

Progress in self-quenching InP-based single photon detectors

Mark A. Itzler*, Xudong Jiang, Bora M. Onat, Krystyna Slomkowski
Princeton Lightwave Inc., 2555 US Route 130 South, Cranbury, NJ 08512

ABSTRACT

The operation of InP-based single photon avalanche diodes (SPADs) in Geiger mode provides great utility for the detection of single photons at near-infrared wavelengths between 1.0 and 1.6 μm . However, SPADs have performance limitations with respect to photon counting rate and the absence of photon number resolution that, at the most fundamental level, can be traced back to the positive feedback inherent in the impact ionization-driven avalanche process. In this paper, we describe the inclusion of negative feedback with best-in-class InP-based single photon avalanche diode (SPAD) structures to form negative feedback avalanche diodes (NFADs) in which many of the present limitations of SPAD operation can be overcome. The use of thin film resistors as monolithic passive negative feedback elements ensures rapid self-quenching with very low parasitic effects. We demonstrate a qualitative difference in the performance of NFADs in the two regimes of small and large negative feedback. With small feedback, we have studied the behavior of the persistent current prior to quenching, for which we have found oscillatory behavior as well as an exponentially distributed duration. For large feedback, we find rapid quenching, accompanied by evidence for a partial discharge of the detector capacitance, leading to charge flows as low as $\sim 3 \times 10^5$ carriers associated with each avalanche event.

Keywords: avalanche photodiode, single photon detector, negative feedback, NFAD, SPAD, InP, InGaAsP

1. INTRODUCTION

For the measurement of single photons, detectors based on avalanche photodiode (APD) structures operating above the APD breakdown voltage V_b often provide the most advantageous combination of performance and practicality. Single photon avalanche diodes (SPADs) operating in Geiger mode based on Si and InGaAsP materials systems have gained increasing relevance for photon counting in the visible and near-infrared wavelength regimes, respectively. More recently, there has been a trend towards increasing the performance and functionality of SPADs using enhanced detector designs that provide self-quenching of the SPAD avalanche pulse. A number of different approaches have been adopted for achieving self-quenching with monolithically integrated structures in both Si and InGaAsP devices, and the resulting devices show promise for overcoming the limitations of canonical SPADs. Appropriately designed self-quenching can dramatically reduce the charge flow associated with each avalanche event, leading to associated reductions in afterpulsing effects and optical crosstalk. Operation of self-quenching devices requires just a fixed dc bias voltage and represents a considerable simplification relative to the complex bias circuitry required for SPADs.

In this paper, we summarize the basic operation of self-quenching detectors, and we describe our specific approach for negative feedback avalanche diodes (NFADs) employing monolithically integrated quench resistors to achieve high-performance free-running photon counting at 1064 nm and 1550 nm. We have studied devices in two different regimes of resistive negative feedback — small feedback ($< 100 \text{ k}\Omega$) and large feedback ($> 500 \text{ k}\Omega$) — and we show significantly different behavior for devices of these two different types. With small feedback devices, there is a critical range of excess bias beyond which the feedback is insufficient to induce rapid quenching of avalanche events. In this limit, the flow of avalanche current can persist for considerable time scales. We have found that at a given value of excess bias, the duration of the persistent current varies randomly and is described by an exponential probability distribution function. Moreover, we have found that the persistent current exhibits pronounced oscillations, and this oscillatory behavior promises to provide a more detailed understanding of the underlying dynamics of avalanche effects. For large feedback devices, quenching is consistently rapid (i.e., there are no persistent current effects). Based on the dependence of avalanche pulse heights and charge flow on excess bias, we find evidence in this large-feedback

* mitzler@princetonlightwave.com, tel: (609)495-2551, fax: (609)395-9113

limit that quenching may be occurring before the complete discharge of the diode capacitance. The prospect of partial discharge of the diode capacitance is of great interest relative to the fundamental goal of reducing the total charge flow associated with each avalanche event. Finally, the highly deterministic behavior of avalanche events in the NFADs shows the potential for photon number resolving detectors created using multi-element NFAD arrays.

1.1 Drawbacks of Single Photon Avalanche Diodes

For many applications, the most practical photodetector available today with single photon sensitivity between 1.0 and 1.6 μm is the InGaAsP/InP single photon avalanche diode (SPAD). However, for general use — especially free-running operation — implementations of InP-based SPADs have typically had limited photon counting rates (e.g., ~ 10 MHz or less), and these detectors do not have the ability to resolve the number of photons in a detected optical pulse. For high-performance photon counting applications such as long distance free-space optical communications[1], quantum cryptography[2], and long-range lidar measurements[3], single photon detectors with much faster counting rates in the range of 100 MHz to 1 GHz will be essential, and the ability to resolve photon number will provide considerable value in many cases.

SPADs are avalanche diode structures designed to operate above their breakdown voltage V_b , so that a single photoexcited carrier can induce a macroscopic current by avalanche breakdown. Many of the shortcomings of conventional SPAD performance are related to the positive feedback inherent in the impact ionization process: if electrons and holes both have finite impact ionization coefficients, then every carrier created by impact ionization tends to create additional carriers, leading to a runaway avalanche. This positive feedback is useful to the extent that it can lead to the very rapid build-up of a macroscopic amount of charge in response to the absorption of just a single photon, providing the SPAD with the functionality of a photon-activated digital switch. However, when charge flows during an avalanche event, some fraction of the carriers are trapped at defect sites in the avalanche region of the structure. If these carriers are detrapped at a later time when the device has already been re-armed, they can lead to dark counts — referred to as “afterpulses” — that are correlated to previous avalanches. Larger avalanches that involve the flow of more charge will result in a greater number of trapped charges and consequently larger afterpulsing effects. The afterpulsing effect can be mitigated by imposing a sufficiently long “hold-off” time before re-arming the SPAD, but this existing solution introduces the present limitation on achievable count rate.

Also related to positive feedback in SPADs is the fact that in virtually all SPAD structures, a non-zero impact ionization coefficient ratio ($k > 0$) leads to significant variation in the amount of charge generated per unit time during an avalanche event. This variation gives rise to the canonical excess noise factor $F(M) \equiv \langle M^2 \rangle / \langle M \rangle^2$ that describes gain variations in avalanche photodiodes operated in linear mode (i.e., below the breakdown voltage). Although $F(M)$ is not directly applicable to Geiger mode operation (i.e., above the breakdown voltage), the related gain variation per unit time gives rise to highly stochastic avalanches. Whereas it would be desirable to add signals from multiple detection regions to determine how many detection events have occurred simultaneously, this avalanche stochasticity prevents adequate quantization of the summed signals for obtaining photon number resolution.

Finally, SPADs require complex external electronic circuitry to control on/off gating of the device (in gated-mode operation) or avalanche quenching (in free-running operation), and for higher count rates, these electronics are increasingly challenging.

1.2 Avoiding the Drawbacks of SPADs with Negative Feedback

Because the probability of afterpulses is proportional to the detrapping of charges trapped during previous avalanche events, a reduction in the afterpulsing effect necessitates that fewer charges are detrapped per unit time. The most straightforward way to achieve this goal is to reduce the number of charges trapped during previous avalanches. Two obvious strategies exist: (i) decrease the number of defect sites at which charges can be trapped, and (ii) reduce the number of charges that could be potentially trapped. The first strategy requires the improvement of the underlying avalanche region material quality. Substantial investment in InP material quality was made in response to the large market for InP-based photodetectors during the telecom bubble of the late 1990s, and further improvements are likely to occur slowly and at great expense. (A similar constraint exists for Si-based SPADs given the maturity of the silicon

material system and the fairly slow progress in its defect density.) Therefore, most practitioners involved in the use of SPADs have adopted the second strategy, which entails reducing the amount of charge that flows through the SPAD structure during each avalanche event.

There has been recent success in reducing avalanche charge flow to achieve high speed arming and disarming of SPADs in the context of periodic operation. One challenge with high frequency operation is that imposing fast gating signals on the capacitive SPAD structure gives rise to large transients that will tend to mask small amplitude avalanche events. With purely sinusoidal gating [4], the gating frequency can be readily filtered out, allowing for the recovery of small-amplitude avalanche responses at high frequency. The use of a “self-differencing” technique [5], in which the transients are suppressed by subtracting adjacent periods of the detector response, also provides for small-avalanche detection at high frequency as long as the subtracted waveform does not contain an avalanche response. While these techniques have demonstrated afterpulsing levels of $\sim 1 - 5\%$ at high count rates (e.g., >100 MHz) and are very promising for certain applications, these demonstrations apply only to periodic gated-mode operation with extremely short (~ 0.1 ns) gates, and these operating conditions can be highly constraining for many applications.

A more general concept for reducing the charge flow per avalanche is the introduction of some form of “negative feedback” to very rapidly lower the internal electric field of the avalanche diode in response to the build-up of an avalanche. The use of negative feedback to achieve self-quenching in avalanche diodes initially received attention in the literature on detectors for nuclear physics experiments [6–11]; these treatments focused exclusively on Si detectors. These Si-based structures have been given a variety of labels, most generically referred to as “silicon photomultipliers” (SiPMs). SiPMs have received growing interest given their prospects for displacing legacy photomultiplier tubes in a number of applications, and several companies have commercialized discrete SiPM detectors with micro-pixelated structures (e.g., see [12] and [13]). In contrast, there have been only a few studies reporting results for non-silicon-based negative feedback devices. One approach achieves self-quenching in an InP-based avalanche diode structure with an epitaxial heterobarrier grown right above the avalanche diode to promote avalanche charge pile-up to reduce the internal electric field, thereby quenching the avalanche.[14] Another report describes an InP-based detector that achieves negative feedback through the use of multi-cell photodetectors and an “internal discrete amplification mechanism” that bears similarity to distributed quenching with micro-pixelated active regions.[15]

Our approach to implementing negative feedback—through monolithic integration of surface thin film resistor elements—is conceptually more similar to designs pursued in silicon-based SiPMs in which polysilicon resistors are integrated on a per-pixel basis (although there is no equivalently standardized integration technology for the InP materials system). The use of custom surface-integrated structures provides us with greater design flexibility relative to other InP-based negative feedback devices reported to date: while these alternative structures incorporate feedback elements through epitaxial integration that must be accomplished during epitaxial growth, our surface-integrated feedback elements are photolithographically fabricated during wafer processing. The use of lithographically defined elements gives us greater latitude for design of experiments and design bracketing during manufacturing, and it also allows the use of less complicated epitaxial structures that have been well-proven through several years of manufacturing commercial SPAD products.

As a final advantage of the NFAD design relative to conventional SPADs, it is worth noting that a well-implemented NFAD is extremely simple to operate: with just a fixed dc bias voltage corresponding to the sum of the diode avalanche breakdown voltage V_b and the desired excess bias V_{ex} , the NFAD will independently execute the entire arm, avalanche, quench, and re-arm cycle and generate an output pulse every time an avalanche event is induced.

2. NFAD DEVICE OPERATION

The introduction of negative feedback using a load resistor is conceptually equivalent to passive quenching, but historically, passive quenching has been introduced through hybrid integration approaches that invariably introduce large parasitic effects.[16] With an ideal monolithic integration, parasitic effects can be negligible, and the amount of charge flow Q required prior to avalanche quenching is in principle determined by the product of the avalanche diode depletion capacitance C_d and the excess bias V_{ex} , i.e., $Q = C_d \times V_{ex}$. With an ideal implementation, each avalanche

involves this same amount of charge flow Q , and Q can be made quite small through appropriate scaling of C_d . (The magnitude of the excess bias V_{ex} will be dictated by operational targets for PDE since PDE increases with larger V_{ex} .) Moreover, as we describe below, we have seen evidence from initial experiments [17] of the potential for limiting the current flow per avalanche to even *less* than $C_d \times V_{ex}$, and we hypothesize that because of sufficiently large negative feedback, we are achieving avalanche quenching with only partial discharge of the diode depletion capacitance C_d .

In Figure 1, we illustrate the basic operational concept for the NFAD. The device is modeled as the series combination of a SPAD and a generalized negative feedback load. We have described the load as the parallel combination of a load resistance R_L and load capacitance C_L , but more general (e.g., non-linear) elements are possible. The SPAD itself is conveniently modelled as having two parallel branches. The first branch consists of the diode depletion capacitance C_d ; the second branch includes the diode dynamic resistance R_d above breakdown, the breakdown voltage V_b , and a switch S that is closed to represent the occurrence of an avalanche. An applied voltage $V_a = V_b + V_{ex}$ is imposed across the entire circuit such that the SPAD is reverse biased beyond its breakdown voltage V_b by the excess bias V_{ex} .

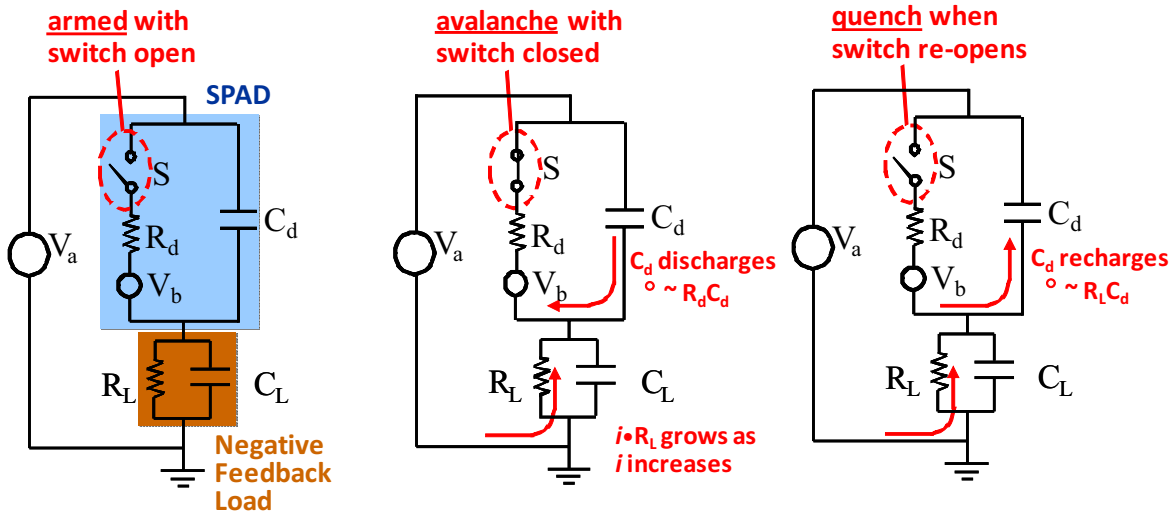


Figure 1. Basic concept of NFAD operation. The device is modeled as the combination of a canonical SPAD with a negative feedback load. The closing of switch S corresponds to the onset of an avalanche event, providing for the discharge of diode capacitance C_d and the shifting of some portion of the excess bias voltage $V_{ex} = V_a - V_b$ from the SPAD structure to the negative feedback load (equivalent to potential drop $i \cdot R_L$). Quenching of the avalanche is represented by the re-opening of switch S , at which point the external source re-charges C_d to re-arm the device.

The NFAD is initially in its armed state with the voltage across the SPAD exceeding V_b by the excess bias V_{ex} . With the switch S open, no current flows. The onset of an avalanche is modeled by the closing of switch S , at which point capacitance C_d discharges through the diode dynamic resistance R_d with a time constant on the order of $\tau_d \sim R_d C_d$. The removal of charge from C_d reduces the voltage across the SPAD structure, although the precise amount of voltage removed depends on the ratio of R_d and R_L . In fact, in steady state, the applied voltage V_a is split between the SPAD structure and the load according to the voltage divider presented by these two resistances in series. The amount of voltage removed from the SPAD structure is precisely the amount of voltage $i \cdot R_L$ developed across the load by the introduction of a current i from the voltage source. When the total current through the SPAD switch S drops to a value smaller than a characteristic quench value i_q , the avalanche will spontaneously quench, represented by the re-opening of switch S . With S open, C_d is re-charged through load resistance R_L , with a recharging time constant $\tau_r \sim R_L C_d$ dictating how long it takes to re-arm the NFAD.

From this description of NFAD behavior, a fundamental design trade-off becomes apparent with the respect to the magnitude of R_L . On the one hand, R_L must be large enough to ensure a small current flow ($\ll i_q$) through the load after the initial capacitive discharge during an avalanche event so that quenching can occur rapidly. On the other hand, because the effective dead time (during which the excess bias is low) is proportional to the recharging time constant

$R_L C_d$, a larger value of R_L leads to a longer re-arming period with consequently lower counting rates. We present further discussion of the impact of R_L being small in Section 4.

3. NFAD DEVICE DESIGN AND MEASUREMENT OF BASIC PROPERTIES

Although NFAD device operation encompasses significant differences from canonical SPAD operation, certain fundamental attributes — such as the trade-off between dark count rate (DCR) and photon detection efficiency (PDE) — will still be dictated by the underlying avalanche diode structure, which is reviewed elsewhere.[18,19] To provide reference data for the underlying SPAD structure upon which the NFAD is based, we fabricated 1550 nm SPAD reference devices without negative feedback elements with device diameters that matched those of the NFAD devices. Using a measurement technique employing short pulse (1 ns) gating [20], we have measured DCR and PDE for 25 μm active diameter SPADs. The DCR of <200 Hz obtained [21] for PDE values of $\sim 15\%$ at 223 K is lower than any 1550 nm SPAD results we have seen reported for comparable conditions, and this performance has been confirmed with measurements by other researchers [22] on other devices from the same process lot. This improved performance relative to earlier generations of PLI 1550 nm SPADs is related to structural changes that included lowered the electric field in the multiplication region for more effective suppression of dark carrier generation by trap-assisted tunneling.

The fabrication of NFADs based on our underlying SPAD avalanche diode structure was carried out using thin film resistor meander lines integrated on the surface of the diode structure. We fabricated two iterations of devices with a range of active area sizes and feedback resistance values, as summarized in Table 1. Compared to 1st iteration devices, the devices from the 2nd iteration had diameters roughly a factor of 2 smaller and feedback resistances roughly a factor of 20 larger. In the discussions that follow, we focus primarily on results obtained from devices representing the extremes of the range of device diameters and feedback resistances obtained. Devices were assembled into either a TO-style package or a 14-pin butterfly-style package with direct fiber coupling to the device. Data were generally obtained for operating temperatures near 230 K.

Table 1: Device Diameters and Typical Feedback Resistances for 1st and 2nd Iteration NFADs

1 st Iteration NFADs			2 nd Iteration NFADs		
Device Type	Device Diameter (μm)	Feedback Resistance ($\text{k}\Omega$)	Device Type	Device Diameter (μm)	Feedback Resistance ($\text{k}\Omega$)
D1F1	34	25	E1G1	22	500
D1F5	34	40	E1G5	22	600
D2F2	46	35	E2G2	22	600
D2F6	46	60	E2G6	22	1100
D3F3	58		E3G3	32	1100
D3F7	58	90	E3G7	32	1900
D4F4	82	95	E4G4	42	1900
D4F8	82	140	E4G8	42	3000

Feedback resistances were experimentally determined by extracting the slope of the linear NFAD current-voltage characteristics measured above the breakdown voltage. As shown in Figure 2, a reference SPAD structure exhibits a very steep current increase above breakdown consistent with a diode dynamic resistance R_d of a few $\text{k}\Omega$; the precise value depends on the device diameter, but for the smallest diameter devices studied (i.e., 22 μm), $R_d \sim 5 \text{ k}\Omega$. In contrast, the NFADs exhibit a much slower linear current increase above breakdown due to the large load resistance. Device type E1G5 has a slope consistent with a 600 $\text{k}\Omega$ resistance, while device type E4G8 exhibits a behavior characterized by a substantially higher resistance of 3000 $\text{k}\Omega$. The much lower resistances of the 1st iteration NFADs were generally too low to provide adequately rapid quenching except for very modest excess bias voltages of ~ 1 V or

less. However, while not appropriate for high performance photon counting operation, these devices exhibited very interesting phenomena related to the flow of avalanche current that will be discussed below.

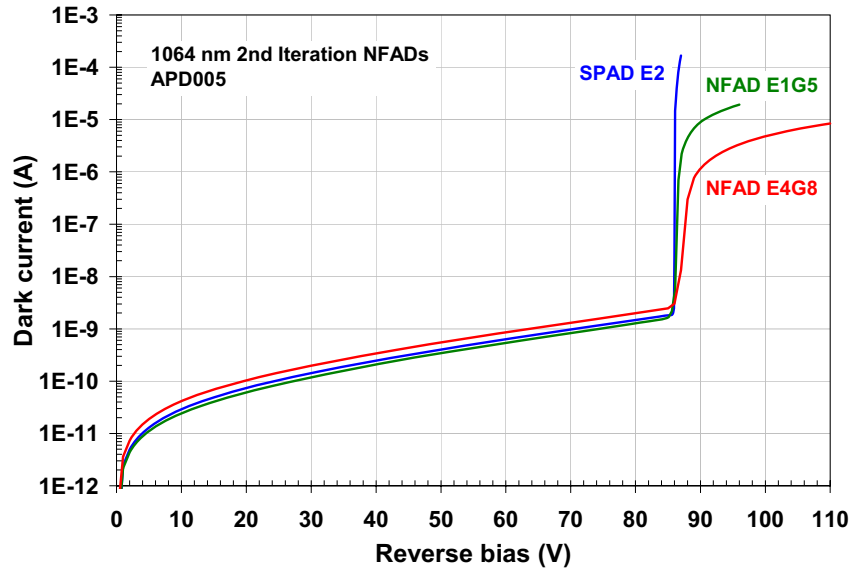


Figure 2. Reverse-bias current-voltage characteristics for a reference SPAD device with structure E2 (blue), an NFAD device E1G5 with feedback resistance $\sim 600 \text{ k}\Omega$ (green), and an NFAD device E4G8 with feedback resistance $\sim 2800 \text{ k}\Omega$ (red).

As mentioned above, NFAD operation is enabled with fairly simple support electronics. In Figure 3, we show a schematic of the measurement set-up used for device characterization. A bias tee is used to apply a fixed DC bias to the NFAD and extract high frequency pulses generated by the device with each avalanche event. Pulses were amplified with 24 dB of electrical gain, and the amplified signal was captured using either a digital oscilloscope or a data acquisition board. Using the digital oscilloscope, we were able to obtain detailed information about single pulse responses using single screen captures. To obtain data sufficient for statistical analysis based on long strings of pulses, we implemented a data acquisition board with very deep memory (255 Mb) that allows for data recording for up to 250 ms with 1 ns resolution.

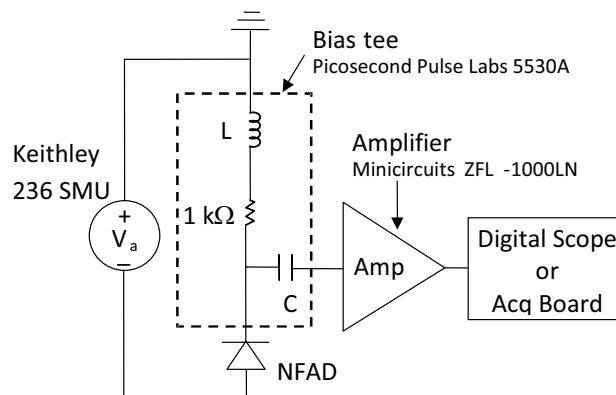


Figure 3. Schematic diagram of NFAD measurement test set-up employing a bias tee and low noise amplifier to record NFAD pulse behavior using a digital oscilloscope or a data acquisition board.

In Figure 3, we illustrate a typical data subset for 4 μ s of data containing several pulses from a 42 μ m active area diameter device of type E4G8 at 230 K and 80 V (excess bias \sim 2 V). I-V data for this device indicates a feedback resistance of 3.1 M Ω . Figure 4 illustrates the details of single dark count pulses from two device types E4G8 and E1G5. Although these two devices had rather different properties (see Table 1), their pulse response behavior is similar. The main portion of the pulse consists of a very narrow peak (FWHM $<$ 2 ns) with a voltage swing of about 45 mV, and the tail of this peak indicates a current decay lasting about 20 ns. Both devices were biased at \sim 2 V excess bias, and measurements were made at 230 K.

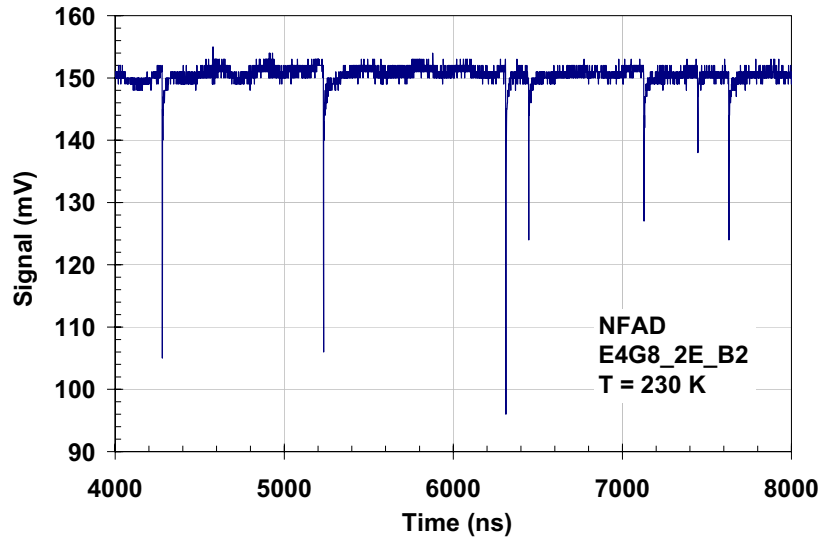


Figure 3. Pulse response from dark counts of NFAD device type E4G8 with feedback resistance of 3.1 M Ω at 230 K and 80 V bias (excess bias \sim 2 V). Data is acquired using a readout board with deep memory (255 Mb) and 1 ns time resolution.

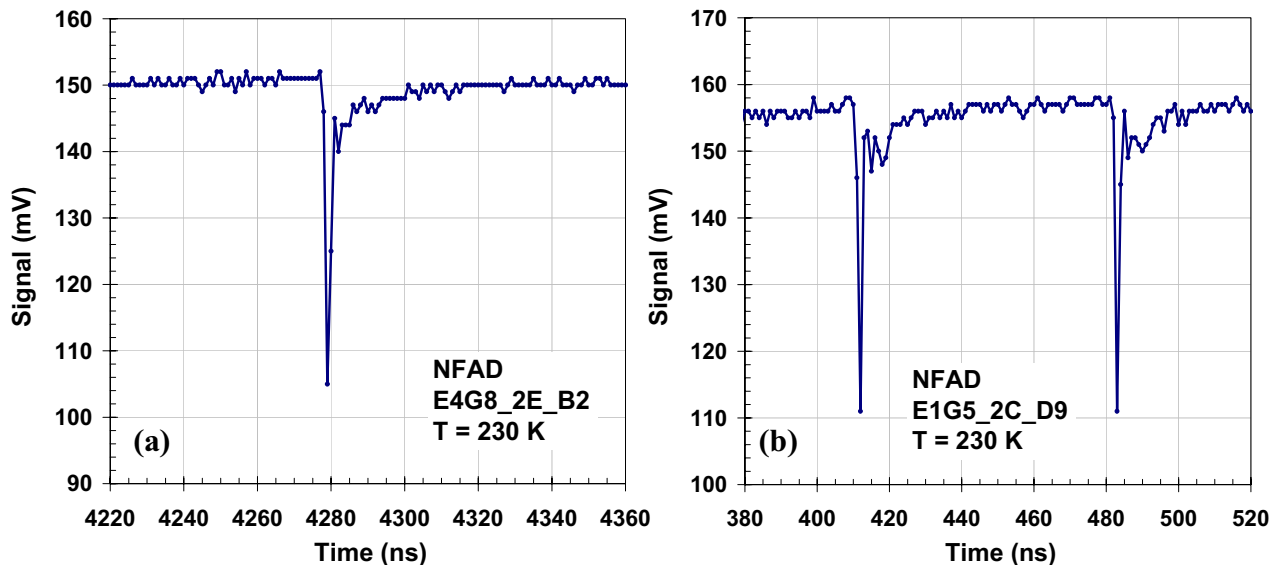


Figure 4. Dark count pulse response from NFADs at 230 K and \sim 2 V excess bias. (a) Device type E4G8 with 3.1 M Ω feedback resistance. (b) Device type E1G5 with 820 k Ω feedback resistance.

The very narrow initial peaks in the pulse responses seen in Figure 4 indicate very rapid quenching that can be understood by considering the shifting of excess bias to the load resistance as avalanche current builds up. The 3 MΩ load resistance of E4G8 requires an instantaneous current of only 0.6 μA to shift ~2 V excess bias to the load resistor. This current amplitude is at least an order of magnitude lower than the “quench current” i_q — determined to be about 10 μA from our 1st iteration devices — below which avalanche quenching is expected to be very rapid. Even for the lower feedback resistance of ~800 kΩ for E1G5, the current required to shift the excess bias of the diode is still well below I_q and promotes rapid quenching. The avalanche process can generate currents on the order of 1 μA in very short times — on the order of nanoseconds — and as long as these currents are sufficient to promote rapid quenching, the pulse response is rapid.

4. NFAD BEHAVIOR WITH SMALL NEGATIVE FEEDBACK

As described in the previous section, a critical aspect of the avalanche quenching process in an NFAD — or in any passive quenching scheme — is that the instantaneous current flow must be reduced below a threshold “quenching” current i_q below which rapid spontaneous quenching is probable.[23,16] The occurrence of spontaneous quenching is intimately related to how many carriers are present in the junction at a given time T_0 . The probability of quenching by some later time $T > T_0$ is given by the joint probability that the sequence of multiplication events for *every* individual carrier in the avalanche region at T_0 dies out by time T . [24] For current flow on the order of i_q , a persistent current may flow for an appreciable period of time before spontaneous quenching occurs.

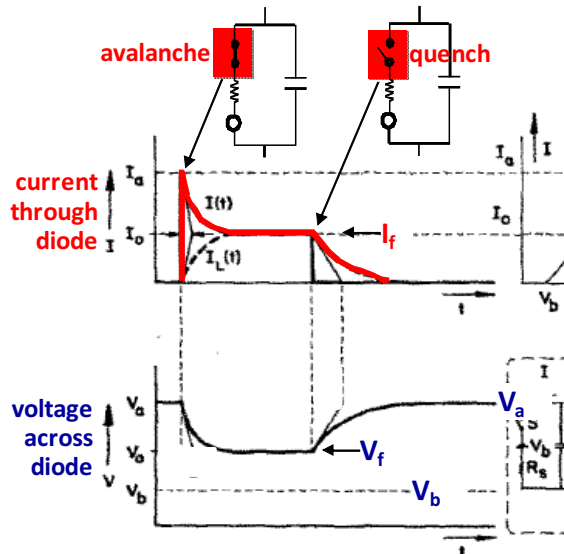


Figure 4. Depiction of time variation of current and voltage in an NFAD. I_f and V_f indicate final steady-state current and voltage, respectively, prior to spontaneous quenching.

A depiction of the time variation of current and voltage in the NFAD is presented in Figure 4. The onset of an avalanche — associated with the closing of switch S, as in Figure 1 — is accompanied by an extremely rapid build-up of avalanche current from the discharge of the diode capacitance. As the capacitive discharge proceeds, voltage across the diode structure decreases (with an equivalent increase in voltage across the load, not illustrated in this figure) while the current flow through the diode switch also decreases. The time-dependent voltage across the diode decays to a final steady-state value V_f dictated by the dividing of the excess bias voltage across the load resistance R_L and the diode dynamic resistance R_d (see Figure 1). The corresponding final steady-state current I_f attains a value that is consistent with V_f and R_L . If $I_f \ll I_q$, then the avalanche can rapidly quench in a time frame much shorter than the time constant for the decay to the steady-state value. If $I_f \sim I_q$, then the time frame in which spontaneous quenching will occur can vary dramatically. Finally, if $I_f \gg I_q$, then the probability of spontaneous quenching will be negligible, and a persistent current will flow until the external bias on the device is removed.

With our 1st iteration NFADs, many devices had small feedback resistances that were only marginally sufficient for inducing avalanche quenching. As the excess bias is increased, avalanche current amplitudes increase, and it is possible to operate these devices in a regime for which $I_f \sim I_q$. In Figure 5, we illustrate a dark count pulse response for device D3F7 device taken at an excess bias of ~1.7 V. Following the initial large amplitude peak in the pulse response, there is persistent current flow for ~120 ns at a voltage level on the order of 15 to 20 mV, after which the current flow spontaneously quenches. (Although the final signal obtained from our test set is a voltage from the amplifier, as shown in Figure 3, this signal is actually proportional to the current flow through the NFAD.) Aside from the inversion of the signal, this behavior is qualitatively consistent with the depiction described in Figure 4.

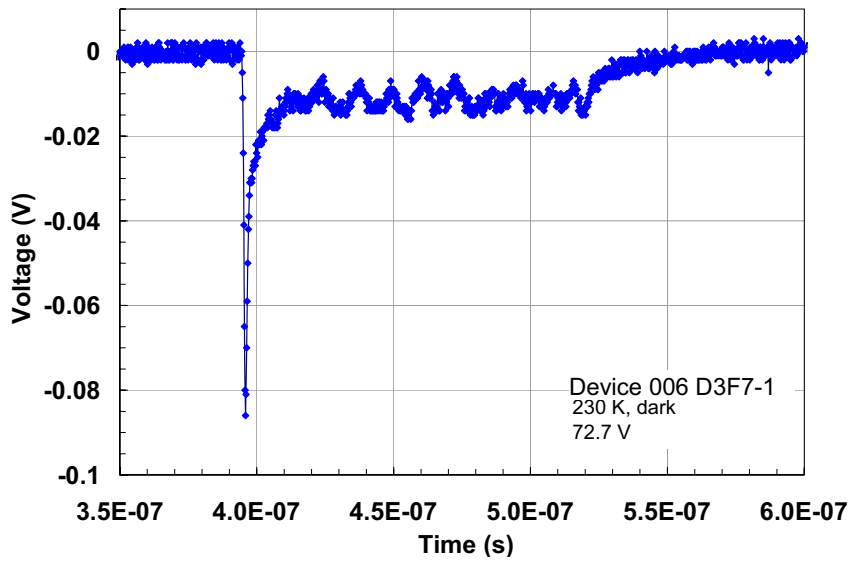


Figure 5. NFAD behavior for an excess bias of ~ 1.7 V for device type D3F7. This particular pulse response shows a rather long period (~ 120 ns) of persistent current flow before spontaneous quenching occurs. The periodic structure found during persistent current flow is categorically different from the noise observed in the absence of current flow (near $V_{\text{sig}} = 0$).

Also evident in the persistent current shown in Figure 5 is pronounced oscillatory behavior of the response signal during this period of persistent current flow, with a strong component at a periodicity of ~ 10 ns. We have seen this phenomenon consistently for every NFAD that exhibits persistent current effects, and we believe that these oscillations are fundamentally related to the carrier dynamics involved with the marginal avalanche that occurs in maintaining the persistent current. A more in-depth model involving the carrier dynamics that give rise to this behavior is reported elsewhere. [25]

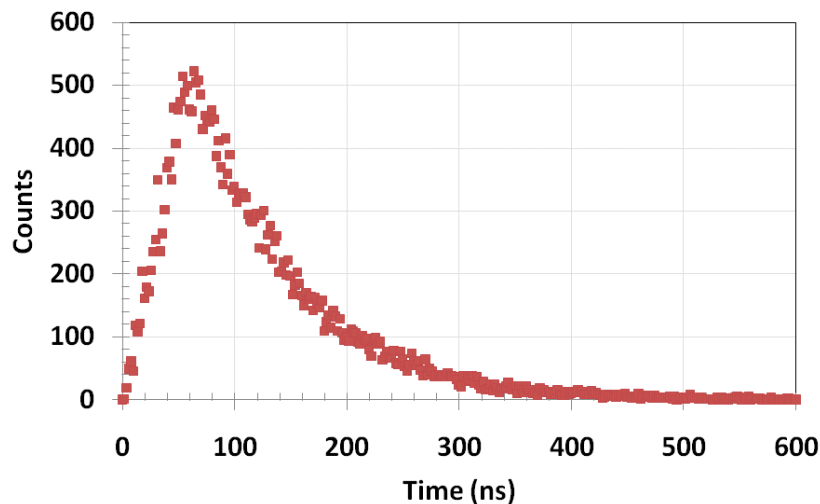


Figure 6. Distribution of persistent current durations for NFAD device type D3F7 at an excess bias of ~ 2.7 V. The distribution is peaked at a duration of ~ 60 ns, with linear decrease for shorter durations, and an exponential fall-off of the probability for increasing duration. Count values have been obtained by binning 1-ns resolution raw data into 2 ns time bins.

In this operating regime with significant persistent current effects, it is also interesting to analyze the probability distribution for the duration of persistent current flow prior to spontaneous quenching. Based on the collection of tens of thousands of pulses, we plot the distribution of persistent current durations in Figure 6 for an NFAD of type D3F7 operated at an excess bias of ~ 2.7 V. This distribution peaks at a value of ~ 60 ns, and there is a linear decrease in probability from this peak to zero for shorter durations. For longer durations, the distribution can be fit quite precisely by an exponential decay. This exponential roll-off is significant since it is predicted [25] by the carrier dynamics modeling used to explain the persistent current oscillatory behavior.

Although there is no practical utility to operating the NFAD when persistent currents are prominent — in fact, this large current flow is highly undesirable since it exacerbates afterpulsing and optical crosstalk — there is still value in studying the device in this limit to confirm fundamental aspects of the device behavior. The oscillatory behavior of the persistent current, as well as the distribution of its duration, comprise experimental results that should be reproduced to confirm the adequacy of a detailed theoretical model of NFAD properties. Other more qualitative NFAD properties can also be verified. For instance, with our present understanding of the persistent current, one concludes that it is not possible for the NFAD to exhibit a detection event (i.e., an avalanche peak) while the persistent current is flowing. In this state, the internal voltage across the diode is essentially equivalent to the breakdown voltage, and there is no excess charge residing on the diode capacitance to give rise to a detectable current pulse. Experimentally, we find this to be the case.

5. NFAD BEHAVIOR WITH LARGE NEGATIVE FEEDBACK

With the fabrication of our 2nd iteration NFADs, we achieved significantly larger feedback resistance than obtained for 1st iteration devices (see Table 1) and consequently eliminated any evidence of persistent current effects. These devices exhibit rapid quenching, as shown in Figure 3 and Figure 4. In this section, we present a comparison of 1st and 2nd iteration results to demonstrate their qualitatively different behavior. From long strings of pulse data obtained over 250 ms, we report distribution data for avalanche pulse height and avalanche charge flow for devices from both iterations. D3F7 is a 1st iteration device with a reasonably large 58 μm active region diameter and a feedback resistance of about 90 k Ω , while E4F8 is the largest of our 2nd iteration devices, with a 42 μm active region diameter and a feedback resistance of ~ 3.0 M Ω . The main point of the data that follow is to show that while the behavior of D3F7 is consistent with canonical passive quenching, E4G8 exhibits unexpected properties that are not anticipated with typical passive quenching circuits.

5.1 Distribution of pulse heights

Figure 7 shows the distribution of avalanche pulse amplitudes for D3F7 and E4G8 at 230 K. In Figure 7(a), the results for D3F7 show that a change in bias voltage from 71.9 V (blue squares) to 72.6 V (red circles) results in a significant increase in pulse height distribution. The shift in the distribution peak from 37 mV to 56 mV is roughly proportional to the change in excess bias voltage. To first order, one would assume that the pulse amplitude would scale with V_{ex}/R_d , where V_{ex} is the excess bias and R_d is the dynamic resistance of the NFAD, and the data in Figure 7(a) tends to confirm this expectation. However, in Figure 7(b), the results for E4G8 show that a very substantial 2.5 V change in bias voltage from 78.0 V (blue squares) to 80.5 V (red circles) results in almost no shift in the distribution. To obtain the same peak in the pulse height distribution after an increase in V_{ex} of 2.5 V is a significant departure from the behavior expected for typical passive quenching.

5.2 NFAD avalanche charge

An analysis of the total charge flow per avalanche for devices D3F7 and E4G8 shows a behavior similar to that just shown for the avalanche pulse height distributions. Figure 8 shows the distribution of avalanche charge flow for D3F7 and E4G8 at 230 K. In Figure 8(a), the results for D3F7 show that a change in bias voltages from 71.9 V (blue squares) to 72.6 V (red circles) results in an increase in avalanche charge flow from 2.5×10^6 e to 6.3×10^6 e. However, in Figure 8(b), the results for E4G8 show that the substantial 2.5 V change in bias voltage from 78.0 V (blue squares) to 80.5 V (red circles) results in no shift in the charge flow distribution, which has a peak value of 1.26×10^6 independent of bias voltage. The smaller 2nd iteration device E1G5 provides a significantly lower charge flow per

avalanche of $\sim 3 \times 10^5 e$, as shown in Figure 9, but exhibits the same independence of the total charge flow on excess bias voltage.

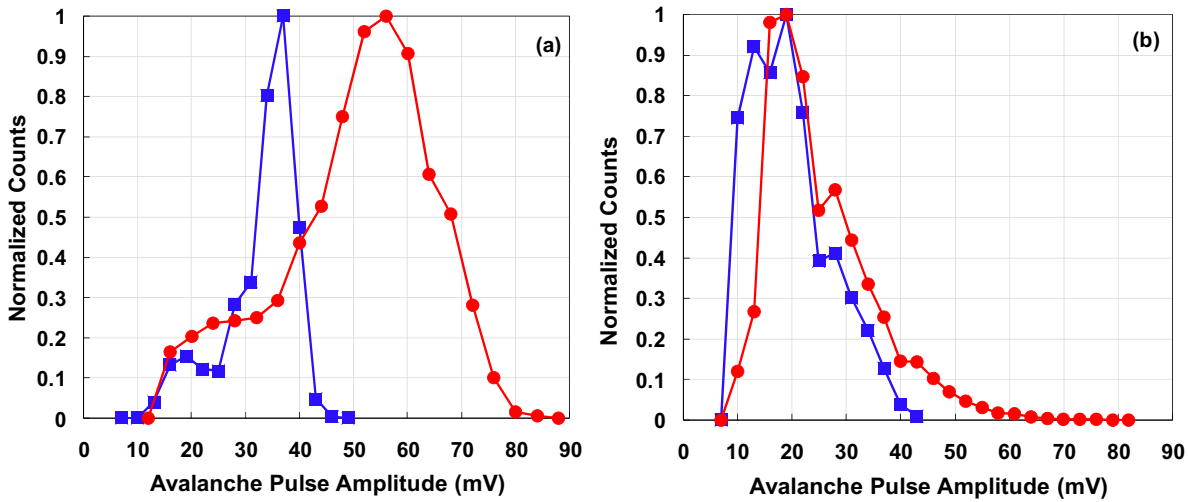


Figure 7. Distributions of avalanche pulse amplitudes for (a) 1st iteration device D3F7 and (b) 2nd iteration device E4G8, both at $T = 230$ K. (a) Distributions obtained with D3F7 for bias voltages of 71.9 V (blue squares) and 72.6 V (red circles). (b) Distributions obtained with E4G8 for bias voltages of 78.0 V (blue squares) and 80.5 V (red circles).

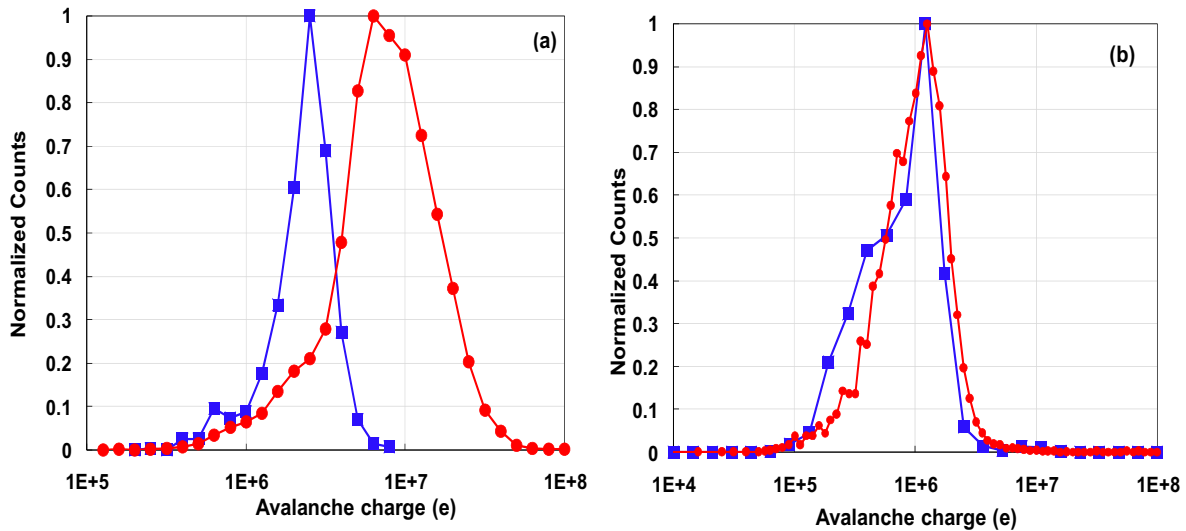


Figure 8. Distributions of charge flow per avalanche for (a) 1st iteration device D3F7 and (b) 2nd iteration device E4G8, both at $T = 230$ K. (a) Distributions obtained with D3F7 for bias voltages of 71.9 V (blue squares) and 72.6 V (red circles). (b) Distributions obtained with E4G8 for bias voltages of 78.0 V (blue squares) and 80.5 V (red circles).

The magnitude of the charge flow per avalanche can be roughly estimated by considering the fact that μA -scale currents across a $M\Omega$ -scale load resistance are sufficient to remove the excess bias and promote rapid quenching. Each $1 \mu A$ of current that flows for 10 ns corresponds to a charge flow of $(1.6 \times 10^{19} e/A)(10^{-6} A)(10^{-8} s) = 1.6 \times 10^5$ carriers. A full discharge of the diode capacitance in removing the excess bias would involve the full capacitance of the diode structure, i.e., $Q = C_d V_{ex}$, and for device E1G5 used to obtain the data in Figure 9, $(80 \text{ fF})(2 \text{ V}) \sim 1 \times 10^6$ carriers. We believe that, with such a large monolithic feedback resistance with very low parasitics, we may be limiting the current to values below the quench current I_q sufficiently rapidly that *only a portion of the total diode capacitance is involved*

in the discharge event. The fact that the apparent charge per avalanche of $\sim 3 \times 10^5$ carriers shown in Figure 9 is 3 to 4 times smaller than estimates based on $Q = C_d V_{ex}$ is consistent with partial capacitive discharge. Additionally, the assumption of complete discharge of the device capacitance with each avalanche leads to an expectation that the charge flow per avalanche should scale proportionally with the excess bias, and we find this not to be the case, as seen in Figure 8(b) and Figure 9.

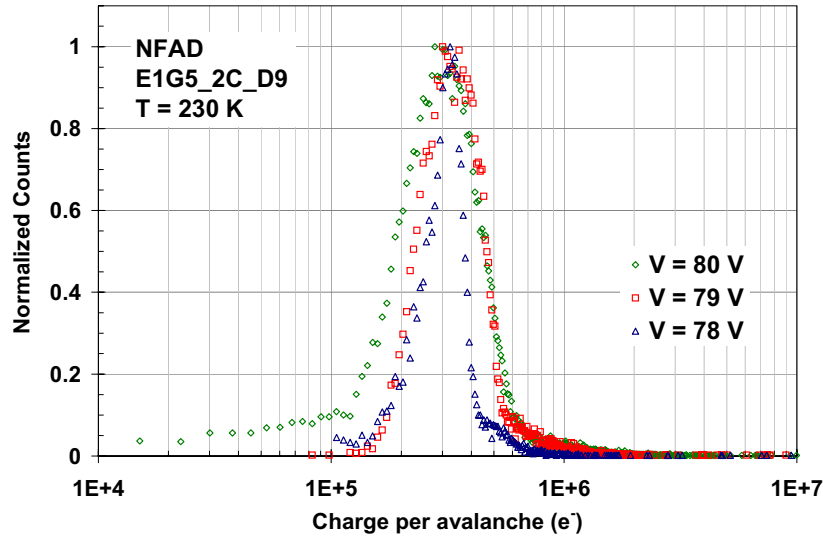


Figure 9. Distribution of charge per avalanche for NFAD E1G5 obtained by integrating over pulse responses. Distributions are presented for three different values of the bias voltage, with breakdown occurring close to 78 V. The bias-independent distributions are peaked at a low value of only $\sim 3 \times 10^5$ carriers per avalanche event.

5. CONCLUSION

To surmount the present obstacles to higher counting repetition rate and photon number resolution using SPADs, we have introduced negative feedback to the conventional SPAD avalanche process to beneficially modify the device avalanche dynamics. The goal of this approach is to achieve extremely consistent passive quenching by the monolithic integration of a suitable negative feedback element, e.g., a suitable thin film resistor in series with the SPAD structure. As we have shown, an appropriate implementation of this concept can lead to rather small avalanches (e.g., $\sim 10^5 - 10^6$ carriers), for which one expects reduced carrier trapping and an associated reduction in afterpulsing effects. For arrayed geometries of NFAD active regions with all anodes connected and all cathodes connected, the single output signal from the array can exhibit quantized current pulses whose amplitudes indicate the number of detectors that have avalanched. Even if one cell of this array experiences some “dead time” following a detection event, remaining cells will still be armed. In this way, multiplexed structures can support not only the resolution of simultaneous photon detection events but also higher counting rates.

ACKNOWLEDGMENTS

We are grateful to Bill Farr and Michael Krainak for valuable discussions concerning the negative feedback avalanche diode concept and its characterization. We acknowledge partial support for this work from JPL and NASA/GSFC.

REFERENCES

- [1] Special Issue on “Free-space laser communications”, *IEEE LEOS Newsletter*, vol. 19 No. 5 (2005).
- [2] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, *Rev. Mod. Phys.*, vol. 74, p. 145 – 195 (2002).
- [3] R. M. Measures, *Laser Remote Sensing - Fundamentals and Applications*, John Wiley & Sons (1984).
- [4] N. Namekata, S. Sasamori, and S. Inoue, *Opt. Express*, **14**, 10043 (2006).
- [5] Z.L. Yuan, B.E. Kardynal, A.W. Sharpe, and A.J. Shields, *Appl. Phys. Lett.* **91**, 041114 (2007).
- [6] D. Bisello, et al., *Nuclear Instruments and Methods in Physics Research A* **360**, p. 83 – 86 (1995); *ibid.* **367**, p. 212 – 214 (1995).
- [7] S. Afanasiev, et al., *Nuclear Physics B* **44**, p. 402 – 405 (1995).
- [8] D. Shushakov and V. Shubin, *Proc. of SPIE* **2397**, p. 544 – 554 (1995); V. Shubin and D. Shushakov, *Proc. of SPIE* **2415**, p. 94 – 103 (1995).
- [9] D. Shushakov and V. Shubin, *Proc. of SPIE* **2699**, p. 173 – 183 (1996).
- [10] G. Bondarenko, et al., *Nucl. Phys. B* **61B**, p. 347 – 352 (1998).
- [11] P. Buzhan, et al., ICFA Instrumentation Bulletin (Fall 2001) [<http://www.slac.stanford.edu/pubs/icfa/fall01.html>]
- [12] <http://sales.hamamatsu.com/en/products/solid-state-division/si-photodiode-series/mppc.php>
- [13] <http://sensl.com/products/silicon-photomultipliers/>
- [14] K. Zhao, A. Zhang, Y.-h. Lo, and W. Farr, *Appl. Phys. Lett.* **91**, 081107 (2007); K. Zhao, S. You, J. Cheng, and Y.-h. Lo, *Appl. Phys. Lett.* **93**, 153504 (2008).
- [15] K. Linga, Y. Yevtukhov, and B. Liang, “Very high gain and low excess noise near infrared single photon avalanche detector: A NIR solid state photomultiplier,” *Proc. of SPIE* **7320**, 7320-0Z (2009).
- [16] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, *Appl. Opt.* **35**, p. 1956 – 1976 (1996).
- [17] X. Jiang, M. A. Itzler, B. Nyman, and K. Slomkowski, “Negative Feedback Avalanche Diodes for Near-infrared Single Photon Detection,” *Proc. of SPIE* **7320**, 7320-11 (2009).
- [18] M. A. Itzler, R. Ben-Michael, C.-F. Hsu, K. Slomkowski, A. Tosi, S. Cova, F. Zappa, R. Ispasoiu, *J. Modern Optics* **54**, no. 2-3, p. 283 – 304 (2007).
- [19] X. Jiang, M. A. Itzler, R. Ben-Michael, and K. Slomkowski, *IEEE J. of Sel. Topics in Quantum Electronics* **13**, p. 895 – 905 (2007).
- [20] D.S. Bethune, W.P. Risk, and G.W. Pabst, *J. Modern Optics* **51**, no. 9-10, p. 1359 – 1368 (2004).
- [21] M. A. Itzler, X. Jiang, B. Nyman, and K. Slomkowski, *Proc. of SPIE* **7222**, 72221K (2009).
- [22] A. Tosi, A. Dalla Mora, F. Zappa, S. Cova, M. A. Itzler, and X. Jiang, *Proc. of SPIE* **7222**, 72221G (2009).
- [23] R. H. Haitz, *J. Appl. Phys.* **35**, p. 1370 – 1376 (1964).
- [24] M. M. Hayat, G. J. Rees, D. A. Ramirez, and M. A. Itzler, *2008 IEEE LEOS Annual Meeting Conference Proceedings*, p. 230 – 231 (2008).
- [25] M. M. Hayat, M. A. Itzler, D. A. Ramirez, and G. J. Rees, “Model for passive quenching of SPADs,” *Proc. of SPIE* **7608**, 7608-83, these proceedings (2009).