Fundamentals of Power and Energy Measurement

Power and energy detectors can be grouped into two broad categories: thermal detectors and quantum detectors. Thermal detectors simply absorb the incident radiation, increasing the detecting element's temperature until the combined effects of conduction and thermal radiation from the detector are in equilibrium with the energy being absorbed by the detector. The two main types of thermal detectors are thermopiles, used primarily to measure the power of cw laser beams, and pyroelectric detectors, used to measure the energy in a laser pulse. Because thermal detectors measure only the heat generated in the detector, they are extremely broadband, typically with a flat spectral response from 200 nm to 20 μm and beyond.

Quantum detectors operate on a completely different principle. The detector is a semiconductor, and the incident radiation excites electrons from the semiconductor's valence band into the conduction band, generating a current that is proportional to the number of photons in the incident radiation. The gap between the two bands is well defined, and only photons with sufficient energy will be able to move an electron from the valence band to the conduction band. The critical energy λ_c is given by

$$\lambda_c = \frac{hc}{E_g} \qquad (11.1)$$

where h is Planck's constant, c is the speed of light in the semiconductor, and E_g is the energy gap. There are two basic types of quantum detectors: photoconductors, used primarily for measurements in the near infrared (to 5 μm), and photodiodes, used primarily for visible and ultraviolet wavelengths. The sensitivity of quantum detectors is very high, allowing measurements in the picowatt range, and their response time can be much, much faster than that of thermal detectors. On the other hand, their wavelength sensitivity is very nonlinear, their effective wavelength range is much narrower, and they are easily damaged by higher power laser beams.

THERMOPILE DETECTORS

A thermopile detector consists of two sets of thermocouples connected in series (see figure 11.1). One set of thermocouples is attached to an absorbing disc (the detector) and the other set is attached to the case (ambient temperature). As incident radiation is absorbed by the disc, the disc's temperature rises, generating a voltage that is directly proportional to the difference in temperature between the disc and the case. The readout unit then interprets the voltage measurement and presents the information in units of watts (power) or joules (energy).

The detector disc is typically made from graphite or aluminum with a durable, black, absorbing finish. The responsivity of the detector primarily depends upon the thermal mass of the detector disc and the thermocouples mounted to it. The energy is absorbed at the surface and must have time to spread through the detector and bring the disc to a uniform temperature. For larger detectors, used primarily for high-power lasers, this can take several seconds; for smaller detectors, a fraction of a second is typical. To counteract this slow response time, manufacturers often incorporate circuitry in the readout electronics that analyzes the slope of the temperature rise and anticipates
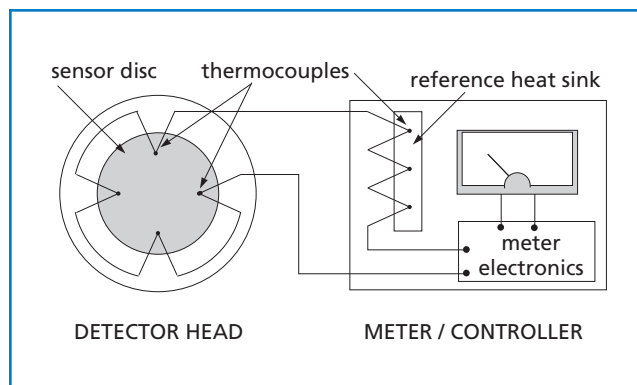

Figure 11.1   **Thermopile Detector**

the equilibrium point. Because of the slow response, thermopile detectors are used primarily for measuring cw output power and average power for pulsed lasers. Peak pulse power and pulse energy measurements are very questionable at rates above a few pulses per second.

PYROELECTRIC DETECTORS

Pyroelectricity and piezoelectricity are closely related. In fact, most piezoelectric crystals also exhibit pyroelectricity. When a pyroelectric crystal is heated or cooled, the stress generated by expansion or contraction generates a voltage across the crystal which is proportional to the rate at which the energy is being absorbed. Either this voltage, or the surface charge on the crystal, can then be measured by a readout unit and converted to watts (peak power) or joules (pulse energy). Unlike a thermopile detector, at thermal equilibrium the voltage across the crystal and the surface charge dissipate, making these detectors unsuitable for cw applications. Detector systems that read the voltage across the crystal are limited to relatively low pulse repetition rates (10's of Hz) because the heat must permeate through the thermal mass of the crystal. Detector systems that measure the surface charge can operate at high pulse rates (>1 kHz) because only the temperature of the crystal at the surface is important; however, at low repetition rates (<10 Hz) the response falls off markedly.

PHOTOCONDUCTORS

Photoconductive detectors are heavily doped semiconductors that have a finite electrical conductivity that increases with temperature. They are used primarily for infrared detection. Typically, the detectors are biased at a fixed voltage, and the current flowing through the detector provides the signal. One problem with photoconductive detectors is that a small change in ambient temperature may significantly increase the dark current (the current flowing through the device when no light is present), swamping the signal. Consequently, either the temperature of the detector must be stabilized or some sort of phase-sensitive detection scheme must be used.

PHOTODIODES

Photodiodes have complex electrical characteristics and can best be understood using the concept of the equivalent circuit—a circuit of individual components (resitors, capacitors, etc.) whose collective (lump-sum) behavior models that of the actual photodiode.

The ideal photodiode can be considered as a current source parallel to a semiconductor diode. The current source corresponds to the current flow caused by the light-generated drift current, while the diode represents the behavior of the junction in the absence of incident light.

An actual photodiode is represented by the equivalent circuit shown in figure 11.2.

In addition to a current source in parallel with a semiconductor diode, a nonconductive layer devoid of carriers (depletion layer) is sandwiched between two conductive layers. This is a classic parallel-plate capacitor, which can support charge separation only in one direction. The effective capacitance, termed the junction capacitance (C_j), is represented in the equivalent circuit by a capacitor in parallel with the other components. The photodiode junction also has finite shunt resistance (R_{sh}). Ancillary parts of the diode (neutral layers, electrical contacts) also give rise to a resistance, usually much smaller than the shunt resistance. This resistance acts between the diode junction and the signal sensing circuit and is therefore termed the series resistance (R_s). The series resistance can usually be assumed to equal zero for modeling and computational purposes.

Photodiode Operation

A photodiode behaves like a photocontrolled current source in parallel with a semiconductor diode and is governed by the standard diode equation:

$$I = I_{\text{photo}} + I_{\text{dk}}\left(e^{qV_0/kT} - 1\right) \qquad (11.2)$$

where

I is the total device current
I_{photo} is the photocurrent
I_{dk} is the dark current
V_0 is the voltage across the diode junction
q is the charge of an electron
k is Boltzmann's constant
T is the temperature in degrees Kelvin.

The I-V (current-voltage) relationship of this equation is shown in figure 11.3. Two significant features to note from both the curve and the equation are that the photogenerated current (I_{photo}) is additive to the diode current, and the dark current is merely the diode's reverse leakage current. Finally, the detector shunt resistance is the slope of the I-V curve (dV/dI) evaluated at $V=0$.
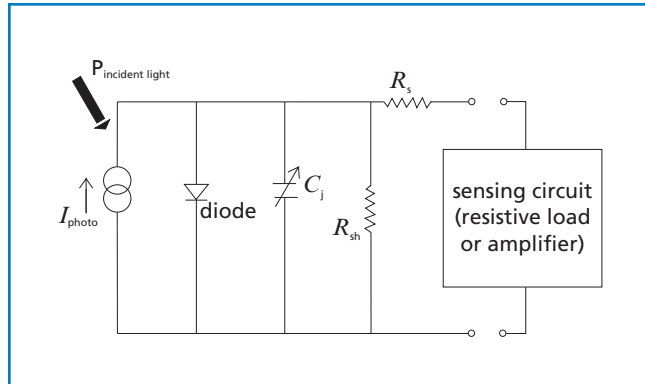

Figure 11.2 **Lumped-sum equivalent-circuit model of a photodiode**

Quantum Efficiency

Photodiodes are quantum devices. Each incoming photon will generate either one or zero units of electron charge which will contribute to the photocurrent. The probability of generating a charge is termed the quantum efficiency (η). Quantum efficiency mainly depends on how efficiently charge carriers are swept across the junction. The overall quantum efficiency of the photodiode is often referred to as external quantum efficiency.
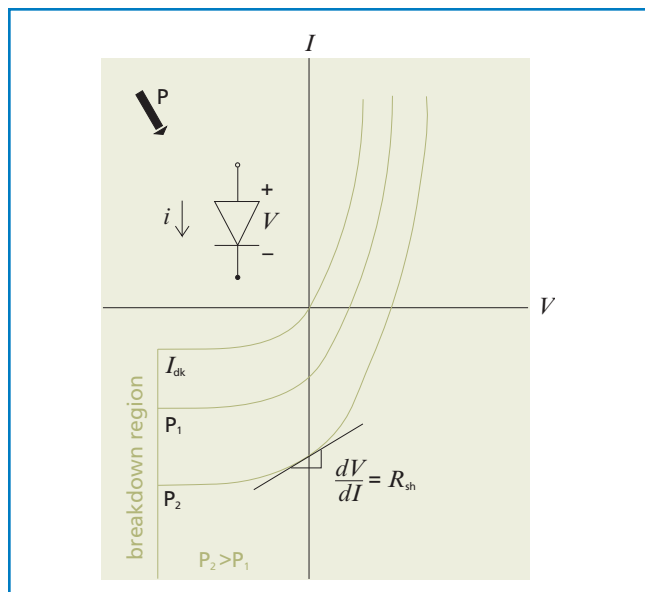

Figure 11.3 **The I-V relationship of a photodiode**

Responsivity

Responsivity ($\mathfrak{R}$) quantifies the photoelectric gain of a detector. Photodiode responsivity is the ratio of the photocurrent (across an effective zero resistance) generated for each watt of incident light power, expressed as amps/watt (A/W). Responsivity depends directly on the quantum efficiency. The maximum theoretical achievable responsivity corresponds to detection of every incident photon (unit quantum efficiency). The energy carried by each photon depends on its frequency according to the equation $E = h\nu$, where ν is the photon frequency (inversely proportional to λ, its wavelength) and h is Planck's constant. Therefore, expressing the responsivity in A/W (as opposed to A/photon) gives this parameter an inherent wavelength dependency:

$$\mathfrak{R} = \frac{q\eta\lambda}{hc}$$

or   (11.3)

$$\mathfrak{R} = \frac{\eta\lambda}{1.24 \times 10^{-6}} \, \text{A/W}$$

where

c is the speed of light

q is the charge of an electron

λ is the wavelength in meters of the photons being detected.

Responsivity has an additional wavelength dependence arising from the variation of quantum efficiency with wavelength. At wavelengths where silicon does not absorb strongly, photons may penetrate more deeply into the device (or pass through it), leading to minority carrier generation too remote from the junction to be detected (i.e., lower quantum efficiency). The typical shape of the silicon photodiode responsivity spectral curve is determined by the absorption spectrum of silicon (see figure 11.4). Photodiode responsivity is usually specified at a single wavelength unless a complete wavelength calibration is performed.

As an example, a silicon photodiode normally will have a high quantum efficiency for light at 800 nm. Assuming a typical η of 0.8 at this wavelength, this leads to a responsivity of 0.52 A/W. At 400 nm a typical η would be only 0.15, which leads to a responsivity of 0.1 A/W.

Responsivity alone is a weak figure of merit because it specifies only the gain of a photodiode, not its associated noise. Signal-to-noise ratio represents the ultimate figure of merit, as discussed later.

Linearity of Response

In many applications it is necessary for the responsivity to remain constant over a wide range of incident light power. Most P-N silicon photodiodes are linear (better than 1%) over seven or eight orders of magnitude. However, when the number of incident photons becomes comparable to the number of electron-hole locations in the active region, the device saturates, resulting in a loss of linearity.
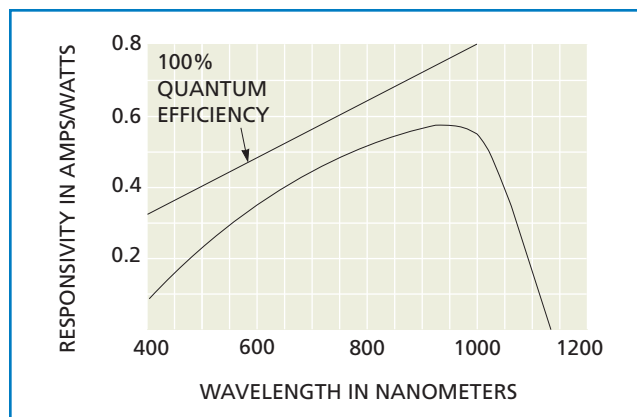

Figure 11.4   Typical responsivity of a CVI Melles Griot silicon photodiode

In the case of the CVI Melles Griot silicon photodiodes, the photocurrent will have a linear relationship to the incident intensity, providing the current density is less than 5×10^{-5} A/mm^2. The maximum allowable radiant intensity is therefore given by:

$$\text{Intensity} < \frac{5 \times 10^{-5}}{\mathfrak{R}} \, \text{W/mm}^2$$

where $\mathfrak{R}$ is the responsivity at the detection wavelength in A/W.

It is important to understand this limitation when using a photodiode to detect laser radiation. The raw output beam from a 1-mW HeNe laser has more than twice the intensity (W/mm^2) necessary to saturate a typical silicon photodiode.

System linearity is also affected by the sensing circuit. Incident light on the photodiode's active area produces a photocurrent which is usually measured by the amount of voltage dropped across an external resistance of known size. As the resultant voltage in the sensing circuit increases, the photodiode becomes forward biased, leading to nonlinear response.

Junction Capacitance

When designing a sensing circuit to maximize the speed or linearity of response, one must know two important electrical characteristics of a photodiode: the junction capacitance (C_j) and the shunt resistance (R_{sh}). Without these, the RC time constant of the complete operating circuit cannot be calculated.

The parallel-plate capacitance across the depletion region gives rise to a junction capacitance which increases with the area of the junction. Since increasing capacitance in a circuit slows its speed of response, photodiodes with smaller active areas are inherently capable of faster response than those with larger active areas. The junction capacitance is a function of the thickness of the depletion layer, which varies with applied bias

(see figure 11.5). Therefore, it is common to specify the junction capacitance at zero external bias. This topic is covered in the discussion on reverse bias operation of photodiodes.

Detector Angular Response

The photocurrent generated from a photodiode is essentially independent of angle of incidence of the incoming radiation when the angle of incidence is less than 30 degrees. Typically, a variation in photocurrent of 1 to 2 percent can be expected, provided the detector's active area is underfilled, i.e., the incoming radiation does not completely cover the device's entire active area. (This condition assumes the photodiode's absorption layer thickness approximately equals the depletion layer thickness in the photodiode junction.)

In circumstances where the photodiode is immersed in a collimated beam of incident light, the device's responsivity will fall off with the cosine of the angle of incidence as follows:

$$\Re_\theta = \Re \cos\theta \qquad (11.4)$$

where $\Re$ is the photodiode responsivity at normal incidence.

Photodiode Speed of Response

Inherent limitations of photodiode response are due to structure and specific junction design, the presence of an externally applied bias, and the wavelength of incoming radiation. The inherent time constant of a photodiode causes a delay in generated photo current (τ). For CVI Melles Griot silicon photodiodes, this time constant is typically 7 to 15 nsec. The photodiode speed of response for various load conditions is discussed in the section on sensing circuits for photodiodes.

Dark Current

The P-N junction of a photodiode does not present an infinite resistance to reverse current flow. Consequently, when a reverse bias is applied to a photodiode, a small current (I_{dk}) flows even in the absence of incident light, as shown in the I-V curve. This dark current increases slowly with increasing reverse bias. Dark current approximately doubles every 8°C.

A large dark current is undesirable since it may represent a significant background above which the actual photocurrent is measured. Furthermore, shot noise associated with the dark current may be significant.

Breakdown Region

There is a maximum reverse bias voltage that can be applied to a photodiode, called the reverse breakdown voltage. At this point, the device is no longer useful as a photodetector and can sustain permanent damage.

Care should be taken to limit the current through the diode in the reverse bias region so that the resultant power dissipated does not exceed 200 mW.

Noise-Equivalent Power and Signal-to-Noise Ratio

Noise-equivalent power (NEP) is the incident light level impinging on a photodiode, which produces a photocurrent equal to the noise level. It is usually regarded as the most significant figure of merit for a photodetector. The NEP is a function of the photodiode's responsivity, the noise of the photodiode and the associated sensing circuit, and the frequency bandwidth over which the noise is measured. In systems applications, the signal-to-noise Ratio (SNR) may be computed by taking the ratio of the incident optical power to the photodiode NEP.

In general terms, NEP is defined as follows:

$$\lambda_c = \frac{hc}{E_g} \qquad (11.5)$$

where $\Re$ is in A/W.

The rms noise current (i_{rms}) is the total integrated noise over the frequencies of interest (f_1 to f_2), defined as follows:

$$i_{rms} = \sqrt{\left[i_1^2(f) + i_2^2(f_N) + \ldots i^2(f) \right] df} \qquad (11.6)$$
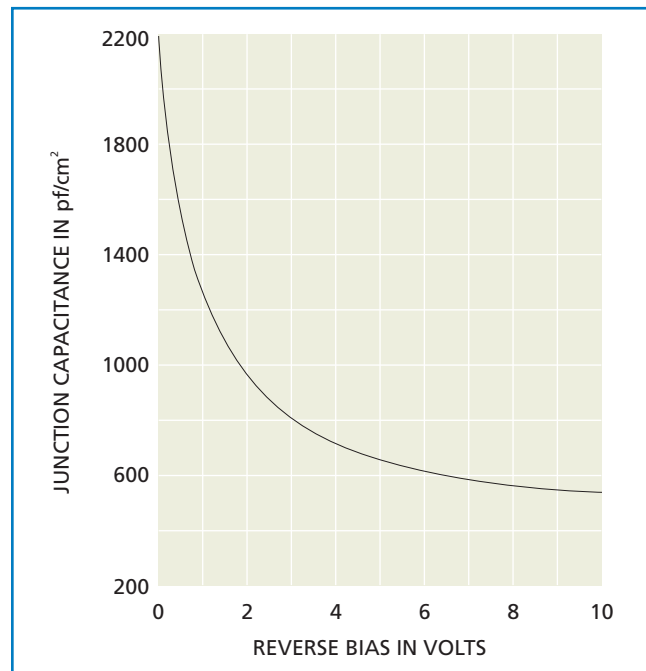

Figure 11.5 **Typical relationship of a silicon photodiode's junction capacitance and applied bias**

If the bandwidth over which the noise is integrated is 1 Hz at frequency f, this is referred to as the noise spectral density at frequency f, with units of W/Hz$^{1/2}$. Since the rms noise is highly dependent upon system application, it is common for noise sources to be specified in terms of noise spectral density at a specific frequency. Since the noise will vary with wavelength, frequency, and temperature, all pertinent parameters must be specified when discussing NEP. The nomenclature NEP (λ, f, T) delineates the specification at wavelength λ, frequency f, and temperature T.

Noise Sources

The lower detection limit for any photodetector is ultimately determined by the device's noise characteristics. There are three main contributions to photodiode noise: Johnson noise, shot noise, and 1/f noise (see figure 11.6.)

Johnson Noise

The statistical fluctuation in the thermal electron-hole pair generation is called thermal noise or Johnson noise (I_{Rsh}). Johnson noise is broadband "white" noise and is expressed as noise per unit bandwidth. The expression for Johnson noise of the photodiode shunt resistance is as follows:

$$I_{R_{sh}} = \sqrt{\frac{4kT}{R_{sh}}} \ \text{A}/\sqrt{\text{Hz}} \tag{11.7}$$

where
k is Boltzmann's constant
T is the absolute temperature in degrees Kelvin
R_{sh} is the shunt resistance in ohms.

As seen from the above formula, Johnson noise increases in proportion to the square root of the temperature.

Shot Noise

Shot noise (I_{sh}) is the statistical noise associated with the photocurrent and dark current (if present). Since shot noise is also broadband, it is expressed as noise per unit bandwidth:

$$I = \sqrt{2q\left(I_{photo} + I_{dk}\right)} \ \text{A}/\sqrt{\text{Hz}} \tag{11.8}$$

where
q is the charge carried by an electron
I_{photo} is the photocurrent
I_{dk} is the dark current.

1/f Noise

The mechanism for 1/f noise is not particularly well understood. Often the characteristics are empirically determined for individual families of photodetectors. The 1/f noise is governed by the following characteristics:

$$I_{1/f}\left(s\right) = I_0 \left(\frac{1}{s}\right)^{\alpha} \ \text{A}/\sqrt{\text{Hz}} \tag{11.9}$$

where
I_0 is typically an inverse function of the active area
α is an empirically derived constant that will vary from 0.25 to 1.0 depending upon the specific construction of the detector
s is the Laplace operator and $= j\omega$.

1/f noise is of concern at frequencies below 100 Hz. For higher frequency applications, photodiode performance is limited by other noise sources. 1/f noise is zero at zero bias.
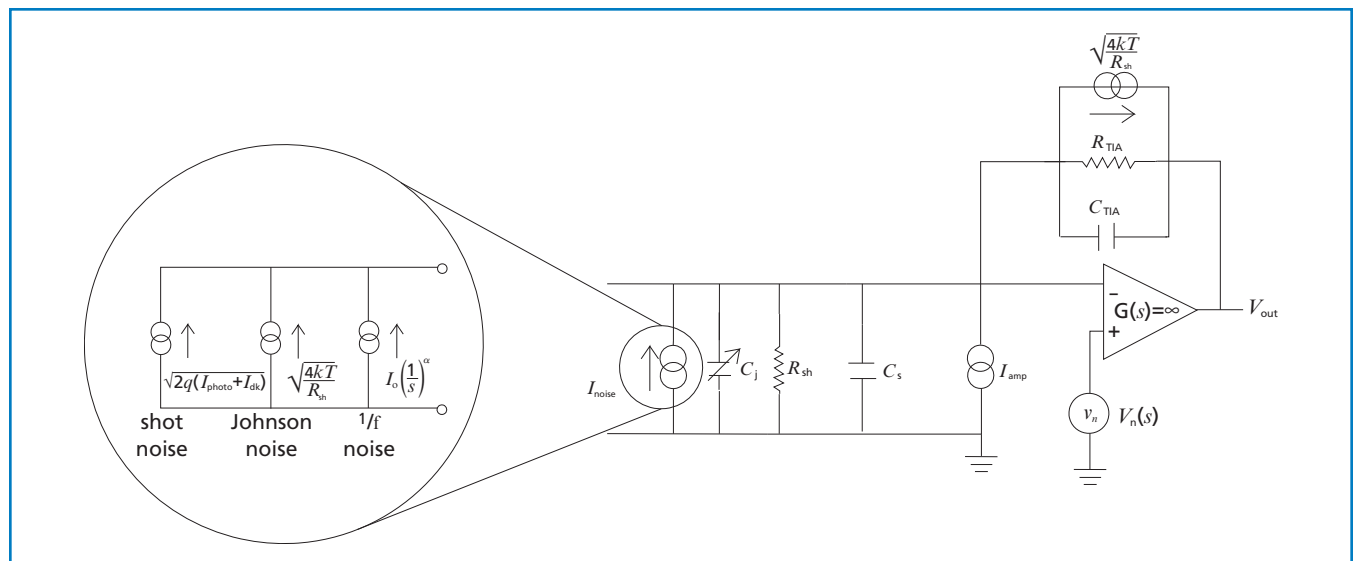

Figure 11.6   **System noise model of a detector and transimpedance-amplifier setup**

Temperature Effects

Like all semiconductor devices, photodiodes are temperature sensitive. The relative population inversion of electrons and holes in the P- and N-layers of the photodiode is directly influenced by temperature, which changes the conductivity and shifts the absorption spectrum. The major photodiode parameters that are sensitive to temperature are shunt resistance, dark current, and, to a lesser extent, responsivity. Typically, the dark current of a silicon photodiode approximately doubles for each 8°C increase or decrease in the device temperature. The shunt resistance approximately doubles for each 6°C change:

These formulas can be used to calculate the shunt resistance and dark current for any temperature from the specified values, which are usually specified at 25°C.

$$I_{dk}(T_2) = I_{dk}(T_1) 2^{\frac{(T_1 - T_2)}{8}} \qquad (11.10)$$

$$R_{sh}(T_2) = R_{sh}(T_1) 2^{\frac{(T_1 - T_2)}{6}} \qquad (11.11)$$

Increasing the temperature of a semiconductor shifts its absorption spectrum to longer wavelengths by reducing the effective band gap. Fortunately, the absorption spectrum of silicon is quite broad. Consequently, the small temperature-induced shifts in the absorption spectrum affect the responsivity significantly only at the edges of the spectral responsivity curve, as shown in figure 11.7.

INTEGRATING SPHERES

In many applications of photodetectors, it is necessary to measure the absolute or relative intensity of a wide-angle beam (divergent source)
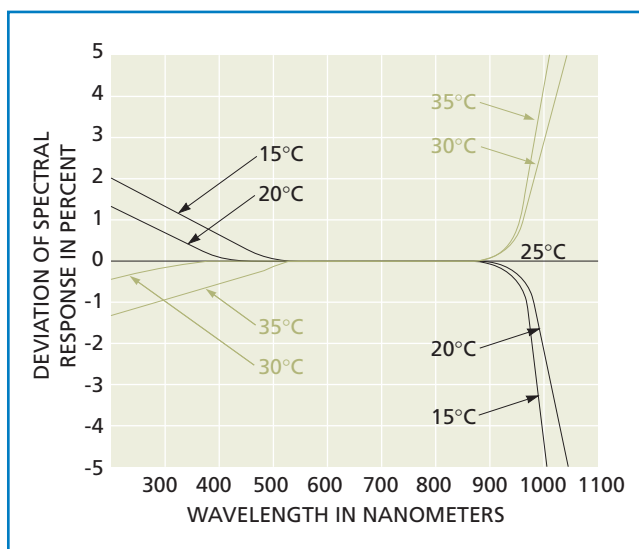

Figure 11.7   **Temperature dependence of silicon photodiode responsivity**

or of an inhomogeneous beam much larger than the active area of a photodetector. Integrating spheres have been used for many years to scramble or average light by multiple diffuse reflections in order to obtain meaningful intensity measurements of these types of sources.

An integrating sphere is a hollow sphere (often aluminum) whose entire inner surface is uniformly coated with a layer of matter which has a highly diffuse reflectance. When light from a source enters an integrating sphere, it loses all memory of direction and polarization. At the exit port, the light intensity is uniform and diffuse. Although other methods have been developed to deal with the problems of averaging the intensity from an inhomogeneous source or of a wide-angle beam, an integrating sphere is the best solution in many applications. In fact, the only real differences between the integrating spheres of today and the spheres of many years ago are the improved quality and stability of the diffuse reflectance coating.

Lambertian Source — Radiance and Irradiance

The two keys to the operation of an integrating sphere are the coating on the inside of the sphere and the spherical shape itself. The coated interior surface is designed to have a highly diffuse reflectance. When a hypothetically perfect diffuse reflector is illuminate with uniform intensity, it behaves like a perfectly diffuse source—a *Lambertian* source (i.e., each unit area reflects light into all available solid angles with equal efficiency).

The complete definition of a Lambertian source is a light source whose radiance is independent of viewing angle. The most well-known Lambertian source is the sun. Radiance is defined as the energy flux per unit projected area per unit solid angle leaving a source or, in general, any reference surface.

Radiance, L, can be expressed as

$$L = \frac{dI}{d\Omega dA \cos\theta} \qquad (11.12)$$

where
   dI is the intensity
   $d\Omega$ is the solid angle
   dA is the unit area of the source
   $\cos\theta$ is the viewing or inclination angle.
   $dA\cos\theta$ is the projected surface area.

To understand the operation of the integrating sphere, consider the light reflected from a uniformly illuminated small area of the interior surface. This unit area is so small that it can almost be considered a point source. This virtual point source radiates equal intensity in all possible directions. The rest of the sphere can be thought of as a target over which the light from this source will be distributed. Two counterbalancing factors ensure that this light will be distributed with equal intensity over the entire target surface. The important intensity parameter for a target surface is the incident energy (flux) per unit area. This is termed the *irradiance* of the target surface.

The inverse square law shows that the light flux per unit area will be a minimum at the most distant point on the sphere. However, the irradiance of a unit area of the target surface depends on its angle of inclination to the source. A large angle of incidence causes the incident flux to be spread over a large target area (cosine projection). A small angle of incidence presents the smallest area to any incident light. The same cosine projection argument applies to the projected area of the source.

The two effects exactly cancel each other so that the light intensity is uniform over the entire area of the sphere. This can be proven by simple geometry, as shown in figure 11.8.

Consider a small unit area of irradiating surface (dA_s), and a small unit area of target surface (dA_t). The intensity of light (irradiance) α on the target due to the source dA_s is given by:

$$\alpha = L\left(\frac{\cos^2\theta}{d^2}\right)dA_s \qquad (11.13)$$

where L is the source radiance and θ is the angle of incidence on dA_t. The d^2 term is from the inverse square law, and the $\cos^2\theta$ term relates to the effective (cosine projected) areas of the source and target.

In a real application, when an extended area of the surface is illuminated with nonuniform intensity, the directly illuminated spot on the sphere can be considered as a large number of minute Lambertian sources. In principle, therefore, the total flux of the beam entering the sphere is spread (averaged) over the entire surface following only a single reflection.

In order to deliver reliable integration and low throughput loss, the coating of the integrating sphere must be a very efficient diffuse reflector. Coatings used on integrating spheres provide reflection efficiencies between 95 and 99 percent. The reflectance must be high in order to minimize absorption loss cau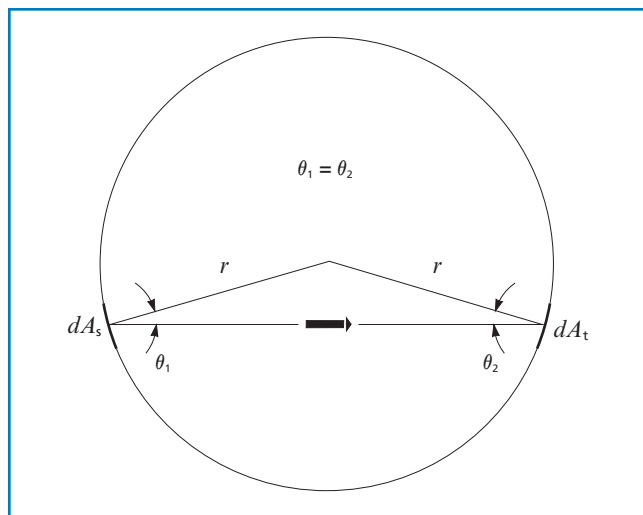sed by multiple reflections, but it must not reflect light in a specular sense (incidence angle=reflectance angle). Any specular component to the reflections off the surface will only help preserve some memory of original angular and spatial distribution of the light source. In many applications, it is also necessary that the properties of the coating be fairly insensitive to wavelength.

Light is not completely averaged or integrated on the first reflection. To prevent erroneous readings that are dependent on input angle, the detector or output port is often shielded with a baffle plate. This small metal plate, also coated with the same diffuse reflectance material, is positioned to preclude any light from the entrance port from reaching the detector after only a single reflection.

Performance Characteristics

Two meaningful figures of merit for integrating spheres are *throughput* and *stability*. Throughput is defined as the ratio of the flux exiting the sphere to the flux entering the sphere. It is determined by the size of the ports (input and exit) relative to the sphere area and by the reflectivity.

$$\text{Throughput} = \frac{A_e R}{\left[1 - R\left(1 - A_p\right)\right]} \qquad (11.14)$$

where

A_e is the area of the exit port divided by total sphere area

R is the reflectance of the coating

A_p is the total area of all ports divided by total sphere area.

Figure 11.9 shows the reflectance of a typical integrating sphere coating.
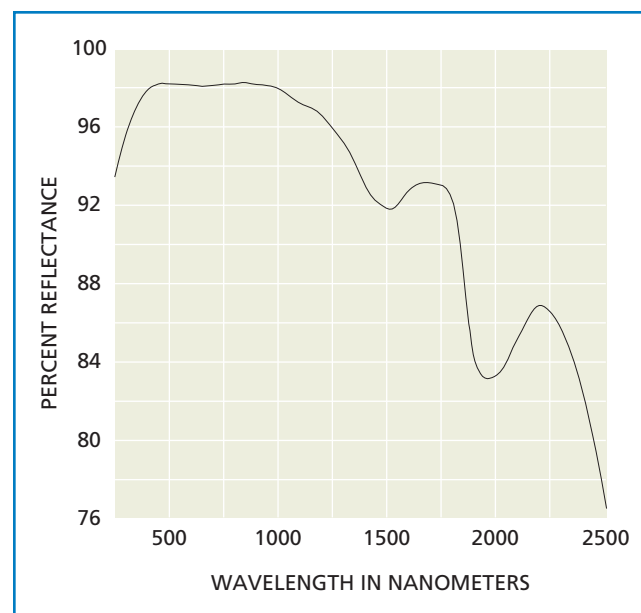

Figure 11.8　**Integrating sphere geometry factors**

Figure 11.9　**Typical reflectance of the integrating sphere coating**

Throughput is very dependent on the reflectance of the sphere surface in a nonlinear manner (see figure 11.10).

Stability is the reciprocal of the change in throughput with respect to the change in reflectivity of the coating. This is an important parameter since the reflectivity of the coating may degrade slightly in certain operating environments over a long period of time, and the reflectivity may be degraded nonuniformly over the sphere surface. A small change in the reflectivity of a low-stability sphere will result in a large change in throughput. Therefore, a low-stability sphere is less likely to give an accurate, reproducible reading than a high-stability sphere.

Stability is related to throughput. A high-throughput sphere is inherently less stable than a lower-throughput sphere because of the nonlinear dependence of throughput on reflectivity. Stable spheres have reflectivity values of 90 percent or less, whereas the low-stability (high-throughput) spheres typically have reflectivities of 97 percent or greater. As an example, a typical low-throughput sphere with a 90 percent reflectivity coating will have an inherent stability factor of 0.5, whereas an identical high-throughput sphere with a 99 percent reflectivity coating will have a stability factor of 0.05.
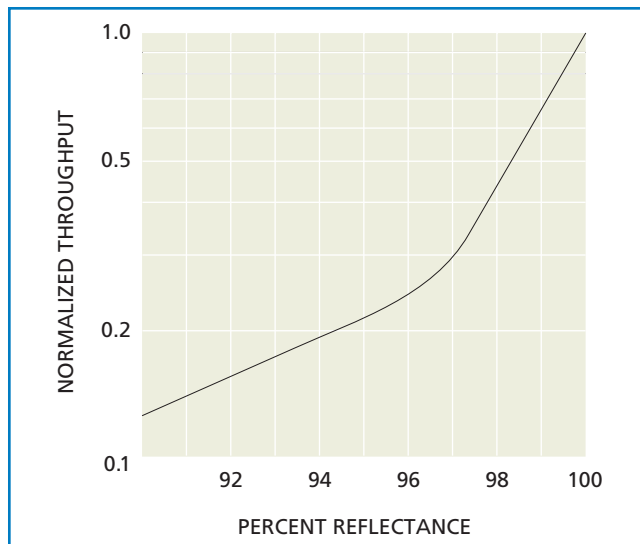

Figure 11.10   **Normalized throughput variation with reflectance for an integrating sphere**

Diverse Applications

Integrating spheres are versatile devices which are used in a wide range of applications. This versatility is best illustrated by the examples outlined below.

1. Diode Laser and Fiber-Optic Power Measurements. Integrating spheres are ideal for collecting and measuring the highly divergent radiation typical of diode lasers and fiber optics.

2. Intensity Measurements of Extended Sources. Tungsten filaments, plasma discharges, and other extended sources can be difficult to collimate or efficiently image onto a photodiode of limited area. The large entrance port of the CVI Melles Griot 13 ISP 005 integrating sphere facilitates coupling in light from extended sources and reduces the need for a complicated optical system.

3. Absolute Radiometry in Anamorphic Laser Systems. An integrating sphere and a large-area detector can accurately measure radiation that is too large to focus onto a detector, or light that is collimated in one plane but defocused in the other. A typical application is measuring the output from laser line projectors.

4. Measurement of Inhomogeneous Beams. The integrating sphere effectively scrambles the input radiation so that any beam inhomogeneity and detector nonuniformities do not affect the accuracy of the measurement.

5. Creation of a Uniform Intensity Profile. An integrating sphere can be used as a transmissive component to homogenize the spatial intensity profile of a beam from an inhomogeneous source. Uniform illumination is required in applications such as calibrating CCD, CMOS, and photodiode arrays.