Single-photon detectors: from traditional PMT to solid-state SPAD-based technology

Silvano Donati, *Life Fellow, IEEE,* and Tiziana Tambosso, *Senior Member, IEEE*

***Abstract*— Solid state photomultipliers (SSPM), based on the multi-element SPAD (single-photon avalanche detector), are nowadays jousted to replace the traditional, vacuum-tube technology photomultipliers (PMT), based on the photocathode and dynode-chain concept. We revisit the milestones and the underlying conceptual issues marking the developments of single-photon-detectors, from the PMT to the SSPM. Then we compare state-of-the-art performances of the two and find that SSPMs are equalling or even surpassing PMTs for response time and sensitivity, while are still lagging for acceptance area, linearity and dark current. We finally draw an overview of applications to: pulse spectrometry, fast waveform analysis, and photon counting.**

***Index Terms* — Photodetection, Photomultipliers, Single photon detectors, Avalanche photodiodes, Semiconductors**

I. Introduction

Single-photon detectors are particularly attractive because they attain or closely approach the quantum limit of sensitivity in the detection of very weak optical signals [1]. Since its invention almost hundred years ago by Slepian [2,3], the photomultiplier tube (PMT) has been the workhorse of single-photon detection in the spectral range extending from deep-UV to near-IR, and has enabled key measurement techniques to be developed, like scintillation counting for nuclear pulse spectrometry [4], fast waveform and lifetime measurements [1], and single-photon counting for measuring extremely weak flux of photons [1,5], down to levels of 0.01 photon/s or 2×10^{-21} W [1], a unequalled sensitivity crucial to a number of applications, ranging from biology [6], astronomy [7], and dating with radionuclides [1].

The PMT commonly uses a thin layer of alkali-metals antimonide [1] as the photosensitive surface, the photocathode, and a chain of multiplying electrodes, the dynodes, to amplify the detected charge up to point it is readily measured at the output anode as a short current pulse following the detection of an individual photon at the photocathode.

Of course, the vacuum-tube nature of the PMT brings about some obvious disadvantages in the practical use, like fragility and limited ruggedness of the multi-electrode dynode chain, the vacuum-tight requirement of enclosure, and need for an in-situ processing of the photocathode, calling for fabrication limited at one at the time.

Yet the PMT has been instrumental to open the way to high sensitivity detection techniques in science and engineering. The readily observed single photon capability, with the CW output current breaking down into a series of equal-waveform short pulses (the single-photon response or SER) as the input intensity is gradually lowered, provided a beautiful experimental evidence of the just introduced Planck's quanta hypotesis.

And, on another side, it is today surprising that the hearth of the device, the photocathode, was readied so easily, on a miracle of cut-and-try fortunate blind finding of the good materials, the alkali antimonides like Cs_3Sb and NaCsKSb (the S-11 and S-20 responses), remarkably, about forty years before the theory of conduction in semiconductor was developed, what is recognized today as the necessary basis of understanding. Equally surprising, in another material, the AgO:Cs (cesiated silver oxide, the early S-1 response), we find an *ante-litteram* technology dicovered by chance of the smart experimentalists of hundred years ago: a truly nanoparticle material.

From the application point of view, PMTs soon unveiled to the first reserachers in photonics that not just *responsivity* - or quantity of response - is important, but even more critical is the input-equivalent-noise per unit bandwidth and surface, the real figure of merit of the detector, as pointed out by R.C.Jones [8] in a classical paper of the 1950, which introduced concepts like *NEP* and *detectivity* that are since then fundamentals in photodetection.

Due to these progress in the system side of application of PMTs, concepts like *thermal*-limited and *quantum*-limited detections were finally understood [1] leading to clarify the focus point: the PMT surpasses other existing detectors beacuse it has *internal gain*, and so makes the thermal or Johnson noise fluctiation introduced by the termination load negligible respect to the quantum or granular noise that the detected current carries in consequence of the quantum nature of detected photons.

In the years between '30 and '70, PMTs were not standing still: from the initial low-gain dynodes - requiring 14-16 stages to attain the $\approx10^8$ gain necessary to see single-photon pulses, new high-gain dynode surfaces with Cs_3Sb and later GaP:Cs allowed to make chains with just 8-12 dynodes [1]. Photocathodes were improved as well, initially with deep-UV CsI, Cs_2Te and metal surfaces (response down to λ= 130-nm) and later on the other side of the spectrum, in the 1970, with the cesiated surface exploiting the NEA (negative affinity emission) concept, and with the modern InGaAs:Cs surfaces, that in reflection [1,9] were able to

Manuscript received...........revised.............accepted...........
Silvano Donati is with the Department of Industrial and Information Engineering, University of Pavia, v. Ferrata 1, 27100 Pavia, Italy.
T.Tambosso is with the Department of Electronics, Da-Yeh University, Dacun 51591, Taiwan. S. Donati is also with the Graduate Institute of Precision Engineering, National Chung Hsing University, Taichung 402, Taiwan. Corresponding author: silvano.donati@ieee.org

extend the response in the near IR up to a photoelectric threshold of about $\lambda \approx 1700$ nm.

Despite the vacuum-tube limitations, up to a few years ago, PMTs were the only photodetector capable of single photon performance and, additionally, they offered remarkable features, like the very large (many cm^2) sensitive area and record low dark current (down to 10^{-18} A/cm^2) at no extra cost.

The only drawbacks of PMT, making a solid-state replacement of it desirable, were the bulky size of the dynode chain - length (for head-on types) or diameter (for circular dynode chain) was typically ranging from 3 to 8 times the size of the photocathode sensitive area - and the relatively high dc voltage required to bias the dynode- chain voltage divider (typ. 0.8 to 3 kV). And, despite some technological improvements appeared in the '90s, with compact MCP (multi channel plate) replacing the dicrete dynode chain, squeezing the length to diameter ratio to about unity, and the integral dc/dc converter requiring only a low voltage to operate the component, the PMT remained a relatively fragile component, difficult to be qualified for use in demanding environment (e.g., space, biomedical) [9].

II. KEY ISSUE: INTERNAL VS EXTERNAL GAIN

Thus, when in the '60 semiconductor photodiodes started to appear, they were hailed to soon become the solid-state replacement of the PMT in single photon detection.

But, there was a subtle impediment to prevent the combination of a photodiode and an (external) high-gain amplifier be equivalent to the PMT, the combination of a photocathode and an (internal) gain.

Indeed, let the photodetected current be I_{ph} ($=\sigma P$, with σ the spectral sensitivity and P the input radiant power) and suppose it is internally amplified by G_{int} before arriving at a load resistance R_L. Then, the signal proceeds in the circuit and get a final amplification by A_{ext}. In this way, the signal voltage is:

$$V_{out} = I_{ph}\ G_{int}\ R_L\ A_{ext} \qquad (1)$$

so, for the signal internal and external gain G_{int} and A_{ext} seem to have the *same* effect. But, V_{out} is actually the *mean* signal, and superposed to it we find a noise v_n^2, due to quantum fluctuations associated to detected photons, and to Johnson noise generated by the load resistance. At the output, the quadratic mean value of these contributions is found to be [1]:

$$v_n^2 = (2eI_{ph}B\ G_{int}^2\ R_L^2 + 4kTB\ R_L)\ A_{ext}^2 \qquad (2)$$

Now G_{int} and A_{ext} have no more the same effect. By taking the ratio of quadratic signal V_{out}^2 and noise v_n^2 we obtain, the SNR^2 (signal-to-noise ratio), or quality factor of our signal, as:

$$SNR^2 = I_{ph} / (2eB + 4kTB / R_L I_{ph} G_{int}^2) \qquad (3)$$

Eq.1c tells us that external amplification A_{ext} doesn't affect the SNR. Moreover, if we wish to preserve the SNR of the incoming photons we have to make the second term of the denominator in Eq.1c negligible respect to the first, and this can only be done by getting a high value of internal gain G_{int} [1]. If we use a photodiode with no internal gain ($G_{int}=1$) and connect it to a load R_L not large enough, then the SNR will be impaired and with a high external amplification we will only be able to observe the huge fluctuation generated by the Johnson-noise load resistance, not single photons.

The condition for the Johnson-noise to be negligible respect to quantum noise (term 2e $I_{ph}B$) is easily seen to be [1]:

$$I_{ph} > (2kT/e) / R_L G_{int}^2 = 50\text{mV} / R_L G_{int}^2 \qquad (4)$$

or, equivalently, we need a load resistance larger than

$$R_L > 50\text{mV} / I_{ph}\ G_{int}^2 \qquad (5)$$

For a photodiode with no internal gain ($G_{int}=1$), at a level of weak signals, down to a typical dark current, for example $I_{ph} = I_{dark} = 0.1$-nA, eq.2b gives $R_L > 500M\Omega$ as the required load. A value so large obviously impairs bandwidth $B=1/2\pi R_L C$ (where C is the stray capacitance seen at the output terminal) and prevents us from seeing a short single-photon pulse. Put another way, if we retain a low value for R_L to exploit bandwidth (for example 50 Ω to match cable impedance) the minimum current we need to be quantum limited (Eq.4) is $I_{ph} > 1$mA, order of magnitudes larger than the single-photon response.

On the other hand, if we are able to provide a sizeable internal gain G_{int} internal to the photodetector, then all the above numbers would be scaled by G_{int}^2, and we can have simultaneously a high sensitivity to small pulses of single photons and a wide bandwidth to observe them.

Thus, the quest for a semiconductor photomultiplier started with the search of a internal-gain mechanism, suitable for incorporation in the detector. Actually, there are several possible phenomena providing gain in semiconductor devices. Bipolar and unipolar transistors were obvious choice, but just to discover that in these devices gain is obtained at the expense of bandwidth, which decreases of the same amount gain is increased [1]. Also photoconductors share the same feature, a possibly high gain obtained at constant gain-bandwidth product [1]. Thus, while these mechanisms can be useful, and indeed they are used in the photo-sensitive counterparts of conventional electronic devices, they are not suitable to reach the high internal gain ($10^7..10^8$) required for single-photon operation together with a large bandwidth.

III. THE ANSWER: THE APD

Finally, in the late '60, a mechanism for internal gain at full bandwidth was identified in the avalanche multiplication that is, the impact ionization of the lattice by carriers at high electric fields, in the region called the second breakdown of the junction [10]. This is the avalanche photodiode (or APD) proposed by McIntyre [11]. In the APD, the gain depends on the ionizing powers α and β of electrons and holes, and on the drift pathlength L of the junction depleted layer. An analysis of the multiplication process, either by transport equations [11] or through a statistical modeling [12], shows that theoretically the gain G_{int} can be very large and even go up to infinity albeit with the assistance of positive feedback. Positive feedback is found because any ionization generates a pair of new charge carriers, one joining the parent charge and gonig the same direction, while the other heads in the opposite direction and will make another ionization backward, so that the process makes a loop. As well known, positive feedback reduces bandwidth and increases noise, and to avoid it we shall work at a moderate value of gain, not exceeding a specific optimal value, found by theory as the ratio of ionization coefficients [1,11,12]:

$$G_{opt} = \alpha/\beta \qquad (3)$$

If we exceed with this gain, using $G>G_{opt}$, bandwidth will be reduced by G/G_{opt}, and the noise figure will increase to a factor $2+ G/G_{opt}$.

Now, the question is: which material has the highest ionization ratio α/β ? Unfortunately, most semiconductors like e.g., Ge, GaAs, InP, and $Ga_{0.47}In_{0.53}As$ have ionization factors α and β differing by not more than a factor of 2. Only Si reaches a good $\alpha/\beta \approx 50...100$ at low fields, and the (less developed) ternary $Ga_{0.94}Al_{0.06}Sb$ reaches a good $\alpha/\beta \approx 100...200$.

Silicon APDs with the so-called reach-through structure [13] have been developed starting from the years '70, for instrumentation applications, exploiting an internal gain of $\approx$100.

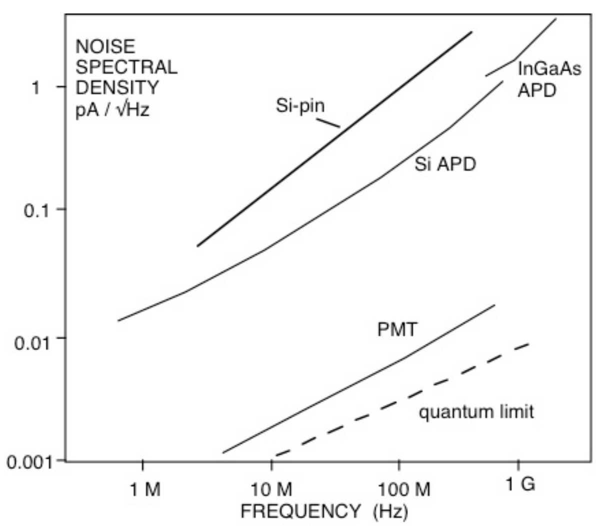


Fig.1 The APD (Avalanche Photo Diode) has a performance better than the corresponding pin-photodiode, but because of the limited optimal gain α/β they are about 2 decades off the performance of a truly single-photon detector, the PMT, which on its turn approaches the quantum limit of detection (from [7]).

Later, also $Ga_{0.47}In_{0.53}As$ with mesa structures on InP APD have been also introduced, with limited (typ. 10..15) gain, for receivers of optical signals in the 3rd window of fibers, as they are advantageous (despite the non-optimal gain and noise figure) and provide an SNR improvement of $\approx$8 dB [1] respect to usual pin photodiode detectors plagued by a huge Johnson noise of R_L.

As no natural semiconductor is available with the desired large α/β ratio, an idea was developed in the '90s [14] trying to circumvent the problem, by synthesis of the desired material with bandgap engineering. For example, in the multistep staircase structure, the composition is varied linearly from GaAs to $Ga_{0.55}Al_{0.45}As$ in each step, so that hole are confined to a low-filed drift and have a low ionization power, whereas electrons go across steps of very high field and raise their α-factor to 500...1000 times the β-factor of holes [1]. Similar behaviour happens in an APD with a multiple quantum well structure [14].

Unfortunately, all these structures are quite time consuming and critical to fabricate, and additionally, they don't go much farther than the optimal gain α/β of 100 of the plain Si-APD.

Therefore, we shall conclude that true single-photon detection cannot be approached with the APD working in the conventional *linear mode*, i.e., when the device is biased at the finite, optimum gain value α/β, and the detected power is read as a current $I_{ph}=\sigma P$.

To substantiate this statement, we can carry out a calculation on detector noise [15] and find out that the minimum number N of photodetected carriers we can detect in a packet, just at the threshold of SNR=1 using either a PIN or an APD with ideal pramplification, is about N=800 and 35, respectively.

So, the situation was looking no-way out in the 1980.

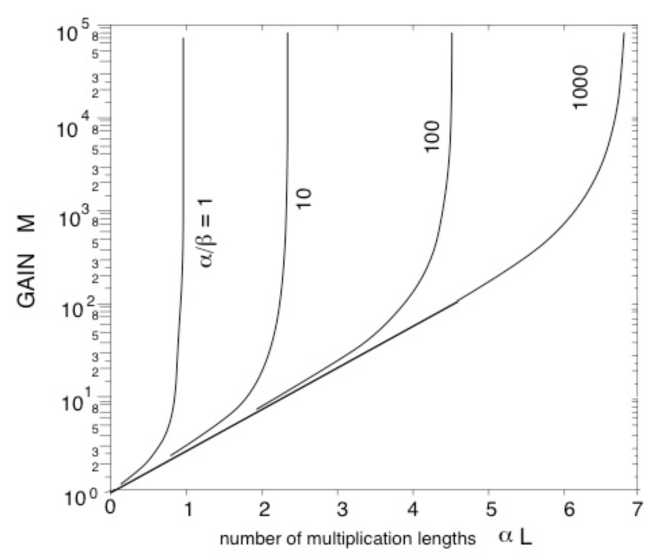


Fig.2 Gain of the APD plotted against αL, the number of multiplications of electrons in the avalanche region, and with α/β as a parameter. At all value of gain, there is an asymptote of infinite gain corresponding to positive feedback giving a loop gain in excess of unity (from [1])

But, as frequently it happens in physics, the next step came from a development almost unrelated, when Cova and co-workers [16] introduced the *Geiger-mode* of operation of the APD, now commonly known as SPAD (single-mode-avalanche-detector).

IV. THE APD BECOMES A SPAD

In the SPAD, we bias the device at a voltage $V_\infty+\Delta V$ slightly in excess of the infinity-gain voltage V_∞ (at which positive feedback makes the gain G_{int} diverge).
This can be done at all α/β ratio, as indicated (Fig.2) by the presence of vertical asymptotes for all α/β in the gain vs αL diagram; in practice, an overdrive ΔV of just a few Volts (or a few percent of V_∞) is usually adequate for proper working.

Once biased at $V=V_\infty+\Delta V$, the SPAD will stay in the "off" state (I=0) until the first charge carrier shows up in the active depletion layer of infinity gain, either a photoelectron detected by absorption of a photon, or a carrier (of dark current) thermally generated. Then, as $G_{int}=\infty$, the current is self-sustained and increases without limits, up to the value permitted by the external circuit. Excessive current is prevented (also to avoid damaging the device) by using a ballast resistance R connecting the SPAD to the dc supply V_{bb}, as shown in Fig.3. When the SPAD is triggered by a charge carrier, it will go short circuit and connect capacitor C, initially charged at V_{bb}, to the output load R_0. As C discharges on R_0, a pulse of voltage amplitude $\approx$-ΔV is generated, decaying with time constant $\tau=R_0C$ (Fig.3). After a few time constants τ, the anode voltage drops below $V< V_\infty$, the infinity gain ceases and current drops to zero.

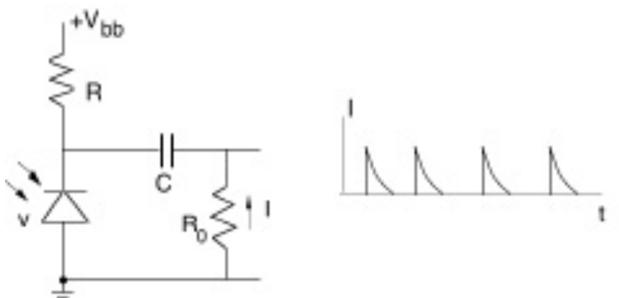


Fig.3 Typical SPAD bias circuit: R is the ballast resistance limiting the dc current through the device and allowing recharge the cathode voltage to V_{bb} with a time constant RC once the SPAD has triggered; C determines the time constant R_0C of current pulse (right) and the recovery time.

More time is however necessary to the SPAD for a complete recovery at the initial voltage $V=V_\infty+\Delta V$, as capacitor C recharges to battery V_{bb} through resistance R, and this will take a few RC time constants. An adequate recovery time is also needed to avoid afterpulsing, i.e., the re-trigger of the SPAD by the charges released by states trapping carriers in previous trigger.

During the recovery time T_R, the SPAD is at $V<V_\infty$ and doesn't respond to any new detected photoelectron or dark current carrier. So, photoelectrons arriving before recovery is completed will be ignored, and multiple photoelectrons in the same pulse time-slot will not produce any change in the pulse waveform respect to the single electron response.
Thus, we have two important differences in the SPAD respect to the PMT: (i) a *dead time* following each detected event, in contrast to the PMT that hasn't any dead time, and (ii) a *photon-at-a-time* type of response to events, with one or more photoelectrons in a burst counted as one, different form the PMT, in which a short burts of N photoelectrons is detected with a pulse N times larger than the SER (single-electron response) [1] pulse.
So, the SPAD is not really the counterpart of the PMT, rather, it is the equivalent of the Geiger-Muller detector (the gas-filled diode biased at a high electric field and driven into discharge by an incoming ray).

Dark current events bring about another limitation, due to the dead time consumed and subtracted to useful signal. In a SPAD with recovery time T_R, the maximum frequency of detectable photons is of the order of $f_{max}\approx 1/T_R$, while dark current sets a minimum frequency at $f_{min}=J_dA_d/e$, where J_d is the dark current density, A_d is the active area and e the electron charge. Given the typical values of J_d, to obtain a good dynamic range f_{max}/f_{min} we need to keep the area A_d very small, typically 10^{-4} to 10^{-2} mm^2. This figures can obviously be improved if we are able to shorten the recovery time T_R, form hundreds of ns of a passive circuit like in Fig.3, to a few ns with active quenching of the device [17]. But, the area still remains minuscule respect to the many cm^2 area of PMTs.

Albeit with some limitations, the SPAD is indeed a real single-photon detector, but its *Geiger-mode* of operation gives an output signal different from the *linear mode* operation of APD or PMT: as illustrated in Fig.4, the output is a sequence of pulses (of equal waveform) each for a detected photon, and frequency of pulses is proportional to the detected power. To recover an analogue replica of the input waveform, the pulse sequence shall be low-pass filtered so as to enhance its mean quasi-dc content.

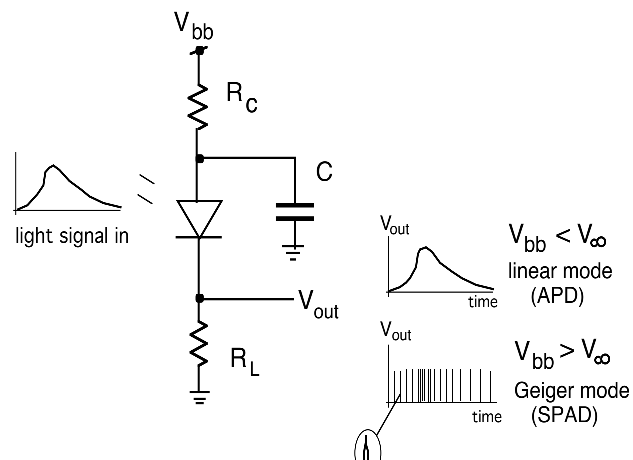


Fig.4 In the linear mode of operation (APD) the output signal waveform is a replica of the input, whereas in the Geiger mode of operation the SPAD supplies a seriers of pulse, all equal in waveform, whose frequency is proportional to the input light power.

After the first demonstrations of the SPAD concept, many groups have contributed internationally to remarkable improvement of its performance [18]. To reduce dark current, the initially used deep junction of the APD reach-through structure has been changed to a shallow junction. This gives the further advantages of easy compatibility to the standard CMOS process [19-21] and a reduced junction transit time and jitter of the pulse response (Fig.5). The only disadvantage of the shallow junction is that long–λ photons are poorly absorbed and spectral responsivity σ(λ) falls off already in the orange/red, much earlier than the λ=1100-nm cutoff of the Si photodiode.

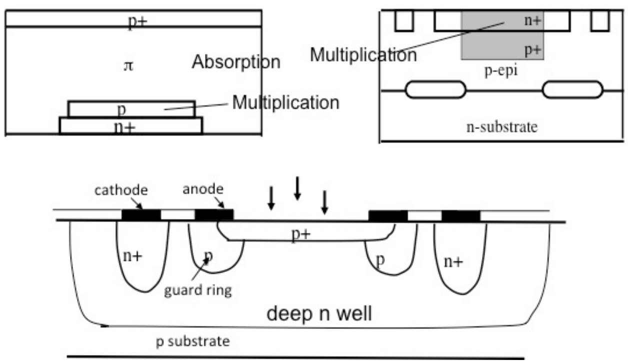


Fig.5 The early APDs [11] (top left) featured a reach-through structure with a deep depletion layer for absorption, required high bias voltage and had long transit time, whereas the first SPADs [16] (top left) used a much thinner absorption layer that reduces bias voltage and improves response time and dark current, at the expense of an earlier red-cutoff. Recent improvements [20,21] (bottom). use a deep p-guard ring to control edge breakdown and a n-trench to reduce diffusion and dark current, in an SPAD structure which is CMOS-compatible.

Thanks to these progresses, the SPAD offers nowadays good performances already as a single-element Si-photodetector. For example, a 50-μm diameter sensitive area SPAD may have typically a quantum efficiency [1] up to η=50-60% in the wavelength range λ=450-650nm, the dark current frequency is a reasonable 1-10 kHz, while the timing resolution is an excellent 30-50 ps, and the dead time is <1-μs. Further, SPAD have been successfully fabricated in other materials with low β/α (e.g. InGaAs on InP substrate) with threshold wavelengths up to 1.6 μm [23].

V. IMAGE SPADS AND THE SMART-PIXEL CONCEPT

Two features of the SPAD lend themselves straight to the *image format* extension of the single element device: the small area, which is suitable to be the pixel of a large array, and the excellent time resolution that allows us to develop time-of-flight and time-resolved analysis on image-format targets, whence the name *3-D* frequently given to this array.

The image-SPAD technology has been recently demonstrated by Charbon and co-workers [20] and by Zappa et al. [21]. Both Groups were able to build a SPAD array of typical size 32x32 (extensions to larget format are under way), and the breakthrough of these work has been that, for the first time, circuits for processing the detected signal have been incorporated internal to the pixel, next to the photosensitive element.

This possibility has been till now discarded in view of the area consumed by the accessory circuits, reducing the fill-factor FF of useful sensitive area and hence the effective sensitivity. But, fine line lithography of present years allow to squeeze circuitry making substantial functions down to a small area and accommodate it around the SPAD sensitive area with little footprint. However, a limit to fine-line lithography is the thinner epilayer decreasing absorpion of photons and hence reduced σ(λ).

On the other hand, we can take advantage of an invariant of radiometry, as introduced in Refs.[24,25], to recover the fill-factor. As detailed in Ref.[1], thanks to invariance of acceptance *a*=AΩ, we can trade an (effective) area increase for an equal solid angle decrease, and be able to recover the fill-factor FF at the expense of a tolerable reduction of the NA (numerical aperture) of the objective lens placed in front of the detector.

So, the combination of a small-area SPAD (typ. 5-μm in diameter) in a pixel of 50-μm diameter accommodating the circuits of avalanche quenching, low-noise preamplifier, time-to-amplitude concerter and line driver [20,26], with a front lens of 50-μm diameter, completely clears the area FF=1% loss, and recovers by a x100 concentration the full sensitive area of a 50-μm diameter receiving surface.

The concentration requires an optical element in front of the SPAD pixel [24], that can be a reflective tilted-paraboloid for optimal performance, or a plano-convex microlens for easy technology. An array of PET-polymer molded microlenses, with 50-μm spacing and 32x32 elements has been used in connection to a 32x32 SPAD array [25] with a demonstrated concentration factor of C=35. This is just the first achieved result, while with better fabrication control, practical C factors up to 200 appear feasible (for use with collimated input signal).

VI. THE SOLID-STATE PHOTOMULTIPLIER

The long dead-time of SPAD is the most serious impediment to approach the PMT capability of detecting 1,2 ...N photons simultaneously. Additionally, we would like the active area is increased, at least to become not that negligible compared to the many cm^2 of PMTs.

Both requirements can be satisfied by the idea introduced by Saveliev and Golovin [27,28] who proposed to segment the receiving area (say 1-mm^2) in an x,y array containing a large number N of individual SPADs (for example N=400 using 50x50 μm^2 individual SPADs or pixels). The SPAD outputs currents are all summed up to a single load R_L (Fig.6), but each of the SPAD has its own separate quenching circuit, integrated around the sensitive area.

Now, let us consider incoming photons that constitute the optical signal to be detected. Due to the statistical nature of the field, each photon arrives at a random position of the x,y array, independent from previous or next photon (fig.6). So, the rate at which one particular SPAD of the array is hit by another photon is reduced by a factor N.

If we take N larger than the product $T_R f_{ph}$ of recovery time and photon rate, each pixel has enough time to recover

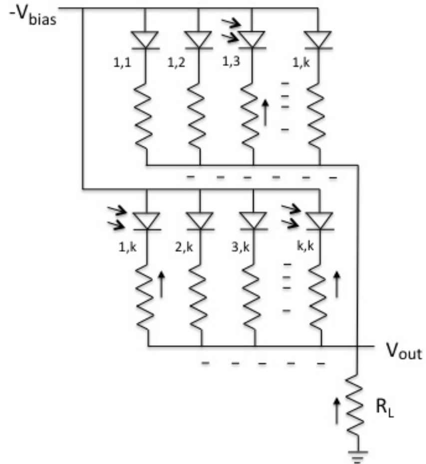


Fig.6 In an SSPM, the device is arrayed into a large number of individual pixels, each with an SPAD. Output currents are all summed up in a common load R_L. If three photons are received simultaneously, because of statistics they will hit separate pixels and trigger different SPADs, thus injecting into load R_L a current 3 times the single-pixel current. To allow separate quenching of each individual SPAD of the array, each pixel has its own ballast resistance.

before next photon arrives.
Thus, dead-time effect are cleared, and we can see the response of a packet of (closely spaced in time) *m* photons hitting the sensitive area, with a peak of output current of amplitude 1,2,...*m* times the single photon amplitude (Fig.7). Different from the PMT, as the charge variance of the SPAD is small, individual packets of 1,2,3...R electrons are resolved even for a Poisson distributed number R of incoming photoelectrons.

Also, as the junction is shallow, transit time and avalanche build-up time are quite fast, and typically the SER jitter is a in the range of τ=100...200-ps FWHM, or τ=42...84-ps rms.

Dark current of the N-element array is the same of the equivalent device area, but now we can readily shrink the individual SPAD to much less than the 50-μm spacing, for example to 5-μm diameter and use a C=100 lens-array concentrator: we then can obtain a reduction of the dark current by a factor C (=100).

VII. Comparing SSPM to PMT

With the recent advances, the SSPM has become a true single-photon photodetector from the functional point of view, capable of detecting single and multiple photons in a packet.

The SER (single electron response) of the SSPM has a temporal jitter (or, random time delay) smaller than that of conventional PMT, because of the much shorter path to be covered internal to the the device (a thin, μm-size depleted layer compared to the cm-size interdynodic distance). Thus, we have consistently an SER rise time τ_r of the order of 100-500 ps, and a jitter τ_0 that can be down to 20-50 ps, better than the 1-3 ns rise time of plain PMTs - yet not much better than the 100-300 ps of MCP (multi-channel-plates) PMT units [1]. This shorter time is an advantage in detection and mesurements on fast waveforms and decay times of weak optical signals [7].

About timing and time sorting, in applications like correlation and time-of-flight measurements, we are interested in the variance σ_t^2 of the timing signal extracted from the photodetector waveform [7].

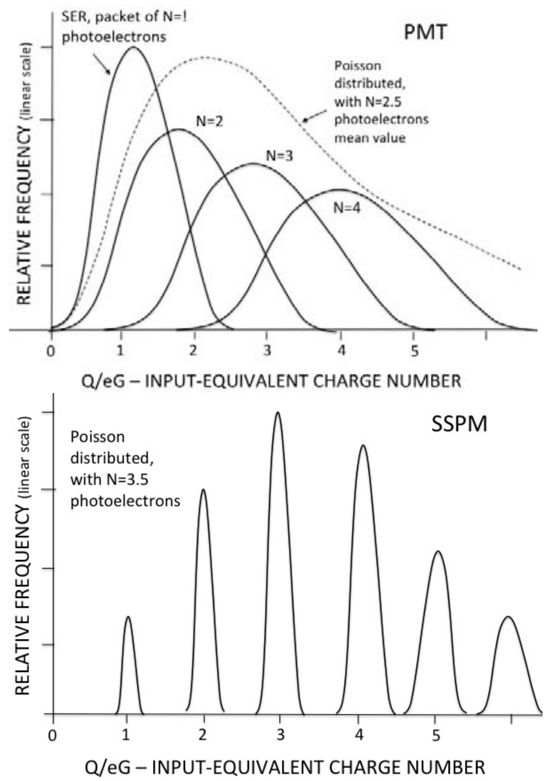


Fig.7 In a PMT, the input-equivalent charge distribution (output charge divided by *e* and gain G) for exactly 1,2, ...R detected photoelectrons has a rather large spread, due to the large variance of the multiplier ε^2_A (≈0.3 typ.), so that on detecting a Poisson-distributed packet of photons we get a broad curve (top, dotted line). In the SSPM the charge variance is smaller (0.05 to 0.1) and we can resolve packets of 1,2,...,m single photoelectrons (bottom) even when the number R is Poisson distributed.

Usually, we perform the timing operation with the aid of a threshold crossing circuit [giving an output pulse when a threshold I_0 is exceeded, $i(t) > I_0$], or by a cantroid sensing circuit [one calculating the center of gravity of the i(t) waveform]. For the centroid timing, the time variance is [1,7]:

$$\sigma_t^2 = [\tau_0^2+\varepsilon^2_A\,(\tau_{SER}/2.36n)^2]\,/\,R \qquad (6)$$

where:
τ_0 is the jitter of the SER response, assumed of rigid shape [is the time-of-flight between photocathode and first dynode [1] for PMT, a value usually in the range 100-200 ps, or the jitter of the avalanche buildup, usually a fraction (0.1 to 0.2) of the 120-ps (typ.) FWHM for the SSPM];
ε^2_A is the variance of the multiplier (≈0.3 typ. for PMT and ≈0.05..0.1 for SSPM) and n the number of dynodes (formally n=1 for the SSPM);
τ_{SER} is the FWHM (full-width-half-maximum) of the SER,

typically 2-5 ns in PMT and 0.5-2 ns in SSPM, depending on output circuit parasitic capacitance;

R is the number of detected photoelectrons, given by $R=\eta F$ in terms of the number of photons F and of the quantum efficiency η, where $\eta=\sigma\, hc/e\lambda$;

In Eq.6, factor 2.36 stands for the FWHM to rms standard deviation ratio of the SER, assumed Gaussian distributed, and the factor *n* is due to the of the time-of-flight averaging on *n* the stages of dynode cascade.

The dependence on 1/R of time variance allows us to make measurements well below the individual contributions τ_0 and τ_{SER}. Typically, with fast circuits, we can go down to a few ps for σ_t in both PMT [7] and SSPM, albeit with a larger photoelectron number for the former, i.e., we may have R of the order of $\approx 10^3$ (PMT) and ≈50 (SSPM).

About possible improvements, in the SSPM the jitter τ_0 is in the 20-50 ps range of values, and no great progress can be expected in a vertical structure, as the parameters governing it (electric field, thickness, saturated speed) are already close to maximum.

In amplitude charge measurements (for application to nuclear pulse spectrometry [1]), the governing parameter is the output number *N* of electrons collected after a light pulse is detected (or, equivalently, the collected charge $Q=eN$). Introducing the variance σ_N^2 of collected number of electrons, and letting $<N>$ for the average value, we have for the relative variance (equal to the inverse of SNR of the measurement) [1]:

$$\sigma_t^2 / <N>^2 = (1+\varepsilon^2_A) / R \qquad (7)$$

As ε^2_A is even smaller in the SSPM respect to the PMT, the SSPM is well suited to this application, provided we are able to couple a scintillator of appreciable volume to collect incoming radiation to the tiny input surface of the SSMP. In contrast to cm-size cylinder-shaped scintillators commonly used with PMTs, using SSMPs we can use a rod- or fiber-shape scintillator and take advantage of the guiding effect by internal reflection at the walls and have substantial detection volume, reported in Refs.[30-32].

In amplitude measurements, the SSMP can read a maximum signal frequency rate up to $\approx N/T_R$, while the minimum signal is $1/T_R$; so we have a dynamic range ≈N, of a few 10^2 typically. In the PMT, the maximum signal is set by the saturation current of the last dynode [1], typ. ≈100 mA, while the single-photon current in high gain operation is a ≈1mA peak. Thus, at high gain ($G=10^8$) the ratio of the two currents sets the dynamic range is a few 10^2. But, if we lower gain by a factor K, the dynamic range will increase of K and then we may easily go up to 10^5 . Linearity is good in both devices up to about one decade below the dynamic range limit, i.e. the integral linearity error may be <1% up to 1 decade below and <5% half decade below the maximum amplitude set by dynamic range.

About wavelength coverage, the spectral sensitivity $\sigma(\lambda)$ of the SSPM is comparable in amplitude to that of PMT photocathodes, but the spectral response is somewhat narrower. By working with a larger overdrive bias ΔV we can increase height and width of the SSPM $\sigma(\lambda)$, but at the expense of an increase of dark current.

Important difference, while PMTs cover the visible and near infrared (NIR) with multialkali antimonides up to λ_t=850-nm and with ternary semiconductors (yet in transmission only) up to λ_t=1700-nm, and nothing is available beyond λ_t=2μm, the SSPM concept and the single-photon performance can be extended and cover the middle- and far-infrared ranges of wavelengths with the appropriate materials (InSb for MIR and HgCdTe or PbSnTe for the FIR). An example of InGaAs on InP SPAD covering the range 0.6-1.4-μm with high quantum efficiency, η=35% and low dark count rate (≈1kHz) has already been reported [23, 33].

Dark current is the weak point of SSPM, presently order of magnitude larger than the pA or fA per cm^2 of sensitive surface area of PMTs. Thus, the photocounting technique with an SSPM can extend the minimum detectable signal well below dark current, typically down to a photoelectron rate F_{min} [1]:

$$F_{min} = \sqrt{(2\eta_d I_d / eT)}/\eta \qquad (8)$$

where η is quantum efficiency, $\eta_d I_d$ is the effective dark current pulses and T the integration time. As $\eta_d I_d \approx 10^{-14}$ A/mm^2 typical current density for an SSPM [32], about 10^5 times larger than in a PMT, the minimum rate at equal T is ≈300 times off the PMT level.

The other performance that has the PMT still in the lead is the active surface area available: very large photocathodes (up to 80 cm diameter as in the Kamiokande experiment [1]) appear still out of reach by SSPM, though we can use the acceptance invariance to increase substantially the effective collecting area at the expense of a reduction of the field-of-view of the device.

VIII. CONCLUSIONS

We have summarized the milestones of development of single-photon detectors, and pointed out the key factors and the conceptual problems that have been solved to open the way of modern SPAD. Based on a comparison of the performances, we have concluded that SSPMs are equalling or even surpassing PMTs for fast response and sensitivity, while are still lagging for acceptance area, linearity and dark current. Finally, we have discussed the use of SSPMs in a number of applications typical of single-photon detectors.

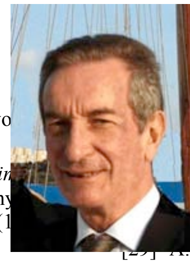
REFERENCES

[1] S. Donati: “*Photodetectors*“, Prentice Hall, Upper Saddle River, N.J., 2000.

[2] B.K. Lubsandorzhiev: "On the history of photomultiplier tube invention" *Nucler Instr and Meth.* 567, (2006) pp.236-238

[3] W.K. Zworykin, G.A. Morton: “*Television*”, J. Wiley & Sons, New York 1939; 2nd edition: Pergamon Press, Oxford, 1955.

[4] S.Donati, E.Gatti, V.Svelto: "The Statistical Behavior of the Scintillation Detector: Theory and Experiments", in: *Advances in Electronics*



Electron Physics, edited by L.Marton, Academic Press, vo... ...9) pp.251-304.

... Poultney: *“Single Photon Detection and Timing: Experim... Techniques”* in: Advances in Electronics and Electron Phy... ...ed by L. Marton, Academic Press, New York, vol.31 (1... pp.39-116.

[6] S. Donati: “Photomultipliers” in: “*Encyclopedia of Biomedical Engineering*”, ed. by M.Akay, J.Wiley and Sons, 2006, ebs092. ISBN: 978-0-471-24967-2

[7] S. Donati: “*Electrooptical Instrumentation*”, Prentice Hall, Upper Saddle River, N.J., 2004.

[8] R.C.Jones: "Performances of Detectors for Visible an Infrared Radiation", in: *Advances in Electronics and Electron Physics*, edited by L.Marton, Academic Press, vol. 4 (1952) pp.2-88.

[9] J.P. Boutot, J. Nussli, D. Vallant: *“Recent Trends in Photomultipliers for Nuclear Physics”*, in: Advances in Electronics and Electron Physics, edited by L. Marton, Academic Press, New York, vol.60 (1983) pp.223-305.

[10] S.M. Sze: “*Physics of Semiconductor Devices*”, 2nd ed., Wiley-Interscience, New York, 1981.

[11] R.J. McIntyre: “*Multiplication Noise in Uniform Avalanche Junctions*”, IEEE Transactions on Electron Devices, vol.ED-13 (1966), pp.164-168.

[12] S.Donati, V.Svelto: “*Statistical Behavior of the Avalanching Photodiode”*, Alta Frequenza, vol.37 (1968), pp.476-486.

[13] R.J. McIntyre, P.P.Webb: "*Low-noise, Reach-Through, Avalanche Photodiodes*" US patent 5,583,352, Dec.1996.

[14] F. Capasso: “*Resonant Tunneling through Double-Barriers, Perpendicular Quantum Transport Phenomena in Superlattices and their Device Applications”*, IEEE Journal of Quantum Electronics, vol.QE-22 (1986) pp.1853-1869.

[15] S. Donati: “*Solved Problems in Photodetectors*“, Prentice Hall, 2000, Problem 5.14, download at http://www-3.unipv.it/donati.

[16] S. Cova et al.: “*20 ps Timing Resolution with Single-Photon Avalanche Photodiodes”*, Rev. of Sc. Instr., 60 (1989) pp.1104-1110.

[17] S.Cova, M. Ghioni, A. Lacaita, C. Samori, F. Zappa: "*Avalanche photodiodes and quenching circuits for single-photon detection*", Appl. Optics vol. 35 (1996), pp.1956-1976

[18] M.Itzler, S.Donati, M.S.Unlu, K.Kato: "*Photodetector and Imaging*", Special Issue of the Journal Selected Topics in Quantum Electronics, vol.10 (2004), pp. 668-840.

[19] E.Charbon, S.Donati “*SPAD Sensors Come of Age*”, OSA Optics and Photonics News, vol.21, (2010), pp.34-41

[20] C.Niclass, A.Rochas, P-A.Besse, E.Charbon: *"Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes"***,** J. Solid State Circ., 2003, vol.40, pp.1847-1854.

[21] L.Panchieri, M.Scandiuzzo, D.Stoppa, G.F.DallaBetta: "*Low-Noise Avalanche in Standard 0.35μm CMOS Technology*" IEEE Trans. Electr.Devices, vol.55 (2008) pp.457-461.

[22] F.Villa, B.Markovic, S.Bellisai, S.Tisa, A.Tosi, F.Zappa, S.Tisa: "*SPAD smart pixel for time-of-flight and time-correlated single-photon counting measurements*", Phot. Techn. Lett., vol.43 (2012), pp.795-804.

[23] M Itzler, X.Jiang, M.Entwistle, K.Slomkoski, A.Tosi, F.Acerbi, F.Zappa, S.Cova: "*Advances in GaInAsP-based Avalanche Diode Single Photon detectors*", Modern Optics, vol.58 (2011), pp.174-200.

[24] S. Donati, G. Martini, E. Randone: "*Improving Photodetector Performance by means of Micro-optics Concentrators*", IEEE J. Lightwave Technologies, vol.29, 2011, pp.661-665.

[25] S.Donati, G.Martini, M.Norgia: “*Microconcentrators to recover fill-factor in image photo-detectors with pixel on-board processing circuits*”, Optics Express, vol.15, 2007, pp. 18066-18074.

[26] E.Charbon, S.Donati: "*An Ultrafast Single-Photon Image Diagnostics Sensor with SPAD Arrays for Industrial and Bio-Applications*", Proc. Adv. laseer Techn (ALT) 2009, Antalya Oct. 4-7 2009; Fizika, Adzerbaijan Journal Physics, vol.16 (2010) pp.64-67.

...Saveliev, V.Golovin: "*Silicon avalanche photodiodes on the base ...metal-resistor-semiconductor (MRS) structures*" Nucl. Instr. and ...eth. A 422, (2000) pp.223-229.

...Saveliev, V.Golovin: "*The recent development and study of silicon ...otomultiplier*" Nucler Instr and Meth. 535, (2004) pp.528-532.

[29] A.G.Stewart, V.Saveliev, J.S.Bellis, D.J.Herbert, P.J. Hughes, J.C. Jackson: "*Performance of 1-mm^2 Silicon Photomultiplier*" IEEE J. Quant. Electr., vol.44 (2008), pp.157-164.

[30] M.D.Petroff, M.Attac: "*High energy particle tracking using scintillating fibers and solid state photomultiplier*" IEEE Trans, Nuclear Science, vol.36 (1989) pp.163-63.

[31] C.Hudson, C.Roy E.Fenyves, H.Hammack, P.Antich: "*Measurement of the energy resolution of a scintillating fiber detector*", Proc.SPIE vol.2281, Scintillating Fiber Technology and Applications II, J.E. Fenyves (editor) 1994, pp. 65-70.

[32] M.A.Karami, M.Gersbach, H.-J. Yoon, E.Charbon: "*A new single-photon avalanche diode in 90-nm standard CMOS technology*", Optics Express, vol.18 (2010),pp.22158-166.

[33] M.A.Itzler, X.Jiang, M.Entwistle: "*Power law dependence of InGaAs/InP SPAD afterpulsing*" J. Modern Optics, vol.59 (2012) pp.1472-80.

Silvano Donati (M’75, SM’98, F’03, LF’09) earned his doctorate in Physics (Laurea in Fisica) *cum laude* from University of Milan, Italy, 1966, and has been Chair professor of University of Pavia from 1980 to 2010. Presently, he is Visiting Professor at National Chung-Hsing University, Taichung, Taiwan, and also Lecturer in Optoelectronics at University of Pavia. He has authored or co-authored 300 papers and holds a dozen patents. He has introduced self-mixing interferometry and chaos-shift-keying cryptography, the topics covered in his Distinguished Lecture talk given in 21 LEOS (now IPS) Chapters in two terms, 2007/08 and 2008/09. He has authored two books, ‘Photodetectors’ (Prentice Hall, 1999) and ‘Electro-Optical Instrumentation’ (Prentice Hall, 2004). He was the founder (1996) and first Chairman (1997-01) of the Italian LEOS Chapter, LEOS VP Region 8 Membership (2002-04) and BoG (2004-06), and Chairman of the IEEE Italy Section (2008-09). He has been Visiting Professor at several Universities of Taiwan, NTU, Taipei, 2005, Sun Yat Sen, Kaohsiung (2007, 2008, 2010), and NCKU, Tainan 2012. He is Life Fellow of the IEEE, and OSA Emeritus Fellow.

Tiziana Tambosso (M’94, SM’02) received the Laurea degree with honors in Electronic Engineering and the PhD in Optoelectronics from University of Pavia, Italy, in 1983 and 1988. Presently, she is Visiting Professor at Da Yeh University, Dacun, Taiwan, lecturing *Green Photonics* and *Laser Instrumentation*. In 2012 she was a lecturer of Master course on *Photovoltaics* at University of Palermo (Italy).

She has been the Head of fiber optic devices group at SIRTI, R&D (1989-93), and project leader at CSELT-TILAB (1993 to 2005) for fiber optic components and amplifiers, photomixing. wireless sensor and RFID. Her main research interests are now on laser instrumentation and green photonics. She holds a dozen of patents and has authored or coauthored about 50 papers on Journals and Conference Proceedings. From 1993 to 1997 Dr. Tambosso acted as Secretary and Vice-Chair of SC86B on fiber optic interconnection devices and passive components. She has co-chaired or TPC chaired, three editions of WFOPC, the Workshop on Fibers and Optical Passive Components of IEEE-LEOS, Pavia (1998 and 2000) and Taipei (2007). She has served as Guest Editor of the IEEE Selected Topics in Quantum Electronics of Sept. 1999 on Fiber Optic Passive Components. Dr. Tambosso has been THE Chairperson of IEEE LEOS Italian Chapter (now IEEE Photonics Italy Chapter) from 2002 to 2012. Presently she is Chapter Activity Coordinator and Section Program Chair of IEEE Italy.