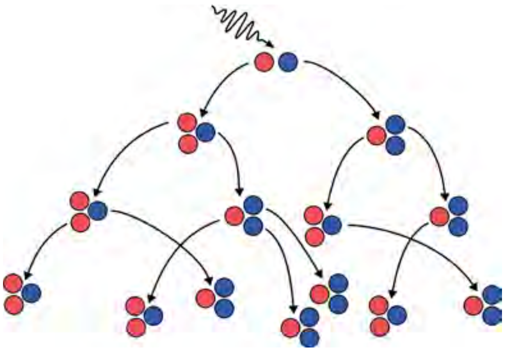
SINGLE PHOTON AVALANCHES DIODES

Angelo GULINATTI*
Dipartimento di Elettronica, Informazione, e Bioingegneria, Politecnico di Milano, 20133 Milano, Italy
***angelo.gulinatti@polimi.it**



Twenty years ago the detection of single photons was little more than a scientific curiosity reserved to a few specialists. Today it is a flourishing field with an ecosystem that extends from university laboratories to large semiconductor manufacturers. This change of paradigm has been stimulated by the emergence of critical applications that rely on single photon detection, and by technical progresses in the detector field. The single photon avalanche diode has unquestionably played a major role in this process.

https://doi.org/10.1051/photon/202412563

Light sensors like pin photodiodes, CCDs, or CMOS image sensors generate a signal with an amplitude proportional to the amount of light hitting the detector. While this signal can eventually be converted in a digital form for easiness of elaboration, they are natively analog detectors. Single photon detectors are radically different, as they generate a standard pulse for each photon hitting the detector. The information about the light intensity is not encoded in the pulse-amplitude but can be obtained by counting the number of pulses in a short amount of time.

In the first part of this article, we will provide a few examples to illustrate why photon counting has become so important in many fields, ranging from fundamental science to consumer electronics. Then we will focus on a specific type of detector, the single-photon avalanche diode or SPAD, that has become very popular thanks to its combinations of remarkable performance, affordability, and easiness of use. We will review its working principle, its performance metrics, the materials, and we will briefly compare it with other single-photon detectors.

WHY COUNTING SINGLE PHOTONS?

The capability of detecting single photons has opened the way to a multitude of exciting new applications that would not have been possible with analog detectors. While the measurement of **faint optical signals** may appear the most obvious domain of application, there are other and even more important fields in which these detectors are essential. The emerging fields of optical **quantum information processing** and quantum sensing rely on the quantum properties of a photon to enhance data communication, processing, and sensing. For example, quantum key distribution (QKD) exploits the impossibility of cloning the quantum state of a photon to detect the presence of an eavesdropper; ghost imaging takes advantage of the correlation between entangled photons to image a scene by measuring photons which have never interacted with the scene itself. All these applications require to take decisions depending for example on which path a photon has taken, in which timeslot it has arrived, or whatever two photons have arrived at the same time or not. The capability of signaling the arrival of a single photon is therefore crucial.

Single photon detectors not only excel in this task, but most of them can also do it with an excellent temporal precision, down to a few tens of picoseconds. This property is widely exploited for the reconstruction of **ultrafast optical** •••

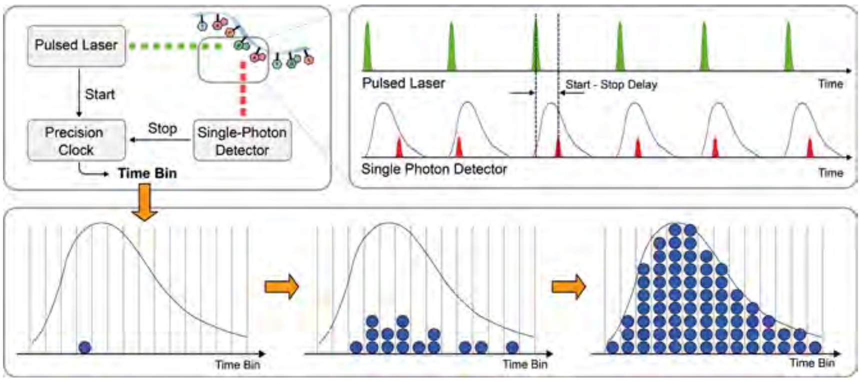


Figure 1: Fluorescence lifetime measured by time-correlated single-photon counting (TCSPC). The sample is excited by a pulsed laser and the delay between the excitation pulse and the emitted photon is measured by a precision clock. By repeating multiple times, it is possible to build a histogram of the delays that reproduces the shape of the optical signal.

signals. A typical example is the measurement of the fluorescence emitted by a sample when excited by a short (ps or fs) laser pulse (Figure 1). If the intensity is low enough (either intrinsically or because it has been purposely attenuated) the quantization of the light will result in (at most) one photon reaching the detector at a random time, with a probability distribution that follows the shape of the optical signal. Therefore, by repeating the experiment multiple times, it will be possible to build a histogram of the arrival times that reproduces the shape of the signal with tens-of-ps details. This technique is known as time-correlated single photon counting (TCSPC). Attaining the same level of detail with conventional techniques is incredibly challenging as it would require an entire acquisition chain (detector, amplifiers, analog-to-digital converters, etc.) operating at tens of gigahertz.

The temporal precision of single-photon detectors is exploited in a multitude of applications, like LiDAR. In this case the 3D features of a scene are reconstructed by measuring the time it takes for a laser pulse to travel from the source to the scene and back to the detector.

SINGLE PHOTON AVALANCHE DIODES

PRINCIPLE OF OPERATION

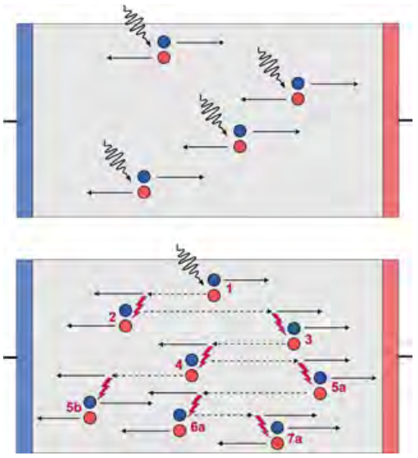
Photodiodes are light sensors that rely on a semiconductor material to convert a photon flux in an electric current. When absorbed in the semiconductor, a photon generates two carriers of opposite charge, an electron and a hole. In a common photodiode, the carriers are accelerated in opposite direction by an electric field and are collected at the electrodes, resulting in a current proportional to the number of incident photons (Figure 2, top). When moving along the detector, the carriers are accelerated by the electric field and, once in a while, they release the acquired energy by colliding with the semiconductor reticle. Usually, this process has no significant effect on the photodiode behavior; but, if the electric field is increased, the energy acquired by the carriers can become high enough to induce ionization, *i.e.* to break the bonds in the reticle and create another electron-hole pair (Figure 2, bottom). This process increases the number of circulating carriers and tends to self-sustain. In fact, as the two carriers move in opposite directions, one will be travelling toward the edge of the device while the other will head toward its center where it can generate additional pairs. The new carriers are in turn accelerated, generate further pairs, and so on.

If the average number of pairs created in each iteration is lower than one, the process comes to an end after some iterations; the electric current recorded at the electrodes is again proportional to the number of incident photons, although a factor M larger than in a conventional photodiode. In this case the detector is referred to as an APD or avalanche photodiode. On the contrary, if the electric field is so high as to create more than one pair per iteration, the multiplication process diverges and the current at the electrodes increases exponentially (until some saturation phenomena occurs). This allows to easily detect the absorption of a single photon and the detector is called single-photon avalanche diode or SPAD. Once this self-sustained avalanche multiplication has been triggered in a SPAD, the absorption of an additional photon does not significantly change the current flowing in it, so the detector is effectively blind. To be able to detect additional photons, the SPAD must be coupled with a quenching circuit that turns the avalanche off and resets the initial operating conditions.

PERFORMANCE METRICS

When selecting a SPAD, there are many performance metrics to look at. Many of these metrics are in trade-off with each other (see Figure 3 for some examples), and the relative importance of each of

Figure 2: In analog photodiodes (top) each photon generates a pair of carriers, which are collected at the electrodes. In SPADs (bottom) a single photon starts an avalanche process resulting in a self-sustained current.



them strictly depends on the applications. So, in single photon detection there is (still) not a one-size-fit-all solution.

Size of the Active Area. The active area, *i.e.* the part of the SPAD which is photo-sensitive, may extend laterally from a few microns to a few hundreds of microns. Diameters ranging from 20 – 100 µm are usually preferred to collect the light from a microscope. Larger diameters may be needed when the light scattered by the sample cannot easily be focused on the SPAD, as it happens for example in two-photon microscopy and diffuse optical tomography. By contrast, smaller diameters are preferred when the light comes from a single-mode fiber or when SPADs are arranged in large arrays to build an image sensor.

Photo Detection Efficiency (PDE). A photon impinging on a SPAD may go undetected either because it is not absorbed in the semiconductor material or because the photogenerated carriers fail to trigger an avalanche. This can be quantified by looking at the photon detection efficiency (PDE), which represents the probability that a photon impinging on the detector active area successfully triggers an avalanche. The PDE is especially important when the number of available photons is limited. An example is the single molecule analysis, where a freely diffusing molecule spends only a limited amount of time in the focus of a confocal microscope. Even more critical are those applications in which coincidences must be measured. In fact, the coincidence of n photon-events is properly recorded only if all the n SPADs involved succeed in detecting the corresponding photon. The probability of success scales therefore with $(\mathrm{PDE})^n$.

Dark Count Rate (DCR). In a semiconductor, an electron-hole pair can occasionally be generated by phenomena different from a photon absorption. For example, impurity atoms favor the thermal breaking of the bonds that keep an electron tied to the reticle, leading to the formation of a pair. Such pairs are as effective in triggering an avalanche as those that have been photogenerated. That means that a SPAD can fire even in dark, and the average number of events recorded in absence of light is called dark count rate (DCR). While the DCR can be assessed and subtracted from the measurement, its statistical fluctuations remain and may hide the useful signal. The lower

●●●

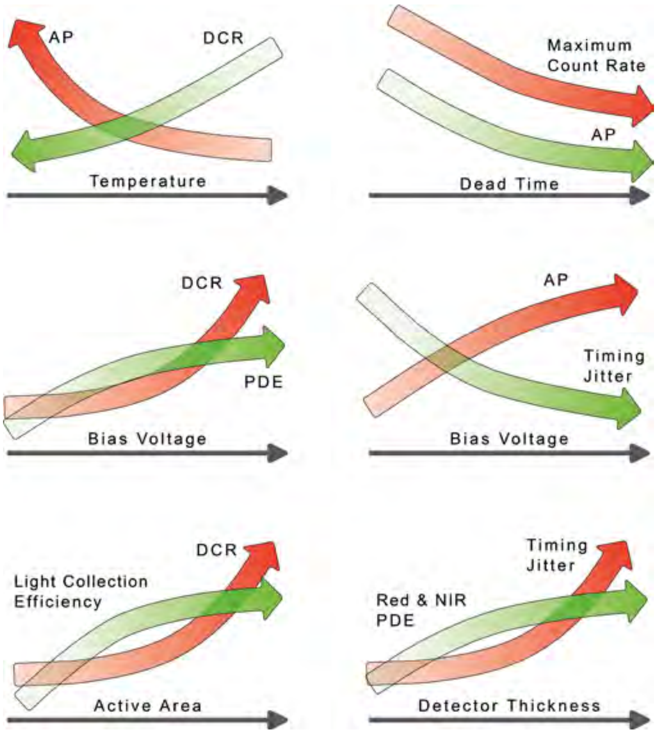


Figure 3: By changing the operating conditions or the design parameters, it is possible to improve some performance metrics at the expenses of others.

is the DCR, the lower are also its fluctuations. A low DCR value is especially important whenever faint light signals must be acquired.

Afterpulsing Probability (AP). During an avalanche, a large number of carriers flows through the device. If one of them is trapped in a defect and is released at a later time, it may trigger an additional avalanche known as afterpulse. Afterpulsing is especially deleterious where a SPAD is used to measure temporal correlations. An example is fluorescence correlation spectroscopy (FCS), where the autocorrelation in the light emitted by a small number of molecules is used to assess molecule properties like the diffusion coefficients.

Dead Time. The detector is inevitably blind for the time it takes to quench the avalanche and reset the initial operating conditions. In addition to this, the detector is intentionally kept off for a time sufficient to release most of the carriers trapped in the previous avalanche without re-triggering a new one. The overall time for which the SPAD remains blind is called dead time. The dead time limits the minimum time-distance at which two photons can be detected. Short dead times are thus needed to operate the detector with high photon fluxes or to measure short interval correlations.

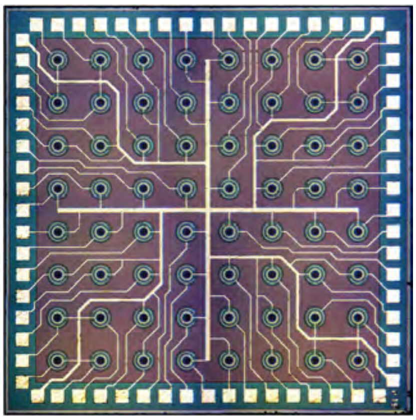
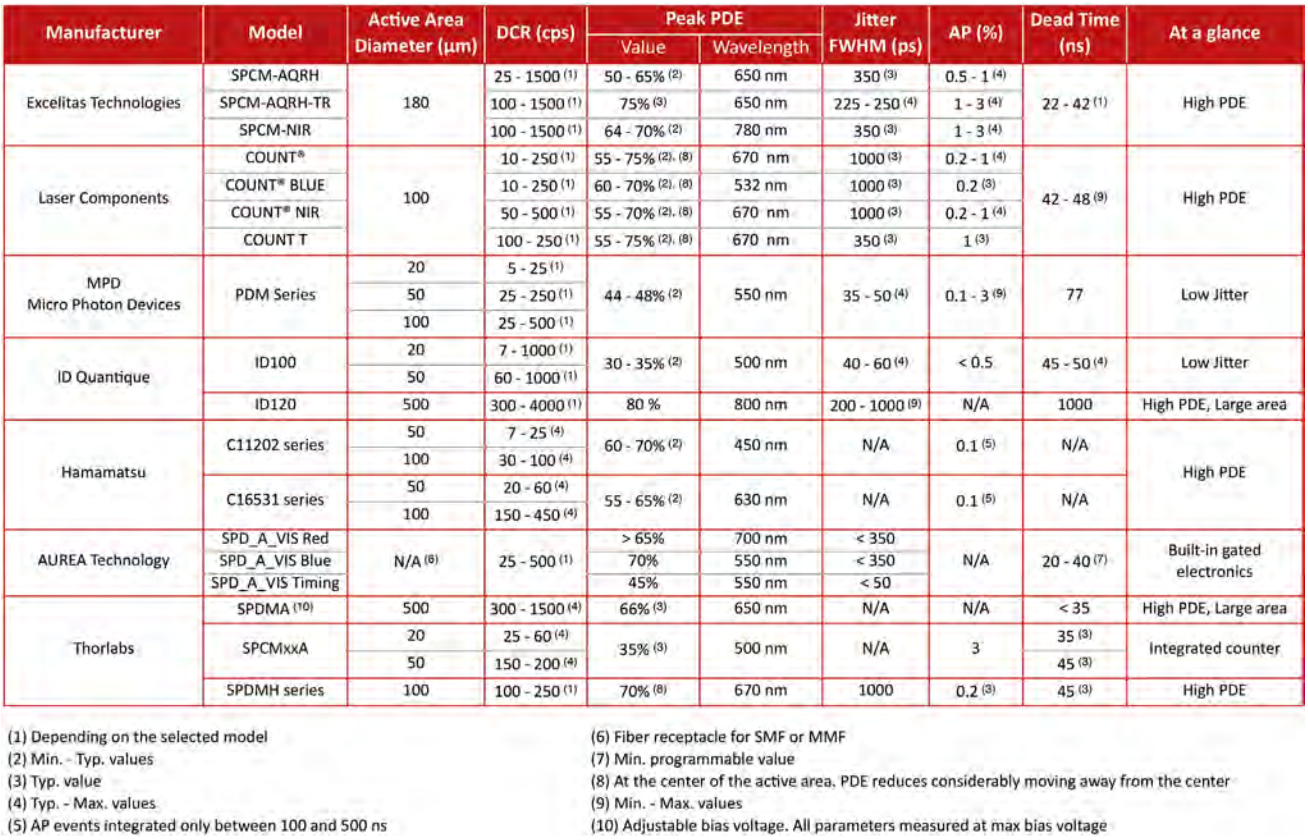


Figure 4: An example of an array of 8 × 8 SPADs.

Timing Jitter. Ideally, the delay between a photon absorption and the generation of the corresponding output pulse should be perfectly constant. However, the randomness of the impact ionization process introduces fluctuations in the way the avalanche-current grows after a photon absorption, resulting in slightly different delays each time the detector is triggered. The statistical distribution of the photon detection delays is called temporal response. The broadening of this curve is usually called timing jitter and it is quantified by providing the full width at half maximum (FWHM). It must be remarked that, although very convenient, the FWHM does not provide a complete description of the phenomenon and the whole temporal response must be inspected to identify the presence of slow components with an amplitude that falls below the half maximum. Timing



Manufacturer	Model	Active Area Diameter (µm)	DCR (cps)	Peak PDE		Jitter FWHM (ps)	AP (%)	Dead Time (ns)	At a glance
				Value	Wavelength				
Excelitas Technologies	SPCM-AQRH	180	25 - 1500 (1)	50 - 65% (2)	650 nm	350 (3)	0.5 - 1 (4)	22 - 42 (1)	High PDE
	SPCM-AQRH-TR		100 - 1500 (1)	75% (3)	650 nm	225 - 250 (4)	1 - 3 (4)		
	SPCM-NIR		100 - 1500 (1)	64 - 70% (2)	780 nm	350 (3)	1 - 3 (4)		
Laser Components	COUNT®	100	10 - 250 (1)	55 - 75% (2), (8)	670 nm	1000 (3)	0.2 - 1 (4)	42 - 48 (9)	High PDE
	COUNT® BLUE		10 - 250 (1)	60 - 70% (2), (8)	532 nm	1000 (3)	0.2 (3)		
	COUNT® NIR		50 - 500 (1)	55 - 70% (2), (8)	670 nm	1000 (3)	0.2 - 1 (4)		
	COUNT T		100 - 250 (1)	55 - 75% (2), (8)	670 nm	350 (3)	1 (3)		
MPD Micro Photon Devices	PDM Series	20	5 - 25 (1)	44 - 48% (2)	550 nm	35 - 50 (4)	0.1 - 3 (9)	77	Low Jitter
		50	25 - 250 (1)						
		100	25 - 500 (1)						
ID Quantique	ID100	20	7 - 1000 (1)	30 - 35% (2)	500 nm	40 - 60 (4)	< 0.5	45 - 50 (4)	Low Jitter
		50	60 - 1000 (1)						
	ID120	500	300 - 4000 (1)	80 %	800 nm	200 - 1000 (9)	N/A	1000	High PDE, Large area
Hamamatsu	C11202 series	50	7 - 25 (4)	60 - 70% (2)	450 nm	N/A	0.1 (5)	N/A	High PDE
		100	30 - 100 (4)						
	C16531 series	50	20 - 60 (4)	55 - 65% (2)	630 nm	N/A	0.1 (5)	N/A	
		100	150 - 450 (4)						
AUREA Technology	SPD_A_VIS Red	N/A (6)	25 - 500 (1)	> 65%	700 nm	< 350	N/A	20 - 40 (7)	Built-in gated electronics
	SPD_A_VIS Blue			70%	550 nm	< 350			
	SPD_A_VIS Timing			45%	550 nm	< 50			
Thorlabs	SPDMA (10)	500	300 - 1500 (4)	66% (3)	650 nm	N/A	N/A	< 35	High PDE, Large area
	SPCMxxA	20	25 - 60 (4)	35% (3)	500 nm	N/A	3	35 (3)	Integrated counter
		50	150 - 200 (4)					45 (3)	
	SPDMH series	100	100 - 250 (1)	70% (8)	670 nm	1000	0.2 (3)	45 (3)	High PDE

(1) Depending on the selected model
(2) Min. - Typ. values
(3) Typ. value
(4) Typ. - Max. values
(5) AP events integrated only between 100 and 500 ns
(6) Fiber receptacle for SMF or MMF
(7) Min. programmable value
(8) At the center of the active area. PDE reduces considerably moving away from the center
(9) Min. - Max. values
(10) Adjustable bias voltage. All parameters measured at max bias voltage

Table 1: Single-pixel detection modules based on silicon SPADs.

jitter and temporal response are important parameters when the photon arrival time must be measured with high accuracy. Examples are fluorescence lifetime imaging (FLIM) and Förster resonant energy transfer (FRET) measurements, where subtle variations in the fluorescence lifetime are used to assess changes in a molecule conformation or in its environment.

When SPADs are arranged in an array (Figure 4), there are additional metrics that may become relevant.

Fill Factor (FF). The active area of a SPAD is typically surrounded by structures which are essential for the proper operation of the detector, but that are not capable of detecting photons. Also, electronic circuits needed to operate the detector and acquire the data are frequently integrated next to each SPAD. Consequently, only a fraction of the pixel is photosensitive. The ratio of the photosensitive area to the entire pixel area is called fill factor (FF). The FF is particularly important in those applications where the SPAD array is flood illuminated, like flash-LiDAR, because it results in a fraction 1-FF of photons being lost. A low FF can be partially mitigated by using micro-lenses, either integrated on the detector chip or added externally. It must be noted that, in the SPAD-array community, it has become quite common to indicate the efficiency of a single SPAD as photon detection probability (PDP), and to use the term PDE to indicate the product of the PDP times the FF.

Optical crosstalk. A small number of photons may be emitted by a SPAD during an avalanche because of the large number of energetic carriers that flow through the device. Occasionally, one of these photons can be reabsorbed in a nearby detector and trigger an avalanche. Such a correlated event is called optical crosstalk. Optical crosstalk is especially detrimental when infrequent coincidences must be measured by using pixels within the same array. This is for example the case in quantum enhanced microscopy, where entangled photons are used to increase the spatial resolution of an optical microscope.

MATERIALS

Silicon is by far the most popular material to fabricate SPADs. Silicon SPADs can be operated at room temperature with low DCR or can be slightly cooled (*e.g.* -5°C) to decrease further the noise. They can detect photons from about 400 to 1000 nm, with a typical peak-PDE around 50% in the green. To extend the sensitivity at longer wavelengths, other semiconductors must be used. Up-to-now the best results have been attained by absorbing the infrared photons in an InGaAs layer and by multiplying the carriers in InP. To mitigate the higher DCR, InGaAs/InP SPADs are typically operated at much lower temperature (*e.g.* -50°C). However, this worsens the afterpulsing problem that intrinsically plagues these devices. Absorption in germanium, combined with multiplication in silicon, is being explored as an alternative to reduce the afterpulsing problem and improve the integrability. Despite significant progresses have been made in the last few years, improvements are still necessary to make SiGe SPADs a viable alternative.

COMPARISON WITH OTHER DETECTORS

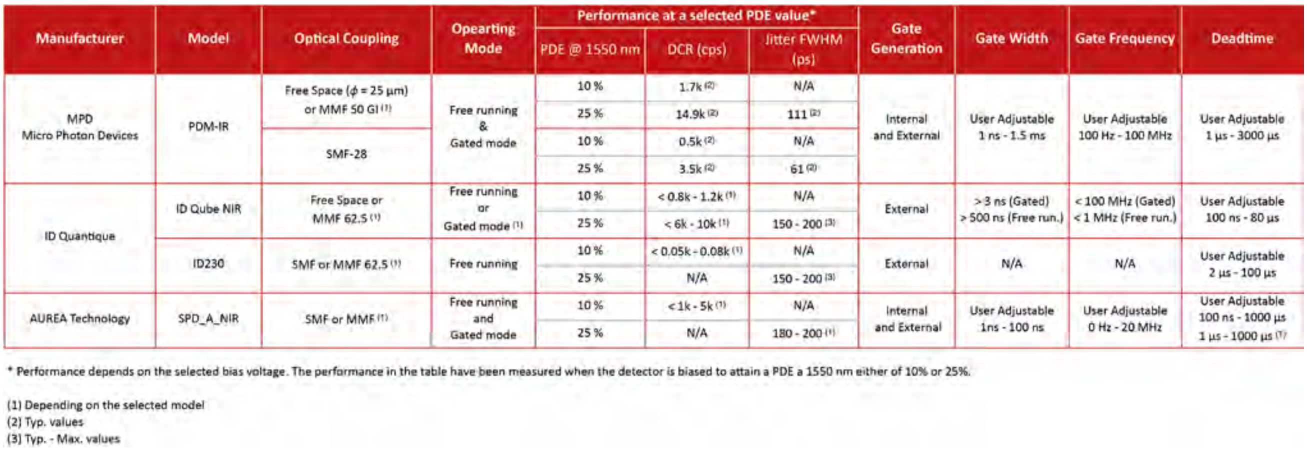
SPADs are not the only single-photon detectors available. The detection of single photons has been possible since the 1930's, thanks to Photomultipliers Tubes (PMTs). In PMTs, an electron is emitted when a photon is absorbed in the photocathode. The photoelectron is accelerated to generate other electrons by hitting a metal target. These are in turn accelerated to generate other electrons and so on, until a large electric pulse is obtained. To work properly, the entire structure must be enclosed in a vacuum tube. Compared to SPADs, PMTs are bulky, cannot form large arrays, and attain a significantly lower detection efficiency, especially at red and near infrared wavelengths. By contrast, PMTs can achieve low DCR even on centimeter-size active areas.

An alternative approach to detect a single photon is to flow a current in a superconductive wire. If the wire is only a few tens-of-nanometer-wide, the absorption of the photon can break the superconductivity, which results in a voltage pulse. ● ● ●

Superconducting nanowire single photon detectors (SNSPDs) attain exceptional performance in terms of PDE, that can extend from deep-UV to mid-IR, DCR, and timing jitter. However, SNSPDs must be operated at temperatures between 1 and 4 K. Whether the improved performance justify the significative increase in cost, complexity and size of the system depends on the specific applications. For example, fundamental science experiments have widely adopted SNSPDs while biological, industrial, and consumer applications favour SPADs.

CONCLUSION

In the last twenty years SPADs have moved from the lab to the market. While the majority of commercial products are single-pixel detectors manufactured in silicon (see Table 1) or, to a smaller extent, in InGaAs/InP (see Table 2), arrays of silicon SPADs have recently started to populate the market (see Table 3). For the coming years, a significant increase in the number and in the performance of the available products is foreseeable. LiDAR for autonomous driving is leading the development of large arrays of SPADs with photon-timing capabilities. Quantum information processing requires detectors with astonishing performance. Safety and security applications are pushing for an extension of the detection spectrum toward the infrared. ●

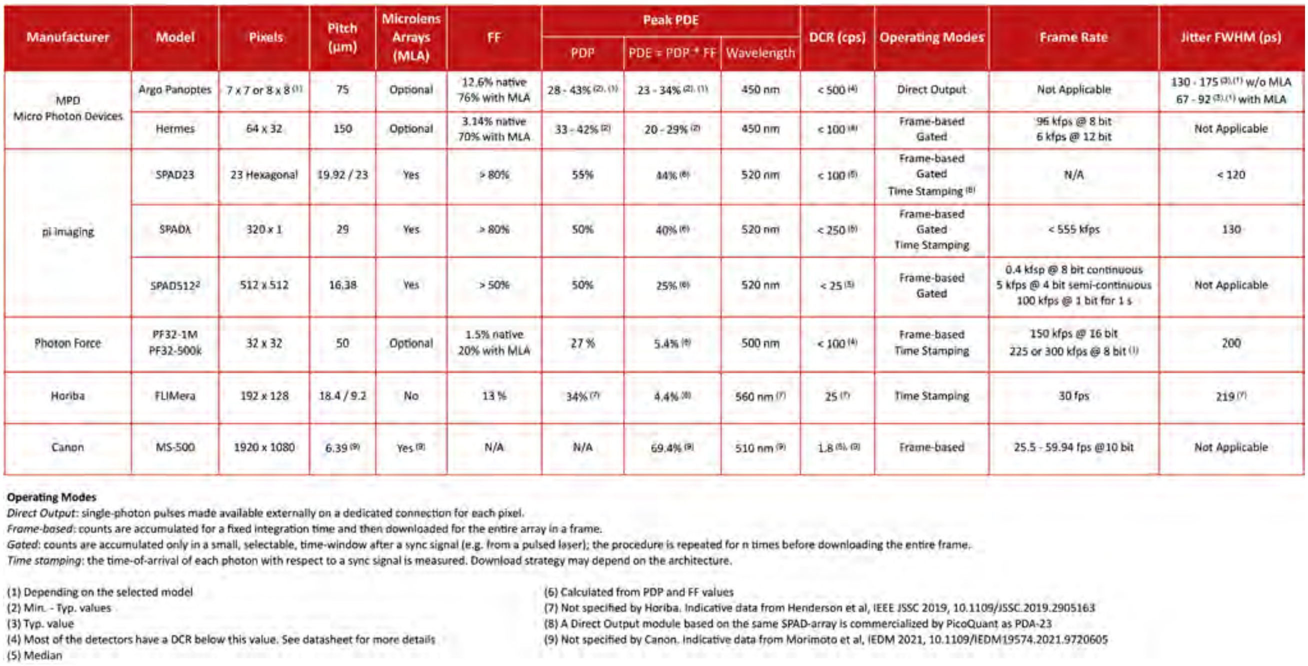


Manufacturer	Model	Optical Coupling	Opearting Mode	Performance at a selected PDE value*			Gate Generation	Gate Width	Gate Frequency	Deadtime
				PDE @ 1550 nm	DCR (cps)	Jitter FWHM (ps)				
MPD Micro Photon Devices	PDM-IR	Free Space (ϕ = 25 µm) or MMF 50 GI (1)	Free running & Gated mode	10 %	1.7k (2)	N/A	Internal and External	User Adjustable 1 ns - 1.5 ms	User Adjustable 100 Hz - 100 MHz	User Adjustable 1 µs - 3000 µs
				25 %	14.9k (2)	111 (2)				
		SMF-28		10 %	0.5k (2)	N/A				
				25 %	3.5k (2)	61 (2)				
ID Quantique	ID Qube NIR	Free Space or MMF 62.5 (1)	Free running or Gated mode (1)	10 %	< 0.8k - 1.2k (1)	N/A	External	> 3 ns (Gated) > 500 ns (Free run.)	< 100 MHz (Gated) < 1 MHz (Free run.)	User Adjustable 100 ns - 80 µs
				25 %	< 6k - 10k (1)	150 - 200 (3)				
	ID230	SMF or MMF 62.5 (1)	Free running	10 %	< 0.05k - 0.08k (1)	N/A	External	N/A	N/A	User Adjustable 2 µs - 100 µs
				25 %	N/A	150 - 200 (3)				
AUREA Technology	SPD_A_NIR	SMF or MMF (1)	Free running and Gated mode	10 %	< 1k - 5k (1)	N/A	Internal and External	User Adjustable 1ns - 100 ns	User Adjustable 0 Hz - 20 MHz	User Adjustable 100 ns - 1000 µs 1 µs - 1000 µs (1)
				25 %	N/A	180 - 200 (1)				

* Performance depends on the selected bias voltage. The performance in the table have been measured when the detector is biased to attain a PDE a 1550 nm either of 10% or 25%.

(1) Depending on the selected model
(2) Typ. values
(3) Typ. - Max. values

Table 2: Single-pixel detection modules based on InGaAs/InP SPADs. These modules are highly configurable, with specific performance that depends on operating conditions. Please see the manufacturer datasheet for further details.



Manufacturer	Model	Pixels	Pitch (µm)	Microlens Arrays (MLA)	FF	Peak PDE			DCR (cps)	Operating Modes	Frame Rate	Jitter FWHM (ps)
						PDP	PDE = PDP * FF	Wavelength				
MPD Micro Photon Devices	Argo Panoptes	7 x 7 or 8 x 8 (1)	75	Optional	12.6% native 76% with MLA	28 - 43% (2), (1)	23 - 34% (2), (1)	450 nm	< 500 (4)	Direct Output	Not Applicable	130 - 175 (3),(1) w/o MLA 67 - 92 (3),(1) with MLA
	Hermes	64 x 32	150	Optional	3.14% native 70% with MLA	33 - 42% (2)	20 - 29% (2)	450 nm	< 100 (4)	Frame-based Gated	96 kfps @ 8 bit 6 kfps @ 12 bit	Not Applicable
pi imaging	SPAD23	23 Hexagonal	19.92 / 23	Yes	> 80%	55%	44% (6)	520 nm	< 100 (5)	Frame-based Gated Time Stamping (8)	N/A	< 120
	SPADλ	320 x 1	29	Yes	> 80%	50%	40% (6)	520 nm	< 250 (5)	Frame-based Gated Time Stamping	< 555 kfps	130
	SPAD512^2	512 x 512	16.38	Yes	> 50%	50%	25% (6)	520 nm	< 25 (5)	Frame-based Gated	0.4 kfsp @ 8 bit continuous 5 kfps @ 4 bit semi-continuous 100 kfps @ 1 bit for 1 s	Not Applicable
Photon Force	PF32-1M PF32-500k	32 x 32	50	Optional	1.5% native 20% with MLA	27 %	5.4% (6)	500 nm	< 100 (4)	Frame-based Time Stamping	150 kfps @ 16 bit 225 or 300 kfps @ 8 bit (1)	200
Horiba	FLIMera	192 x 128	18.4 / 9.2	No	13 %	34% (7)	4.4% (6)	560 nm (7)	25 (7)	Time Stamping	30 fps	219 (7)
Canon	MS-500	1920 x 1080	6.39 (9)	Yes (9)	N/A	N/A	69.4% (9)	510 nm (9)	1.8 (5), (9)	Frame-based	25.5 - 59.94 fps @10 bit	Not Applicable

Operating Modes
Direct Output: single-photon pulses made available externally on a dedicated connection for each pixel.
Frame-based: counts are accumulated for a fixed integration time and then downloaded for the entire array in a frame.
Gated: counts are accumulated only in a small, selectable, time-window after a sync signal (e.g. from a pulsed laser); the procedure is repeated for n times before downloading the entire frame.
Time stamping: the time-of-arrival of each photon with respect to a sync signal is measured. Download strategy may depend on the architecture.

(1) Depending on the selected model
(2) Min. - Typ. values
(3) Typ. value
(4) Most of the detectors have a DCR below this value. See datasheet for more details
(5) Median
(6) Calculated from PDP and FF values
(7) Not specified by Horiba. Indicative data from Henderson et al, IEEE JSSC 2019, 10.1109/JSSC.2019.2905163
(8) A Direct Output module based on the same SPAD-array is commercialized by PicoQuant as PDA-23
(9) Not specified by Canon. Indicative data from Morimoto et al, IEDM 2021, 10.1109/IEDM19574.2021.9720605

Table 3: Detection modules based on silicon SPADs arrays. Please refer to the manufacturer datasheet for more details on performance and operating modes.