

Optical Engineering

SPIDigitalLibrary.org/oe

Signal-to-noise ratio of Geiger-mode avalanche photodiode single-photon counting detectors

Kimberly Kolb

Signal-to-noise ratio of Geiger-mode avalanche photodiode single-photon counting detectors

Kimberly Kolb*

Rochester Institute of Technology, Center for Detectors, 17-3177, 74 Lomb Memorial Drive, Rochester, New York 14623

Abstract. Geiger-mode avalanche photodiodes (GM-APDs) use the avalanche mechanism of semiconductors to amplify signals in individual pixels. With proper thresholding, a pixel will be either “on” (avalanching) or “off.” This discrete detection scheme eliminates read noise, which makes these devices capable of counting single photons. Using these detectors for imaging applications requires a well-developed and comprehensive expression for the expected signal-to-noise ratio (SNR). This paper derives the expected SNR of a GM-APD detector in gated operation based on gate length, number of samples, signal flux, dark count rate, photon detection efficiency, and afterpulsing probability. To verify the theoretical results, carrier-level Monte Carlo simulation results are compared to the derived equations and found to be in good agreement. © 2014 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.OE.53.8.081904](https://doi.org/10.1117/1.OE.53.8.081904)]

Keywords: Geiger-mode avalanche photodiode; signal-to-noise ratio; afterpulsing; Monte Carlo; photon counting.

Paper 131669SS received Nov. 1, 2013; revised manuscript received Feb. 14, 2014; accepted for publication Feb. 25, 2014; published online Mar. 31, 2014.

1 Introduction

Array-based Geiger-mode avalanche photodiodes (GM-APDs) have been developed by Massachusetts Institute of Technology Lincoln Laboratory and characterized by the Center for Detectors to determine their suitability for space-based imaging applications, specifically for exoplanet missions.¹ A GM-APD can count single photons, which is very useful for low-light imaging applications. In Geiger-mode (digital) operation, the APD simply records whether or not an avalanche occurred in a given exposure window (a gate) and repeats the process over many samples. From these samples, an avalanche probability can be calculated, and from that probability, the estimated flux is derived.² Since this operation is fundamentally different from charge-coupled device (CCD) or CMOS detectors, the form of the signal-to-noise ratio (SNR) expression is fundamentally different as well. To derive the expression, the detection cycle and unique sources of noise must be fully understood. This paper derives the expression for the SNR of measurements from a GM-APD detector measuring intensity via gated operation. The final form of the expression depends on dark count rate (DCR), afterpulsing probability (p_{aft}), photon detection efficiency (PDE), gate length, number of gates (n_{gates}), and signal flux. The inclusion of afterpulsing probability also results in a means to estimate the signal when significant afterpulsing is present. The equation is also useful for choosing the best operating conditions for such a detector (and similar detectors), which might include small amounts of afterpulsing as a trade-off for higher SNR with higher duty cycle.

It is important to note that while the majority of GM-APD applications use arrival time-based measurements or analog avalanche totals to count photons, this detector measures intensity by measuring the avalanche probability during a set exposure window (usually on the order of

microseconds). SNR has been presented for time-based measurements previously for similar devices,³ but it is important to emphasize that the operation is fundamentally different for the device presented here. For this detector, the measurement is actually of the probability of an avalanche given a certain exposure window, and all resolution of time is lost. Timing jitter is irrelevant, and the output from each pixel is digital. Afterpulsing effects on noise have been investigated by others,⁴ though no integration of afterpulsing statistics into SNR for a gated, digital device has been presented. Given the fundamental difference in operation between most GM-APDs and the device presented in this paper, a new expression for SNR must be derived from fundamentally different first principles.

Noise equivalent power (NEP) for this detector is not presented in this paper, given the lack of a closed-form solution for the sensitivity. Most currently accepted NEP expressions for GM-APDs are irrelevant, given the inclusion of the time-measurement paradigm.⁵ Other accepted expressions for NEP are derived with the assumption that the intensity output from the pixel is analogue, not digital (i.e., number of avalanches is given rather than avalanche probability), which results in a completely different relationship between signal and noise.⁶

1.1 Avalanche Mechanism

At reverse bias voltages above some breakdown voltage (V_{BR}), a self-sustaining avalanche can be established. Greater detail on the avalanche behavior of this device has been previously presented by its developers.⁷ An avalanche resulting from one photogenerated electron is indistinguishable from one initiated by a larger number of photogenerated electrons arriving simultaneously.

It is important to note that even in Geiger mode, there is a probability that the avalanche may dwindle in its earliest

*Address all correspondence to: Kimberly Kolb, E-mail: kem2691@rit.edu

stages and result in a nondetectable signal. While gain is an important metric for linear-mode operation, Geiger-mode operation is better characterized by the probability that the avalanche will become self-sustaining, referred to as the avalanche initiation probability. This probability can be calculated independently in theory.⁸ However, the avalanche initiation probability is a component of the overall PDE, the measurement of which is usually sufficient for testing and characterization purposes.

1.2 Detection Cycle

For this device, the detection cycle is clocked externally and reset at regular intervals. Each exposure is composed of five distinct stages, repeated many times over. The first stage is the arming of the device, when the bias on the pixel is increased above the breakdown voltage. A set delay (the second stage) is then observed, which constitutes the exposure gate. After the gate, a recording pulse is asserted (the third stage) that transfers the state of the pixel (1 or 0) to the read-out circuit. Immediately after the recording pulse is complete, the pixel is forcefully disarmed (the fourth stage), meaning that the voltage is set below the breakdown voltage. A final delay (the fifth period) is observed after the disarm signal, called the hold-off time. This delay is usually to mitigate afterpulsing events. At the end of the hold-off time, the pixel is armed again and the cycle repeats. In an ideal device, the forced disarm would be unnecessary because avalanches between gates would not affect the occurrence of avalanches during the gates. In practical use, forced disarm is required because of the afterpulsing mechanism, which can induce an avalanche in a subsequent gate with a characteristic exponential decay probability.

1.3 Afterpulsing

Afterpulsing occurs when a trap (an intermediate energy state that exists in the band gap of the material) releases a carrier that initiates an avalanche in the absence of a photo-generated electron or dark count. Carriers are released from traps at random times, with the average detrapping time defined as the trap lifetime. The length of the trap lifetime depends on the type of trap and its energy level. Avalanching carriers can remain in these states longer than the hold-off time of the device, becoming free again during the next detection cycle. The electric field immediately accelerates the newly mobile electron or hole, which may initiate an avalanche of carriers in the absence of photon signal.⁹

1.4 Monte Carlo Simulations

A Monte Carlo simulation was created to calculate the observed SNR for any detector characteristics or operational settings. The simulation works from a detailed list of inputs, including DCR, p_{aft} , PDE, gate length (t_{gate}), signal flux, and n_{gates} . Carriers are generated from a random number Poisson distribution for each gate, based on the PDE, signal flux, and DCR. For gates where the number of generated carriers is >0 , the simulation records a 1. The total number of avalanches (total number of 1's) is divided by n_{gates} to calculate the avalanche probability given the DCR [see Eq. (6)]. The experiment is repeated 10^7 times, and then the mean and standard deviation of the results are calculated, giving the SNR.

To add afterpulsing, a simple binomial random number generator is used to generate a carrier with probability p_{aft} , given that the simulation recorded a 1 in the previous gate. In this case, the avalanche probability is calculated with Eq. (29) [which simplifies to Eq. (6) if $p_{\text{aft}} = 0$], and SNR is calculated by dividing the mean by the standard deviation of the 10^7 trial results.

2 SNR Derivation Neglecting Afterpulsing

The first step in the derivation of any SNR expression is defining the processes of the detection cycle (in the case of photon-counting detectors, the physical meaning of the string of 1's and 0's). To simplify the relationships, the initial assumptions are that PDE is unity and there are no other sources of noise (these nonidealities will be added later). Instead of counting each 1 as a single photon, the ratio of 1's to the total number of gates is the probability of one or more photons arriving within a single gate. The probability of n_p photons in a specific interval for an average number of photons per gate λ_p is

$$P(n_p) = \frac{e^{-\lambda_p} \lambda_p^{n_p}}{n_p!}. \quad (1)$$

For a Poisson distribution, since each trial is memoryless (independent),

$$\begin{aligned} P(n_p = a \text{ or } n_p = b) &= P(n_p = a) + P(n_p = b) \\ &= e^{-\lambda_p} \left(\frac{\lambda_p^a}{a!} + \frac{\lambda_p^b}{b!} \right), \end{aligned} \quad (2)$$

and the total probability that n_p equals zero or any positive integer is 1, or

$$\sum_{n_p=0}^{\infty} \frac{e^{-\lambda_p} \lambda_p^{n_p}}{n_p!} = 1. \quad (3)$$

The ratio of 1's to the total number of gates can, therefore, be defined as

$$\begin{aligned} \frac{n_{1's}}{n_{\text{gates}}} &= P(n_p \geq 1) = 1 - P(n_p = 0) = 1 - \frac{e^{-\lambda_p} \lambda_p^0}{0!} \\ &= 1 - e^{-\lambda_p}. \end{aligned} \quad (4)$$

Since dark carrier generation is also a Poisson process, the average number of dark carriers per gate, λ_d , is additive in the exponential function. Changing n_p to n to represent the total number of carriers per gate (assuming dark current is the only source of noise and that there is no photon loss) and λ_p to $\lambda = \lambda_p + \lambda_d$ to reflect the average number of total carriers per gate,

$$\begin{aligned} \frac{n_{1's}}{n_{\text{gates}}} &= P(n \geq 1) = 1 - P(n = 0) = 1 - \frac{e^{-\lambda} \lambda^0}{0!} = 1 - e^{-\lambda}; \\ \lambda &= \lambda_p + \lambda_d. \end{aligned} \quad (5)$$

From this simplified solution, the estimate of the average number of photons per gate is

$$\hat{\lambda}_p = -\ln\left(1 - \frac{n_{1's}}{n_{gates}}\right) - \hat{\lambda}_d. \quad (6)$$

In practice, $\hat{\lambda}_d$ may be estimated by applying Eq. (4) to measurements made under dark conditions. The estimate of the total number of photons incident on a pixel over an exposure (the fluence) with PDE less than unity is

$$\text{PDE} \cdot \hat{\lambda}_{tot} = \left[-\ln\left(1 - \frac{n_{1's}}{n_{gates}}\right) - \hat{\lambda}_d\right] \cdot n_{gates}. \quad (7)$$

The variance of the estimate is slightly more complicated. To simplify the equations, let $n_{1's}/n_{gates} = x$ and $\text{PDE} \cdot \hat{\lambda}_{tot} = y$. The variance of x is the variance of a binomial distribution based on the number of gates (or trials), n_{gates} , and the probability of an avalanche during a gate, $p(\text{gate} = 1) = p$. Note that x is only an estimate of p , so they are denoted differently

$$\sigma_x^2 = \frac{n_{gates} \cdot p(1-p)}{n_{gates}^2} = \frac{p(1-p)}{n_{gates}}. \quad (8)$$

The variance of the total number of 1's is the Bernoulli trial variance, $p(1-p)$, multiplied by the number of trials, n_{gates} . However, x is the ratio of total number of 1's to the total number of gates, so the variance must be divided by the square of the scale variable, resulting in Eq. (8).

Now the variance of the estimate (which is a function of x) must be defined. To begin, the expected value of $y(x)$ is defined in terms of the probability density function of x , or $f(x)$, as follows:

$$E[y(x)] = \int_{-\infty}^{\infty} y(x)f(x)dx. \quad (9)$$

If $f(x)$ is concentrated about the mean (a valid assumption here since x is the outcome of a series of Bernoulli trials, a binomial distribution), then $f(x)$ is assumed to be negligible outside the range $(\mu - \epsilon, \mu + \epsilon)$, where μ is the mean of $f(x)$ and ϵ is a bounding variable, and $y(x)$ becomes $y(\mu)$

$$E[y(x)] = y(\mu) \int_{\mu-\epsilon}^{\mu+\epsilon} f(x)dx = y(\mu). \quad (10)$$

This estimate may be improved by a polynomial expansion.¹⁰

$$y(x) = y(\mu) + y'(\mu)(x - \mu) + \dots + y^n(\mu) \frac{(x - \mu)^n}{n!}. \quad (11)$$

Inserting Eq. (11) into Eq. (10) and neglecting higher-order terms for a parabolic approximation,

$$E[y(x)] = y(\mu) + y''(\mu) \frac{(x - \mu)^2}{2} = y(\mu) + y''(\mu) \frac{\sigma_x^2}{2}. \quad (12)$$

To find the variance of y ,

$$\sigma_y^2 + \mu_y^2 = E[y^2(x)], \quad (13)$$

$$\begin{aligned} \mu_y^2 &= E[y(x)]^2 = \left[y(\mu) + y''(\mu) \frac{\sigma_x^2}{2}\right]^2 \\ &= y^2(\mu_x) + y(\mu_x)y''(\mu_x)\sigma_x^2 + [y''(\mu_x)]^2 \frac{\sigma_x^4}{4}, \end{aligned} \quad (14)$$

$$\begin{aligned} E[y^2(x)] &= y^2(\mu_x) + \left[\frac{d^2}{dx^2}y^2(\mu_x)\right] \frac{\sigma_x^2}{2} E[y^2(x)] \\ &= y^2(\mu_x) + \sigma_x^2[|y'(\mu_x)|^2 + y(\mu_x)y''(\mu_x)]. \end{aligned} \quad (15)$$

Since σ_x is always <1 , the σ_x^4 term in Eq. (14) is negligible. Substituting the simplified Eqs. (14) and (15) into Eq. (13) and solving for σ_y^2 , the variance of y is shown in Eq. (16):¹⁰

$$\sigma_y^2 = |y'(\mu_x)|^2 \sigma_x^2. \quad (16)$$

Equation (16) is the variance of the estimate of the number of photons. The derivative of $y(x)$ [Eq. (6)] is

$$y'(\mu_x) = \frac{1}{1 - \mu_x}. \quad (17)$$

Recall that μ_x is the mean of the observed avalanche probability distribution $p(\text{gate} = 1) = p$. Substituting Eqs. (8) and (17), the total variance for the exposure is

$$\sigma_{tot}^2 = |y'(\mu_x)|^2 \sigma_x^2 n_{gates}^2 = \frac{P}{(1-p)} n_{gates}. \quad (18)$$

Since the output of each gate records either a 1 or a 0 in the case of a photon-counting mode detector, each gate may be modeled as a Bernoulli trial with the probability of a 1 equal to $p = \mu_x$ and the probability of a 0 equal to $q = 1 - p$. Assuming that only one electron is necessary to register a 1, p is defined by Eq. (5). Now σ_y^2 may be evaluated as a function of the average number of photogenerated carriers per gate (λ_p) and the average number of dark-current-generated carriers per gate (λ_d), substituting Eq. (5) and incorporating PDE

$$\sigma_{tot}^2 = \left(\frac{E\left[\frac{n_{1's}}{n_{gates}}\right]}{1 - E\left[\frac{n_{1's}}{n_{gates}}\right]} \right) n_{gates} = [e^{(\text{PDE} \cdot \lambda_p + \lambda_d)} - 1] n_{gates}. \quad (19)$$

Therefore, the SNR for an intensity measurement using a gated GM-APD with no afterpulsing is

$$\text{SNR} = \frac{\text{PDE} \cdot \lambda_p \cdot n_{gates}}{\sqrt{\frac{P}{(1-p)} \cdot n_{gates}}}, \quad (20)$$

where $p = 1 - e^{-(\text{PDE} \cdot \lambda_p + \lambda_d)}$.

PDE is the photon detection efficiency (the probability that a photon will be absorbed and the resulting carrier will initiate an avalanche), λ_p is the number of photons absorbed per gate, and λ_d is the number of dark current carriers generated per gate. Together, PDE, λ_p , and λ_d give the probability, p , that a gate will record an avalanche. n_{gates} is the number of gates in the exposure (note that $\text{SNR} \propto \sqrt{n_{gates}}$).

Any dead time losses are built into the signal and the noise because the estimates are always based on multiples of per-gate fluence. For example, if the estimate of the mean number of photons per gate is 2 (assume that PDE is 1), the number of gates is 100, but the duty cycle is 50%, then the estimate of the total number of photons in the exposure is $2 \cdot 100 = 200$ (instead of the total incident fluence of 400).

Figure 1 shows an overlay of the Monte Carlo results and the analytical solution in Eq. (20) for SNR versus the total fluence, normalized to the ideal shot noise limit of SNR. Ideal SNR is the shot-noise limited case, where $\text{SNR} = \sqrt{\text{Fluence}}$ (fluence is the number of signal photons arriving at the pixel during the exposure). The simulation agreed with the theoretical data in both mean and standard deviation.

The roll-off at low fluence is due to background counts (DCR), which start to contribute significantly to the noise when the number of background electrons per gate (10^{-5} in Fig. 1) is $\sim 10\%$ of the signal (photogenerated) electrons. The roll-off at high fluence is due to high avalanche probability, where the proportional variance of the estimate increases sharply due to the exponential relationship between avalanche probability and photon fluence. As the gate photogenerated electron (PE) fluence approaches and surpasses 1, the SNR drops sharply. In this regime, a very small change in measured avalanche probability translates to a very large change in estimated fluence. The curve peaks at $\sqrt{\text{PDE}}$ because the data are normalized to $\sqrt{\text{Fluence}}$ (the shot-noise limited case).

3 SNR Derivation Including Afterpulsing

Since there is no read noise in a digital GM-APD detector, Eq. (19) correctly models the variance of the estimate neglecting afterpulsing. Some researchers choose to blank the gates following a recorded event to reduce the afterpulsing noise.¹¹ This may be useful for instances of high afterpulse probabilities, but the drawback is a significant decrease in the number of gates (the number of samples) and the duty

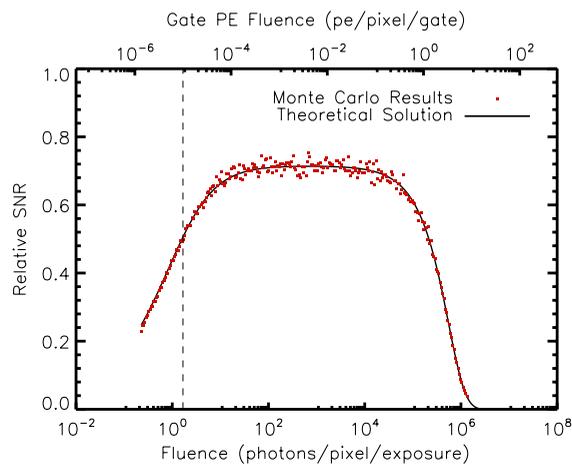


Fig. 1 This plot shows Monte Carlo results versus analytical solution for the relative signal-to-noise ratio (SNR) of a Geiger-mode avalanche photodiode (GM-APD) in photon-counting mode over a range of fluence values. Dark count rate is 1 Hz. The dashed vertical line notes the fluence at which photogenerated signal and noise contributions are equal. Gate length is $10 \mu\text{s}$, exposure time is 1 s, photon detection efficiency is 60%, and duty cycle is $\sim 85\%$. Relative SNR is normalized to the ideal SNR, the shot-noise limited case where $\text{SNR} = \sqrt{\text{Fluence}}$.

cycle of the device. Depending on the afterpulsing probability, blanking may not increase the SNR of the measurement. Therefore, it is beneficial to quantify the contribution of afterpulsing to the noise. To start, the probability of a gate equal to 1 must be amended to include the probability of an afterpulse carrier. The new definition of p (the probability of a gate equal to 1) is

$$\begin{aligned} P(\text{gate} = 1) &= P(\text{afterpulse} \cup 1|\lambda) \\ &= P(\text{afterpulse}) + P(1|\lambda) \\ &\quad - P(\text{afterpulse})P(1|\lambda). \end{aligned} \quad (21)$$

Here, $P(\text{afterpulse})$ is the probability that one or more afterpulse carriers are present (or become present) during the gate, and $p(1|\lambda)$ is the probability that one or more photogenerated or dark carriers are present during the gate (where $\lambda = \text{PDE} \cdot \lambda_p + \lambda_d$). The probability of an afterpulse carrier being present during the gate, while dependent on the previous gate avalanche probability, is statistically independent of the current gate. Therefore, they are not mutually exclusive and the probability of either happening is not a simple sum. This leads to the subtraction of the cross term in Eq. (21). $p(1|\lambda)$ is equal to Eq. (4), but the derivation of $P(\text{afterpulse})$, which will be referred to as p_{aft} going forward, is more involved. The derivation of avalanche probability given a certain afterpulse probability requires a few assumptions. For this derivation to be valid, the following must be true.

1. The pixels are disarmed at the end of the gate and no avalanche events occur between gates (see Sec. 1.2).
2. There is no (or insignificant) dependence on gates previous to the gate immediately preceding to the gate of interest (i.e., the probability of an avalanche in the current gate is only a function of the state of the gate immediately before it and the photo- and dark carrier generation process).
3. There is no significant delayed cross talk from neighboring pixels (i.e., afterpulsing occurs only as a result of the same pixel's previous state, not a neighboring pixel's previous state).

Physically, p_{aft} is the integral of the exponential decay function of afterpulse arrival time from the beginning of the gate to the end of the gate (i.e., from $t_{\text{hold-off}}$ to $t_{\text{hold-off}} + t_{\text{gate}}$ if $t = 0$ is the time of the previous avalanche). With passive quenching and clocking relative to the avalanche, this estimate has zero error. However, with gated clocking of the circuit, p_{aft} theoretically changes from gate to gate because the effective quench time is dependent on when the previous gate's avalanche occurred. In gated operation, the precise avalanche arrival time is unknown, though an average value of p_{aft} may be sufficient since the measurement involves many gates over the course of the exposure.

The first assumption listed above is easily confirmed by determining the method of operation of the device in question, while the third can be confirmed with avalanche correlation tests between neighboring pixels. The second assumption does imply some constraints in the amount of afterpulsing in order for this analysis to be relevant. The assumption hinges on the behavior and amount of traps in

the pixel. If the exponential decay function of detrapped carriers extends past two arm periods, then there is a certain probability that an avalanche will occur due to an avalanche in a gate more than one arm period before. If one assumes that 1% is a negligible probability for an afterpulse-induced avalanche two arm periods later, then it is easy to calculate the maximum decay lifetime allowed for assumption 2 to be valid. For example, in an exposure with $t_{\text{gate}} = 10 \mu\text{s}$ and an 85% duty cycle, the total arm period is $\sim 11.8 \mu\text{s}$. If we assume that dark- or photo-induced avalanches occur near the beginning of the gate on average (Poisson arrival statistics), then the effective time between populating a trap and the beginning of the next gate is roughly equal to the arm period. Therefore, the decay lifetime must be less than or equal to $2[-\ln(0.01)]t_{\text{armperiod}}$, or in this case, the decay lifetime must be $\leq 5.1 \mu\text{s}$. For valid results, the measured p_{aft} value (using the expression derived in this paper) should be no greater than the integral of the exponential decay over the following gate, or $p_{\text{aft}} = 0.09$. However, as t_{gate} decreases and duty cycle increases, the maximum valid p_{aft} also increases, especially if the average avalanche time moves further from the beginning of the gate.

Having established the assumptions made and their implications, the derivation can continue. To begin, the avalanche probability must be redefined to include afterpulsing probability. The probability of an afterpulse in the first gate ($n = 0$) is zero, since there were no previous gates. Therefore, the probability of an avalanche is

$$P_0(\text{gate} = 1) = 0 + (1 - e^{-\lambda}) - 0 = 1 - e^{-\lambda}. \quad (22)$$

For the second gate ($n = 1$), the probability becomes more complicated

$$\begin{aligned} P_1(\text{gate} = 1) &= p_{\text{aft}}(1 - e^{-\lambda}) + (1 - e^{-\lambda}) \\ &\quad - p_{\text{aft}}(1 - e^{-\lambda})(1 - e^{-\lambda})(\text{gate} \\ &= 1) = (1 - e^{-\lambda})(1 + p_{\text{aft}}e^{-\lambda}). \end{aligned} \quad (23)$$

The first term, $p_{\text{aft}}(1 - e^{-\lambda})$, is the probability that an afterpulse carrier is present in the second gate, and the second term, $(1 - e^{-\lambda})$, is the probability that a photogenerated electron or dark carrier is present in the second gate (recall that $\lambda = \text{PDE} \cdot \lambda_p + \lambda_d$). The third term is the cross term that must be subtracted since the first two terms (probabilities) are independent of one another [see Eq. (21)].

Moving on to the third gate ($n = 2$), a pattern begins to emerge

$$\begin{aligned} P_2(\text{gate} = 1) &= p_{\text{aft}}(1 - e^{-\lambda})(1 + p_{\text{aft}}e^{-\lambda}) + (1 - e^{-\lambda}) \\ &\quad - (1 - e^{-\lambda})(1 + p_{\text{aft}}e^{-\lambda})(1 - e^{-\lambda}) \\ &= (1 - e^{-\lambda})[1 + p_{\text{aft}}e^{-\lambda} + (p_{\text{aft}}e^{-\lambda})^2]. \end{aligned} \quad (24)$$

That pattern continues for the fourth gate ($n = 3$)

$$\begin{aligned} P_3(\text{gate} = 1) &= p_{\text{aft}}(1 - e^{-\lambda})[1 + p_{\text{aft}}e^{-\lambda} + (p_{\text{aft}}e^{-\lambda})^2] \\ &\quad + (1 - e^{-\lambda}) - (1 - e^{-\lambda})[1 + p_{\text{aft}}e^{-\lambda} + (p_{\text{aft}}e^{-\lambda})^2] \\ &\quad \times (1 - e^{-\lambda}) = (1 - e^{-\lambda})[1 + p_{\text{aft}}e^{-\lambda} \\ &\quad + (p_{\text{aft}}e^{-\lambda})^2 + (p_{\text{aft}}e^{-\lambda})^3]. \end{aligned} \quad (25)$$

The expression can now be simplified to sum notation.

$$P(\text{gate} = 1) = (1 - e^{-\lambda}) \sum_{n=0}^N (p_{\text{aft}}e^{-\lambda})^n, \quad (26)$$

where $p(\text{gate} = 1)$ is the avalanche probability, N is the total number of gates in the exposure (n_{gates}), and n is the number of an individual gate. p_{aft} is the probability that an afterpulsing carrier is present during a given gate, and $\lambda = \text{PDE} \cdot \lambda_p + \lambda_d$. For large values of N , the upper limit of the sum can be assumed infinite since $p_{\text{aft}}(1 - e^{-\lambda})$ is always < 1 , and higher-order terms will be very small. Making these assumptions, the sum in Eq. (26) becomes a Maclaurin series that converges

$$\sum_{n=0}^{\infty} (p_{\text{aft}}e^{-\lambda})^n = \frac{1}{1 - p_{\text{aft}}e^{-\lambda}}. \quad (27)$$

Therefore,

$$P(\text{gate} = 1) = \frac{1 - e^{-\lambda}}{1 - p_{\text{aft}}e^{-\lambda}}, \quad (28)$$

for large values of n_{gates} . This probability (the probability of one or more electrons being present in a gate given a certain afterpulsing probability) is based on a compound Poisson distribution that skews from the standard distribution given the same mean.⁴

When $p_{\text{aft}} = 0$, the probability of a gate equal to 1 is simply Eq. (5). Similarly, when $\lambda = \text{PDE} \cdot \lambda_p + \lambda_d \gg 1$, the probability of a gate equal to one is Eq. (5) again; the additional number of gates triggered due to afterpulse carriers approaches zero because the other carrier generation rates are very high. Figure 2 shows the relationship between avalanche probability, gate fluence, and afterpulse probability.

Since $p_{\text{aft}} < 1$ and $e^{-\lambda} \leq 1$, Eq. (26) quickly converges for large values of N . As long as the total number of gates is $> 10^3$, the error due to assuming an average avalanche probability will be negligible even for very high values of p_{aft} .

To find a new estimate for the average number of photons per gate, Eq. (5) must be solved again, but with the new expression for $p(n \geq 1)$. n_1 's/ n_{gates} (avalanche probability) is now equal to Eq. (28), and the estimate of the mean number of photogenerated carriers per gate (λ_p) is now

$$\hat{\lambda}_p = -\ln \left[\frac{1 - \frac{n_{\text{ones}}}{n_{\text{gates}}}}{1 - p_{\text{aft}} \frac{n_{\text{ones}}}{n_{\text{gates}}}} \right] - \hat{\lambda}_d. \quad (29)$$

Going back to Eq. (16), the variance of the estimate is a function of the first derivative of the estimate and the variance of n_1 's/ $n_{\text{gates}} = x$. The Markov chain will be used to derive this variance.

The Markov chain is a method of ascertaining the distribution of predicted events that are dependent on the present state and nothing else.⁶ In this case, the present state is the n 'th gate value, while the predicted value is for the $(n + 1)$ 'th gate (dependent on the present state). This particular process is discrete, which simplifies the derivation.

To start the derivation, the probabilities of every state transition must be defined. Figure 3 shows a state diagram for the case of afterpulsing and gate values.

P_{01} is the probability of a 1 in the $(n + 1)$ 'th state given a 0 in the n 'th state, and so on for the other probabilities. The sum of the probabilities leaving a state must equal 1. Equations (30) to (33) show the expressions for each probability (λ is the total electron fluence from photon and dark current processes during a single gate):

$$P_{01} = 1 - e^{-\lambda}, \quad (30)$$

$$P_{00} = 1 - P_{01} = e^{-\lambda}, \quad (31)$$

$$P_{11} = p_{\text{aft}} + P_{01} - p_{\text{aft}}P_{01} = 1 - e^{-\lambda}(1 - p_{\text{aft}}), \quad (32)$$

$$P_{10} = 1 - P_{11} = e^{-\lambda}(1 - p_{\text{aft}}). \quad (33)$$

Since afterpulsing has no effect on gates following a recorded zero (per the assumptions in this derivation), the probability for p_{00} and p_{01} are straightforward Poisson probabilities. For p_{11} , the probability of an afterpulse or a λ -generated carrier is the sum of both probabilities minus the cross term. The equation for p_{10} follows from p_{11} . These probabilities comprise the matrix M , which is defined to calculate the probability of a one and zero based on the probabilities for the previous gate [see Eq. (34)]. M is defined such that its columns are composed of the probabilities leaving each state in Fig. 3. M is used to complete Eq. (34), which defines a useful relationship between the avalanche probabilities defined above

$$M = \begin{bmatrix} P_{00} & P_{10} \\ P_{01} & P_{11} \end{bmatrix} \times \begin{bmatrix} p_0(n+1) \\ p_1(n+1) \end{bmatrix} = \begin{bmatrix} P_{00} & P_{10} \\ P_{01} & P_{11} \end{bmatrix} \begin{bmatrix} p_0(n) \\ p_1(n) \end{bmatrix}. \quad (34)$$

In a steady-state approximation (many gates), the probability of 1 or 0 is independent of the outcome of a gate many gates before. Therefore, $p_0(n+1)$ and $p_0(n)$ approach p_0 , just as $p_1(n+1)$ and $p_1(n)$ approach p_1 , as shown in Eq. (35)

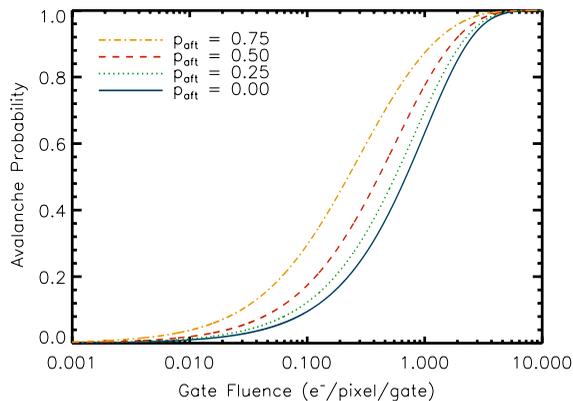


Fig. 2 This plot shows avalanche probability versus gate fluence (in electrons per pixel per gate) for various afterpulsing probabilities. At low and high fluence levels, the avalanche probability converges for all values of p_{aft} .

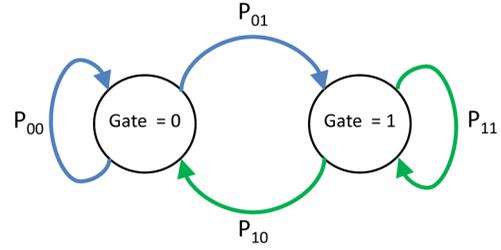


Fig. 3 A state diagram for gate values from the n 'th to $(n + 1)$ 'th gates is shown.

$$\begin{bmatrix} P_0 \\ P_1 \end{bmatrix} = \begin{bmatrix} P_{00} & P_{10} \\ P_{01} & P_{11} \end{bmatrix} \begin{bmatrix} P_0 \\ P_1 \end{bmatrix}. \quad (35)$$

The determinant of M is zero as a result of the defined state relationships, and $p_1 + p_0 = 1$. Using these relationships, Eq. (35) can be solved to find p_1 and p_0 (the steady-state probabilities for 1 and 0, respectively). The expression for p_1 in Eq. (36) should match the expression previously derived in Eq. (28), as a check:

$$\begin{aligned} \begin{bmatrix} P_0 \\ P_1 \end{bmatrix} &= \begin{bmatrix} e^{-\lambda} & e^{-\lambda}(1 - p_{\text{aft}}) \\ 1 - e^{-\lambda} & 1 - e^{-\lambda}(1 - p_{\text{aft}}) \end{bmatrix} \begin{bmatrix} P_0 \\ P_1 \end{bmatrix} \rightarrow P_1 e^{-\lambda}(1 - p_{\text{aft}}) \\ &= P_0(1 - e^{-\lambda}), P_1 = \frac{1 - e^{-\lambda}}{1 - p_{\text{aft}}e^{-\lambda}}, \\ P_0 &= 1 - P_1 = \frac{e^{-\lambda}(1 - p_{\text{aft}})}{1 - p_{\text{aft}}e^{-\lambda}}. \end{aligned} \quad (36)$$

To calculate the variance of the number of counts, the standard definition of variance will be used and interpreted in terms of M and the state probabilities, as shown in Eq. 37

$$\begin{aligned} \sigma_c^2 &= E[c^2] - (E[c])^2 = \sum_{i=1}^N \sum_{j=1}^N \overline{p_1(i)p_1(j)} - \overline{p_1(i)}\overline{p_1(j)} \\ &= \sum_{k=-(N-1)}^{N-1} (N - |k|)(P_1 p_1(k) - P_1^2) \\ &= (N - 0)P_1(1 - P_1) + 2P_1 \sum_{k=1}^{N-1} (N - k)[p_1(k) - P_1] \\ &= NP_1 P_0 + 2P_1 \sum_{k=1}^{N-1} (N - k)[p_1(k) - P_1], \end{aligned} \quad (37)$$

where N is the number of gates (n_{gates}), $k = i - j$, and $p_1(k)$ is the probability of a 1 given that a 1 was recorded k gates before. $p_1(k)$ is also, by definition, the bottom right corner term of M^k , or

$$p_1(k) = [0 \quad 1]M^k \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (38)$$

To find a closed-form solution for M^k , and therefore for $p_1(k)$, M must be diagonalized, or a matrix A must be found such that

$$A^{-1}MA = D \quad M^k = A^{-1}D^kA, \quad (39)$$

where the columns of A are the eigenvectors of M and D is a diagonal matrix of the eigenvalues of M .¹² The eigenvalues of M and the matrices A and D are defined in Eqs. (40) to (42)

$$\text{eigenvalues} = \begin{bmatrix} 1 \\ p_{\text{aft}}e^{-\lambda} \end{bmatrix}, \quad (40)$$

$$A = \begin{bmatrix} \frac{e^{-\lambda}(1-p_{\text{aft}})}{1-e^{-\lambda}} & -1 \\ 1 & 1 \end{bmatrix}, \quad (41)$$

$$D = \begin{bmatrix} 1 & 0 \\ 0 & p_{\text{aft}}e^{-\lambda} \end{bmatrix}. \quad (42)$$

Since D is a diagonal matrix, D^k [Eq. (39)] is calculated simply as a matrix of the k 'th power of the individual terms. Referring back to Eqs. (38) and (39), $p_1(k)$ can be calculated as in Eq. (43)

$$\begin{aligned} p_1(k) &= [0 \quad 1]A^{-1}D^kA \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \frac{(1-e^{-\lambda}) + e^{-\lambda}(1-p_{\text{aft}})(p_{\text{aft}}e^{-\lambda})^k}{1-p_{\text{aft}}e^{-\lambda}} \\ p_1(k) &= \frac{p_{01} + p_{10}(p_{\text{aft}}e^{-\lambda})^k}{1-p_{\text{aft}}e^{-\lambda}}. \end{aligned} \quad (43)$$

As k approaches infinity (many gates), $p_1(k)$ approaches p_1 , the steady-state value of avalanche probability. Substituting Eq. (43) into Eq. (37), and using some geometric series identities, the variance of the total number of counts is

$$\begin{aligned} \sigma_c^2 &= P_1P_0 \\ &\times \left\{ n_{\text{gates}} + 2 \frac{p_{\text{aft}}e^{-\lambda}[(p_{\text{aft}}e^{-\lambda})^N + n_{\text{gates}}(1-p_{\text{aft}}e^{-\lambda}) - 1]}{(1-p_{\text{aft}}e^{-\lambda})^2} \right\}. \end{aligned} \quad (44)$$

Since n_{gates} is assumed to be large ($>10^3$ as expressed previously), the expression simplifies to

$$\sigma_c^2 = n_{\text{gates}}P_1P_0 \left(1 + 2 \frac{p_{\text{aft}}e^{-\lambda}}{1-p_{\text{aft}}e^{-\lambda}} \right). \quad (45)$$

It is interesting to note that the variance in Eq. (45) is a scaled version of the standard binomial variance ($n_{\text{gates}}p_1p_0$), just as the carrier distribution per gate is a skewed version of the standard Poisson distribution when afterpulsing is significant. Equation (45) is the variance of the total number of counts, but the variance of x (the avalanche probability) is scaled by the total number of gates

$$\sigma_x^2 = \frac{\sigma_c^2}{n_{\text{gates}}^2} = \frac{P_1P_0}{n_{\text{gates}}} \left(1 + 2 \frac{p_{\text{aft}}e^{-\lambda}}{1-p_{\text{aft}}e^{-\lambda}} \right). \quad (46)$$

Equation (16) provides the correct expression for the variance of the estimate, substituting the new avalanche probability given in Eq. (28)

$$\hat{\lambda}_p = y(\mu_x) = -\ln \left(\frac{1-\mu_x}{1-p_{\text{aft}} \cdot \mu_x} \right). \quad (47)$$

Solving Eq. (29) in terms of λ and assigning $y = \lambda$ and $\mu_x = p(\text{gate} = 1)$ from Eq. (28), the total variance for the entire measurement is

$$\begin{aligned} \sigma_y^2 &= |y'(\mu_x)|^2 \sigma_x^2 n_{\text{gates}}^2, \\ \sigma_y^2 &= n_{\text{gates}}P_1P_0 \left[\frac{1-p_{\text{aft}}}{P_0(1-p_{\text{aft}} \cdot P_1)} \right]^2 \left(1 + 2 \frac{p_{\text{aft}}e^{-\lambda}}{1-p_{\text{aft}}e^{-\lambda}} \right), \end{aligned} \quad (48)$$

where n_{gates} is the total number of gates in the exposure, p_1 is the avalanche probability, and p_0 is the probability that a gate does not record an avalanche ($1-p_1$). p_{aft} is the probability that an afterpulse carrier is present during a gate, and $\lambda = \text{PDE} \cdot \lambda_p + \lambda_d$ (the sum of the mean number of photo-generated electrons and mean number of dark carriers present in any given gate). When p_{aft} is zero, the variance of the estimate simplifies to Eq. (16) as expected. Combining Eq. (48) with the first half of Eq. (19), the SNR of a GM-APD in photon-counting mode is

$$\text{SNR} = \frac{\text{PDE} \cdot \lambda_p \cdot n_{\text{gates}}}{\sqrt{P_1P_0 \left(1 + 2 \frac{p_{\text{aft}}e^{-\lambda}}{1-p_{\text{aft}}e^{-\lambda}} \right) \left[\frac{1-p_{\text{aft}}}{P_0(1-p_{\text{aft}} \cdot P_1)} \right]^2 n_{\text{gates}}}},$$

where

$$\begin{aligned} P_1 &= \frac{1-e^{-\lambda}}{1-p_{\text{aft}}e^{-\lambda}}, \quad P_0 = \frac{e^{-\lambda}(1-p_{\text{aft}})}{1-p_{\text{aft}}e^{-\lambda}}, \\ \lambda &= \text{PDE} \cdot \lambda_p + \lambda_d. \end{aligned} \quad (49)$$

Note that Eq. (49) simplifies to Eq. (20) (SNR neglecting afterpulsing) when $p_{\text{aft}} = 0$.

Figure 4 shows an overlay of Monte Carlo simulation results and the theoretical solution according to Eq. (49) for the same detector. The simulation agreed with the theoretical data in both mean and standard deviation.

It is important to note that, as explained in the beginning of Sec. 3, an actual value of $p_{\text{aft}} = 0.75$ or even 0.25 is unlikely for most operating conditions if the assumptions stated at the beginning of Sec. 3 hold. However, since the simulation is based on the same assumptions as the derived expression for SNR, and the inputs are given without regard to feasibility, the comparison of simulated to calculated results in Fig. 4 is valid. The exaggerated values of p_{aft} more easily illustrate the overall trends in SNR behavior across a range of fluence values as the afterpulsing probability changes.

The earlier onset of roll-off at high fluence for larger values of p_{aft} is due to an effective decrease in saturation level. Given the same fluence, the avalanche probability will increase with increasing afterpulse probability. The roll-off at low fluence is still due to background noise (DCR). While the relative SNR still has a maximum of $\sqrt{\text{PDE}}$ for the case of $p_{\text{aft}} = 0$, the maximum for cases where $p_{\text{aft}} > 0$ decreases.

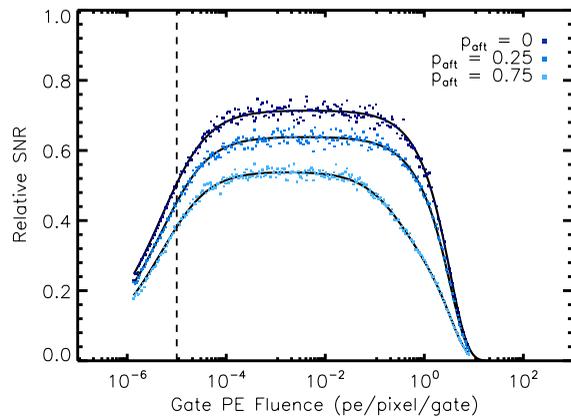


Fig. 4 This plot shows Monte Carlo results (individual points) and analytical solution (corresponding solid lines) for the relative SNR of a GM-APD in photon-counting mode versus gate fluence for multiple afterpulsing probabilities. The dashed vertical line notes the fluence at which photogenerated signal and noise contributions are equal. Relative SNR is normalized to the ideal SNR, the shot-noise limited case where $\text{SNR} = \sqrt{\text{Fluence}}$.

4 Conclusions

An expression was derived for the SNR of a given intensity measurement using a GM-APD in gated (photon-counting) mode. The expression is a function of signal level, PDE, DCR, afterpulsing probability, gate length, and the number of gates sampled (n_{gates}). The theoretical results agreed with carrier-level Monte-Carlo simulations across a variety of input values. Notably, afterpulsing probability has been integrated into the SNR equation and an expression for estimating the signal given significant afterpulsing is established. This is significant because it allows theoretical comparisons between current technologies [electron-multiplying CCDs (EMCCDs), linear-mode APDs (LM-APDs), CCDs] and GM-APD array-based imaging detectors, even when afterpulsing is significant. The expression does have limitations, however, given the assumptions made in the derivation. The most important of these assumptions is that afterpulsing is only dependent on the state of the gate immediately prior to the gate in question (which excludes the possibility of an avalanche due to a trap releasing a carrier from two or more gates prior). For most devices and operations, this assumption is reasonable, though measured results should be checked against the limits imposed by the stated assumptions.

The equation is also useful for choosing the best operating conditions for such a detector (and similar detectors), which might include small amounts of afterpulsing as a trade-off for higher SNR with higher duty cycle. Given any combination

of detector characteristics and operational settings, the scientific imaging potential of a GM-APD in gated mode can now be evaluated. The derived expression is an important tool going forward in the search for the next great detector for photon-counting applications.

Acknowledgments

This work was supported by NASA Headquarters under the NASA Earth and Space Science Fellowship Program—Grant NNX13AO54H. The author would like to thank Dr. Joong Lee of the RIT Center for Detectors and Dr. Brian Aull of the MIT Lincoln Laboratory for their insights into mathematical methods.

References

1. D. F. Figer et al., "A photon-counting detector for exoplanet missions," *Proc. SPIE* **8151**, 81510K (2011).
2. B. F. Aull et al., "Adaptive optics wavefront sensors based on photon-counting detector arrays," *Proc. SPIE* **7736**, 773610 (2010).
3. P. Gatt, S. Johnson, and T. Nichols, "Geiger-mode avalanche photodiode lidar receiver performance characteristics and detection statistics," *Appl. Opt.* **48**(17), 3261–3276 (2009).
4. S. Vinogradov et al., "Probability distribution and noise factor of solid state photomultiplier signals with cross-talk and afterpulsing," in *IEEE Nuclear Science Symp. Conf. Record*, pp. 1496–1500, IEEE Operations Center, Piscataway, NJ (2009).
5. R. Hadfield, "Single photon detectors for optical quantum information applications," *Nat. Photonics* **3**(12), 696–705 (2009).
6. A. Lacaite et al., "Single-photon detection beyond $1\mu\text{m}$: performance of commercially available InGaAs/InP detectors," *Appl. Opt.* **35**(16), 2986–2996 (1996).
7. R. F. Pierret, *Semiconductor Device Fundamentals*, Addison-Wesley Publishing Company, Reading, MA (1996).
8. R. J. McIntyre, "On the avalanche initiation probability of avalanche diodes above the breakdown voltage," *IEEE Trans. Electron Devices* **20**(7), 637–641 (1973).
9. W. Kindt and K. de Langen, "Integrated readout electronics for Geiger mode avalanche photodiodes," in *Solid-State Circuits Conf.*, pp. 216–219, IEEE Operations Center, Piscataway, NJ (1998).
10. A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York (2002).
11. R. Ben-Michael, M. A. Itzler, and B. Nyman, "Afterpulsing effects in $1.5\mu\text{m}$ single photon avalanche photodetectors," in *19th Annual Meeting of the IEEE Lasers and Electro-Optics Society*, pp. 783–784, IEEE Operations Center, Piscataway, NJ (2006).
12. M. Anthony and M. Harvey, *Linear Algebra: Concepts and Methods*, Cambridge University Press, New York (2012).

Kimberly Kolb is a PhD student at the Rochester Institute of Technology. She received her BE degree in microelectronic engineering in 2008 from the Kate Gleason College of Engineering at the Rochester Institute of Technology and her MS in imaging science from the Chester F. Carlson Center for Imaging Science at the same university in 2011. She won a NASA Earth and Space Science Fellowship for her dissertation proposal, "Single photon-counting detectors for NASA astronomy missions," wherein she proposes to evaluate and compare three types of semiconductor-based photon-counting detectors: Geiger-mode avalanche photodiodes (APDs), linear-mode APDs, and EMCCDs.