Crosstalk Analysis of Integrated Geiger-mode Avalanche Photodiode Focal Plane Arrays

Lincoln Laboratory, Massachusetts Institute of Technology, 244 Wood St., Lexington, MA, USA 02420-9108

ABSTRACT

Arrays of photon-counting Geiger-mode avalanche photodiodes (APDs) sensitive to 1.06 and 1.55 \( \mu \text{m} \) wavelengths and as large as 256 x 64 elements on 50 \( \mu \text{m} \) pitch have been fabricated for defense applications. As array size, and element density increase, optical crosstalk becomes an increasingly limiting source of spurious counts. We characterize the crosstalk by measurement of emitted light, and by extracting the spatial and temporal focal plane array (FPA) response to the light from FPA dark count statistics. We discuss the physical and geometrical causes of FPA crosstalk, suggest metrics useful to system designers, then present measured crosstalk metrics for large FPAs as a function of their operating parameters. We then present FPA designs that suppress crosstalk effects and show more than 40 times reduction in crosstalk.

Keywords: avalanche photodiodes, Geiger-mode, APD, optical crosstalk, single-photon counters

1. INTRODUCTION

Avalanche photodiodes (APDs) operating in the Geiger mode (Gm) can be used to detect single infrared photon arrival with subnanosecond accuracy.\(^1\) Arrays of GmAPDs can be constructed and hybridized with silicon read-out integrated circuits (ROICs) and micro optics to form a focal plane array (FPA) for use in imaging or other applications.\(^2\) MIT Lincoln Laboratory has fabricated large arrays of APDs sensitive to 1.06 and 1.55 \( \mu \text{m} \) wavelengths as large as 256 x 64 elements and down to 50 \( \mu \text{m} \) pitch for defense applications.

One limitation of densely packed GmAPD FPAs is optical crosstalk, where the process of detection produces light which couples to nearby APDs in the array and induces spurious events. As array size, density and performance increase, optical crosstalk becomes an increasingly limiting source of these spurious counts.

This paper discusses the physical and geometrical causes of FPA crosstalk, suggests metrics useful to system designers, then presents measured crosstalk metrics for large FPAs as a function of their operating parameters. We then present FPA designs that suppress crosstalk effects and show a greater than 40 times reduction in crosstalk.

2. FPA OPERATION

In this section we outline the design of our APD devices and review the physics behind APD operation, touch on well-established APD performance metrics, and describe the operation of the ROIC and optical stack-up hybridized to our FPA.

2.1 APD Design

A schematic of a single APD is depicted in figure 1. The devices are grown in a multi-wafer OMVPE (organometallic vapor phase epitaxy) reactor on InP substrates. Our group has previously presented details on device design and fabrication which can be found in Donnelly, et. al.\(^2\) The device is of the separate absorber-and-multiplier design.
2.2 Emitted cross-talk spectrum

Within the absorber layer of the device, an incident photon creates charge carriers. After creation, the carriers drift through the field stop layer into the multiplier, where impact ionization produces a macroscopic (1-3 mA) current pulse. During avalanche, the large quantities of carriers in the multiplier have a broad energy distribution, and so emit light due to recombination. Finkelstein et. al.\(^3\) estimate the conversion efficiency in InP to be \(3 \times 10^{-4}\) photons/carrier.

We measured the breakdown spectrum of our InP based APDs to have both a large band edge component at about 950 nm and a broad, possibly thermal component. A spectrum is presented in figure 2.

![InP APD Breakdown Emission Spectrum](image)

Fig. 2. Measured spectrum of InP APD breakdown emission. Corrections for grating efficiency, detector cut-off, etc are included. The broadband component fits roughly a 3000K blackbody spectrum.

A 1200°C blackbody source was used to calibrate the luminescence measurement system. Corrections for grating efficiency, detector cut-off, etc are included in the measured emission spectrum. Two primary components are observed: A near band-edge component and a broadband component that fits roughly a 3000K blackbody spectrum.
2.3 Standard APD performance metrics

Three interdependent figures of merit describe the utility of individual GmAPDs: PDE, Dark Count Rate (DCR), and reset time. PDE, or detection efficiency, is simply the probability that a detection event results from a photon incident on the focal plane, although it is sometimes quoted without coupling and fill-factor type losses for a photon inside the detector. Typical values range from 20-50%, depending on a variety of factors. Dark count rate measures the rate at which the devices fire in the absence of any light, due to charge carriers within the device from thermal, tunneling or other low-probability processes. Dark counts follow Poisson statistics. That is, they have an equal probability of occurring per unit time that the device is active, and are measured in mean number of events per second. Dark counts give both the source events and the background to our crosstalk measurement, as detailed later. InP GmAPD DC values tend to range from $10^4 - 10^6$ Hz at room temperature. After an avalanche event, the likelihood of a subsequent dark count is elevated if the APD is immediately rearmed above its breakdown voltage. Finally, the reset or dead time is the minimum time the APD requires after an event to avoid a likely dark count, called an afterpulse, upon rearming. Typically, this time is on the order of a microsecond at room temperature.

2.4 FPA design

Each GmAPD in a focal plane array (FPA) is separately armed and timed. A simplified diagram of the pixel arming and readout is given in figure 3.

![Fig. 3. Simplified circuit diagram of single pixel arming and readout](image)

The arming circuit is designed such that capacitive coupling limits the total charge available to pass through the APD. During breakdown, an arming capacitor discharges through a current limiting resistor. At some point during avalanche, depletion of the available charge restricts the APD discharge, which follows an exponential-like decay profile, shown in figure 4, until the digital logic registers the event and engages the active quench.

![Fig. 4. Simulated current response of an APD. The decay time constant is less than a nanosecond, but the exponential tail continues until the active quench is asserted several nanoseconds later.](image)
The digital timing circuit for each pixel times a photon arrival (or dark count) into a discrete time bin. One version of the timing circuit used on our FPAs is described by B. F. Aull, et al. To read out the digital time stamps for the array, the ROICs used throughout this article are designed to operate in a framed mode, depicted in figure 5.

![Diagram of framed ROIC operating timeline](image)

**Fig. 5.** Framed ROIC operating timeline. Photon time of arrival is captured, the individual device is disabled for the remainder of the range gate, and the digital timing value is read out after the end of the range gate.

In framed operation, the entire array is armed together, allowed to observe for a time, then disarmed, digitally read out to a buffer off-chip, and re-armed together. Each pixel fire is separately digitally timed into discrete arrival time bins. Once a pixel fires, it cannot register a second event within a frame. Hence, the pixel is blocked — it is insensitive to any arriving photons. This blocking property is important in the crosstalk context because in order to accurately determine the probability of a crosstalk event, the algorithm needs to take into account which nearby devices are armed and available to register an event.

### 3. CAUSES OF CROSSTALK

In this section we discuss how photons generated during a detection event can be detected as spurious counts within a focal plane array.

#### 3.1 Optical Coupling

The broad spectrum light produced during avalanche propagates outward into the device substrate. In both mesa and planar devices, this light can reflect off the back wafer surface and trigger other APDs in the neighborhood of the primary APD. The process is illustrated in figure 6.

![Diagram of optical coupling](image)

**Fig. 6.** Hot carrier electroluminescence generated in the APD avalanche region can couple to nearby devices through the front optical surface of the InP APD wafer.

The degree of optical coupling depends on the geometry of the array; the size and spacing of the active area, and the thickness of the InP or Si substrate. Another primary factor is the interface at the front optical interface (often called the back side, since growth and fabrication focus on the other side) at which the total internal reflection angle is readily reached. For an InP-air interface, the angle is 17.7°, which strongly increases the coupling efficiency for the next-
nearest-neighbor and farther pixels in our FPAs. Calculating the precise coupling for all devices is best done with optical design software. Rech, et. al. \(^5\) have published a treatment of this for G-M silicon APD arrays made from planar devices.

4. CROSSTALK MEASUREMENT

This section describes the crosstalk signature in FPA data frames, proposes metrics for quantifying crosstalk, describes one specific algorithm for extracting these metrics, and presents measured crosstalk results for our FPAs. Since there is a background random firing rate, it is impossible to establish causality – crosstalk photon or dark count – for any given event. Instead, we can measure how events are correlated at relative distances and time delays.

4.1 Goals and definitions

The ultimate goal of quantifying crosstalk is to specify the specific probability that a firing pixel at location \(x\) and time bin \(t\), named the primary event, causes a secondary event at another pixel at another location at some later time, \(x', t'\). There are several problems with calculating this number directly for all \(x\). The probability of firing at any one location and at any one time bin is very small. If we assume the optical coupling between pixels is identical throughout the array, and that the distribution of detection efficiencies for a given operating condition is well behaved, we can integrate behavior from many primary fires into statistics representative of typical crosstalk behavior for every pixel throughout the array.

Specifically, we propose two quantitative measures: The expected number of counts within a certain distance and delay from an originating fire \(N_{\rho\tau}\) (assuming every pixel is active), and the probability \(P_{\rho\tau}(n)\) of \(n\) or more correlated counts within a region \(\rho\) and delay range \(\tau\). In this section, we define an algorithm for measuring these statistics, discuss their interpretation and present measured data from focal plane arrays (FPAs).

First, we define a coordinate system for discussing source and secondary events, diagrammed in figure 7. Within a FPA of size \((nx, ny)\), a source event happens at some coordinate \(\vec{x}\) and in time bin \(t\), both constrained to integer values. Similarly, a candidate secondary event occurs at a different coordinate \(\vec{x}', t'\).

![Fig. 7. Coordinate system definition for relating a primary fire within a FPA at coordinate \(x\) to a secondary fire at \(x'\), mapped to a space where every primary event occurs at the center.](image)

Under the assumption that the crosstalk response is similar at different parts of the FPA, we would like to measure the probability that a pixel offset \(\vec{x}' - \vec{x}, t' - t\) from any event fires. This array of probabilities becomes very much like a discrete point spread function (PSF) for the pixel array, both in space offset and in time delay. This PSF measures the “leakage” probability that a single photon event casts into the FPA around it. The dimensions of this array range from \(\vec{x}' - \vec{x} = -nx, -ny\) to \(nx, +ny\), and in the time delay dimension from \(t' - t = 0\) to some max gate width \(\tau\).

Within this discrete PSF is included random background events not caused by crosstalk just due to the dark count rate of the pixels. This manifests as a background probability uniform over \(\vec{x}' - \vec{x}, t' - t\). We can measure this background by
looking at the PSF far removed either in distance or in time from the initiating event, when hot-carrier emitted photons are unlikely to reach or after the primary event and any likely descendents have been quenched. In practice, we use the rate some $10^+$ exponential decay constants after a primary fire as the background rate after the rate has subsided to a visually apparent background value. Distance removed rates have the disadvantage of less accumulated statistics and an increased sensitivity to FPA uniformity. If the total number of events is not dominated by a broad pattern of pixels with elevated dark count rates, the background reasonably close to the center of the PSF comes from a broad sampling of pairs of pixels, and the variation in the background rate is dominated by statistical fluctuations in the measurement. Since the measurement is itself a Bernoulli random process and the probabilities are small, the standard deviation of the background is nearly equal to the square root of the mean background level.

4.2 Crosstalk Measurement Algorithm

The primary consideration in designing the algorithm to estimate the PSF probabilities is to correctly account for both the number of times a secondary pixel fires and the number of times it could have fired but did not. This involves keeping track of every fire in relation to every other fire within a frame. Below, we use the convention that an array coordinate with a square bracket is a single coordinate location that is looped over inside the algorithm, while non-indexed coordinate variables represent a list of one or more sets of coordinates. It may be clearer to read the pseudocode below.

Within a single frame, three tracking arrays are used as the algorithm advances through the time bins: The first array tracks the active pixels with a simple Boolean value for each physical pixel; The second tracks the cumulative number of correlated fires which occur for each PSF offset and time delay; The third array tracks the cumulative number of active pixels of every spatial offset in relation to every potential source event. After initializing these arrays, the algorithm loops over every time bin $t[j]$ containing an event and for each primary event in that bin $\tilde{x}[i], t[j]$, finds secondary events $\tilde{x}', t'$, calculates offset vectors $\tilde{x}' - \tilde{x}[i], t' - t[j]$. After calculating these offset positions, the algorithm increments the cumulative fires array at $\tilde{x}' - \tilde{x}[i], t' - t[j]$, increments the cumulative chances array at $- \tilde{x}$ using the current active pixels array, then sets the active pixels at $\tilde{x}$ to disabled in the active pixels array. Pseudo code is shown below.

```
FOR all frames in set
  Initialize active pixel map
  FOR time bins $t[j]$ that contain any hits
    Find coordinates $x$ of source hits in $t[j]$ 
    Find coordinates $x'$ of secondary hits between $t[j]$ and $t[j] + \tau$
    FOR all source hits $x[i]$ in time bin $t[j]$
      Increment events array at coordinates $x' - x[i], t' - t[j]$
      Increment chances array at $- x$ with active pixel map
  ENDFOR
  Update active pixel map at $x$ as inactive
ENDFOR
```

In practice, we precalculate the hit offsets by bin and store them in order to avoid repeated searching and calculation. The end result is two arrays, a correlated number of events array (size 2nx-1, 2ny-1, tau) and a cumulative active pixels (or “chances”) array (size 2nx-1, 2ny-1, 1). Since our algorithm accumulates statistics independently from a large number of frames, it parallelizes very well. After calculation, we expand the chances array to tau by subtracting the previous time bin’s events from the number of chances for the next time bin: $\text{chances}[:, :, 1] = \text{chances}[:, :, 1] - \text{events}[:, :, 1]$. One subtlety of crosstalk fires that occur in the same time bin (i.e. $t' - t = 0$) is that we do not know which event is a primary and which is a result. The algorithm counts each pair as a primary and as a result. To deal with this, we divide the event count at $t' - t = 0$ by 2. Each pair of events is counted equally likely to be the source and the induced result. After this correction, the events array is then divided element-by-element by the chances array to form a PSF probability estimate array. A sample PSF is presented in figure 8.
Fig. 8. Cross-talk PSF data integrated over $\tau=21$ time bins. Source data is dark-count rate statistics for a 128x32 focal plane array. The x-y coordinates are the number of pixels away from a GmAPD that caused the primary firing. For example, a pixel in the corner of a 128x32 array can be as far as (31,127) pixels in (x,y) away from a pixel in the opposite corner. The color bar is linear, with white denoting ~1.5% chance of a nearby pixel firing due to cross talk.

A large number of frames, typically thousands, is required to measure an adequate number of events resulting in a good crosstalk-to-background ratio. In order to determine the quantity of source data is necessary, we can consider that each offset in the PSF is the sum of as many Bernoulli random variables as there are events under consideration. This makes each offset a binomial random variable, with mean $n_{\text{event}}p_{\text{fire}}$ and variance $n_{\text{event}}p_{\text{fire}}(1-p_{\text{fire}})$. This is not surprising, since it goes to the values expected for a Poisson process for small $p_{\text{fire}}$. This implies for any meaningful signal, we need at least as many events as the inverse of the probability of a count within integration time $\tau$, where $P_{\text{event}} = \sqrt{DCR \cdot \tau}$.

For a given hypothetical primary event, the PSF is a large array of Bernoulli random variables, each with a different probability of firing and each contingent on the earlier variables at the same offset. This three dimensional PSF probability estimate is awkward to use to compare FPAs. A two-dimensional metric or even a scalar value would be much more useful as a specification or a comparison between operation at different temperatures and bias voltages.

4.3 Figures of merit with fewer than three dimensions

We come back to the two quantitative measures defined at the beginning of the section and work on how to derive them from the PSF. We would like to add the PSF variables in space offset and time delay to derive the probability mass function (PMF) of a composite random variable $P_{\text{pt}}(n)$ that tells us the probability of any $n$ pixels firing within a certain time and spatial offset from a primary fire, assuming that all pixels around the primary are active. Once we have this we can calculate the expectation value of $n$ in the normal way to get $N_{\text{pt}}$.

First, consider the time delay dimension. Since there can be maximally one event per frame at a certain spatial offset, we add them by treating them as a sort of Markov chain, as shown in fig 9.

Fig. 9. Simple Markov chain model of successive probabilities for a single spatial offset over time delay.

To find the probability of an event occurring, we add the time based probabilities in the following simple way:
This addition gives us a single random variable with the probability of an event at any time within \( \tau \) time bins for a single specific spatial offset. Using this method, we can add up \( \tau \) time bins for all the spatial offsets and get a two-dimensional PSF which displays nicely (figure 8) and helps us understand the geometrical factors leading to crosstalk.

Now, consider the spatial offset. For two independent integer valued random variables \( X \) and \( Y \) with PMFs \( p_X \) and \( p_Y \), we can calculate the PMF of the random variable \( Z=X+Y \) by:

\[
p_Z(z) = \mathbb{P}(X + Y = z)
\]

\[
p_Z(z) = \sum_x p_X(x)p_Y(z-x)
\]

Casting this equation so that we wish to derive the PMF of a point-spread region with possible hits \( N \) from a smaller region with possible hits \( N' \) and a single pixel offset with hits \( M \), we have the equation

\[
p_N(n) = \sum_{m=(0,1)} p_M(m)p_{N'}(n-m).
\]

Computationally, this is the iterated convolution of the PMFs of the component variables.

The result of these computations is a PMF for a region and time which gives a measured probability of zero, one, two, etc. events occurring within a region and time related to a primary event. From this PMF, we can calculate our figures of merit.

### 4.4 System-applicable figures of merit

The algorithm described above produces a 3D spatial and temporal array of the probability of cross-talk from a nominal detector element. This section presents a simplified figure of merit which we routinely use to characterize different FPAs. The figure of merit can be used to estimate the performance impact for many system applications. For example, in a ladar scene it is cross-talk firings in the spatial vicinity of the primary photon event that can affect range precision and angle-angle resolution.

Since the functional form and time constant of the optical crosstalk only changes marginally from array to array, we choose the region \( \Omega \) to consist of a 9x9 element subarray around a primary firing element, with the primary element at the center of the 9x9 subarray. From the region of interest's PMF, we then compute the metric referred to as \( \mathbb{P}(>1) \) or \( \mathbb{P}(\geq2) \), the probability of more than one sympathetic firing occurring in the 9x9 subarray. This probability depends strongly on the voltage bias which, in turn, determines the PDE. Therefore, we arbitrarily constrain \( \mathbb{P}(>1) \) to be less 10% and quote the PDE at that setting of the GmAPD voltage bias. For example, a given 128x32 array may be quoted as having median PDE = 25% with \( \mathbb{P}(>1) = 10% \). Stated differently, the system designer knows the median PDE of the array and also that 90% of the time there will be only one or zero sympathetic firings in a given 80-element subarray. He or she can use that information to estimate blockage loss and noise impact that would affect the system link budget.

### 5. DESIGNS FOR IMPROVED CROSSTALK

There are a number of ways to suppress optical crosstalk. We loosely categorize them as electrical, physical, and spectral schemes.

#### 5.1 Electrical design to reduce optical crosstalk

Since the amount of light emitted by the APD is directly proportional to the total charge passed through the device during breakdown, electrical schemes seek to adjust the device, ROIC, or biasing scheme to reduce the total current, and thus the emitted light. Designing the circuit to have a high series resistance, a low arming capacitance, or quickly quenching the avalanche thus reducing the time current flows through the are all viable approaches. We do not present any results in this area.
5.2 Physical design to reduce optical crosstalk

Optical schemes seek to reduce the optical coupling between neighboring pixels by manipulating the optical path between the source and result APDs. This involves manipulating the device size, wafer geometry, or optical interfaces. Most obviously, reducing the size of the APDs and increasing the separation between them has a strong reductive affect on crosstalk, although it entails design trade-offs with system optics and detection efficiency that must be carefully considered. For planar rather than mesa devices some groups etch trenches between APDs to block direct device to device crosstalk through the substrate.\(^7\) For our devices, the patterned metallization on the illumination side of the FPA (figure 10) partially suppresses internal reflection and thus reduces crosstalk to more than half the PSF offset locations.

![Diagram of an FPA with patterned metallization](image)

Fig. 10. (a) Schematic of additional (processing) backside/optical front patterned layer. (b) pattern of optical power reflected back to the device plane from a circular emitter at the position labeled primary. The white circles are the approximate active area of neighboring APDs.

Now we proceed to show the results of similar arrays with and without backside metallization. Comparison charts with PSFs and the derived scalar metrics are shown in figure 11. We expect the crosstalk to scale as the product of the overbias (available charge on arm node from capacitor) and the quantum efficiency (chance of an event from crosstalk photon), which is a nonlinear function of overbias\(^2\). As the crosstalk gets worse, \(P(>1)\) approaches 1 and \(N\) passes 1. \(N\) passing 1 implies the possibility of a self-sustaining chain reaction. In practice, FPAs are able to operate usefully beyond \(N=1\) due to sizeable fractions of blocked pixels disrupting the chain. The first thing to note when looking at figure 11 is that the suppressed array reaches \(N=1\) at a much higher overbias, so a much greater range of operation is available to the designer, including bias ranges with better detection efficiency. Exactly how much improvement is seen depends on which metric at which overbias, but the metrics show roughly two orders of magnitude (100x) suppression of crosstalk.
5.3 Spectral control to reduce optical crosstalk

For applications such as ladar or communications interested in a narrow range of wavelengths, it is possible to design a filter into the FPA that passes the wavelength of interest and attenuates other regions of the broad hot carrier avalanche emission spectrum, as seen in fig 13. For devices not illuminated through the substrate, this filter need not even pass the wavelength of interest, merely attenuate light in the sensitive region of the device. Inclusion of the filter should be possible in 3D integrated devices where a filter is hybridized to the devices or grown epitaxially as we do with InP or as done by Ghioni et. al. in silicon on SOI. For our devices, this filter layer is a lattice matched quaternary layer, grown such that the band edge absorbs wavelengths 10s of nanometers shorter than the detection wavelength, effectively creating a long-pass filter beneath every APD. Light is absorbed both as it is emitted and before it is coupled into a secondary APD.

Fig. 12. Schematic showing filter layer, labeled in blue. Also shows the metal layer at the bottom surface of the wafer.
InP APD Breakdown Emission + 1.06 um Absorber Spectra

Fig. 13. (a) Schematic showing 1 µm p+ InGaAsP lattice-matched filter layer, colored in green. (b) Avalanche emission spectrum, overlaid with filter layer and APD absorber cut-off wavelengths.

Fig. 14. Comparison of crosstalk metrics between fully packaged FPAs with and without crosstalk filter layer and backside metallization. The first column of charts show $P(>0)$, $P(>1)$, and the corresponding background values. The second column of charts show $N$. Device size is different from previous comparison. Both are evaluated over $p>=9x9$ region and $\tau = [0 \, 20\, \text{ns}]$.

Charts comparing derived scalar metrics are shown in figure 14. These devices are a smaller (20 µm nominal diameter) than the previous comparison. The conventional and crosstalk suppressed designs are on differently doped substrates, which may have different absorption characteristics. The data were also taken at different temperatures, 0°C and 20°C respectively. Low temperatures reduce the dark count rate and therefore the background rates significantly. To first order, cutoff and absorber band edges shift together with temperature, so temperature should only have a small effect on measured crosstalk. A possible source of error is an unusually broad distribution of breakdown voltage across the array leading to a low stated value for the breakdown voltage and an artificially low value for $V_{\text{overbias}}$, since the breakdown voltage is measured relative to the with the lowest values.

Depending on the metric, this comparison shows between 9 and 20 times improvement.
REFERENCES


