Chapter 3

Photomultiplier Tubes

Sergey V. Polyakov
National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

Chapter Outline Head

3.1 INTRODUCTION

Photomultiplier tubes (PMTs), also known as photomultipliers, are remarkable devices. While a PMT was the first device to detect light at the single-photon level, invented more than 80 years ago, they are widely used to this day, particularly in biological and medical applications. Modern PMTs deliver low noise and low jitter detection over a wide dynamic range. However, they offer limited detection efficiency, especially for longer wavelengths. We discuss the physical mechanisms behind the photon detection in PMTs, the history of their development, and the key characteristics of PMTs in photon-counting mode.

3.2 BRIEF HISTORY

There are two distinct phenomena that are fundamental to the operation of a photomultiplier tube. The first is the photoelectric effect. A range of materials emit electrons when illuminated with light. The requirement is that the photons

Single-Photon Generation and Detection, Volume 45. http://dx.doi.org/10.1016/B978-0-12-387695-9.00003-2

have energy that is equal to or exceeding the so-called workfunction of the photoelectric material. The second is the phenomenon of secondary emission. When an incident electron possesses sufficient energy, it can knock out multiple electrons when hitting a surface, or when passing through a medium. In a PMT, an incident photon excites a photoelectron from the photocathode; then the photoelectron is accelerated by an electric field between the photocathode and the first dynode, where it knocks out multiple electrons. Those electrons are subsequently accelerated to a second dynode where each electron knocks out a few more electrons. This acceleration and multiple electron emission process is repeated on subsequent dynodes, with the number of dynode stages typically ranging from about 10 to about 20, yielding a final pulse of $\approx 10^6$ electrons at the PMT anode.

Heinrich Hertz discovered the photoelectric effect in 1887 [1]. Later, Albert Einstein established that the mechanism behind this effect is unequivocally quantum, and that individual photons of light transfer their energy to single electrons [2]. The photoelectric effect therefore provides a way to turn photons into an electrical signal. Of course, the electrical signal from just one electron is too tiny to be detected, as it would be within the noise of an ordinary amplifier. Therefore, special steps for noise reduction are necessary. Luckily, single-photon detection does not require faithful amplification: so long as we can distinguish presence of an electrical signal due to a photon detection from the noise in the electrical circuit, the height and shape of an electrical signal can vary from one detection to another.

This problem was resolved using the mechanism of secondary emission, discovered by L. Austin and H. Starke in 1902. Secondary emission occurs only if incoming (primary) electrons possess enough energy to liberate the electrons in the target material. The electrons emitted in this process are called secondary electrons. The average number of secondary electrons emitted in the process (and hence an amplification coefficient) rapidly increases with the energy of primary electrons, and then saturates. Saturation properties depend on the secondary emission surface. This effect is easily understood. As the energy of primary electrons grows, the number of excited electrons in the secondary-emissive material increases, but that happens at a greater depth. The deeper the electrons are excited, the lower is the chance that they can reach the surface and escape. As a result, the fraction of electrons that can escape saturates, and then decreases. Most surfaces saturate with primary electrons energy of ≈ 1 keV.

While the noise properties of an amplifier based on secondary emission are compatible with the needs of single-photon detection (practically, no secondary electrons can be generated in the absence of primary electrons), the limited gain of this process is a significant obstacle. In 1930, Leonid Kubetsky found a way to overcome this gain limitation. He proposed a multistage secondary emission device, [3] where the output of one stage is sent onto a subsequent secondary emission stage. The first high-gain amplifier implementation of this scheme was reported by Vladimir Zworykin and colleagues [4]. They characterized a

three-stage amplifier and reported a maximum measured gain of 60. Within less than a year, Kubetsky reported a measured amplification gain of "about 1000" [5].

Contemporary with these developments, a much improved photoelectric material was found, Cesium-antimony (Cs_3Sb) [6], which has much higher emission probability for the photoelectron (or *quantum efficiency*, discussed later in this Chapter in detail). Further commercial development of PMTs was led by RCA in the United States and Hamamatsu in Japan.

3.3 PRINCIPLE OF OPERATION

A typical PMT, as illustrated in Fig. 3.1, contains a photocathode, several dynodes, and an anode in a sealed glass envelope with a high vacuum inside. Photons incident on a PMT undergo the following steps:

1. Photons enter the tube through the input window.
2. Photons excite electrons, some of which are emitted from its surface into the vacuum. These emitted electrons are called photoelectrons.
3. Photoelectrons are focused with a focusing electrode onto the first secondary-electron emission surface, called a dynode. The number of photoelectrons is multiplied via the secondary-electron emission effect.
4. The secondary emission is repeated several times, on each of the dynodes in the device. Thus, high-gain amplification is achieved.
5. Secondary electrons from the last dynode are collected on the anode.
6. A current spike is detected by detection electronics.

In the following we discuss details of the two major elements of a PMT: photocathodes and dynodes. Further discussion of PMT engineering and usage can be found in a recent comprehensive handbook [7].

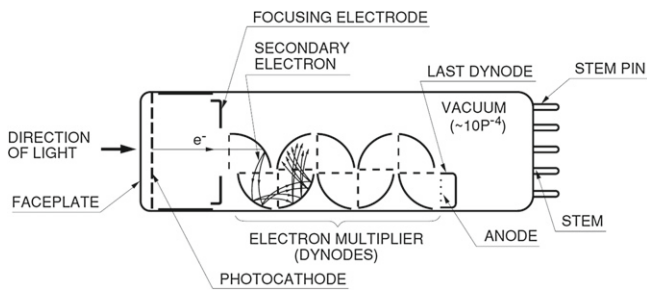

FIGURE 3.1 Typical photomultiplier tube (PMT). *Reproduced with permission from [7].*

3.3.1 Photoelectron Emission and Photocathodes

To be emitted to vacuum, an electron needs to overcome the vacuum level barrier, also known as the material work function. Such a barrier occurs at the surface of any solid. As a photon strikes a photocathode, electrons within the material can acquire some of the photons' energy and become excited. Then, as an excited electron diffuses through the material, it may come to the material's surface and escape to the vacuum provided that its energy is greater than the material's workfunction (i.e. sufficient to overcome the vacuum level barrier). The entire emission scenario can be quantitatively described as a series of probabilistic events. Thus, the *quantum efficiency*, or the ratio of photoelectrons to the incident photons, is a product of the probabilities of each of the steps. This quantum efficiency may be presented as:

$$\eta_{\mathrm{QE}} = (1 - R)\frac{P_{\mathrm{e}}(\lambda)}{k}\frac{1}{1 + 1/kl}P_{\mathrm{s}}, \tag{3.1}$$

where R is a reflection coefficient of incident light, k is the absorption coefficient, P_{e} is the probability that a photon can excite an electron of a sufficient energy, l is a mean escape length of the electrons, P_{s} is the probability that an electron that reaches the surface escapes into vacuum, and λ is the wavelength of the photon.

Another important factor contributing to the photocathode performance is its operation mode. Photocathodes may work in reflection or transmission, depending on the tube configuration. In reflection mode, the electrons are emitted from the illuminated side of the photocathode, and the input light normally comes through the side of the PMT's glass vacuum envelope. Transmission-mode photocathodes are used for head-on photomultiplier tubes. In this latter case, photoelectrons are emitted from the back side of the photocathode. In this configuration the light comes through the end of the cylinder-shaped glass envelope and often the photocathode is deposited directly on the inside of the glass.

Note that the parameters in the equation for quantum efficiency depend on both the choice of material for the photocathode and its operation mode. Photocathodes are generally made of compound alkali metals, or semiconductors activated with alkali, because these materials offer low vacuum level barriers. Cs-I and Cs-Te are used for ultraviolet light, for wavelengths as short as 115 nm. The short wavelength cutoff is due to limitations of glass used to seal the tube rather than the photocathode material. The photocathodes themselves can operate with even shorter wavelengths if the input window is removed. Both of these compounds are insensitive to visible and infrared light. Bialkali (Sb-Rb-Cs, Sb-K-Cs) and multialkali (Sb-Na-K-Cs) photocathodes are sensitive to the photons of visible light (up to 700 nm for bialkali and up to 850 nm for multialkali photocathodes). To extend sensitivity even further, to about 1000 nm, semiconductors, such as GaAsP, GaAs, and InGaAs activated

with alkali (Cs), are used as photocathodes. Relatively recently, [8,9], spectral sensitivity has reached telecom wavelengths (wavelengths up to 1700 nm) in specially engineered semiconductor-based structures based on PN junctions (albeit with relatively low efficiencies).

Table 3.1 summarizes typical peak quantum efficiencies and sensitivity ranges of the PMTs with typical photocathode materials. Note that in most cases, the short wavelength sensitivity cutoff is due to window transmittance limitations, while the long wavelength sensitivity cutoff is due to properties of a photocathode material.

From Table 3.1, we see that quantum efficiency of photocathodes, and hence of photomultipliers is highest in the ultraviolet. And while spectral sensitivities have expanded into the near infrared, the low quantum efficiencies there make PMTs much less attractive than other technologies, many of which are discussed in the following chapters.

3.3.2 Secondary Emission, Dynodes

As we have outlined, the amplification of the photocurrent in photomultipliers is done via a secondary emission. To optimize secondary emission, an electrode made of stainless steel, copper-beryllium or nickel is coated by a so-called secondary-emissive material. These include alkali-antimonide, beryllium oxide, magnesium oxide, gallium phosphide, and arsenide-gallium phosphide. Figure 3.2 shows typical secondary emission ratios δ obtainable with different secondary emissive surfaces. δ is defined as an average number of secondary electrons per primary electron. Secondary emission ratios depend on the energy of the primary electron, therefore, to use secondary emission as an amplification mechanism for photoelectrons efficiently, primary electrons need to be accelerated. Acceleration is accomplished by applying an electric field between the dynodes. As follows from Fig. 3.2, δ grows proportionally to the accelerating potential V, and then it saturates after V reaches a certain value, typically on the order of 1000 volts. Therefore, typical amplification is 10–100 per an accelerating surface. As discussed in the historical section of this chapter, to arrange for a higher amplification, multiple secondary emission surfaces must be used. In modern photomultipliers up to 20 dynodes are used in a sequence.

In a PMT, the accelerating electric field is created by applying electrical potentials to dynodes. The dynode shape is an important engineering consideration because the electric field and thus electron trajectories depend on the dynode shape. The goal here is improving collection efficiency of primary electrons, and reducing the uncertainty in latency time of the device (electron transit time spread).

The collection efficiency of the first dynode (that is, the dynode that immediately follows the photocathode) is most critical, because any inefficiency at this stage adversely affects the detection efficiency and overall performance of the photomultiplier. Typically, the collection efficiency of the first dynode

TABLE 3.1 Properties of Typical Photocathodes

	Reflection Mode		Transmission Mode	
Photocathode Material	**Spectral Range, (nm)**	**Peak Quantum Efficiency (%)**	**Spectral Range, (nm)**	**Peak Quantum Efficiency (%)**
Cs-I	115 to 200	26 @ 125 nm	115 to 200	13 @ 130 nm
Cs-Te	115 to 320	37 @ 210 nm	115 to 320	14 @ 210 nm
Sb-Cs	185 to 750	25 @ 280 nm	–	–
Bialkali	185 to 750	30 @ 260 nm	160 to 650	27 @ 390 nm
Multialkali	185 to 900	30 @ 260 nm	160 to 850	25 @ 280 nm
Ag-O-Cs	–	–	400 to 1200	0.36 @ 740 nm
GaAs(Cs)	185 to 930	23 @ 300 nm	380 to 890	14 @ 760 nm
InGaAs(Cs)	300 to 1040	16 @ 370 nm	–	–
InP/InGaAs(Cs)	300 to 1700	1 @ 1200 nm	950 to 1700	2 @ 1550 nm

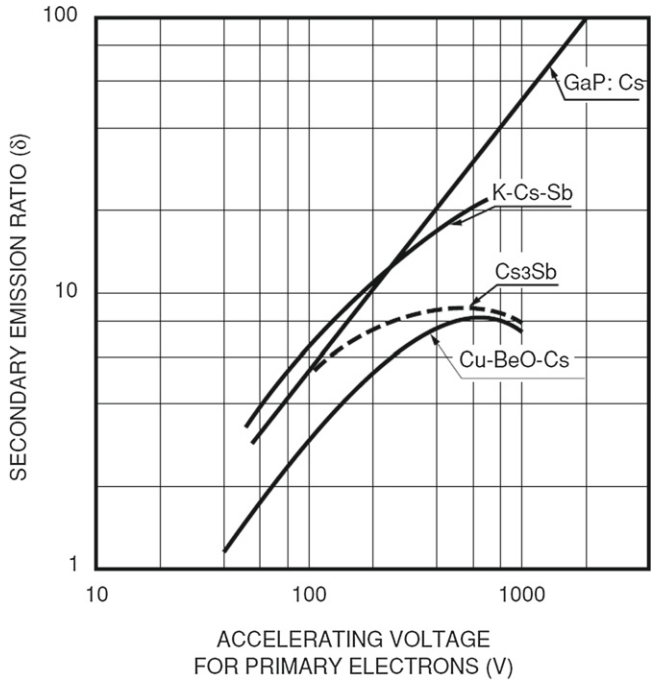

FIGURE 3.2 Secondary emission ratios as a function of an accelerating potential for typical dynodes. *Reproduced with permission from [7].*

is optimized to be better than 60–90%. It is common to engineer a better collection efficiency by focusing the trajectories of photoelectrons between a photocathode and the first dynode with a focusing electrode. If the goal is to minimize electron transit times, collection efficiency can be sacrificed in favor of shorter distances between dynodes and/or higher electric fields. A variety of dynode shapes are commercially available, with different designs leading to differences in device characteristics. In addition to temporal response and secondary-electron collection efficiency mentioned earlier, spatial and angular uniformity of detection efficiency depend on a dynode shape. Next we discuss timing characteristics of the PMTs, which are very important for many single-photon applications.

Let us define the relevant timing characteristics used to describe temporal performance of a photomultiplier, see Fig. 3.3. This figure depicts the typical electrical signal of a photomultiplier in response to a delta-function-like optical input (which could be a single photon). The latency time (also known as electron transit time) is determined by the time it takes electrons to traverse the tube from a photocathode to an anode. The latency time varies from shot to shot, because photoelectrons and secondary-electron beams may take different paths in a tube. Therefore it is relevant to not only characterize average latency but also

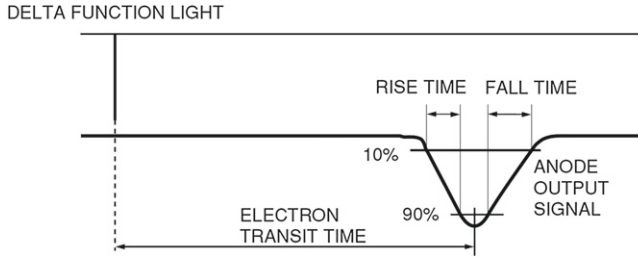

FIGURE 3.3 Temporal characterization of an output from a PMT. *Reproduced with permission from [7].*

its variation (or timing jitter), sometimes referred to as transit time spread (TTS) in photomultiplier literature. It is characterized by full width at half maximum (FWHM) of the latency time distribution histogram. This is not to be confused with pulse width, the overall width of the electrical output pulse, also measured by its FWHM. The rise time of the electrical signal is defined as the time it takes the output pulse to rise from 10 to 90 percent of the peak pulse height. The fall time is defined analogously, i.e. as time it takes for the electrical signal to change from 90 to 10 percent of the peak output pulse height.

In Table 3.2, we list most common dynode types and provide typical temporal, uniformity, and collection efficiency values. Figure 3.4 shows graphical depictions of photomultiplier tube geometries and dynode shapes. We see that PMT's timing characteristics depend strongly on dynode design. In addition, PMT characteristics may vary depending on the dynode's size and shape. We also see that there are tradeoffs between configurations that minimize TTS and maximizing collection efficiency or spatial response uniformity. Thus, the most suitable dynode shape should be chosen depending on a particular application.

3.4 PHOTON COUNTING WITH PHOTOMULTIPLIERS

Let us discuss how a photomultiplier device can be used for single-photon detection and photon counting. We have seen that a photocathode is sensitive to single photons, and that single photoelectrons can be sufficiently amplified with a sequence of dynodes. At the same time, current in the absence of the optical input (dark current) is small, because secondary emission occurs very seldom with no primary electrons present. These are the three most important prerequisites for photon counting. However, additional considerations are also required.

One obvious issue is a photomultiplier's response to a delta-function-like input. This is because a single photoelectronic event on a photocathode may be treated as such an input. The electrical current at the output is no longer a

TABLE 3.2 Characteristics of Photomultipliers vs Dynode Geometry

Dynode Type	Rise Time, (ns)	Fall Time, (ns)	Pulse Width, (ns)	Latency, (ns)	TTS, (ns)	Uniformity	Collection Efficiency
Linear-focused	0.7 to 3	1 to 10	1.3 to 5	16 to 50	0.37 to 1.1	Poor	Good
Circular-cage	3.4	10	7	31	3.6	Poor	Good
Box-and-grid	<7	25	13 to 20	57 to 70	<10	Good	Very good
Venetian blind	<7	25	25	60	<10	Good	Poor
Fine mesh	2.5 to 2.7	4 to 6	5	15	<0.45	Good	Poor
Metal channel	0.65 to 1.5	1 to 3	1.5 to 3	4.7 to 8.8	0.4	Good	Good

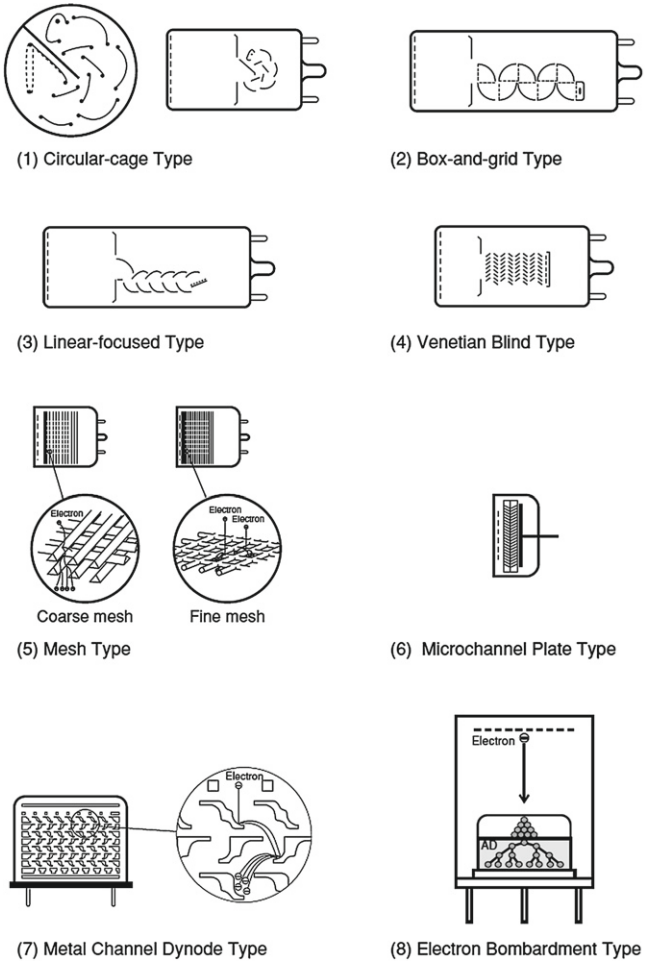

FIGURE 3.4 Typical dynode shapes. *Reproduced with permission from [7].*

delta-function, and has a finite width (sometimes referred to as an instrument response function), characterized by FWHM, $\Delta\tau$. It therefore becomes difficult to distinguish single photons that come at a rate close to or exceeding $1/\Delta\tau$, because the output pulses overlap in time. The second issue is that the overall gain of each photoelectron significantly varies. To first order, the number of secondary electrons released from a dynode obeys a Poisson distribution with an average number of secondary electrons equal to δ, the secondary-electron multiplication factor. This process is repeated as many times as there are dynodes. Thus, the uncertainty in the number of secondary electrons gives rise to a rather wide distribution of output pulse heights, making it impossible

to reliably discriminate between a detection event due to absorption of just one photon versus that due to absorption of multiple photons. Given these complications, a typical use of photomultipliers (and many other detectors, as we will see in the following chapters of the book) is to distinguish between an input with no photons from that of one or more photons. Such detectors are referred to as *non-photon-number resolving* (or, sometimes, as *click/no-click* detectors). Thus, a click/no-click detector distinguishes events when one or more photons have been collected from when no photons have been detected with an exceptional signal-to-noise ratio. The leading edge of an output photoelectronic current provides a reliable (to within the TTS) "timestamp" for a detection event, useful in many applications, particularly for quantum information technology.

To characterize the performance of a photomultiplier in the photon-counting regime, a pulse-height distribution is measured. A typical pulse-height distribution of a photomultiplier is shown in Fig. 3.5a. This pulse-height distribution is measured with and without an input optical signal. The latter measurement can be used to characterize the so-called dark current. The origin of dark current is typically thermal-electron emission that occurs on the photocathode and the dynodes. Obviously, the thermal electrons from the dynodes go through fewer stages of amplification, resulting in pulses with lower amplitude than those due to photoelectrons.

Typically, a photomultiplier output is analyzed with a discriminator set at some threshold level L. Once the electrical output exceeds the threshold, a detection occurs. By setting an appropriate threshold, most spikes of current consistent with the amplification of thermal electrons that escaped from dynodes (especially those dynodes that are closer to the anode) may be filtered out. However, for any reasonable value of L, some dark-current events, particularly these originated from a photocathode and first few dynodes, will still exceed the threshold, and get counted. Detection events due to dark current are called *dark counts*. Obviously, given a detection event, it is generally impossible to determine for sure if that event is a dark count, or it is due to a true detection of a photon. One can only determine the probability that a detection event is (or is not) a dark count, based on the light level and the average number of dark counts per second. As we see in Fig. 3.5a, the dark count rate decreases as the level of threshold L increases. At the same time, a large fraction of legitimate counts are rejected when that threshold level is set too high.

The overall number of counted events for each measurement (with and without optical input) as a function of threshold level is presented in Fig. 3.5b. This figure is obtained by the integration of a histogram in Fig. 3.5a from L to ∞, as a direct consequence of a threshold detection. This trivial observation leads us to an important conclusion: while intrinsic *quantum efficiency* of the photosensitive material (in this case, a photocathode) is a constant value for a given material, *detection efficiency* of the device is not: by changing the detection threshold the overall detection efficiency is changed, as can be understood from

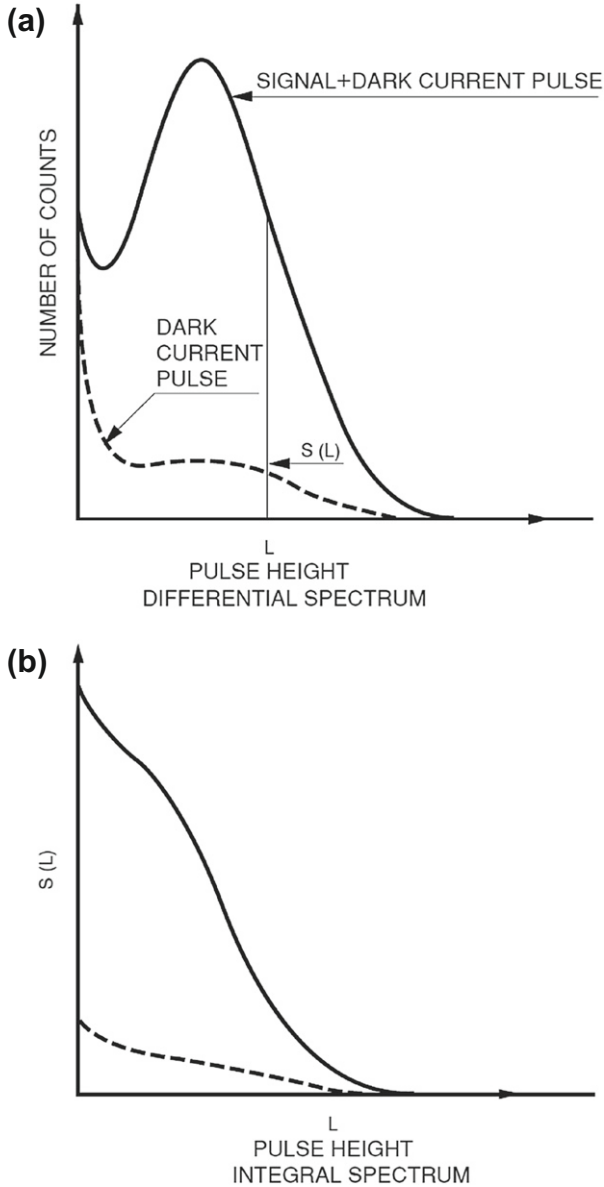

FIGURE 3.5 (a) Pulse-height histogram for characterization of a PMT; (b) total detected pulses of a PMT vs. a level for a discriminating threshold (integrated pulse-height histogram). Depending on where a discriminating threshold is set, a certain detection efficiency and dark count rates are achieved. *Reproduced with permission from [7].*

Fig. 3.5b. We see that with the same input optical power the total number of counts $S(L)$ decreases as the threshold L increases, which reduces detection efficiency.

Both dark-current and signal-current pulse-height distributions should be taken into account when selecting an optimal threshold. Obviously, the choice of L is application dependent. One possible application of photon counting is to establish the radiant power of a weak optical source. In this case, the signal-to-noise ratio should be maximized by selecting an appropriate L. Other applications call for maximum detection efficiency. In such cases, L can be lowered to improve detection efficiency at the expense of signal-to-noise ratio. The trade-offs between characteristics similar to those seen in these examples are quite common to other detector technologies, and will be addressed in the following chapters. Note that the ultimate goal for detector development is to find a technical solution which improves *all* characteristics of single-photon detection, as opposed to sacrificing one of the performance metrics in favor of the other.

Another important consequence of using a discriminator is the possibility of the output pulse overlap. If the two pulses are too close to be resolved, only one count will be recorded. That is, immediately after a single-photon detection, the detection system becomes effectively blind to the new incoming photons. The time after an initial detection during which no new input photons can be detected is the *dead time*. As explained in Chapter 2, dead time is a common issue with many different single-photon detectors, and the mechanisms behind dead time differ among the various technologies. In photomultipliers, both pulse length and discrimination threshold affect the dead time. A typical dead time of photomultipliers is on the order of 10 ns, but can be as short as 1 ns. This compares favorably with many other single-photon-detection technologies, whose dead times range from 10s of ns to microseconds. Short dead times are important for measurements at high photon rates. Note here that, in practice, dead time can also be limited by external factors such as pulse -counting electronics, rather than the detector itself.

Due to dead time, when measuring the intensity of a weak classical source (for example, an attenuated laser or an incoherent white light source), the true photon flux is underestimated. This effect becomes noticeable at photon detection rates $\approx 1\,\%$ of the maximal counting rate $f_{\max} = 1/t_{\text{dead}}$, where t_{dead} is the dead time. This saturation effect makes single-photon detectors highly nonlinear. For an uncorrelated (Poisson-like) source, such as a weak coherent light beam, one can readily apply a simple correction factor, which allows estimation of the number of photons that were missed during the dead time: $N_{\text{true}} = N_{\text{measured}}/(1 - N_{\text{measured}} t_{\text{dead}})$. This correction is good (to within 1 %) up to $f_{\max}/10$. When the source of light is correlated (such as a single-mode thermal source or a non-classical source of light), the formula above cannot be used.

The last issue covered in this section is *afterpulsing*. Afterpulses are false detection events that occur after a prior detection event. Afterpulses are common to many single-photon detectors, but are caused by different underlying effects

for different types of detectors. In PMTs, there are two effects that cause afterpulses. The first mechanism is elastic scattering of the electrons on the first dynode. Usually, the characteristic time delay for this kind of afterpulses is short: on the order of a nanosecond. For most photon-counting PMTs this is not a problem, because such an afterpulse is generated in the immediate vicinity of the event that induced it, and this afterpulse is not detected by the electronics because of the dead time. There are, however, afterpulses that are delayed with respect to the initial pulse long beyond the dead time. These are caused by the ionization of residual gas molecules in the tube. The positive ions may return to the photocathode and knock out electrons that are amplified as if they are legitimate events, and cause false counts. This process occurs over a time range scaling from several hundreds of nanoseconds to tens of microseconds, which may create serious problems for photon counting, and particularly to precision measurements with PMTs, because it is impossible to attribute particular detections to a true detection versus an afterpulse (c.f., dark counts). Instead, a statistical approach is used: the probability of an afterpulse may be estimated, and a statistical correction can be applied.

As with many single-photon detectors, the properties of PMTs vary not only among the different types of PMTs, but also from one device to another. Moreover, they may change depending on the age of the particular PMT, voltage used as an accelerating potential, history of previous light exposure, and other factors.

3.5 CONCLUSION

In this chapter, we have described the first photon-counting technology, invented over 80 years ago. Remarkably, this technology is still used widely today. In addition to reviewing how PMTs function, we saw how physical properties of the device (instrument response function, quantum efficiency, etc.) result in a set of properties that are specific to the photon-counting detectors (detection efficiency, dark count rate, dead time, afterpulsing). These properties are universal, applicable to photon-counting detectors with other technologies. In the following chapters, other technologies of single-photon detection will be introduced.

REFERENCES

[1] H. Hertz, Ann. Phys. 267, 983–1000 (1887).
[2] A. Einstein, Ann. Phys. 17, 132 (1905).
[3] L. Kubetsky, USSR, Author's Certificate No. 24040, Priority date 4 August 1930 (1930).
[4] V. Zworykin, G. Morton, and L. Malter, Proc. IRE 24, 351 (1936).
[5] L. Kubetsky, Proc. IRE 25, 421 (1937).
[6] P. Gorlich, Z. Angew. Phys. 101, 335 (1936).
[7] Photomultiplier Tubes, Basics and Applications, 3rd ed., Hamamatsu Photonics K.K. (2007).
[8] M. Niigaki, T. Hirohata, T. Suzuki, H. Kan, and T. Hiruma, Appl. Phys. Lett. 71, 27 (1997).
[9] K. Nakamura and H. Kyushima, Jpn. J. Appl. Phys. 67, 5 (1998).