The Benefits of Photon Counting

Photon counting is a digital method of measuring light intensity with time. Unless the bandwidth of the change of intensity is very low and the intensity high, Photon Counting is the best method of quantifying changing light intensity. Until recently, photon counting has dominated light measurement for low intensity, but with the increase of bandwidth in new devices, counting of much higher light intensities is possible.

Let us consider a simple comparison of a 1 cm photocathode Photomultiplier Tube (PMT) and a 1 cm Pin Diode. The PMT with a gain of one million and photon counting typically ranges from 4 $x10^{-19}$ watts to 24 pico-watts, whereas the Pin diode, which has no gain, would have a range of perhaps one picowatt to ten milliwatts. The PMT upper limit is determined by the maximum count rate, and that can be increased by a factor of 100 over that commonly used. The PIN diode dynamic range comes with a greatly reduced risetime of 250 nanoseconds compared to one-half nanosecond for the PMT. To get a GHz analog bandwidth from a Pin diode, one has to reduce the active diameter to about 150 micro-meters with a noise of 275 pico-watts. The main reason the PIN diode is not selected is that both wide bandwidth and sensitivity are simultaneously needed.

Pitfalls

This document is to acquaint users with the many pitfalls of photon counting at very high speed which can be avoided. There are many application notes for single photon counting that are quite excellent, but most of them make it seem that photon counting is a very simple thing. After all, is not in the interest of companies to expose the weaknesses of their products. Once you have the product, there is not much you can do to correct the deficiencies. We will begin by examining the limitations of each type of detector, and try to illustrate ways of minimizing the weaknesses of the devices.

APD detectors

Let's begin with avalanche photo-diodes called APD's or sometimes Solid State Photomultipliers. APDs have gain and can detect a single photon in the Geiger mode where the avalanche photo diode completely breaks down at a single photon impact. APD's cannot detect single photons in the non-Geiger mode because the gain is too low and the intrinsic noise levels are too high. Many people purchase APDs only because the stated quantum efficiency of the APD is very high, and the rise time is specified to be very fast. What is not generally realized is that once the APD detects a photon, the avalanche process completely fills the semiconductor device with carriers that must be swept out before another photon can be detected which limits the count rate severely. Clearing the APD of carriers is usually facilitated by applying an electrical pulse to sweep out the carriers, but that causes a long dead time after each photon. If the sweep out pulse doesn't get all the carriers out of the traps, the APD will after-pulse producing false photon counts. Complicated software programs just to statistically correct for the large number of after-pulses are necessary. As a rule of thumb, the higher the quantum efficiency, the greater the after-pulsing and the greater the count rate, the greater the *probability* of after-pulsing. At 3 megahertz count rate, a 3 % after-pulse rate has been reported in the literature as well as a 30 ns dead time. Naturally these excess counts are quite annoying.

An important example of how this type of detector would have problems is in Bathymetry or lidar probing of a water body. Probing the coastal areas near land is of great interest for shipping and commercial applications, but the splash or unavoidable Fresnel reflection caused by change of index and the large overload at penetration of the water causes APDs to become dead for very long times. The dead times for APDs overloaded in this manner may be in the tens of microseconds or hundreds of meters. This severely reduces the application of APDs for bathymetry.

Another example would be probing through trees to the forest floor with a lidar. The branches and leaves would tend to produce large overloads that would make the lidar return dead for significant distances making the ground invisible.

These examples illustrate that a Geiger mode APD is most useful for extremely low light levels with extremely low photon count rates. Semiconductors have both majority and minority carriers, and so there are two lifetimes of the carriers in the device. The lifetime of the minority carrier is much longer than that of the majority carrier in general. This means that the APD rise time is very fast but the fall time has long tails, so the analog mode is not ideal. There is no known way to sweep all carriers out in nanosecond times to prevent after-pulsing. We shall return to APD non-Geiger mode later.

Photomultipliers

Photomultipliers have been around for more than 50 years, and are still in quite common use. They are electron multipliers with a photo-cathode where a photoelectron is emitted by the photoelectric effect that is cascade amplified to a detectable level. The amplification process of the electron multiplier is one of the quietest and fastest of any kind of amplifier in existence. A single photoelectron consisting of 1.602×10^{-19} coulombs may be amplified to hundreds of micro amperes . The conversion from coulombs to micro-amperes occurs because the infinitesimally short single photoelectron event is spread by amplification. In this process, the single photoelectron strikes a dynode which then emits approximately 6 more electrons that move to the next dynode where the process is repeated. Since the number of amplified electrons that each dynode produces varies, the output pulse height is a Poisson distribution. In other words, the pulse height for a single photoelectric event can vary over more than three orders of magnitude. The very smallest single photon events fall below the noise level of the system and are missed.

So far the only slight disadvantage of the photomultiplier tube is that single photon pulses are spread according to the structure of the amplification dynodes. A major advantage of the photomultiplier tube is that the photo cathode may be very large compared to the size of the APD of comparable bandwidth. One disadvantage of the photomultiplier tube is that the photo cathode quantum efficiency for the omission of a single photoelectron is usually less than 50% whereas in APD may have an efficiency 20 or 30% higher.

The real disadvantage of the photomultiplier tube is that the electron multipliers cannot pass high current without heating the dynodes and causing excess noise or possible damage to the tube. Direct sunlight is instantly fatal. The practical maximum average dynode current limit is typically 10 µA. Assuming a typical gain of 1 million for the dynodes, the number of coulombs delivered to the anode is 1.602^{-13} per photoelectron which yields an average of 62 million photons per second in the safe operating current range of the photomultiplier tube. Even worse, the photomultiplier tube becomes very nonlinear and creates excess noise at currents far below this count rate. A typical sustained photon count rate for photomultiplier is in the order of 10 million photo electrons per second.

It may seem obvious that there *is* a way to count photons at a very high rate. If the photomultiplier tube produces extremely narrow pulses and has very low gain, it will have both wide bandwidth and high potential count rate without damaging the dynodes. Such tubes were produced for a short period, but customers found it too difficult to use the very low output, and manufacturing was stopped. One such tube with five dynodes was able to detect 50 photons per second in the presence of a background **average** of 300 million photons per second without damage.

A new tube with low gain and single photon full width half height of 430 picoseconds and and average allowable current of 100 microamperes will be available in about 60 days. This tube should be able to count at GHz photon rates. (3-1-16)

The most common tubes still produced with low gain and sub-nanosecond pulse widths are micro-channel plate tubes which have a drastically smaller allowable average current that negate the possibility of high photon count rates. Micro-channel tubes are used for very good timing capability for low light levels due to their faster response times.

Hybrid detectors

There is another type of single photon detector that is commercially available which has few of the disadvantages of either the APD or the photomultiplier. A hybrid photomultiplier in this case is not an APD plus a semiconductor amplifier, rather it is a photomultiplier merged with an APD single electron amplifier. A large photo cathode emits a single photoelectron into a multiple kilovolt voltage difference giving the single photoelectron a great deal of energy. This energy is deposited in the APD operating in a non-Geiger mode which produces a gain of 1500 from the electron bombardment. The avalanche gain of 100 intrinsic to the APD bias raises the total electron gain to 150,000 with typical output voltage into 50 ohms of 1 mV.

There are numerous benefits of the hybrid detector. The first is that the photo cathode is very large, and this makes it much easier to actually get the photons onto the detector. A typical gigahertz **size** APD might only be 100 µm in diameter which makes it very difficult to collect photons efficiently. The second and most amazing advantage is that the single photo electron voltage output only varies about 10% from photoelectron the photoelectron. This means that the uncertainty of the Poisson distribution in the photoelectron amplification with a photomultiplier tube is completely eliminated. The third advantage is that the overload recovery time for the hybrid tube is a short as one nanosecond to return to single photon counting sensitivity. The overload recovery time for a photomultiplier tube is much faster than that for an APD, but it is still at least tens of nanoseconds.

The main advantage of the hybrid tube is that the sub-nanosecond single photon pulse width and lower gain means that count rates approaching a gigahertz may easily be achieved. Since the pulse height from a single photon is constant, overlapping pulses or pulses that fall on top of each other may be resolved in the electronics resulting in demonstrated count rates of five GHz.

There are some minor disadvantages of the hybrid photomultiplier. The most obvious is the exceptionally small 1 mv signal into 50 ohms necessary to maintain the wide bandwidth. This problem is straightforwardly solved by selecting very low noise amplifiers with high gain. A less obvious inconvenience is that the breakdown voltage of the APD decreases with decreasing temperature. Practically, this necessitates monitoring the temperature of the APD to ensure that the reverse bias voltages never set too high for the current temperature. The avalanche gain also increases with decreasing temperature, so some allowance must be made for the fact that the gain will tend to decrease in a warm electronics environment.

Pitfall table Potential Problems

Pin Diode	Small Pin diodes with fast response have sub-millimeter areas. Impossible to do Photon counting because of intrinsic Noise Sun safe, but background offset can obscure fast pulses
APD - Linear	Photon counting is impossible due to noise. Sensitive, but large RF voltage spikes occur in the presence of background light. Temperature-gain sensitive with long majority and minority tails from the sweep-out of carriers after input light pulses.
APD Geiger PMT	Photon counting with large dead times and many false after-pulses
Micro-Channel Tube	Fast, but *Extremely* low current damage for very low count rates Photocathode has lower Quantum Efficiency than semiconductors No filters exist to allow microchannel use in sunlight background
Electron PMT	Nanosecond photon pulse widths; 10 MHz Maxumum count rate due to limited anode current of 10 ua; Poisson distribution of pulse heights mixes noise with signal.
Low Gain PMT	GHz count rates but Poisson distribution of photon pulse heights; High Quantum Efficiency versions not yet available
Hybrid PM	Sub-nanosecond pulse width but temperature sensitivity to linear APD. System has fewest limitations and GHz capability

General difficulties in Photon Counting

Any time electronic circuits move beyond the tens of megahertz, circuitry becomes more complicated. Impedance matching, stray capacitance, and ringing become severe problems. The anode of the photomultiplier for example, is simply a wire supported by glass hanging in free space which does not intrinsically have a 50 ohms impedance, and connecting the anode to a co-actual cable will cause an impedance mismatch and ringing of the output. The faster the signal, the worse the ringing. If the photomultiplier tube is terminated into an impedance which is too high, a reflected pulse of the same polarity will be added to the incident pulse usually resulting in a much wider pulse or one that is actually produced by many multiple pulses. If the impedance of the termination is less than the photomultiplier natural impedance, an inverted reflected pulse will be added to the photomultiplier output which may have the appearance of a damped oscillation. Due to the difficulties of eliminating ringing and other types of impedance mismatches, photon counting amplifier discriminations often take a shortcut of simply putting out a 10 ns wide pulse for each photoelectron so that ringing will not cause false multiple counts. The problem with this is that real photo electrons that are only separated by a nanoseconds are completely missed.

Ringing

The problem of impedance mismatch ringing is seldom mentioned. Since the ringing damps out due to losses, the most common solution is just to extend the output pulse width of the photon discrimination so that multiple pulses are not produced. Of course this decreases the available count rate. Ringing also means that the lower level for threshold detection of photons cannot be lower than the highest ringing pulse from the highest photon event. To set the discrimination level too low would be to double count the higher photon events. All detector manufacturers produce curves to aid selection of the optimum threshold relative to the dark current pulses to minimize false counts and maximize the collection of actual photon counts. Only a few address techniques to minimize ringing.

A comment seems in order. The idea of using a 10ns pulse to avoid ringing may not seem to be so bad, but signals tend to come in bursts. So one might get three photons in 10 ns and then very few for a long time. Counting the three photons as one could lead to a mis-calibration of the peak intensity which could be very significant. It is **always** better to match the output logic pulse width to the photon pulse width. Avoid the incorrect logic that says, 'if the maximum count rate is 10 MHz before the tube overloads the anode, the 10 ns pulse width has no real effect.' PMTs can handle much larger currents up to GHz count rates for brief times, so a tube that has a sub-nanosecond photon pulse width could *briefly* count photons at a GHz. Using a 10 ns pulse output cripples the tube by a factor of 100. It is important to differentiate between *average* count rates and *peak* count rates. The purpose of this paper is to help facilitate very large *average and peak* count rates by squarely addressing the problems.

DC level shift

Another problem of all these detectors is that monopolar pulses are produced by the detectors. DC coupled amplifiers have much more noise than AC amplifiers and usually much lower bandwidth because of the way they must be constructed. As the photon count rate increases in an AC coupled amplifier, the average DC level due to mono-polar pulses shifts to balance the charges across the capacitor which changes the effective discrimination level. In order to combat this problem and have a constant discrimination level, many systems will use constant fraction timing techniques imported from high-energy nuclear physics. These techniques usually do not work well at sub-nanosecond pulse widths due to amplifier and transformer limitations, so single delay line clipping is used to restore the baseline. Essentially a short circuited coaxial cable is attached to the PMT output with a T. This also has the side effect of increasing the pulse width of the single photoelectron but maintains the threshold completely stable over all pulse rates.

Analog Noise

The problem of noise has not even been addressed. Each of the detectors has an excess noise factor which essentially increases the noise in the system with photon count rate. The practical effect of this is that background light intensities which do not contain any signal information significantly raise the effective noise level of the detector so that signal recovery is difficult or impossible. This may mean that photon counting at some level below the background rate may never recover a signal no matter how long it is averaged. In practice, the number of orders of magnitude below the average background that a signal may be recovered is very limited.

The noise in a photomultiplier tube corresponds primarily to the number of false photon pulses produced in the output in the photon counting mode. If a photomultiplier is used in the analog mode, the dynode current will heat the dynodes producing excess electrons producing excess current all the way down the dynode chain. This produces a false DC signal. The excess current also produces non-linearity that limits in dynamic range by stressing the dynode bias resistors. There is a general rule that the noise in a circuit goes up as the square root of the bandwidth. In the analog mode, the noise goes up quickly with bandwidth and has larger effects at small frequencies. In the digital or photon counting mode, increasing the bandwidth typically does not increase the noise significantly as the false counts occur from spontaneous thermal emission from the photocathode. For example, suppose the tube produces a 10 nanosecond single photon pulse width. If a different dynode structure is used that produces a 0.75 nanosecond single photon pulse width, the bandwidth has been increased by ten-fold but the noise is essentially the same false spontaneous emission rate from the photo-cathode.

Limiting Amplifier

A limiting amplifier is a high gain amplifier with constant output. Unlike fiber optical amplifiers, Marina photonics designed the constant-pulse limiting amplifier for random, monopolar pulses of either positive of negative amplitude. Whether the amplifier is set up to receive negative pulses or positive pulses is chosen at the factory. A separate multiple resistor and capacitor biasing chain is used separately for each polarity for high frequency stability and noise elimination.

The input to the CPA is an extremely low noise (0.5-0.6db), very wide dynamic range amplifier (45dBm typical) with 4 GHz bandwidth. This is followed by a 1.5 GHz differential input/output, variable gain limiting amplifier to match the input signal level optimally. The variable gain amplifier has more than 30 db of gain and more than 30db of gain adjustment. The final output chip is a very high gain differential Logic chip with bipolar output capable of driving 50 ohm cables into DC terminations. The rise and fall time maximums are less than 200 ps, and the chip jitter is less than one picosecond. [1] A Typical 65 db maximum gain of the entire system allows better than 300 uv sensitivity for full output of 600 mv. Both outputs must be terminated in 50 ohms either AC or DC coupled.

[1] All specifications taken from the manufacturers' data sheets

Pin Diode Trigger

It is often the case that a trigger signal is needed with low jitter from a pulsed optical path such as a laser. Even if a laser is Q switched, the Q switch pulse may have high jitter relative to the actual laser output pulse. Extremely high gain and sensitivity of the PE16VA variable gain amplifier allows one to pick off a small portion of almost any optical signal with an appropriate pin diode to get subnanosecond triggering with low jitter. Since the output is a constant amplitude, most circuits will directly accept the output of the PE16VA.

A pin diode coupled with the PE16VA variable gain amplifier will enable one to detect a brief laser pulse or other brief light source such as a xenon strobe and produce a standard output suitable for triggering other devices. The maximum repetition rate of the light source is inversely proportional to the size of the diode. That is, the larger the diode, the lower the maximum repetition rate. Larger pin diodes are more convenient because they collect more light and produce a larger signal due to the larger amount of light collected, but the long fall-time caused by the capacitance of the pin diode discharging into 50 ohms prevents rapid re-triggering. Surprisingly, the rise time of many large pin diodes is not strongly dependent on its size, so sub-nanosecond rise-times are possible even for very large areas.

How it works

The PE16VA amplifier board contains a built-in 5 V bias network for the pin diode so that the diode can be used in a fully depleted mode to reduce the capacitance of the diode and increase the bandwidth. Several models of the PE16VA amplifier with pin diodes will be offered.. The spectral response of the various diodes is usable from 350 nm to 1100 nm with peak response at 980 nm and 750 nm. The sizes available are 0.5 mm, 1.1 mm, 2.4 mm, and 3.6 mm.

One unexpected feature is that a light pulse of sub-nanosecond width produces logic outputs up to 40 MHz for the 3.6 mm Si diode. This is because the rise time into the PIN diode is fast into 50 ohms, and the long tail discharging the capacitance and sweeping out the carriers in the diode has no effect on the positive edge. Rather, the pulse width is somewhat stretched, but the initial edge change on the output is very sharp and stable.