What is an SiPM and how does it work?

Slawomir Piatek, PhD, Hamamatsu Corporation & New Jersey Institute of Technology
October 7, 2016

Related content: **A technical guide to silicon photomultipliers (MPPC)**

What is a silicon photomultiplier?

A silicon photomultiplier (SiPM) is a solid-state photodetector that in response to absorption of a photon can produce a current pulse several tens nanoseconds long containing 10^5 to 10^6 electrons. Therefore, a SiPM has a gain. This gain is comparable to that of a photomultiplier tube (PMT).

Structure

A SiPM is a pixelated device where each pixel, or a microcell, is a series combination of an avalanche photodiode (APD) and a quenching resistor (R_Q). All of the microcells are connected in parallel; thus, a SiPM has two prongs: an anode and a cathode. The three panels in the figure below show, from left to right, a cross section of three microcells in a possible architecture, top view — the side of impinging light — of the device, and an equivalent electrical circuit.

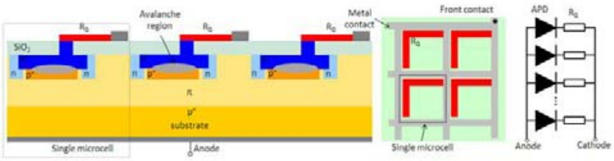

Figure 1. This figure depicts a typical structure of a SiPM. It does not correspond to the actual structure of the Hamamatsu product.

In practical SiPMs, the microcells of identical size are arranged in a rectangular pattern. Depending on the device, the size of a microcell varies from 10 μm to 100 μm and the number of microcells per device ranges from several hundreds to several tens of thousands. SiPMs have active areas ranging from 1 mm^2 to 6 mm^2 and spectral sensitivities from UV to IR, peaking in the visible (400 nm - 500 nm).

Operation

A SiPM is externally biased so that the voltage on each APD is above its breakdown voltage. Thus each APD operates in Geiger mode. The difference between the biasing voltage and the breakdown voltage is known as overvoltage — the main adjustable parameter controlling operation of the device. If a SiPM absorbs a photon, the resulting charge carrier (an electron or hole depending on the structure) can trigger an avalanche in the gain region (shown as the gray oval in the figure) within the p^+ - n^+ structure. The avalanche can produce 10^5 - 10^6 carriers; this constitutes the gain. The role of the quenching resistor is to restore the APD back to the Geiger mode. The gain of a SiPM depends linearly on the overvoltage and the value of junction capacitance. Even though a SiPM is a pixelated device, it is not an image sensor. It does not store charge like a CCD does. It is an analog device producing a time-varying output signal that is measured in real time.

Characteristics

There are several key parameters characterizing a SiPM. These are breakdown voltage, gain versus overvoltage relation, photon detection efficiency, crosstalk probability, afterpulsing probability, and dark count rate.

Applications

A SiPM most resembles a PMT in its operation and characteristics. Compared to a PMT, a SiPM requires a lower operating voltage; has immunity to magnetic fields; is sturdier and more rugged; and is smaller in size. For these and other reasons, the SiPM has begun to replace PMTs in low-light applications such as positron emission tomography (PET), light detection and ranging (LIDAR), or radiation detection in high-energy physics.

How does a SiPM work?

A silicon photomultiplier (SiPM), though pixelated, is a photodetector that produces an analog output signal in real time. The output is a time sequence of waveforms (or current pulses), which have a discrete distribution of amplitudes: A, 2A, 3A, etc. The histogram of the amplitudes depends on the intensity and time-characteristics of the incident light. This note is a closer look into how a single microcell of a SiPM generates a waveform in response to light.

Operation

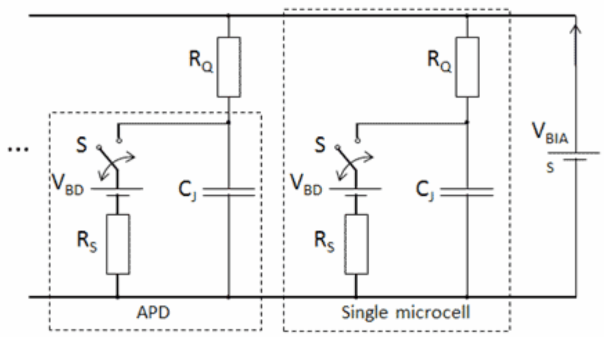

Figure 2. A simplified equivalent circuit of a SiPM (two representative microcells) biased with an external voltage source V_{BIAS}.

All of the microcells in Figure 2 are in parallel. A microcell is a series combination of an avalanche photodiode (APD) and a quenching resistor R_Q. The circuit modeling the APD is a parallel combination of a capacitor C_J (representing the junction capacitance) with a series combination of a switch S, voltage source V_{BD} (equal to the breakdown voltage), and a resistor R_S (representing the resistance of the entire APD during a discharge).

In the absence of light and ignoring dark counts, S is open and the voltage on C_J is V_{BIAS} ($V_{BIAS} > V_{BD}$) — the microcell is in a light-sensitive state and the APD is in Geiger mode. When the microcell absorbs a photon, an electron-hole pair forms; one of the charge carriers drifts to the avalanche region, where it can initiate an avalanche. At the instant the avalanche begins, S closes causing C_J to be discharging through R_S ($R_S \ll R_Q$) with the time constant $R_S C_J$. The voltage on C_J decreases, lowering the probability of impact ionization. For optimal R_Q, the probability eventually becomes so small that the avalanche quenches. At this moment, S opens and V_{BIAS} recharges C_J with the time constant $R_Q C_J$.

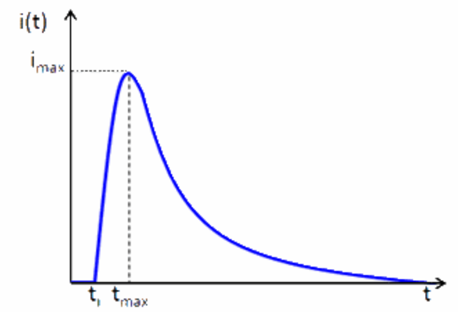

Figure 3. A graph of the current flowing through the terminals of the SiPM during the process discussed above.

In Figure 3, the avalanche begins at t_i and quenches at t_{max}. The leading edge of the pulse starts at t_i and the rise is proportional to $\left[1 - exp\left(-\frac{t}{R_s C_J}\right)\right]$. The pulse reaches the maximum value $i_{max} \approx \frac{V_{BIAS}-V_{BD}}{R_Q+R_S} = \frac{\triangle V}{R_Q+R_S}$ at t_{max} and then declines as $exp\left(-\frac{t}{R_Q C_J}\right)$. Integrating the current pulse with respect to time gives the total charge Q that has transferred between the terminals of the SiPM. Since a single charge carrier caused the transfer, the gain $M = \frac{Q}{e}$. A detailed analysis shows that $Q = C_J \triangle V$ and, thus, $M = \frac{C_J \triangle V}{e}$. If two or more microcells "fire" nearly simultaneously in response to incident light, the output signal is a linear superposition of the pulses, each like the one depicted in the figure, giving an amplitude 2A, 3A, and so on. In contrast, if the same pixel absorbs simultaneously two or more photons, the output pulse is identical to that shown in the figure and the amplitude is A.

Gain of a silicon photomultiplier

An avalanche photodiode in a SiPM operates in Geiger mode. If a charge carrier triggers a discharge in a microcell, the consequence is a current pulse with the total amount of charge Q flowing through the terminals of the SiPM. The gain μ is defined as a ratio of Q and the fundamental charge $e = 1.6 \times 10^{-19}$ C. Typical values of μ are in the range of 10^5 - 10^7. This note is a closer look at the gain.

Anatomy of a 1 p.e. waveform

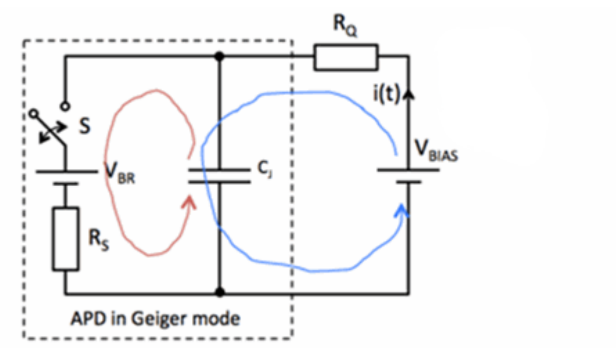

Figure 4. A simplified electrical equivalent circuit of a microcell externally biased with V_{BIAS}.

Figure 4 shows a simplified electrical equivalent circuit of a microcell that is externally biased with V_{BIAS}. In the absence of light and dark counts, the microcell is in Geiger mode: the switch S is open, the voltage on the junction capacitance C_J is V_{BIAS}, and no current flows through the quenching resistor R_Q or the series resistance R_S. At the instant a charge carrier triggers a discharge (avalanche), the switch S closes and C_J begins to discharge through R_S (shown in the figure with the red swirl) causing a voltage drop on R_Q and, thus, a current through the terminals of the SiPM.

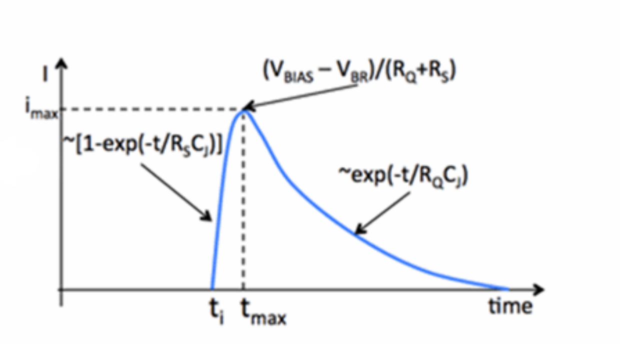

Figure 5. Graph of current pulse.

Figure 5 shows the current pulse begins at t_i. The current rapidly increases until it reaches the maximum value $i_{max} = (V_{BIAS} - V_{BR})/(R_Q + R_S)$ at t_{max}, which is on the order of ~1 ns. At t_{max}, the voltage on the APD drops to approximately V_{BR}, which is not enough to sustain the discharge. Quenching occurs. The junction capacitance C_J begins recharging, causing the voltage across R_Q to be decreasing and, thus, the current through the terminals of the SiPM to be also decreasing exponentially with the characteristic time $\tau = R_Q C_J$. For typical values of R_Q and C_J, $\tau \approx 10$ ns.

The amount of charge in the pulse is approximately $Q = i_{max} \cdot \tau$. The gain is then:

Equation 1

$$\mu = \frac{Q}{e} = \frac{i_{max}\tau}{e} = \frac{1}{e} \cdot \frac{(V_{BIAS} - V_{BR})}{(R_Q + R_S)} \cdot R_Q C_J$$

But $R_Q \gg R_S$, thus

Equation 2

$$\mu \approx \frac{(V_{BIAS} - V_{BR})\ C_J}{e} = \frac{\Delta V \cdot C_J}{e}$$

where $\Delta V \equiv V_{BIAS} - V_{BR}$ is known as "overvoltage." The equation for the gain shows that μ varies linearly with ΔV and C_J and, somewhat surprisingly, it does not depend on R_Q. The value of C_J is fixed by the specific architecture of the SiPM; thus, ΔV is the adjustable parameter that controls μ. There is no explicit dependence of μ on temperature; however, because V_{BR} does depend on temperature, so does μ for a fixed

V_{BIAS}. To eliminate the temperature dependence of the gain, the SiPM should be operated so that ΔV is constant.

In the linear relationship between the gain and overvoltage shown by Figure 6 for three SiPMs, note that for a given ΔV, the gain increases with the size of a microcell and, thus, with C_J.

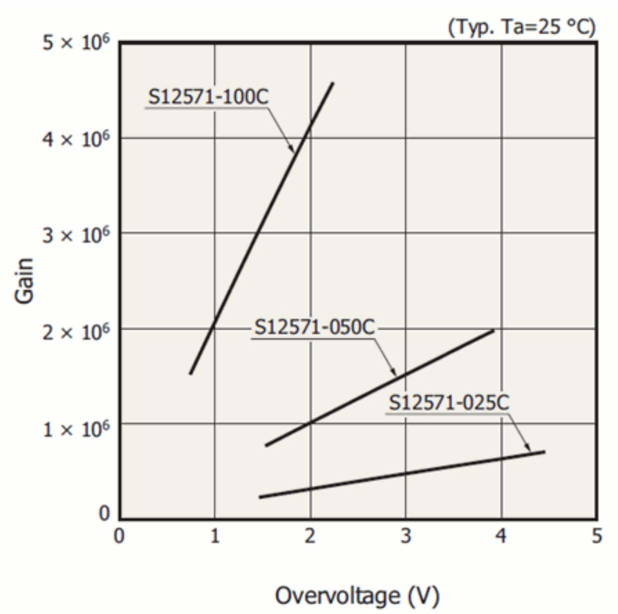

Figure 6. The linear relationship between the gain and overvoltage for three SiPMs manufactured by Hamamatsu: S12571-100C, S12571-050C, and S12571-025C.

Learn more about how temperature affects SiPM performance.

What is the photon detection efficiency (PDE) of a SiPM?

Photon detection efficiency (ξ) is a probability that a silicon photomultiplier (SiPM) produces an output signal in response to an incident photon. It is a function of overvoltage ΔV and wavelength λ of the incident light, and can be expressed as a product $\xi(\Delta V,\lambda) = f \cdot \eta \cdot P_G$. In the equation, f is a geometrical fill factor, η quantum efficiency, and P_G probability of Geiger discharge. Photon detection efficiency is a key characteristic of a SiPM; below is its more detailed discussion.

Geometrical fill factor

A SiPM is a pixelated device. In a given SiPM, all of the pixels, or microcells, are square, of the same size, and arranged in a rectangular (tiled) pattern. The total area occupied by the microcells is known as an active area. Depending on the design of a SiPM, the active areas range from 1 x 1 mm^2 to 6 x 6 mm^2. Because certain regions of a microcell have no photosensitivity, only some fraction of the active area is sensitive to light. This fraction is known as the geometrical fill factor. The values of f range from ~30% to ~80%, with the larger value for the larger size of a microcell.

Quantum efficiency

This parameter is the probability that an incident photon produces a charge carrier (electron or hole) capable of triggering Geiger discharge in the avalanche section of the depletion region. The layered structure of a SiPM can be optimized through the Beers-Lambert law so that the absorption of light occurs in the depletion layer for the desired range of wavelengths. An electric field in the depletion region separates the photo-generated electron-hole pair and injects one of the charge carriers into the avalanche region. Outside of the depletion region, the electric field is weak and the photo-generated electron-hole pair is likely to recombine. The above discussion implies that η is a function of wavelength.

Probability of Geiger discharge

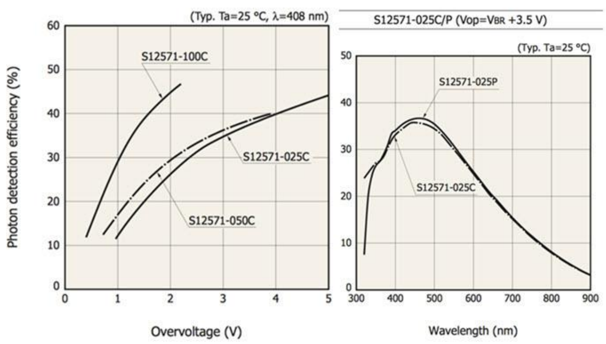

Figure 7. Samples of plots showing how ξ depends on ΔV and λ.

Once in the avalanche region, the injected charge carrier gains kinetic energy from the electric field and if the kinetic energy exceeds the ionization threshold, may be able to impact-ionize silicon atoms, thus triggering an avalanche. The likelihood of this scenario depends on three factors: 1) the strength of the electric field, which is controlled by ΔV; 2) the width of the avalanche region, which also depends on ΔV and 3) collisional cross section of charge carriers with phonons and the specific history of these interactions. If ΔV < 0, the likelihood of Geiger discharge is exceedingly small. As ΔV increases above 0, the kinetic energy of a charge carrier can be above the ionization threshold energy and, moreover, the width of the avalanche region widens, making the likelihood increasingly higher. Inelastic scattering of charge carriers with phonons increases with ΔV, which limits steeply the likelihood of Geiger discharge rising with ΔV. The figures are samples of plots showing how ξ depends on ΔV and λ.

How to measure breakdown voltage in a SiPM

The construction of a silicon photomultiplier (SiPM) consists of microcells (pixels) connected in parallel. Each microcell is a series combination of an avalanche photodiode (APD) and a quenching resistor. In a normal application, an external voltage source V_{BIAS} biases the SiPM so that each APD operates in Geiger mode: the voltage on the APD, equal to V_{BIAS}, is larger than its breakdown voltage, V_{BD}. The difference $V_{BIAS} - V_{BD}$ is known as **overvoltage ΔV**, which is one of the most important parameters affecting the operation of a SiPM. Thus, knowing the value of V_{BD} is crucial because it determines the required value of V_{BIAS} for the desired ΔV. V_{BD} is the key characteristic of a SiPM.

Measuring V_{BD}

Several techniques of measuring V_{BD} exist. This note describes one that relies on the reverse-bias I-V characteristic. Figure 8 shows the experimental setup. The SiPM is in darkness, inside of a temperature-controlled "black box." It is reverse-biased with a voltage source V_S and the picoammeter A measures the current.

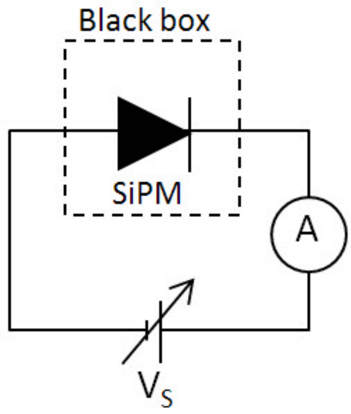

Figure 8. Diagram of experimental setup for measuring V_{BD} inside a temperature-controlled "black box."

Figure 9 shows a possible I-V characteristic around V_{BD}.[1] For $V_S << V_{BD}$ (not shown), the (dark) current I monotonically increases with V_S. The dark current has two contributions: 1) the bulk current due to thermally generated charge carriers, primarily in the depletion region, and 2) the surface current due to defects at the Si-SiO_2 interface. Both contributions increase with V_S. As V_S approaches V_{BD}, the carriers generated in the bulk begin to have enough energy to impact ionize Si atoms in the avalanche section of the depletion region — an APD has a gain larger than 1. The current now increases more rapidly with each voltage step, reaching the highest rate of increase when $V_S = V_{BD}$. For $V_S > V_{BD}$, the APDs operate in Geiger mode, with the gain linearly proportional to ΔV. In addition, crosstalk and afterpulsing contribute to the net current. The current increases with the voltage but at the smaller rate than when V_S was approaching V_{BD}. However, as V_S increases further, another region develops on the I-V characteristic where the current sharply increases with V_S. Here, the density of charge carriers is so high that the quenching resistors cannot restore the APDs back to Geiger mode and the SiPM becomes an Ohmic device. Plotting $d(\log I)/dV_S$ versus V_S reveals the voltages at which the maximum rate of the current increase with voltage occurs; these voltages correspond to the peaks on this plot.

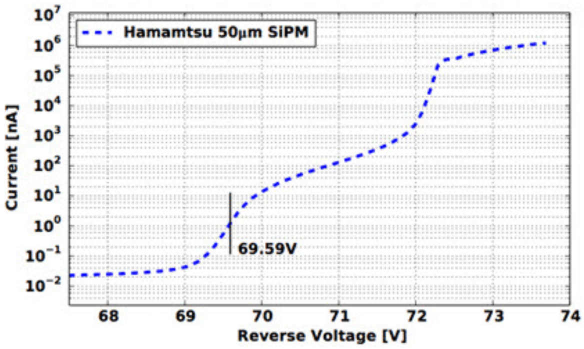

Figure 9. Plot of a possible I-V characteristic around V_{BD}.

Figure 10 shows such a plot for the I-V characteristics listed in this scenario. The first peak (at the lower voltage) exists at V_{BD}. This is how V_{BD} can be measured. The precision of the measurement depends on the voltage step used while performing the sweep. A smaller step yields more accurate determination of V_{BD}. In practice, one would first determine the approximate value of V_{BD} with coarse voltage steps. A finer sweep around V_{BD} would follow. Repeating the finer sweeps and averaging the results would further improve the accuracy of the measured V_{BD}.

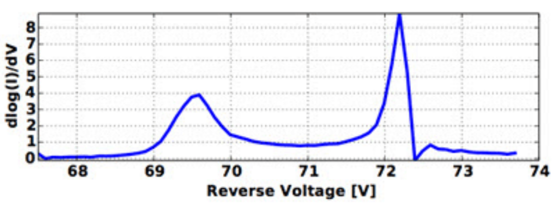

Figure 10. Plot of d(log I)/dV_s versus V_s for the I-V characteristics plot in Figure 9.

S/N in a continuous wave operation of a SiPM

Analogously to a photomultiplier tube, a SiPM can operate in two distinct modes: continuous wave and photon counting. In the former, the output pulses are neither detected nor counted individually; instead, an analog output current is measured. In the latter, the individual pulses are either counted (digital photon counting) or integrated to yield the charge released in Geiger discharge (analog photon counting). The choice of the detection technique depends, among other factors, on the frequency and duration of the output pulses. If the pulses are crowded and overlapping, the continuous wave mode is appropriate. If the pulses are distinguishable, then photon counting — analog or digital — is preferable.

Continuous mode operation

If the individual output current waveforms have a significant overlap, the continuous wave operation of a SiPM is warranted. Suppose a monochromatic light with average power P illuminates the active area of a SiPM. In response, the average signal current is:

Equation 3

$$I_s = \mu \frac{P\lambda e}{hc} \xi$$

where μ and ξ are the gain and photon detection efficiency of the SiPM, respectively; λ is wavelength of the incident light; e is the fundamental charge; h is Planck's constant; and c is the speed of light. The total output current, I_{out}, is a superposition of dark current, I_d, signal current, I_S, and the "excess current" due to crosstalk and afterpulsing. In the first-order approximation, the excess current is $(P_{ct} + P_{ap})I_S$, where P_{ct} and P_{ap} are the probabilities of crosstalk and afterpulsing, respectively. Thus, $I_{out} = I_d + I_S + (P_{ct}+P_{ap})I_S$. This equation assumes that I_d already includes the contributions from crosstalk and afterpulsing.

For the measured values of I_{out} and I_d, the photo-generated current is:

Equation 4

$$I_s = \frac{I_{out} - I_d}{(1 + P_{ct} + P_{ap})}$$

Equating equations 3 and 4 and solving for P yields for the incident light power:

Equation 5

$$P = \frac{hc}{\lambda e} \frac{1}{\xi} \frac{1}{\mu} \left(\frac{I_{out} - I_d}{1 + P_{ct} + P_{ap}} \right)$$

Signal-to-noise ratio

If I_S is measured under the conditions that the measurement bandwidth is B (in Hz) and the excess noise figure is F, then the signal-to-noise ratio is given by

Equation 6

$$S/N = \frac{I_s}{\sqrt{i_d^2 + i_s^2}}$$

where $i_d = \sqrt{(2e\mu BFI_d)}$ and $i_S = \sqrt{(2e\mu BFI_S)}$ are expressions for the shot noise of dark current and signal current, respectively, and I_S is given by Equation 4. Note that the expressions for the dark current and photocurrent shot noise contain only one power of μ because both the dark current and the photocurrent are already amplified by μ.

Excess noise figure

The excess noise figure F must be known for the above equation for S/N to be useful. The theoretical understanding of F is still lacking because numerous factors may contribute to F in a complex way. It is likely that F is a function of overvoltage, input light level, crosstalk probability, afterpulsing probability, microcell-to-microcell gain variations, and stochastic nature of the gain mechanism in a single microcell.

Optical crosstalk in a SiPM

Optical crosstalk occurs when a primary discharge (avalanche) in a microcell triggers secondary discharges in one or more adjacent microcells. The secondary discharge may be nearly simultaneous with the primary (direct or prompt crosstalk) or delayed by several 10's of ns (delayed crosstalk). Optical crosstalk is an example of a correlated noise: it can be present only if a primary discharge is present. The primary discharge can be due to 1) absorption of a photon, 2) thermal generation of a charge carrier in the multiplication region, 3) injection of a charge carrier, thermally generated outside of the avalanche region, into the avalanche region, or 4) crosstalk-induced secondary discharge becoming the primary discharge for subsequent crosstalk events. If not corrected for, crosstalk makes the output signal higher than that implied by the amount of the incident light.

Mechanism

The figure below depicts the mechanism for the prompt (P-CT), delayed (D-CT), and no (No-CT) crosstalk. The primary avalanche in the middle pixel creates three representative photons. One of them moves directly to the avalanche region of the microcell on the right and triggers a simultaneous secondary avalanche there. This is a direct or prompt crosstalk (P-CT). The other photon creates a charge carrier in the vicinity of the avalanche region of the microcell on the left. The charge carrier diffuses to the avalanche region, triggering a secondary avalanche that is delayed with respect to the primary. This is a delayed crosstalk (D-CT). The third photon leaves the SiPM; no crosstalk occurs (No-CT). The majority of photons produced by the primary discharge does not produce crosstalk.

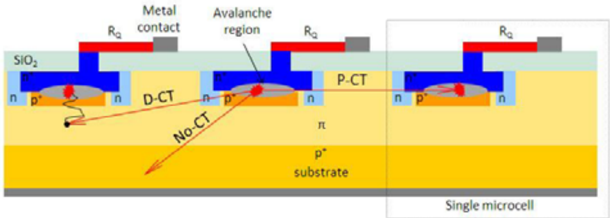

Figure 11. This diagram depicts the mechanism for the prompt (P-CT), delayed (D-CT), and no (No-CT) crosstalk. (It also shows a typical structure of a SiPM, but it does not correspond to the actual structure of the Hamamatsu product.)

Dependence

The probability of optical crosstalk depends on several factors: 1) the size of a microcell, 2) the layered architecture of a SiPM, and 3) the difference between the bias voltage and the breakdown voltage, or the overvoltage. For a given SiPM, the first two factors are fixed; therefore, the overvoltage is the parameter that primarily controls the probability. The probability increases with increasing overvoltage.

Consequences

1. If a SiPM is used to measure the incident photon flux, whether it is by photon counting or continuous wave operation, crosstalk causes an appearance of a higher incident flux.
2. Crosstalk is the main contributor to the excess noise figure.
3. Crosstalk effectively limits the practical setting for the gain.

Measuring crosstalk probability in a SiPM

A Geiger discharge in a microcell of a SiPM emits isotropically about a few tens of photons that have enough energy to create an electron-hole pair. Most of these photons either escape the SiPM or create pairs far from the avalanche region. However, there is a finite probability that a photon creates a pair in, or close to, an avalanche region of a nearby microcell. One of the charge carriers from the pair may trigger a secondary avalanche — this avalanche is known as crosstalk. It is also possible that not only one but two, or more, photons emitted by the primary discharge trigger secondary discharges. The probability of crosstalk for a given SiPM is a function of overvoltage ΔV. This note describes a method of measuring crosstalk probability P_{CT}.

Output waveforms

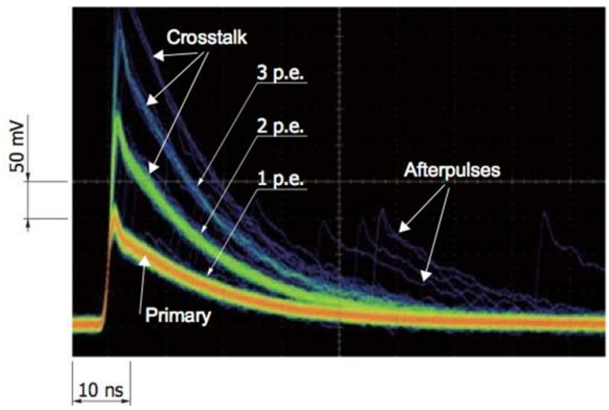

Figure 12. An oscilloscope trace with the display set to persistence mode.

In the absence of crosstalk, the output waveform resembles the one labeled 1 p.e. in the figure above. If the primary avalanche is accompanied by one secondary crosstalk discharge, the resulting waveform, labeled 2 p.e. in the figure, is a superposition of two 1 p.e. waveforms. If it is accompanied by two, a 3 p.e. waveform results, and so on. The figure is an oscilloscope trace with the display set to persistence mode. The SiPM is in darkness; the 1 p.e. waveforms are due to dark counts. The color gradient, from warmer (red) to cooler (blue), is proportional to the "density" of waveforms, that is, the frequency of occurrence. The most common events are 1 p.e., followed by 2 p.e., 3 p.e., etc. All of the waveforms that are different from 1 p.e. are due to crosstalk.

Measuring crosstalk probability

The primary waveforms (1 p.e.) are due to dark counts; whereas, 2 p.e., 3 p.e., etc. are due to crosstalk. Let $N_{1p.e.}$, $N_{2p.e.}$, $N_{3p.e.}$ be the number of 1 p.e., 2 p.e., 3 p.e., etc. waveforms counted within some observation time τ. Therefore, the dark count rate is $(N_{1p.e.} + N_{2p.e.} + N_{3p.e.} + ...)/\tau$, whereas the probability of crosstalk is $P_{CT} = (N_{2p.e.} + N_{3p.e.} + N_{4p.e.} + ...)/(N_{1p.e.} + N_{2p.e.} + N_{3p.e.} + ...)$.

A relatively simple procedure to measure this probability is to count the number of waveforms with the setup shown in Figure 13.

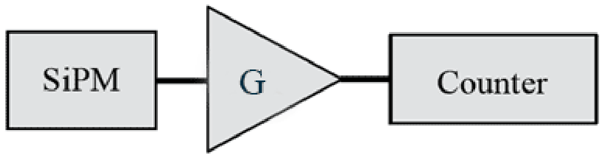

Figure 13. Diagram of a simple setup to measure probability of crosstalk.

The SiPM operates in the dark. The wide-bandwidth amplifier G converts the output current signal to a voltage signal and amplifies the signal's strength. The output of the amplifier is an input to a digital counter such as Tektronix FCA3100. The counter counts a signal if its amplitude exceeds a certain user-set level (lower discrimination threshold). The experimental data consist of the number of counts per unit time as a function of threshold level. Plotting this data produces the characteristic "staircase" shown in Figure 14.

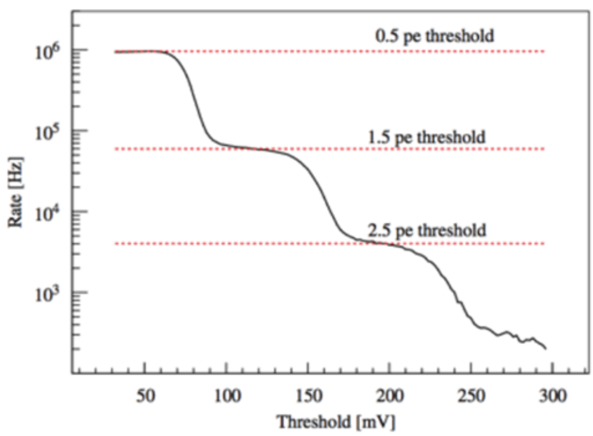

Figure 14. Plot of experimental data producing characteristic "staircase."

The vertical axis plots the count rate (frequency) and the horizontal axis the threshold level. The frequency corresponding to the top red line is due to all detectable waveforms, whereas that for the red line below the top is due to crosstalk waveforms. Referring to the first frequency as $v_{0.5p.e.}$ and the other as $v_{1.5p.e.}$, the probability of crosstalk is $P_{CT} = v_{1.5p.e.}/v_{0.5p.e.}$.

References

1 Adopted from "Study of the Silicon Photomultipliers and Their Applications in Positron Emission Tomography," PhD Dissertation by Chen Xu.

About the author

Slawomir S. Piatek, PhD, is a senior university lecturer of physics at New Jersey Institute of Technology. In his role as scientific consultant at Hamamatsu Corporation, he has developed a photonics training program for engineers and is involved in popularizing SiPM as a novel photodetector by writing and lecturing about the device.