CHAPTER 4

OPERATING CHARACTERISTICS AND CONSIDERATIONS

Photomultiplier characteristics that need to be considered in most applications include sensitivity, time and frequency response, stability, linearity, and possible environmental effects at a certain gain.

Before choosing a photomultiplier for a given application, however, it is well to establish that a photomultiplier is in fact the best type of detector for that application. Alternatives, such as a vacuum or semiconductor photodiode plus a high-gain, low-noise amplifier, may offer advantages in size, power supply, or cost. Other criteria on which the choice may depend include spectral sensitivity, frequency response, and output current range. Leaving all these out of consideration, however, the area where a photomultiplier clearly excels is in its ability to detect very low-level light; and in particular, below a certain threshold, to do so with a better signal-to-noise ratio than any alternative detector. In many applications this is decisive. For the light level to be detected and the required signal-to-noise ratio, a photomultiplier may be the only choice.

4.1 Signal-to-noise ratio comparison: photomultiplier vs photodiode

Whether the detector is a photomultiplier, or a photoemissive cell, or a photodiode plus amplifier, the critical factor governing signal-to-noise ratio is the quantity of light received, which often depends on the size of the sensitive surface. To make the following comparison independent of size, therefore, it will be based on the assumption of equal photocurrents: I_k for the photomultiplier, and I_d for the photodiode.

Equation 3.42 gave the signal-to-noise ratio of a photomultiplier with cathode current I_k, gain G, load resistance R_L, and anode dark-current noise (referred to the cathode) $i_{k,n}$. The equivalent expression for a photodiode, plus low-noise, high-gain amplifier connected as a current-voltage converter with feedback resistance R_f is

$$\frac{S}{N} = \frac{I_d}{\sqrt{B_N \left(\frac{4kT}{R_f} + 2eI_d \right) + \overline{i_{d,n}^2} + \overline{i_{o,n}^2}}} \tag{4.1}$$

where $i_{d,n}$ is the intrinsic noise current of the detector, and $i_{o,n}$ the input equivalent noise current specified by the amplifier manufacturer. (The amplifier input equivalent noise voltage, $e_{o,n}$, also specified by the manufacturer, is usually negligible and is here disregarded.) Equating Eq.3.42 with Eq.4.1 gives the photocurrent at which both

detectors have like signal-to-noise ratios,

$$R_f I_k = R_f I_d = \frac{2kT}{e(a - 1)} \left(1 - \frac{R_f}{G^2 R_L} \right) + R_f \frac{\overline{i_{d,n}^2} + \overline{i_{o,n}^2} - \overline{i_{k,n}^2}}{2e(a - 1)B_N} \tag{4.2}$$

If $G^2 R_L >> R_f$, this simplifies to

$$R_f I_k = R_f I_d = \frac{2kT}{e(a - 1)} + R_f \frac{\overline{i_{d,n}^2} + \overline{i_{o,n}^2} - \overline{i_{k,n}^2}}{2e(a - 1)B_N} \tag{4.3}$$

The two terms of Eq.4.3 correspond to two limiting cases.

If R_f is low (say, less than 10^5 Ω), the second term is negligible compared with the first. For a = 1.5 (a being 1 + V_G, where V_G is the photomultiplier gain variance), the signal *voltage* below which the photomultiplier has the better signal-to-noise ratio is

$$R_f I_k = R_f I_d \leq \frac{2kT}{e(a - 1)} \approx 100 \text{ mV}$$

The photocurrent or incident flux to which this corresponds can be lowered only by increasing the feedback resistance R_f. But that can be done only at the expense of bandwidth.

If R_f is high (say, more than 10^7 Ω), the first term becomes negligible compared with the second. Then, since $i_{d,n}$ and $i_{k,n}$ are also negligible compared with $i_{o,n}$, the voltage threshold below which the photomultiplier has the better signal-to-noise ratio tends toward

$$R_f I_k = R_f I_d \leq R_f \frac{\overline{i_{o,n}^2}}{2e(a - 1)B_N}$$

which, for $\sqrt{\overline{i_{o,n}^2}/B_N}$, corresponds to a *photocurrent*

$$I_k = I_d \approx 10^{-7} \text{ A}$$

Thus, at very low light levels requiring a high feedback resistance R_f, the signal-to-noise ratio of a photodiode is limited by the amplifier input equivalent noise current $i_{o,n}$.

In Eqs 4.2 and 4.3 the following practical values apply.

For a photomultiplier

$$\sqrt{\overline{i_{k,n}^2}/B_N} < 10^{-14}\ \text{A/Hz}^{1/2}\ \text{with S1 cathode}$$
$$< 10^{-16}\ \text{A/Hz}^{1/2}\ \text{with other cathodes}$$

for a vacuum photodiode

$$\sqrt{\overline{i_{d,n}^2}/B_N} < 10^{-14}\ \text{A/Hz}^{1/2}\ \text{with S1 cathode}$$
$$< 10^{-16}\ \text{A/Hz}^{1/2}\ \text{with other cathodes}$$

for a low-noise silicon photodiode

$$\sqrt{\overline{i_{d,n}^2}/B_N} \leq 5 \times 10^{-15}\ \text{A/Hz}^{1/2}$$

and for a low-noise operational amplifier with FET input

$$\sqrt{\overline{i_{o,n}^2}/B_N} > 10^{-13}\ \text{A/Hz}^{1/2}$$

Figure 4.1 shows the signal-to-noise ratio, based on these values and at 1 Hz bandwidth, as a function of photocurrent for a photomultiplier and for a photodiode plus low-noise amplifier. For the photomultiplier, G^2R_L is taken as parameter, and for the photodiode R_f. The intersections of the curves with the horizontal axis (S/N = 1) indicate the photocurrents at which the signal is no longer distinguishable from the noise. The intersections of the dashed curves with the solid ones indicate the photocurrents at which the signal-to-noise ratio of a photodiode becomes competitive with that of a photomultiplier.

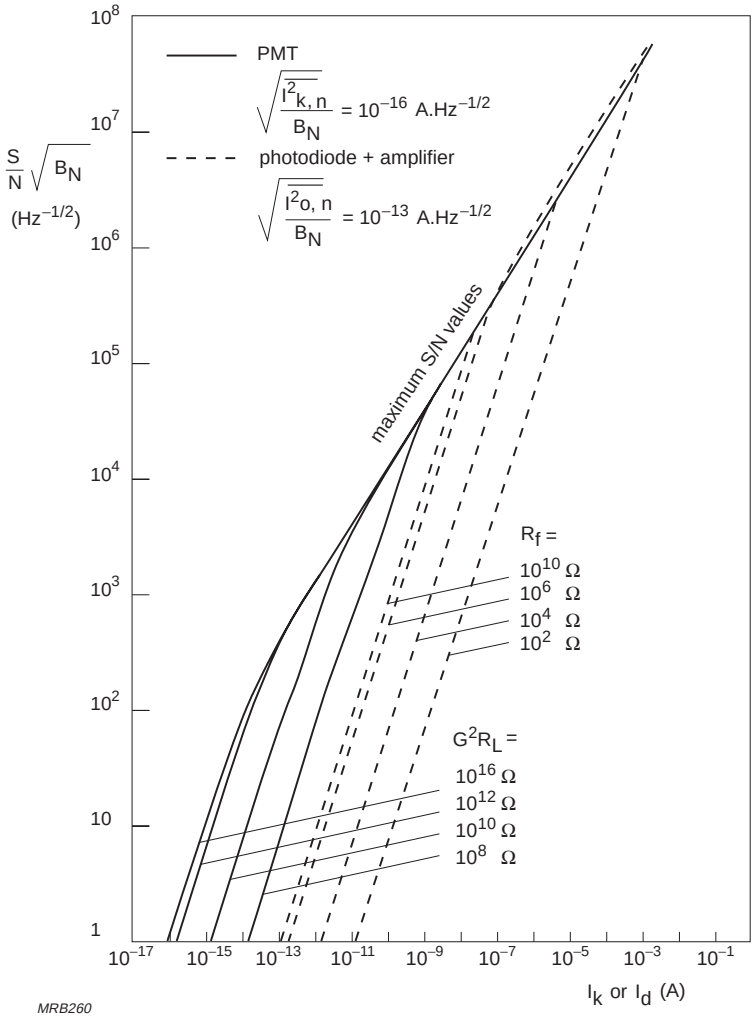


Fig.4.1 Signal-to-noise ratio as a function of photocurrent for a photomultiplier (solid lines) and a photodiode plus low-noise amplifier (dashed line)

4.2 Photomultiplier selection criteria

Points to consider in choosing a tube for a given application are photocathode characteristics, response speed, and number of stages.

Photocathode spectral sensitivity and size. If the light to be detected is monochromatic, choose a tube whose maximum sensitivity is as close as possible to the same wavelength. If it is not monochromatic, look for the best match between sensitivity and spectral distribution, using the matching factor described in A6.3. Bear in mind, though, that the greater the sensitivity in the red, the greater the thermionic emission. And that relative variations of sensitivity versus temperature and from tube to tube, are largest close to the photoemission threshold.

The choice of photocathode diameter depends partly, but not entirely, on the size of the incident light beam or source. Beam or source size determines the minimum practical diameter but not the maximum. If the equipment or installation does not impose strict constraints on size or weight, tubes with medium-diameter cathodes are often preferable to those with small ones. They are generally more stable and have higher permissible anode currents. Dark current does not vary in strict proportion with cathode diameter; and, if the dark current of a tube with a large-diameter cathode is inconveniently high, it can be reduced by reducing the *effective* cathode diameter in the way described in §5.8.2.

Response speed. If the rise time required is about one nanosecond or less, or the required bandwidth more than about 100 MHz, choose a fast-response tube. Such tubes also have the best time resolution.

If the rise time required is a few nanoseconds, or the required bandwidth between 50 and 100 MHz, a standard tube with linear-focusing dynodes is a good choice. Tubes with venetian-blind dynodes are comparatively slow and not suitable for bandwidths of more than about 10 MHz.

Number of stages. General-purpose tubes usually have eight or ten stages and a gain of 10^3 to 10^7 at an applied voltage of 600 to 1800 V. Lowering the voltage impairs fast response and linearity, so if lower gain is required, choose a tube with fewer stages. For gain higher than 10^7, choose one with more stages. The additional stages enable interdynode voltages to be kept moderate and so prevent dark current from becoming excessive. A tube with twelve stages will safely give a gain of 10^8; setting the gain higher than that limits the output current pulse range (§5.5.3) and is seldom justified. A fast-response tube operating at a gain of 10^9 with type A voltage distribution (§5.2.1) approaches its linearity limit even with single-electron pulses.

4.3 Factors affecting sensitivity

The anode sensitivity of a photomultiplier varies according to the part of the cathode surface from which photoemission originates. The variation can be mapped by scanning the cathode with a narrow beam of light and plotting the resulting anode current variation. Figure 4.2 is a plot of the anode sensitivity variation measured relative to one diameter of a 32 mm SbKCs cathode; the dashed line shows the corresponding variation of cathode sensitivity measured along the same diameter with the tube connected as a diode. Comparison of the two curves illustrates the relative variation of the two components of anode sensitivity given in Eq.2.11: namely, the

photocathode sensitivity S_k and the collection efficiency η of the cathode/first-dynode space.

Figure 4.2 was plotted with the scanning beam normal to the cathode; a different angle of incidence would give a different sensitivity contour.

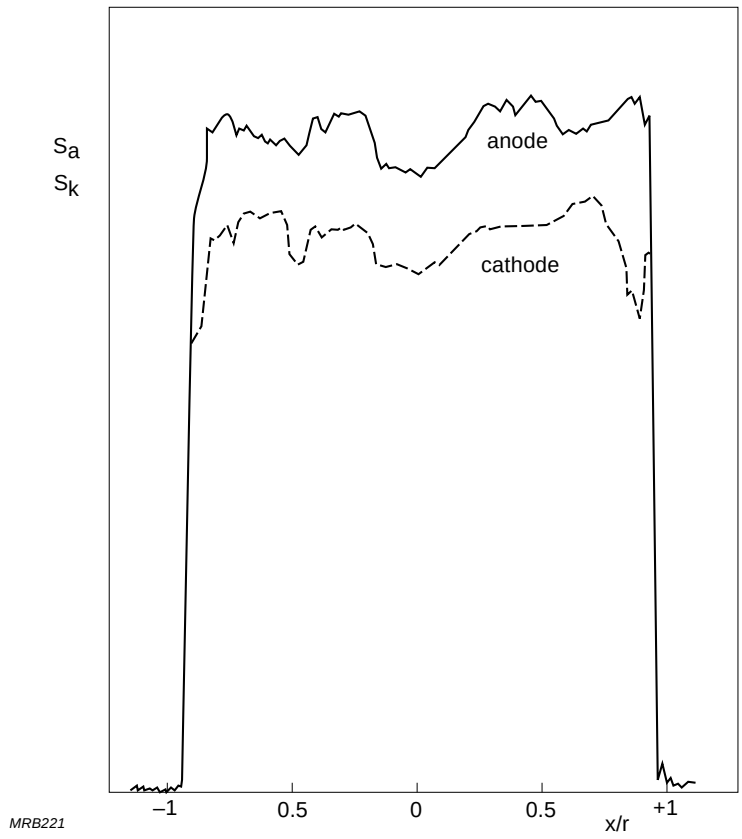


Fig.4.2 Example anode and cathode sensitivity variation measured at λ = 424 mm along one cathode diameter of a photomultiplier with a 32 mm SbKCs cathode. Vertical axis, relative sensitivity (arbitrary scale); horizontal axis, distance from cathode centre, normalized with respect to cathode radius. Curves offset for clarity

4.3.1 Wavelength

Cathode sensitivity variations depend very much on wavelength. They are usually low in the spectral range corresponding to maximum sensitivity and increase towards the ends of that range. SbCs (S11) and bialkali cathodes are commonly used near their maximum sensitivity wavelengths, where uniformity of sensitivity is generally good. Trialkali (S20, S20R) cathodes, however, are often used near their threshold wavelengths, where uniformity is less good. Figure 4.3 shows the sensitivity variation of an S20 cathode measured at wavelengths of 424 nm, 629 nm, and 800 nm along a diameter aligned with the dynodes; note how the good uniformity at 424 nm deteriorates at the longer wavelengths. For this reason, the measurement of uniformity

of response for tubes with bialkali cathodes with, for example, green (560 nm) LEDs can give misleading results.

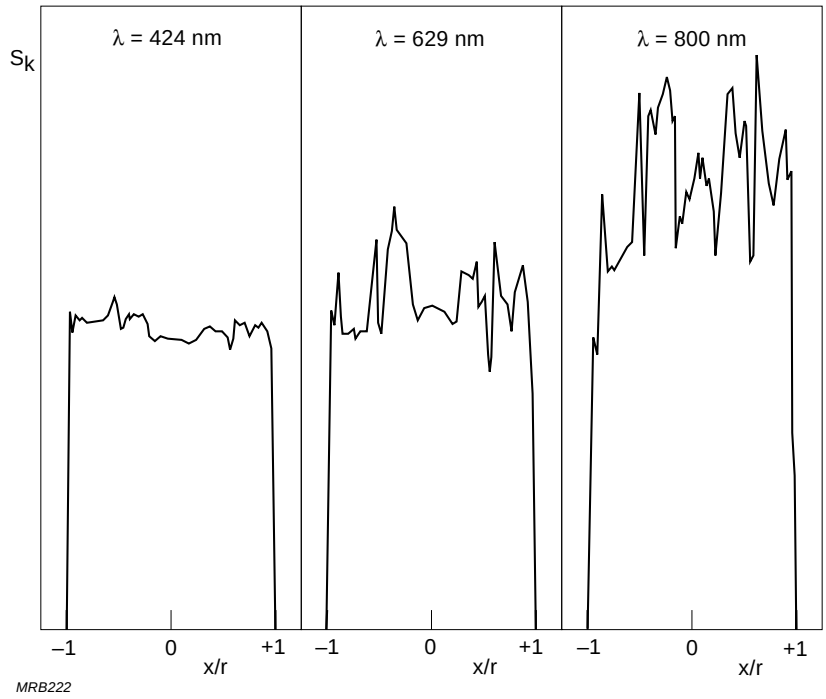


Fig.4.3 Examples of cathode sensitivity variation measured along one diameter of a $SbNa_2KCs$ cathode at three wavelengths

4.3.2 Collection efficiency

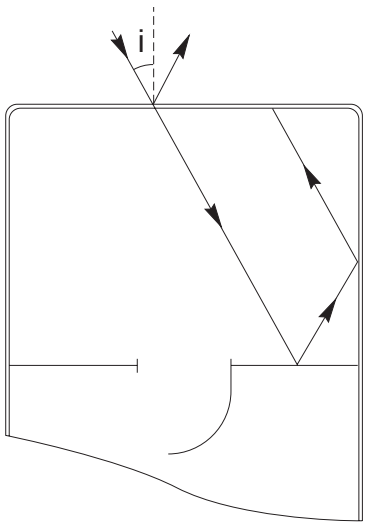
The electron-optical input system is meant to direct all photoelectrons, regardless of their points of origin or initial velocities, onto the useful area of the first dynode. Design features that can contribute to this include: a spherically curved photocathode, a large-area first dynode (e.g. venetian blind), and an electrode in the input system whose potential can be adjusted to compensate for assembly tolerances.

In tubes with focusing dynodes the uniformity of collection is generally best along a diameter perpendicular to the plane of symmetry. The ratio of the first-to-second dynode voltages is also important, for it influences the effective area of the second dynode. To optimize performance, the second dynode potential, and that of any adjustable focusing electrode in the input system, must be carefully adjusted with the tube biased for maximum gain and preferably with the cathode fully illuminated.

Magnetic fields, including the earth's, to which a photomultiplier is exposed have an adverse effect on collection efficiency.

4.3.3 Angle of incidence

Depending on its angle of incidence, light that passes through a semitransparent photocathode may either be reflected back to it from the interior of the tube (Fig. 4.4) or impinge on some internal photoemissive surface. In either case, the resulting photoelectrons may be collected by the multiplier and contribute to the anode signal. This is an important cause of the variation of apparent anode sensitivity with angle of incidence. Figure 4.5(a) shows examples of anode sensitivity contours measured on the same tube at three angles of incidence. As the transmission coefficient of the photoemissive layer varies with wavelength, so does the amount of light that penetrates to the interior of the tube. Thus, the variation of apparent sensitivity with angle of incidence is also a function of wavelength.



MRB223

Fig.4.4 Light that passes through a semitransparent cathode may be reflected back to it by internal surfaces

Increasing the thickness of the photoemissive layer reduces the amount of light it transmits and, hence, the amount of internal reflection. This effect is clearly evident in a comparison of the angle-of-incidence related sensitivity variations of the thin S20 and the thick S20R photocathode.

In some cases nearly all internal reflection effects can be eliminated by frosting (sandblasting or etching) the outside of the input window, without any loss of absolute sensitivity (Fig.4.5(b)). This is not effective, however, when the tube is optically coupled to a scintillator or light guide by means of a matching compound such as silicone grease.

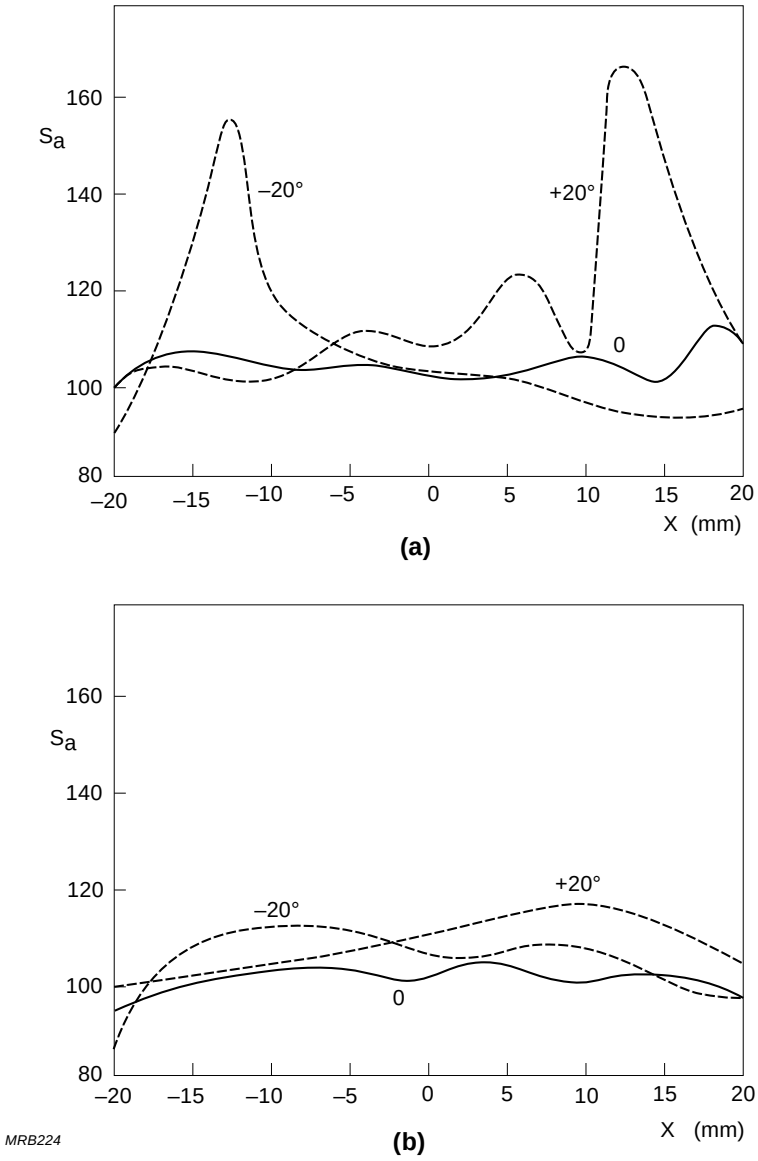


Fig.4.5 Examples of relative anode sensitivity variation at three angles of incidence, measured at λ = 629 mm along one diameter of a tube with $SbNa_2KCs$ cathode; with (a) polished and (b) frosted input window. x is the distance from the cathode centre

4.4 Time characteristics

4.4.1 Pulse response: determining factors

When a light pulse of infinitesimal duration (delta-function pulse) excites the photomultiplier, the duration of the resulting anode pulse is not infinitesimal. The lengthening is due to electron transit time variations in the cathode/first-dynode space and in successive stages of the electron multiplier. The probability distribution of the variations in each stage is called the *specific response* of the stage. The individual

stages are not, however, statistically independent; there is some correlation between the transit time spreads in successive stages, and this makes it difficult to calculate the pulse response of the tube from the response of the respective stages. However, if the probability distribution of electrons arriving at the anode is assumed to be approximately gaussian with a standard deviation σ_R, then the response $R_\delta(t)$ to a delta-function light pulse is given by

$$R_\delta(t) = \frac{1}{\sigma_R\sqrt{2\pi}} \exp\left\{-\frac{(t - t_t)^2}{2\sigma_R^2}\right\} \tag{4.4}$$

where t_t is the mean transit time.

To a good approximation, the variance σ_R^2 of the pulse response can be represented by the sum of the variances of the response of each stage:

$$\sigma_R^2 = \sigma_{k,d1}^2 + \sigma_{d1,d2}^2 + \ldots + \sigma_{dN-1,dN}^2 + \sigma_{dN,a}^2 \tag{4.5}$$

Generally, it can be assumed that all stages beyond the second have equal response; hence,

$$\sigma_R^2 = \sigma_{k,d1}^2 + \sigma_{d1,d2}^2 + (N - 1)\,\sigma_{d,d}^2 \tag{4.6}$$

and the response pulse width (FWHM) is

$$t_w = 2.36\,\sigma_R \tag{4.7}$$

This is least when the response is due to a single photoelectron, for then there are no transit time differences in the cathode/first-dynode space and the term $\sigma_{k,d1}^2$ in Eq.4.6 vanishes.

Electrodynamic effects in the anode collection space can also affect the pulse response, altering the anode pulse shape and increasing t_w. Among these are the electromagnetic effect of electron movement close to the anode collector grid and the generation of high-frequency currents due to oscillation of electrons about this grid. The often imperfect matching of the anode to the output transmission line can also give rise to oscillations at the pulse trailing edge (‘ringing’).

Effect of applied voltage. The transit-time fluctuations that affect pulse response have two main causes:

– the initial velocity spread of electrons emitted by different electrodes; the contribution of this cause varies as $1/V_{d,d}$ (where $V_{d,d}$ is the interdynode voltage).

– the difference in transit time due to different points of emission from the same dynode; this contribution varies as $1/\sqrt{V_{d,d}}$.

Thus, the lengthening of the response pulse at each stage depends on a factor between $V_{d,d}^{-1}$ and $V_{d,d}^{-1/2}$. Provided the permissible limits are observed, increasing the voltage per stage is an effective way to improve the pulse response.

Effect of wavelength. The photoelectron energy distribution depends on the incident light wavelength. This, therefore, also affects transit time fluctuations, but only in the electron-optical input system and the first multiplier stage. Moreover, as all stages of the tube contribute about equally to the lengthening of the response pulse, the overall effect of wavelength is only moderate. The existence of other causes of response pulse lengthening, apart from transit time fluctuations, further diminishes the overall significance of wavelength.

4.4.2 Pulse response: measurement

Single-electron response. There are two ways to measure the response to emission of a single photoelectron:

– with continuous light so attenuated that the average interval between successive photoelectrons is much greater than the least interval the measuring set-up is able to resolve;
– with light pulses so attenuated that the probability of each pulse giving rise to only one photoelectron is much greater than the probability of its giving rise to more than one.

If fluctuations in the number of photons per light pulse follows a Poisson distribution, so will the number, $n_{k,i}$, of photoelectrons emitted in response to them:

$$P(n_{k,i}) = \frac{(\bar{n}_{k,i})^{n_{k,i}}}{n_{k,i}!} \exp(-\bar{n}_{k,i})$$

Thus the probability of no photoelectron being emitted is $P(0) = \exp(-\bar{n}_{k,i})$; the probability of only one being emitted is $P(1) = -\bar{n}_{n,k}\exp(-\bar{n}_{n,k})$; and the probability of more than one being emitted is $P(n_{k,i} > 1) = 1-P(0)-P(1)$. Since the ratio $P(n_{k,i} > 1)/P(1)$ tends toward $\bar{n}_{k,i}/2$ as $\bar{n}_{k,i}$ tends toward zero, it is evident that it is possible to so attenuate the light pulses that the probability of more than one photoelectron being emitted per pulse is negligible compared with that of only one being emitted. In practice single-electron operation is obtained by so attenuating the light that less than one anode pulse occurs per hundred light pulses. The ratio $P(n_{k,i} > 1)/P(1)$ is then less than 5×10^{-3}.

Multi-electron response is measured with subnanosecond pulses – as from fast LEDs, semiconductor lasers, spark sources, or Cherenkov sources – which are short compared with the several nanosecond pulse-widths t_w of present-day photomultipliers.

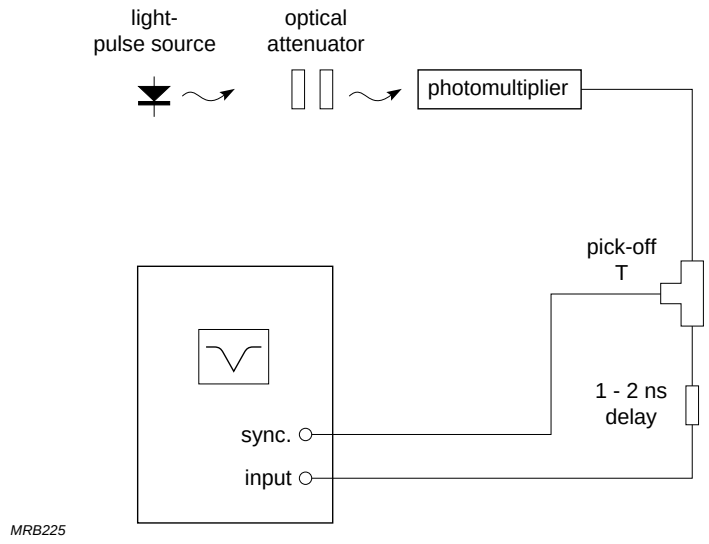


Fig.4.6 Set-up for measuring pulse response

Figure 4.6 shows the usual set-up for measuring pulse response. The oscilloscope may be synchronized by a signal with 1 or 2 ns lead time taken from one of the last dynodes, or by one taken from a pick-off T that precedes a 1 or 2 ns delay as shown. The response R*(t) measured on the oscilloscope is the convolution of the illumination function L(t), the pulse response $R_\delta(t)$ of the photomultiplier, and the pulse response S(t) of the measuring set-up (transmission line, delay line, oscilloscope):

$$R^*(t) = L(t) * R_\delta(t) * S(t) \tag{4.8}$$

When the measurement is made under single-electron conditions, L(t) approximates a delta function and Eq.4.8 simplifies to

$$R^*(t) = R_\delta(t) * S(t)$$

The pulse response of the measuring set-up must be accurately known; estimates may introduce significant error. It is preferable to use a set-up whose response is known to approximate a delta function. Then, in the single-electron case

$$R^*(t) \approx R_\delta(t)$$

and in the multi-electron case

$$R^{*}(t) \approx L(t) * R_{\delta}(t) \tag{4.9}$$

R*(t) then has a standard deviation

$$\sigma_R^{*} = \sqrt{\sigma_L^2 + \sigma_R^2} \tag{4.10}$$

where σ_L and σ_R are the standard deviations of the illumination function L(t) and the photomultiplier response $R_{\delta}(t)$.

4.4.3 Transit time differences

Transit time differences are due mainly to differences in electron path lengths but also, in part, to differences in electric field strength. The largest originate in the cathode/first-dynode space, where path-length differences are greatest; the electron multiplier, where path lengths are more nearly equal, does not make so important a contribution.

Effect of applied voltage. Like transit-time fluctuations, transit-time differences vary inversely as the square root of inter-electrode voltage. As it is the cathode/first-dynode space that contributes most, it is here that the most can be gained by applying the maximum permissible voltage.

Effect of wavelength. Incident light wavelength does not significantly affect transit time differences. In fact, even its effect on overall transit time is small. At wavelengths from 250 nm to 900 nm the initial energy of the photoelectrons is only a fraction of an electron-volt, which is negligible compared with the several hundred electron-volts they may gain from the electric field between the cathode and first dynode. The observed variation in transit time as a function of wavelength amounts to only about 1 ps/nm.

Measurement of transit-time differences. Transit time is determined by measuring the interval between signals known to be synchronous with light pulses at the cathode and the resulting anode pulses. Transit-time differences can be measured, and mapped, by focusing the light pulses on different parts of the cathode and noting the corresponding transit times. Figure 4.7 is an example of the transit-time differences measured with reference to perpendicular cathode diameters of a fast-response photomultiplier. Figure 4.8 shows a set-up for measuring both transit-time differences and transit-time spread.

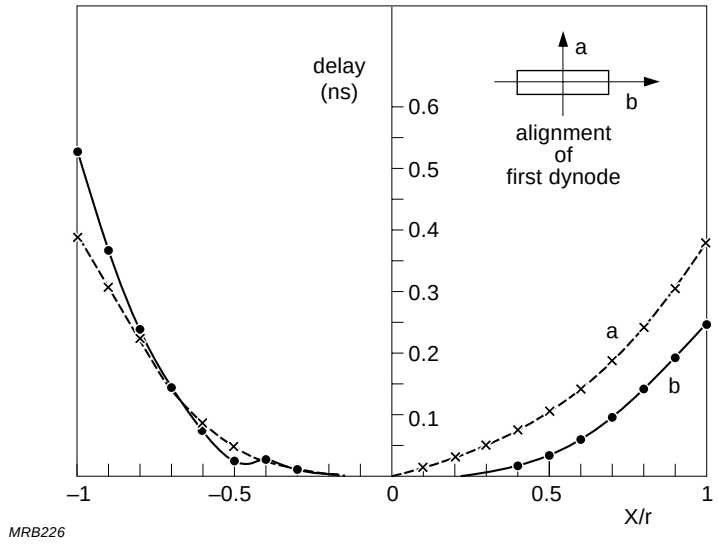


Fig.4.7 Transit-time differences of a fast-response photomultiplier as functions of distance from the cathode centre, measured along perpendicular diameters

4.4.4 Transit-time spread

The pulse-to-pulse fluctuation of transit time (jitter) limits the accuracy of photomultiplier-based time measurements. For single-electron operation it has been suggested that the variance σ_{tt}^2 of the total transit-time fluctuation is

$$\sigma_{tt}^2 = \sigma_{k,d1}^2 + \sigma_m^2 \tag{4.11}$$

where $\sigma_{k,d1}^2$ and σ_m^2 are the variances of the transit-time fluctuations in the cathode/first-dynode space and the electron multiplier. If all stages of the multiplier except the first are identical,

$$\sigma_m^2 = \frac{\sigma_{d1}^2}{g_1}(1 + v_{g1}) + \frac{\sigma_{dd}^2}{g_1(g - 1)}(1 + v_g) \tag{4.12}$$

where g_1 is the gain of the first stage, g the average gain of subsequent stages, v_{g1} and v_g are the relative variances of those gains, and σ_{d1}^2 and σ_{dd}^2 are respectively the variances of the transit-time fluctuations in the first and subsequent stages. Equations 4.11 and 4.12 show that the stages that predominate in determining the transit-time spread are the cathode/first-dynode space and the first multiplier stage.

Transit-time fluctuations in the cathode/first-dynode space have two components: a *chromatic* one due to the spread of photoelectron initial velocities, and a *geometric* one due to path-length differences.

The *chromatic component* has a variance σ_{vi}^2 which can be resolved into components σ_{vn}^2 and σ_{vt}^2 corresponding to the normal and tangential components of initial velocity:

$$\sigma_{vi}^2 = \sigma_{vn}^2 + \sigma_{vt}^2 \tag{4.13}$$

If the probability distribution of the initial velocities is known, σ_{vn} and σ_{vt} can be found from Eq.1.1. For fast-response tubes σ_{vi} is between 50 and 500 picoseconds, depending to some extent on incident-light wavelength and applied voltage.

The *geometric component* has a variance σ_{cb}^2 which can be calculated if the relation between photoelectron points of origin and transit-time differences can be formulated. If, for instance, the transit time varies roughly as the square of the distance of the point of origin from the cathode centre, then $\sigma_{cb} \approx 0.3\ \Delta t_{tmax}$, where Δt_{tmax} is the transit-time increment due to origination at the maximum distance from the centre.

For the total contribution of the electron-optical input system

$$\sigma_{k,d1}^2 = \sigma_{vn}^2 + \sigma_{vt}^2 + \sigma_{cb}^2 \tag{4.14}$$

For fast-response photomultipliers, $\sigma_{k,d1}$ is usually between 150 and 350 ps, depending on the cathode/first-dynode voltage and the incident-light wavelength.

The contribution of the electron multiplier, σ_m, is mainly from the first stage, for two reasons:

- the number of secondary electrons there is smaller than in subsequent stages, so the standard deviation tends to be larger;
- the role of the first stage as a coupling zone between the electron-optical input system and the iterative part of the multiplier also tends to increase the standard deviation.

Again, there is a chromatic component and a geometric one. The chromatic component is due to the spread of secondary electron initial velocities. The geometric one, which may predominate and can be an important factor in the effect of wavelength on overall transit-time spread, is due to the scatter of electrons on the first dynode. It thus depends on the primary-electron velocity spread. In fast-response tubes σ_m is between about 150 ps and 250 ps, depending on voltage and wavelength.

Effect of applied voltage. The spread of initial velocities varies as $1/V_{d,d}$, and the spread of transit-time differences as $1/\sqrt{V_{d,d}}$. Furthermore, the stage gain g in Eq.4.12 varies as a power of $V_{d,d}$ between 0.65 and 0.75. Hence the overall transit time spread varies as $V_{d,d}^{-n}$, where n is between 0.5 and 1.

Effect of wavelength. Wavelength affects the energy distribution of the photoelectrons and, hence, their initial velocity spread, which increases as wavelength decreases. Although wavelength changes have little effect on the geometric component in the electron-optical input system, they do affect the chromatic component there, σ_{vi}, and the geometric component σ_{cb} in the first multiplier stage. The overall effect depends on the relative importance of these contributory effects. With a bialkali cathode, a wavelength increase from 400 nm to 560 nm decreases the transit-time spread about 40%.

Measurement of transit-time spread. Transit-time spread is measured by recording the intervals between a clocked series of light pulses and the corresponding series of anode pulses (Fig.4.8). The transit time probability distribution depends on the mean number of photoelectrons, $\overline{n}_{k,i}$, emitted per light pulse, the variance being greatest for single-electron operation. The measured probability distribution can also depend to some extent on the statistics of photon emission because the timing reference chosen is a light pulse.

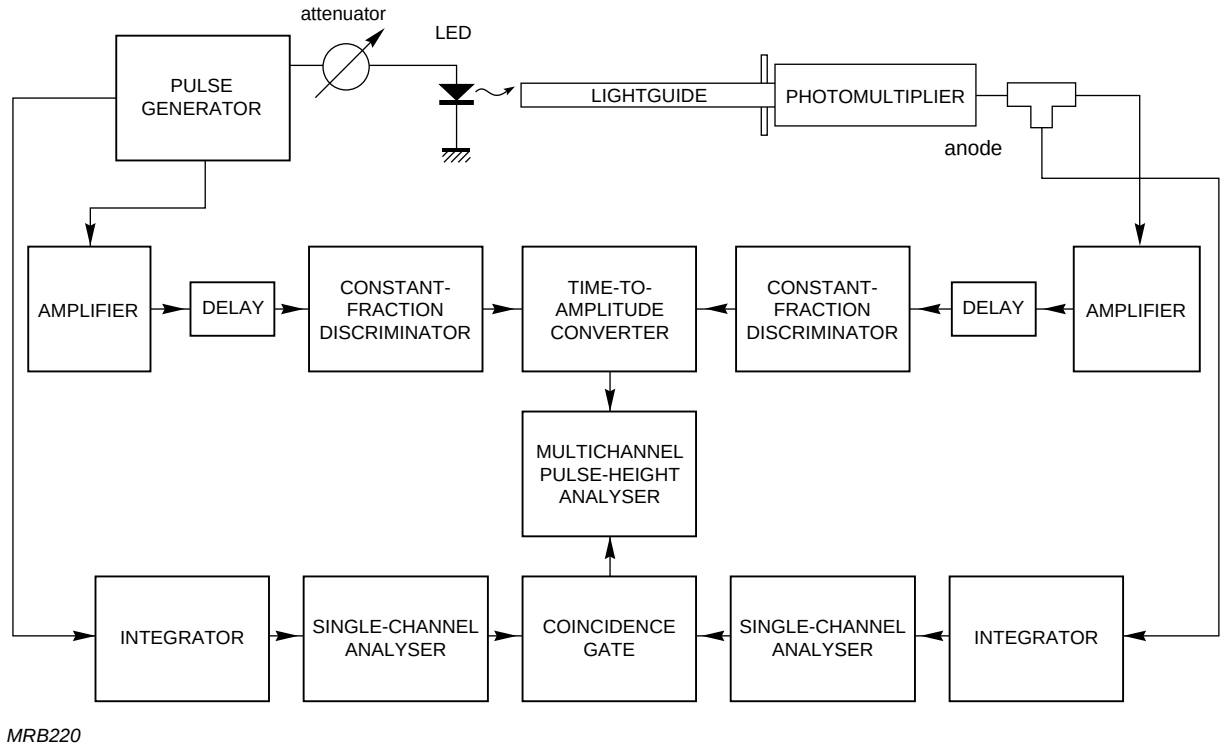


Fig.4.8 Set-up for measuring transit-time differences and transit-time spread

Single-electron operation. Let L(t) be the probability distribution of the instants of photon emission, such that

$$\int_0^\infty L(t)\ dt = 1$$

and R(t) the probability distribution of the occurrence of the corresponding anode pulses referred to the instants of photoelectron emission. Then the measured probability distribution of the transit time will be

$$R^{*}(t) = L(t) * R(t) \tag{4.15}$$

The single-electron time resolution, designated R_t^*, is the FWHM of the $R^*(t)$ curve. Provided the light pulses with which it is measured are very short in comparison with R(t), R_t^* is an accurate measure of the transit time spread. Figure 4.9 is a single-electron time resolution curve of a fast-response photomultiplier.

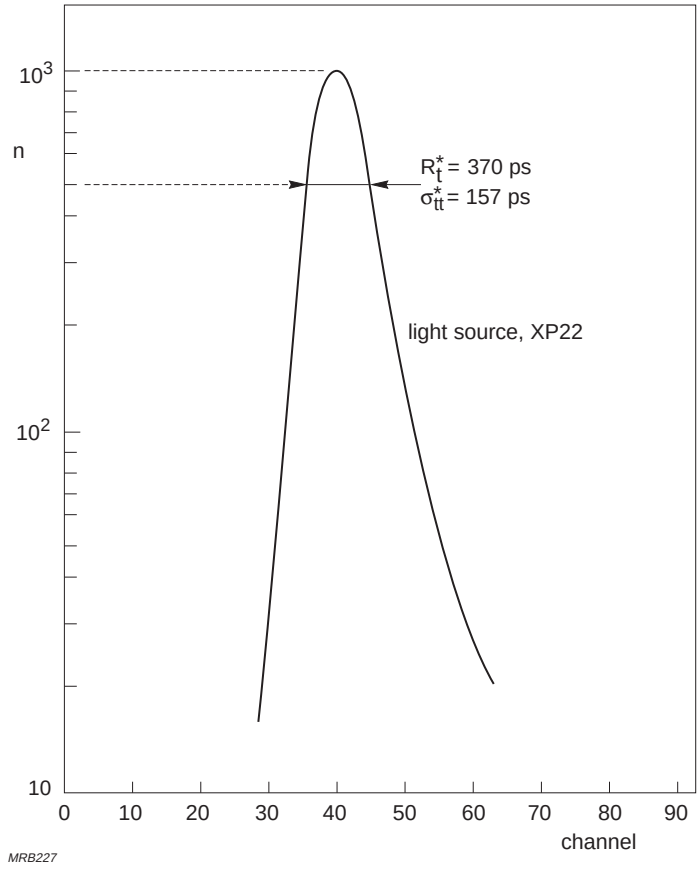


Fig.4.9 Time resolution of a fast-response photomultiplier, for single-electron pulses originating from a single point on the cathode at an illumination wavelength of 560 nm, measured according to the method of Fig.4.8. Vertical axis, number of pulses per channel (arbitrary scale); horizontal axis, channel number (about 40 ps per channel)

If L(t) and R(t) are approximately gaussian, with variance σ_L^2 and σ_{tt}^2, then $R^*(t)$ will also be gaussian, with a variance

$$\sigma_{tt}^{*2} = \sigma_L^2 + \sigma_{tt}^2 \tag{4.16}$$

and FWHM

$$R_t^* = 2.36\ \sigma_{tt}^* \tag{4.17}$$

Multi-electron operation. Provided L(t) and R(t) are gaussian,

$$\sigma_{tt}^{*2} = \frac{\sigma_L^2 + \sigma_{tt}^2}{\bar{n}_{k,i}} \tag{4.18}$$

For fast-response tubes in single-electron operation at an illumination wavelength of 400 nm, R_t^* is usually less than 1 ns; in operation with an average of 10 photoelectrons per light pulse this is divided by $\sqrt{10}$ and becomes less than 320 ps.

4.4.5 Frequency response

The frequency response G(f) is important in applications involving modulated light. Its upper limit is mainly due to statistical effects and imperfect matching of the output to the external circuit. The frequency response can be derived from the pulse response $R_\delta(t)$ via the Fourier transform

$$G(f) = \int_{-\infty}^{\infty} R_\delta(t)\ \exp(-j2\pi ft)\ dt \tag{4.19}$$

The narrower the pulse response, the higher the cut-off frequency. If the pulse response is gaussian with a variance σ_R^2,

$$R_\delta(t) = \frac{1}{\sigma_R\sqrt{2\pi}}\ \exp\left\{ -\frac{(t - t_t)^2}{2\ \sigma_R^2} \right\} \tag{4.20}$$

and, from Eq.4.19,

$$G(f) = \exp\left[-2(\pi f\sigma_R)^2 - j2\pi ft_t \right] \tag{4.21}$$

the absolute value of which is

$$|\ G(f)\ | = \exp\left[-2(\pi f\sigma_R)^2 \right] \tag{4.22}$$

This corresponds to a 3 dB bandwidth

$$B_{3dB} = \frac{0.133}{\sigma_R} \tag{4.23}$$

The bandwidth can also be expressed as a function of the step-response rise time $t_{r,\varepsilon}$. If the pulse response is gaussian, with a variance σ_R^2 (and FWHM $t_w = 2.36\ \sigma_R$), the rise time in response to a unit step $\varepsilon(t)$ is (§A4.1.2)

$$t_{r,\varepsilon} \approx 1.11\ t_w = 1.11 \times 2.36\ \sigma_R \tag{4.24}$$

whence, from Eq.4.23,

$$B_{3dB} \approx \frac{0.35}{t_{r,\varepsilon}} \tag{4.25}$$

For a photomultiplier with $t_w \approx 3$ ns, $B_{3dB} \approx 105$ MHz.

(Some authors have suggested other ways of describing the step response of a photomultiplier yielding a comparable 3 dB bandwidth.)

Note. In accordance with IEC standard, Publications 306-4 (1971) and 462 (1974), the pulse response is specified in terms of delta-function rise time (t_r) and FWHM (t_w); this is not to be confused with the step-response rise time $t_{r,\varepsilon}$ used for calculating the bandwidth.

4.5 Linearity

The degree of proportionality between the number of electrons collected at the anode and the number of incident photons is called *charge linearity*. The degree of proportionality between incident flux and anode current is called *current linearity*; in this relationship therefore, *time* is an additional parameter. Limits on both charge and current linearity are set by internal and external factors.

4.5.1 External factors affecting linearity

Power supply. Changes in interelectrode voltages affect gain (Fig.4.10) by influencing the dynode secondary emission factors and the electron trajectories.

Divider current. When the electrode voltages are derived from a resistive divider across a stabilized power supply, the anode current I_a tends to lessen the potential between the last dynode and the anode. This upsets the voltage distribution throughout the divider and causes an increase of gain comparable to what would be caused by increasing the high voltage by the same amount.

The current through an iterative divider ($R = R_1 = R_2 \ldots = R_N$) when there is no anode current is

$$I_p = \frac{V_{ht}}{(N + 1)R} \tag{4.27}$$

where V_{ht} is the high voltage and R is the common value of the resistors. When there is an anode current, the currents in the resistors are altered as shown in Fig.4.11. To keep V_{ht} constant, I_p must assume a new value

$$I_p' = I_p + \Delta I_p = I_p + \frac{1}{N + 1} \sum_{i = 0}^{N} I_i \tag{4.28}$$

From Eq.2.7 the gain at divider current I_p' is

$$G = KV_{ht}^{N\alpha} = KR^{N\alpha} \prod_{i = 0}^{N-1} (I_p + \Delta I_p - I_i)^{\alpha} \tag{4.29}$$

and the gain when the anode current is zero is

$$G_m = KR^{N\alpha} I_p^{N\alpha}$$

Thus, the ratio of the gain at divider current I_p to that when the anode current is zero is

$$\frac{G}{G_m} = \prod_{i = 0}^{N-1} \left(1 + \frac{\Delta I_p - I_i}{I_p} \right)^{\alpha} \tag{4.30}$$

or, if we neglect terms higher than the first order

$$\frac{G}{G_m} = \left\{ 1 + \sum_{i = 0}^{N-1} \left(\frac{\Delta I_p}{I_p} - \frac{I_i}{I_p} \right) \right\}^{\alpha}$$

Substituting the value of ΔI_p from Eq.4.28 gives

$$\frac{G}{G_m} = \left\{ 1 + \frac{I_N}{I_p} \left(\frac{N}{N + 1} - \frac{1}{N + 1} \sum_{i = 0}^{N-1} \frac{I_i}{I_N} \right) \right\}^{\alpha} \tag{4.31}$$

As the quantity under the summation is always less than unity, the quantity between the inner parentheses is positive and the gain G is an *increasing* function of the ratio $I_N/I_p = I_a/I_p$.

Throughout the range of variation $I_N \leq I_p$ the ratio I_N/I_p is nearly independent of the gain of each stage; so, by setting

$$\frac{I_i}{I_{i-1}} = g_i = g$$

and noting that $g^N >> 1$, Eq.1.31 can be simplified to

$$\frac{G}{G_m} = \left\{ 1 + \frac{I_a}{I_p} \left(\frac{N}{N+1} - \frac{1}{(N+1)(1-g)} \right) \right\}^{\alpha}$$

Or, for large values of N,

$$\frac{\Delta G}{G} = \frac{G - G_m}{G} \approx \alpha \frac{N}{N+1} \frac{I_a}{I_p} \qquad (4.32)$$

Equation 4.32 expresses the relative variation of the gain as a function of the ratio of the anode current I_a to the nominal divider current at $I_a = 0$, provided the decrease of voltage across the terminals of the last stage does not impair collection efficiency. Under these conditions (and provided no internal factors work against it), the ratio $\Delta G/G$ has the same sign as I_a/I_p: an increase of I_a results in an increase of gain (Fig.4.12). This increase (or overlinearity) as a function of I_a is largely independent of N. For $\alpha = 0.7$, $N = 10$, and $I_a/I_p = 0.1$, it amounts to about 7%.

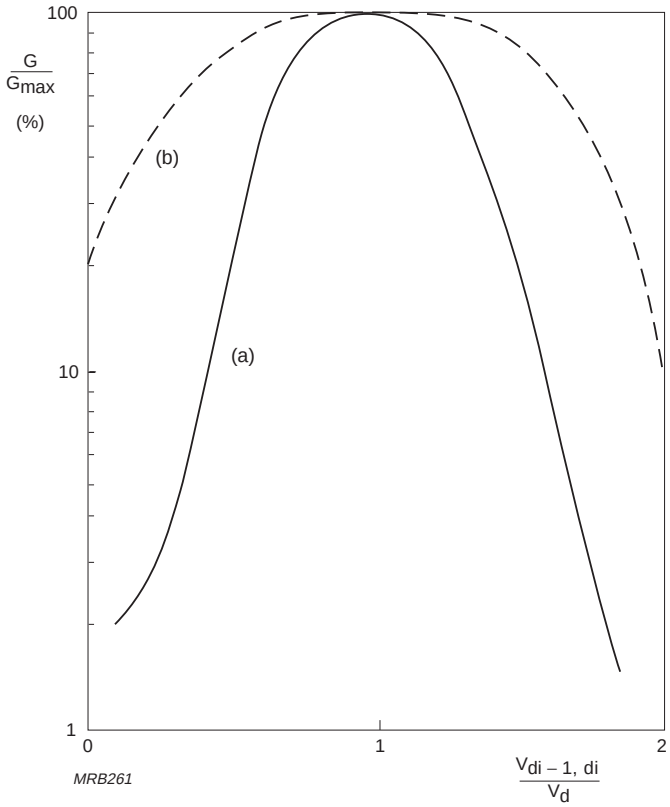


Fig.4.10 Gain variation as a function of the voltage between successive dynodes, for a tube with (a) linear focusing dynodes, (b) venetian-blind dynodes. V_d = nominal interdynode voltage

When the ratio I_a/I_p approaches unity, Eq.4.32 no longer holds. The voltage drop in the last stage, which increases with I_a, becomes too great and collection efficiency declines rapidly, leading to an abrupt decrease of gain. Internal factors may accentuate this effect (§4.5.2). To maintain linear operation in resistive voltage dividers, a good rule is to ensure that the ratio $I_a/I_p \leq 0.01$ (see Fig.4.12 and §5.2.2).

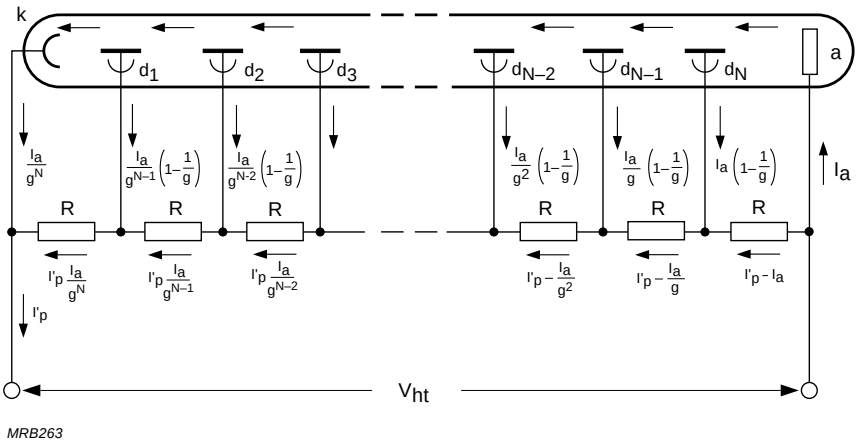


Fig.4.11 Current distribution in an iterative voltage divider

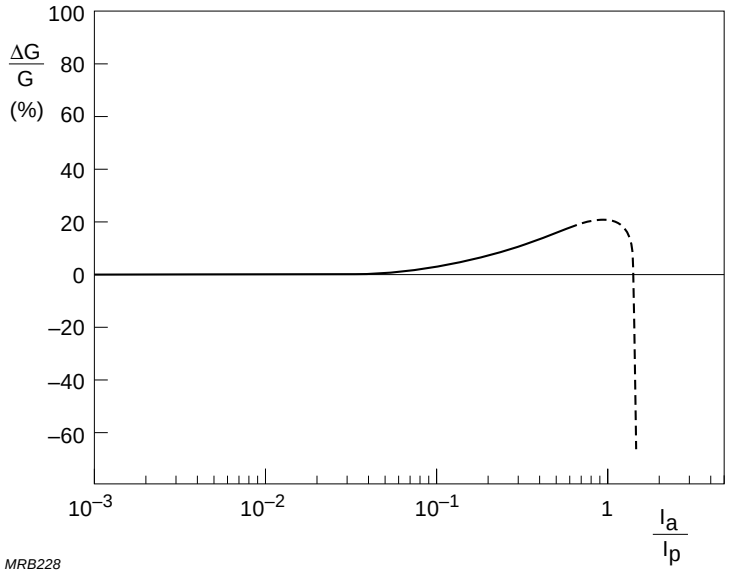


Fig.4.12 Gain variation (overlinearity) as a function of the ratio I_a/I_p

The maximum value of the ratio $\Delta G/G$ depends mainly on the voltage across the tube and how the drop between anode and last dynode affects the voltage distribution among the first stages.

To ensure that the ratio $I_a/I_p \leq 0.01$, I_p may be increased, but this solution is limited by power dissipation which may not exceed a few watts.

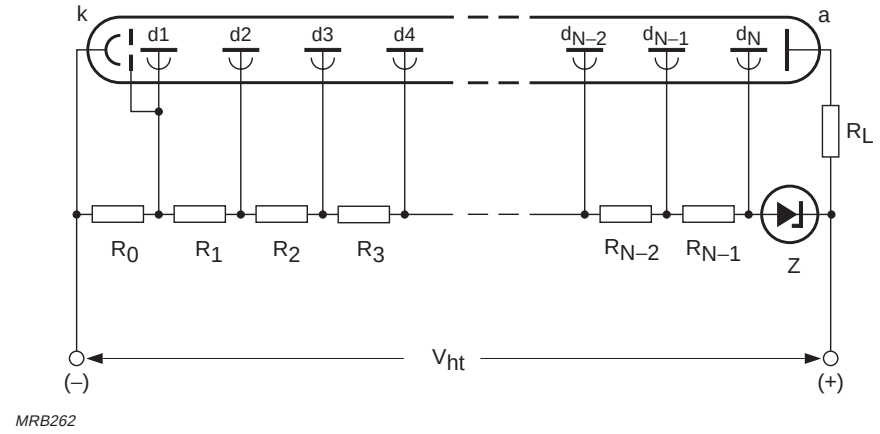


Fig.4.13 Voltage divider with zener diode for improved linearity

The dependence of gain on anode current can be lessened by substituting a zener diode for the resistance R_N and possible for R_{N-1} (Fig.4.13); the overlinearity is then eliminated and the anode current at which the rapid decrease of gain occurs becomes about ten times higher.

A voltage divider composed solely of zener diodes is never used: firstly, because it makes it impossible to adjust the gain by adjusting the supply voltage; and secondly, because it provides no current-limiting action to guard against the consequences of accidental overexposure of the cathode. In high-current applications, therefore, active dividers (employing transistors) are often preferred.

Reservoir capacitors. When the anode current can reach high values for only a small fraction of the time (short-pulse operation), it is preferable to connect reservoir (or decoupling) capacitors to the dynodes (§5.2.5). The charge stored by the capacitors must be sufficiently large compared with that supplied by each dynode when pulses pass through the tube so that the dynode potentials will not vary by more than one or two volts. Calculation of the required capacitance values differs according to whether the decoupling is parallel or series; in the latter case voltage variations are

cumulative. When space or insulation considerations limit the size of the capacitors that can be used, their values must be calculated with especial care, as inadequate decoupling affects charge linearity rather than current linearity, which could cause misleading results.

Damping resistors. When a photomultiplier is operating in the pulse mode, a high-frequency spurious oscillation superimposed on the anode pulses may be observed, even with pulses as wide as a few hundred nanoseconds. This oscillation, which affects the linearity characteristic of the tube, usually producing an overlinearity, may appear abruptly when the anode current exceeds a certain level. One way of overcoming this effect is to connect a 50 Ω non-inductive resistor in series with each of the last two or three dynodes. Fast-response photomultipliers with plastic bases have such resistors built-in. For other types they must be wired into the socket, between the base and the decoupling capacitors.

Anode load. The voltage developed across the anode load subtracts from the last-dynode to anode voltage and, if it is not negligible compared with that, may affect linearity. As the load voltage rarely exceeds a few volts, however, this is seldom the case.

4.5.2 Internal factors affecting linearity

Space charge. At high currents, space charge can influence the electron trajectories, causing collection losses; at still higher currents it can cause some electrons to return to the surfaces from which they originate. The condition resembles that of a space-charge limited diode with parallel-plane electrodes, for which the relation between current density J_s (in A/cm^2) and electric field is given by the Child-Langmuir equation

$$J_s = 2.2 \times 10^{-6} \frac{V^{3/2}}{d^2} \tag{4.33}$$

where V is the interelectrode voltage in volts, and d the interelectrode distance in centimetres.

The current density is normally highest between the last dynode and the anode. To ensure a high field there, the anode is positioned close to the surface of the last dynode and made in the form of a grid through which the electrons pass on their way from the next-to-last dynode. Then, it is the field between the next-to-last dynode and the anode, which is 3 to 5 times lower, that sets the limit for current linearity in most photomultipliers.

That limit can be raised by using a progressive instead of an equal voltage distribution in the last stage (§5.2.1), so as to raise the voltage between the last two stages to as much as 300 V or more. To maintain correct focusing between dynodes without unduly increasing the gain, the interelectrode voltages are progressively decreased in the anode to cathode direction so that the nominal value applies at the terminals of the first stages. For tubes with focusing dynodes, the data sheets give, in addition to the conventional voltage distribution, one or two examples of recommended progressive distributions. Using these, the maximum pulse current in linear operation can be increased from 10 – 50 mA to 100 – 300 mA. For some special tubes, linear pulse operation can be obtained at currents of more than 5 A.

For tubes with venetian-blind or box-and-grid dynodes, the maximum pulse currents for linear operation are smaller (10 – 50 mA) because of the very low electric fields between all dynodes other than the last.

For most tubes, the current linearity limit due to space charge varies as V_{ht}^{n}, where n is between 2 and 3. This is merely approximate, but when the limit at one voltage is known from the published data it gives a practical indication of the limit at another voltage, especially if the onset of saturation is progressive. If linearity is not important, the maximum anode current that can be obtained before saturation is several times greater than the maximum for linear operation.

The space charge phenomena that limit current linearity last for times comparable to the transit times between dynodes, that is, 1 to 2 ns. Even when linearity errors are severe, there is no charge accumulation and the errors are strictly related to the electron current passing between the last dynodes.

Current linearity is important when pulses are wide compared with the pulse response of the tube; when they are of the same order as the pulse response, it is no longer relevant. The significant parameter then is charge linearity. Depending on the shape of the pulses, higher peak anode currents, can be obtained under short-pulse conditions while still maintaining good charge linearity.

Data sheets specify only the current linearity limit, not the charge linearity limit, and for a worst-case situation with anode pulses about 100 ns wide.

Cathode resistivity. The electron-optical input system is designed on the assumption that the cathode is an equipotential surface. Any departure from that condition is likely to alter the electron trajectories and affect the collection efficiency of the first dynode. This is what happens, at least in the case of semitransparent cathodes having

no underlying conductive layer, when the cathode current is too large in relation to the surface resistivity.

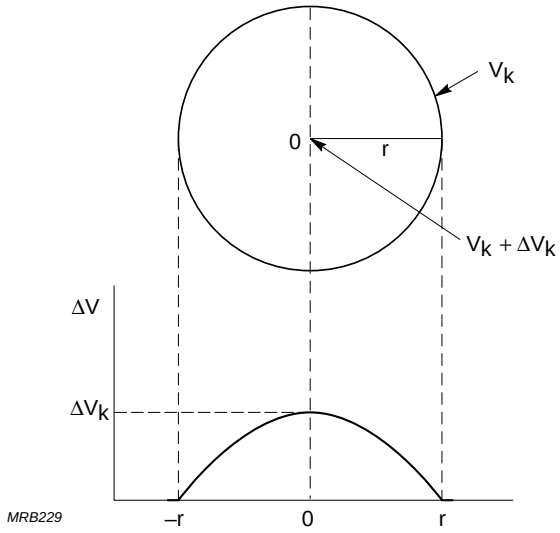


Fig.4.14 Potential distribution due to cathode resistivity; r is the cathode radius

Consider a circular cathode of uniform sensitivity, uniformly illuminated and emitting a total current I_k (Fig.4.14). Let $R_\square$ be its surface resistivity (the bulk resistivity divided by the thickness); the potential difference between the centre and the edge is then

$$\Delta V_k = \frac{R_\square \, I_k}{4 \, \pi} \tag{4.34}$$

If it exceeds a few volts, this potential difference increases the input-system convergence and causes loss of electrons emitted from the cathode edge. More complicated effects occur when only small areas of the cathode are illuminated. These lead to a dynamic variation of gain as a function of cathode current; in other words, to linearity errors. However, for the cathode current normally encountered, such phenomena are practically significant only with bialkali cathodes. For these, the surface resistivity is about 10^{10} Ω at ambient temperature, decreasing slightly with illumination but increasing rapidly as temperature decreases. A tube with a 45 mm diameter bialkali cathode exhibits a non-linearity of a few percent at a mean cathode current of about 10 nA at ambient temperature; at −30°C the same non-linearity occurs at a current of only 0.1 nA. In tubes with larger cathode diameters, the currents at which comparable non-linearity occurs are even lower. In tubes with S11 and SbRbCs cathodes, comparable non-linearity at ambient temperature occurs at currents respectively about 100 times and 10 times higher.

Fortunately, the distributed capacitance of the cathode (about 1 pF) is sufficient to store a charge of about 10^{-12} C. At a gain of 10^5, this corresponds to an anode pulse of 100 mA amplitude and 1 μs duration; cases in which cathode resistivity actually presents a problem are therefore fairly uncommon.

Gain drift. Gain may undergo more or less reversible variations when the mean anode current varies. Although this too constitutes a linearity error, by convention it is treated as an instability (§4.6).

4.5.3 Linearity measurement

Many methods of linearity measurement have been developed but all are limited to an accuracy no better than 2%. Two types of gain drift may interfere with the measurement:

– long-term, time-dependent drift (§4.6.1)
– short-term shift due to changes of illumination (§4.6.2).

To avoid these, the measurement must be made quickly and with a mean current not exceeding a few microamperes. The measurement should result in determining the anode current at which space charge limiting starts to become evident, avoiding all other causes of linearity limiting. The methods described below are for measuring either current or charge linearity.

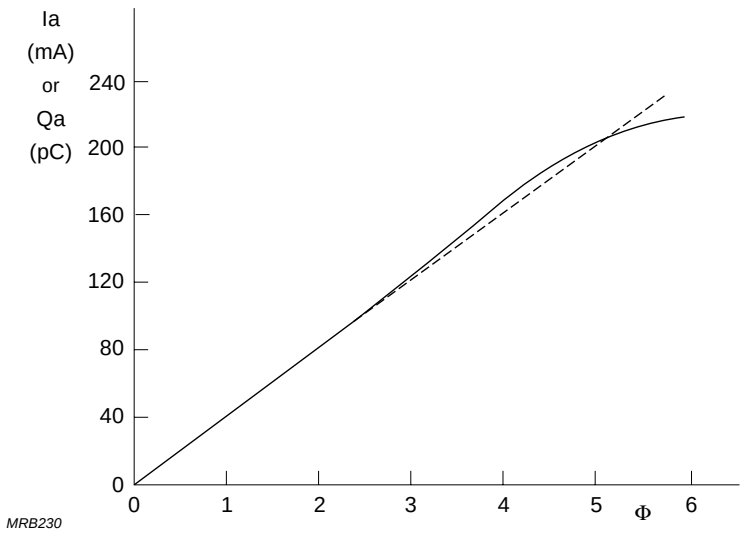


Fig.4.15 Typical current or charge linearity characteristics of a photomultiplier operating from a supply with type B voltage division (flux Φ in arbitrary units)

Figure 4.15 shows a typical linearity curve, in which a slight overlinearity appears before saturation. Such overlinearity is often observed with voltage dividers designed for delaying the onset of saturation at high current levels (§5.2.1). It can be corrected by adjusting the voltages of the stages immediately preceding the last, but at the cost of lowering the current threshold beyond which saturation occurs (§5.2.3).

XY method. This method makes use of an oscilloscope having identical X and Y deflection factors (Fig.4.16). The anode pulse of the photomultiplier under test deflects the beam parallel to one axis, and the anode pulse of a reference photomultiplier operating in its known linear region deflects it parallel to the other. The measurement is therefore one of instantaneous current linearity. Both tubes are excited simultaneously by light pulses of a few hundred nanoseconds.

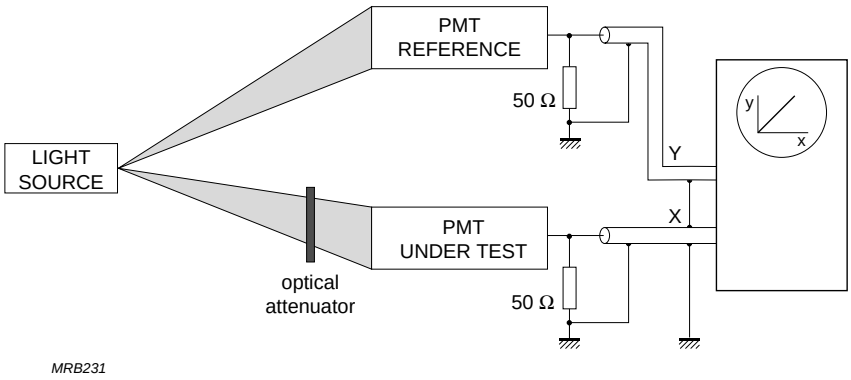
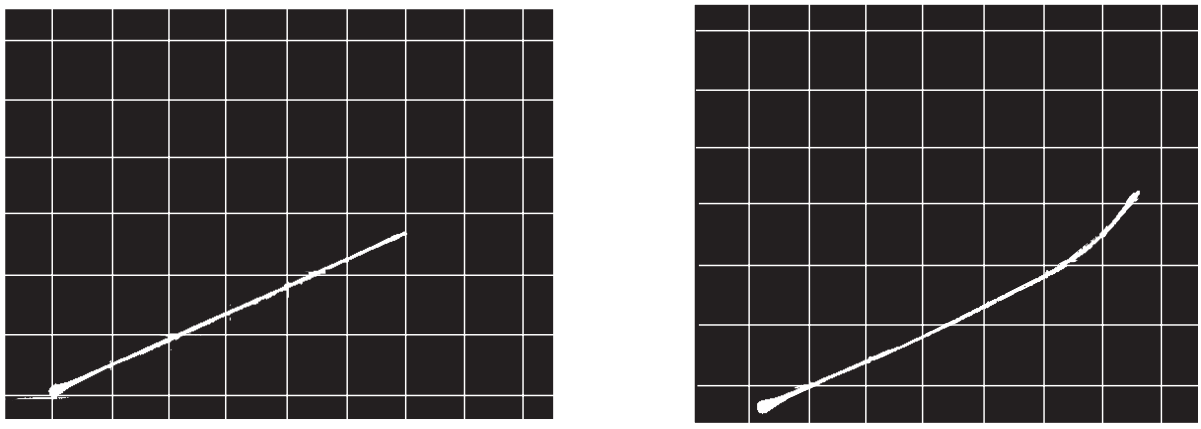


Fig.4.16 Set-up for the XY-method of measuring linearity



(a) (b)

Fig.4.17 Oscillogram obtained by the method of Fig.4.16 showing (a) linear and (b) non-linear response. Scales, 50 mA/div

At each level of pulse current corresponding to the setting of an optical attenuator, an oscilloscope trace is obtained. This may have, for example, the shape of Fig.4.17(b), with a linear part and a curved part; the linearity error is the percentage by which the curved part of the trace departs from the line of the straight part. This is a measurement of *integral linearity*. In Fig.4.17(b), the tube under test is linear within 5% up to 200 mA and within 10% up to 300 mA.

Dual pulse method. The flux from a practically monochromatic source (for example a LED) can be calibrated with great accuracy. Two such light sources controlled by separate generators send light pulses to the photomultiplier under test. The width of the pulses is about 100 ns and their amplitudes are in a fixed ratio, for example 2:1. Provided the timing of the two generators is known, the pulses from each can be recognized by the processing electronics. The pulse repetition frequency must decrease with increasing pulse height to ensure that the mean anode current of the photomultiplier remains constant (at a value < 1 μA) throughout the whole range of pulse amplitude variation.

The method consists in increasing the light-pulse amplitudes, while maintaining their ratio (for example, by the use of neutral filters), and monitoring the height or charge ratio of the current pulses supplied by the photomultiplier. In this way, a pulse-current value can be determined beyond which the height or charge ratio differs by a given amount from the initial ratio. This is a *differential linearity* measurement. An advantage of it is that it does not require the use of a reference photomultiplier.

The ratio of the photomultiplier pulses can be measured with a multichannel pulse-height analyser, or with an oscilloscope, but not so accurately.

Composite radiation method. This method consists in exposing a scintillator-photomultiplier combination to radiation from a composite γ-source emitting several known energies between a few hundred keV and a few MeV (Fig.4.18). The relative heights of the integrated current pulses of the photomultiplier are then measured with a multichannel pulse-height analyser. When the mean amplitudes of the pulses corresponding to each γ line are plotted against energy, the resulting curve has an initial linear part followed by a curved part indicative of either overlinearity or saturation. The anode pulse height beyond which the curve departs from straightness by a given amount is a measure of the integral linearity of the scintillator-photomultiplier combination, and from this it is possible to determine the linearity of the photomultiplier alone. (As inorganic scintillators are not linear at low energies, pulses corresponding to γ-energies of less than a few hundred keV should be disregarded.)

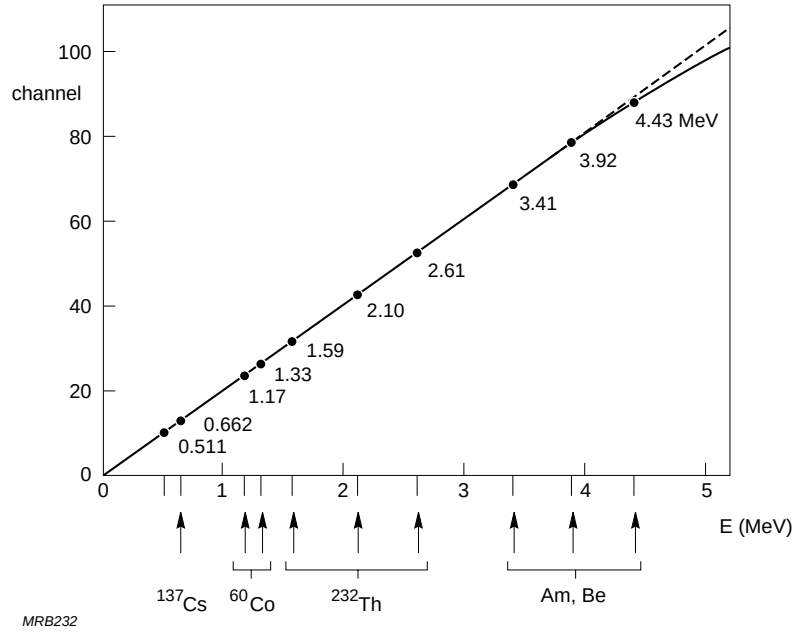


Fig.4.18 Distribution of γ-ray energies emitted by a source consisting of ^{60}Co, ^{137}Cs, ^{232}Th and AmBe as used in measuring photomultiplier linearity. Vertical scale, channel number of multichannel pulse-height analyser

The composite radiation method is especially applicable to high-gain tubes: at a gain of 10^7 an energy of 1 MeV absorbed in a NaI(Tl) scintillator gives an anode pulse of about 30 mA. A major drawback of the method is the near-impossibility of adjusting the amplitude of the light pulses. The amplitude of the anode pulses can therefore be adjusted only by varying the gain. Since this means varying the high voltage, it can itself affect the linearity to be measured.

Method using bursts of three pulses. In this method, a special pulse generator, giving burst of three calibrated pulses of increasing amplitude is used to drive one LED. The light reaches the cathode, passing through a neutral optical filter giving a five-fold attenuation. The anode pulses feed a multichannel analyzer (Fig.4.19 (a)). and the registered channel numbers p_1, p_2, and p_3 become p_1', p_2' and p_3' when the filter is withdrawn and an five-fold *electrical* attenuator is placed before the multichannel analyser. An on-line calculator can be used to calculate the ratios $a = p_2/p_1$ and $a' = p_2'/p_1'$ and to calculate the linearity deviation $(a' - a)/a$. In the same way, the ratios $b = p_3/p_1$ and $b' = p_3'/p_1'$ give the linearity deviation $(b' - b)/b$.

A typical test setup could, for example, comprise the following:
- burst of three pulses of 50 ns giving, when the filter is ON, three anode pulses of 6, 20 and 30 mA peak value at the operating gain
- time between each pulse of 100 μsec
- burst frequency of a few kilohertz.

When the filter is OFF, the anode pulses should be 30, 100 and 150 mA.

Fig.4.19(b) shows the results of this test. The linearity deviation is then calculated between the channel numbers corresponding to the ideal values 100 – 30 mA on the one hand, and 150 – 30 mA on the other hand, with respect to the tube linearity at very low level. One advantage of this method (which measures the differential linearity) is that the result is not affected by the possible gain shift due to the change of mean anode current that occurs when the filter is withdrawn.

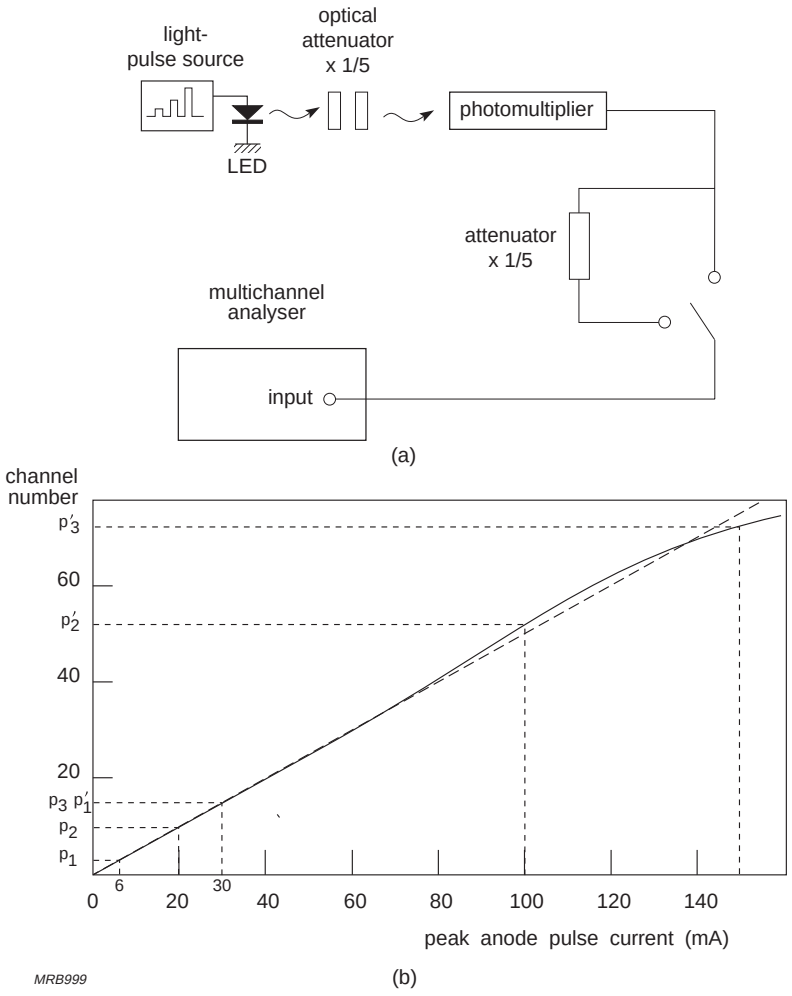


Fig.4.19 Pulse-linearity test; (a) block diagram of the 'three-pulses method'; (b) example of an experimental linearity curve obtained with the three-pulses method (channel number in arbitrary units)

4.6 Stability

The term 'stability' is used to describe the relative constancy of anode sensitivity with time, temperature, mean current, etc. The most important departures from constancy are:

– **long-term drift**, which is a time-dependent variation of gain under conditions of constant illumination
– **short-term shift**, which is a variation of gain following a change in mean current.

4.6.1 Long-term drift

Two modes of long-term drift can be distinguished, according to whether the mean anode current is high or low.

High-current drift; operating life. Certain more or less irreversible effects are observable at mean anode currents larger than about 10 μA. After long storage (e.g. a few months), a photomultiplier exhibits a large drift of gain for the first one or two days of operation. For some thousands of hours after that the gain is relatively stable, then it slowly decreases as a function of the total charge handled. The rate of these variations varies roughly as the anode current of the tube.

Operating life, defined as the time required for anode sensitivity to be halved, appears to be a function of the total charge delivered. Values of 300 to 1000 coulombs are typical. If the incident flux is reduced (by, say, 90%) or cut off completely, or if the supply voltage is switched off for several days, the following sequence can be observed when the original operating conditions are restored: first, a certain recovery of sensitivity accompanied by a renewed initial drift; then, a tendency to catch up fairly quickly with the slow decline of sensitivity at the point at which it was interrupted.

Figure 4.20 illustrates the relative gain variation of a photomultiplier operating at a mean anode current of 30 μA. The initial drift, which can be considered an ageing period, is between 20% and 40%. The duration of the ageing period depends on the anode current; at 10 μA it is about 24 hours. As long as the mean current does not fall below about 100 nA, ageing is still observable though very slow.

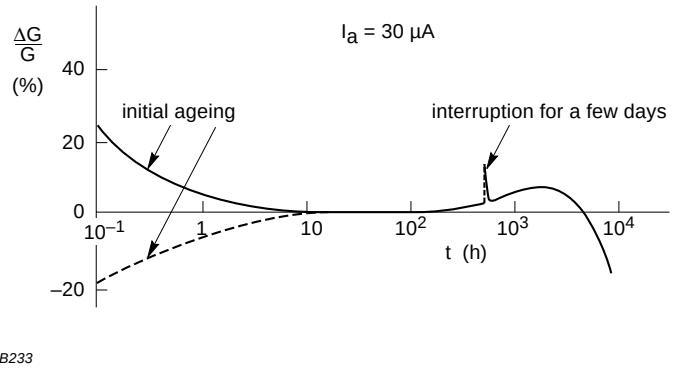


Fig.4.20 Relative gain variation of a photomultiplier operating at high average current

In most cases, if the gain is high and the cathode current low, the variations of anode sensitivity reflect variations of gain due to changes in the surface state of the dynodes. This is commonly attributed to the mobility under electron bombardment of the alkali metal molecules (mainly potassium and caesium) with which they are coated, though the exact mechanism is probably more complicated than the literature suggests.

When mean anode current is only a few microamperes, total charge delivered is no longer the decisive factor for operating life. Other effects, such as helium migration through the glass or internal migration and diffusion balances, determine the end of useful life, which is then measured in years and is independent of the mode of operation. The experience of many users would even seem to indicate that continuous, uninterrupted operation results in better long-term stability of performance characteristics than storage.

Photomultipliers with S1 cathodes deserve separate mention. Even at anode currents of only a few microamperes they exhibit large short-term drift which is independent of the gain adjustment. This drift is reversible, and the process can be speeded up by heating the tube for a few hours at the maximum permissible temperature, which suggests that there is some exchange of molecules between the dynodes and surfaces not subject to electron bombardment.

Low-current drift. When a photomultiplier is switched on and subjected to more or less constant illumination, its gain changes over the first few hours or days. The amount of change differs from type to type and even from one specimen to another of the same type. In most cases, though, the rate of change quickly decreases to as low as one per cent a month (Fig.4.21), and the higher the current the quicker the gain stabilizes. It is sometimes worthwhile to speed the process by operating the tube initially at a current up to ten times higher than that expected in the intended application. It is also advisable to leave the tube switched on even when it is idle. If the tube is stored for a time comparable with its former operating period, the gain change reverses and is repeated when the tube is again put into service.

The ANSI[1)] test, which is used to characterize this type of drift, employs a scintillator and a ^{137}Cs source positioned so as to produce a fixed count rate between 10^3 and 10^4 per second. After a stabilization period of 30 to 60 minutes, the height

[1)] ANSI: American National Standards Institute N42.9-1972

of the ^{137}Cs peak (662 keV) is recorded every hour for the next 16 hours and the mean gain deviation (MGD) calculated according to the formula:

$$MGD = \frac{\sum_{i=1}^{17} | p - p_i |}{17} \cdot \frac{100}{p} \qquad (4.35)$$

where p is the mean height of the peak averaged over the 17 readings and p_i the height corresponding to the *ith* measurement.

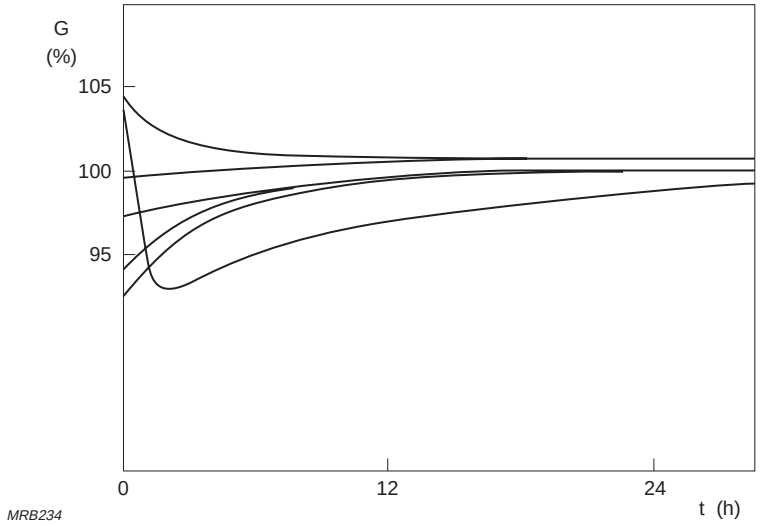


Fig.4.21 Examples of initial low-current drift

This type of drift is not related to the high-current long-term drift previously described. Though its major cause is also related to change in the structure of the emissive surfaces, other factors, such as the charge distribution at insulator surfaces (e.g. dynode spacers), may also play an important part. The drift is much less (MGD < 1%, typically) in tubes with multialkali cathodes and CuBe or SbCs coated dynodes than in those with S11 cathodes or AgMg dynodes.

The ANSI test specification does not mention the anode sensitivity at which the test is to be performed. However, when a figure for long term stability is given, the mean anode current during the test must be specified. Values of about a microampere are generally used because they are broadly representative of most applications. For convenience, the scintillator and source used in the ANSI test may be replaced by a LED.

Figure 4.22 gives some examples of anode sensitivity variation curves having the same maximum deviation but different MGD values. For some applications, one may want to know the stability over a long period (for example, a month). It is possible to determine an MGD over such a period, but the measurement is more difficult because of the likelihood of drift in the measuring system itself. For such measurements, a radioactive source in combination with a scintillator is preferable to a LED because its long-term stability is much better.

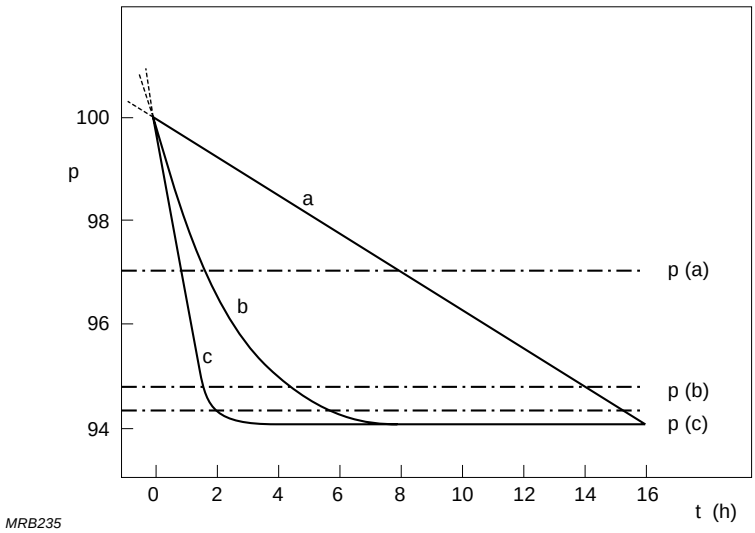


Fig.4.22 Anode sensitivity curves showing the same absolute change over 16 hours but different values of mean gain deviation (MGD) according to the ANSI method. Curve a: MGD = 1.6%; curve b: MGD = 1.1%; curve c: MGD = 0.75%; p(a), p(b) and p(c) are the corresponding mean heights of the pulse peak averaged over 17 readings

4.6.2 Short-term shift

When the flux to which a photomultiplier is exposed gives rise to a mean anode current of less than 10 μA, the gain is usually sufficiently stabilized after about 10 or 15 minutes for its long-term drift to be disregarded. If the flux is then changed abruptly, the anode current, instead of assuming a new value abruptly, starts a new drift phase before stabilizing again (Fig.4.23). Thus, the gain becomes a function (often an increasing one) of the mean value of the anode current reckoned over an interval of a second or longer.

For most photomultipliers, the time required to stabilize the gain after changing the average flux is around a second. But in some cases, and especially for tubes with S11 cathodes, this fast shift is augmented by one with a much longer time constant (about an hour). Figure 4.24 gives an example of shift with a single, short time constant; and Fig.4.25 an example with two time constants, one short and one long.

Two methods are used to measure the gain shift due to a change of average flux. The ANSI test uses a ^{137}Cs radioactive source and NaI(Tl) scintillator. After a stabilization time of at least 15 minutes, the position of the ^{137}Cs absorption peak is recorded at a count rate of 10^4 per second. The source is then moved to reduce the rate to 10^3 counts per second and the new position of the peak is recorded. The shift is characterized by the relative shift of the peak, $\Delta p/p$.

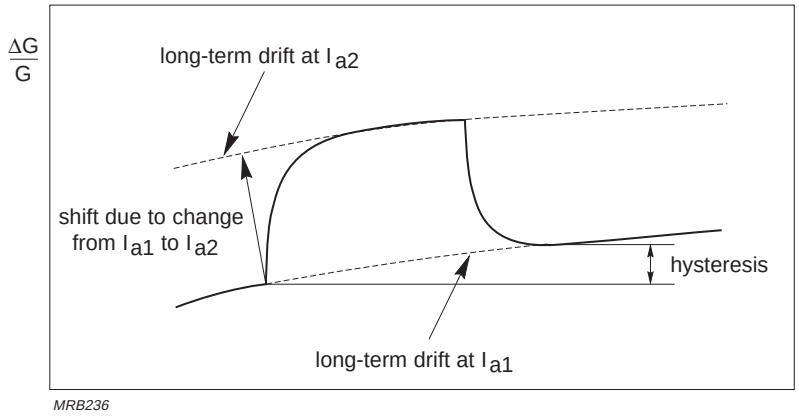


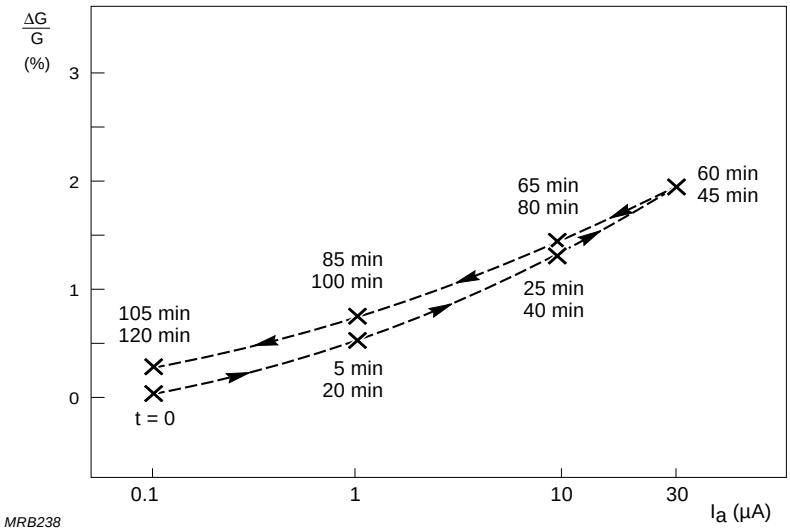
Fig.4.23 Long-term gain drift and short-term shift due to change of operating conditions

The ANSI test specification does not mention the anode sensitivity at which the test is to be performed. For the stability figures to have meaning, the extreme values of the anode mean current must be quoted. To take account of typical photomultiplier applications, the test is usually performed between 300 nA and 30 nA, 1 μA and 100 nA or even between 10 μA and 100 nA.

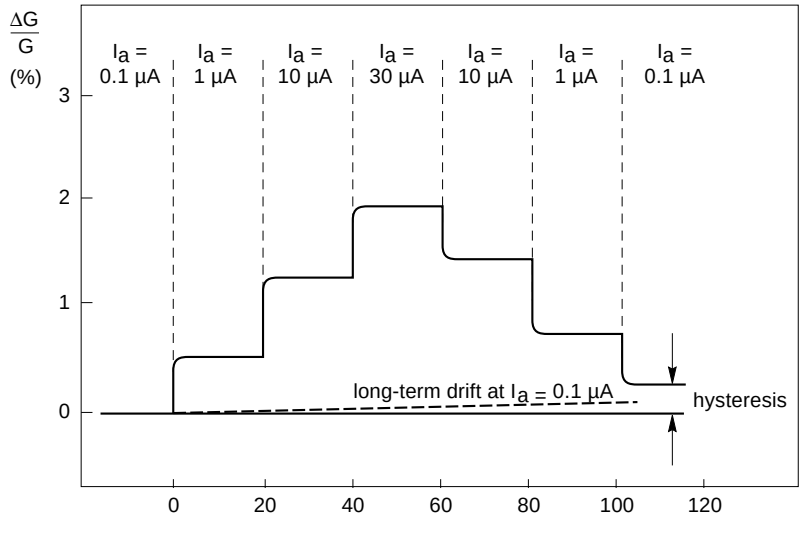
Another method, easier to set up, uses two independent LEDs that illuminate the photomultiplier simultaneously. One emits pulses of adjustable intensity and frequency, or simply a continuous flux of adjustable intensity, for setting the mean anode current to any desired level throughout the applicable range. The other emits pulses of fixed intensity and frequency. The mean height of the anode pulses due to these is a measure of the relative gain at the set level.

One cause of shift may be that charges on internal insulators (dynode spacers) are affected by the passage of electrons (scatter, for example), and that this in turn modifies the focusing between stages. In present-day photomultipliers, particularly those with venetian-blind dynodes, careful design of the electrode structure practically eliminates this effect. But even so, variations in gain due to variations in secondary

emission can still be observed, which suggests the influence of phenomena at the level of the emissive layer itself.



(a)



(b)

Fig.4.24 Gain shift of a photomultiplier with a single, short stabilization time constant: (a) as a function of mean anode current, (b) as a function of time. Measured 5 minutes and 15 minutes after illumination changes made every 20 minutes

Tubes with bialkali cathodes and CuBe venetian-blind or SbCs coated dynodes are usually considered the most stable in respect of shift, gain variations of less than 1% being common for anode current variations of ten to one (from 100 nA to 1 μA).

Figures 4.24 and 4.25 show that the gain does not return exactly to its original value when the flux does. This hysteresis reflects an interaction between long-term and short-term stability parameters. During prolonged operation, the higher the current the quicker the hysteresis tends to disappear. Here again, accelerated ageing at medium to high current has a useful effect.

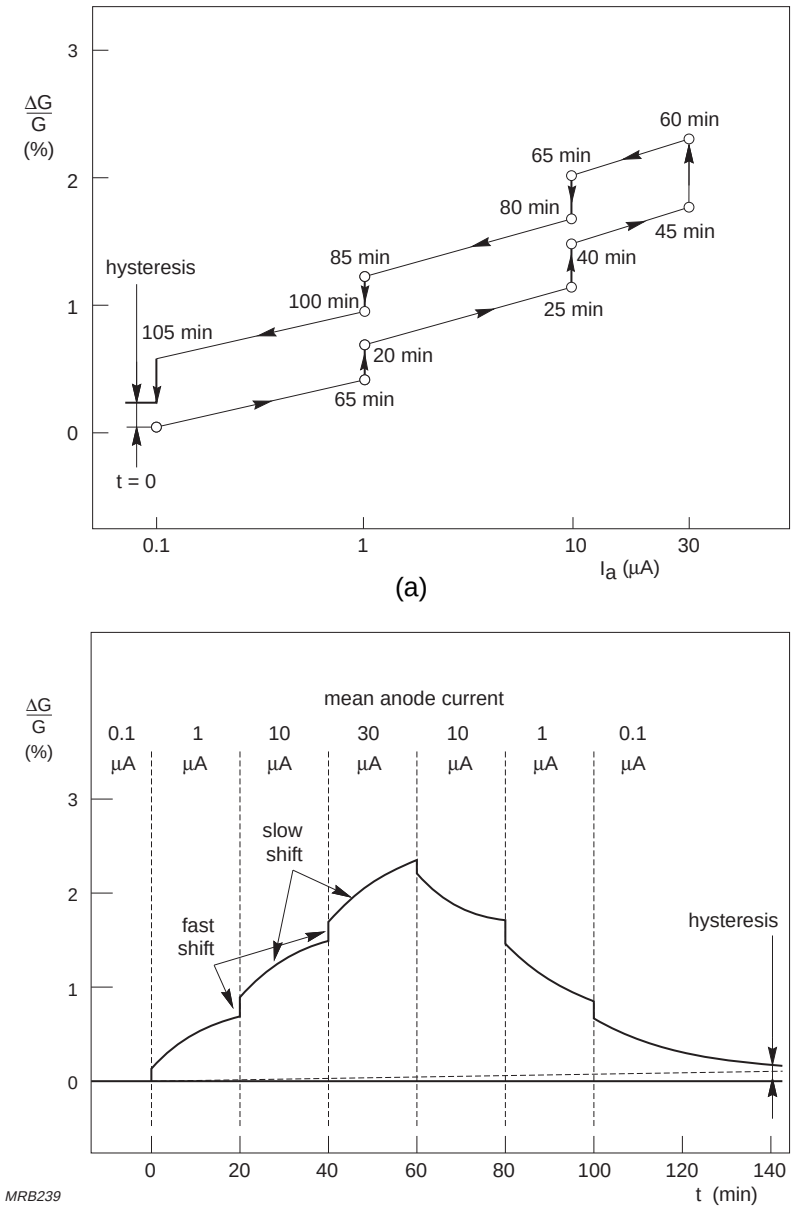


Fig.4.25 Gain change of a photomultiplier with one short and one long stabilization time constant: (a) as a function of mean anode current, (b) as a function of time. Measuring conditions as for Fig.4.24

4.7 Afterpulses

Afterpulses are spurious pulses that appear in the wake of true pulses. They can be observed on an oscilloscope during detection of very short flashes such as scintillation and laser pulses. As they are time-correlated with the true pulses, they are particularly inconvenient in chronometry and time spectrometry applications using coincidence techniques; in counting applications they often spuriously limit the number of true pulses that can be registered. Afterpulses have two main causes – luminous reactions and ionization of residual gases – which can be distinguished by the interval that separates the afterpulse from the true pulse.

4.7.1 Luminous reactions

When the electrodes are bombarded by electrons they emit photons. Although the luminous efficiency of these impacts is very low, in some tubes it is possible for light emitted even by the last stages to make its way back to the photocathode and so give rise to afterpulses. With respect to the true pulses from which they originate, theses are delayed by the sum of the light and electron pulse transit times, typically 20 to 100 ns.

4.7.2 Ionization of residual gases

Ionization of gas traces also gives rise to afterpulses. The traces may be of residual gases left in the tube after evacuation or desorbed by materials of its structure, or of helium that has migrated through the glass. Ionization afterpulses are delayed by a few hundred nanoseconds to as much as several microseconds with respect to the true pulses from which they originate.

Ionization afterpulses can be subdivided according to whether they originate in the electron-optical input system or the electron multiplier.

Primary photoelectrons in the electron-optical input system can generate positive ions that are accelerated towards the cathode and there give rise to emission of one or more secondary electrons. The transit time of the ions depends more on the input-system electric field and the mass of the ions than on the distance from the cathode at which they originate. The usual ions are H_2^+, and He^+, and CH_4^+; in a photo-multiplier with a 50 mm cathode, these have transit times of about 0.3 μs, 0.4 μs and 1 μs, respectively, at a cathode to first-dynode voltage of 300 V. The amplitude of the resulting afterpulses, relative to the true pulses, increases very rapidly with increasing cathode to first-dynode voltage.

The amount of helium entering the tube depends on the kind of glass used, its surface area, and the ambient helium partial pressure (typically 0.7 Pa ≈ 5×10^{-3} torr). For applications in which higher helium partial pressures are encountered, tubes with soft (lime) glass envelopes are preferred.

Ionization afterpulses originating in the electron multiplier come mainly from the last stages, where the electron current is largest. The relative amplitude of these pulses is independent of the cathode to first-dynode voltage but it does increase rapidly with gain. The delays observed for afterpulses of this type amount to a microsecond or more (e.g. about 3 μs for Cs^+ ions).

4.7.3 Afterpulse ratio

Each true pulse may be followed by one or more afterpulses (Fig.4.26). The relation between true pulses and afterpulses is quantified by an *afterpulse ratio* which may be stated as

- the ratio of the *number* of afterpulses to the number of true pulses, or
- the ratio of *charge* transferred by the afterpulses to charge transferred by the true pulses,

in either case expressed as a percentage.

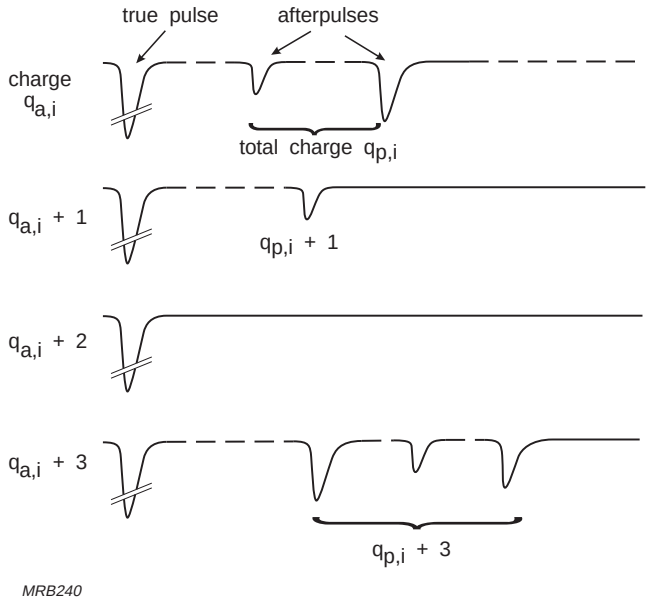


Fig.4.26 Examples of the development in number and charge of afterpulses

In specifying the afterpulse ratio, the interval in which afterpulses are counted following each true pulse must also be stated. The ratio can vary greatly depending on the measuring conditions and can be interpreted differently according to the application.

Afterpulse ratio stated as a charge ratio is given by the expression

$$F_{apq} = \frac{\sum q_{p,i}}{\sum q_{a,i}} \times 100 \qquad (4.36)$$

where $q_{a,i}$ is the charge transferred by true pulses and $q_{p,i}$ the charge transferred by afterpulses. The ratio is usually less than 1% and, as long as the gain is not too high, does not vary much with the number of true pulses or the amount of charge they contain. When the charge transferred by each true pulse is very small (e.g. in single-electron operation), that transferred by each afterpulse may be as large or even larger. However, as proportionally fewer true pulses are then followed by afterpulses, the charge ratio remains about the same.

The afterpulse ratio can be reduced by decreasing the cathode to first-dynode voltage as far as the application will permit, or by working with minimum gain, or both.

The effects of afterpulses can be minimized by using coincidence techniques, by blanking the photomultiplier for a set interval after each true pulse (§5.10), or simply by using electronics with sufficiently long dead time.

4.8 Environmental considerations

Environmental factors – chiefly temperature, magnetic fields, background radiation, and atmosphere – can affect the operation of a photomultiplier in varying degrees, temporarily or permanently. To a large extent the effects can be guarded against or compensated.

4.8.1 Temperature

By the nature of their photoemissive and secondary emissive materials, photo-multipliers are also sensitive to temperature variations. These affect three of the main characteristics:

- spectral response (the shape of the curve)
- dark current (the thermionic component)
- anode sensitivity and gain (secondary emission coefficients).

Changes in characteristics due to temperature variations within the permissible limits are usually reversible, though there may be some hysteresis that disappears only gradually.

Effects of temperature on the photoemission and secondary-emission surfaces are complex. They depend not only on the composition of the surfaces but also, to some extent, on the type of tube; and even between tubes of the same type there are appreciable differences. However, tendencies and average values can be identified.

In storage as well as use, photomultipliers must be kept within the temperature limits specified in their data sheets, usually −30 °C to 80 °C (50 °C for types with S1 cathodes). Beyond those limits effects such as sublimation of the cathode or stresses in the glass may occur. **Always consult the manufacturer before considering operation of a photomultiplier outside its published temperature limits**.

Effect on spectral sensitivity. The spectral sensitivity characteristic does not vary much with temperature. The greatest relative variation is usually observed close to the photoemission threshold. For a given application, therefore, it is advisable to choose a tube with a type of cathode that makes it possible to operate far from the threshold.

The dependence of monochromatic sensitivity on temperature differs from one type of cathode to another. For a given type it is defined (in percent per degree at 20 °C) by the ratio

$$\alpha_k = \frac{\Delta S_k}{S_{k(20°C)}\Delta\theta} \times 100 \qquad (4.37)$$

where ΔS_k is the change of cathode sensitivity observed over a temperature interval $\Delta\theta$. For commonly used photocathodes α_k usually varies continuously with wavelength, in some cases passing through the zero and changing sign.

The value of α_k and its variation with wavelength depend not only on the composition of the cathode but also on the structure of the tube, and therefore differ considerably from one type to another. The figures given below are merely indicative.

S11 cathode. Of all photocathodes, this type exhibits the widest variation of α_k as a function of wavelength; see Figs 4.27 and 4.28. The sign change of two of the specimens in Fig.4.28 at about 580 nm wavelength is supported by other observations but is not the rule; the temperature coefficient of many S11 cathodes keeps the same sign at least throughout the range of practical wavelengths (400 nm to 620 nm).

Bialkali SbKCs cathode. The temperature coefficient is very low in wavelength range 400 nm to 500 nm (Fig.4.29) where sensitivity is maximum, and may go to zero there or change sign. In Fig.4.30, where the coefficient is plotted for temperature intervals −20 °C to 20 °C, and 20 °C to 60 °C, it is nowhere greater than 0.15% in the range 400 nm to 500 nm. Furthermore, its variation with temperature is very small at short wavelengths.

S20 and S20R trialkali cathodes. These are characterized by a negative temperature coefficient throughout most of the useful spectrum. Figure 4.31 shows the variation of relative sensitivity as a function of temperature, of both types for several wavelengths; and Fig.4.32, the variation of temperature coefficient as a function of wavelength. Note that type S20R, whose response extends farther into the red, has a larger temperature coefficient than type S20.

Whatever the type of cathode, the residual change of sensitivity (hysteresis) following a temperature cycle is usually much less than 1% and can be disregarded.

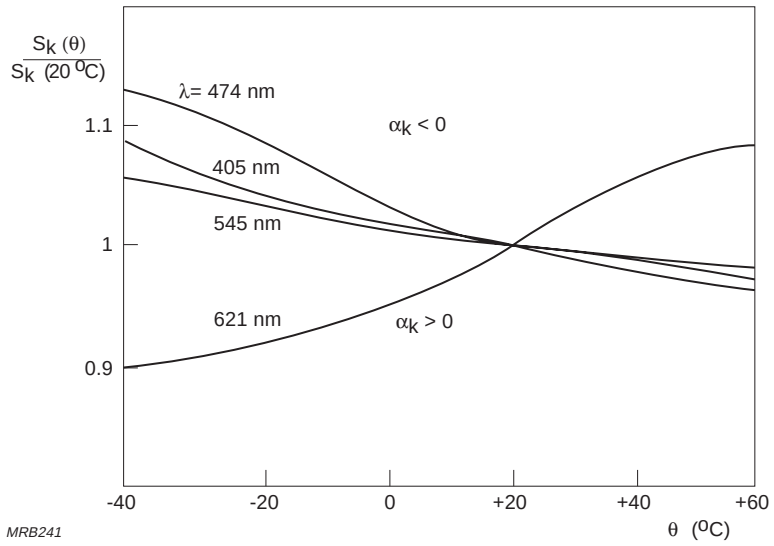


Fig.4.27 Relative variation of sensitivity of a type S11 cathode as a function of temperature, with wavelength as parameter

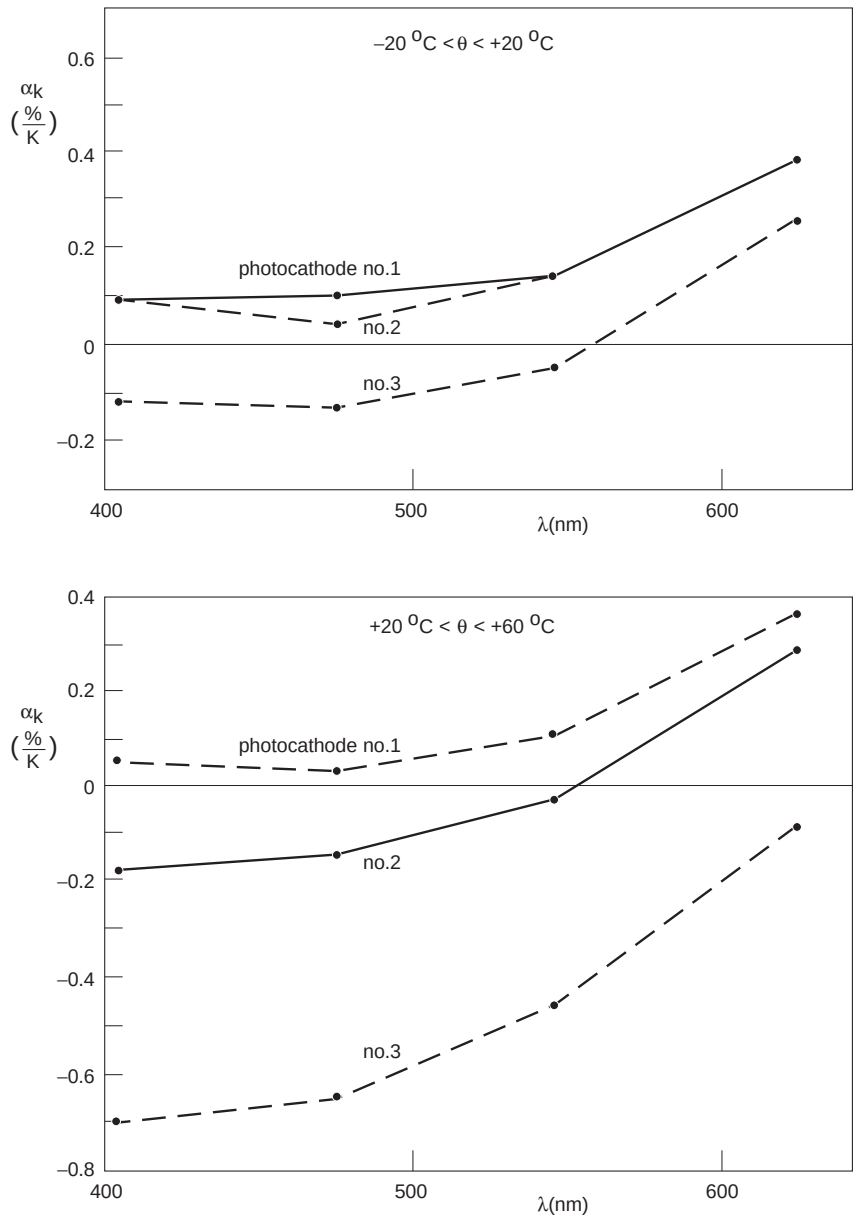


Fig.4.28 Examples of the variation of the temperature sensitivity coefficient α_k of type S11 cathodes as a function wavelength

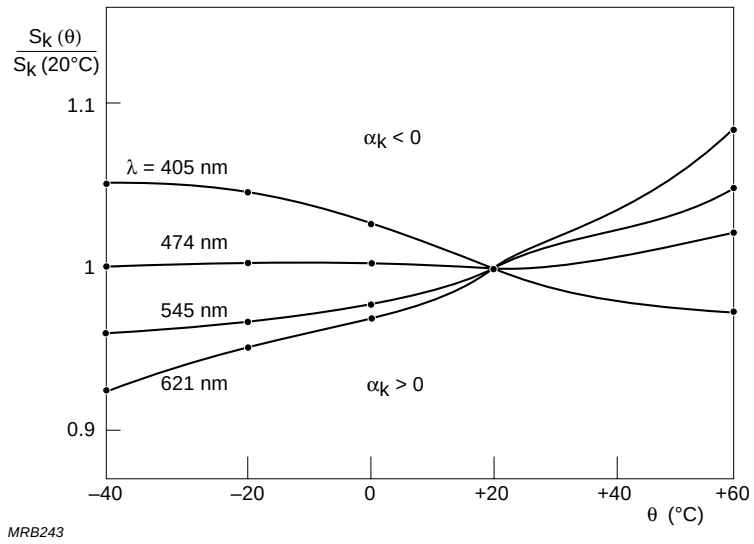


Fig.4.29 Relative variation of sensitivity of a SbKCs bialkali cathode as a function of temperature, with wavelength as parameter

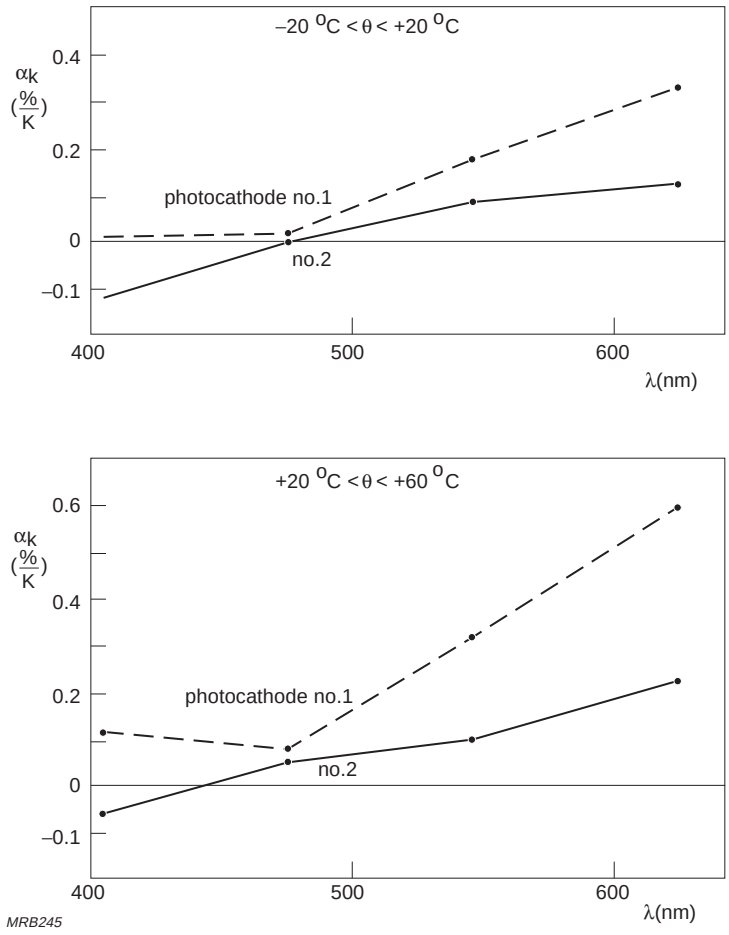


Fig.4.30 Examples of the variation of the temperature sensitivity coefficient α_k of SbKCs bialkali cathodes as a function of wavelength

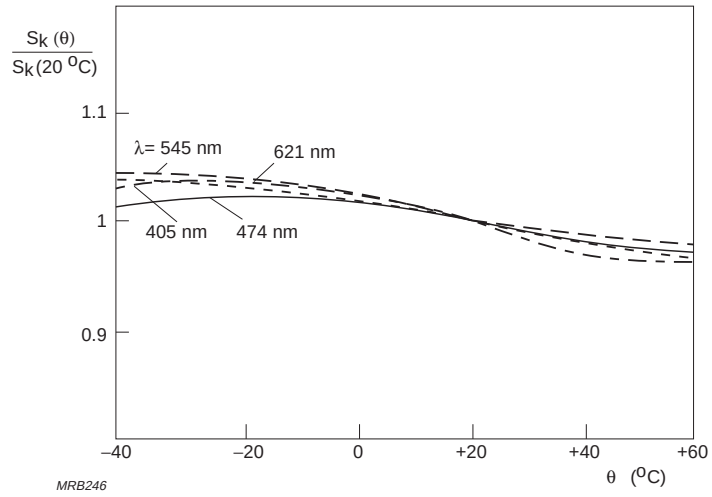


Fig.4.31 Relative variation of sensitivity of type S20 cathode as a function of temperature, with wavelength as parameter

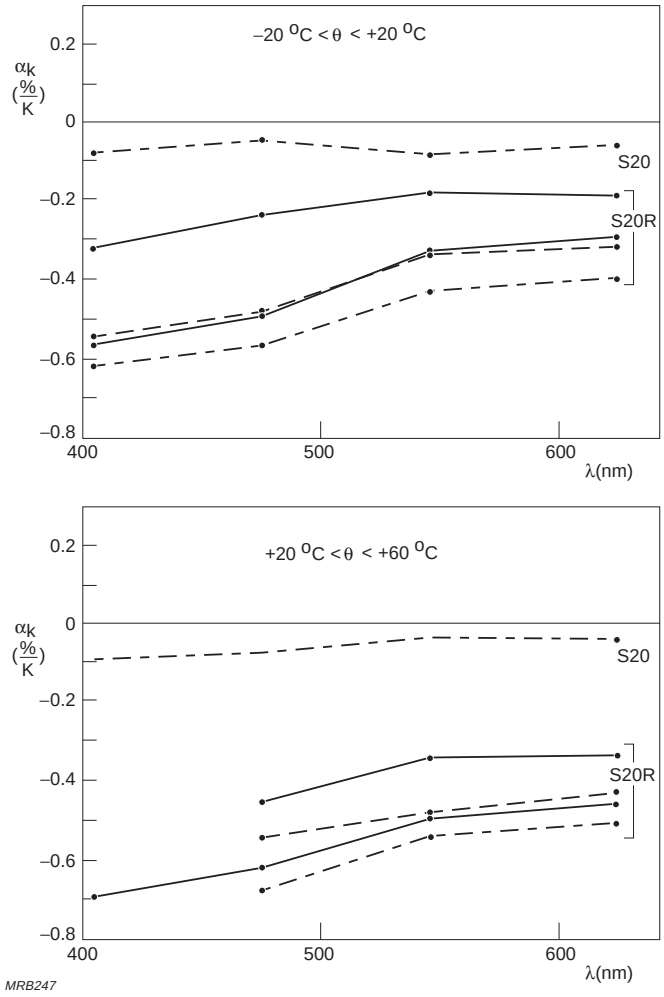


Fig.4.32 Examples of the variation of the temperature sensitivity coefficient α_k of type S20 and S20R cathodes as a function of temperature

Effect on cathode resistivity. The resistivity of photocathodes varies inversely with temperature. This can limit the minimum operating temperature, especially of bialkali SbKCs cathodes which, at room temperature, have a resistivity a hundred to a thousand times greater than that of S11 and S20 cathodes (Fig.4.33). The practical minimum for bialkali cathodes is −30 °C (if cathode current is more than 0.1 nA); and for all other types, about −100 °C.

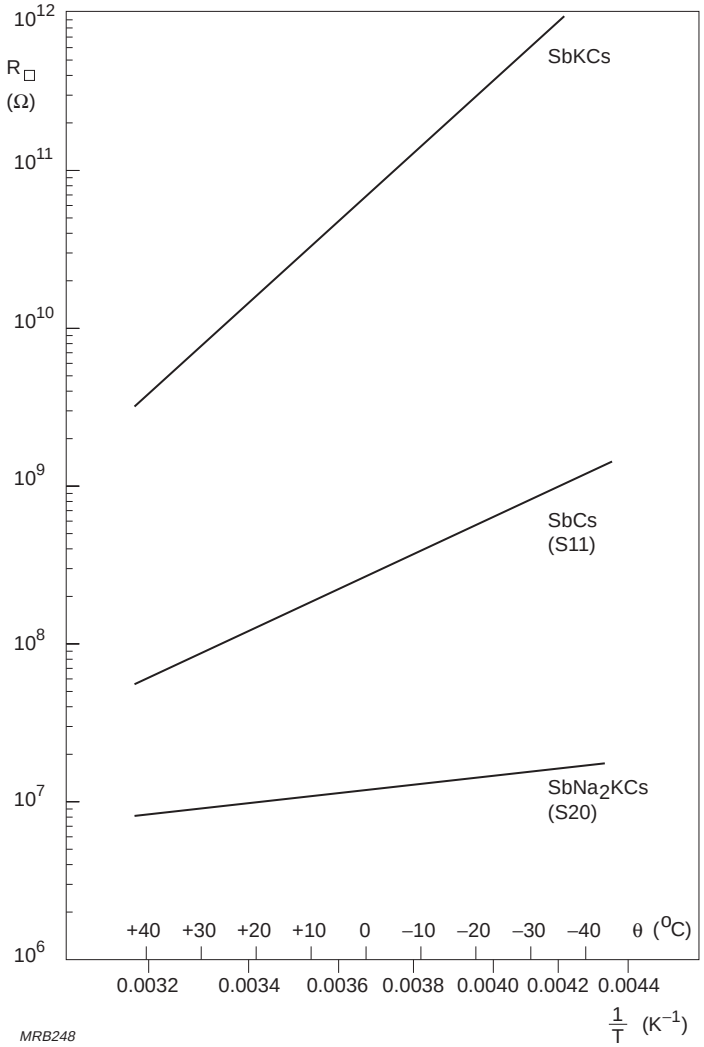


Fig.4.33 Surface resistivities of three photoemissive materials as functions of temperature

Effect on the dark current. The thermionic component of the dark current obeys Richardson's law, Eq.3.1; therefore, both dark current and dark pulse rate increase strongly with temperature. The rate of increase depends mainly on the cathode material but may differ considerably from one tube to another of the same type. This

is because some components of the dark current, such as thermionic emission, field emission, leakage current, etc., have different temperature coefficients and differ in relative importance from tube to tube. For the same reason, the dark current and dark pulse rate do not usually continue to decrease below a certain temperature, and in some cases may even increase, particularly in tubes with multialkali cathodes.

Table 4.1 Thermionic emission characteristics of common photocathodes

Type of cathode	Thermionic emission at 20 °C (A/cm^2)	Temperature rise for which thermionic current doubles (K)	Minimum operating temperature (°C)
AgOCs (S1)	10^{-13} – 10^{-11}	5 – 7	–100
SbCs (S11)	10^{-16} – 10^{-15}	6 – 15	–20
SbKCs	10^{-19} – 10^{-17}	4 – 5	–20
$SbNa_2KCs$ (S20)	10^{-19} – 10^{-15}	4	–40
$SbNa_2KCs$ (S20R)	10^{-17} – 10^{-15}	4	–40

Effect on gain and anode sensitivity. Dynode secondary emission also varies with temperature (though less so than cathode thermionic emission) and correspondingly affects gain. The temperature coefficient of gain is usually negative and depends not only upon the composition of the dynodes but also upon that of the cathode and, to some extent, the structure of the multiplier. For CuBe and AgMg dynodes the coefficient is about –0.1% per degree with bialkali and trialkali (S20) cathodes, and between –0.5% and –1% per degree with S11 cathodes. It is smaller in tubes with venetian-blind dynodes than in those with focusing dynodes.

Variations in anode sensitivity reflect variations in both cathode sensitivity and multiplier gain. At certain temperatures and wavelengths these may be equal and opposite, cancelling each other.

After a temperature cycle, anode sensitivity usually does not return exactly to its previous value. The hysteresis is mainly in the multiplier gain (cathode sensitivity hysteresis being negligible) and tends to disappear after long storage. Figure 4.34 shows the relative cathode sensitivity, gain, and anode sensitivity hysteresis of a tube with bialkali cathode and CuBe venetian-blind dynodes.

Gain hysteresis hardly varies from tube to tube. It is not cumulative and, in fact, tends to diminish after repeated temperature cycles. As an example, a tube with venetian-blind dynodes showed a −2% gain change after a first −40 °C to 60 °C to −40 °C cycle, but only −0.8% after a second cycle.

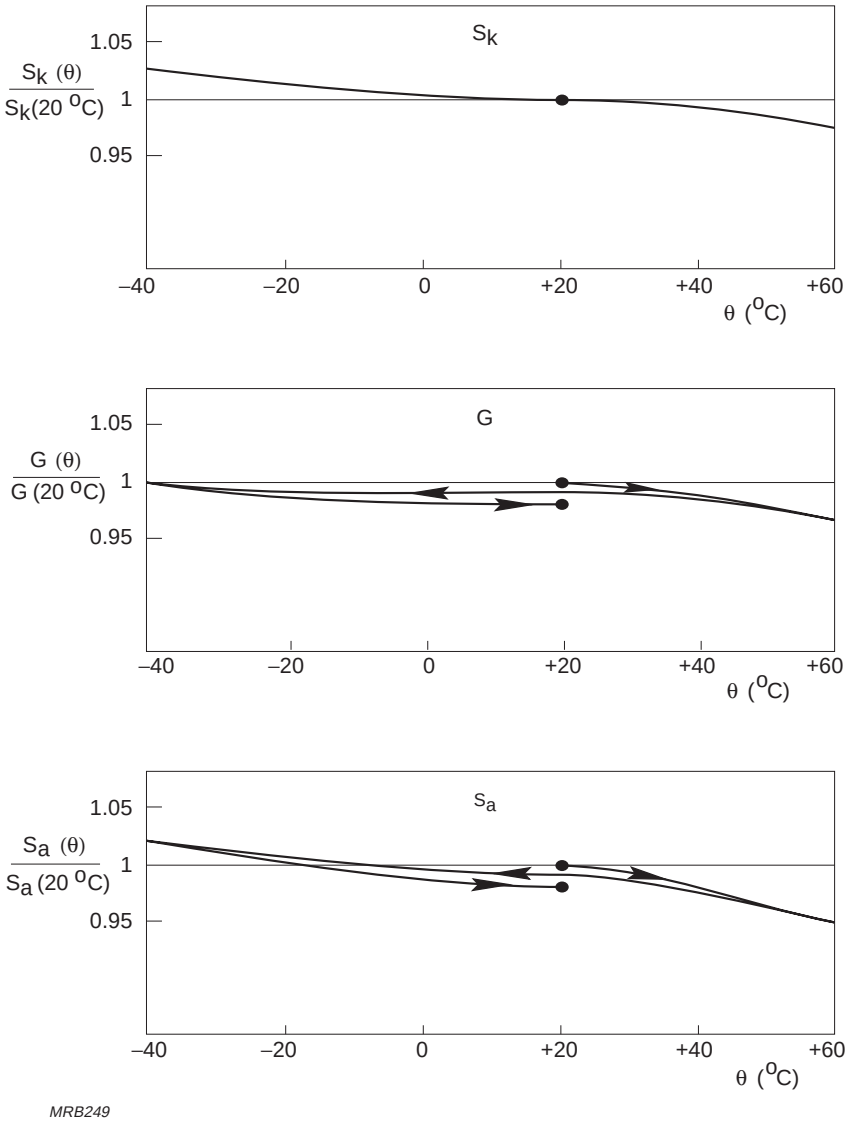


Fig.4.34 Temperature variation of cathode and anode sensitivity and gain of a tube with a bialkali cathode at a wavelength of 405 nm, showing hysteresis.

4.8.2 Magnetic fields

Magnetic fields even as weak as the earth's affect photomultiplier performances. This can be demonstrated by rotating a horizontally mounted tube about its main axis. The resulting variation of anode sensitivity is due to the varying effect of the earth's field on the electron trajectories, and the corresponding variation of collection efficiency in all stages. Highly focused tubes, in which the electron impact areas on the dynodes are small, are the most sensitive to magnetic effects; a transverse flux density of a few tenths of a millitesla can reduce gain by 50%. In a tube with venetian-blind dynodes, the field required to produce the same effect would be up to three times as large.

Magnetic influence is greatest in the electron-optical input system, where electron trajectories are longest. Increasing the voltage across the input system increases the energy of the electrons and decreases the sensitivity to magnetic fields. A photomultiplier tube is least sensitive to magnetic fields parallel to its axis.

Tubes with linear focusing dynodes are most sensitive to magnetic influence when the field is parallel to the dynodes (axis y in Fig.4.35). (For precautions against magnetic effects see §5.9.)

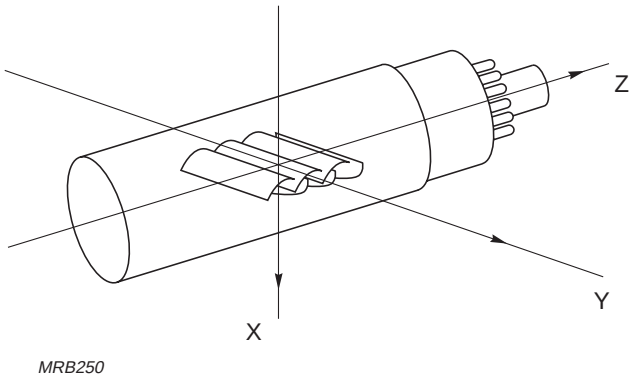
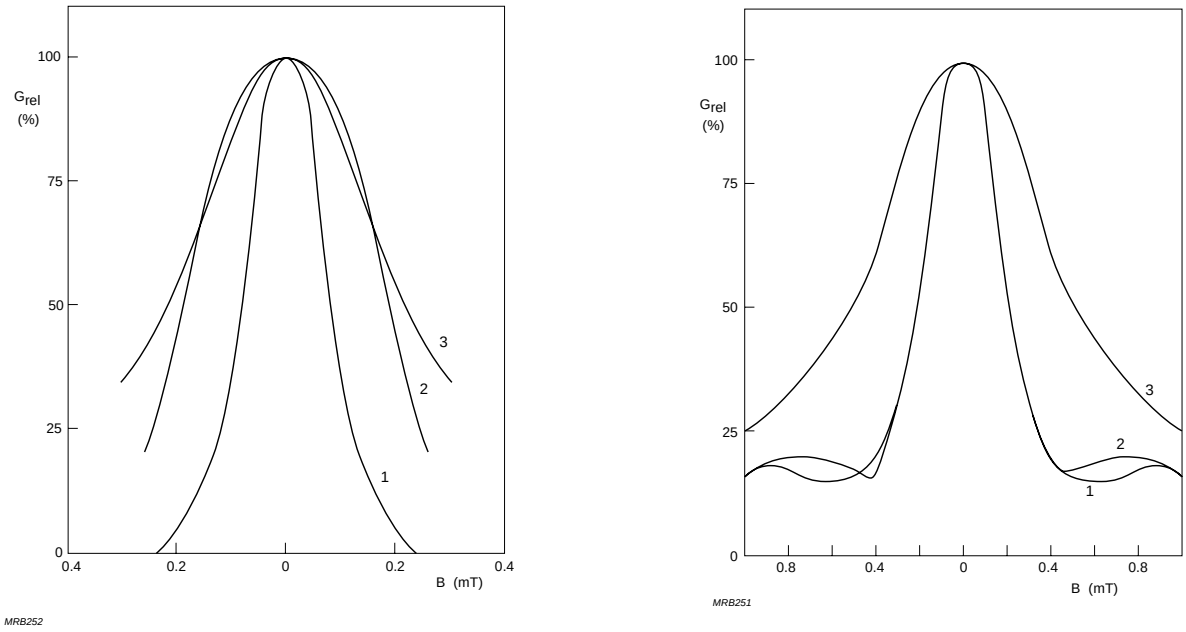


Fig.4.35 Axes used in measuring magnetic sensitivity

Magnetic sensitivity is measured relative to three perpendicular axes (Figs 4.35 and 4.36). Data sheets give either the measured sensitivity curves or the values of magnetic flux density parallel to each axis at which gain is halved. The data are for optimum operating conditions. Magnetic sensitivity is greater when electrode potentials are not optimum, as is the case when gain is deliberately decreased by defocusing a dynode or the accelerating electrode.

Strong fields may permanently magnetize some parts of a photomultiplier, lastingly affecting its performance. If that happens, the tube can be demagnetized with a coil producing a flux density of about 10 mT, at 50 Hz.



(a) (b)

Fig.4.36 Relative gain variation as a function of magnetic field: (a) for a tube with linear focusing dynodes, (b) for a tube with venetian-blind dynodes.
curve 1: field aligned with y-axis (Fig.4.35)
curve 2: field aligned with x-axis
curve 3: field aligned with z-axis

4.8.3 Radiation

Ionizing radiation can give rise to secondary effects such as dark current increase or, if the radiation is sufficiently intense, permanent loss of sensitivity and gain.

Dark current increase. Background radiation interacts with parts of the tube, mainly the glass, to produce light pulses that increase the dark current or dark pulse rate. The effect can be demonstrated by bringing a radioactive source close to the tube; for high-energy γ-radiation (more than 1 MeV) the dark current will increase considerably, owing to pair production in the glass.

Light pulses due to background radiation have two causes: scintillation and the Cherenkov effect.

Scintillation is due to interaction of low-energy α- and β-radiation with the glass of the envelope; such radiation may come from the surroundings or from the glass itself (e.g. from traces of ^{40}K). The scintillation efficiency is very low – about 10 photons

per MeV, not all of which reach the cathode – so the resulting pulses are mostly single-electron.

Cherenkov effect occurs when charged particles traverse a medium – e.g. the glass of the envelope – at more than the speed of light in that medium. Such particles may be due to cosmic radiation or may be generated in the medium itself by high-energy γ-radiation. The resulting anode pulses are of large amplitude, corresponding to 10 to 20 photoelectrons; moreover, they are followed by low-amplitude afterpulses due to phosphorescence of the glass excited by the ultraviolet content of the Cherenkov emission.

Cherenkov effect due to cosmic radiation can be a troublesome source of dark pulses in many applications; however, there are several ways of keeping it, or its influence on the measurement, to a minimum.

- *Thin input window*. The amplitude of light pulses due to the Cherenkov effect is proportional to the thickness of the glass: about 500 photons/cm in the wavelength range 300 nm to 600 nm. Input windows should therefore be as thin as possible to help keep the Cherenkov effect low.
- *Inverting the tube*. The direction of cosmic rays is usually more or less vertical. Positioning the tube with the input window down reduces their effect; Cherenkov radiation then tends to be emitted outwards, and only the fraction reflected back by the glass-air interface reaches the cathode. If the window is up, all of it reaches the cathode. (Mounting the tube horizontally gives an intermediate result. The glass cross-section presented to cosmic rays is much smaller but the distance the rays travel in the glass is much larger, so the count rate is lower but the pulse amplitude higher).
- *Window material*. Glass that limits transmission of ultraviolet light attenuates that component of the Cherenkov radiation. Glass windows give less Cherenkov-effect dark current than fused silica windows.
- *Anti-coincidence*. With an anti-coincidence system using two detectors it is possible to eliminate a large part of the dark current due to cosmic radiation. An anti-coincidence detector above the photomultiplier, or better still, completely surrounding it, makes it possible to cancel all pulses registered simultaneously.
- *High discrimination*. As pulses due to cosmic rays are of high amplitude, this characteristic can often be used as a basis for discrimination. When low-energy radiation is to be detected, for instance, the counting window can be centred on the relevant energy band so as to exclude high-amplitude pulses.

Materials in the surroundings may have a significant effect. Cosmic rays interact with them to produce showers of secondary particles that give rise to trains of noise pulses in the photomultiplier. A concrete roof a few centimetres thick is not enough to block cosmic radiation but, on the contrary, may be a troublesome source of noise due to secondary particles (spallation noise).

Gain and sensitivity loss. Permanent loss of gain and sensitivity is a serious risk only in very high radiation environments, such as parts of extraterrestrial space where electron flux can reach 10^{10} electrons per square centimetre per second. Under those conditions the emissive properties of the dynodes change and the input window darkens, affecting the transmission at shorter wavelengths. Lithium fluoride and lime glass windows are more sensitive than others to prolonged radiation. The photoemissive layer is relatively unaffected, probably because its absorption coefficient for ionizing radiation is low.

Permanent alteration of gain and sensitivity becomes noticeable only after exposure doses of about 10^4 rad.

4.8.4 Atmosphere

Humidity. Because of the high voltages used, operation in a damp atmosphere can lead to insulation problems. Condensation gives rise to leakage currents which increase the dark current. Local insulation breakdowns may also occur. Take particular care to avoid condensation on the glass, at the pins, and especially inside the plastic base. If moisture does get into the base, it will be necessary to drill a hole in the base key to enable it to escape.

Ambient pressure. Photomultipliers can operate satisfactorily at low ambient pressure, but precautions against flashover at the pins are necessary at pressures below 10 kPa (≈ 75 torr). *For operation or storage at high ambient pressure, consult the manufacturer; permissible pressures differ from type to type and are not usually given in the data sheets.*

Helium partial pressure. Glass is permeable to helium, the rate of penetration being proportional to the helium partial pressure. Of the glasses used in photomultipliers, lime glass (soft glass) is the least permeable; borosilicate glass and fused silica (hard glasses) are, respectively, about 100 and 1000 times more permeable. Helium intrusion increases the afterpulse factor and shortens the life: a tube with a fused silica window in a helium partial pressure of 100 kPa has a useful life of only a few days; this is a hazard to be guarded against in helium-cooled high-energy physics

experiments. The partial pressure of atmospheric helium is normally about 0.7 Pa, which is low enough to allow an average useful life of some ten years. Finally, note that helium penetration increases with temperature.

4.8.5 Mechanical stress

Like all electron tubes, photomultipliers should be protected against mechanical and temperature stress. Vibration or shock transmitted to the dynodes can modulate the gain (microphony). Especially robust types are available for use in hostile environments.

APPENDIX 4. SIGNAL TRANSFER IN LINEAR SYSTEMS

A4.1 Pulse and step responses

A transmission system having an input signal $U_1(t)$ and an output signal $U_2(t)$, is called linear if its response to the sum of two input signals acting together is equal to the sum of its responses to the two acting separately.

When a unit-area input pulse $U_1(t)$ of width Δt and amplitude $1/\Delta t$ is applied, the output pulse $U_2(t)$ is wider than Δt (Fig.A4.1). As Δt tends toward zero, $U_1(t)$ tends toward the delta function $\delta(t)$, such that

$$\lim_{\Delta t \to 0} \int_0^{\Delta t} \delta(t)\,dt = 1$$

The corresponding output pulse $U_2(t)$ then represents the *pulse response*, $R_\delta(t)$.

The *step response* $R_\varepsilon(t)$ is the response to a unit step $\varepsilon(t)$ such that

$\varepsilon(t) = 0$, for $t < 0$
$\varepsilon(t) = 1$, for $t \geq 0$

The unit step is related to the delta function by the expression:

$$\delta(t) = \frac{d}{dt}\,\varepsilon(t) \qquad \text{(A4.1)}$$

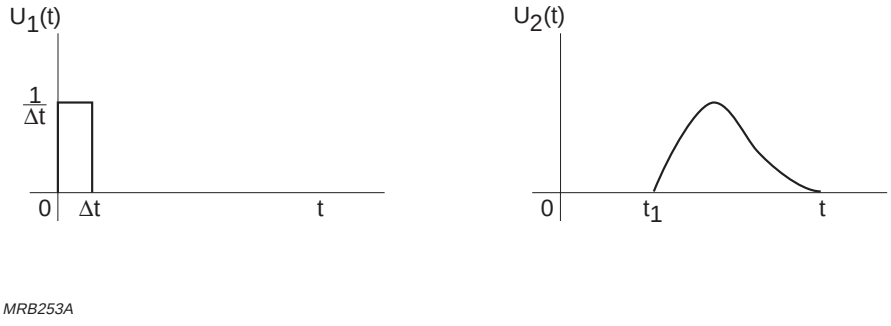


Fig.A4.1 Response to a rectangular pulse

A4.1.1 Superposition principle

An arbitrary input signal $U_1(t)$ can be represented by the superposition of very narrow adjacent pulses of width ds (Fig.A4.2). The response of the system to one such pulse occurring at an instant δ is (to within the second order) equal to the product of the pulse area and the pulse response of the system, shifted by the time s:

$$dU_2(t) = U_1(s)ds \;.\; R_\delta(t - s)$$

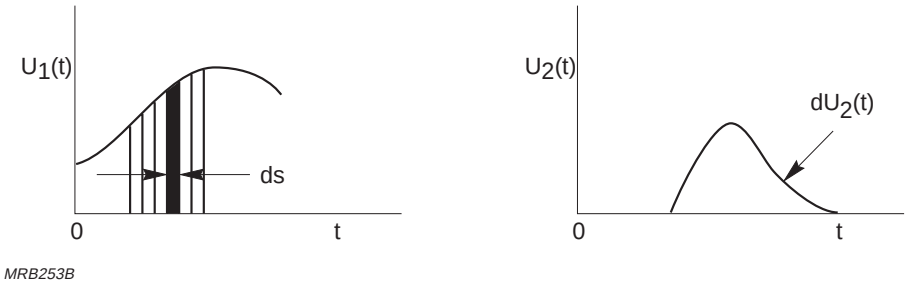


Fig.A4.2 Superposition of elemental pulses

Provided the system is linear, its response to the input signal $U_1(t)$ is the summation of the individual responses to the elementary pulses; thus,

$$U_2(t) = \sum_s dU_2(t)$$

or, letting ds tend toward zero,

$$U_2(t) = \int_{-\infty}^{+\infty} U_1(s)\; R_\delta(t - s)\; ds \qquad \text{(A4.2)}$$

This integral is known as the convolution product and is written

$$U_2(t) = U_1(t) * R_\delta(t) \qquad \text{(A4.3)}$$

Interchanging the variables s and t-s gives an equivalent form of Eq.A4.2,

$$U_2(t) = \int_{-\infty}^{+\infty} U_1(t - s)\; R_\delta(s)\; ds \qquad \text{(A4.4)}$$

i.e., the convolution product is commutative.

From Eqs A4.2, A4.3 and A4.4 it is possible to calculate the step response $R_\varepsilon(t)$ if the pulse response $R_\delta(t)$ is known:

$$R_\varepsilon(t) = \varepsilon(t) * R_\delta(t) \tag{A4.5}$$

This equation is useful when it is difficult to measure the step response directly. With photomultipliers, for example, it is easier to simulate repetitive delta-function pulse inputs than it is to simulate repetitive unit-step inputs.

Using the superposition principle and similar reasoning to the above, it can be shown that the output of a linear system can also be expressed as a function of the step response $R_\varepsilon(t)$ by the relation:

$$U_2(t) = \frac{d}{dt}\left[U_1(t) * R_\varepsilon(t)\right] \tag{A4.6}$$

This is equivalent to Eq.A4.3 and can be derived from it via Eqs A4.1 and A4.5.

A4.1.2 Rise time and FWHM

Equation A4.5 can be simplified by noting that the pulse response $R_\delta(t)$ and the step response $R_\varepsilon(t)$ are zero for $t \leq 0$; thus,

$$R_\varepsilon(t) = \int_0^t R_\delta(s)\, ds \tag{A4.7}$$

Several workers have proposed that the pulse response $R_\delta(t)$ of a photomultiplier can be accurately represented by a function of the type

$$R_\delta(t) = \varepsilon(t)\, \frac{\sqrt{m+1}}{m!\sigma_R} \left(\frac{\sqrt{m+1}}{\sigma_R}\, t\right)^m \exp\left(-\frac{\sqrt{m+1}}{\sigma_R}\, t\right) \tag{A4.8}$$

where $m = 8$ (solid line in Fig.A4.3(a)). The dashed line in Fig.A4.3(a) represents the step response $R_\varepsilon(t)$ derived from the pulse response $R_\delta(t)$ via Eq.A4.5 or A4.7. The step response is usually characterized in terms of the rise time $t_{r,\varepsilon}$ between 10% and 90% of maximum (points A and B). Figure A4.3(a) illustrates an interesting consequence of Eq.A4.7: the rise time between points A′ and B′, where the slope of the step response is half of its maximum, is equal to the full width at half maximum t_w of the pulse response. Points A′, B′ are located close to points A, B respectively.

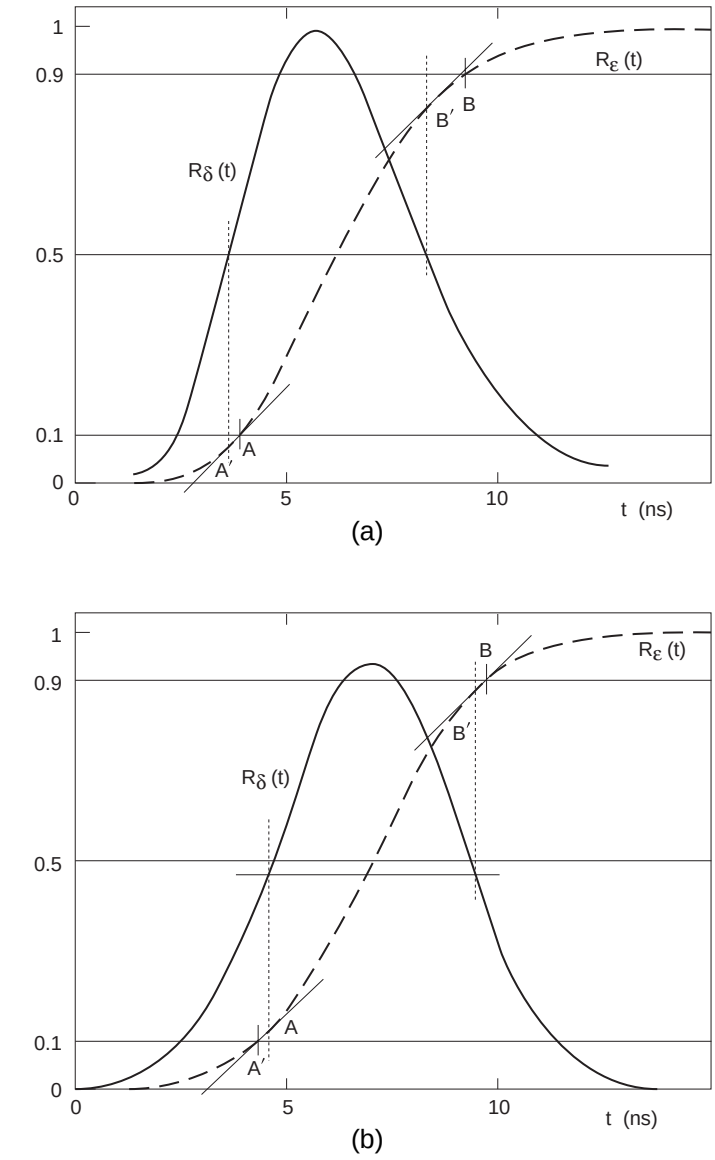


Fig.A4.3 Pulse response (a) of the type $at^m exp\ (-bt)$; (b) Gaussian pulse response

Another case to consider is the gaussian function (Fig.A4.3(b)) with standard deviation σ_R

$$R_\delta(t) = \varepsilon(t)\ \frac{1}{\sigma_R\ \sqrt{2\pi}}\ \exp\left\{-\ \frac{(t\ -\ t_t)^2}{2\sigma_R^2}\right\} \qquad \text{(A4.9)}$$

which is sometimes assumed for the pulse response of a photomultiplier, even though it is symmetrical; it is easier to manipulate than Eq.A4.8 and facilitates approximation. For the gaussian function points A′ and B′ are very close to points A and B, and the

step response rise time is nearly equal to the pulse response full width at half maximum:

$$t_{r,\varepsilon} \approx 1.11\ t_w \tag{A4.10}$$

Gaussian light pulse. The photomultiplier pulse response will be as given by Eq.A4.9, and the convolution of this with a gaussian light pulse is:

$$R^*(t) = \frac{1}{2\pi\sigma_L\sigma_R} \int_0^t \exp\left\{-\frac{(s - t_L)^2}{2\sigma_L^2}\right\} \exp\left\{-\frac{(t - t_t - s)^2}{2\sigma_R^2}\right\} ds \tag{A4.11}$$

or:

$$R^*(t) = \frac{f(t)}{2\pi\sigma_L\sigma_R} \exp\left\{-\frac{(t - t_t - t_L)^2}{2(\sigma_L^2 + \sigma_R^2)}\right\} \tag{A4.12}$$

It can be shown that if $(t_L + t_t)$ is greater than about twice the quantity $\sqrt{\sigma_L^2 + \sigma_R^2}$, the function f(t), which increases monotonically, quickly approaches the asymptote

$$\sqrt{\left(2\pi\ \frac{\sigma_L^2\sigma_R^2}{\sigma_L^2 + \sigma_R^2}\right)} \tag{A4.13}$$

When t is in the range $t_L + t_t \pm 2\sqrt{\sigma_L^2 + \sigma_R^2}$, the function f(t) approaches this asymptote to within less than 1%; $R^*(t)$ can then be written in the form

$$R^*(t) = \frac{1}{\sqrt{2\pi\left(\sigma_L^2 + \sigma_R^2\right)}} \exp\left\{-\frac{(t - t_t - t_L)^2}{2\left(\sigma_L^2 + \sigma_R^2\right)}\right\} \tag{A4.14}$$

which is a gaussian function with variance $\sigma_L^2 + \sigma_R^2$.

A4.2 Time resolution

Pulse transit time in a photomultiplier (Fig.A4.4) fluctuates from pulse to pulse. As the instant to which it is referred is the instant of illumination, however, its probability density distribution cannot be considered independently of the statistics of photon emission. Illumination that takes the form of, say, a delta-function pulse gives a different density distribution than illumination by a pulse of arbitrary width.

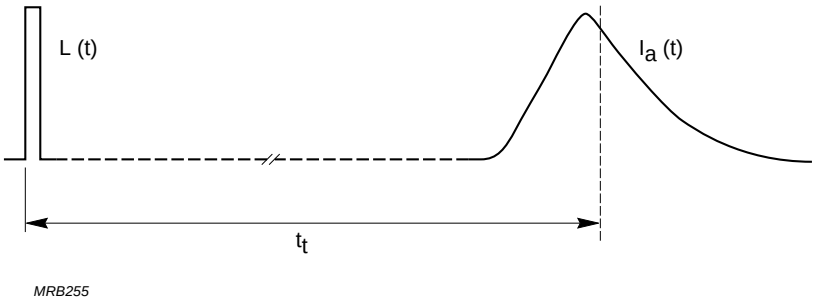


Fig.A4.4 Pulse transit time t_t.

A4.2.1 Delta-function light pulse

It is convenient to normalize the illumination function with respect to the mean number of photoelectrons per pulse, $\bar{n}_{k,i}$, thus

$$L(t) = \bar{n}_{k,i}\ \delta(t)$$

The pulse transit time can be defined as the interval separating the occurrence of the light pulse and a definable reference point on the ensuing current pulse – say, its centre of gravity. Let R(t) represent the probability density of the transit time t_t when the tube operates repeatedly under single-electron conditions (one photoelectron per pulse, Fig.A4.5); the mean transit time is $\bar{t}_t$ and the variance $\sigma^2_{tt(1)}$. Now assume that the cathode emits $n_{k,i} >> 1$ electrons per pulse, and that each of these gives rise to an identical anode pulse; the sum of the $n_{k,i}$ anode pulses is then the total response of the tube. The transit times of the individual pulses obey the R(t) probability density; however, they conform to it exactly only if their number is infinite. In actuality, therefore, the multi-electron pulse transit time fluctuates about a mean value that must be determined with reference to a large number of light pulses.

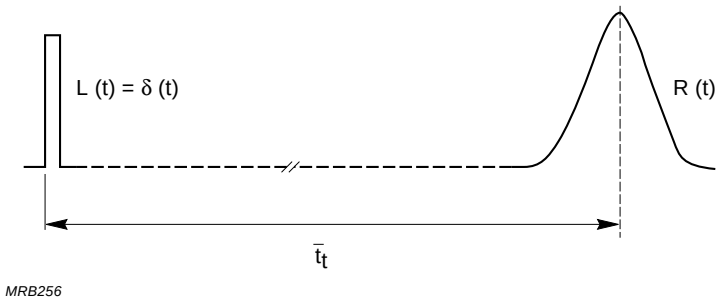


Fig.A4.5 Pulse transit-time probability distribution

To determine the variance of the fluctuations when each light pulse releases $n_{k,i}$ photoelectrons, consider the period during which the $n_{k,i}$ elementary responses arrive at the anode to be divided into infinitesimal intervals dt (Fig.A4.6). The fraction $dn_{k,i}$

of current pulses arriving at the anode during a single interval dt fluctuates from one interval to the next because the $n_{k,i}$ elementary pulses are randomly divided amongst the intervals; and also because the number $n_{k,i}$ itself fluctuates from one light pulse to the next. The first of these factors predominates; if the number $n_{k,i}$ is large enough, its fluctuations are negligible compared with those of the fraction $dn_{k,i}$ arriving during a single interval dt. To simplify calculation without affecting the generality of the results, we can therefore replace $n_{k,i}$ by its mean value $\bar{n}_{k,i}$ and $dn_{k,i}$ by $n_k(t)$. A variation $\Delta n_k(t)$ in this number shifts the centre of gravity of the transit time distribution by an amount

$$\Delta t_t = \frac{\Delta n_k(t)}{\bar{n}_{k,i}} (t - \bar{t}_t) \text{ if } \Delta n_k(t) << \bar{n}_{k,i}$$

or,

$$\sigma_{tt}^2 = \frac{\sigma_{nk}^2(t)}{\bar{n}_{k,i}^2} (t - \bar{t}_t)^2$$

where σ_{tt}^2 and $\sigma_{nk}^2(t)$ are the variances of t_t and $n_k(t)$.

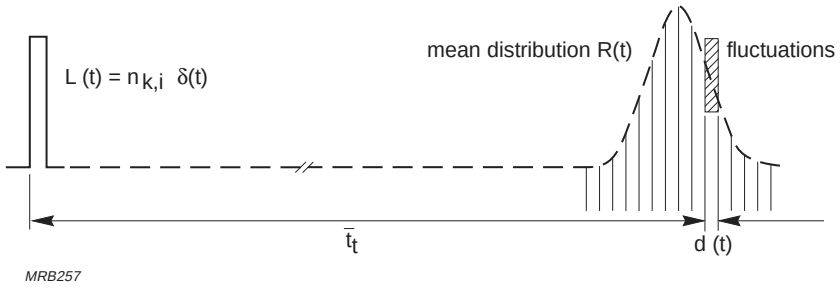


Fig.A4.6 Transit-time distribution of $n_{k,i}$ elementary response pulses

If we assume that the fluctuations in the number $n_k(t)$ during the intervals dt are absolutely independent of each other, then $n_k(t)$ obeys a Poisson law and we can write

$$\sigma_{nk}^2(t) = \bar{n}_{k,i} R(t)\,dt$$

whence

$$\sigma_{tt}^2(t) = \frac{R(t)\,dt}{\bar{n}_{k,i}} (t - \bar{t}_t)^2$$

or, by integrating over all the intervals dt:
If we assume R(t) to be gaussian with a mean value $\bar{t}_t$ and a variance $\sigma_{tt(1)}^2$

$$\sigma_{tt}^2(n_{k,i}) = \frac{1}{\bar{n}_{k,i}} \int_{-\infty}^{\infty} (t - \bar{t}_t)^2 \, R(t)\,dt \qquad \text{(A4.15)}$$

$$R(t) = \frac{1}{\sigma_{tt(1)}\sqrt{2\pi}} \exp\left\{ - \frac{(t - \bar{t}_t)^2}{2\sigma_{tt(1)}^2} \right\} \qquad \text{(A4.16)}$$

and Eq.A4.15 becomes:

$$\sigma_{tt}^2(n_{k,i}) = \frac{\sigma_{tt(1)}^2}{\bar{n}_{k,i}} \qquad \text{(A4.17)}$$

That is: when $n_{k,i}$ photoelectrons are emitted in response to a delta-function light pulse, the transit-time variance is equal to the single-electron pulse transit time variance divided by $n_{k,i}$.

A4.2.2 Arbitrary light pulse

For a light pulse of measurable width, transit time must be referred to definable points on both the light pulse and the ensuing anode pulse (Fig.A4.7).

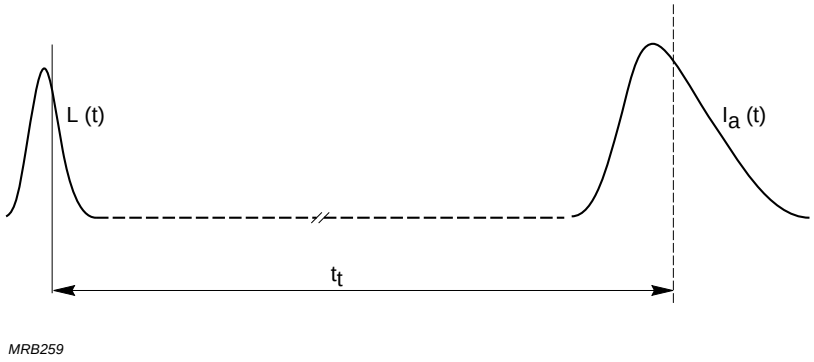


Fig.A4.7 Reference points for transit-time measurement must be clearly definable

Consider first the single-electron case. Relative to the chosen reference point, the instant of photoemission fluctuates from pulse to pulse with a density distribution corresponding to the illumination function L(t) which, in this case, may be normalized to unity.

$$\int_0^\infty L(t)\,dt = 1 \tag{A4.18}$$

As photoemission and electron multiplication are separate events in cascade, the density distribution of the times after the reference point on the light pulse when anode pulses occur is given by the convolution product

$$R^*(t) = L(t) * R(t) \tag{A4.19}$$

If L(t) and R(t) are approximately gaussian, with standard deviations σ_L and $\sigma_{tt(1)}$, then $R^*(t)$ is also gaussian and its variance is

$$\sigma_{tt(1)}^{*2} = \sigma_L^2 + \sigma_{tt(1)}^2 \tag{A4.20}$$

Now consider the case when $n_{k,i} >> 1$ photoelectrons are emitted per light pulse. As before, these can be assumed to give rise to $n_{k,i}$ elementary anode pulses the transit times of which will obey the density distribution of Eq.A4.19. The reasoning applied in the case of the delta-function light pulse can then be extended to the arbitrary light pulse by substituting $R^*(t)$ for R(t), and Eq.A4.15 can be rewritten

$$\sigma_{tt}^{*2}(n_{k,i}) = \frac{1}{\bar{n}_{k,i}} \int_{-\infty}^{\infty} (t - \bar{t}_t)^2 R^*(t)\,dt \tag{A4.21}$$

and the transit time variance is

$$\sigma_{tt}^{*2}(n_{k,i}) = \frac{\sigma_L^2 + \sigma_{tt(1)}^2}{\bar{n}_{k,i}} \tag{A4.22}$$

The foregoing reasoning disregards the effect of gain fluctuations on the shape of the anode pulse and, hence, the position of a reference point such as its centre of gravity. It has been shown that, for large values of $\bar{n}_{k,i}$ gain fluctuations with a relative variance v_G increase both the delta-pulse and the arbitrary-pulse transit time variances by a factor $(1 + v_G)$.