write

$$\frac{\langle N \rangle}{i_d} = \frac{\langle i \rangle}{e} \quad (22:35)$$

and the gr noise current generator can be written as

$$\langle i_N^2 \rangle = \frac{4e \langle i \rangle (i_0 = i_d)}{(1 + 4\frac{1}{2} f^2 i_0^2)} \quad (22:36)$$

Thus, we expect gr noise to be less in a detector with a rapid recombination time and also less in heavily doped semiconductors where the number of charge carriers will be greater at a given average current $\langle i \rangle$ than in an intrinsic material.

Thus, in a semiconductor detector there are contributions from 1=f noise, gr noise, and thermal noise. The relative contribution of these noise sources varies with frequency in the schematic way shown in Fig. (22.4).

22.3 Detector Performance Parameters

22.3.1 Noise Equivalent Power

It is important to consider under what conditions of operation the performance of a detector will be primarily limited by generation-recombination noise, since for all but low frequencies gr noise generally dominates over 1=f and Johnson noise. This can be carried out by considering the noise equivalent power (NEP) of the detector, which for a good detector is a practical measure of the magnitude of the gr noise. The NEP is the rms value of a sinusoidally modulated light signal falling on the detector that gives rise to an rms electrical signal equal to the rms noise voltage. The NEP is usually specified in terms of a blackbody source, a

reference bandwidth, usually 1 Hz, and the modulation frequency of the radiation.

For example NEP (500K, 900, 1) implies a blackbody illuminator whose temperature is 500K, a 900 Hz modulation frequency, and a reference bandwidth of 1 Hz. If the illuminating signal has intensity I (W/m^2) and falls on a detector of area A then we can write

$$NEP = \frac{IA}{\Delta f} \frac{V_N}{V_S} \tag{22:37}$$

where V_S , V_N are the signal and noise voltages, respectively, measured with a bandwidth Δf .

We can rearrange Eq. (22.37) to give the equivalent intensity needed to generate a S=N ratio of 1 in a 1 Hz bandwidth as

$$I(S=N = 1; \Delta f = 1Hz) = \frac{NEP}{A} \tag{22:38}$$

For example a detector with a NEP of $10^{-12} W Hz^{-1/2}$ needs a total power of $10^{-12} W$ of blackbody power to fall onto its sensitive surface to generate a signal equal to the detector noise. The photon noise limit for such a detector would correspond to a received intensity of

$$I \text{ (photon-noise-limited, } S=N = 1; \Delta f = 1Hz) = 2h\nu$$

So the relative magnitudes of the detector noise limited minimum detectable light signal and that set by photon noise is

$$\frac{I \text{ (detector noise limited)}}{I \text{ (photon noise limited)}} = \frac{NEP}{2Ah\nu} \tag{22:39}$$

Strict equality does not hold in Eq. (22.39) as the NEP is generally defined for blackbody radiation and the quantum $h\nu$ implies an appropriate average frequency for the incoming radiation.

22.3.2 Detectivity

Many detectors exhibit an NEP that is proportional to the square root of the detector area; so a detector-area-independent parameter, the detectivity D^* is frequently used, specified by

$$D^* = \frac{\sqrt{A}}{NEP}; \tag{22:40}$$

where A is the area of the detector. D^* is specified in the same way as NEP: for example, $D^* (500K, 900, 1)$. To specify the variation in response of a detector with wavelength, the spectral detectivity is often used. Thus, the symbol $D^*(\lambda; 900; 1)$ would specify the response of the detector to radiation of wavelength λ , modulated at 900Hz and detected with a 1Hz bandwidth. D^* is measured in units $cm Hz^{1/2} W^{-1}$.

The performance of photodiodes used in optical communication systems is frequently specified in terms of their responsivity R , usually specified in A/W , which characterizes the response of the detector to unit irradiance. At high data rates the actual S/N performance of these detectors will depend on the wide-band amplification electronics with which they are used. Specification of their performance in terms of NEP or D^* becomes less relevant.

22.3.3 Frequency Response and Time Constant

The frequency response of a detector is the variation of responsivity or radiant sensitivity as a function of the modulation frequency of the incident radiation. The frequency variation of R and the time constant of the detector are generally related according to

$$R(f) = \frac{R(0)}{(1 + 4^{1/2} f^2 \tau^2)^{1/2}} \quad (22:41)$$

For frequencies above $1=2^{1/2}\tau$ the response is falling off significantly and at high enough modulation frequencies the detector will provide a dc output proportional to the average intensity.

22.4 Practical Characteristics of Optical Detectors

The development of optical detectors has occurred, in common with various branches of electronics, by a series of advances through the use of gas-filled tubes and vacuum tubes to various semiconductor devices. However, whereas in general electronics the vacuum tube is now reserved for specialized applications, vacuum-tube optical detectors such as the photomultiplier are still in widespread use.

22.4.1 Photoemissive Detectors

Photoemissive detectors include gas-filled and vacuum photodiodes, photomultiplier tubes, and photo-channeltrons. These are all photon detectors that utilize the photoelectric effect. When radiation of frequency ν falls upon a metal surface, electrons are emitted, provided the photon energy $h\nu$ is greater than a minimum critical value \bar{A} , called the work function, which is characteristic of the material being irradiated. A simplified energy-level diagram illustrating this effect for a metal-vacuum interface is shown in Fig. (22.5) (a). For most metals, \bar{A} is in a range from 4-5 eV ($1 \text{ eV} \cdot \lambda = 1.24 \mu\text{m}$), although for the alkali metals it is lower,

Fig. (22.5).

for example, 2.4 eV for sodium and 1.8 eV for cesium. Pure metals or alloys, particularly beryllium-copper, are used as photoemissive surfaces in ultraviolet and vacuum-ultraviolet detectors.

Lower work functions, and consequently sensitivities that can be extended into the infrared, can be obtained with special semiconductor materials. Fig. (22.5)(b) shows a schematic energy-level diagram of a semiconductor-vacuum interface. In this case the work function is defined as $\hat{A} = E_{\text{vac}} - E_F$, where E_F is the energy of the Fermi level. In a pure semiconductor, the Fermi level is in the middle of the band gap, as illustrated in Fig. (22.5)(b). In a p-type doped semiconductor, E_F moves down toward the top of the valence band, while in n-type material it moves up toward the bottom of the conduction band. Consequently, \hat{A} is not as useful a measure of the minimum photon energy for photoemission as it is for a metal. The electron affinity \hat{A} is a more useful measure of this minimum energy for a semiconductor. Except at absolute zero, photons with energy $h\nu > \hat{A}$ cause photoemission. Photons with energy $h\nu > E_g$ lead to the production of carriers in the conduction band; this leads to intrinsic photoconductivity, which is the operative detection mechanism in various infrared detectors, such as InSb.

(a) Vacuum Photodiodes. Once electrons are liberated from a photoemissive surface (a photocathode), they can be accelerated to an electrode positively charged with respect to the cathode | the anode | and generate a signal current. If the acceleration of photoelectrons is directly from cathode to anode through a vacuum, the device is a vacuum photodiode. Because the electrons in such a device take a very direct path from anode to cathode and can be accelerated by high voltages | up to several kilovolts in a small device | vacuum photodiodes have the fastest response of all photoemissive detectors. Risetimes of 100 ps or

less can be achieved. External connections and electronics are generally the limiting factors in obtaining short risetimes from such devices. However, vacuum photodiodes are not very sensitive, since at most one electron can be obtained for each photon absorbed at the photocathode. In principle, of course, the limiting sensitivity of the device is set by its quantum efficiency. Practical quantum efficiencies for photoemissive materials range up to about 0.4.

If the space between photocathode and anode is filled with a noble gas, photoelectrons will collide with gas atoms and ionize them, yielding secondary electrons. Thus, an electron multiplication effect occurs. However, because the mobility of the electrons moving from cathode to anode through the gas is slow, these devices have a long response time, typically about 1 ms. Gas-filled photocells are no longer competitive with solid-state detectors in practical applications.

(b) Photomultipliers. If photoelectrons are accelerated in vacuum from the photocathode and allowed to strike a series of secondary electron emitting surfaces, called dynodes, held at progressively more positive voltages, a considerable electron multiplication can be achieved and a substantial current can be collected at the anode. Such devices are called photomultipliers. Practical gains of 10^9 (anode electrons per photoelectron) can be achieved from these devices for short light pulses. Because of their very high gain, photomultipliers can generate substantial signals when only a single photon is detected: for example, an anode pulse of 2-ns duration containing 10^9 secondary photoelectrons produced from a single photoelectron will generate a voltage pulse of 4 V across 50 ohms. This, coupled with their low noise, makes photomultipliers very effective single-photon detectors. Photomultiplier D^* -values can range up to 10^{16} $\text{cm}^2\text{Hz}^{1/2}\text{W}^{-1}$. Surprisingly, the dark-adapted human eye, which can detect bursts of about 10 photons in the blue, comes close to this sensitivity.

The time response characteristics of photomultiplier tubes depend to a considerable degree on their internal dynode arrangement. The response of a given device can be specified in terms of the output signal at the anode that results from a single photoelectron emission at the photocathode. This is illustrated in Fig. (22.6). Because electrons passing through the dynode structure can generally take slightly different paths, secondary electrons arrive at the anode at different times. The anode pulse has a characteristic width called the transit-time spread, which usually ranges from 0.1 to 20 ns. The time interval between photoemission at the cathode and the appearance of an anode pulse is called the transit

Fig. (22.6).

Fig. (22.7).

time, and is usually a few tens of nanoseconds. The transit time and transit-time spread also fluctuate slightly from one single-photoelectron-produced anode pulse to the next.

There are four main types of dynode structure in common usage in photomultiplier tubes; these are illustrated in Fig. (22.7). The circular cage structure is compact and can be designed for good electron collection efficiency and small transit-time spread. This dynode structure works well with opaque photocathodes, but is not very suitable for high-amplification requirements where a larger number of dynodes is required. The box-and-grid and Venetian-blind structures offer very good electron-collection efficiency. Because they collect multiplied electrons independently of their path through the dynode structure, a wide range of secondary-electron trajectories is possible, leading to a large transit-time spread and slow response. Typical response times of these types of tube are 10-20 ns.

The focused dynode structure is designed so that electrons follow

Fig. (22.8).

paths of similar length through the dynode structure. To accomplish this, electrons that deviate too much from a specified range of trajectories are not collected at the next dynode. These types of tube offer short response times, typically 1-2 ns.

Venetian-blind tubes can easily be extended to many dynode stages and have a very stable gain in the presence of small power supply fluctuations. These tubes also have an optically opaque dynode structure, which contributes to their exhibiting very low dark-current noise when operated under appropriate conditions.

(c) Photocathode and Dynode Materials. The performance of a photomultiplier depends not only on its internal structure, but also on the photoemissive material of its photocathode and the secondary-electron-emitting material of its dynodes. The wavelength dependence of various commercially available photocathode materials is shown in Fig. (22.8); some radiant sensitivities and quantum efficiencies are given in Table 22.1. The short-wavelength cutoff of a given material depends on its work function. This cutoff is not sharp because, except at absolute zero, there are always a few electrons high up in the conduction band available for photoexcitation by low-energy photons. Some few of these electrons, because of their thermal excitation, will be emitted without any photo-stimulation. This contributes the major part of the dark current observed from the photocathode. Materials that have low work functions, and are consequently more red- and infrared-sensitive, have higher (often much higher) dark currents than materials that are optimized for visible and ultraviolet sensitivity.

Photocathodes are available in both opaque and semitransparent forms, depending on the model of phototube. In the semitransparent form the photoelectrons are ejected from the thin photoemissive layer on the op-

Table 22.1 Characteristics of Photoemissive Surfaces.

Cathode	Radiant Sensitivity (mA/W)			Peak Quantum Efficiency (%)	Peak Wavelength (nm)	Dark Current ^a (A/cm ²)
	515 nm	694 nm	1.06 ¹ μm			
S-1 Cs-O-Ag	0.6	2	0.4	0.08	800	9 × 10 ⁻¹³
S-10 Cs-O-Ag-Bi ^b	20	2.7	0	5	470	9 × 10 ⁻¹⁶
S-11 Cs ₃ Sb on MnO ^c	39	0.2	0	13	440	10 ⁻¹⁶
S-20 (Cs)Na ₂ K ₂ Sb (tri-alkali)	53	20	0	18	470	3 × 10 ⁻¹⁶
S-22 (bi-alkali)	42	0	0	26	390	1.6 × 10 ⁻¹⁸
GaAs ^b	48	28	0	14	560 ^e	10 ⁻¹⁶
GaAsP ^d	60	30	0	19	400 ^e	3 × 10 ⁻¹⁵
InGaAs ^d	{	{	4.3	{	{ ^e	3 × 10 ⁻¹⁴
InGaAsP ^d	{	{	{	47	300 ^e	2.5 × 10 ⁻¹³
S-25 (ERMA) ^f	53	26	0	25	430	1 × 10 ⁻¹⁵
Cs-Te (solar blind)	{	{	{	15	254	2.5 × 10 ⁻¹⁷

Note: Table shows typical values, but these can vary greatly from one manufacturer to another.

^a At room temperature.

^b Cathode designated S-3 is similar.

^c Several types of CsSb photocathodes exist where the CsSb is deposited on different opaque and semitransparent substrates and various window materials are used. These photocathodes have the designations S-4, 5, 13, 17, and 19 as well as S-11.

^d Negative-electron-affinity (NEA) photoemitters.

^e May show no wavelength of maximum quantum efficiency; quantum efficiency falls with increasing wavelength. However, exact spectral characteristics will depend on thickness of photoemitter and whether it is used in transmission or not.

^f Extended-red S-20.

posite side from the incident light. In both types the photocathode has to perform two important functions: it must absorb incident photons and allow the emitted photoelectron to escape. The latter event is inhibited if photons are absorbed too deep in a thick photoemissive layer, or if the photoelectron suffers energy loss from scattering in the layer.

If the emitted photoelectron has too much energy, it can excite a further electron across the band gap. This pair-production phenomenon inhibits the release of photoelectrons from the photoemissive layer, and accounts for the ultraviolet cutoff characteristics of the different materials shown in Fig. (22.8). With reference to Fig. (22.5) it can be shown that for pair production to occur the incident photon energy must be greater than $2E_g$. Photoelectrons have the best chance of escaping, and the photocathode its highest quantum efficiency, for materials where $\hbar\nu < E_g$.

Practical photoemissive materials fall into two main categories: classical photoemitters and negative-electron-affinity (NEA) materials. Classical photoemitters generally involve an alkali metal or metals, a group-V element such as phosphorus, arsenic, antimony, or bismuth, and sometimes silver and/or oxygen. Examples are the Ag-O-Cs(S1) photoemitter, which has the highest quantum efficiency beyond about 800 nm of any classical photoemitter, and Na_2KSbCs , the so-called tri-alkali (S-20) cathode.

NEA photoemitters utilize a photoconductive single-crystal semiconductor substrate with a very thin surface coating of cesium and usually a small amount of oxygen. The cesium (oxide) layer lowers the electron affinity below the value it would have in the pure semiconductor, achieving an effectively negative value. Examples of such NEA photoemitters are GaAs (CsO) and InP (CsO). NEA emitters can offer very high quantum efficiency and extended infrared response. GaAs (CsO), for example, has higher quantum efficiency in the near infrared than an S-1 photocathode.

The performance of the dynode material in photomultiplier tubes is specified in terms of the secondary-emission ratio δ as a function of energy. For a phototube with n dynodes the gain is δ^n . In the past, the commonest dynode materials were CsSb, AgMgO, and BeCuO. The last is also used as the primary photoemitter in windowless photomultipliers that are operated in vacuo for the detection of vacuum-ultraviolet radiation. BeCuO has the desirable property that it can be reactivated after exposure to air. The above materials have δ -values of 3-4. Newer NEA dynode materials have much higher δ -values: in particular, that of GaP can range up to 40 for an incident-electron input energy of 800 eV. With such high δ -values a photomultiplier tube needs fewer dynodes for a given gain, which means a more compact and faster-response tube can be built. In very many commercial photomultipliers, the first dynode at least is now frequently made of GaP. This offers improved characteriza-

Fig. (22.9).

tion of the single-photoelectron response of the tube, which is important in designing a system for optimum signal-to-noise ratio.

The accelerating voltages are supplied to the dynodes of a photomultiplier by a resistive voltage divider called a dynode chain. The relative resistance values in the chain determine the distribution of voltages applied to the dynodes. The total chain resistance R determines the chain current at a given total photocathode-anode applied voltage. Many photomultiplier tubes have one or more focusing electrodes between the photocathode and the first dynode. The voltage on these electrodes can be adjusted to optimize the collection of photoelectrons from the photocathode.

The actual response-time behavior of the photomultiplier can be determined by observing its single-photoelectron response. This is done by looking at the anode pulses with a fast oscilloscope. The photocathode does not need to be illuminated for this to be done; sufficient noise pulses will usually be observed. The pulses should appear as in Fig. (22.6). These anode pulses reflect the time distribution and number of secondary electrons reaching the anode following single (or multiple) photoelectron emissions from the photocathode. If the height distribution of anode pulses is measured, a distribution such as is shown in Fig. (22.9) (a) will probably be seen. The idealized distribution shown in Fig. (22.9) (b) may be seen from newer tubes with GaP dynodes, which have a very high \pm -value.

22.4.2 Photoconductive Detectors

Photoconductive detectors can operate through either intrinsic or extrinsic photoconductivity. The physics of intrinsic photoconductivity is

Fig. (22.10).

illustrated in Fig. (22.10)(a). Photons with energy $h\nu > E_g$ excite electrons across the band gap. The electron-hole pair that is thereby created for each photon absorbed leads to an increase in conductivity which comes mostly from the electrons. Semiconductors with small band gaps respond to long-wavelength radiation but must be cooled accordingly; otherwise thermally excited electrons swamp any small photoconductivity effects. Table 22.2 lists commonly available intrinsic photoconductive detectors together with their usual operating temperature and the limit of their long-wavelength response, λ_0 , together with some representative figures for detectivities and time constants. Note that silicon and germanium are also operated in both photovoltaic and avalanche modes.

If a semiconductor is doped with an appropriate material, impurity levels are produced between the valence and conduction bands as shown in Fig. (22.9)(b). Impurity levels that are able to accept an electron excited from the conduction band are called acceptor levels, whereas impurity levels that can have an electron excited from them into the conduction band are called donor levels. Thus, in Fig. (22.10)(b) photons with energy $h\nu > E_A$ excite an electron to the impurity level, leaving a hole in the valence band and thereby giving rise to p-type extrinsic photoconductivity. Photons with energy $h\nu > E_D$ will excite an electron into the conduction band, giving n-type extrinsic photoconductivity. For example, gold-doped germanium has an acceptor level 0.15 eV above the valence band and is an extrinsic p-type photoconductor, as is copper-doped germanium, which has an acceptor level 0.041 eV above the valence band. These are the two most commonly used extrinsic photoconductive detectors, responding out to about $9 \mu\text{m}$ and $30 \mu\text{m}$ respectively. Curves showing the variation of their D^{eff} with wavelength are given in Fig. (22.11).

Fig. (22.11).

Table 22.2 Intrinsic Photoconductive Detectors

Semiconductor	T (K)	E _g (eV)	λ ₀ (μm)	D ² (max) (cmHz ^{1/2} W ⁻¹)	τ
CdS	295	2.4	0.52	3.5 × 10 ¹⁴	50 ms
CdSe	295	1.8	0.69	2.1 × 10 ¹¹	10 ms
Si	295 ^a	1.12	1.1	2 × 10 ¹²	50 ps
Ge	295 ^a	0.67	1.8	10 ¹¹	10 ns
PbS	295	0.42	2.5	2 × 10 ¹¹	0.1-10 ms
	195	0.35	3.0	4 × 10 ¹¹	0.1-10 ms
	77	0.32	3.3	8 × 10 ¹¹	0.1-10 ms
PbSe	295	0.25	4.2	1 × 10 ⁹ - 5 × 10 ⁹	1 ¹ s
	195	0.22	5.4	1.5-4 × 10 ¹⁰	30-50 ¹ s
	77	0.21	5.8	3 × 10 ¹⁰	50 ¹ s
InSb ^b	77	0.22	5.5-7.0	3 × 10 ¹⁰	0.1-1 ¹ s
Hg _{0.8} Cd _{0.2} Te	77	0.1	12-25	10 ⁹ -10 ¹¹	> 1 ns

^a Increased sensitivity can be obtained by cooling.

^b More commonly operated in a photovoltaic mode.

All photoconductive detectors, whether intrinsic or extrinsic, are operated in essentially the same way, although there are wide differences in packaging geometry. These differences arise from differing operating temperatures and speed-of-response considerations. A schematic diagram which shows the main construction features of a liquid-nitrogen-cooled photoconductive or photovoltaic detector is given in Fig. (22.12). Uncooled detectors can be of much simpler construction | for example, in a transistor or °at solar cell package.

One feature of the cooled detector design shown in Fig. (22.12) is wor-

Fig. (22.12).

Fig. (22.13).

thy of note. The field of view of the detector is generally restricted by an aperture, which is kept at the temperature of the detection element. This shields the detector from ambient infrared radiation, which peaks at 9.6 μ m. For detection of low-level narrow-band infrared radiation the influence of background radiation can be further reduced by incorporating a cooled narrow-band filter in front of the detector element. The filter will only radiate beyond the cutoff wavelength of the detector, and it restricts transmitted ambient radiation to a narrow band.

Fig. (22.13) shows a basic biasing circuit commonly used for operating photoconductive detectors. R_d is the detector dark resistance. It is easy to see that the change in voltage, ΔV , that appears across the load resistor R_L , for a small change ΔR in the resistance of the detector is

$$\Delta V = \frac{i V_0 R_L \Delta R}{(R_d + R_L)^2} \quad (22:42)$$

This is at a maximum when $R_L = R_d$. Thus it is common practice to bias the detector with a load resistance equal to the detector's dark

resistance. The bias voltage is selected to give a bias current through the detector that gives optimum detectivity.

A few photoconductive detectors are worthy of brief extra comment. Silicon and germanium are much more commonly used for photodiodes, frequently in an avalanche mode. These devices are discussed in the next section. Lead sulfide detectors have high impedance, 0.5 to 100 M Ω , and slow response, but are sensitive detectors in the spectral region between 1.2, and 3 μm and can be used uncooled. $D^*(\lambda)$ curves for these detectors and lead selenide are shown in Fig. (22.13).

Table 22.3 lists the characteristics of some extrinsic photoconductive detectors. Because of the small energy gaps involved in these detectors, they all operate at cryogenic temperatures. The use of extrinsic photoconductivity for the detection of far-infrared radiation requires the introduction of appropriate doping material into a semiconductor in order to generate an acceptor or donor impurity level extremely close to the valence or conduction bands, respectively. Well-characterized impurity levels can be generated in germanium in this way using gallium, indium, boron, or beryllium doping, but the long-wavelength sensitivity limit is restricted to about 124 μm . Longer-wavelength-sensitive, extrinsic photoconductivity can be observed in appropriately doped InSb in a magnetic field. However, bulk InSb can be used more efficiently for infrared detection in a different photoconductive mode entirely^{[22:5];[22:6];[22:7]}. Even at the low temperature at which far-infrared photoconductive detectors operate ($\sim 4\text{K}$), there are carriers in the conduction band. These free electrons can absorb far-infrared radiation efficiently and move into higher-energy states within the conduction band. This change in energy results in a change of mobility of these free electrons, which can be detected as a change in conductivity. These hot-carrier-effect photoconductive detectors can be used successfully over a wavelength range extending from 50 to 10,000 μm ; they have detectivities up to $2 \times 10^{12} \text{cmHz}^{1/2} \text{W}^{-1}$ and response times down to 10 ns or less. They are frequently operated in a large magnetic field (several hundred kA/m or more).

22.4.3 Photovoltaic Detectors (Photodiodes)

In a photovoltaic detector photoexcitation of electron-hole pairs occurs near a junction when radiation of energy greater than the band gap is incident on the junction region. Extrinsic photoexcitation is rarely used in photovoltaic photodetectors. The internal energy barrier of the junction

Table 22.3 Extrinsic Photoconductive Detectors

Semiconductor	Impurity	T(K)	λ_0 (μm)	D^{a}	τ
Ge: Au ^a	p-type	77	8.3	$3 \times 10^9 - 10^{10}$	30 ns
Ge: Hg ^b	p-type	<28	14	$1 - 2 \times 10^{10}$	>0.3 ns
Ge: Cd	p-type	<21	21	$2 - 3 \times 10^{10}$	10 ns
Ge: Cu ^b	p-type	<15	30	$1 - 3 \times 10^{10}$	>0.4 ns
Ge: Zn ^{a,b}	p-type	<12	38	$1 - 2 \times 10^{10}$	10 ns
Ge: Ga	p-type	<3	115	2×10^{10}	>1 ¹ s
Ge: In	p-type	4	111		<1 ¹ s
Si: Ga	p-type	4	17	$10^9 - 10^{10}$	>1 ¹ s
Si: As	n-type	<20	22	$1 - 3.5 \times 10^{10}$	0.1 ¹ s

^a Sometimes also contain silicon.

^b Sometimes also contain antimony.

causes the electron and hole to separate, creating a potential difference across the junction. This effect is illustrated for a p-i-n junction in Fig. (22.1). Other types of structure are also used, such as p-i-i-n, Schottky-barrier (a metal deposited onto a semiconductor surface) and heterojunctions. The p-i-n and p-i-i-n structures are the most commonly used. All these devices are commonly called photodiodes. The characteristics of some important photodiodes are listed in Table 22.4. Important photodiodes include silicon for detection of radiation between 0.1 and 1.1 μm and detectors based on the InGaAs(P) system for the region between 0.9 and 1.7 μm , which encompasses the important fiber optical communication wavelengths of 1.3 μm and 1.55 μm (see Chapter 25). Typical spectral responsivities for some of the materials are shown in Fig. (22.15). Other photodiodes with more specialized applications include germanium between 0.4 and 1.8 μm , indium arsenide between 1 and 3.8 μm , indium antimonide between 1 and 7 μm , lead-tin telluride between 2 and 18 μm , and mercury-cadmium telluride between 1 and 12 μm . Some typical curves of $D^{\text{a}}(\lambda)$ for these photodiodes are shown in Fig. (22.16). These spectral response regions are not all necessarily covered by a detector operating at the same temperature; for example, InSb responds to 7 μm at 300 K but to wavelengths no longer than 5.6 μm at 77 K. The wavelength response of PbSnTe and HgCdTe depends also on the stoichiometric composition of the crystal. All these photodiodes have very high quantum efficiency, defined in this case as the ratio of photons absorbed to mobile electron-hole pairs produced in the

Fig. (22.14).

Fig. (22.15).

Fig. (22.16).

junction region. Values in excess of 90% have been observed in the case of silicon.

When a photodiode detector is illuminated with radiation of energy greater than the band gap, it will generate a voltage and can be oper-

Table 22.4. Photovoltaic Detectors (Photodiodes)

Semiconductor	T (K)	Wavelength		D^{a} (max) (or NEP) ^b	τ
		Range	(μm)		
Si	300	0.2-1.1		$\cdot 2 \times 10^{13}$	>6ps
InGaAs ^a	300	0.9-1.7		6×10^{14} (NEP, $\text{W}/\text{Hz}^{1/2}$)	20ps
GaAsP ^a	300	0.3-0.76		2×10^{15} (NEP, $\text{W}/\text{Hz}^{1/2}$)	
Ge	300	0.4-1.8		10^{11}	0.3 ns
InAs	300	1-3.8		$\cdot 4 \times 10^9$	5 ns-1 μs
InAs	77	1-3.2		4×10^{11}	0.7 μs
InSb	300	1-7		1.5×10^8	0.1 μs
InSb	77	1-5.6		$\cdot 2 \times 10^{11}$	>25 ns
PbSnTe	77	2-18		$\cdot 10^{11}$	20 ns-1 μs
HgCdTe	77	1-25		10^9 - 10^{11}	>1.6 ns

^a Precise performance depends on stoichiometry. Quaternary versions of these detectors based on InGaAsP are also used.

^b The D^{a} or NEP of these detectors are typical values. The actual S/N performance of these detectors will in practical applications depend on the inevitable additional noise added by amplifier electronics^[22:8].

ated in the very simple circuit illustrated in Fig. (22.17) (a). However, it is much better to operate a photodiode detector in a reverse-biased mode, as shown in Fig. (22.17) (b), where positive voltage is applied to the n-type side of the junction and negative to the p-type. In this case, the observed photosignal is seen as a change in current through the load resistor. The difference between the two modes of operation can be easily seen from Fig. (22.18) (a), which shows the $I_{\text{p}} - V$ characteristic of a photodiode in the dark and in the presence of illumination. A photodiode responds much more linearly to changes in light intensity and has greater detectivity when operated in the reverse-biased mode. Ideal operation is obtained when the diode is operated in the current mode with an operational amplifier that effectively holds the photodiode voltage at zero | its optimum bias point. A simple practical circuit which can be used to operate a photodiode in this way is shown in Fig. (22.18). In this circuit, the bias voltage V_{B} is not necessary, but for many photodiodes will improve the speed of response, albeit at the expense of an increase in noise. The p-i-n structure is most commonly used in these devices because its performance, in terms of quantum efficiency (number of useful

Fig. (22.17).

Fig. (22.18).

carriers generated per photon absorbed) and frequency response, can be readily optimized. These devices have very low noise and fast response. In practice, the limiting sensitivity that can be obtained with them will be determined by the noise of the associated amplifier circuitry.

If the reverse bias voltage on a photodiode is increased, photoinduced charge carriers can acquire sufficient energy traversing the junction region to produce additional electron-hole pairs. Such a photodiode exhibits current gain and is called an avalanche photodiode (APD). It is in some respects the solid-state analog of the photomultiplier. Avalanche photodiodes are noisier than $p\text{-}i\text{-}n$ photodiodes, but because they have internal gain, the practical sensitivity that can be achieved with them is greater. Because of their importance $p\text{-}i\text{-}n$ photodiodes and APDs are worthy of more detailed discussion.

22.4.4 $p\text{-}i\text{-}n$ Photodiodes

In a simple $p\text{-}i\text{-}n$ junction photodiode using the structure shown in

Fig. (22.19).

Fig. (22.1) reverse bias leads to a current that increases linearly with incident optical power over up to 9 orders of magnitude, say from 1 pW to 1 mW. In order for electron-hole pairs created by photon absorption to appear as useful current in the external current they must be swept out of the junction region and collected at the electrodes before they have a chance to recombine. This is best accomplished if as much of the photon absorption as possible occurs in a thick depletion layer that is close to the electrodes, as shown in Fig. (22.19). Under reverse bias there are very few mobile charge carriers in the depleted i-layer. There is a build up of electrons on the heavily doped p⁺ side of the device and of holes on the heavily doped n⁺ side. Thus, the static electrical state is that of a parallel plate capacitor of capacitance

$$C = \frac{\epsilon_0 \epsilon_r A}{d} \quad (22:43)$$

For a typical device with $d = 20^{-4} \text{ m}$; $A = 10^{-8} \text{ m}^2$; $\epsilon_r = 12$; $C = 0.05 \text{ pF}$. The intrinsic time constant of the diode driving a 50 ohm load is 2.7ps. These are high speed devices. There are very many variations in the detailed construction of p-i-n photodiodes^{[22:8];[22:9];[22:10]}. It is possible to use heterostructures in these devices, for example the p⁺-i-n⁺ layers can be GaAlAs/GaAs/GaAs or (InGaAsP)₁/(InGaAsP)₂/InP*. If the layer through which incident radiation enters has a larger bandgap than the absorbing (intrinsic) layer, then long wavelength photons will not be absorbed in the surface layer. As is common in these layered semiconductor structures (see also Chapter 13), additional highly doped layers may be included adjacent to contact (metal) electrodes. If a metal

* The subscripts indicate different stoichiometries. For example (InGaAsP)₁ could be InP or InGaAs.

Fig. (22.20).

contact layer is placed on a layer that is not sufficiently heavily doped a rectifying Schottky diode results. Such structures can actually be used as semiconductor detectors themselves, particularly for ultraviolet detection. A thin, transparent gold layer is placed on a substrate of GaAsP or GaP, as shown in Fig. (22.20).

22.4.5 Avalanche Photodiodes

When the reverse bias of a photodiode is increased sufficiently the internal electric field can accelerate photo-generated charge carriers to sufficient energy that they can excite additional electron-hole pairs. Both electrons and holes can contribute to this process, which is shown schematically for an electron in Fig. (22.21). A schematic layout of a typical avalanche photodiode (APD) is shown in Fig. (22.21). In this structure electron-hole pairs are created initially mostly in the lightly p-doped intrinsic layer (called a μ layer), because the $n^+ ; p$ junction region is very thin. There is sufficient voltage across the μ layer for photo-generated electrons and holes to drift rapidly across it. At the $n^+ ; p$ junction near the positive electrode there is a large field gradient and efficient avalanche multiplication occurs. The n-doped region at the edge of the p-doped region is called a guard ring. It prevents edge effects at the boundary between the $n^+ ; p$ and μ layers. This keeps the avalanche region electric field uniform: there are no high fields at the edge where an avalanche could become destructive breakdown. The performance of an APD is characterized by its multiplication factor M , which is the number of electron-hole pairs generated by the dominant carrier in the detector | electrons in silicon, holes in InGaAs/InP devices. In the latter we see once again the use of a heterostructure, which

Fig. (22.21).

allows control of where incident radiation is absorbed. Incident radiation can pass through the larger bandgap InP and then be absorbed in the InGaAs. It is important in both APDs, and p-i-n photodiodes for the highest response speed that electron-hole pairs be produced in a fully depleted region of the device. If the charge carriers are generated in a region where electron or holes are present in significant concentration, the motion of the photogenerated carriers is slowed by diffusion through these undepleted regions. The fastest response is obtained if the photogenerated carriers are pulled out of the depleted region by the bias field at the maximum possible speed. This occurs at the so-called saturation velocity, v_s , which is typically on the order of 10^5 m/s.

To achieve saturation velocity requires an applied field in the depleted region on the order of 1MV/m. In a $50 \mu\text{m}$ thick depleted $\frac{1}{4}$ layer, as shown in Fig. (22.22), this requires a reverse bias of 50V. A rough estimate for the speed of response will then be 0.5 ns. Of course, to determine the precise response of the detector it would be necessary to include the actual spatial distribution of photo-generated carriers in the device^[22:8]. The lateral dimensions of the device must also be kept small so as to minimize capacitance effects on the speed of response.

22.5 Thermal Detectors

Thermal detectors, in principle, have a detectivity that is independent of wavelength from the vacuum ultraviolet upward. However, the absorbing properties of the "black" surface of the detector will, in general, show some wavelength dependence, and the necessity for a protective window on some detector elements may limit the useful spectral bandwidth of the device. Most commonly available thermal detectors although by no

Fig. (22.22).

means as sensitive as various types of photodetectors, achieve spectral response very far into the infrared | to the microwave region in fact | while conveniently operating at room temperature. Each of these detectors is discussed briefly below. Putley^[22:6] gives a more detailed discussion.

(a) Thermopile. Thermopiles, although they are one of the earliest forms of infrared detector, are still widely used. Their operation is based on the Seebeck effect, where heating the junction between two dissimilar conductors generates a potential difference across the junction. An ideal device should have a large Seebeck coefficient, low resistance (to minimize ohmic heating), and a low thermal conductivity (to minimize heat loss between the hot and cold junctions of the thermopile). These devices are usually operated with an equal number of hot (irradiated) and cold (dark) junctions, the latter serving as a reference to compensate for drifts in ambient temperature. Both metal (copper-constantan, bismuth-silver, antimony-bismuth) and semiconductor junctions are used as the active elements. The junctions can take the form of evaporated films, which improves the robustness of the devices and reduces their time constant, although this is still slow (0.1 ms at best). Because a thermopile has very low output impedance, it must be used with a specially designed low-noise amplifier, or with a step-up transformer^[22:11].

(b) Pyroelectric Detectors. These are detectors that utilize the change in surface charge that results when certain asymmetric crystals (ones that can possess an internal electric dipole moment) are heated. The crystalline material is fabricated as the dielectric in a small capacitor, and the change in charge is measured when the element is irradiated. Thus, these devices are inherently a.c. detectors. If the chopping frequency of the input radiation is slow compared to the thermal relaxation time of the crystal, the crystal remains close to thermal equilibrium

and the current response is small. When the chopping period becomes shorter than the thermal relaxation time, much greater heating and current response results. The responsivity of the detector in this case can be written as

$$R = \frac{p(T)}{\frac{1}{2}C_p d} \quad (AW^{-1}); \quad (22:44)$$

where $p(T)$ is the pyroelectric coefficient at temperature T ; d is the spacing of the capacitor electrodes, and $\frac{1}{2}$ and C_p are the density and specific heat of the crystal, respectively.

The equivalent circuit of a pyroelectric detector is a current source in parallel with a capacitance, which can range from a few to several hundred picofarads. For optimum performance, the resultant high impedance must be matched to a high-input-impedance, low-output-impedance amplifier. These detectors can have response times as short as 2 ps. Their detectivities are comparable with those of thermopiles, and they also have °at spectral response. Consequently, they can replace the thermopile in many applications as a convenient, room temperature, wide-spectral-sensitivity detector of infrared and visible light.

(c) Bolometer. The resistance of a solid changes with temperature according to a relation of the form

$$R(T) = R_0[1 + \alpha(T - T_0)]; \quad (22:45)$$

where α is the temperature coefficient of resistance, typically about 0.05 K^{-1} for a metal, and R_0 is the resistance at temperature T_0 .

A bolometer is constructed from a material with a large temperature coefficient of resistance. Absorbed radiation heats the bolometer element and changes its resistance. Bolometers utilize metal, semiconductor, or almost superconducting elements. Metal bolometers utilize fine wires (platinum or nickel) or metal films. The mass of the element must be kept small in order to maximize its temperature rise. Even so, the response time is fairly long ($\sim 1 \text{ ms}$). Semiconductor bolometer elements (thermistors) have larger absolute values of α and have largely replaced metals except where very long-term stability is required.

It is usual to operate bolometer elements in pairs in a bridge circuit, as shown in Fig. (22.22). One element is irradiated, while the second serves as a reference and compensates for changes in ambient temperature. Thermistors have a negative $I - V$ characteristic above a certain current and will exhibit destructive thermal runaway unless operated with a bias resistor. It is therefore usually best to operate the thermistor at currents below the negative-resistance part of its $I - V$ characteristic.

Fig. (22.22).

Fig. (22.24).

(d) The Golay Cell. In a Golay cell (named for its inventor, M.J.E. Golay)^[22:12], radiation is absorbed by a metal film that forms one side of a small sealed chamber containing xenon (used because of its low thermal conductivity). Another wall of the chamber is a flexible membrane, which moves as the xenon is heated. The motion of the membrane is used to change the amount of light reflected to a photodetector. The operating principle and essential design features of a modern Golay cell are shown in Fig. (22.24). Although these detectors are fragile, they are very sensitive and are still widely used for far-infrared spectroscopy.

22.6 Detection Limits for Optical Detector Systems

In a practical detector system the detector itself is coupled to various electronic devices such as amplifiers, filters, pulse counters, limiters, discriminators, phase-locked-loops, etc. It is beyond our scope here to deal with all the consequent realities of how the optical detection limit is in-

°uenced by the additional noise contributions of these devices. Nor can we deal in detail with how various ways of encoding information onto a light signal can be used to enhance the signal-to-noise ratio of the overall detection system. For additional details the reader should consult the more specialized literature^{[22:8];[22:13]; [22:17]}. However, in a well-designed optical detection system the performance should be primarily limited only by the characteristics of the detector itself. Therefore, in our discussion of fundamental detector limits we will deal with only the unavoidable noise sources of the detector and its associated resistors, as shown schematically in Fig. (22.3).

22.6.1 Noise in Photomultipliers

Noise in photomultipliers comes from several sources:

- (a) Thermionic emission from the photocathode.
- (b) Thermionic emission from the dynodes.
- (c) Field emission from dynodes (and photocathode) at high interdynode voltages. In this phenomenon the potential gradient near the emitting surface is sufficiently great to liberate electrons from the material.
- (d) Radioactive materials in the tube envelope, for example ⁴⁰K in glass.
- (e) Electrons striking the tube envelope and causing °uorescence
- (f) Electrons striking the dynodes and causing °uorescence
- (g) Electrons colliding with residual atoms of vapor in the tube, cesium for example, and causing °uorescence
- (h) Cosmic rays

Of these sources thermionic emission from the photocathode is generally the most important. It increases with the area of the photocathode. Thermionic emission from the dynodes leads to smaller anode pulses than for electrons originating at the photocathode. Therefore, when a photomultiplier is used in photon counting, where individual anode pulses are counted, discrimination against some noise can be effected by counting only within a specific range of anode pulse lengths, as shown in Fig. (22.9). Small anode pulses are likely to originate with noise source (b). Large anode pulses are likely to originate with noise sources (c) - (h).

22.6.2 Photon Counting

Photomultipliers are extremely sensitive detectors of ultraviolet, visible and near-infrared radiation. They have the unique ability to provide a macroscopic signal output from a single photoelectron liberated at their photocathode. In photon counting the anode pulses are counted individually and the pulse count rate in the case of illumination is compared with the dark count rate. The signal-to-noise ratio of this process can be further enhanced if the arrival time of signal photons can be isolated to a time window of width ζ , for example by synchronizing the pulse counting electronics to the excitation of the phenomenon leading to the light.

Let us suppose that the (weak) light source to be detected emits \bar{N}_1 photons per second that are absorbed in the photocathode. Consequently, the average number of signal anode photoelectron counts per second is \bar{N}_1 . The number of signal counts in a time interval, which can be assumed to occur randomly within this time period, is $\bar{N}_1\zeta$, with a variance

$$\overline{(N_1 - \bar{N}_1)^2} \zeta = \bar{N}_1 \zeta \tag{22:46}$$

If there are zero dark counts the relative accuracy with which \bar{N}_1 can be determined is

$$\frac{\overline{(N_1 - \bar{N}_1)^2} \zeta}{\bar{N}_1 \zeta} = \frac{1}{\bar{N}_1 \zeta} \tag{22:47}$$

Therefore, by counting for a long time good accuracy can be obtained: for $\bar{N}_1 \zeta = 10^6$ counts the accuracy is 0.1%.

If in addition there are dark counts that are detected randomly at a rate \bar{N}_2 per second, then the variance in the number of dark counts in a time ζ is

$$\overline{(N_2 - \bar{N}_2)^2} \zeta = \bar{N}_2 \zeta \tag{22:48}$$

The relative accuracy with which the number of signal counts can be determined is

$$\frac{\overline{(N_1 + N_2 - \bar{N}_1 - \bar{N}_2)^2} \zeta}{\bar{N}_1 \zeta} = \frac{1 + \bar{N}_2/\bar{N}_1}{\bar{N}_1 \zeta} \tag{22:49}$$

An estimate of the limiting sensitivity of photon counting can be obtained by setting the relative accuracy to unity and choosing $\zeta = 1$ sec.

* For a more detailed discussion of this point see Chapter 24.

The minimum sensitivity is then

$$(\bar{N}_1)_{\min} = \frac{1}{2} \left(1 + \sqrt{1 + 4\bar{N}_2} \right); \quad (22:50)$$

and since generally $\bar{N}_2 \gg 1$,

$$(\bar{N}_1)_{\min} \approx \sqrt{\bar{N}_2}; \quad (22:51)$$

So, for example with a 500 nm source, and a photomultiplier with quantum efficiency of 20% and 10^2 dark counts per second the minimum detectable optical power is

$$P_{\min} = 50 \text{ h}^\circ = 2 \times 10^{-17} \text{ W} \quad (22:52)$$

22.6.3 Signal-to-noise Ratio in Direct Detection

In many applications of photomultiplier tubes the anode pulses are integrated to give a fluctuating analog current. The shot noise originating at the photocathode is, from Eq. (22.20)

$$\langle i_N^2 \rangle_c = 2e(\bar{i}_c + \bar{i}_d) C f; \quad (22:53)$$

where \bar{i}_c is the average photocathode current produced by a light source and \bar{i}_d is the average photocathode dark current. This noise is multiplied by the amplification of the electron number by interaction with the N dynodes of the tube. If each dynode has a secondary emission multiplication efficiency μ the overall gain of the tube is*

$$G = \mu^N \quad (22:54)$$

Because of statistical fluctuations in the secondary emission process and in addition because electrons can originate thermionically from the dynodes, the noise at the anode is further increased by a noise factor F . The overall noise appearing at the anode is

$$\langle i_N^2 \rangle_A = 2eG^2F(\bar{i}_c + \bar{i}_d) C f \quad (22:55)$$

For $\mu = 4$, and a 14 stage tube $G = 2.6 \times 10^8$, and typically $F = (\mu + 1) = 1.08$.

A photomultiplier is a current source: to convert this current to a detected voltage the amplified photoelectron current passes through an anode resistor of value R . This anode resistor may be part of the photomultiplier circuit, or may be provided in whole or in part by the input impedance of a following amplifier stage. The Johnson noise from the

* The first dynode often has a larger μ value than the others, but we will assume that μ is an average value.

resistor is

$$\langle i_N^2 \rangle_R = \frac{4kT \Delta f}{R} \tag{22:56}$$

To optimize detection of a signal it is common practice to amplitude modulate the signal at some angular frequency ω_m so as to permit synchronous detection at this frequency^[22:11]. The optical power reaching the photocathode can be represented in this case as

$$P = P_0 (1 + m \sin \omega_m t); \tag{22:57}$$

where m is a modulation parameter.

The average photocathode current is

$$\bar{i}_c = \frac{e \gamma P_0}{h\nu}; \tag{22:58}$$

and

$$i_c(t) = \bar{i}_c [1 + m \sin \omega_m t] \tag{22:59}$$

At the anode the time-varying part of this current is

$$i_s(t) = \bar{i}_c G m \sin \omega_m t \tag{22:60}$$

The signal-to-noise ratio (S=N) at the input to the electronics in Fig. (22.3) is therefore

$$\frac{\langle i_s^2 \rangle}{\langle i_N^2 \rangle_A + \langle i_N^2 \rangle_R} = \frac{\bar{i}_c^2 G^2 m^2}{2eG^2 F (\bar{i}_c + \bar{i}_d) \Delta f + 4kT \Delta f R} \tag{22:61}$$

The noise from the photomultiplier tube is usually sufficiently large that the Johnson noise can be neglected. If it is assumed that $\bar{i}_d \gg \bar{i}_c$; then for $m = 1$ and $S=N = 1$ we have, from Eqs. (22.58) and (22.61)

$$(P_0)_{\min} = \frac{2h\nu}{e^{1-2}} (F \bar{i}_d \Delta f)^{1-2}; \tag{22:62}$$

Example: Typical values for a good PMT will be $\gamma = 0.2$; $F \approx 1$; $\bar{i}_d \approx 10^{-15}$ A. For a 530 nm source Eq. (22.62) gives $(P_0)_{\min} = 3 \times 10^{-16}$ W:

22.6.4 Direct Detection with p-i-n Photodiodes

The shot noise from a p-i-n photodiode is

$$\langle i_N^2 \rangle_1 = 2e(\bar{i}_s + i_d) \Delta f; \tag{22:63}$$

where \bar{i}_s is the average signal current, and i_d is the dark current (usually specified for a low noise photodiode in units of nA/√Hz or pA/√Hz).

The average signal current is, in a similar way to Eq. (22.58)

$$\bar{i}_s = \frac{e \gamma P_0}{h\nu}; \tag{22:64}$$

where γ is the quantum efficiency, which is much larger than for a photomultiplier: values 0.7-0.8 are common. For the simple equivalent current

discussed previously (Fig. (22.3)) there is an additional Johnson noise contribution of magnitude

$$(i_N^2)_2 = \frac{4kT C f}{R} \quad (22:65)$$

The overall S=N ratio for direct detection of an unmodulated signal is

$$\frac{\langle i_s^2 \rangle}{\langle i_n^2 \rangle_1 + \langle i_n^2 \rangle_2} = \frac{(e P_0 h \nu)^2 R}{2eR(e P_0 h \nu + i_d) C f + 4kT C f} \quad (22:66)$$

This S=N ratio would be reduced by a factor $m^2=2$ for a modulated signal (cf. Eq.(22.61)).

If shot-noise dominates over dark current and Johnson noise, then the shot-noise-limited S=N ratio is

$$\frac{S}{N} = \frac{P_0}{2h \nu C f} \quad (22:67)$$

In a practical application using a photodiode there will be additional stages of electronic amplification that add noise. It is common to characterize the effect of an electronic circuit on the noise by its noise figure, F_N .

The noise figure can be defined conveniently as

$$F_N = \frac{\text{noise power at output of circuit}}{\text{amplified noise power at the input}} \quad (22:68)$$

For input Johnson noise the mean-square, amplified, output noise current is

$$\langle i_N^2 \rangle_1 = \frac{4kT G C f}{R}; \quad (22:69)$$

where G is the power gain of the amplifier within the frequency band being considered. The actual output mean-square noise current is

$$\langle i_N^2 \rangle_2 = \frac{4kT G F_N C f}{R} \quad (22:70)$$

It is as if the amplifier were noiseless but the input Johnson noise is increased by the noise figure.

The noise figure is frequently quoted in dB, a noise figure of 3 dB would represent a doubling of the output noise over the value expected from a noiseless amplifier. The term noise temperature, T_i , is also used defined by

$$F_N = 1 + \frac{T_i}{T_{\text{ambient}}}; \quad (22:71)$$

Example: We illuminate an InGaAs photodiode with a responsivity of 0.8A/W at 1.3 μ m in an optical communication link in which there is 30 dB of loss between a 10mW source and receiver. The system bandwidth is 100 MHz, the dark current is 5nA (equivalent to 0.5 pA/Hz).

We assume that the amplification electronics has a noise figure of 6 dB (relative to a 50 ohm input). We note the following:

Received power $P_0 = 10^{-3} \times 10\text{mW} = 10^{-1} \text{ W}$;

Signal current $\bar{i}_s = 8 \text{ A}$

Dark current $\bar{i}_d = 5\text{nA}$ (dominated by \bar{i}_s)

Signal Power = $(\bar{i}_s)^2 R = 3.2 \text{ nW}$

Shot noise power = $2e(\bar{i}_s + \bar{i}_d) \zeta f R = 1.28 \times 10^{-14} \text{ W}$

Johnson noise power = $4kT \zeta f = 4(1.38 \times 10^{-22}) (300) (10^8) = 1.66 \text{ pW}$

Effective Johnson noise power including noise figure = $10^{0.6} \times 1.66 \text{ pW} = 6.6 \text{ pW}$

In this case the Johnson noise is dominant, the effective signal-to-noise ratio is

$$\frac{S}{N} = \frac{3.2 \times 10^{-9}}{6.6 \times 10^{-12}} = 485:$$

22.6.5 Direct Detection with APDs

In an APD there is a multiplication of the number of charge carriers by a factor M. This multiplication can result from secondary ionizations produced by both electrons and holes. It is desirable that one or other of these charge carriers should have a significantly greater secondary ionization coefficient* than the other^[22:18]. The current in an APD increases by the factor M but the associated shot noise increases further because in a given avalanche process M will fluctuate, taking values M; M ± 1; M ± 2, etc. The mean square noise current thereby increases, not by a factor M², but by a factor FM², where F is called the noise factor. In silicon APDs F typically lies in the range 2-20. Therefore, the shot noise becomes

$$\langle i_N^2 \rangle_1 = 2eFM^2(\bar{i}_s + i_d) \zeta f \tag{22:72}$$

Both photo- and dark-generated carriers contribute to the shot noise. The overall S=N ratio is modified from Eq. (22.66) and becomes

$$\frac{\langle i_s^2 \rangle}{\langle i_N^2 \rangle_1 + \langle i_N^2 \rangle_2} = \frac{M^2 (e^{\zeta} P_0 = h^{\circ})^2 R}{2eRFM^2(e^{\zeta} P_0 = h^{\circ} + i_d) \zeta f + 4kT \zeta f} \tag{22:73}$$

Eq. (22.73) shows that the S=N ratio improves with increasing multiplication as the Johnson noise contribution becomes less important until

* The secondary ionization coefficient is the number of secondary electron-hole pairs produced per unit length by a specific charge carrier (electron or hole) in travelling through the material.

the avalanche shot noise becomes dominant. The avalanche limited $S=N$ ratio is

$$\frac{S}{N} = \frac{P_0}{2Fh\nu Cf}; \quad (22:74)$$

a reduction of the signal-noise-ratio from the quantum limited value by the noise factor.

The NEP can be computed for both p-i-n photodiodes and APDs by setting $S=N = 1$ in equations like (22.66) and (22.73) and thereby the minimum input power for $S=N = 1$ determined. For $P_0 = P_{\min}$, we expect the dark current to be larger than the signal current so Eq. (22.73) becomes

$$\frac{S}{N} = \frac{(Me^{-P_0/h\nu})^2}{[2eFM^2i_d + 4kT=R]Cf} \quad (22:75)$$

For $S=N = 1$ and a 1 Hz bandwidth $P_0 = P_{\min} = \text{NEP}$, which gives for an APD

$$\text{NEP}(W=\overline{P_{\text{Hz}}}) = \frac{h\nu}{Me^{-P_0/h\nu}}(2eFM^2i_d + 4kT=R)^{1/2}; \quad (22:76)$$

The equivalent result for a p-i-n diode can be obtained by setting $M = F = 1$.

Example. For an InGaAs APD with $F=10$, $M=100$, $R=50$ ohm, $i_d = 2\text{nA}$, and responsivity 0.8 A/W , the quantum efficiency is

$$\eta = \frac{0.8(6.626 \times 10^{-34})(3 \times 10^8)}{(1.6 \times 10^{-19})(1.55 \times 10^{-6})} = 0.64 \quad (22:77)$$

From Eq. (22.73) the NEP is

$$\begin{aligned} \text{NEP}(W=\overline{P_{\text{Hz}}}) &= \\ &= \frac{(2 \times 1.6 \times 10^{-19} \times 10^5 \times 2 \times 10^{-9} + 4 \times 1.38 \times 10^{-22} \times 300=50)^{1/2}}{0.8 \times 100} \\ &= 4.7 \times 10^{-13} W=\overline{P_{\text{Hz}}}; \end{aligned}$$

Note that in this case the thermal noise is dominant.

22.7 Coherent Detection

We have seen several examples in this chapter of how the signal-to-noise ratio of an optical detection system is limited by noise, in particular by Johnson noise and the electronic noise of the amplification stages that follow an optical detector. Even if these sources of noise were not present the signal-to-noise ratio would be limited by shot noise. In the direct detection schemes discussed so far the quantum noise limit set

Fig. (22.25).

by shot noise is rarely attainable. However, it is possible to achieve the quantum limit for detection by the use of coherent detection. In this scheme the detector is illuminated simultaneously by the signal light and by a second source of light called the local oscillator, (l.o.) which must be phase coherent with the signal. The degree of phase coherency that is required to make this scheme work well has been discussed in detail by Salz^[22:19].

A useful rule of thumb is that the local oscillator must be phase coherent over a time long enough to receive the information being transmitted. In an optical communication system in which binary information is being detected the required phase coherence time is on the order of the pulse duration being detected.

The schematic way in which coherent detection is carried out is shown in Fig. (22.25). For optimum performance not only must signal and l.o be phase coherent, but their phase fronts and polarization states must be matched at the detector surface. This corresponds, for example, to a situation in which two linearly polarized TEM₀₀ Gaussian laser beams, which are coaxial and of equal spot size, coincide in a beam waist at the detector. Deviations from this ideal geometry, either through spot size mismatch, angular or lateral misalignment, decrease the efficiency of the detection process.

We represent the electric field of the signal beam at the detector surface as

$$E_s(t) = E_1 \cos(\omega_1 t + \hat{A}_1) \quad (22:78)$$

and of the local oscillator as

$$E_{lo}(t) = E_2 \cos(\omega_2 t + \hat{A}_2) \quad (22:79)$$

The detector responds to the intensity of the light falling on it. There-

fore, the detector current is

$$i(t) = \frac{RA}{Z} [E_s(t) + E_{lo}(t)]^2; \quad (22:80)$$

where R is the detector responsivity, A is the effective detector area, and Z is the characteristic impedance of the medium in front of the detector. Substitution from Eqs. (22.78) and (22.79) into (22.80) gives

$$i(t) = \frac{RA}{Z} [E_1^2 \cos^2(\omega_1 t + \phi_1) + E_2^2 \cos^2(\omega_2 t + \phi_2) + 2E_1 E_2 \cos(\omega_1 t + \phi_1) \cos(\omega_2 t + \phi_2)] \quad (22:81)$$

By the use of well-known trigonometrical identities* the detector current can be written as

$$i(t) = \frac{RA}{Z} \left[\frac{E_1^2}{2} + \frac{E_1^2}{2} \cos 2(\omega_1 t + \phi_1) + \frac{E_2^2}{2} + \frac{E_2^2}{2} \cos 2(\omega_2 t + \phi_2) + E_1 E_2 \cos[(\omega_1 + \omega_2)t + \phi_1 + \phi_2] + E_1 E_2 \cos[(\omega_1 - \omega_2)t + \phi_1 - \phi_2] \right] \quad (22:82)$$

The detector does not respond to the rapidly oscillating terms at frequencies ω_1 ; ω_2 , and $(\omega_1 + \omega_2)$ in Eq. (22.82) y, so the detector current is

$$i(t) = \frac{RA}{Z} \left[\frac{E_1^2}{2} + \frac{E_2^2}{2} + E_1 E_2 \cos[(\omega_1 - \omega_2)t + \phi_1 - \phi_2] \right]; \quad (22:83)$$

We are assuming that the difference frequency $(\omega_1 - \omega_2)$ is within the response range of the detector. The generation of this difference frequency is called optical mixing. If $\omega_1 = \omega_2$ it is a homodyne mixing process, otherwise it is a heterodyne process. Similar processes occur in the operation of FM radio receivers. Note that the first two terms in Eq. (22.83) correspond to the detector current resulting from signal and local oscillator independently.

The detector shot noise is

$$\langle i_N^2 \rangle = \frac{2eRA}{Z} \left(\frac{E_1^2}{2} + \frac{E_2^2}{2} \right) + i_d^2 \quad (22:84)$$

where i_d is the detector dark current. In addition we expect a Johnson

* $\cos^2 x = \frac{1}{2}(1 + \cos 2x)$; $\cos X + \cos Y = 2 \cos\left(\frac{X+Y}{2}\right) \cos\left(\frac{X-Y}{2}\right)$.

y In a more advanced analysis that reflects the fact that optical detectors respond to a light signal by absorbing photons these high frequency terms do not even appear. See Chapter 24 for a further discussion.

noise contribution of

$$\langle i_N^2 \rangle_2 = \frac{4kT \zeta f}{R} \tag{22:85}$$

It is usual to operate a coherent detection scheme in which $E_2 \gg E_1$: a weak signal is mixed with a strong local oscillator. Therefore, the effective signal current is

$$i_s(t) = \frac{RAE_1E_2}{Z} \cos[(\omega_1 + \omega_2)t + \hat{A}_1 + \hat{A}_2] \tag{22:86}$$

Eq. (22.86) provides essential insight into the desirability of coherent detection for detection of a weak coherent optical field E_1 . The signal current generated when this weak optical field is mixed with a local oscillator field E_2 is multiplied by the magnitude of the local oscillator field.

If $\omega_1 \approx \omega_2$ Eqs. (22.84)-(22.86) give a signal-to-noise ratio

$$\frac{\langle i_s^2(t) \rangle}{\langle i_N^2 \rangle_1 + \langle i_N^2 \rangle_2} = \frac{R^2 A^2 E_1^2 E_2^2}{2Z^2} \cdot \frac{eRA}{Z} (E_1^2 + E_2^2) + i_d + \frac{4kT}{R} \zeta f \tag{22:87}$$

With a sufficiently powerful local oscillator its shot noise dominates and the dark current contribution and Johnson noise can be neglected. In this case, for $E_2 \gg E_1$ Eq. (22.87) reduces to

$$\frac{S}{N} = \frac{RAE_1^2}{2eZ \zeta f} \tag{22:88}$$

Noting that the average signal power is

$$P_s = \frac{AE_1^2}{2Z} \tag{22:89}$$

and that $R = e/h\nu$, Eq. (22.88) becomes

$$\frac{S}{N} = \frac{P_s}{h\nu \zeta f} \tag{22:90}$$

It might appear from a comparison of Eqs. (22.67) and (22.90) that ideal heterodyne detection has quantum limited performance twice that of direct detection. However, to make such a comparison one must assume that it is possible to use the same signal detection bandwidth ζf in both schemes. If an unmodulated CW laser beam of power P_s is to be detected then Eq. (22.67) would predict that the direct detection (dd) photon-noise limited signal-to-noise ratio would be

$$\frac{S}{N}_{dd} = \frac{P_s}{2h\nu \zeta f} \tag{22:91}$$

However, in practice one could not hope to achieve this S=N ratio because the resultant dc detector signal would be overwhelmed by 1=f noise.

It is always necessary to modulate the signal in some way. For example, if the signal power is

$$P = P_0(1 + m \sin \omega_m t) \quad (22:92)$$

the detected signal at frequency ω_m could be band-pass filtered. In this case the effective signal to noise ratio is

$$\frac{S}{N_{dd}} = \frac{m^2 P_s}{4h\nu C f} \quad (22:93)$$

Particular care must be taken in comparing a homodyne receiver with a heterodyne receiver. If both signal and local oscillator are CW signals at the same frequency it is meaningless to talk of detection of the signal, since signal and local oscillator are essentially indistinguishable. To discuss the signal-to-noise ratio it is essential to consider how the signal beam is being modulated, since it is through modulation that information is transferred. To illustrate this we consider a situation in which the signal beam is phase modulated so that

$$E_s(t) = E_1 \cos(\omega_1 t + m \sin \omega_m t) \quad (22:94)$$

In this case the information being transferred is the sinusoidal phase modulation at frequency ω_m . For small values of m , which is called the modulation depth, Eq. (22.94) can be rewritten as*

$$E_s(t) = E_1 [J_0(m) \cos \omega_1 t + J_1(m) \cos(\omega_1 + \omega_m)t + J_1(m) \cos(\omega_1 - \omega_m)t] \quad (22:95)$$

The signal beam has acquired sidebands at frequencies separated by $\pm \omega_m$ from the carrier frequency ω_1 .

The mixing process, in this case, can be written as

$$i(t) = \frac{RA}{Z} [E_1 \cos(\omega_1 + m \sin \omega_m t) + E_2 \cos(\omega_2 + \omega_1 t)]^2 \quad (22:96)$$

By the use of Eq. (22.95), and with the assumption that $m \ll 1$, so that $[J_1(m)]^2 \ll J_1(m)$, Eq. (22.95) reduces to

$$i(t) = \frac{RA}{Z} \left[\frac{E_2^2}{2} + \frac{E_1^2 J_0^2(m)}{2} + 2J_1(m)E_1E_2 \sin[(\omega_1 - \omega_2 - \omega_1)t] \sin \omega_m t \right] \quad (22:97)$$

after the (non-existent) high-frequency terms are eliminated. The useful signal current contained in Eq. (22.97) is

$$i_s(t) = \frac{2RAJ_1(m)}{Z} E_1E_2 \sin[(\omega_1 - \omega_2 - \omega_1)t] \sin \omega_m t \quad (22:98)$$

The information term $\sin \omega_m t$ is present as a low-frequency modulation

* See Chapter .

of the intermediate frequency (i.f.) ($\omega_1 \pm \omega_2$). In a similar way to Eq. (22.87) the signal-to-noise ratio is

$$\frac{\langle i_s^2(t) \rangle}{\langle i_{N>1}^2 \rangle + \langle i_{N>2}^2 \rangle} = \frac{R^2 A^2 J_1^2(m) E_1^2 E_2^2}{Z^2 \left[\frac{eRA}{Z} E_2^2 + i_d + \frac{4kT}{R} \right] C f}; \tag{22:99}$$

where we have made the assumption that $E_1 \ll E_2$, as is usually the case in a heterodyne receiver.

For a sufficiently powerful local oscillator Eq. (22.99) gives

$$\frac{S}{N} = \frac{RAJ_1^2(m)E_1^2}{eZCf}; \tag{22:100}$$

For homodyne detection, we would optimize performance by setting $A_2 = 1/2$ in Eq. (22.98). In this case the signal-to-noise ratio is

$$\frac{S}{N}_{\text{homodyne}} = \frac{2RAJ_1^2(m)E_1^2}{eZCf} \tag{22:101}$$

In this situation the homodyne $S=N$ is twice the value achieved by heterodyne detection. It should be stressed, however, that the $S=N$ in coherent detection schemes depends on the precise modulation scheme, and also on the contribution to the signal-to-noise ratio of the demodulation scheme used to extract the information term ($\sin \omega_m t$ in Eq. (22.98), from the i.f. signal at $(\omega_1 \pm \omega_2)$ ^{[22:16];[22:20]}. A discussion of the merits of various demodulators, involving band-pass filters, discriminations, limiters, and phase-locked loops is beyond our scope here.

We can note that the signal power in the two sidebands in Eq. (22.95) can be written as

$$P_s = \frac{E_1^2 J_1^2(m)}{Z} \tag{22:102}$$

so that in terms of the sideband power the heterodyne $S=N$ is, as before

$$S=N = \frac{P_s}{h^\circ C f}; \tag{22:103}$$

Example: An interesting example of the coherent detection process is provided under conditions of suppressed carrier operation. If the modulation depth in Eq. (22.94) is set to a value $m = 2.405$, then $J_0(m) = 0$ and the carrier is suppressed. For a detector with $\gamma = 0.6$, and bandwidth 100 MHz operating at 1.55 μ m the minimum detectable signal power | defined to correspond to $S=N = 1$ is

$$P_s = \frac{h^\circ C f}{1}; \tag{22:104}$$

which in this case gives

$$P_s = \frac{(6.626 \times 10^{-34})(3 \times 10^8)(10^8)}{0.6(1.55 \times 10^{-6})} = 21\text{pW}$$

Fig. (22.26).

22.8 Bit-Error Rate

We conclude this chapter with a discussion of a final performance parameter that is widely used to characterize the performance of optical communication systems. This is the bit-error-rate, used to describe the performance of an optical detection system receiving binary-encoded signals. In its simplest form such a scheme has a binary "one" represented by the detection of an optical pulse by the detector. A binary "zero" is represented by the failure to detect such a pulse. The "ones" and "zeros" constitute a bit stream at a prescribed clock rate, as shown schematically in Fig. (22.26)(a) and (b). The detection of a "one" is characterized by the detector signal raising above a threshold level γ_1 . The detection of a "zero" is characterized by the detector signal failing to rise above a lower threshold level γ_2 . In the presence of additive noise, represented schematically by Fig. (22.26)(c), the overall detector signal will appear as shown schematically in Fig. (22.26)(d). The questions that then arise are: when will the noise reduce the detector signal below γ_1 when a "one" is present, and when will the noise be above level γ_2 when a "zero" is being detected. Both of these possibilities lead to a bit detection error, as shown schematically in Fig. (22.26)(e). One simple way to model this process is to assume that the detector gives rise to a noise current that fluctuates in a Gaussian fashion | more sophisticated analyses would examine more closely the statistical nature of both the detection and noise processes and of the different ways in which the binary signal is encoded^{[22:16];[22:20]}.

For a detector noise current with variance $\langle i_N^2 \rangle$ that is Gaussian distributed about zero the probability that the noise is below a set level

is

$$P(i_N < i_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-i_1} e^{-x^2/2} dx \quad (22:105)$$

where $i_1^2 = \langle i_N^2 \rangle$.

On average a simple binary stream of "ones" and "zeros" will have equal numbers of ones and zeros. Therefore, for a large number N of transmitted bits there will be N/2 "ones" and N/2 "zeros". There will be a bit error whenever a "one" is present but the noise current is $< i_1$; there will also be a bit error whenever a "zero" is present but the noise is above level i_2 . Therefore, we can write the probability of error as

$$p_e = \frac{1}{2} [P(i_N < i_1) + P(i_N > i_2)] \quad (22:106)$$

A simple result can be obtained if $i_1 = i_2 = i_s/2$, where i_s is the detector current corresponding to a "one". In this case

$$p_e = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-i_s/2} e^{-x^2/2} dx + \frac{1}{\sqrt{2\pi}} \int_{i_s/2}^{\infty} e^{-x^2/2} dx; \quad (22:107)$$

which becomes

$$p_e = \frac{1}{\sqrt{2\pi}} \int_{i_s/2}^{\infty} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{i_s/(2\sqrt{2})}^{\infty} e^{-t^2} dt \quad (22:108)$$

Eq. (22.108) can be written in terms of the error function, erf(z), defined by the relation^[22:21]

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \quad (22:109)$$

where erf z → 1 as z → ∞. Therefore,

$$p_e = \frac{1}{2} \left[1 - \text{erf}\left(\frac{i_s}{2\sqrt{2}}\right) \right] = \frac{1}{2} \text{erfc}\left(\frac{i_s}{2\sqrt{2}}\right) \quad (22:110)$$

and finally

$$p_e = \frac{1}{2} \text{erfc}\left(\frac{i_s}{2\sqrt{2}}\right) \quad (22:111)$$

The signal-to-noise ratio can be characterized by

$$\frac{S}{N} = \left(\frac{i_s}{\text{RMS noise current}} \right)^2; \quad (22:112)$$

so

$$p_e = \frac{1}{2} \text{erfc}\left(\frac{1}{\sqrt{2}} \sqrt{\frac{S}{N}}\right) \quad (22:113)$$

Fig. (22.27) shows a plot of this probability of error, or bit-error-rate, as a function of S=N. A frequently used performance measure is $P_e < 10^{-9}$, which implies S=N = 141, or 21.5dB.

Fig. 22.27).

In a practical test of bit-error-rate a random, long sequence of binary pulses (10^{15} ; 10^{21}) would be transmitted from source to receiver and the received bit sequence compared with a record of the transmitted sequence.

References

- [22.1] van der Ziel, A. (1954). Noise Prentice-Hall, New York.
- [22.2] Nyquist, H. (1928). Phys. Rev. 32, 110.
- [22.3] Ramo, S., Whinney, J.R. & Van Duzer, T. (1989). Fields and Waves in Communication Electronics, 2nd edition, Wiley, New York.
- [22.4] Bleaney, B.I. & Bleaney, B. Electricity and Magnetism, Oxford U.P., Oxford.
- [22.5] Putley, E.H. (1969). "Indium Antimonide Submillimeter Photoconductive Detectors," Appl. Opt. 4, 649-656.
- [22.6] Putley, E.G. Optical and Infrared Detectors, "Thermal Detectors".
- [22.7] R.J. Keyes, Ed., Topics in Applied Physics, Vol. 19, Springer-Verlag, Berlin, 1977.
- [22.8] J. Gowar, (1984), Optical Communication Systems, Prentice-Hall, Englewood Cliffs, New Jersey.
- [22.9] H. Kressel, Editor, Semiconductor Devices for Optical Communication, Topics in Applied Physics, Vol. 39, Springer-Verlag, Berlin, 1982.
- [22.10] H. Melchior, M.B. Fisher, and F.R. Arams, "Photodetectors for Optical Communication Systems, Proc. IEEE, 58, 1466-1486, 1970.
- [22.11] J.H. Moore, C.C. Davis, and M.A. Coplan, Building Scientific Apparatus, 2nd Edition, Addison Wesley, Reading, Massachusetts, 1989.
- [22.12] M.J.E. Golay, "A Pneumatic Infra-Red Detector," Rev. Sci. Instrum., 18, 357-362, 1947.
- [22.13] H. Taub, and D.L. Schilling, (1971), Principles of Communication Systems, McGraw-Hill, New York.
- [22.14] H. Kressel, (Editor) (1982), Semiconductor Devices for Optical Communication, Topics in Applied Physics, Vol. 39, Springer-Verlag, Berlin..
- [22.15] P.K. Cheo, (1990), Fiber Optics and Optoelectronics, 2nd Edition, Prentice-Hall, Englewood Cliffs, New Jersey..
- [22.16] R.M. Gagliardi and S. Karp, Optical Communications, Wiley, New York, 1976.

- [22.17] H.B. Killen, (1991), Fiber Optic Communications, Prentice-Hall, Englewood Cliffs, New Jersey.
- [22.18] R.J. McIntyre, "Multiplication noise in uniform avalanche diodes," IEEE Trans. Electron. Devices, ED-13, 164-168, 1966.
- [22.19] J. Salz, "Coherent lightwave communications," AT&T Tech. J. 64, 2153-2209, 1985.
- [22.20] Paul E. Green, Jr., Fiber Optic Networks, Prentice-Hall Englewood Cliffs, New Jersey, 1993.
- [22.21] M. Abramowitz and I.A. Stegun, Handbook of Mathematical Functions, Dover, New York, 1965.

Figure Captions

- Fig. (22.1). Photoexcitation at a p ; n junction.
- Fig. (22.2). Circuits used in an analysis of Johnson noise, (a) antenna connected to a resistor, (b) including the radiation resistance of the antenna.
- Fig. (22.3). Equivalent circuit of a shot-noise-limited detector driving a resistive load.
- Fig. (22.4). Schematic variation of noise with frequency in a semiconductor detector.
- Fig. (22.5). Band structure of (a) a metal vacuum interface and (b) a pure-semiconductor-vacuum interface (\bar{A} = work function; \hat{A} = electron affinity; E_g = band-gap energy; E_p = Fermi level).
- Fig. (22.6). Typical anode pulse produced by a single photoelectron emission at the cathode of a photomultiplier tube, t is the time following photoemission. The transit-time T , the transit-time speed cT , and the peak anode current i_A^0 all fluctuate from pulse to pulse.
- Fig. (22.7). Schematic diagram of the internal structure of various types of photomultiplier tube: (a) squirrel cage; (b) box and grid; (c) Venetian blind; (d) focused dynode. D=dynode; PC - photocathode.
- Fig. (22.8). Wavelength dependence of a radiant sensitivity of several commercially available photocathode materials.
- Fig. (22.9). Schematic photomultiplier anode pulse-height distributions: (a) two forms likely to be observed in practice; (b) idealized form from tube with dynodes having a high and well-defined secondary emission coefficient. The best signal/noise ratio would be obtained in a photo-

counting experiment by collecting only anode pulses in a height range roughly indicated by the shaded region AB.

Fig. (22.10). Mechanism for (a) intrinsic photoconductivity; (b) extrinsic photoconductivity.

Fig. (22.11). $D^{\pi}(\lambda)$ as a function of wavelength for various photoconductive detectors. (Courtesy of Hughes Aircraft Company).

Fig. (22.12). Radiation-shielded liquid-nitrogen-cooled photoconductive or photovoltaic infrared detector assembly.

Fig. 22.13). Simple biasing circuit for operating a photoconductive detector with modulated radiation.

Fig. (22.14). D^{π} as a function of wavelength for various lead-salt photoconductive detectors. (Courtesy of Hughes Aircraft Company).

Fig. (22.15). Spectral responsivities of some important near-infrared photodiodes.

Fig. (22.16). D^{π} as a function of wavelength for various photovoltaic detectors. (Courtesy of Hughes Aircraft Company).

Fig. (22.17). Photovoltaic detector operated in (a) open-circuit mode; (b) reversed-biased mode.

Fig. (22.18). (a) Current-voltage characteristics of a photodiode; (b) current-mode-operated photodiode.

Fig. (22.19). Schematic construction of a p-i-n photodiode. The lightly n-doped intrinsic layer is frequently designated as a ω -layer.

Fig. (22.20). Illustration of avalanche electron-hole pair production by an electron crossing a reverse-biased p-n junction. The Fermi energies are separated by the brass voltage V_B across the junction. An electron crossing the junction from A to B has sufficient energy to excite an electron hole pair and simultaneously face to energy C.

Fig. (22.21). Schematic construction of a planar n-i-p-i-p⁺ APD with a guard ring and showing typical dimensions.

Fig. (22.22).

Fig. (22.22). Bridge operating circuit for thermistor bolometers using compensating shielded thermistor, with device construction shown.

Fig. (22.24). Schematic design of the Golay detector. The top half of the line grid is illuminated by the LED and imaged back on the lower half of the grid by the ω exible mirror and meniscus lens. Any radiation-induced deformation of the ω exible mirror moves the image of the line grid and changes the illumination reaching the photodiode.

Fig. (22.25). Schematic arrangement for optical mixing.

Fig. (22.26). Binary signal transmission showing clock, data, the effect of added noise and the resultant detection of "ones" and "zeros".

Fig. (22.27). Bit-error-rate as a function of signal-to-noise ratio in a simple binary detection system.