The second problem with PHA's is the long time (several microseconds) they require to increment the selected memory cell. The analyzer is "dead" (i.e., unable to respond to another pulse) during this time. At modest count rates ($\sim 10^4$/sec), a significant fraction of the elapsed clock time is dead time. This not only causes the apparent counting rate to be a nonlinear function of the true pulse rate (cf., Section 1.4.2.2.2); it also can grossly distort distributions from which a "background" has been subtracted. For example, suppose we want to find the distribution of pulses due only to light striking the cathode. We cannot observe this directly, for the dark pulses are always mixed in and measured along with the light pulses. Thus, if the number of dark pulses between heights h and $(h + dh)$ per unit time is $d(h)\,dh$, and the number of pulses in the same interval produced by a light source is $s(h)\,dh$, the distribution observed with the light on for a time t is

$$p(h) \cdot t_l = [s(h) + d(h)] \cdot t_l, \tag{1.5.1}$$

where $t_l < t$ is the "live time." Now if we also observe the dark pulses for a time t, we obtain the distribution $d(h) \cdot t_d$, where $t_d < t$ is the live time for the dark measurement. Because $t_l < t_d$, the difference

$$p(h) \cdot t_l - d(h) \cdot t_d = s(h) \cdot t_l - d(h) \cdot (t_d - t_l) \tag{1.5.2}$$

is clearly *not* proportional to $s(h)$, because of the $d(h)$ term. Because $t_d > t_l$, and $d(h)$ generally rises very rapidly at small h, the difference given by Eq. (1.5.2) can be driven to (or below!) zero at small h, creating a spurious maximum at moderate heights even if $s(h)$ is actually a monotonically decreasing function.

Clearly, it is necessary that all measurements be made for the same live time, not the same clock time.[101] Fortunately, most PHA's built today have live-timers. However, most simple pulse-counting systems used in photometry do not; so the above difficulty can still occur in measurements made with the discriminator-type systems described below.[101a]

The pulse-height distributions measured with an analyzer are sometimes called "differential pulse-height spectra." This is a misnomer, as they are (strictly speaking) frequency functions rather than spectra, and are "differential" only by contrast to the more properly named "integral" distributions, which include a broad interval of pulse height.

[101] C. W. McCutchen, *Phil. Mag.* (*8th ser.*) **2**, 113 (1957).
[101a] C. Smit and C. Th. J. Alkemade, *Appl. Sci. Res.* (*B*) **10**, 309 (1963).

The integral distributions are usually measured by counting all pulses greater than a threshold height h_t in a single-channel counter. The count rate is thus

$$P(h_t) = \int_{h_t}^{\infty} p(h)\,dh. \tag{1.5.3}$$

In statistical terms, $P(h)$ is a (complementary) distribution function, while $p(h)$ is a frequency function. [For example, if $p(h)$ were the Gaussian error frequency function, $P(h)$ would be the complementary error function, usually denoted by erfc.]

The device that rejects, or discriminates against, the pulses with $h < h_t$, is called a *discriminator*. Often two discriminators are used in anticoincidence, so that only pulses with $h_1 < h < h_2$ are counted; the system is then described as a "window" counter, with lower and upper thresholds h_1 and h_2. Most discriminators are designed for speed rather than for a sharp cutoff, so the thresholds are actually somewhat fuzzy.

If an integral distribution $P(h)$ is measured for many closely spaced values of h, the differential frequency function $p(h)$ can be recovered by numerical differentiation. However, it may not be the same as the $p(h)$ measured directly with an analyzer. Aside from the obvious problem of statistical (counting) errors, there are several possible explanations for this.

(a) There may be nonlinearity in one system or the other (see above).

(b) The faster time resolution of the discriminator system may resolve correlated pulse pairs that are counted as one larger pulse by the PHA. This is especially true of afterpulses due to dynode glow feedback to the PMT cathode; these usually come less than a hundred nanoseconds after the parent pulse.[44]

(c) At moderate light levels, pulse overlap[70] may be severe enough to distort the PHA measurements, but not those from the faster discriminator.

(d) Impedance mismatches (usually at cable connectors) can produce an exponentially decaying train of echo pulses after a large pulse. Such a train has a frequency function of the form[30,31] $p(h) = a/h$; it may be recorded by a fast discriminator system, but the echo pulses are not usually resolved from the parent pulse by the slower PHA.

Finally, one must remember that the shape of the distribution depends on electrode voltages,[35,39,83] on the area of the cathode illuminated,[35,39] and the wavelength of the illuminating light,[35,83] as well as on the temperature,[39,102] overall voltage, and other operating conditions. The wavelength dependence is due partly to an interaction between the wavelength dependence of the cathode sensitivity map and the position dependence of the pulse-height distribution; partly to the variation of mean electron energy with wavelength; and partly to the appearance of multiple-photoelectron emission (due to pair production) in the ultraviolet.[35] The shape of the contribution from cosmic rays depends on tube orientation.[30,41]

Because of these complications, there are several tests for internal consistency that should be made during measurements of pulse heights. (a) The gain of any amplifier used should be changed, if possible, by a known factor k. Then if we use subscripts 1 and 2 to denote measures made before and after the gain change, respectively, we should have $P_1(h) = P_2(kh)$ and $p_1(h) = p_2(kh)/k$, for all h† (this check is easy to make if k is a small integer). (b) The apparent light-pulse distribution $s(h)$ obtained by subtracting $d(h)$ from $p(h)$ should have a fixed shape, independent of light intensity. A convenient method of testing this is to find $s(h)$ for two light intensities differing by at least a factor of two, and plot $\log s(h)$ [or $\log S(h)$] against h; the two plots should coincide when displaced vertically. If only integral counts are obtained, the ratio $[S_1(h)/S_2(h)]$ should be independent of h (where now 1 and 2 refer to the two different light levels). The light source must be varied so as to keep its spectral content fixed, because of the dependence of $s(h)$ on color; varying the current through a lamp filament is thus *not* a suitable method unless a narrow spectral passband is used. (c) One should always make sure that very few small pulses are due to amplifier noise, pickup, line transients, etc. If many pulses are recorded with the PMT voltage off (or at a very low setting), the trouble must be eliminated, or higher dynode gain and higher discrimination settings must be used to get above the instrumental noise. This extraneous contribution to $d(h)$ rises precipitously as $h \to 0$,[30,104] so that a rather small increase in threshold reduces it from a horrendous level to an innocuous one.

[102] K. Haye, *J. Phys. Radium* **24**, 86 (1963).
[103] J. Rolfe and S. E. Moore, *Appl. Opt.* **9**, 63 (1970).
[104] G. C. Baldwin and S. I. Friedman, *Rev. Sci. Instrum.* **36**, 16 (1965).

† To a fair approximation,[30,31] a change in dynode gain produces a similar effect. Thus, an increase in tube voltage[98] or a drop in temperature[49] causes $P(h)$ to become stretched out toward higher h; but $p(h)$ becomes squashed down as well as stretched out, because the pulses are spread over more channels in the PHA.[30,103]

1.5.1.2. Interpretation. The pulse heights produced by various mechanisms were mentioned in Section 1.4.2, and are discussed at length in Ref. 30 and elsewhere; so only a brief review will be given here. We begin at the largest sizes and work down. Pulse heights are referred to mean light-pulse height

$$\bar{h}_s = \int_0^\infty h\, s(h)\, dh \Big/ \int_0^\infty s(h)\, dh. \qquad (1.5.4)$$

Most pulses with normalized heights greater than about 10 are Cerenkov pulses due to cosmic rays and radioactive decays. These occur only in end-window tubes with high-efficiency cathodes; they should not occur with either S-1 or opaque-backed cathodes. The CR component ($\sim$1–2/cm^2/minute) varies with orientation: very large ($>100\bar{h}_s$) pulses can occur in a horizontal tube; more, but smaller ones, occur in a tube facing up; and they are smallest in size and number in a face-down tube.[41] Occasionally, large power-line transients can cause very large pulses (e.g., when large electric motors are started up).

Pulses with heights greater than 2 or 3 and less than 10 or 20 are due to gas ions striking the cathode; a typical height[40] is about 4. They occur $\sim$1 μsec after light pulses, a few percent of the time; they also occur in the dark. They appear to make up most of the dark current in many tubes (except those with S-1 cathodes); they account for both the temperature-dependent and the time-dependent components of dark current.[30] Both gas ions and dynode glow should produce a contribution to $d(h)$ that is proportional to h^{-2} at smaller pulse heights, as is observed in many tubes.[30,39] Typically, some 10^2 ion pulses appear, at room temperature, per square centimeter of cathode area per second.

The expected distribution $s(h)$ due to single photoelectrons ($h \approx 1$) was discussed in Section 1.4.1.3.3. The actual distribution is more smeared out, partly because of dynode nonuniformity,[29] and partly because of small afterpulses produced in the dynode system,[30] which contribute a $1/h^2$ component at $h \ll 1$.

The pulse-height distribution due to thermionic or field emission from the cathode is similar to that caused by photoelectrons. It is not identical, however, owing to a different spatial distribution (most of the thermionic emission probably comes from local hot spots that do not contribute strongly to photoemission) and a lower initial electron energy ($\sim kT$, or $\sim$0.03 eV at 300 K), cf., the dependence of $s(h)$ on wavelength. Spontaneous emission from dynodes[30] contributes a $1/h$ component to $d(h)$; there is some indication that exposing a tube to a uniform gamma-ray flux produces such a component by direct stimulation of dynodes. Echo pulses[31] should also have a $1/h$ distribution.

At small ($h \ll 1$) pulse heights, the PMT should contribute only the $1/h$ and $1/h^2$ components mentioned above. However, the smallest pulses actually observed are usually due to amplifier noise. The simplest model for such noise is white, Gaussian noise, such as Johnson noise in the anode load resistor or in the input resistance of the preamplifier. (Johnson noise is just the Rayleigh–Jeans tail of blackbody radiation, observed at radio and audio frequencies. A resistor is an absorbing—hence "black"—body at these frequencies.) In fact, this noise is what makes photomultipliers necessary in the first place: without the very large gain provided by the dynodes, the individual photoelectrons would be buried in the thermal noise of the circuitry.

Obviously, circuit-noise pulses are important if they are comparable to other noise pulses in height and frequency ($\sim 10^3$/sec at room temperature, ~ 1/sec in cooled tubes). With a 1-MHz bandwidth, we are concerned about pulses that occur with probability 10^{-3} at most; at 100 MHZ, only 10^{-8} of the random noise pulses can be important. Thus if we had pure Gaussian noise, we would be concerned about events more than 3 standard deviations from the mean; but this region is notoriously non-Gaussian in real systems. Thus the noise of concern is likely to be due to rare events unrelated to thermal noise.

1.5.1.3. Time Distribution. Here we are concerned with the temporal randomness or correlation of pulses. We must treat light and dark pulses separately; and we should consider the two-dimensional correlation in a plot of pulse height against time of occurrences.

If photons enter the cathode "at random" (i.e., if the number of photons in a given time interval obeys Poisson statistics), and if photoelectrons are produced "at random" (i.e., each photon has the same, fixed probability of producing an emitted photoelectron), it can be shown[105] that the time distribution of photoelectrons also obeys Poisson statistics. Actually, because photons obey Bose–Einstein statistics, they tend to clump together instead of arriving independently and at random. The clumping occurs on a time-scale comparable to the coherence time of the light; for "white" light this is very short, but for quasi-monochromatic (e.g., laser) light it can exceed the resolution time of pulse-counting systems, so an excess of close pulse pairs can be seen. The clumping can also be detected as coincident (i.e., temporally unresolved)

[105] D. L. Fried, *Appl. Opt.* **4**, 79 (1965).

pairs of photoelectrons produced in two detectors illuminated by the same coherent wave. In laboratory experiments, the coincidences are usually detected by nuclear pulse-counting methods, and the wave is divided into two coherent parts by a beam-splitter; in the Hanbury-Brown "intensity-correlation" interferometer,[106] coincidences are detected by radio-frequency correlation techniques employing analog multipliers, and wave-front division is used to obtain two coherent beams. In both cases, the same principles apply. As this field has been reviewed recently,[107] no further mention will be made here. In any case, the deviations from Poisson statistics are quite small in ordinary photometric work.[48,49,98]

Although the light pulses are randomly distributed in time, to a good approximation, the dark pulses sometimes are not, especially in cooled tubes.[46a,48] Because each large radioactive or cosmic-ray event is followed by a host of small pulses,[43–46a] and because such events dominate the dark noise of cooled alkali-antimonide tubes,[41,47] the correlated clumps of dark pulses (especially in tubes with broad spectral response[46,48]) are probably due to such events. Presumably, the small daughter events are mostly single-electron photoresponses to induced window phosphorescence.[43,44] Unfortunately, no detailed study of the temporal and height dependence of these correlated pulses has yet been made.

The occurrence of clumps of pulses with a characteristic time scale τ should produce excess (nonwhite) noise at frequencies lower than $\sim 1/\tau$. PMT frequency-spectra have been studied by several authors.[98,108,109,109a] No excess noise has been reported at frequencies greater than 0.1 Hz ($\tau \approx 10$ sec), although one should expect some in the dark noise of cooled tubes with "nonstatistical" (i.e., correlated) dark noise[48] at frequencies up to perhaps 10^2–10^3 Hz. Whether the excess fluctuations observed for $\tau \geq 20$ sec should be regarded as "flicker noise" or merely a slight gain instability[98] is perhaps a matter of terminology. The "excess noise" level here corresponds to gain variations on the order of 10^{-3} with a time scale of a minute or so; small temperature fluctuations (~ 0.1°C) could play a part, as well as cesium migration and other drift or fatigue effects. Similar $1/f$ noise has been observed by Smit *et al.*[109a] in the dark count rate of an uncooled 1P28.

[106] R. Hanbury-Brown, J. Davis, L. R. Allen, and J. M. Rome, *M.N.R.A.S.* **137**, 375, 393 (1967) and references therein.

[107] C. L. Mehta, *Progr. Opt.* **8**, 375 (1970).

[108] A. H. Mikesell, *Publ. U. S. Naval Obs., 2nd Ser.* **17**, 143 (1955).

[109] R. C. Schwantes, H. J. Hannam, and A. van der Ziel, *J. Appl. Phys.* **27**, 573 (1956).

[109a] C. Smit, C. Th. J. Alkemade, and W. F. Muntjewerff, *Physica* **29**, 41 (1963).

As is well known,[110] low-frequency noise can be avoided by "chopping." This usually means a rapid symmetrical alternation between the unknown signal source being measured, and a reference (usually, dark) signal. Many solid state detectors and amplifiers have considerable $1/f$ noise in the low audio-frequency range, so that chopping frequencies near 1 kHz are common. The effect of chopping is to frequency-shift the measurements from a region near 0 Hz (dc) to a region around the chopping frequency, where the $1/f$ noise is negligible. The chopped signal is then demodulated by synchronous rectification, which restores the dc baseband.

Chopping is advantageous in situations that are dominated by additive low-frequency noise. This is true of most photoconductive detectors, for example, which are nearly always operated with symmetrical, square-wave chopping between signal and dark states. In principle, multiplicative noise (e.g., gain variations) can also be eliminated by chopping, if the demodulation process takes ratios instead of differences, and if the reference is a standard source instead of darkness. However, the difficulty of establishing a constant reference source, and the complexity of taking ratios, have usually restricted this technique to a few specialized instruments, such as densitometers.

As $1/f$ noise is negligible in photomultipliers down to 0.1 Hz or so, the rapid chopping that is so useful with other detectors is not advantageous.[98] Instead, occasional dark or background readings, usually made by hand, may suffice to establish reference levels. This can be regarded as a very low-frequency chopping, with a very asymmetrical duty cycle.

Finally, in many applications a photomultiplier is not limited by either additive (dark) noise or multiplicative noise, but by photon noise. In this case, symmetrical chopping is disadvantageous, because it degrades the photoelectron statistics by discarding half of the events. Of course, one still has to establish zero (dark) and gain (reference) levels; but only a small fraction of the total time may be needed for this purpose. We shall take a closer look at photon-noise-limited detection in the next section.

1.5.2. Detection: the Signal/Noise Ratio

When the light falling on a detector is strong enough that statistical fluctuations in the number of detected photons are larger than the dark noise, the detector is described as *photon-noise limited*. Also, it is often

[110] R. H. Dicke, *Rev. Sci. Instrum.* **17**, 268 (1947).

possible to distinguish a photon-noise component, even when dark noise is not negligible. Thus we need a concise way of describing the photon noise.

For comparison, let us consider an idealized, perfect, photon detector: every photon received produces an output signal, and the output is zero in the dark. Clearly, the signal/noise ratio depends only on the statistical fluctuations in the incident beam of photons. If the beam is weak enough, photon correlations can be neglected and Poisson statistics are a good approximation. Thus if g is the average number of photons observed per unit time, the probability of observing k photons in a particular unit-time interval is given by Eq. (1.4.10), and the standard deviation (i.e., noise) of a series of such measurements is $g^{1/2}$. The ratio of signal g to noise $g^{1/2}$ is thus $g^{1/2}$. For example, to obtain a standard error of 1%, we must count photons until 10^4 of them have been detected, on the average.

Now let us make the slightly more realistic assumption that photons are detected with some reduced probability, $q < 1$; we can call q the *quantum efficiency* of this idealized detector. The average number of detected photons per unit time is $N = g \cdot q$; its standard deviation is (again assuming Poisson statistics) $N^{1/2}$, so the signal/noise ratio for the less efficient detector is $N^{1/2} = g^{1/2}q^{1/2}$. Thus the signal/noise ratio is reduced by the square root of the quantum efficiency. For example, a detector with an efficiency of $\frac{1}{4}$ gives half the signal/noise ratio of a perfect detector.

Now suppose we have a real photon-noise-limited detector, with a signal/noise ratio half that of the perfect detector. It is natural to describe this real detector as having an effective quantum efficiency of $\frac{1}{4}$. As we deal only with the signal/noise ratio, it does not matter whether the detector is operated as a photon counter, or whether it has an analog output. Thus the concept of an equivalent (or, "detective") quantum efficiency can be applied to dc as well as pulse-counting measurements with photomultipliers.

Obviously, the detective quantum efficiency (DQE) of a photomultiplier cannot be higher than the quantum efficiency of the tube's cathode, at a given wavelength. In fact, it is always less; for some electrons (typically $\sim$20%) are lost in the dynodes, and not all photoelectrons are equally weighted[111] in the detection process. The weighting depends on the height of each anode pulse, and on the method of detection. As we shall

[111] W. A. Baum, *in* "Astronomical Techniques" (W. A. Hiltner, ed.), Chapter 1. Univ. of Chicago Press, Chicago, Illinois, 1962.

see, the DQE is typically about half the cathode quantum efficiency in photon-noise-limited PMT's.

1.5.2.1. DC vs. Pulse Counting†

1.5.2.1.1. General Relations. We now investigate the effects of the weighting functions corresponding to different detection schemes. In pulse counting, all pulses between two heights (or all above a threshold level) are counted equally (weight 1) and all others are ignored (weight 0). In dc or charge integration, each pulse is weighted by its height (charge). Each weighting function gives different results for signal and dark pulses, because of the different distribution with height. Thus, it is convenient to consider the two extreme cases of strong signal (where dark pulses are negligible) and weak signal (where the signal pulses are almost negligible), for each weighting function. Both of these limits are unrealistic, since in practice the dark pulses always dominate the small pulse end of the distribution, and signal pulses must make an appreciable contribution or the time required to make an observation is unreasonably long. Nevertheless, both cases are helpful in understanding practical situations.

In the following analysis, lower-case letters are used to denote pulse-height distributions, and upper-case symbols denote counting rates, or measurements integrated over a range of pulse heights (e.g., anode currents). The letters S, D, and P indicate signal (light alone), dark, and signal plus dark (i.e., the quantity actually measured with light on the tube), respectively. It is important to realize that the quantities analyzed are all *rates*, such as counts per unit time, or charge per unit time (current). If a quantity Q is measured for a time t, the total number of pulses observed will be proportional to Qt, and its standard deviation will be proportional to $(Qt)^{1/2}$; thus the statistical estimate (mean $\pm$ standard deviation) of the rate Q will be the observed total $[Qt \pm (\alpha Qt)^{1/2}]$ divided by t, or $[Q \pm (\alpha Q/t)^{1/2}]$. (The factor α absorbs the proportionality factors.)

If it is necessary to distinguish between rates estimated from a unit-time observation and rates estimated from an observation of duration t, the time interval (1 or t) will be used as a subscript. Obviously, the mean values are the same in either case, so this notation is applied only to estimated standard errors. In the above example, the expected error in Q estimated from an observation of unit time is $\sigma_{Q,1} = (\alpha Q)^{1/2}$, and the

† Most of the material in this section is taken from Ref. 30.

error in an estimate of the rate Q obtained from an observation of duration t is $\sigma_{Q,t} = (\alpha Q/t)^{1/2}$. Notice that the error in the estimated rate decreases with the square root of the observation time; to halve the error, we must observe four times as long.

In general we observe a pulse distribution

$$p(h) = s(h) + d(h), \tag{1.5.5}$$

where s and d are, respectively, the signal and dark pulse-height distributions, and $s(h)$ is proportional to the amount of light falling on the tube per unit time. If our weighting function is $w(h)$, the quantity measured with the light on for a time t_{L} is

$$P \times t_{\mathrm{L}} = t_{\mathrm{L}} \times \int_0^\infty w(h) \times p(h)\, dh = (S + D) \times t_{\mathrm{L}}, \tag{1.5.6}$$

where

$$S \equiv \int_0^\infty w(h)\, s(h)\, dh, \tag{1.5.7}$$

and

$$D \equiv \int_0^\infty w(h)\, d(h)\, dh \tag{1.5.8}$$

is the rate measured with the light off. (We assume that the parent distributions are independent of time, so that P, S, and D refer to unit times.)

Now consider the statistical fluctuation of a measurement of duration t. If there are on the average $[p(h) \times dh \times t]$ pulses with heights between h and $h + dh$, the variance in this quantity is $\sigma_p{}^2 \times t = [p(h) \times dh \times t]$. If there is no correlation between pulses of different heights, the variance in the measured quantity is

$$t\sigma_{P,1}^2 = t \int \{\partial P/\partial[p(h)]\}^2 \times \sigma_p{}^2 = t \int_0^\infty \{w(h)\}^2 \times p(h)\, dh, \tag{1.5.9}$$

where the subscript 1 refers to unit time.

If the total observation time is t, of which the light is on for a fraction f and off for a fraction $1 - f$, the total measurement with the light on is $[P \times ft \pm (ft\sigma_P{}^2)^{1/2}]$; with the light off, it is $\{D \times (1 - f)t \pm [(1 - f) \times t \times \sigma_D{}^2]^{1/2}\}$. Hence, the variances of P and D are, respectively,

$$\sigma_{P,t}^2 = \sigma_{P,1}^2/(ft) \tag{1.5.10}$$

and

$$\sigma^2_{D,t} = \sigma^2_{D,1}/[(1-f)\times t]. \tag{1.5.11}$$

From Eq. (1.5.6), we see that the estimated light flux on the tube is proportional to

$$S = P - D, \tag{1.5.12}$$

so that

$$\sigma^2_{S,t} = \sigma^2_{P,t} + \sigma^2_{D,t} = (\sigma^2_{P,1}/ft) + [\sigma^2_{D,1}/(1-f)t], \tag{1.5.13}$$

where

$$\sigma^2_{P,1} = \int_0^\infty [w(h)]^2\, p(h)\, dh \tag{1.5.14}$$

and

$$\sigma^2_{D,1} = \int_0^\infty [w(h)]^2\, d(h)\, dh. \tag{1.5.15}$$

The value of f which minimizes $\sigma^2_{S,t}$ is found by setting the derivative of Eq. (1.5.13) with respect to f equal to zero; this gives

$$(1-f)^2\sigma^2_{P,1} = f^2\sigma^2_{D,1} \tag{1.5.16}$$

or

$$[f/(1-f)] = \sigma_{P,1}/\sigma_{D,1}. \tag{1.5.17}$$

Hence,

$$f = \sigma_{P,1}/(\sigma_{P,1} + \sigma_{D,1}) = 1/[1 + (\sigma_{D,1}/\sigma_{P,1})]. \tag{1.5.18}$$

In the weak-signal limit, $p(h) \to d(h)$ so $(\sigma_{D,1}/\sigma_{P,1}) \to 1$ and $f \to \frac{1}{2}$; in the strong-signal limit, $p(h) \gg d(h)$ so $(\sigma_{D,1}/\sigma_{P,1}) \to 0$ and $f \to 1$. These results for the limiting cases are well known; McCutchen[101] even stated in 1957 that "the best use of a given total counting time requires that it should be divided between experiment and background runs as the square root of the recorded count rates," which is just Eq. (1.5.17) if we recall that $\sigma_{P,1} = P^{1/2}$ in a counting experiment (and similarly for D). However, because of the dependence of $\sigma_{D,1}$ and $\sigma_{P,1}$ on $w(h)$, the optimum value of f for other cases will depend on the weighting function used.

If the measurement is made against a background light level, as in astronomical photometry of faint stars against the night sky, the background distribution,

$$b(h) = s_B(h) + d(h), \tag{1.5.19}$$

must be substituted for $d(h)$ in Eqs. (1.5.5)-(1.5.18). Here, $s_B(h)$ is proportional to the background light. If $s_B(h)$ is larger than the signal $s(h)$, we again have the low signal value $f = \frac{1}{2}$, even if $s(h) \gg d(h)$, as may happen with a cooled photomultiplier. However, in this case the distributions $b(h)$ and $p(h)$ are similar apart from a scale factor, so that $(\sigma_{B,1}/\sigma_{P,1})$ is nearly independent of $w(h)$; this similarity introduces some features of the strong signal case.

We now examine the effects of different weighting functions in detail.

1.5.2.1.2. Pulse Counting. As has been pointed out above, the small-pulse end of $d(h)$ generally is at least as steep as $1/h$. Therefore, since the pulse-counting weighting function

$$w_{pc}(h) = \begin{cases} 0, & h < l, \\ 1, & l \le h \le u, \\ 0, & h > u, \end{cases} \tag{1.5.20}$$

converts the integrals of Eqs. (1.5.6)-(1.5.9) to the form

$$n_p(l, u) = \int_l^u p(h)\, dh, \tag{1.5.21}$$

we must have $l > 0$ to prevent divergence. Physically, the finite number of dynodes would in principle provide such a cutoff, but this is usually far below amplifier noise and load resistor noise. As a practical matter, a finite lower threshold is always required in pulse counting; the strong signal limit is physically unattainable for this weighting function.

What lower threshold l should be used to achieve maximum signal-to-noise ratio? The signal-to-noise ratio ϱ is

$$\begin{aligned}\varrho_t &= S/\sigma_{S,t} = n_s(l,u)\times[\sigma^2_{S,t} + \sigma^2_{D,t}]^{-1/2} \\ &= t^{1/2}\times n_s(l,u) \\ &\quad \times \{(1/f)\, n_s(l,u) + \{(1/f) + [1/(1-f)]\}\times n_d(l,u)\}^{-1/2},\end{aligned} \tag{1.5.22}$$

where n_s and n_d are defined analogously to n_p [see also Eq. (1.5.21)]. Here we have decomposed the sum

$$n_p(l, u) = n_s(l, u) + n_d(l, u) \tag{1.5.23}$$

and collected the n_d terms. We already know from Eq. (1.5.18) that

$$f(l, u) = \frac{(n_s + n_d)^{1/2}}{(n_s + n_d)^{1/2} + (n_d)^{1/2}}. \tag{1.5.24}$$

We could substitute Eq. (1.5.24) into Eq. (1.5.22) and eventually solve for the optimum discriminator settings l and u in the general case. However, in practice it is usually desirable to set $f = \frac{1}{2}$ or $f \approx 1$; the small loss in ϱ is compensated by the increase of operational efficiency due to convenience.[101] In general, we select l and u to maximize ϱ for the weakest signal we wish to measure. We must keep l and u fixed for all signal levels, or [due to the dependence of $s(h)$ on wavelength] the effective spectral response will vary. Optimizing for weak signals means we have less than optimum signal-to-noise for strong ones, but this is not a problem since we still have much better signal-to-noise on strong signals than on weak ones.

Having selected a value for f, which should be $\frac{1}{2}$ for genuinely weak signals and should be closer to 1 for stronger ones, we must set

$$\partial\varrho(u, l)/\partial u = \partial\varrho(u, l)/\partial l = 0 \tag{1.5.25}$$

to find optimum discriminator settings u and l. This gives

$$d(l)/s(l) = d(u)/s(u) = (1 - f) + 2(n_d/n_s). \tag{1.5.26}$$

For the weak signal limit, both $[d(h)/s(h)]$ and $(n_d/n_s) \to \infty$, so the f term is negligible. This gives

$$d(l)/s(l) = d(u)/s(u) = 2(n_d/n_s), \tag{1.5.27}$$

independent of f. (However, we already know we should pick $f = \frac{1}{2}$ in this case.) This condition is not easy to determine, since n_d and n_s are both functions of u and l.

The choice is simplified if we use the fact that both $d(h)$ and $s(h)$ are, in general, rapidly decreasing functions of h. Therefore, we are not far wrong to set

$$n_d(l, u) \approx n_d(l, \infty) = \int_l^\infty d(h)\, dh \tag{1.5.28a}$$

and

$$n_s(l, u) \approx n_s(l, \infty) = \int_l^\infty s(h)\, dh. \tag{1.5.28b}$$

The functions on the right are simply the usual "bias curves" or "integral pulse-height distributions." If we define

$$R(h) = 2s(h)/d(h) \tag{1.5.29}$$

and

$$Q(l, u) = n_s(l, u)/n_d(l, u), \tag{1.5.30}$$

Eq. (1.5.27) corresponds to

$$R(l) = R(u) = Q(l, u). \tag{1.5.31}$$

Notice that $R(h)$ is twice the ratio of the signal and dark distributions, and $Q(l, u)$ is the ratio of signal counts to dark counts (sometimes erroneously called the signal-to-noise ratio). Equation (1.5.28) means that $Q(l, u) \approx Q(l, \infty)$, so

$$R(l_1) = Q(l_1, \infty) \tag{1.5.32}$$

defines a good first approximation to l. Then

$$R(u_1) = R(l_1) \tag{1.5.33}$$

defines an approximate upper cutoff, so we can determine a second approximation l_2 from

$$R(l_2) = Q(l_2, u_1) \tag{1.5.34}$$

and so on. The process converges very rapidly; an example is given in the Appendix of Ref. 30.

Notice that the result is independent of the signal strength used to compute R and Q, as both sides of Eqs. (1.5.29)–(1.5.34) are proportional to S.

The first approximation l_1 corresponds to Morton's[112] condition that the slope of the signal pulse integral distribution be half that of the dark pulse distribution, when the curves are plotted on semilog paper. For,

$$d[\log n_d(h, \infty)]/dh = -d(h)/n_d(h, \infty) \tag{1.5.35}$$

and similarly for $[n_s(h, \infty)]$, so that Morton's condition that

$$d[\log n_s(h, \infty)]/dh = \tfrac{1}{2} d[\log n_d(h, \infty)]/dh \tag{1.5.36}$$

gives Eq. (1.5.32) if $h = l_1$.

If a single level discriminator is used instead of a window counter, $u = \infty$ and $l = l_1$.

In the case of moderately large signals, we may suppose that Q is

[112] G. A. Morton, *Appl. Opt.* **7**, 1 (1968).

very large, so that we can neglect the (n_d/n_s) term in Eq. (1.5.26). The optimum discriminator settings are then given by

$$d(l)/s(l) = d(u)/s(u) \approx (1 - f), \tag{1.5.37}$$

which is the fraction of the time we have reserved for counting dark pulses. We know that $d(h)/s(h)$ becomes very large as $h \to 0$ or $h \to \infty$; therefore, the only question regarding the applicability of Eq. (1.5.37) is whether the signal is indeed large enough to make $(n_d/n_s) \ll (1 - f)$, which is already a small quantity. Another way of saying this is that if the dark count is not negligible, we should spend more time measuring it; i.e., increase $(1 - f)$. The stronger the signal, the smaller we can make Eq. (1.5.37); in the limit of no dark pulses, we would have $l \to 0$ and $u \to \infty$, counting all the pulses. This would clearly give the maximum signal-to-noise ratio possible: we cannot do better than to count every photoelectron equally.

Even if our weakest signal is not strong, we may be able to use Eq. (1.5.37) to determine approximate discriminator settings l_1 and d_1, and again use an iterative technique to find the best values

$$\frac{d(l_k)}{s(l_k)} = \frac{d(u_k)}{s(u_k)} = (1 - f) + 2\,\frac{n_d(l_{k-1}, u_{k-1})}{n_s(l_{k-1}, u_{k-1})}. \tag{1.5.38}$$

All of the above assumed pulses are independent. However, both the signal and dark distributions contain afterpulses and other induced pulses such as photoemission from dynode glow, etc. These induced events are certainly not statistically independent of the primary events. Any correlation between pulses of heights h_1 and h_2 should appear as a cross-product term in the integrand of Eq. (1.5.9). Another way of treating this problem is to regard an induced pulse as increasing the weight of the parent pulse.

Let us consider the large afterpulses,[40] which may typically be about four times the height of a primary photoelectron pulse and occur with probability ~0.05 about 0.3 μsec after the primary pulse. If the counting equipment used is relatively slow, the two pulses will not be resolved and will be treated as one large pulse. Fast counters, on the other hand, will count both pulses. In this case we can regard 0.95 of the n_i independent pulses as having weight 1, and $0.05n_i$ as having weight 2, since these are counted twice. The observed count is then

$$\begin{aligned} n_{\text{obs}} &= 0.95n_i + 2\times 0.05n_i \\ &= 1.05n_i, \end{aligned} \tag{1.5.39}$$

and the variance [see Eq. (1.5.14)] is

$$\begin{aligned} \sigma^2_{\text{obs}} &= 0.95n_i + 2^2\times 0.05n_i \\ &= 1.15n_i \approx 1.1n_{\text{obs}}. \end{aligned} \tag{1.5.40}$$

Thus the observed variance is about 10% larger than would be expected from the total count. Such a small deviation from ideal counting statistics would be hard to measure.

The large afterpulses could be rejected by the upper level discriminator in a fast system. In a slow system, on the other hand, they would appear (unresolved from their parent pulses) as legitimate signal pulses and should be counted. This comparison shows that the pulse height distributions should be measured with the same equipment that is used for light detection, since different systems will measure different pulse height distributions from the same tube.

The situation is much worse in the case of cosmic-ray afterpulses, since typically about 10 afterpulses may be produced per cosmic ray.[42–46] In this case $n_{\text{obs}} \approx 10n_i$ and $\sigma^2_{\text{obs}} \approx 10^2 n_i = 10n_{\text{obs}}$, if the cosmic rays dominate the dark pulses. This agrees with Rodman and Smith's data[48] on refrigerated S-20 tubes, in which σ^2 was five to ten times higher than the total number of pulses counted per sample time. Their typical dark count rates of ~20 pulses/sec are a few dozen times the expected cosmic ray rate for a 2-in. (5-cm) tube; the agreement is very good if about half of Rodman and Smith's dark counts were due to cosmic rays, and the other half to spontaneous events originating within the tube.

Finally, one should realize that the $1/h^2$ component of both $s(h)$ and $d(h)$ is mostly due to daughter pulses that are not statistically independent; their numbers should fluctuate by about the same factors as the smaller number of parent pulses that caused them. Such large fluctuations are observed; in fact, the difficulty of obtaining reproducible results at small heights has prevented most investigators from publishing any data on this region. If the lower discriminator threshold is set too low, these small pulses will add to the effective weights of the parent pulses.

Clearly, the effective weighting function can be much more complicated than expected, if correlated pulses are present. The rapid increase in correlated pulses with extended red or uv response means that tubes with wider spectral response than actually needed should be avoided.

Before leaving the subject of pulse counting, we offer a prescription for adjusting a pulse-counting system to optimize signal/noise ratio. A practical method must allow for correlated pulses and other nonideal weighting effects, and should use the actual counting equipment used for photometry (i.e., should not require a PHA). Fortunately, experimental signal/noise functions have rather broad maxima for most tubes, so the exact discriminator setting is not critical. We assume that "strong" and "weak" constant-light sources are available, and that the PMT housing is light-tight, so that an accurate dark measurement can be made.

After allowing adequate warm-up or cool-down time to stabilize the (dark) PMT and electronics, turn the tube voltage down to a very low value (less than $\frac{1}{3}$ normal operating voltage), or off, to look for electronic noise. Run the lower discriminator level down until noise pulses appear in the counter; if a window counter is used, the upper edge should be set as high as possible. Find the discriminator level at which a few noise pulses per minute occur; then raise the discriminator to at least twice this level. Or, if no amplifier noise is seen, set the discriminator to a threshold level near $\frac{1}{2}$ V (most discriminators do not function reliably much below this level).

Having set the discriminator to a low but reliable level, increase the high voltage in steps of about 10% (50- or 100-V steps are usually convenient). At each step, count both the dark pulses and the (dark + light) pulses with the strong light on, each for a time $t \geq 10$ sec. (The strong light should produce a count rate much higher than the dark rate, but not so high as to cause fatigue or nonlinearity. About 10^4 counts/sec is suitable. A 1-mm pinhole placed 3 km from an ordinary 60-W incandescent lamp transmits about the right amount of light, for most tubes, so an attenuation of 10^6 is needed if the lamp is placed 3 m ($\sim$10 ft) away from such a pinhole. Thus the required strong light is really quite dim by laboratory standards.) Do not increase the tube voltage beyond the recommended maximum value, or the point at which the count rates exceed 10^5/sec, whichever comes first.

If the light was bright enough, the light count rate P should exceed the dark count rate D by a large factor at each voltage, so that $S = P - D \approx P$. Then a plot of $(P/D^{1/2})$ as a function of tube voltage should be a fair approximation to $(S/D^{1/2})$, the expected signal/noise ratio if all pulses are independent. The maximum in this function should be a fair first approximation to the optimum high voltage. In general, this value should be near the manufacturer's "typical" operating voltage for the tube. Excessively high voltages tend to increase dark noise; excessively low ones lead to poor photoelectron collection efficiency.

Set the high voltage 50–100 V above the indicated optimum, and reduce the light until the (light *minus* dark) count rate S is about half the dark rate D (weak light level). The high voltage, or the discriminator level, can now be adjusted in smaller steps ($\sim$2% in voltage, or $\sim$20% in discriminator level) to find the point that actually maximizes (S/σ_s), as estimated statistically from repeated measurements. For example, suppose we make n ($\geq$20) paired measurements of P and D at each setting. For each pair we compute $S_i = P_i - D_i$; the ratio of the mean ($\bar{S} = \sum S_i/n$) of the S_i to their standard deviation $\{s_S = [\sum (S_i - \bar{S})^2/(n-1)]^{1/2}\}$ is a good estimate of (S/σ_s) for weak signals. If 5 or 6 settings are used, and 20 bright and dark readings of 10 seconds' duration are made at each setting, the whole optimization can be done in less than 1 hr. Such a small investment of time is well worth while to achieve optimum results at low light levels.

If a window-type counter is used, the upper threshold setting is not critical, as very few very large pulses occur in any case. Usually, a setting that excludes 1% of the strong-light count (or 10 times the lower-threshold value, whichever is larger) is satisfactory.

Finally, if the maximum in (S/σ_s) is rather broad and flat, it may be desirable to pick an operating point slightly off the peak, so as to improve stability of operation. Maximum stability against gain changes occurs when $d[\ln P(h)]/d[\ln h] = h/P(dP/dh)$ is least, for a given light level; thus, one wants to operate at the flattest part of a log–log plot of count rate, $P(h)$, against discriminator level h. In general, this depends on light level, and cannot be optimized for both strong and weak lights. However, as "dark" or "background" checks are usually made more frequently than standard-source (gain) checks, it is usually best to optimize for gain stability at the higher light levels.

1.5.2.1.3. Direct Current Methods. In dc or charge-integration photometry,

$$w_{\mathrm{dc}}(h) = h. \tag{1.5.41}$$

Thus, the measured quantity per unit time is

$$P = \int_0^\infty h \times p(h)\, dh = \mu_{1,p} \tag{1.5.42}$$

with variance

$$\sigma_{p,1}^2 = \int_0^\infty h^2 \times p(h)\, dh = \mu_{2,p}\,. \tag{1.5.43}$$

The signal-to-noise ratio is therefore

$$\begin{aligned} \varrho_t = S/\sigma_{S,t} &= \mu_{1,s} \times [\sigma_{S,t}^2 + \sigma_{D,t}^2]^{-1/2} \\ &= (t)^{1/2} \mu_{1,s} \{(1/f)\mu_{2,s} + \{(1/f) + [1/(1-f)]\}\mu_{2,d}\}^{-1/2}, \end{aligned} \tag{1.5.44}$$

where $\mu_{1,s} = S$ is the first moment of the signal pulse distribution, and the μ_2's are the second moments defined as in Eq. (1.5.43). Note that $\mu_{i,s}$ is proportional to S for all i.

In the weak signal limit we can ignore $\mu_{2,s}$ in Eq. (1.5.44) and have

$$\varrho_{\text{weak}} \to S(t/4\mu_{2,d})^{1/2} \tag{1.5.45}$$

for $f = \frac{1}{2}$. Here, $(2\mu_{2,d})^{1/2}$ is the dark-current noise; the dark current itself is $\mu_{1,d}$.

Compared to pulse counting, dc photometry discriminates against the small dark pulses originating in the dynodes. Thus, for spontaneous dynode emission, $d(h) \propto 1/h$ for $h < h_0$ and the dark current

$$\int_0^{h_0} h/h\, dh$$

is finite. Even the induced dynode pulses proportional to $1/h^2$ give a finite dark current, since each dynode contributes equally to the anode current and there are only a finite number of dynodes.

The noise contribution from dynodes is even smaller. Since spontaneous emission from D1 is amplified g times less than cathode emission, the noise power contributed by D1 emission is g^2 (typically 10–20) times less than that due to cathode emission. Even for induced emission, each dynode contributes g times less noise than its predecessor, so the total is finite and comes mainly from the first stages. Thus, the dark noise is dominated by the very large ion, cosmic-ray, and radioactive-decay pulses.

In a cooled tube, the dark *noise* can be due mainly to cosmic rays (and nearly independent of temperature), even though the dark *current* is mainly due to ions (and is still temperature-sensitive).[41] In computing the CR noise, the small afterpulses should be added to the weight (i.e., height) of the initial Cerenkov pulse in proportion to their own heights; at most, this doubles the noise computed from the giant pulses alone. Thus, although most of the dark current usually comes from the dynodes, most of the dark-current noise comes from the cathode. The dc value of dark current is therefore not a good indicator of the tube quality for low-level work, even if dc leakage is negligible. This is particularly true for a refrigerated end-on tube, where cosmic rays are relatively more important.

In the strong signal limit, we can ignore the $\mu_{2,d}$ term in Eq. (1.5.44) and have

$$\varrho_{\text{strong}} \to S(ft/\mu_{2,s})^{1/2}. \tag{1.5.46}$$

We can safely set $f \approx 1$ here, so

$$\varrho \approx [\mu_{1,s}/(\mu_{2,s})^{1/2}](t)^{1/2}. \tag{1.5.47}$$

An ideal quantum detector of efficiency q achieves a signal-to-noise ratio

$$\varrho_{\text{ideal}} = Nqt/\sigma_{Nqt} = Nqt/(Nqt)^{1/2} = (Nqt)^{1/2} \tag{1.5.48}$$

in observing a stream of N photons per second. If we normalize $s(h)$ so that

$$s(h) = Nq_k s_0(h), \tag{1.5.49}$$

where $s_0(h)$ is the probability distribution of signal pulses and q_k is an effective (i.e., allowing for lost electrons) cathode quantum efficiency, we see that Eq. (1.5.47) becomes

$$\varrho \approx (Nt)^{1/2} \times (q_k \mu_{1,s_0}^2/\mu_{2,s_0})^{1/2}. \tag{1.5.50}$$

Thus, the detective quantum efficiency of dc photometry is just the cathode quantum efficiency times a degradation factor,

$$\Delta_{\text{dc}} = \mu_{1,s_0}^2/\mu_{2,s_0}, \tag{1.5.51}$$

which is the ratio of the square of the mean pulse height to the mean square pulse height. Since the detective quantum efficiency for pulse counting, in the strong signal limit, approaches the cathode quantum efficiency q_k, Eq. (1.5.51) also gives the ratio of dc to pulse-counting detective quantum efficiencies for strong signals.

Prescott[29] has shown that $s_0(h)$ probably belongs to a family of functions bounded by the Poisson distribution on one hand and by the exponential distribution on the other. Hence, we can use these two limiting forms to place bounds on Δ_{dc}.

For the exponential distribution e^{-h} we have $\mu_1 = 1$ and $\mu_2 = 2$, so $\Delta_{\text{dc}} = \frac{1}{2}$, a result first published by Lynds and Aikens,[113] although it was probably discovered earlier by Baum. For the Poisson distribution, we have to pick the mean value. We will surely have an overestimate for Δ_{dc} if we set $\mu_1 = g_1$, the gain of the first dynode; for we then neglect the additional broadening of $s_0(h)$ by multiplication statistics at the following dynodes. Since the Poisson distribution has the property that

$$\sigma^2 = \mu_2 - \mu_1{}^2 = \mu_1, \tag{1.5.52}$$

[113] C. R. Lynds and R. I. Aikens, *Publ. Astr. Soc. Pacific* **77**, 347 (1951).

we have

$$\Delta_{\mathrm{dc}} \leqq \mu_1{}^2/(\mu_1{}^2 + \mu_1) = g_1/(g_1 + 1). \tag{1.5.53}$$

Hence, even for strong signals, pulse counting is more efficient than dc photometry. However, the pulse-counting advantage is smaller for tubes with narrower pulse-height distribution [contrary to the myth that says tubes with narrow $s(h)$ and/or $d(h)$ are more suitable for pulse counting]. Tubes such as the 1P21/931-A or ITT tubes, which have a more nearly Poisson signal pulse distribution, will therefore be slightly better for dc photometry than the EMI venetian blind tubes. These results agree with Baum's[111] experimental data, which showed a DQE advantage for pulse counting of about 1.3 for a 1P21 in the strong signal limit, and factors approaching 2.0 for other types. The relatively small advantage of pulse counting at high light levels agrees with Nakamura and Schwarz's statement[114] that "simple dc measurements are about as good as pulse counting for light levels above the dark current [equivalent] level."

Afterpulsing is more important for dc work than for pulse counting. If we again assume a 5% rate of afterpulses four times the average signal pulse height, we have an anode current proportional to

$$n_{\mathrm{obs}} \approx 0.95n_i + (4 + 1)\times 0.05n_i = 1.2n_i, \tag{1.5.54}$$

where $(4 + 1)$ is the total height of an afterpulsing event, and n_i is the number of photoelectron pulses $(= Nq_k t)$. Ignoring the spread in pulse heights within each group, we have

$$\sigma^2_{\mathrm{obs}} \approx 0.95n_i + (4 + 1)^2\times 0.05n_i = 2.2n_i \approx 1.83n_{\mathrm{obs}}. \tag{1.5.55}$$

Thus, a typical rate of afterpulsing can roughly double the relative noise power, and halve the DQE. This may explain why 1P21's usually give best dc signal/noise ratios at relatively low overall voltages (700–800 V): these relatively gassy tubes must be run at low potentials to keep ion production and afterpulsing to a minimum.[115]

One argument frequently offered[112] in favor of pulse counting is that it is relatively insensitive to gain drifts. This argument assumes that the discriminator level can be set at a point on $s(h)$ where nearly all signal pulses are counted, and where $ds/d(\ln h)$ is small, so that a small fractional change in effective discriminator level l produces a very small fractional

[114] J. K. Nakamura and S. E. Schwarz, *Appl. Opt.* **7**, 1073 (1968).
[115] R. W. Engstrom, *J. Opt. Soc. Amer.* **37**, 420 (1947).

change in $n(l, u)$. These assumptions can only be approximately met for strong signals, however. In the weak signal limit, we are always fighting the $1/h^2$ component of the dark pulses, so that a fractional change in l produces a comparable change in $n_d(l, u) \approx 1/l$. Thus, pulse counting and dc detection have similar sensitivities to gain changes, if the weak-signal advantage of pulse counting is to be realized. One can, however, choose to sacrifice weak-signal performance in order to obtain less sensitivity to gain changes at high light levels. Even so, the "peak-to-valley" ratio at the single-electron maximum of $p(h)$ is usually not very high, owing to the induced (dynode) pulses; so the decrease in gain sensitivity is rarely as large as a factor of ten (i.e., 10% change in gain will still produce a 1% change in signal).

1.5.2.2. Other Methods. Various other techniques have been proposed, involving still other weighting functions. Among these may be mentioned the measurement of the shot-noise power instead of the current at the anode. In this case the signal depends on the second moment of $p(h)$, and the noise depends on the fourth moment. Consequently, the shot-noise method is still less sensitive to the small dynode pulses (which are always a minor problem in dc detection), and still more sensitive to the large CR and ion pulses. The analysis has been carried out,[30] but will not be repeated here as it shows no large advantages for this method. This scheme is more complicated to use than either dc or pulse counting, and suffers from both the limited dynamic range (nonlinearity) of pulse counting (due to the nonlinear weighting), and the dc disadvantages of zero drift and analog (rather than digital) output.

Symmetrical chopping followed by synchronous detection has also been tried; as pointed out earlier, the lack of $1/f$ noise in photomultipliers deprives this technique of its main advantage. Of course, chopping does eliminate the zero-drift problems of most dc amplifiers; but this is better done by chopping the (filtered) anode current (as is done in some solid state operational amplifiers), rather than the light, which throws away half the photons. Chopping the light fixes f at $\frac{1}{2}$, since the dark current is observed during the half cycles when the light is cut off. No separate dark reading is then necessary. However, the value $f = \frac{1}{2}$ is inefficient at higher light levels. At moderate light levels, Nakamura and Schwarz[114] found little difference between synchronous detection and pulse counting; one should bear in mind that pulse counting may also reject an appreciable fraction of the light pulses, if the discriminator rejects most of the dark pulses.

Finally, one can ask what weighting function would maximize the signal/noise ratio ϱ. In the case of a noiseless photomultiplier $[d(h) \equiv 0]$, we maximize the signal/noise ratio ϱ by counting all pulses equally. Also, in a noiseless tube with a background light level $b(h) \propto s(h)$, we readily see that all pulses should be counted equally.

In the case of the window counter, the condition $R(u) = R(l)$ shows that it is not the pulse height or pulse rate that is important, but the probability that a pulse is a signal pulse; if all pulses are equally likely to be signal pulses, all should be counted equally.

This suggests that we should take

$$w(h) = s(h)/p(h). \tag{1.5.56}$$

In the weak signal limit, $s \to 0$ and $p \to d$, so we may adopt

$$w_0(h) = s_0(h)/d(h), \tag{1.5.57}$$

where $s_0(h)$ is defined by Eq. (1.5.49). In fact, we can show that this function does maximize[30] the weak-signal signal-to-noise ratio ϱ. We may regard this result as an example of the principle of matched filtering, in which a weighting scheme proportional to the probability of success gives optimum detection.

However, the optimum weighting scheme, which would require detailed pulse-height analysis followed by computer processing, has hardly any advantage over pulse counting. Clearly, it has maximum advantage when $s(h)$ and $d(h)$ have very different shapes; for if they had the same shape, we should simply count all pulses (equal weighting) to achieve maximum ϱ. So, we shall consider an idealized tube in which

$$s_0(h) = e^{-h}$$

and

$$d(h) = \delta \times h^{-2}$$

for $h \leq h_{\max}$; we suppose $h_{\max} \approx 4$ if ion pulses are the largest dark pulses, and $h_{\max} \approx 10$–100 if cosmic-ray and gamma-ray pulses are important.† We adopt these forms for $s(h)$ and $d(h)$ because they are typical of many real tubes, and because this choice makes s and d very different in shape and hence produces large changes for different weighting functions. For tubes (such as the ITT tubes) in which $s(h)$ and $d(h)$

† In some of the following we shall take $h_{\max} \approx \infty$, where only a small error is involved.

are similar, we have essentially the large signal or background situation, in which different weighting functions produce rather similar results, and pulse counting gives the best results.

For pulse counting, we find $l_1 = \frac{1}{2}$; the successive approximations are $u_1 = 5.19$; $l_2 = 0.520$; $u_2 = 5.07$; $l_3 = 0.520$. The corresponding values of ϱ^2 as a function of l and u are

$$\varrho_{\mathrm{pc}}^2(\tfrac{1}{2}, \infty) = tN^2q_k^2/8e\delta = (tN^2q_k^2/\delta)\times 0.046 \tag{1.5.58a}$$

and

$$\varrho_{\mathrm{pc}}^2(0.52, 5.07) = (tN^2q_k^2/\delta)\times 0.050. \tag{1.5.58b}$$

Thus, a window counter, used optimally, would give about 10% higher efficiency than a simple discriminator. In order to reject the optimum amount of dark noise, we can only count about 60% of the actual signal pulses; if a wide dynamic range is required, the detective quantum efficiency for strong signals cannot exceed $0.6q_k$. For weak signals, the noise equivalent input is about

$$N_{0,\mathrm{pc}} = (Se\delta)^{1/2}/q_k \tag{1.5.59}$$

for the simple discriminator, and about 5% lower for the optimum window counter.

With dc detection, Eq. (1.5.45) gives

$$\varrho_{\mathrm{dc}} = Nq_k/(t/4\delta h_{\max})^{1/2}, \tag{1.5.60}$$

so

$$N_{0,\mathrm{dc}} = 2(\delta h_{\max})^{1/2}/q_k. \tag{1.5.61}$$

Thus, the dc performance for weak signals depends on the largest dark pulses, and may be either better or worse than pulse counting, depending on whether $h_{\max}$ is less or greater than $2e \approx 5.4$. Nakamura and Schwarz compared dc and pulse counting in an EMI 9558 at -45°C, and found only about a factor of two difference in noise equivalent input. This gives $h_{\max} \approx 20$, which seems reasonable since cosmic-ray noise dominates under these conditions. As mentioned before, $q = q_k/2$ for this case at large signals. Thus, pulse counting is only 20% more efficient than dc detection in the large signal limit, if the discriminator is optimized for weak signals. (We must use the same discriminator setting for both; because the pulse-height dependences of spectral and spatial response, data taken at different discriminator settings are not directly comparable.)

Now let us look at the results expected for the optimum weighting function $w_0(h) = h^2 e^{-h}/\delta$. We have

$$W_0 = \int_0^\infty S_0^2(h)/d(h)\, dh = \int_0^{h_{\max}} (e^{-2h}/\delta h^{-2})\, dh = 1/4\delta, \quad (1.5.62)$$

so

$$\varrho_{\mathrm{opt}}^2 = N^2 q_k^2 t/16\delta \quad (1.5.63)$$

and†

$$N_{0,\mathrm{opt}} = 4(\delta)^{1/2}/q_k. \quad (1.5.64)$$

Thus, in the case we have considered, the advantage of using optimum weighting is small; the noise equivalent input is lower by a factor of only $(e/2)^{1/2} = 1.17$ than in pulse counting.

The price paid for optimum weak-signal detection is some decrease in strong signal detective quantum efficiency. In the strong signal limit,

$$\sigma_{S,t}^2 = \sigma_{S,1}^2/ft = \int_0^\infty w_0^2(h) \times s(h)\, dh/ft$$

$$= \int_0^\infty (h^2 e^{-h})^2 N q_k e^{-h}\, dh/ft\delta^2$$

$$= Nq_k/ft\delta^2 \int_0^\infty h^4 e^{-3h}\, dh = Nq_k/ft\delta^2 \times 8/81, \quad (1.5.65)$$

and

$$S = \int_0^\infty w_0(h)\, s(h)\, dh$$

$$= \int_0^\infty S_0(h)/d(h) \times Nq_k\, S_0(h)\, dh$$

$$= Nq_k \times W_0 = Nq_k/4\delta. \quad (1.5.66)$$

So, for $f \approx 1$,

$$\varrho^2 \approx S^2/\sigma_{S,t}^2 = (N^2 q_k^2/16\delta^2)/(8Nq_k/81t\delta^2) = Nq_k t \times 81/128. \quad (1.5.67)$$

The detective quantum efficiency for large signals is thus $(81/128)q_k = 0.633 q_k$. For the model we have chosen, this represents less degradation at high light levels than for any of the other detection methods (weighting functions). Thus, optimum weighting should produce better results at all light levels. The improvement in going from pulse counting to optimum detection is relatively small, however.

† Equation (1.5.64) was derived assuming $h_{\max} = \infty$. The error made is approximately $0.5 h_{\max}^2 \exp(-2h_{\max})$, which is ~ 0.01 for $h_{\max} = 3$ and 0.003 for $h_{\max} = 4$.

It is desirable to monitor the dc anode current even in non-dc detection, as gain drifts on the order of a factor of two may occur at microampere anode currents, due to dynode fatigue. Such large signals usually produce a temporary rise in dark noise, also.

A major limitation of any nonlinear weighting system is limited dynamic range; when a significant fraction (say, 1%) of pulses overlap, a similar degree of nonlinearity results. If we try to extend the range by switching over to dc methods for strong signals, we must remember that the apparent spectral response will change when we change weighting functions. The same problem occurs if we try to avoid saturation and fatigue effects by changing the tube voltage; the apparent change in red/blue response ratio can be several percent.

Some detection schemes involve combinations of techniques, such as those involving chopping and synchronous detection, or pulse counting followed by a rate meter and analog recording.[116] Apart from the loss of photons if chopping is used, such methods inherently offer the same signal/noise ratio as the simpler techniques, depending on which moment of the pulse-height distribution is measured (zeroth for pulse counting; first for current measurements; etc.).

1.5.2.3. Analog Versus Digital Recording. When high precision (~1% or better) is required, or when a large number of measures are to be averaged together to improve the signal/noise ratio, there are substantial advantages in digitizing the signal and recording all data in digital form. Obviously, all data must eventually be reduced to digital form for quantitative analysis; but many people, especially spectroscopists, like to see a graphical (analog) display. There is no doubt that a strip-chart display is valuable as a gross indication of equipment stability or malfunction, but such a qualitative diagnosis does not require very high resolution or linearity. The range of signal/noise ratios that can usefully be displayed on a chart only extends from 5:1 up to perhaps 100–200 to one, however.[103] At lower S/N, the signal is not visible in the noise; above this range, the noise becomes invisible against the signal.

Even in the useful S/N range of analog recording, it is inefficient to try to recover numerical data from a chart record.[117] Even an experienced worker can extract only about half or a third of the original information

[116] D. L. Akins, S. E. Schwarz, and C. B. Moore, *Rev. Sci. Instrum.* **39**, 715 (1968).
[117] A. T. Young, *Observatory* **88**, 151 (1968).

from a chart record. That is, the error of reading the chart is at least as large as the standard deviation of the mean deflection obtained by averaging over a centimeter of chart, even if the person reading the chart has years of experience; a novice makes much larger errors.

The chart-reading error appears to be independent of the filter used to smooth the analog record, however; a given steady signal can be read with equal accuracy, whether recorded with a short or a long time constant. Thus, provided that the filtering does not remove information, the observer may choose whatever smoothing he likes.

In addition to the substantial errors introduced by trying to digitize an analog record manually, one must consider the economics involved. A digitizer and printer or punch cost less than the annual salary of a chartmeasurer, as well as being faster and more reliable.

Finally, a digital system has a better filter function than an analog system with *RC* filtering. The digital system, which integrates for a time τ, not only averages symmetrically and uniformly in the time domain (in contrast to the one-sided and unequal weighting of the *RC* filter); it also gives statistically independent samples, instead of the partially correlated analog data, which are asymmetrically distorted by the "memory" of earlier values. On the other hand, one must remember that the equivalent bandwidth Δf of the *RC* filter is $1/(4RC)$, while that of the integrator is $1/(2\tau)$; thus, the two systems include the same noise bandwidth only if $RC = \tau/2$, not τ as is often erroneously assumed. The noise bandwidths and other characteristics of these and other filters are well discussed by Robben.[98]

To sum up, it is best to use a digital system to record data for analysis, while maintaining an analog record as a crude real-time indicator of data quality. (We may add that the advantages of digitized data have even been recognized in the analysis of photographic spectra.[61,118])

[118] G. I. Thompson, *Publ. Roy. Obs. Edinburgh* **5**, no. 12, 245 (1967); **7**, no. 2, 19 (1970). See also W. K. Bonsack, *Astron. Astrophys.* **15**, 374 (1971).

2. OTHER COMPONENTS IN PHOTOMETRIC SYSTEMS*

2.1. Optical Systems

2.1.1. The Telescope and Atmosphere

2.1.1.1. The Structure of a Star Image. In measuring the light of a star, a small diaphragm is placed in the focal plane of the telescope to exclude, as far as possible, unwanted light from neighboring stars and from the sky. The smaller the diaphragm, the better this interference can be rejected. However, if too small an aperture is used, a significant fraction of the measured star's light is also excluded. Because this excluded light depends on variable atmospheric and instrumental factors, it cannot be accurately corrected for, and errors will result in the measured data. Thus there is an optimum diaphragm size, which depends on the image structure and the precision and accuracy required.

2.1.1.1.1. The Telescope Alone: Diffraction and Stray Light. The diffraction pattern in the focal plane of a perfect telescope with a uniform circular aperture (pupil) is proportional to

$$I(r/\pi) = [2J_1(r)/r]^2, \tag{2.1.1a}$$

where r is the distance from the center of the image in (angular) units of λ/D rad; λ is the wavelength of observation; D is the diameter of the entrance pupil, and J_1 is the Bessel function of the first kind. However, most telescopes used for photometry have an annular aperture, obstructed by a secondary mirror. If the fraction of the diameter obscured is t, the ideal diffraction pattern has the form

$$\begin{aligned} I(r/\pi) &= 4[J_1(r)/r - t^2J_1(tr)/(tr)]^2/(1-t^2)^2 \\ &= 4[J_1(r) - tJ_1(tr)]^2/[r(1-t^2)]^2. \end{aligned} \tag{2.1.1b}$$

* Part 2 is by Andrew T. Young.

METHODS OF EXPERIMENTAL PHYSICS:

1. Classical Methods
   Edited by Immanuel Estermann
2. Electronic Methods, Second Edition (in two parts)
   Edited by E. Bleuler and R. O. Haxby
3. Molecular Physics, Second Edition (in two parts)
   Edited by Dudley Williams
4. Atomic and Electron Physics—Part A: Atomic Sources and Detectors, Part B: Free Atoms
   Edited by Vernon W. Hughes and Howard L. Schultz
5. Nuclear Physics (in two parts)
   Edited by Luke C. L. Yuan and Chien-Shiung Wu
6. Solid State Physics (in two parts)
   Edited by K. Lark-Horovitz and Vivian A. Johnson
7. Atomic and Electron Physics—Atomic Interactions (in two parts)
   Edited by Benjamin Bederson and Wade L. Fite
8. Problems and Solutions for Students
   Edited by L. Marton and W. F. Hornyak
9. Plasma Physics (in two parts)
   Edited by Hans R. Griem and Ralph H. Lovberg
10. Physical Principles of Far-Infrared Radiation
    L. C. Robinson
11. Solid State Physics
    Edited by R. V. Coleman
12. Astrophysics—Part A: Optical and Infrared
    Edited by N. Carleton

Volume 12

Astrophysics

PART A: Optical and Infrared

Edited by

N. CARLETON

Smithsonian Astrophysical Observatory
and Harvard University
Cambridge, Massachusetts

1974

ACADEMIC PRESS · New York and London

A Subsidiary of Harcourt Brace Jovanovich, Publishers

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

Library of Congress Cataloging in Publication Data
Main entry under title:

Optical and infrared astronomy.

(Methods of experimental physics, v. 12: Astrophysics, pt. A)
Includes bibliographical references.
1. Astronomical spectroscopy. 2. Photometry, Astronomical. 3. Spectrum, Infra-red. I. Carleton, Nathaniel, Date ed. II. Series: Methods of experimental physics, v. 12, pt. A. III. Series: Astrophysics, pt. A.
QB465.067 522'.6 73-17150
ISBN 0–12–474912–1

PRINTED IN THE UNITED STATES OF AMERICA

CONTENTS