PHOTOMULTIPLIERS

SILVANO DONATI
University of Pavia
Pavia, Italy

Despite being invented more than 80 years ago, the photomultiplier is still unrivaled, even by the most recent semiconductor photodetectors, in its ability to detect weak optical signals down to the single-photon level. The photomultiplier tube (PMT) readily provides a wide dynamic range and an excellent linearity, with a good speed of response and a very low excess noise. PMTs are made available from manufacturers in a variety of formats (from <1 cm outlines to large sensitive areas, up to $50 \times 50 \, cm^2$). Also, an ample choice of spectral responses exists, ranging from deep ultraviolet (UVC, down to $\lambda \approx 100 \, nm$) to near infrared (NIR, up to $\lambda \approx 1700 \, nm$ with the new NEA photocathodes).

A disadvantage of PMTs is the relatively bulky and fragile structure; however, this is mitigated in the most recent design incorporating an integrated multiplier chain based on the MCP (multichannel plates) technology. In addition, the high voltage required to supply the tube is a hindrance, but units with a built-in voltage multiplier offer an adequate solution.

Thus, in a large number of applications to instrumentation, PMTs offer the most viable and satisfactory solution to measurements on light signals. Just to quote a few, PMTs are recognized as the working horse for measurements of fast-pulsed waveforms, energy spectrometry, decay-times, time-of-flight, particle and radiation energy, dating with radionuclides, and photon counting.

This article is structured as follow. After a short introduction to PMT operation (Sect. 1), in Sect. 2 we analyze the elements of the PMT — photocathode, electron optics, and multiplying chain and related parameters. Section 3 is devoted to summarize the theoretical results of PMT response. Then, we discuss specifications and characteristics of PMTs (Sect. 4). Finally, in Section 5 we review the most widespread applications of PMTs to measurements and the performance we can obtain.

1. OVERVIEW OF PMT OPERATION

The basic elements and functions in a PMT can be described with reference to Fig. 1 as the following (1):

- Photocathode (either in transmission or in reflection) to convert the photon flux of the incoming optical signal to a photoelectron flux.
- Electron optics, to collect and focus the photoelectrons on the first dynode of the multiplier chain.
- Dynode chain, to provide a high gain of multiplication in the number of photoelectrons.
- Anode, to collect the amplified charge and serve as the output electrode.

Applied to photocathode and first dynode, as well as across successive dynodes, one typically finds a voltage of 100–200 V, which is adequate for electrons to hit the next electrode with enough energy to be multiplied, and to ensure high collection efficiency.

When a primary photoelectron hits the first dynode, it produces the emission of g electrons because of secondary emission. Secondary electrons leaving the dynode are accelerated to the next dynode, producing g^2 electrons. The process continues up to the last dynode, so that for n dynodes, we find g^n electrons at the anode, which is the gain of the dynode chain, $G = g^n$, a very high quantity even for small g: For example, for $g = 3.5$, with $n = 8$–14 dynodes, the gain goes up to $G = 2.2 . 10^4 \ldots 4 . 10^7$, respectively.

At a voltage $V_{ak} = 100 \ldots 200 \, V$ between electrodes, the time-of-flight between dynodes separated by $d_{[cm]}$ can be calculated from expression:

$$\tau_d = d(2m/eV_{ak})^{1/2} = 33.7_{[ns]} \cdot d_{[cm]}(V_{ak})^{-1/2},$$

and the typical result is $\tau_d \approx 3 \, ns$. Then, after $n = 10$ dynodes, we get a total transit time $T_{del} = n\tau_d = 30 \, ns$, which is the *delay* of the SER (single-electron response), the current pulse obtained at the anode following a single photoelectron released at the photocathode. Another even more important quantity is the *duration* of the SER, that comes about from the transit time spread, $\Delta\tau_d$, because of nonidealities of the geometry. A typical value is $\tau_{pulse} \approx 2 \ldots 3 \, ns$ in an 8–12-stage PMT.

The peak current of the SER is found (approximately) by dividing the total charge eg^n by the duration $\Delta\tau_d$. Thus, we get: $i_p = eg^n/\Delta\tau_d = 1.6 \cdot 10^{-19} \cdot 4 \cdot 10^7/2 \cdot 10^{-9} = 3.2 \, mA$, a current large enough for an easy handling by electronic circuits. Thus, the individual current pulses can be watched at a PMT output by an oscilloscope (as shown in Fig. 2) are just the single photons that have been detected by our sensor.

2. BASIC ELEMENTS OF THE PMT

In this section, the principle of operation and performance of elements constituting the PMT is described, then the most commonly employed structures of the PMT as well as some recent variants of them are reviewed.

2.1. Photocathode

The photocathode is the thin film on which incoming photons impinge, at the entrance window of the PMT. By the photoelectric effect, the photons energy is traded for an electron-hole pair, and electrons (or photoelectrons) are emitted in the vacuum where they become available for multiplication at the dynode chain.

According to the incidence surface, *reflection* (or opaque) photocathodes are distinguished when electrons come out from the same side of photons incidence; and *transmission* (or semitransparent) photocathodes are distinguished when electrons are emitted from the surface side opposite to it (2).
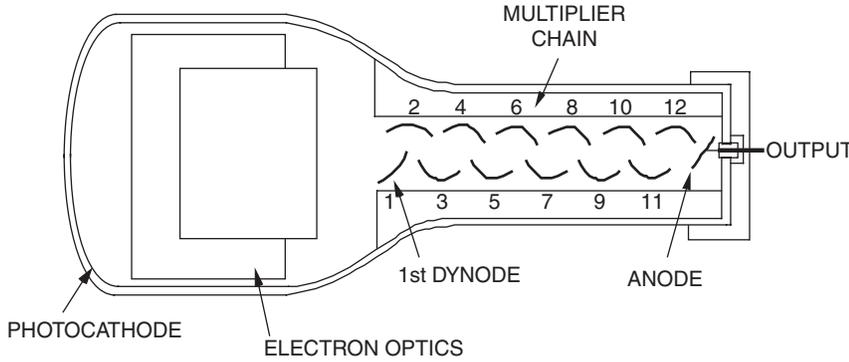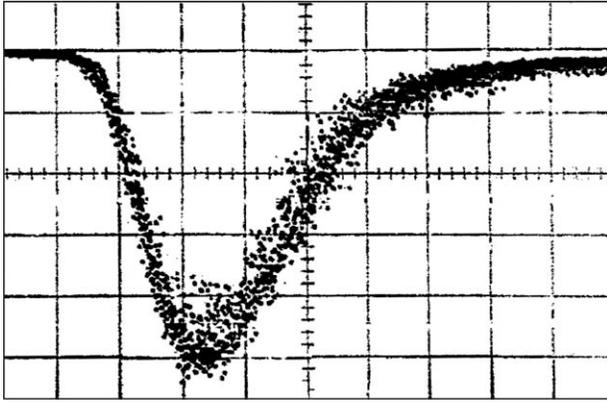
Figure 2. Single-electron-response (SER) current pulse at the anode of a 12-dynode PMT, as seen at the sampling oscilloscope (time scale: 1 ns/div, amplitude scale: 0.5 mA/div).

Photoemission is a crucial phenomenon in PMT. It can be described as the sequence of (1) photon absorption, and generation of an electron-hole pair; (2) diffusion of the electron to the surface; and (3) emission of the electron in the vacuum.

In a simple picture, the energy levels of a reflection photocathode can be represented as in Fig. 3 (1). Conduction and valence bands are separated by the energy gap E_g, the vacuum level is the minimum electron energy to leave the material, whereas the pair production E_p is the level at which ionization yields new electron-hole pairs. The work function and the activation energy are indicated by E_W and E_A.

Internal to the photocathode, optical power decays exponentially according to the Lambert–Beer law: $P(z) = P_0 \exp{-\alpha z}$. Here, $\alpha = \alpha(\lambda)$ is an absorption coefficient specific of the material, and $L_{abs} = 1/\alpha$ is the attenuation length representing the 1/e depth of penetration.

For the photoelectron to be emitted, two conditions must be satisfied. First, the energy condition, by which the generated electron must be raised to at least at the vacuum energy level to be able overcome the potential barrier and leave the material, must be satisfied. In terms of photon energy $h\nu$, the energy condition is $h\nu \geq E_g + E_A$, and letting $\nu = c/\lambda$, and expressing λ in μm and energy in

eV, we get the *photoemission threshold wavelength* λ_t as

$$\lambda_t = hc/(E_g + E_A) \text{ or, } \lambda_{t[\mu m]} = 1.24/E_{[eV]}.$$

The second condition deals with the path traveled by the electron to reach the surface. The path is a random zigzag because of scattering within the atomic lattice.

Scattering can be quasi-elastic, which causes a small energy loss (cases b and d in Fig. 3) and a long free path L_{dif}, or it is anelastic (or ionizing), with a large loss ($>E_g$) and a short free path L_{dif} as the electron is above E_p (case c material Fig. 3). For a high escape probability, it is required that L_{dif} is larger than the absorption length in the, or $L_{dif} > 1/\alpha$.

Last, to obtain a wide spectral range of response, the pair level Ep must always be higher than the vacuum level, or $E_p \gg E_A$

The picture of band diagram (Fig. 3) can be refined to include surface states, which are because of unintentional doping or to defects in the first atomic, and are responsible of band bending at the surface. Bending is upward if the defects are *p*-type on a *n*-type substrate and is downward if the defects are *n*-type on a *p*-type substrate (Fig. 4). In the last case, we get a NEA (negative electron affinity) photocathode, in which the barrier E_A at the surface is virtually lowered to zero or negative. Photogenerated electrons can tunnel from conduction band to vacuum if the barrier thickness is small (a few atomic layers). Thus, in all photocathodes, the final cesiation plays the important role of helping saturate defects and lower the surface barrier. NEA photocathodes have been demonstrated in GaAs (gallium arsenide) treated with cesium oxide, $GaAs:Cs_2O$, with response up to 900 nm, and in $In_{1-x}Ga_xAs:Cs$ (indium gallium arsenide), with photoresponse up to 1700 nm (3).

In conclusion, desirable properties for photocathode materials are: (1) work function $E_W = E_g + E_A$ comparable with photon energy hν in the spectral range of interest; (2) absorption length $1/\alpha$ less or comparable to the scattering length L_{sc} (for high η); (3) the largest possible pair-production energy E_p (for wide spectral response).

Early photocathodes were metallic, with work functions of several eV's and thresholds in the UV. In the 1930s, with the discovery of alkaline antimonide, efficiency of a few percent in the visible was obtained. Additional progress in the 1960s, with bi- and tri-alkaline
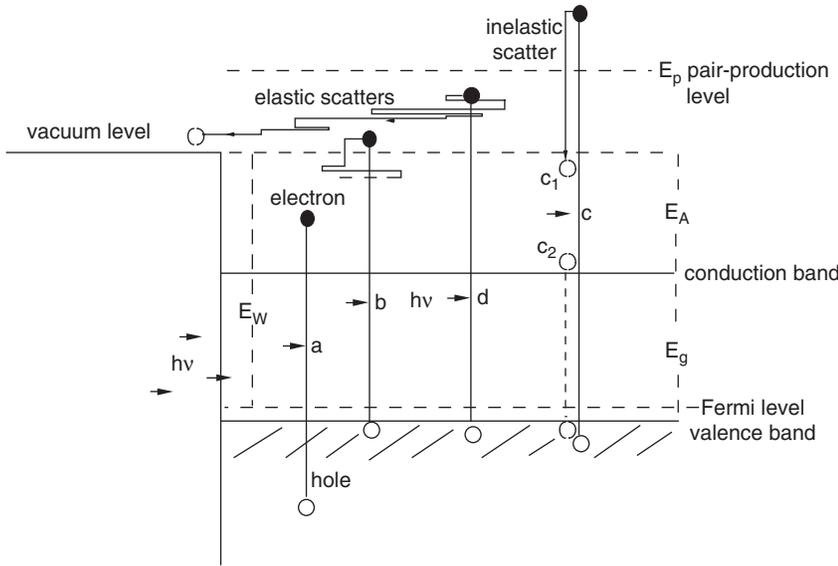
Figure 3. Energy levels in the photocathode. Conduction and valence bands are separated by the energy gap E_g, and E_A is the activation energy to reach the vacuum level. Condition $h\nu \geq E_g + E_A$ holds for the photons energy to produce a photoelectron. Not all the generated electrons can diffuse to the surface and escape. Electron d is emitted because generated not far from the surface and with enough energy to withstand elastic scattering, other electrons have too small an energy (like a and b) or are generated too deep in the material. Electron c from an energetic photon is raised above the pair-production energy, where a fast inelastic scatter that produces a new electron follows (c_2), but neither c_1 nor c_2 can be emitted. (Adapted from Ref. 1, with permission of the Publisher).
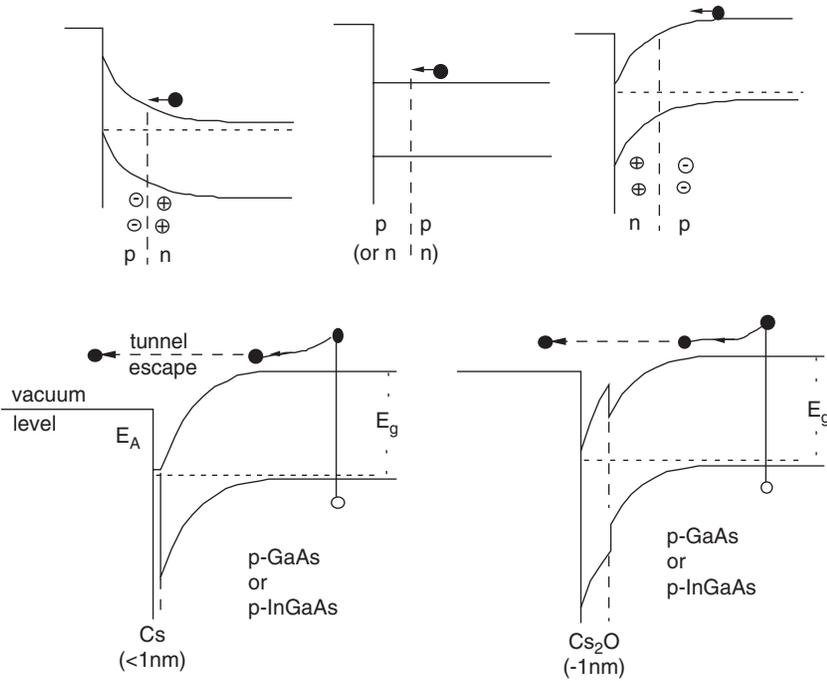
Figure 4. Band bending at the photocathode surface: a p-type surface layer atop on an n-type substrate curves the bands upward, hindering emission. An n-type layer on a p-type substrate curves the bandwidths downward, easing the emission of the electron from the conduction band to the vacuum by tunneling. The effect, called NEA (negative electron affinity), is taken to advantage in cesiated photocathodes.

antimonide, extended thresholds to near IR (800–900 nm). In the 1970s, the NEA concept was introduced, and long-wavelength response improved marginally. In recent years, the practicable threshold for PMTs has reached the 1700 nm of the third window of optical fibers.

The absorption condition favors materials with large α on a wide spectral range, such as semiconductors with a direct-bandgap structure, whereas in indirect-bandgap materials, like Si and Ge, a lattice vibration phonon is required to conserve momentum in the photon absorption. Then, absorption has a much smaller probability per unit length (or α) and efficiency is very low.

In applications, the photocathode response is characterized by the *spectral sensitivity* $\sigma = \sigma(\lambda)$, defined as the ratio of the photoemitted current I to the incident radiant power P: $\sigma(\lambda) = \text{I/P}$. One can also use the *quantum efficiency* $\eta = \eta(\lambda)$, given by the ratio F/F' of the photoemitted electrons rate, to received photons rate. Writing the detected current as $\text{I} = e\text{F}$ and the radiant power as $\text{P} = \text{F'}h\nu$ the ratio $\alpha = \text{I/P}$ follows as:

$$\sigma(\lambda) = \eta e/h\nu = \eta\lambda/(hc/e) = \eta(\lambda/1.24)\,[\text{A/W}].$$

2.1.1. Properties of Common Photocathodes. The most important group of materials is the alkaline metals antimonide. In single-metal antimonides such as Cs_3Sb, K_3Sb and Na_3Sb, both peak quantum efficiency η_{max} and threshold λ_t increase with the atomic weight (i.e., in the

order Li-Na-K-Rb-Cs) where as pair-production threshold E_p decreases (and the UV response is reduced).

Cesium antimonide Cs_3Sb has a spectral response (designated as S-11) close to the eye when deposited on a borosilicate-glass window. On fused silica, it goes deep in the UV, to 160 nm (S-4, S-5). Also extended to UV are: cesium telluride (Cs_2Te) with cutoff at 180 nm and cesium iodide (CsI) reaching 120 nm on LiF or MgF windows. These photocathodes with a threshold λ_t smaller than visible are called *solar-blind* and generally withstand (as opposed to all other photocathodes) natural illumination without damage.

Bi-alkaline photocathodes (CsK_2Sb, Na_2KSb, $CsRb_2Sb$) have a little bit wider photoresponses, and low dark currents (especially Na_2KSb) that make them attractive for single-photon counting.

Tri-alkaline photocathodes (or multi-alkali) Na_2KSb:Cs has a still larger η_{max}, and a response extended up to $\approx 800\,nm$ and to $\approx 850\,nm$ in the so-called S-25 or S-20 extended-red-multi-alkali (ERMA). Using a fused silica window, it reaches 160 nm in the UV. For the wide spectral range, the S-25 photocathode is very commonly used in spectrometry and wide-band applications.

Although sensitivity is poor ($\eta < 1\%$), the old silver-cesium-oxide Ag-O-Cs (S-1 response) is still important in low-cost applications of IR up to 1100 nm. Last, NEA photocathodes in GaAs and ternary compounds offer an improvement with respect to tri-alkaline in the near IR up to about 900 nm.

A conductive underlayer is necessary in transmission photocathodes, especially those with high resistivity like Na_2KSb. The ohmic-contact layer shall be transparent, and even very thin metal layers cannot be used (with the exception of W in the UV).

Tin oxide and indium oxide (or ITO, indium-tin oxide) are the best suited for ohmic-contact and yield a satisfactory resistance (typically $1\,k\Omega$) with transmission of $\approx 95\%$.

Figure 5 shows the spectral responses of popular photocathodes in different materials. In the visible and near-infrared, alkaline photocathodes are best suited because of the high spectral sensitivity (up to 100 mA/W peak) and quantum efficiency (up to 40%). Transmission and reflection photocathodes have nearly the same shape of spectral response, but with ≈ 20–50% larger $\sigma(\lambda)$ and a more extended long-λ response in reflection types. In contrast, transmission photocathodes have the access window coincident with the photosensitive surface and can collect more light (i.e., use an objective lens with larger numerical aperture) (1,2,4).

2.1.2. Photocathode Technology. Photocathode technology is a thin-film, high-vacuum process (1,4). Because of alkali metals reactivity to oxygen, the photocathode is fabricated under a high vacuum (typ.10^{-6}torr) condition. The glass envelope is fabricated first, outgassing it at $400^{\circ}C$ for several hours to eliminate adsorbed gas from the walls. Glasses free from Pb, Ni, W are used, because these metals react with Cs and produce compounds of low work function that increase the dark current. Alkali-metal vapors are produced by reduction of a salt of the metal. A typical mixture is made by one part of cesium chromate ($CsCrO_4$) and two parts of Si or Al as the reducing agent. The mixture is brought to $700^{\circ}C$ by radio-frequency heating. Deposition rate is controlled by the temperature and duration of the process.

For multicomponent photocathodes, the metal is made to react with other components, by heating the glass envelope to 140–180°. The process is controlled by visual inspection (aspect and color) and by measuring the photocurrent in the final steps of fabrication.

2.1.2.1. Monoalkali Photocathodes. The prototype is CsSb, the best known photocathode produced since the 1930s. As resistivity is high, a sublayer of SnO or ITO (indium-tin oxide) is first formed as the entrance window. First, a film of antimonium is the deposited, whose thickness is tuned to an optimal value of $\approx 6\,nm$ (≈ 20 atomic layers) by looking at a decrease of optical transmission, to $\approx 75\%$ of the initial value. Then, Cs is evaporated as the envelope is kept at $130^{\circ}C$ to promote reaction with Sb. Evidence of the reaction is the appearance changing from metallic to reddish. The photoresponse gradually increases up to a peak beyond which irreversible decrease occurs. At the peak, the evaporation of cesium is blocked and, on cooling, a small quantity of oxygen is admitted for a final improvement of photoresponse near the threshold.

2.1.2.2. Other Monoalkali Photocathodes. Photocathodes prepared like Cs_3Sb are the antimonide of Rb, K, Na, and Li, whose work functions of, respectively, 2.09, 2.22, 2.29, and 2.50 eV, give a response progressively shifted toward the blue and UV, with gradually decreasing quantum efficiencies.

2.1.2.3. Bi- and Tri-Alkali Photocathodes. These photocathodes are more critical to fabricate, and call for an accurate control of composition to achieve the best sensitivity. In the bi-alkali Na_2KSb, sodium and potassium are in 1:2 ratio, with a slight excess of Sb for the p-type doping. The Na_2KSb photocathode is prepared by reacting K with Sb to build up a layer of K_3Sb. Then, Na vapor is admitted and heated as the crystalline structure changes from hexagonal to cubic.

Adding small quantities of oxygen or cesium to the bi-alkali layer, the work function is further lowered and the threshold is improved. Thus, the spectral responses of S-20 and S-20 ERMA are process variants compared with the Na_2KSb rather than composition variants.

2.1.2.4. Cesiated Silver-Oxide Photocathode. This very old and still used photocathode is obtained by first evaporating a thin layer of Ag, typically 10-20 nm (50% optical transmission), and oxidizing it by sputtering with oxygen ions in a low-pressure discharge until the transmissions returns to 100% (Ag_2O is transparent). Finally, the layer is cesiated at $130^{\circ}C$ until the response is maximized. The photoemission mechanism appears to be promoted by Cs_2O for the peak in the blue, and to metallic Ag for the IR peak, with tunneling from Ag to Cs_2O, with possible nanoparticle effects.
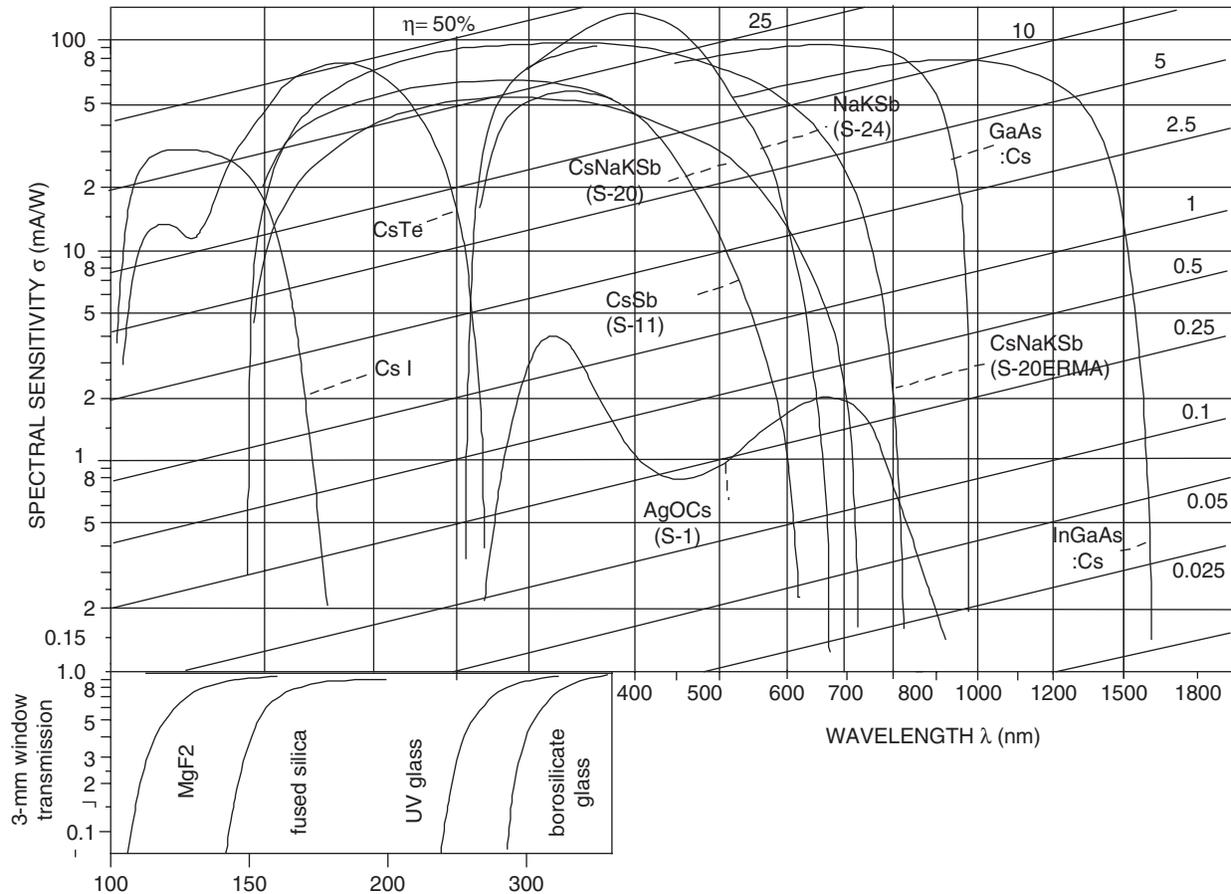
Figure 5. Spectral sensitivity $\sigma(\lambda)$ vs. wavelength of reflection photocathodes. Lines of constant quantum efficiency η are also shown. The EIA standard designations are indicated in brackets. Inset shows the cutoff of common window materials (adapted from Ref. 1, with permission of Prentice-Hall).

2.1.2.5. NEA Photocathodes. The most recent (1980) photocathodes, still progressing to extend single-photon techniques in the IR. Different from common photocathodes, they are based on single-crystal materials grown with the modern technologies of semiconductors.

Best known are gallium arsenide (GaAs) and the ternary compounds $Ga_xIn_{1-x}As$, which allow tuning of the bandgap energy through the control of composition x of the constituents in the III-V crystal lattice cell. All photocathodes are cesiated to take advantage of band bending at the surface, thus the NEA name. Substrate for the photocathode is a relatively thick wafer of the material, thus all NEA photocathodes are of the *reflection* type, although no reason against transmission types exists.

With the recent progress in ternary compounds grown on InP substrates and of bandgap engineering technologies, it appears feasible to develop long-λ photocathodes with responses up to the middle infrared. Indeed, a photocathodes with 1700-nm cutoff are already available, and further work should be able to provide substantial extension of the IR coverage.

2.2. The Electron Optics

The electron optics is placed between photocathode and first dynode (Fig. 1), in form of one or more electrodes properly shaped and kept at increasing voltages (2,4). Its scope is to collect all of the photoelectrons leaving the photocathode and focus them on the much smaller first dynode surface, with high efficiency and the smallest time spread. The collection efficiency ηc is important because it combines with the quantum efficiency η_{ph} of the photocathode to give the effective electron/photon ratio $\eta_{ph}\eta_c$. The time spread $\Delta\tau_d$ of the photocathode to first dynode flight is the most important factor affecting the speed of response.

2.3. The Electron Multiplier

In secondary emission, a primary electron of some 100–200 eV energy excites the crystal lattice, raising electrons to high-energy states favorable for emission. As a result of the relatively high energy, ionization is multiple and produces several electron-hole pairs.

For a secondary electron to be emitted, it shall have an energy $E_s \geq E_g + E_A$ (1–2 eV) sufficient to overcome the potential barrier at the surface, and a diffusion length larger
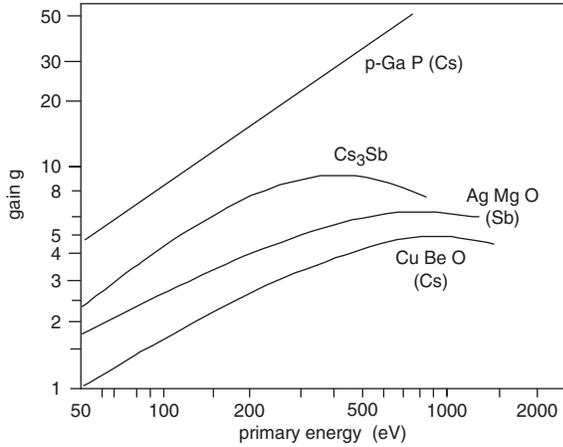
Figure 6. Secondary emission gain of some dynode material as a function of the energy of the impinging primary electron (or interdynode voltage)

than the depth of generation from the surface, much alike in photoemission.

Should the primary energy E be fully used, we get a number of $N_{max} = E/(E_g + E_A)$, typically 50–100, which is never approached because: (1) several electrons are produced with a too low energy (between E_g and $E_g + E_A$) or little excess energy; (2) energy is wasted in quasi-elastic scattering with the lattice. As the energy E of the primary electron is increased, the point of generation in the material is deeper, and when depth approaches diffusion length, the probability of emission starts decreasing and a maximum of the multiplication gain is found.

As a result of the random aspect of the multiplication process, all the above pictures apply to the mean values of quantities. For example, the gain g has the meaning of an average number of secondary obtained at the output for one primary electron on the dynode. The probability distribution p(n) of the number n of secondary generated by a primary electron is found experimentally to follow the Poisson distribution $p(n) = g^n \, e^{-g} /n!$, from which the mean value $\langle n \rangle$ and variance $\sigma^2 = \langle \Delta n^2 \rangle$ are $\langle n \rangle = g$ and $\sigma^2 = g$.

The secondary emission gain of commonly used dynode materials is plotted in Fig. 6 as a function of primary energy (or interdynode voltage). As one can see, CsSb has a fairly high gain at low voltages but withstands only $<80°C$ and has a moderate current density ($\approx 200 \, \mu A/cm^2$). The cesiated oxides of copper-beryllium and Ag-Mg alloys are more resistant to temperature and have a high output current density ($>100 \, mA/cm^2$). Last, cesiated gallium phosphide has a NEA band structure (as in Fig. 4) and a diffusion length of some μm (instead of $\approx 20 \, nm$ of the other materials), which combine to yield a high gain per unit voltage and a much higher saturation level.

2.4. Common Photomultiplier Structures

In Fig. 7, (2) a few preferred structures for the PMT are reported. With reflection photocathodes, the *squirrel-cage* structure (a) is a popular choice and is still used for the

good collection efficiency and compactness. Its limitation is the maximum number of dynodes, ≤ 10.

More difficult to assemble is the venetian-blind structure (c) that, with a large area and a good collection efficiency even without an electron-optics, is fairly compact but not the fastest. The linear-chains (d and e), also called box-and-grid, are often preferred as a good compromise between ease of fabrication, access to the photocathode, good collection area, and overall size.

In all PMTs, the anode electrode is not critical and can be simply made by a thin metal plate. In front of it, one may place a fine grid acting as an electrostatic screen between last dynode and anode, so that the SER duration is shortened. In addition, to avoid extra frequency cutoff from parasitics, in fast PMTs, the anode is made part of a coaxial-like structure ending in the output connector matched to $50 \, \Omega$.

2.4.1. Photocathode-Related Parameters.
2.4.1.1. Spectral Response. Spectral response is specified by the spectral sensitivity, $\sigma = \sigma(\lambda) = I/P$, ratio of photogenerated current I to incident optical power P, or by the quantum efficiency $\eta = \eta(\lambda)$, ratio of emitted photoelectrons to collected photons.

Usually, the individual $\sigma(\lambda)$ of a device is not supplied by manufacturers, and reference is made to average spectral curve as Fig. 5. Sometimes, devices are supplied with the individual *luminous sensitivity* $\sigma_L = I/P_L$ (in A/lm) for a standard source (typically the W ribbon-lamp at 2850 K). Of course, the luminous sensitivity cannot be traced exactly to $\sigma(\lambda)$, but is useful as a relative scale-factor to compare different devices.

2.4.1.2. Temperature Coefficient of the Spectral Sensitivity. This quantity is expressed by the relative variation per unit temperature difference, or $\alpha = (1/\alpha)d\sigma/dT$. Experimentally, α_σ is negative (typically -0.3–$0.5\%/°C$) except near the photoelectric threshold, where it becomes positive and increases fast. A wavelength $\lambda_{\sigma 0}$ is then found at which $\alpha_\sigma = 0$, and this $\lambda_{\sigma 0}$ is slightly above the wavelength of peak response.

2.4.1.3. Dark Current. Of great importance in low-level and counting applications of PMTs, the total dark current is composed of several contributions: (1) dispersion current through the insulating structure; (2) dark current of electrodes other than the photocathode; and (3) photocathode dark current. If the first two terms are properly controlled and made negligible, the ultimate limit of dark current is set by the dark photoelectrons. Their current density is given by Richardson's law: $J_d = (4\pi em/h^3) \, (kT)^2 \exp(-E_W/kT)$, where E_W is the work function.

As $E_W \approx h\nu$ near threshold λ_t, one can also write $J_d = (4\pi em/h^3)(kT)^2 \exp(-hc/\lambda_t kT)$, and this equation shows that dark current increases with increasing threshold λt. Therefore, in photon-counting applications, one will choose the spectral response the least IR-extended to minimize Jd. Figure 8 is a plot of dark current in common PMTs along with the theoretical values from the above equations. As one can see from the diagram, cooling substantially reduces dark current.
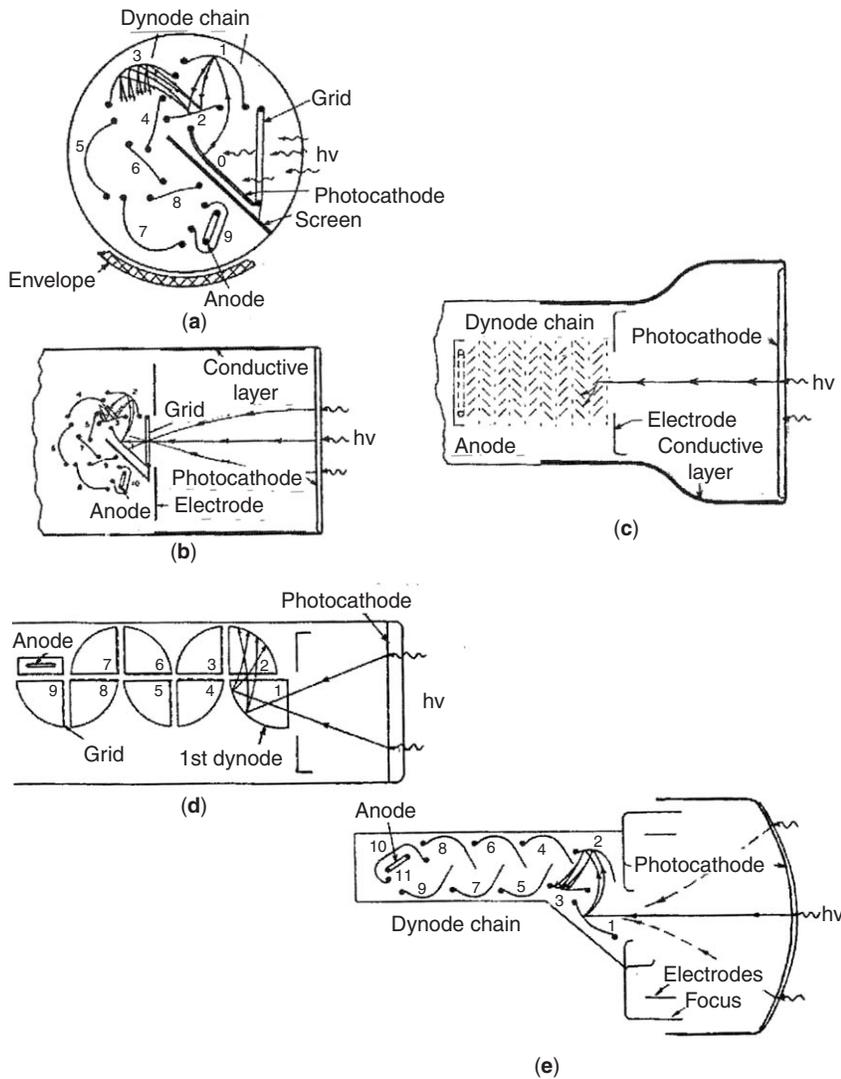
Figure 7. PMT structures: (a) squirrel-cage for reflection photocathodes and (b) for transmission photocathodes; (c) venetian-blind dynodes; (d) box and grid; (e) linear dynode-chain (adapted from Ref. 2, by courtesy of Burle Technologies).

2.4.1.4. Uniformity of Response. In different portions of the photocathode and in image tubes, the $\sigma(\lambda)$ may spread up to $\pm 5\%$ at the peak of response, and up to $\pm 20\%$ near λ_t. Near the edge of the photocathode surface, the uniformity is worse than at the center.

2.4.1.5. Linearity and Saturation. About compliance to the ideal linear $I = \sigma P$ relation, no experimental evidence exists of appreciable linearity error (say, $> 0.1\%$) on a dynamic range from the dark current (pA or fA per cm^2) up to about $10\, mA/cm^2$. At large input power, photocurrent will saturate, at $\approx 100\, mA/cm^2$, because of charge storage effects.

2.4.1.6. Fatigue and Degradation. Photoemission is a nondestructive process in itself, so photocathode and dynodes have a potentially unlimited lifetime. However, several factors lead to fatigue and degradations. Excessive illumination is a source of catastrophic degradation, because power dissipation and heating of photocathode or dynodes (even for few minutes) reactivate the diffusion process with an irreversible alteration of the structure,

especially in cesiated types. When the tube is turned off, excessive illumination may cause fatigue, with a sensitivity loss and a dark current increase, which, however, recovers in 10–100 min after returning to the dark condition.

Residual gas in the tube envelope is another important factor of progressive degradation. Photoelectrons in transit to positive electrodes ionize the residual gas atoms, produce positive ions that are accelerated toward the photocathode, and hit it with high kinetic energy, sufficient to pull away clusters of atoms. This process is the same as sputtering, or cathode evaporation. Thus, fabrication of a PMT starts with a well-outgassed envelope ($\approx 10^{-7}$ torr). When adequately fabricated and correctly used, photocathodes may reach very respectable lifetimes, and MTTF (mean time to failure) of $5 \cdot 10^4$ hours (working) and of $3 \cdot 10^5$ (nonworking) are currently obtained.

2.4.2. Special PMT Structures. The basic PMT structure with photocathode, electron optics, and dynode chain described in the previous section is by far the most common in commercial devices. However, through the years, several variant special types have been developed to improve
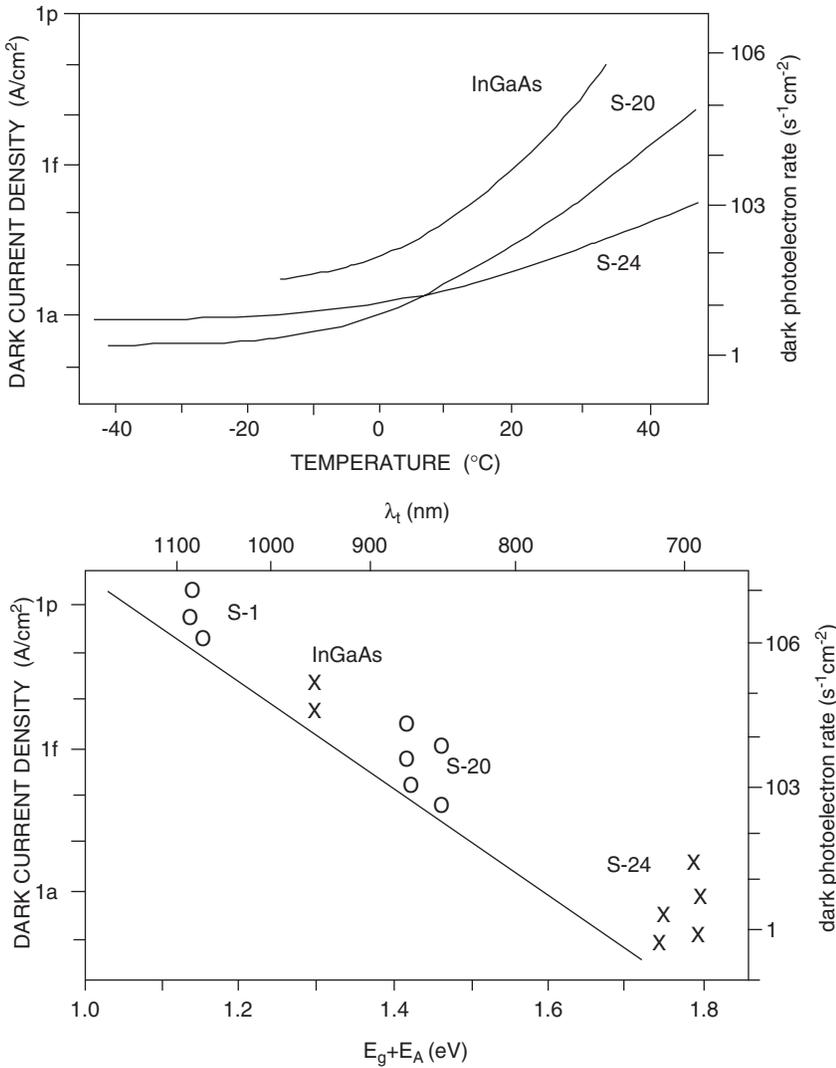
Figure 8. (top) Dark current density and dark photoelectron rate of some PMTs as a function of the work energy. (bottom) The temperature dependence of the dark current density (adapted from Ref. 1 by courtesy of Prentice-Hall).

parameters like speed of response, amplitude statistics, or simply reduce overall size.

2.4.2.1. Miniaturized PMTs. Miniaturized PMTs are not really different devices, but they include all ancillary circuits (bias, dc/dc converter, etc.) and have overall dimensions comparable with those of a solid-state photodiode circuit, yet with the single photon capability of a PMT.

Another class is that incorporating a MCP (microchannel plate) multiplier as the substitute to the dynode chain (Fig. 9). The microchannel is simply a hollow small (0.1-mm diameter.) tube, with the internal walls deposited with a secondary emitter. Electrons entering the microchannel make a zigzag path multiplying themselves at each hit on the walls.

Making a plate of several microchannels put side by side, we get a thin (≈ 5 mm) multiplier chain that, from point to point, is equivalent to as many dynode chains. With respect to a dynode multiplier, now the number of hits (or equivalent stages) is not a constant, which worsens the statistics considerably. However, as the outline size is so strongly reduced, and considering that at the
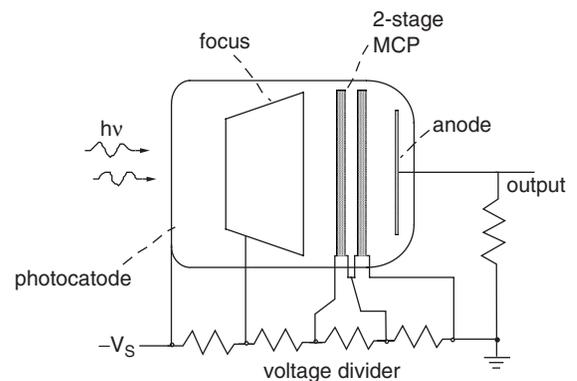
Figure 9. A two-stage MCP photomultiplier.

increased number of primary photoelectrons the statistics return to good values, the MCP-PMT is preferred in several applications.

The typical structure of a MCP PMT is shown in Fig. 9. Two or three multiplying stages are used, conjugated to

one another by proximity focusing, and to the photocathode by an eventual electron optics. Compared with a single-high gain stage, the two-stage is preferred because it screens better the region of high and low electron density, preventing positive feedback from ionization of residual gas. Ionization is a source of feedback because the positive ions reach the multiplier input and generate new electrons.

Instability then sets in for example, if the PMT gain is 10^8 and barely 10^{-8} ions/electron go back and ionize at the input. This phenomenon is aggravated, in a glass microchannel PMT, by the tendency of MCP to release, in the useful life, gas molecules adsorbed on its surface when bombarded by ions. With a suitable stagger, opposite in the adjacent stages (chevron-type, Fig. 9), ions produced in the second stage cannot reach the input of the first. So, if the gain per stage is kept low (for example, $<10^4$), the ion feedback effect becomes negligible.

MCP PMTs have typical multiplication gains of 10^4 (1-stage), 10^6 (2-stages), and 10^8 (3-stages). With interelectrode voltages of about $1000\,V$, the SER duration is $\approx 500\,ps$ with rise times in the range 100–200 ps. A typical value of the multiplier variance is $\varepsilon_{Amc}^2 = 5 - 20$, considerably worse than in dynode PMTs, but which can be tolerated when the number of detected photoelectrons is large.

Another advantage of the MCP-PMT structure is the electrostatic conjugation of input/output faceplate, because electrons emitted from a certain photocathode pixel land to a corresponding pixel on the anode. Now, if the anode is segmented into a number of elements separately accessible from as many external electrodes, a *multianode*-MCP-PMT, capable of high gain and mltiple readout of the different areas of the photocathode, is obtained.

3. THE PMT RESPONSE, GAIN, AND NOISE

In this section, the main responses describing PMT are reviewed. The treatment is statistical to assess accuracy of the measurements performed by the PMT. The main results are reported without derivations, but the interested reader may find exhaustive treatment in Refs. 1 and 5. The responses to consider are:

– *Integral (or charge) response*. The number N of electrons collected at the output, after a photoelectron impinge on the first dynode or a photon is detected by the photocathode, is a random variable characterized by its mean value $\langle N \rangle$ and variance σ_N^2.

The multiplication process is described by the Poisson statistics. At the ith dynode, an individual primary electron is multiplied to m secondary electrons leaving the dynode, and mean and variance of the random number m are $\langle m \rangle = g_i$, and $\sigma_m^2 = g_i$. When the packet of electrons leaving the first dynode is tracked up to the last dynode and the anode collects the N electrons, we get for the mean

and variance of the number N:

$$\langle N \rangle = g_1 g_2 \cdots g_n,$$

$$\sigma_N^2 = \sum_{i=1,n} (g_1 g_2 \cdots g_i)(g_{i+1} \cdots g_n)^2.$$

These expressions can be justified by simple arguments. It is reasonable that the mean number is the product of the mean gain at each dynode. About variance, it is the sum of n contributions from the dynodes, of the form $(g_1 g_2 \ldots g_i)(g_{i+1} \ldots g_n)^2$. Indeed, at the ith dynode, one finds an average of $g_1 g_2 \ldots g_{i-1}$ electrons arriving from the preceding dynode, each being multiplied by g_i and adding a variance contribution g_i. The quadratic fluctuation goes multiplied by the squared gain incurred between $(i+1)$th dynode and anode, or by $(g_{i+1} \ldots g_n)^2$, and the above expression for σ_N^2 follows.

Taking all the dynode gains as equal except the first, that is, $g_2 = g_3 = \ldots = g_n = g$, the equations can also be written as:

$$\langle N \rangle = g^n,$$

$$\sigma_N^2 = \langle N \rangle^2 (1/g_1)[1 + 1/g + 1/g^2 + \cdots + 1/g^{n-1}]$$

$$\approx \langle N \rangle^2/[g_1(1 - 1/g)]$$

$$= \approx \langle N \rangle^2 g/g_1(g - 1).$$

The quantity:

$$\varepsilon_A^2 = g/g_1(g - 1) = \sigma_N^2/\langle N \rangle^2$$

is relative variance of the dynode chain and is called the *variance of the multiplier*. The $1/g_1$-dependence shows that first dynode is the most important in determining ε_A^2, where the others add less fluctuation, as their contribution is summarized by a factor $g/(g-1)$ not far from unity [it is, for example, $g/(g-1) = 1.33$ for $g = 4$].

As the second step, consider a light pulse on the photocathode containing F photons. Mean and variance of the number of electrons NF collected at the anode is found (1) as $\langle N_F \rangle = \langle \eta F \rangle \langle N \rangle$ and the variance is $\sigma_{NF}^2 = \langle \eta F \rangle \sigma_N^2 + \sigma_{R^2} \langle N \rangle^2 = \langle N_F \rangle^2 (1 + \varepsilon_{A^2})/\langle \eta F \rangle$

Comparing with the relative variance of the F-photons packet, (i.e., $\sigma_F^2/\langle F \rangle^2 = 1/\langle F \rangle$), one can see that the PMT has a noise figure given by $NF^2 = [\sigma_{NF}/\langle N_F \rangle^2]/[\sigma_F^2/\langle F \rangle^2] = (1 + \varepsilon_A^2)/\eta$

Of the two contributions to NF, one is the $1/\eta$ factor because of the photocathode, we choose one with the maximum η at the signal wavelength; the other $1 + \varepsilon_A^2$ summarizes the dynode multiplication statistics adds noise (term $\varepsilon_A^2 = 0.33$ for $g = 4$).

– *Impulse (or current) response*. For an electron impinging on the first dynode at time $t = 0$, the anode response is the time-dependent function known as the SER (single-electron response, Fig. 2). More in general, following an illumination $\Phi(t)$of the photocathode, one will get a current i(t) out the anode. As

$i(t)$ is a random waveform, it is characterized as the second-order statistic by mean $\langle i(t) \rangle$ and variance $\sigma_i^2(t)$. As $i(t) = en(t)$, we may restrict ourselves to consider the number of electrons $n(t)$.

The process of interdynode flights is schematized by means of the probability density function $f(t)$, where $dp = f(t)dt$ is the differential probability that the time of flight is between t and $t + dt$. It is intuitive that the mean value $n(t)$ is $\langle n(t) \rangle = \langle n_T \rangle f(t)$, where n_T is the random number of secondary electrons, with mean and variance: $\langle n_T \rangle = g$, $\sigma_{n_T}^2 = g$. Compounding the flight-and-multiplication statistics on the n-dynode chain (1), one obtains for the mean value: $\langle i(t) \rangle = SER(t) = g_1 g_2 \ldots g_n f_1(t) * f_2(t) * \ldots * f_n(t)$, where * stands for convolution integral operation. This expression is intuitive, as the n-stage cascade of transfer function has an impulse response given by the convolution of individual impulse responses, and a gain given by the product of individual gains. The variance is (1):

$$\sigma_{SER^2}(t) = \sum_{i=1,n} g_1 \ldots g_i \left[g_{i+1} \ldots g_n \right]^2$$

$$f_1(t) * f_2(t) * \cdots * f_i(t) * [f_{i+1}(t) * \cdots * f_n(t)]^2.$$

This expression is similar to that of the integral response, with weights decreasing with the stage number i, each of them being scaled by the gain g_i in respect to the preceding one.

An important simplification of this equation is obtained by the assumption of *rigid* SER, that is, take the generic term in the summation as equal to the square of the mean waveform, or $f_1(t) * \ldots * f_i(t) * [f_{i+1}(t) * \ldots * f_n(t)]^2 \approx [f_1(t) * f_2(-t) * \ldots * f_n(t)]^2$. Then, the variance becomes the nice result:

$$\sigma_{SER^2}(t) \approx \varepsilon_A^2 \langle SER(t) \rangle^2.$$

The approximation is named rigid SER because it ascribes the fluctuations only to the amplitude of the SER, neglecting shape fluctuations.

One may extend the results to the detection of a light pulse $\Phi(t)$ incident on the photocathode. If F are the photons contained in the pulse, and assume that the number of photons is Poisson distributed, one finds:

$$\langle I(t) \rangle = \langle \eta F \rangle \Phi(t) * f_0(t) * \langle SER(t) \rangle \text{ and}$$

$$\sigma_I^2(t) \approx (1 + \varepsilon_A^2) \langle \eta F \rangle \Phi(t) * f_0(t) * \langle SER(t) \rangle^2.$$

With these expressions, one can calculate the average waveform of the output current $I(t)$ from the PMT as the convolution of the SER(t) response and the waveform of illumination $\Phi(t)$. With a variety of methods, one can also make the inverse calculation, the deconvolution of the SER from the output $I(t)$ to obtain the waveform of illumination $\Phi(t)$ presented at the input. Interesting to note, the accuracy of the inversion is very good for duration of $\Phi(t)$ large or comparable with the SER duration $\Delta\tau$. One can also go down to a fraction of $\Delta\tau$ reaching $\approx 100\,ps$ as

the practical limit of the minimum time interval on which a detail of the $\Phi(t)$ waveform can be resolved.

Fluctuations of the PMT response are characterized by $\sigma_I^2(t)$, which, in the rigid-SER approximation, reduces to the convolution of $\Phi(t)$ and the square of the SER, multiplied by the scale factor $(1 + \varepsilon_A^2)\langle \eta F \rangle$. This factor has to be compared with the square of the mean $\langle I(t) \rangle$ scale factor, (i.e., $\langle \eta F \rangle^2$) yielding for the signal-to-noise ratio $(1 + \varepsilon_A^2)/\langle \eta F \rangle$, once again the same S/N ratio of the charge measurement.

Last, a term $f_0(t)$ is found in mean and variance, the time-of-flight between photocathode and first dynode. This quantity can be neglected when $\Phi(t)$ is slower or comparable with the SER, but not when resolution or the deconvolution calculations is pushed to the lower limits.

- *Frequency Domain Response*. This response, pertaining to the permanent sinusoidal regime, is described by the mean transfer function $F(\omega)$ and the noise spectral density s. Recalling that the transfer function $F(\omega)$ of a device is the Fourier transform of the impulse response $U_\delta(t)$, and that the noise power spectrum $S^2(\omega) = di^2/d\omega$ is the Fourier transform of the autocorrelation $K_U(\tau)$, one can use the results of the previous section and write the frequency response of the multiplying chain as $F(\omega) = \mathscr{F}\{SER(t)\}$, where $\mathscr{F}$ stands for Fourier transform. In the case of a Gaussian waveform, for example, we have $F(\omega) = \langle N \rangle \exp(-\omega^2 n \sigma_t^2/2 + i\omega n t_0)$, and the high-frequency cutoff (3dB point) is at frequency $f_T = 0.83/2\pi\sigma_t\sqrt{n} = 0.13/\sigma_t\sqrt{n}$.

When one detects a mean current $I = I_{ph} + I_b$ at the first dynode input, in other words the sum of a signal I_{ph} and a dark current I_b, the noise power spectral density $S^2(\omega)$ is computed as $S^2(\omega) = 2e\,(I_{ph} + I_d)\,(1 + \varepsilon_A^2)\,|F(\omega)|^2$.

Analogously, the current from excitation of the photocathode is $F(\omega) = \mathscr{F}\{f_0 * \langle SER(t) \rangle\} \approx \mathscr{F}\{\langle SPR(t) \rangle\}$, and for the detection of a mean light power P, the spectral density of noise is $S^2(\omega) = 2e\,(\eta eP/h\nu + I_d)\,(1 + \varepsilon_A^2)\,|F(\omega)|^2$.

As a remark, the PMT is close to the quantum detection regime in its pass band for $I_{ph0} \geq I_d$, (i.e., for an input light power larger than the breakpoint value $P_{i0} \geq (h\nu/\eta e)I_d$). In this case, the signal-to-noise ratio is $(S/N)^2 = I_{ph}/[2eB(1 + \varepsilon_A^2)] = P/[2(h\nu/\eta)B(1 + \varepsilon_A^2)]$ and the noise figure is $NF^2 = (1 + \varepsilon_A^2)/\eta$.

The upper frequency f_T of dynode-chain PMTs is usually located at a few hundreds MHz, and that of MCP-PMTs up to a few GHz, consistent with the SER duration $\Delta\tau$. This performance may look poor, in view of the tens to hundreds GHz attained by some modern semiconductor photodiodes [e.g., the SAM or UTC structures (1)]. However, it has to be noticed that the bandwidth limitation comes just from the dispersion of the multiplying structure, not from the photocathode that has a response time as short as $\approx 0.5\,ps$. Thus, by strongly accelerating the emitted photoelectrons and reading the optical signal by a nondispersion method, one is able to build an ultrafast cousin of the PMT, the so-called *streak-camera tube* (1). Invented in 1935, the streak-camera has been perfected to

resolve details in optical waveforms down to 0.5–2 ps, yet preserving all the features specific of photocathodes, including single-photon capability and the wide spectral response (deep UV to near IR).

– *Time Sorting and Measurements*. As a result of the fast response, PMTs are ideally suited for time-of-flight measurements and temporal localization down to the subnanosecond scale, with important applications in biological diagnostics, spectroscopy, and telemetry. The PMT response is characterized by the delay T between a short light pulse detected at $t = 0$ and the output anode pulse, where $\langle T \rangle$ is the mean delay and the variance σ_T^2.

In principle, temporal localization is performed by a decision circuit, which switches on at a time correlated with the pulse waveform. In a time-of-flight measurement, a start and a stop time-localization exit, the difference of which is the desired time of flight.

The decision circuit is usually a Schmitt discriminator with a switching threshold S_o (Fig. 10), and the sorting signal is generated when the signal crosses S_o at time T_o.

If the amplitude fluctuation $\Delta S(t)$ of the signal around its mean value $\langle S(t) \rangle$ is represented by an uncertainty band $\langle S(t) \rangle + \sigma_S$ and $\langle S(t) \rangle - \sigma_S$, the switching time will likewise exhibit fluctuations ΔT with an uncertainty band represented by $T_o + \sigma_T$ and $To - \sigma_T$. With the reasonable hypothesis of small amplitude fluctuations σ_S around the mean $\langle S(t) \rangle$, one can apply the statistical principle of the linear regression and calculate σ_T as the ratio of amplitude standard deviation σ_S to mean signal slope: $\sigma_T^2(T_0) = \sigma_S^2(T_0)/\langle |dS/dt|_{t=T_0} \rangle^2$.

Using the results of the previous section, one can write the accuracy of a time measurement on an SER pulse (dynode chain response) following single-photon detection as:
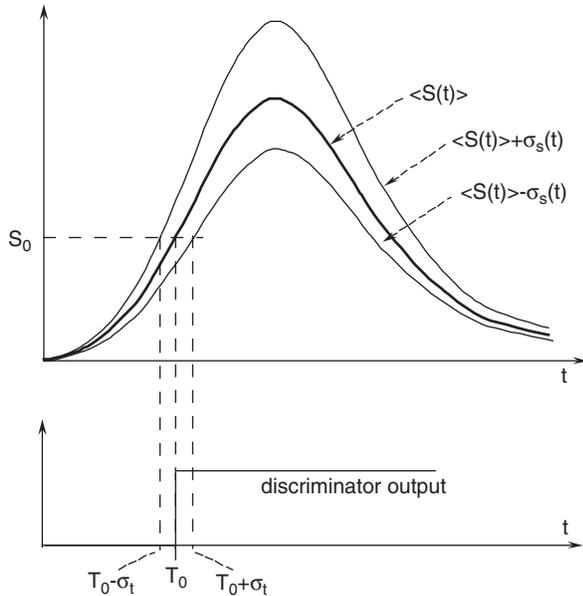
$$\sigma_{T1}^2 = \varepsilon_A^2 \langle SER(t) \rangle^2 / [d/dt \langle SER(t) \rangle],$$

and for light pulses much faster than the SER, using $\Phi(t) = \delta(t)$, we get:

$$\sigma_T^2 = [(1 + \varepsilon_A^2)/R]\{f_0(t) * \langle SER(t) \rangle\}^2 / \{d/dt [f_0(t) * \langle SER(t) \rangle]\}^2$$

As an example, for Gaussian-distributed time-of-flights $f_o...,f_n$, each with a variance σ_t, using a threshold placed at the maximum SER slope, $\sigma_{T\delta}^2 = [(1 + \varepsilon_A^2)/R](n + 1)\sigma_t^2$ is obtained. Thus, time resolution $\sigma_{T\delta}$ obtained is given by the SER time-width $\sqrt{(n+1)}\sigma_t$ and decreases as the square root $\sqrt{R}$ of the mean number of detected photons. Typical values of the accuracy $\sigma_{T\delta}$ are 0.3–1 ns for $R = 1$, and 3–10 ps for $R = 10^4$.

– *Correlation response*. In correlation measurements, are described the statistical properties of light fields by the correlation function $\rho(\tau)$ defined as: $\rho_L(\tau) = \int_{t=0-\infty} L(t)L(t+\tau)dt$, where $L(t)$ is the input light-signal power, assumed random and stationary.

Two functions are needed to compute the output correlation from the input light correlation, the autocorrelation function associated with the mean single-photon response, $\langle SPR(t) \rangle = f_0(t)*\langle SER(t) \rangle$:

$$\rho_{\langle SPR \rangle}(\tau) = \int_{t=0-\infty} \langle SPR(t+\tau) \rangle dt,$$

and the autocorrelation function associated with the fluctuations:

$$\rho_{\Delta SPR}(\tau) = \rho_{\Delta SPR}(\tau) + \varepsilon_A^2 \int_{t=0-\infty} f_0(t) * \langle SER(t) \rangle \langle SER(t') \rangle.dt.$$

With these functions, one finds the compounding of correlations and the PMT output as:

$$\rho_1(\tau) = (\eta e)^2 F \rho_{\Delta SPR}(\tau) + (\eta e)^2 F^2 [1 + \gamma(\tau)] * \rho_{\langle SPR \rangle}(\tau).$$

Also in this case, by a simple deconvolution, one can go down to resolve field correlation on a time scale comparable with the SER duration $\Delta\tau(\approx 2 - 5ns)$, and, using the last expression, to a small fraction of $\Delta\tau$.

4. PMT PERFORMANCES

Now review the main parameters affecting PMT performances.

About *photocathodes*, a wide choice of formats is available (e.g., tubes with 12, 18, 25, 38, or 50 mm diameter are standard). Special devices can reach 200 and 500 mm diameter with hemispherical surfaces or a hexagonal shape suitable for side-by-side packing with little wasted area. Special units with unusually large sizes have been fabricated for nuclear physics and biology applications. Ex-

Figure 10. Timing of a PMT signal by a threshold-crossing circuit with trigger at a level S_0.

cluding specific considerations, it is always better to select the smallest photocathode area as permitted by the application at hand, because not only size and cost are minimized, but also the dark current and perhaps the time response (2).

Number of dynodes and gain. Usually, the number of dynodes of the multiplying chain is between 8 and 12. Special units can have n = 6 (intense pulse detection) or n = 14 (for VUV and in some obsolete specimens). No advantage exists in using a number of dynodes exceeding n = 12, which already gives adequate gain for single-photon detection.

Nor is it customary to go below n = 6, the minimum to obtain a gain ($g^n > 10^4$) large enough to render further amplification unnecessary and making the load-resistance noise negligible. The typical dependence of gain $G = g^n$ from supply voltage V_{al} and number of stages n is shown in Fig. 11 for common dynodes. The gain dependence on voltage is taken to advantage for an easy trimming of the PMT gain, even on a range of a decade or more.

Best performance is obtained by the least possible number of dynodes, so that the PMT works at the highest voltage V_{al}, which makes the dynode gain g_i be high and the time dispersion σ_{ti} low, improving the charge and current responses (the factor ε_A^2) as well as the time response (variance σ_T^2). A few PMTs use a GaP:Cs first dynode (Fig. 7) to get high g1 gain and thus smaller multiplier variance $\varepsilon_A^2 = g/g_1(g - 1)$.

In addition, the SER duration and delay decrease and benefit the speed of response; the slope dG/dV_{al} is less, and the need for a stabilized supply V_{al} to obtain a stable G is relaxed (6,7).

PMTs with microchannel plates have gains up to 10^7 (3-stage) but with a large variance ε_A^2; they are useful in se-

lected applications where the limited dynamic range of output current (100 nA–1 μA) is acceptable.

SER waveform and related parameters. Parameters used to describe the SER waveform (Fig. 2) of a fast PMT (Fig. 12): the delay time τ_d between excitation and the response peak; the time duration Δ_τ or pulse full-width half maximum (or FWHM); the rise time τ_r, or the time interval during which the pulse rises from 10% to 90% of its peak value.
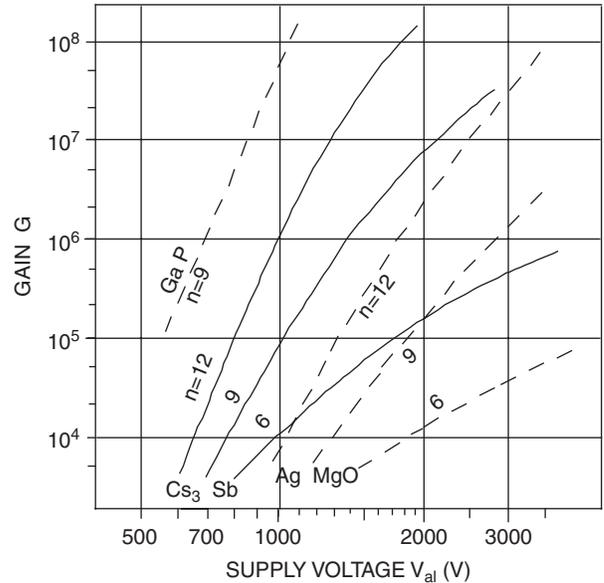
Figure 11. The gain of the dynode chain as a function of supply voltage for some commonly used dynode materials and number of stages (from Ref. 1, reproduced by courtesy of Prentice-Hall).
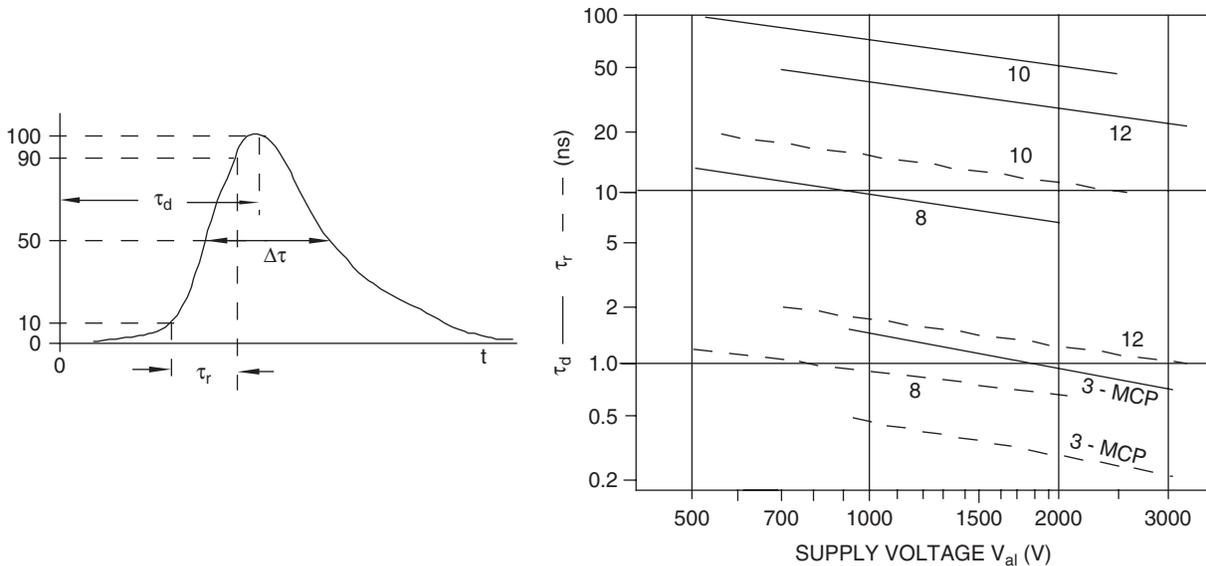
Figure 12. (left) Definition of the SER characteristic time parameters and, (right) the SER rise time τ_r (dashed lines) and delay time τ_d (solid lines) of typical PMTs, as a function of the total supply voltage, and with the number of dynodes as a parameter. (From Ref. 1, reproduced with permission by Prentice-Hall).

All these quantities have a dependence of the type $\tau = k(V_{al}/n)^{-1/2}$ indicating the value of choosing a low n and a high V_{al}. In a photon-counting application (6), the SER peak current i_p is kept in the mA range, so that the voltage across the 50 Ω load is hundreds of mV and the signal can be treated directly by subsequent circuits.

Figure 12 (right) illustrates the rise and delay times of some common PMTs with different numbers of stages, as a function of the supply voltage. The trend $V_{al}^{-1/2}$ is generally well matched; the range of V_{al} spans about a decade, with slow specimens (those with n = 10) that may have τ_r = 10 ns and τ_d = 50 ns, and the fast ones (n = 8 and 12) with rise time τ_r down to 0.6–1 ns (better at small n). Units with MCPs can reach τ_r = 150–300 ps.

Linearity, Dynamic Range and Saturation. Ideally, the output current I_u is a linear function of the radiant detected power P, that is, $I_u = G\sigma P$, where σ is the spectral sensitivity and G is the gain. The integral linearity error ε_l is defined as the relative deviation ΔI_u of the current from the nominal expected value $I_u = G\sigma P$, or $\varepsilon_l = \Delta I_u/I_u$.

In the PMT, ε_l is usually very small (i.e., < 0.1%) over a wide dynamic range (several decades), which is the error found in the normal region of use of the device, from the minimum level comparable with dark current I_b to the large-signal regime.

For large signals, first an increase of ε_l is found, then the output signal I_u levels off, with a different behavior in a direct current (dc) and pulsed-signal regime.

In dc, the maximum current density of dynodes is never attained, because it leads to excessive dissipation. If 100 mW/cm² is the dissipation tolerated by a dynode before damage, for typical values of 200 V and 0.5 cm² , one obtain a current limit of 100 μA. It is the last dynode to sustain the largest dissipation and would be destroyed first, because it has the largest current in the chain.

The dynode bias voltages are commonly obtained with a resistive divider, as in Fig. 13. Letting all the resistances equal, the dynode current is limited to the current flowing in the divider, $I_p = V_{al}/(n+1)R$. When I_n approaches I_p, a linearity error ε_l is generated. The error is positive for small I_u because I_n is outgoing from the dynode, and the

voltage drop across the resistances of preceding stages *increases* the voltage and, hence, gain. For larger signals, the anode-to-last dynode voltage is appreciably reduced, and the anode collection efficiency drops because of space charge effects. Thus, ε_l becomes negative and a saturation of the output current is finally reached.

Typical maximum dc anode current in a PMT is in the range of 100 to 500 μA for dynode structures. When compared with the dark and dispersion anode currents of 0.1–10 nA (typically), the above values indicate a dynamic range of linearity (in direct current) of 4–5 decades.

In the pulsed regime, one can allow for much larger peak currents without incurring excessive dissipation.

Of course, the dynode voltages cannot vary during the pulse, which is ensured with a capacitance bypass of the last (2–3) dynodes, those supplying the largest currents. Small capacitors (C $\approx$ 1000 pF typically) are placed in parallel to R to ensure an effective bypass up to tens of μs.

In MCP PMTs, the allowable specific dissipation in the microchannel is smaller ($\approx$ 1 mW/cm²) than for discrete dynodes and the maximum dc currents are but $\approx$ 100 nA/cm². As bypass capacitors cannot be used, peak current is limited by the charge stored in the stray capacitance (typically $\approx$ 1 nC/cm2), with a slow recovery time in the range of ms.

Resolution in Charge and Time Measurements. This accuracy is described by $\varepsilon_A^2 = g/[g_1(g-1)]$, the relative variance of the charge distribution at the anode for R = 1 photoelectron at the input. One can also refer to resolution ρ_A of amplitude measurement, defined as the full-width half maximum (FWHM) of the charge distribution. Then, it is $\rho_A = 2.36\,\varepsilon_A/\sqrt{R}$ for a Gaussian statistics.

Dark current in a PMT, at the anode $I_{d/a}$, is the sum of three terms: $I_{d/a} = GI_{d/ph} + G'I_{d/d} + V_{al}/R_{is}$, where $GI_{d/ph}$ is the dark current of the photocathode amplified by the gain G, the dark currents of the dynodes is summarizes by $G'I_{d/d}$, and the last term is the leak through the insulation resistance R_{is}.

In a well-designed structure, R_{is} is high enough to make the last term negligible. Contribution V_{al}/R_{is} is a dc component and can be filtered out when one takes only
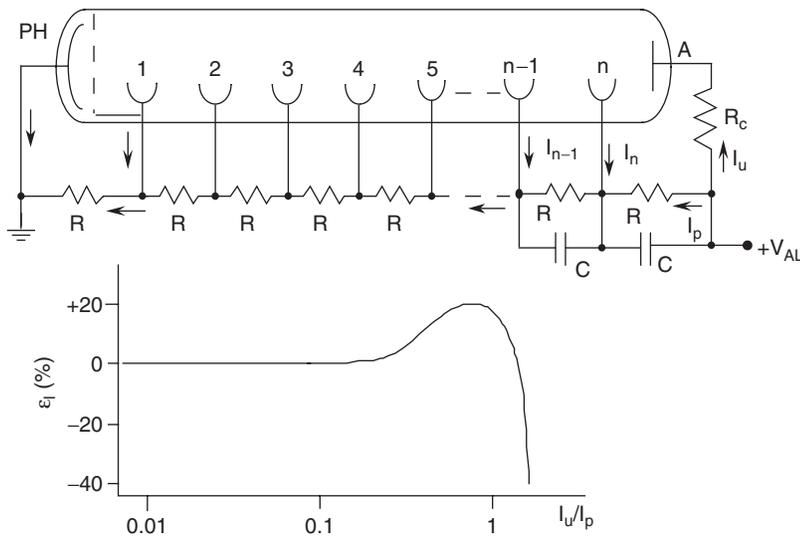
Figure 13. Schematic of the dynode voltage divider (top) and linearity error in dc current near saturation (bottom) (from Ref. 1, reproduced with permission of Prentice-Hall).

single SER pulses. As a result of the different charge distribution compared with the SER, the dynode contribution is significantly reduced by validating the output pulses through a proper window of discrimination. This handling, used in single-photon counting, attains the intrinsic limits of photocathode dark current.

Bias circuits are for general-purpose applications, where equal resistances are used in the dynode divider (typically $R = 0.1–1 M\Omega$), with bypass capacitors ($C \approx 1000 pF$) across the last two or three stages. This way, the maximum gain is achieved at a given supply voltage V_{al}. However, if a specific performance is to be optimized, voltages shall be distributed differently.

To get the best amplitude and time resolution, R_1 is set 2 to 5 times R, so that both ε_A^2 and σ_{t0}^2 are minimized. Dynamic range, linearity, and saturation are improved, in both dc and pulsed regimes, increasing the final two resistors to 3–6 times R.

A negative voltage applied to the photocathode is preferred to a positive applied to the anode. In this way, the output signal is referred to ground (and not to the high-voltage V_{al}), and dc decoupling is no longer needed.

In all PMTs, a magnetic screen is commonly used, made of soft ferromagnetic material with high permeability μ_r (e.g., mumetal, permalloy) in the form of a tube to be slipped on the envelope. This item is usually supplied by the manufacturer and has the exact size for each device. For high static (dc) magnetic fields, a loss of gain is incurred because electron paths are deflected and the next electrode experienced a collection loss. In ac, the same effect shows up as a modulation of the output current. The typical susceptivity for a unshielded PMT is 2–10%/gauss for axial fields and, 10–50%/gauss for transversal fields. Special units designed for immunity to high fields are available with a 10^2 reduction of the above figures.

5. APPLICATIONS OF PMTS

In this section, the common applications of PMTs to instrumentation and related measurement areas, are reviewed and some examples of design and the performance obtained are illustrated.

5.1. Detection of Weak Signals with Moderate Bandwidth

In the detection of weak signals in dc or with low-frequency content, the PMT offers good performance of sensitivity and wide spectral range (extended to the extreme UV) together with a small dark current and large sensitive area.

Tubes with n = 6–9 stages are commonly used, and the load resistance R_c is kept as high as possible compatible with the stray capacitance C_a of anode and output wiring to ground, and with the desired bandwidth B: $R_c = 1/[2\pi C_a B]$. For example, with $C_a = 10 pF$ and $B = 100 kHz$, $R_c = 160 k\Omega$. In this way, the delta-impulse response is a negative exponential with time constant $\tau = R_c C_a$ ($= 1.6 \mu s$ in our case). The output signal is given by $V_a = R_c I_a = R_c \sigma G P$, where P is the detected radiant power and σ is the spectral sensitivity. Establishing a limit of anode dc current at $I_{amax} = 1 \mu A$, the maximum output signal is $V_a = R_c I_{amax} = 160 mV$.

5.2. Measurement of Fast Waveforms

The PMT is the best-suited device for the detection of fast (ns range) optical signals, as it offers both a high gain G and a large bandwidth B, and a quantum-limited performance at weak signals. The gain-bandwidth product GB of the PMT surpasses by several decades that of other types of photodetectors, and offers a low-noise figure. These features make the PMT ideal in the spectral range of response even with small quantum efficiencies, in several applications of physics, biology, and engineering (7).

To ensure an adequate output signal amplitude $V_a = R_c \sigma G P$ in subsequent circuit handling (for example, be at least $\approx 100 mV$), one needs a larger gain $G = g^n$, obtained either with more dynodes or with larger inter dynode voltages.

The best speed of response is obtained with the gain G given by the smallest number n of dynodes, fed at the largest allowed voltage. With a $G = 10^8$ gain, an optical mean power $P = 1 pW$ provides a photoelectron rate $\sigma P/e = 3.10^5 s^{-1}$ (at $\lambda = 550 nm$), and each photoelectron is detected as a pulse of peak amplitude $Ge/t_d = 5 mA$ (having taken $\Delta\tau = 3 ns$).

If the waveform is pulsed and contains R electrons in the SER duration, with a saturation level at a peak of $I_s = 100 mA$, a dynamic range $R = I_s \Delta\tau/Ge = 20$ is obtained. For even stronger pulses, one may reduce G accordingly (at fixed I_s), for example, with R = 2000 photoelectrons, a gain $G = 10^6$ will suffice (which requires n = 8 dynodes), etc.

5.3. Time Measurements

The time accuracy σ_T^2 primarily depends on the response times associated with SER and f_0 and scales as the square-root number of detected photoelectrons. Requirements for output structure, gain, and number of dynodes are those for fast waveforms. In addition, it is useful to reduce the dispersion σ_{t0} and keep a high-gain g_1, with a higher voltage at the first dynode.

In a typical fast PMT with $\Delta_\tau = 3 ns$ and n = 12, one can have $\sigma_{t0} = 0.4 ns$ for the dispersion of the first dynode and $\sigma_{ti} = 0.9 ns$ for those from other dynodes. Then, assuming $g_1 = 6$ and $g = 4$, $\sigma_T = \{\sigma_{t0}^2 + [g/(g-1)g_1]\sigma_{ti}^2\}^{1/2} = \{0.16 + 0.18\}^{1/2} = 0.58 ns$ is the intrinsic limit in a time measurement with a single photoelectron (1,5–7). This figure is typical of what can be obtained with a CFT (constant fraction timing) commercial circuit, while for a fixed-threshold timing accuracy it would be $\sigma_T = \{\sigma_{t0}^2 + n[g/(g-1) g_1]\sigma_{ti}^2\}^{1/2} = 1.52 ns$.

For pulses $\Phi(t)$ faster than the SER, this figure scales down as $\sqrt{R}$ with an increase in the number of photoelectrons per pulse R, going down easily to the limits of resolution allowed by fast electronic circuits ($\approx 10 ps$), even for relatively weak optical pulses with, for example $R = 10^3$. For slow pulses, the intrinsic limit to accuracy σ_T is determined by the characteristic time τ_Φ associated with the waveform $\Phi(t)$, still with the dependence $\tau_\Phi/\sqrt{R}$ for the number of photoelectrons per pulse R.

5.4. Photocounting Techniques

In photocounting (3,6), SER pulses coming from the photocathode are descriminated from all other pulses (from dynodes) and because of spurious actions (i.e., ion bombardment and radiation contaminants in the envelope-like ^{40}K of glass).

By eliminating all these pulses, the residual counting rate achieves the very low level inherent in the dark emission. A simple method, currently used for discrimination, is that of measuring the total charge Q at the anode, and accept only those events that have the charge Q falling between two thresholds Q_1 and Q_2 properly selected.

A circuit to implement the photocounting technique is shown in Fig. 14. The anode output is passed on to the counter through a linear gate, enabled by the result of charge amplitude discrimination. Charge discrimination is best performed on the last dynode current pulse, made available by deriving a small resistance in series to the dynode. The dynode pulse is well correlated to the anode-pulse charge (by a factor $-1 + 1/g_n$) and, additionally, it leads the anode pulse by the time-of-flight t_n (a few ns). Thus, the integrate-and-dump amplifier has a time t_n available to supply the result of discrimination at the enable input E before the anode pulse arrives at the linear gate input. If t_n does not suffice, an additional delay t_r will be added by means of a delay line.

After the linear gate, a shaper is used to shorten the pulse as much as possible to avoid pile-up of successive pulses, before going to the counter. A last function performed (not shown in Fig. 14) is that of dumping the integrator after a time T_{rec} from each pulse, to recover the zero level for the next pulse to be discriminated.

Requirements for PMTs in a photocounting regime are: a SER amplitude in the mA range (to get ≈ 100 mV in circuits), requiring $G \approx 10^7$–10^8 and therefore $n \approx 12$ dynodes; a high first dynode gain to get a good discrimination efficiency; and a voltage divider adequate to have a short Δ_τ of SER.

The recovery time T_{rec} of the integrator determines the maximum photon rate acquired for the photocounting as $F = 1/T_{rec}$, hence a dynamic range (in power) $P = e/T_{rec}\sigma$ (with $\sigma = 20$ mA/W and $Trec = 10$ ns, it is $P = 0.8$ nW). The sensitivity performance for weak signals is then determined by the dark current I_d of the photocathode, which can go down to fA or aA at ambient temperature depending on the threshold of the material.

The sensitivity of the photocounting technique can be now evaluated (1). The total counting $N = N_s + N_d$, given by the sum of signal and dark currents, has a mean value $\langle N \rangle = \langle N_s \rangle + \langle N_d \rangle = \eta . \eta_p$ FT $+ (\eta_d I_d$ /e)T, where η is photocathode quantum efficiency, and η_p, η_d are the counting discrimination efficiencies of photon and dark pulses. Both contributions are distributed according to Poisson statistics and, therefore, the variance is $\sigma_N^2 = \langle N_s \rangle + \langle N_d \rangle$. For $I_d = 0$ and $\eta\eta_p = 1$, the signal-to-noise ratio intrinsic to input signal electrons is $(S/N)_i^2 = \langle N \rangle^2/\sigma_N^2 = \langle N_s \rangle$, whereas with a finite dark current, $(S/N)^2 = \langle N_s \rangle^2 / [\langle N_s \rangle + \langle N_d \rangle]$ is obtained.
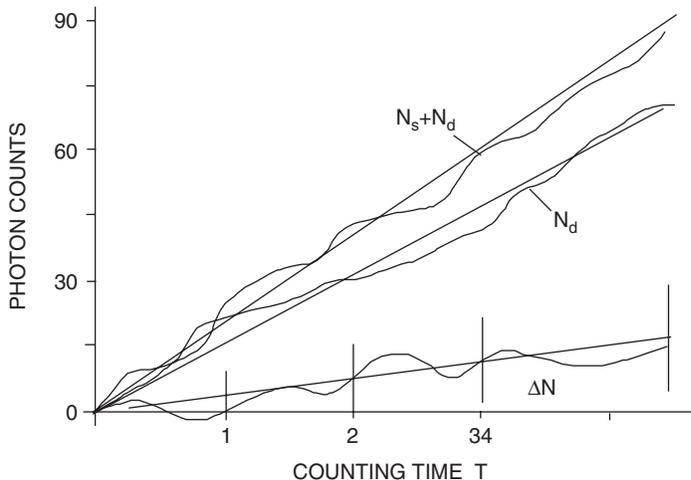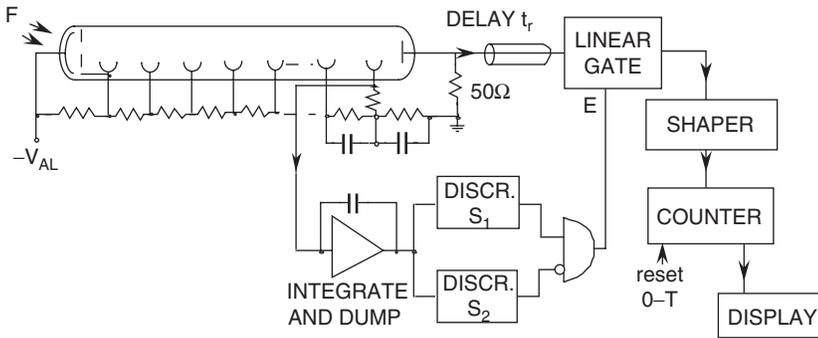
Figure 14. Basic functional scheme of photocounting (top), and an example of measurement, showing evolution of signal plus dark, dark only, and result after dark subtraction (bottom). Vertical bars on ΔN indicate the $\pm 0.5\sigma_{\Delta n}$ confidence level. (from Ref. 1, reproduced with permission of Prentice-Hall).

With dark counting rates I_d/e of a few electrons/cm^{2}s, the minimum measurable radiant power is $P = Fh\nu/\eta.\eta_p \approx 10^{-18}$ W (at $\eta.\eta_p = 0.1$, near the peak of photocathode response). This performance is unsurpassed by other photodetectors.

A refinement is to subtract dark counting N_d with a second measurement performed with $N_s = 0$ (no light in) and lasting as above, a time T. Subtracting the dark gives as a result $\Delta N = \langle N_1 \rangle - \langle N_2 \rangle = \langle N_s \rangle + \langle N_d \rangle - \langle N_d \rangle = \langle N_s \rangle$. For the variance of ΔN, recall that for a difference it is the sum of variances, so that $\sigma_{\Delta N}^2 = \langle N_s \rangle + 2\langle N_d \rangle$ and $(S/N)^2 = \langle N_s \rangle^2 / [\langle N_s \rangle + 2\langle N_d \rangle]$.

Figure 14 (bottom) illustrates the typical outcome of a photocounting measurement, plotting the signal plus dark and the dark alone. Both graphs increase with time T, but the randomness makes their difference surely positive only after a certain time; in the figure, condition $\langle N_s \rangle > \sigma_{\Delta N}$ is reached at the right end of the diagram.

The minimum signal detectable in photocounting with dark subtraction is obtained by setting the condition S/N $= 1$, thus $\langle N_s \rangle^2 = \langle N_s \rangle + 2\langle N_d \rangle$. When the signal trying to be recovered is weaker than the dark, that is $\langle N_s \rangle \ll 2\langle N_d \rangle$, $\langle N_s \rangle_{min} = \sqrt{[2\langle N_d \rangle]} = \sqrt{(2\eta_d I_d \ T/e)}$ is obtained.

As $\langle N_s \rangle = \eta\eta_p$ FT, it corresponds to a minimum detectable photon rate F_{min} given by:

$$F_{min} = \sqrt{(2\eta_d I/eT)}/\eta\eta_p,$$

from which one can see that F_{min} improves as the inverse square root of integration (or counting) time T. Letting all ηs $= 1$, F_{min} becomes the geometrical mean $\sqrt{(I_d/e).(2/T)}$ of the dark pulse rate I_d /e and of 2/T (i.e., two counts in the integration time T).

Typically, the minimum detectable rate is evaluated as $F_{min} = \sqrt{(2 \cdot 0.7 \cdot 1/T)}/0.35 \cdot 0.7 = 4.83/\sqrt{T}$ photon/s. For an 8 h integration period, it would yield $4.83/\sqrt{28800} = 0.028$ photon/s or also $P_{min} = h\nu F_{min} = 1.3 \cdot 10^{-20}$ W, which is the power collected from a m $= 28$th magnitude star with a 1m^2 telescope aperture.

About photocounting, it is sometimes claimed that the sPAD (single-photon avalanche photodiode) has the same capability of PMTs in detecting and counting single photons. The SPAD is an avalanche photodiode (1) based on a pin structure fabricated in a suitable ($E_g \approx h\nu$) semiconductor. It is biased a little bit beyond the infinite-gain voltage of the linear-regime avalanche multiplication (1). The SPAD can indeed detect single photons, but only in a Geiger-counter mode, that is, with a long dead-time following each detected event and no capability of linear response. In fact, the response to, say, R $= 10$ simultaneous photons is the same as for R $= 1$ photon, whereas in PMTs, it is ten times as large. The SPAD also has a limited spectral range, difficult to be extended into the UV, and a very minute sensitive area, a tiny fraction of mm^2 compared with cm^2 of the smallest PMT.

5.5. Nuclear Radiation Spectrometry

In the context of nuclear physics, PMTs are widely used in a number of measurements, such as energy spectrum of radiation, particle identification, time of flight, and coincidence detectors. When the PMT is optically coupled to a scintillator, the combination is called a *scintillation detector* (5).

The scintillator is a sort of converter that, when excited by energetic radiation (γ-rays, fast electrons, neutrons, etc.), generates a fast light pulse containing a number of photons R proportional to the energy E_γ of the absorbed quantum, i.e., $R = \kappa E_\gamma$, with κ being a conversion factor of the material (typically $\kappa \approx 1$ photoelectron/keV).

Common scintillators are crystals like NaI:Tl (thallium-activated sodium iodide), CsI:Tl, BiGeO (bismuth germanate), or organic and polymeric compounds known as anthracene, stilbene, etc. Scintillators' emission spectrum is centered in the interval of 400–550 nm to match the common photocathode response (particularly, S-11 and S-20), and their impulse response waveform has a nearly exponential decay, with time constants ranging from a few ns (organic) to 100 ns (crystals) (1,5).

The scintillator base is placed in optical contact on the input window of the PMT, while its walls are treated with white paint (TiO_2), to help rediffuse photons toward the PMT. From the measurement of the total charge $Q = \eta eGR$ of the anode pulse, one goes back to the energy of the detected quantum as $E_\gamma = Q/\kappa\eta eG$. Using the integral response accuracy, the charge S/N ratio $\sigma_Q/Q = \sqrt{[(1 + \varepsilon_A^2)/R]}$ coincides with the S/N ratio of the energy σ_E/E_γ, hence an energy resolution $\sigma_E = E_\gamma\sqrt{[(1 + \varepsilon_A^2)/R]} = \sqrt{[(1 + \varepsilon_A^2)/\kappa]}$. For example, the 1.46 MeV peak of ^{40}K is resolved by NaI:Tl with $\sigma E \approx 40$ keV. In this measurement, it is the low value of κ that sets the resolution, whereas the factor ε_A^2, the noise added by the PMT, is small.

Energy spectrometry measurements (7) are carried out with the setup reported in Fig. 15. By integration of the anode current in a charge amplifier, one first gets a quasistep waveform whose amplitude is proportional to the charge. The integrated pulse is passed to a pulse-shaping circuit, to limit pulse width and maximize the pulse acceptance rate without superposition of waveforms.

Finally, the signal is sent to a multichannel analyzer (MCA), an instrument that classifies the pulses according to their peak amplitude, stores the event in the memory, and displays the resulting distribution.

5.6. Dating with Radionuclides

Energy spectrometry has an application in dating with radionuclides, especially with carbon, which is based on the decay of ^{14}C, a radioisotope that is produced in the atmosphere by reaction of nitrogen with slow neutrons and secondary components of cosmic rays and is present in atmospheric carbon dioxide with a concentration confidently assumed constant through the last million years.

Dating consists of measuring the content of the isotope ^{14}C in organic substances, where it has been initially absorbed from atmosphere during their lifetime, and since then has decayed with a half-life of $T_{1/2} = 5700$ years with the emission of 160 keV electrons.

The measurement is performed by looking at the rate of detection of 160 keV electrons from the specimen, which is proportional to the residual content of ^{14}C in it.
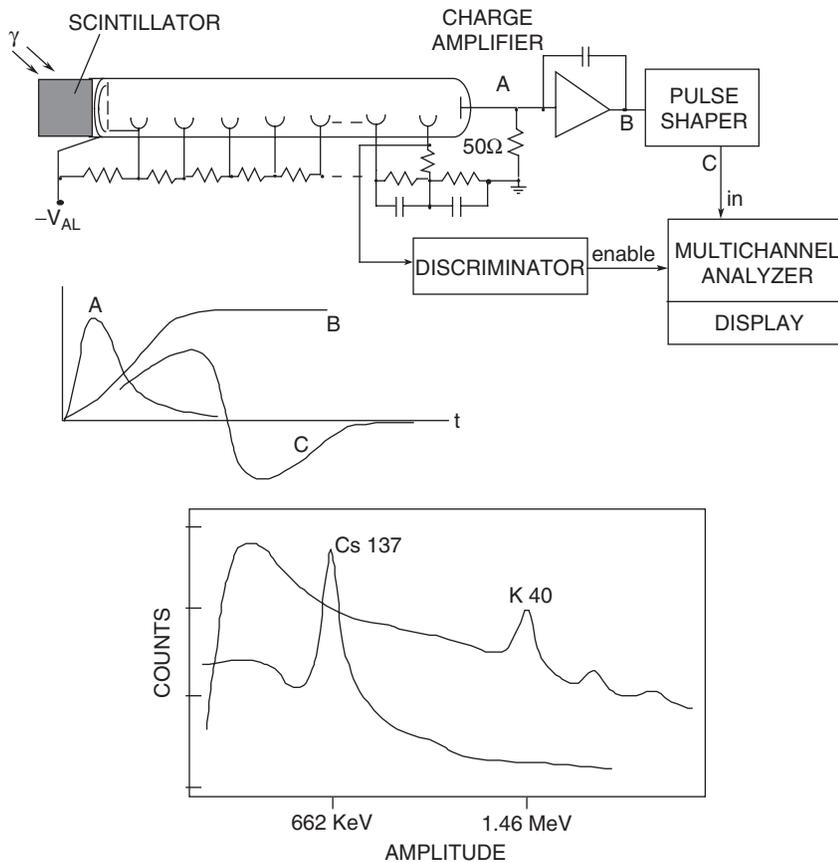
Figure 15. Scheme of energy spectrometry measurements (top) and signal waveforms (middle). The energy spectrum measured with the scintillation detector (bottom) reveals radionuclides species (energy signature) and their concentrations (counts intensity) (from Ref. 1, reproduced with permission of Prentice-Hall).

With an energy spectrometry as in Fig. 1–15, one can distinguish such electrons from the background of radiation (which has a different energy distribution), and resolve concentrations down to $c \approx 10^{-3}$ of the initial value c_0. As concentration decays according to the simple expression $c/c_0 = 2^{-T/T_0}$, for the age T: $T = T_{1/2} \log_2 (c_0/c)$ is obtained, with a practical limit in dating to about 40,000 years.

Also, other radionuclides, such as ^{87}Rb, ^{232}Th, ^{235}U, and ^{40}K, are currently used to cover much larger time spans (up to 10^6–10^9 years) for use with inorganic samples where ^{14}C is absent, again with the same method and measurement setup.

BIBLIOGRAPHY

1. S. Donati, *Photodetectors*, Upper Saddle River, NJ: Prentice-Hall, 2000.

2. Burle Industries, *Photomultiplier Handbook*, vol. PMT-62. Lancaster, PA: Burle, 1987.

3. R. L. Bell and W. E. Spicer, Compound photocathodes: a new family of photoemitters with greatly improved performance. *Proc. IEEE* 1970; 58:1788–1801.

4. L. M. Biberman, and S. Nudelman, *Photoelectronic Imaging Devices*, vols. 1 and 2. New York: Plenum Press, 1971.

5. S. Donati, E. Gatti, and V. Svelto, The statistical behaviour of the scintillation detector: theory and experiments. In: L. Marton, ed., *Advances in Electronics and Electron Physics*, vol. 26. New York: Academic Press, 1969, pp.251–304.

6. S. K. Poultney, Single photon detection and timing: experiments and techniques. In: L. Marton, ed., *Advances in Electronics and Electron Physics*, vol. 31. New York: Academic Press, 1972, pp. 39–116.

7. J. P. Boutot, J. Nussli, and D. Vallant, Recent trends in photomultipliers for nuclear physics, In: L. Marton, ed., *Advances in Electronics and Electron Physics*, vol. 31. New York: Academic Press, 1983, pp. 223–305.