Journal of Modern Optics
Vol. 59, No. 17, 10 October 2012, 1472–1480

Taylor & Francis
Taylor & Francis Group

Power law temporal dependence of InGaAs/InP SPAD afterpulsing

Mark A. Itzler*, Xudong Jiang and Mark Entwistle

Princeton Lightwave Inc., 2555 US Route 130S., Cranbury, NJ 08512, USA

(*Received 18 February 2012; final version received 25 May 2012*)

The characterization and analysis of afterpulsing behavior in InGaAs/InP single photon avalanche diodes (SPADs) is reported for gating frequencies between 10 and 50 MHz. Gating in this frequency range was accomplished using a matched delay line technique to achieve parasitic transient cancellation, and FPGA-based data acquisition firmware was implemented to provide an efficient, flexible multiple-gate sequencing methodology for obtaining the dependence of afterpulse probability P_{ap} on hold-off time T_{ho}. We show that the detrapping times extracted from the canonical exponential fitting of $P_{ap}(T_{ho})$ have no physical significance, and we propose an alternative description of the measured data, which is accurately fit with the simple power law behavior $P_{ap} \propto T_{ho}^{-\alpha}$ with $\alpha \sim 1.2 \pm 0.2$. We discuss the physical implications of this functional form, including what it may indicate about trap defect distributions and other possible origins of this power law behavior.

Keywords: single photon avalanche diode; SPAD; photon counting; Geiger mode; InGaAs/InP; afterpulsing

1. Introduction

There has been substantial progress in the past decade in the performance and implementation of single photon avalanche diodes (SPADs) based on the InGaAsP material system for the detection of photons in the wavelength range 0.9–1.7 μm [1,2]. The fundamental tradeoff between photon detection efficiency (PDE) and dark count rate (DCR) has reached a level of performance that is sufficient for many applications (see, for example, [3]) with convenient and cost-effective packaging employing thermoelectric cooling. The timing accuracy of these detectors has also been demonstrated to be highly precise, with timing jitter of 100 ps or less for typical operating conditions [1,4]. At present, the most problematic issue pertaining to many applications for InP-based SPADs is the limitation in counting rate posed by the phenomenon known as afterpulsing.

A SPAD is operated by applying a bias voltage to the device that exceeds its avalanche breakdown voltage V_b. In this metastable state, a single photoexcited carrier entering the high-field multiplication region of the device can generate an avalanche of carriers by impact ionization. This mode of operation, known as the Geiger mode, results in a macroscopic pulse of charge that is easily detectable by appropriate electronic threshold circuitry following the detector. This process provides a simple, effective means for detecting single photons. However, if charge carriers created during the avalanche process become trapped at atomic defect sites in the multiplication region, the subsequent detrapping of these carriers at a later time can trigger spurious additional avalanches known as afterpulses. Afterpulses are a type of dark count, but unlike other dark counts mechanisms – such as thermal excitations or tunneling effects – that occur randomly in time, afterpulses are strongly correlated to previous avalanches during which trap sites were populated. The most straightforward mitigation of afterpulsing effects is to force the SPAD to be inactive by biasing it below V_b for a 'hold-off' time that is sufficiently long to allow all trapped carriers to be detrapped and swept from the multiplication region without the possibility of triggering afterpulses. However, imposing long hold-off times limits the rate at which photons can be detected.

During the past several years, the most prevalent strategy employed to mitigate afterpulsing has been to limit the amount of charge flow that occurs during each avalanche. The rationale has been that smaller avalanches with less charge flow lead to less carrier trapping and consequently reduced afterpulsing. The most successful implementation of this afterpulsing reduction strategy has been the use of gated quenching with very short (~100–200 ps) gate durations. Using either sine-wave gating [5–7] or self-differencing

*Corresponding author. Email: mitzler@princetonlightwave.com

ISSN 0950–0340 print/ISSN 1362–3044 online

http://dx.doi.org/10.1080/09500340.2012.698659
http://www.tandfonline.com

techniques [8,9] to eliminate the transient effects inherent in high-frequency gate modulation, this approach has been used to achieve impressive GHz-rate gated operation of InGaAs/InP SPADs.

However, these recent advances in afterpulse reduction have been restricted to periodic gating with very short gates with photon arrivals synchronized to the gating frequency. More generally, it would be highly desirable to have asynchronous, or 'free-running', operation with very fast (ns-scale) quenching and reset following each photon detection. Among recent studies, there have been reports of free-running behavior achieved using hybrid passive quenching with low bias operation [10] and device structures incorporating monolithically integrated negative feedback to provide avalanche self-quenching [11,12]. Nevertheless, to date these implementations have been constrained to rather modest values of PDE by the onset of afterpulsing at higher bias. Beyond clever circuit approaches such as the short-gating techniques, further progress towards resolving afterpulsing limitations will require a better understanding of the microscopic origins of this phenomenon. The predominant device design for InGaAsP-based SPADs used today [2,13,14] employs an InP multiplication region, and very little is known about the trap sites that might be responsible for afterpulsing effects in these devices.

With the objective of improving our understanding of the nature of the defects that give rise to afterpulsing in InGaAs/InP SPADs, we have extended a gating scheme based on matched delay line transient cancellation to characterize the dependence of afterpulsing probability P_{ap} on hold-off time T_{ho} at frequencies between 10 and 50 MHz. Additionally, we have implemented field-programmable gate array (FPGA) control of this circuitry to generalize the standard time-correlated carrier counting [15] (or 'double-pulse') method for afterpulse measurements. We show that the canonical approach to fitting such measured data – in which it is assumed that one or just a few defects with exponential carrier detrapping dictate the afterpulsing behavior – does not yield physically meaningful results for detrapping time constants. Instead, we show that our data, as well as similar afterpulsing data obtained by other groups, is accurately described by a power law $P_{ap} \propto T_{ho}^{-\alpha}$ with $\alpha \sim 1.2 \pm 0.2$. We then propose a simple model for calculating the dependence $P_{ap}(T_{ho})$ based on the assumption of a defect population with a distribution of detrapping rates, and we show that the experimental observed power law behavior is consistent only with specific types of defect distributions. Finally, we discuss other possible explanations for the observed power law dependence of P_{ap} on T_{ho}.

2. 50 MHz gating with FPGA-based data acquisition

2.1. Matched delay line transient cancellation

To perform the measurements described in this paper, we have adopted a short-gate technique first developed by Bethune and Risk [16,17] in which two matched coaxial delay lines – one inverting and the other non-inverting – are used to cancel the parasitic transient responses resulting from the rapid swing in bias voltage across the capacitance of the SPAD at the beginning and end of the gate period. The SPAD bias control circuitry applies periodic excess bias gates consisting of a fixed voltage swing $\Delta V \sim 4\,V$ with rise and fall times of $\sim$0.1 ns and a gate plateau ranging from $\sim$0.5 to 5 ns. The voltage swing ΔV is added to a dc bias level $V_{dc} < V_{b,}$ and the excess bias V_{ex} is set by adjusting the dc bias level so that $V_{ex} = V_{dc} + \Delta V - V_b$. Earlier implementations of this scheme [16,18] were limited to $\sim$5 MHz due to sizable avalanche pulses giving rise to excessive afterpulsing at frequencies above this range. Through the implementation of shorter delay lines, more precise design of the hybrid circuitry used for transient cancellation, and the reduction of other circuit parasitics, we have achieved good performance for gating frequencies up to 50 MHz, with afterpulsing of $\sim$2–3% for 10% PDE using 1 ns duration gates [2].

2.2. Multiple-gate sequencing as extension of time-correlated carrier counting

The most widely used technique for characterizing SPAD afterpulsing effects is the time-correlated carrier counting method [15], also referred to as the 'double-pulse' method. The basic scheme is to apply a first gating pulse at time T_1 to arm the SPAD and induce an avalanche during this gate by illuminating the SPAD with a coincident optical pulse. A second gating pulse is then applied after a hold-off time T_{ho} at a later time $T_2 = T_1 + T_{ho}$ without illumination to observe if a count occurs at T_2. Sufficient repetitions of this double-pulse sequence provide a statistical measure of the increase in the count rate at T_2 above the intrinsic DCR due to afterpulses correlated to the earlier avalanches occurring at T_1. This increase in DCR at T_2 is a quantitative measure of the afterpulsing probability $P_{ap,}$ and by systematically varying $T_{ho,}$ one obtains the relation $P_{ap}(T_{ho})$.

The double-pulse method is effective for characterizing the dependence of the afterpulse probability P_{ap} on hold-off time $T_{ho,}$ but it is time-consuming to step through successive values of $T_{ho,}$ and it does not measure higher order effects related to afterpulses initiated by previous afterpulses. Using FPGA-based firmware, we have generalized the double-pulse method to implement a more flexible scheme in

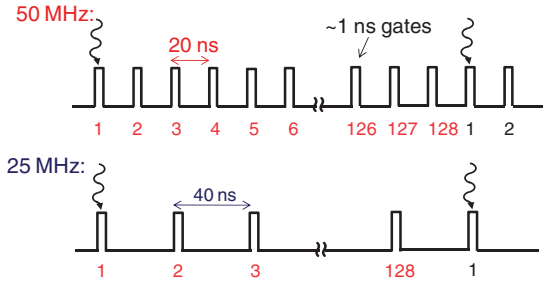

Figure 1. Illustrative examples of high frequency gating sequences implemented using firmware in field programmable gate arrays (FPGAs). Photon arrivals are synchronized with the first gate (labeled '1') in a programmable sequence of gates, and in this example, counts can be recorded in all of the gates 2–128 following photon detection at gate 1. (The color version of this figure is included in the online version of the journal.)

which each illuminated gate at T_1 is followed by an arbitrary number of additional gates at multiples of the minimum time separation ΔT between gates – i.e. $T_n = T_1 + n\Delta T$. We illustrate this gating scheme in Figure 1 with $n = 128$ for two gating frequencies of 50 MHz ($\Delta T = 20$ ns) and 25 MHz ($\Delta T = 40$ ns).

Using the FPGA-controlled multiple-gate sequencing with the improved matched delay line transient cancellation circuitry, we measured the afterpulse probability P_{ap} per 1-ns gate as a function of hold-off time $T_{ho,}$ as shown for four different gating frequencies in Figure 2. For each gating frequency f_g, points are obtained at timing intervals $\Delta T = 1/f_g$. Unlike the legacy double-pulse method, these data can include higher order afterpulsing effects in which one afterpulse induces a later afterpulse. The higher order effects can also be eliminated from the data set by considering only the earliest count among the T_n gates in each multiple-gate sequence and ignoring counts at later gates. For this particular data set, since the maximum values of P_{ap} are only ~1% even for the shortest hold-off time used (i.e. 20 ns for the 50 MHz gating in Figure 2(*d*)), higher order effects will have a probability of $\sim 1 \times 10^{-4}$ and can be neglected when considering afterpulsing probabilities on a per-gate basis. However, we do find a measurable dependence on gating frequency of the background DCR using the multiple-gate sequencing technique described above. Specifically, for 10 and 50 MHz operation, the measured DCR is 1.7×10^{-6} and 2.6×10^{-6} per 1-ns gate, respectively. Data obtained at 25 MHz and 40 MHz (not shown in Figure 2) accurately interpolate between the 10 and 50 MHz values; only the 33 MHz DCR value falls outside this trend, being somewhat lower than expected. These results suggest that higher afterpulsing at higher frequencies is increasing the apparent background DCR value by about 35% for an increase

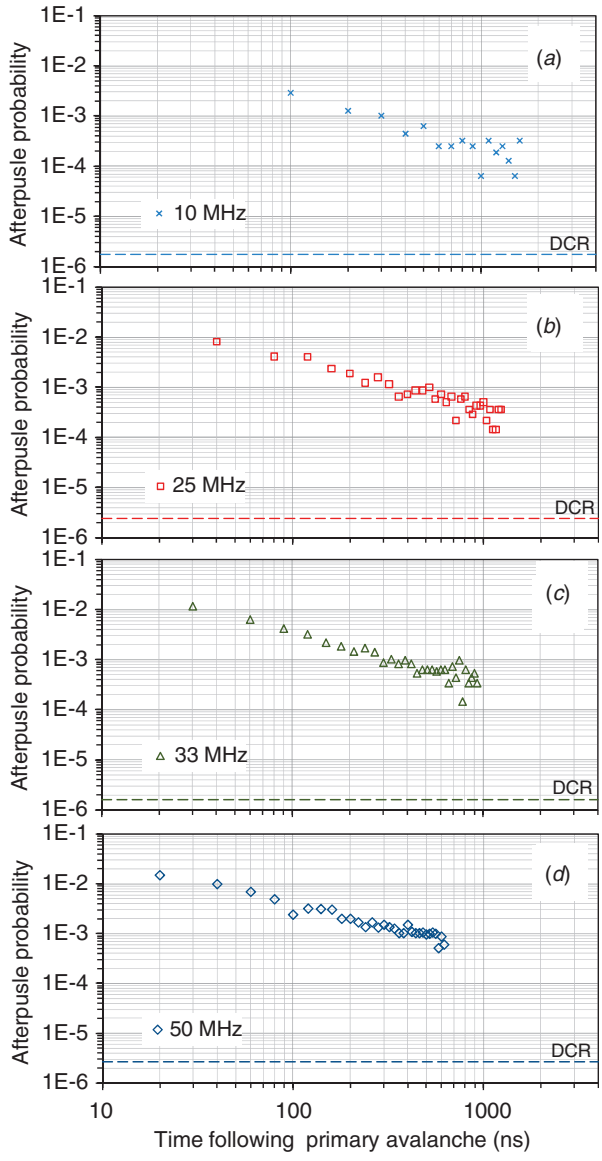

Figure 2. Measured results for afterpulse probability per 1 ns gate as a function of the hold-off time following a photon-induced primary avalanche with PDE = 20%. (*a*)–(*d*) correspond to gating frequencies of 10, 25, 33, and 50 MHz, respectively, implemented using FPGA firmware to establish the multiple-gate sequencing illustrated by Figure 1. Dashed lines indicate the background DCR measured at each frequency. All data were obtained at 223 K. (The color version of this figure is included in the online version of the journal.)

in frequency from 10 MHz to 50 MHz using this new technique.

The data in Figure 2 also show some variation with gating frequency in the temporal decay of the afterpulsing probability to the measured background DCR value. Using the fitting behavior described below in Section 3.2 to extrapolate the data in this figure, these decay times to reach the background DCR for the

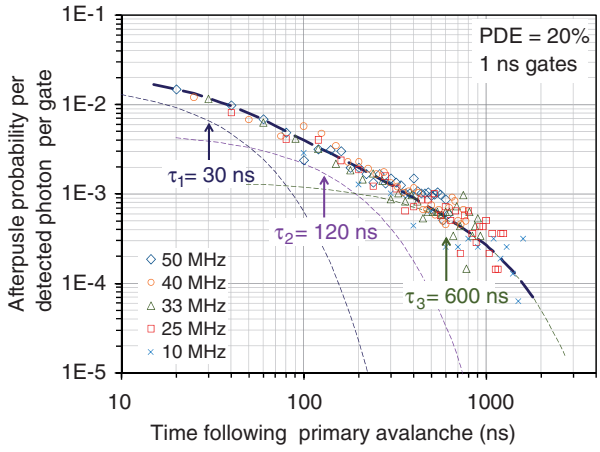

Figure 3. Symbols indicate measured afterpulse probability per 1 ns gate as a function of the hold-off time following a photon-induced primary avalanche for five different frequencies, as in Figure 2. Light dashed curves indicate piece-wise exponential fits (with indicated detrapping times τ_n), and the heavy dashed curve indicates the resulting fit given by the sum of the light dashed curve fits. (The color version of this figure is included in the online version of the journal.)

10 MHz and 50 MHz gating frequencies are estimated to be ~80 μs and ~300 μs, respectively. As with the slight increase in background DCR for higher frequency operation, we believe that higher order afterpulsing effects are responsible for the slower decay to the background DCR value for higher gating frequencies. It is also worth noting that these decay times are rather long because the background DCR values are so low.

3. Functional analysis of afterpulsing data

3.1. *Fitting based on legacy assumption of a few dominant traps*

The $P_{ap}(T_{ho})$ obtained for the different frequencies (as in Figure 2) all follow the same functional form, as can be seen by plotting the measured data for all frequencies on the same graph. In Figure 3, the symbols correspond to the measured data for frequencies of 10, 25, 33, 40, and 50 MHz.

Essentially, all attempts to model the behavior of $P_{ap}(T_{ho})$ in previous literature on SPADs start with the basic assumption that the defect traps which lead to afterpulsing are characterized by an exponential detrapping time τ. In the simplest case for which there is just one dominant type of trap in the SPAD multiplication region, the afterpulsing probability would be characterized by $P_{ap}(T_{ho}) = A_0 + A_1\exp(-T_{ho}/\tau)$, where A_0 and A_1 are constants. If a sufficiently narrow range of hold-off times is used, then a reasonable fit can be obtained using this single-exponential functional form. For instance, this analysis was applied in [19] to afterpulsing measurements with hold-off times in the range of 1 to 10 μs, for which $\tau \sim 0.9$ μs provided a best fit.

However, for all data sets obtained using wider ranges of hold-off time, the single-exponential form does not provide a good fit. It has then been common practice to assume that additional traps with different time constants are also present, and analysis proceeds with a generalized form

$$P_{ap}(t) = A_0 + A_1 e^{-T_{ho}/\tau_1} + A_2 e^{-T_{ho}/\tau_2} + A_3 e^{-T_{ho}/\tau_3} + \cdots, \quad (1)$$

where A_0 is the background dark count rate, and A_i and τ_i are the exponential pre-factor and detrapping time constant, respectively, associated with the ith trap type. In [20], such a procedure was used to fit afterpulsing data with hold-off times ranging from 1.25 to 100 μs by using three time constants ($\tau_1 \sim 0.5$ μs, $\tau_2 \sim 6.1$ μs, and $\tau_3 \sim 99$ μs). Similarly, in [21], hold-off times ranging from 0.02 to 50 μs were used, and fitting of the resulting afterpulsing data required four detrapping time constants ($\tau_1 \sim 0.07$ μs, $\tau_2 \sim 0.9$ μs, $\tau_3 \sim 4.2$ μs, and $\tau_4 \sim 33$ μs). For the measured $P_{ap}(T_{ho})$ data in Figure 3 spanning hold-off times ranges from 20 ns to 1.5 μs, we can follow a similar procedure to obtain the fit shown by the heavy dashed line. Obtaining a good fit over this range of hold-off times requires the use of three exponential terms indicated by light dashed curves, and the corresponding time constants $\tau_1 \sim 30$ ns, $\tau_2 \sim 120$ ns, and $\tau_3 \sim 600$ ns are indicated in the figure.

We now contend that in all of these cases, this exercise in fitting measured afterpulsing data with the minimum of exponentials required to obtain a good fit is entirely arbitrary and that the *extracted detrapping time constants have no physical significance*. Throughout essentially all of the literature in which this procedure is employed, each additional factor of ~5 to 10 in experimental hold-off times used to make the measurements requires the addition of another exponential term, yielding another time constant which supposedly indicates the prevalence of another trap type. However, it is evident from this whole fitting procedure that the *extracted values for the detrapping times depend entirely on the range of hold-off times used* in the data set. To emphasize this point, we repeat the exponential fitting process illustrated in Figure 4 for a subset of the data spanning the somewhat narrower hold-off time range of 60 ns to 900 ns. Fitting this subset of the same exact experimental data, the simplest form of Equation (1) to provide a good fit requires just two exponentials with time constants of

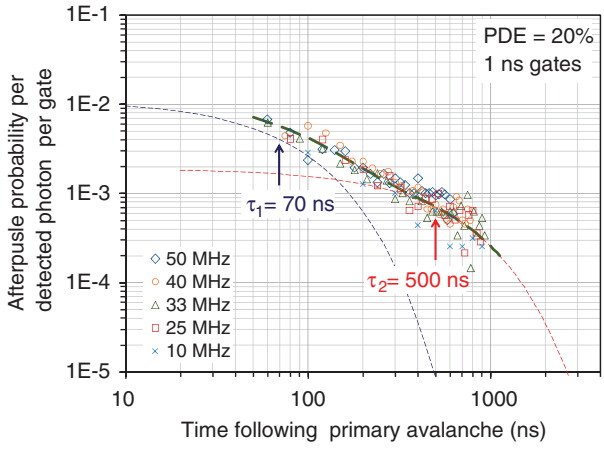

Figure 4. Exponential fitting as in Figure 3 to a subset of the measured results in Figure 3 from 60 ns to 900 ns. Although data points in this range are identical to those in Figure 3, the fitting result depends entirely on what subset of data points is used, and so the resulting detrapping values have no physical significance. (The color version of this figure is included in the online version of the journal.)

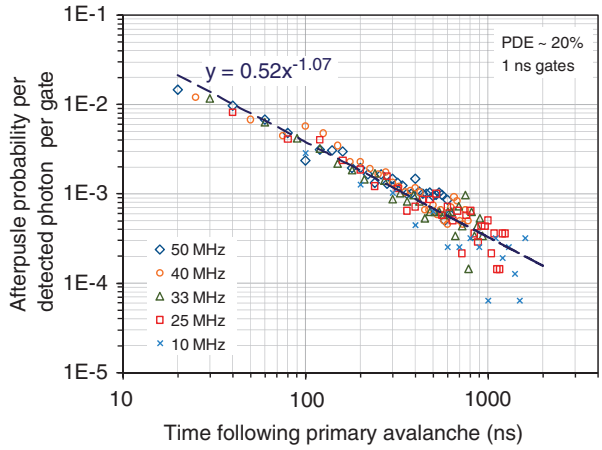

Figure 5. Same measured data (symbols) for $P_{ap}(T_{ho})$ as in Figure 4 with a power law fit of the functional form $P_{ap} = CT_{ho}^{-\alpha}$ (dashed line). (The color version of this figure is included in the online version of the journal.)

$\tau_1 \sim 70$ ns and $\tau_2 \sim 500$ ns. Clearly, these results bear no relationship to the values obtained earlier in Figure 3, and in neither case do the values obtained for τ_n have physical significance.

3.2. *Fitting based on power law behavior*

Although the exponential fitting procedure described above is motivated by reasonable physical expectations – i.e. the dominance of a small number of defect trap types – its arbitrary nature makes it impossible to ascribe physical significance to its output. We are then confronted with the question of whether an alternate fitting procedure can be found that is less arbitrary and perhaps provides more meaningful clues to the underlying physical origin of the afterpulsing behavior.

In Figure 5, we present the same experimental data plotted in Figure 4 and show that these data are described quite accurately over two orders of magnitude of T_{ho} by a simple power law (dashed line) of the form $P_{ap} = CT_{ho}^{-\alpha}$ where C and α are constants. The best fit to this data set is provided by $\alpha \sim 1.07$.

It is natural to then ask whether other afterpulsing data reported in the literature exhibits the same functional form. In Figure 6, we show measured data (symbols) originally reported by Restelli and Bienfang [22] at NIST in Gaithersburg, as well as results obtained by the group at the University of Virginia [23]. Over at least two orders of magnitude in hold-off time, a simple power law dependence (dashed lines) provides a very accurate fit to both sets of data, with

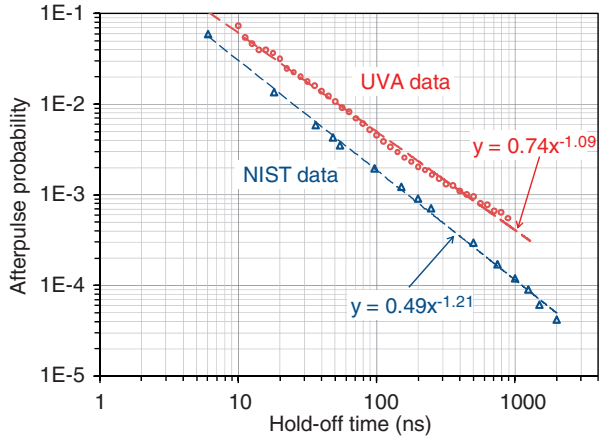

Figure 6. Measured afterpulsing dependence on hold-off time from NIST [19] and Univ. of Virginia [20] showing similar power law behavior to that of Figure 5. Both sets of data were taken using the double-pulse method with 1 ns gates. Operating temperatures for NIST and UVA data were 253 K and 230 K, respectively. (The color version of this figure is included in the online version of the journal.)

respective exponent values of $\alpha \sim 1.21$ and $\alpha \sim 1.09$. Data shown in this figure were acquired with 1 ns gating, but measurements incorporating a range of different gate widths (from 0.5 to 3 ns) yield similar values of α with a variation of no more than ± 0.06 relative to the 1 ns gating values. We also performed a similar analysis of afterpulsing data obtained from Namekata and Inoue [24] at Nihon University, and again, the data exhibited power law behavior, although with a somewhat larger exponent of $\alpha \sim 1.38$. Based on this analysis of $P_{ap}(T_{ho})$ data obtained from

InP/InGaAs SPADs by four different research groups, it appears that the afterpulsing dependence on hold-off time is well described by a power law behavior $P_{ap} \propto T_{ho}^{-\alpha}$ with $\alpha \sim 1.2 \pm 0.2$.

The power law dependence of P_{ap} on T_{ho} is immediately evident from the simple straight-line plot obtained when the data is graphically represented in a log–log format. It is interesting to note that much of the earlier data for $P_{ap}(T_{ho})$ reported in the SPAD literature has been plotted in a log–linear format, and so the power law functional form was not readily apparent.

4. Possible physical significance of afterpulsing power law behavior

Given the demonstrated power law behavior of the afterpulsing dependence on hold-off time, it is natural to ask whether this behavior provides detailed information concerning the nature of the defects giving rise to afterpulsing. The literature on deep level traps in InP (e.g. the survey by Anderson and Jiao [25]) shows that a rather dense spectrum of trap levels has been identified. These results suggest that instead of assuming just a few dominant trap levels, it may be more physically meaningful to develop a model that incorporates exponential detrapping from a broad distribution of trap levels.

We first define the inverse of the detrapping time constant τ to be the detrapping rate $R \equiv 1/\tau$ and then consider that the afterpulse probability is directly related to the temporal change in trap occupation N given by $\mathrm{d}N/\mathrm{d}t \sim R\exp(-T_{ho}R)$. Next, we assume that a broad spectrum of trap levels results in a distribution of detrapping rates $D(R)$. To obtain the afterpulse probability, it is then necessary to integrate over this detrapping rate distribution, so that

$$P_{ap} \sim \int_{\tau_{max}}^{\tau_{min}} \mathrm{d}R D(R) R \mathrm{e}^{-T_{ho}R}. \quad (2)$$

To establish consistency with experimental measurement conditions, we choose the integration limits $\tau_{min} = 10\,\mathrm{ns}$ and $\tau_{max} = 10\,\mu\mathrm{s}$. However, different choices for these limits result in only modest quantitative differences in the model output.

Our strategy is to apply the model to obtain $P_{ap}(T_{ho})$ using different assumptions for the form of the detrapping rate distribution $D(R)$ to determine whether there are specific forms of $D(R)$ that yield the power law behavior found experimentally. The simplest case is that of a single trap type which is described by a delta function distribution $D(R) = \delta(R - R_0)$, where the single detrapping time $\tau_0 = 1/R_0$. This leads to an exponential dependence of P_{ap} on T_{ho}, and we have already determined that a simple exponential fit does not accurately describe experimental data. The case of a Gaussian distribution of detrapping rates can be viewed as a broadened delta function, and using the model, we find a similar exponential dependence for $P_{ap}(T_{ho})$ which again does not match the measured data.

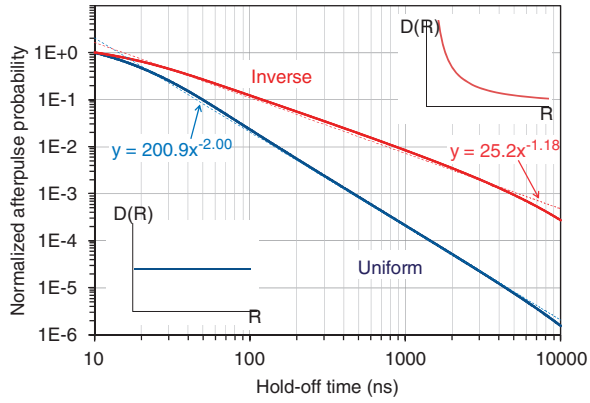

Figure 7. Solid curves indicate model results for normalized afterpulsing probability per unit time as a function of hold-off time following a primary avalanche. Results for the 'inverse' and 'uniform' detrapping rate distributions are labeled correspondingly. Light dashed lines show power law fits to the model results. Insets show the functional shape of the detrapping rate distributions $D(R)$ of the 'inverse' distribution (upper right) and 'uniform' distribution (lower left). (The color version of this figure is included in the online version of the journal.)

We obtain a more interesting result if we assume a uniform distribution of detrapping rates $D(R) = C$ for some constant C; physically, this corresponds to a collection of traps in which the number of traps with any given value of detrapping rate is always the same. After integrating over the range of hold-off times from 10 ns to 10 µs, the model provides the $P_{ap}(T_{ho})$ dependence shown in Figure 7 by the lower solid line labeled 'uniform'. The dashed line power law fit to the model output shows that the uniform distribution does provide the power law behavior $P_{ap} \propto T_{ho}^{-\alpha}$ but with $\alpha \sim 2.0$, which is significantly different from the values $\alpha \sim 1.2 \pm 0.2$ that we have obtained from fitting the experimental data.

Among all the other distributions we have considered, the only distribution for which the model in Equation (2) provides results that are consistent with our experimental results is an 'inverse' distribution given by $D(R) \propto 1/R$. The model output for this distribution is plotted as the upper solid line in Figure 7 labeled 'inverse', and the dashed line fit shows that it is accurately described by a power law $P_{ap} \propto T_{ho}^{-\alpha}$ with $\alpha \sim 1.18$, in good agreement with our experimentally determined power law behavior of $\alpha \sim 1.2 \pm 0.2$.

5. Discussion concerning model results and power law behavior

5.1. Potential insights from distribution modeling

Our modeling approach based on the assumption of a distribution of detrapping rates $D(R)$ is consistent with materials characterization results establishing a spectrum of trap levels in InP [25]. It is relevant to note that while 'narrow' distributions such as the delta function and Gaussian distributions yield exponential behavior for $P_{ap}(T_{ho})$, the wider distributions such as the uniform and inverse distributions exhibit power law behavior. This is natural given mathematical relationships by which integrations over a range of exponential functions can give power law functions. The fairly consistent value $\alpha \sim 1.2$ of the exponent of the power law found experimentally by us and a number of other groups provides a means for constraining the number of physically plausible trap distributions. In particular, the agreement of the measured data with the model results for the inverse distribution $D(R) \propto 1/R$ suggests that there is a higher density of traps with a slower release rate.

It is also important to point out, however, that we have no basis for identifying the inverse distribution as a unique solution for obtaining the power law behavior with $\alpha \sim 1.2$. There may be a more analytical approach that can provide a class of distributions which have this general property. It does seem likely that the general shape of distributions which exhibit this behavior will lie between 'narrow' distributions (e.g. delta function and Gaussian), which exhibit exponential behavior, and 'wider' distributions (e.g. uniform), which exhibit a power law dependence but with larger exponents than those found experimentally.

5.2. Alternate explanations for power law behavior

Notwithstanding the plausibility that a broad distribution of trap levels exists in the InP multiplication region, there may be other physical mechanisms that establish a distribution of detrapping rates. For instance, it has been pointed out that since afterpulsing occurs in the presence of very high electric fields ($\sim 5 \times 10^5$ V/cm), Poole–Frenkel effects can significantly change the detrapping behavior at multiplication region defects by enhancing the carrier emission probability [26]. Non-uniformity in the electric field amplitude across the multiplication region would induce a consequent variation in detrapping rates, and the resulting distribution could lead to the power law dependence of $P_{ap}(T_{ho})$ described in our discussion above.

It is also possible that there are other explanations for non-exponential temporal behavior in detrapping processes. Jonscher proposed a model [27] for the dynamics of deep level detrapping in the space charge regions of p–n junctions that invoked a dipole-like response leading to the prediction of power law behavior in detrapping effects. He points out that the observed functional form for the dielectric response of detrapping could have a number of causes such as many-body interactions, 'screened hopping' processes, or correlations between detrapping transitions. These speculations are qualitative, and they unfortunately do not provide information on the distributions of traps. But this earlier study makes clear that more complex 'correlated' detrapping mechanisms that deviate from the usual assumption of non-interacting independent traps can lead to power law behavior in the temporal behavior of detrapping.

Finally, we have also analyzed afterpulsing measurements obtained for Si SPADs [28], and these results seem to be qualitatively different from the results found for InGaAs/InP SPADs. For afterpulsing data taken over two orders of magnitude in hold-off time, the behavior of $P_{ap}(T_{ho})$ for Si SPADs falls between power law and exponential. It is possible that the distribution of detrapping rates in Si multiplication regions is considerably narrower than in InP regions due to a narrower spectrum of trap levels. The implication of this result – that the nature of defects in Si SPADs may be categorically different than for InP-based devices – is certainly plausible given the greater maturity of the Si materials system.

6. Conclusions

To carry out this study of the dependence of afterpulsing in InP/InGaAs SPADs on hold-off time, we increased the gating repetition rate of a matched delay line scheme for transient cancellation to 50 MHz. Additional improvements have already been made to achieve 100 MHz gating with this technique, and we believe further significant increases in photon counting rate are feasible with the matched delay line approach. To provide efficiency and flexibility to this measurement system, we have implemented FPGA-controlled electronics to count afterpulses for an arbitrary number of sequential gates following a photon-induced avalanche. This multiple-gate sequencing serves as a generalization of the traditional double-pulse method for characterizing the dependence of the afterpulsing probability on hold-off time. Moreover, similar FPGA-based circuitry will provide greatly enhanced capabilities for future photon counting instrumentation. In this present study, the multiple-gate sequencing approach proved to be noisier than well-implemented double-pulse measurements, but we hope to make

future improvements to enable the acquisition of comparably lower noise data. Additionally, although FPGA-based electronics provide substantial flexibility for gating bias control, our use of the matched delay line scheme for achieving higher frequency gating of 50 MHz does restrict us to rather short gate durations on the scale of a few nanoseconds.

Using our measured data for $P_{ap}(T_{ho})$, we have shown that the legacy analysis consisting of fitting the data with a series of exponentials to extract detrapping times supposedly associated with a few specific trap levels has no physical significance. The detrapping times extracted by this analysis are arbitrary and depend entirely on the range of hold-off times used in the measurements.

We instead show that $P_{ap}(T_{ho})$ can be quite accurately fit with a simple power law $P_{ap} = CT_{ho}^{-\alpha}$. To assess the generality of this result, we have analyzed afterpulsing data on InP/InGaAs SPADs obtained by a number of other groups, and we find that the power law description of these data is very consistent, with an exponent of $\alpha \sim 1.2 \pm 0.2$.

In considering that the literature on InP indicates a broad spectrum of trap levels in this material, we developed an afterpulsing model that assumes a distribution of detrapping rates $D(R)$. We found that for appropriate choices of the detrapping rate distribution, $P_{ap}(T_{ho})$ exhibits power law behavior for at least three orders of magnitude in hold-off time. Of all the distributions used in evaluating the model, the only one that provided good quantitative agreement with the measured data is an 'inverse' distribution corresponding to $D(R) \propto 1/R$. It is very possible that this distribution is not unique in providing good agreement with experimental results, but it establishes a general description of the trap population assuming that this afterpulsing model is correct. We also point out that other physical mechanisms could give rise to a distribution of detrapping rates, such as a variation in electric field amplitude in the multiplication region. Finally, there are other possible explanations for the observed power law behavior of $P_{ap}(T_{ho})$ if more complex 'correlated' detrapping mechanisms were to exist in the InP multiplication regions of these SPADs.

Further work with the afterpulsing model described in this paper is necessary to determine what additional predictions it may yield for experimental verification. Future efforts towards this end will be necessary to provide more evidence that the assumption of a distribution of traps is in fact a realistic explanation for the measured power law temporal dependence of afterpulsing in InP/InGaAs SPADs.

Acknowledgements

We are grateful to Alessandro Restelli, Josh Bienfang, Joe Campbell, Naota Namekata, and Massimo Ghioni for providing us with raw data from afterpulsing measurements made in their respective laboratories to analyze for this study. We also thank Prof. Ghioni for insightful discussions concerning afterpulsing in silicon SPADs.

References

[1] Itzler, M.A.; Ben-Michael, R.; Hsu, C.-F.; Slomkowski, K.; Tosi, A.; Cova, S.; Zappa, F.; Ispasoiu, R. *J. Mod. Opt*. **2007,** *54,* 283–304.
[2] Itzler, M.A.; Jiang, X.; Entwistle, M.; Slomkowski, K.; Owens, M.; Tosi, A.; Acerbi, F.; Zappa, F.; Cova, S. *J. Mod. Opt*. **2011,** *58,* 174–200.
[3] Single-Photon Technologies. Selected Papers from the 4th International Workshop on Single and Entangled Photons: Sources, Detectors, Components, and Applications. *J. Mod. Opt*. **2011,** *58*(3–4), 169–337.
[4] Zappa, F.; Tosi, A.; Cova, S. *Proc. SPIE* **2007,** *6583,* 65830E.
[5] Namekata, N.; Sasamori, S.; Inoue, S. *Opt. Express*. **2006,** *14,* 10043–10049.
[6] Namekata, N.; Adachi, S.; Inoue, S. *IEEE Photonics Technol. Lett*. **2010,** *22,* 529–531.
[7] Zhang, J.; Eraerds, P.; Walenta, N.; Barreiro, C.; Thew, R.; Zbinden, H. *Proc. SPIE* **2010,** *7681,* 76810Z.
[8] Yuan, Z.L.; Kardynal, B.E.; Sharpe, A.W.; Shields, A.J. *Appl. Phys. Lett*. **2007,** *91,* 041114.
[9] Yuan, Z.L.; Sharpe, A.W.; Dynes, J.F.; Dixon, A.R.; Shields, A.J. *Appl. Phys. Lett*. **2010,** *96,* 071101.
[10] Warburton, R.E.; Itzler, M.; Buller, G.S. *Appl. Phys. Lett*. **2009,** *94,* 071116.
[11] Zhao, K.; You, S.; Cheng, J.; Lo, Y.-H. *Appl. Phys. Lett*. **2008,** *93,* 153504.
[12] Itzler, M.A.; Jiang, X.; Onat, B.M.; Slomkowski, K. *Proc. SPIE* **2010,** *7608,* 760829.
[13] Donnelly, J.P.; Duerr, E.K.; McIntosh, K.A.; Dauler, E.A.; Oakley, D.C.; Groves, S.H.; Vineis, C.J.; Mahoney, L.J.; Molvar, K.M.; Hopman, P.I.; Jensen, K.E.; Smith, G.M.; Verghese, S. *IEEE J. Quantum Electron*. **2006,** *42,* 797–809.
[14] Pellegrini, S.; Warburton, R.E.; Tan, L.J.J.; Ng, J.S.; Krysa, A.B.; Groom, K.; David, J.P.R.; Cova, S.; Robertson, M.J.; Buller, G.S. *IEEE J. Quantum Electron*. **2006,** *42,* 397–403.
[15] Cova, S.; Lacaita, A.; Ripamonti, G. *IEEE Electron Device Lett*. **1991,** *12,* 685–687.
[16] Bethune, D.S.; Risk, W.P. *IEEE J. Quantum Electron*. **2000,** *36,* 340–347.
[17] Bethune, D.S.; Risk, W.P.; Pabst, G.W. *J. Mod. Opt*. **2004,** *51,* 1359–1368.
[18] Ben-Michael, R.; Itzler, M.A.; Nyman, B.; Entwistle, M. In *2006 Digest of the LEOS Summer Topical Meetings*; IEEE, 2006; pp 15–16.
[19] Jensen, K.E.; Hopman, P.I.; Duerr, E.K.; Dauler, E.A.; Donnelly, J.P.; Groves, S.H.; Mahoney, L.J.; McIntosh,

K.A.; Molvar, K.M.; Napoleone, A.; Oakley, D.C.; Verghese, S.; Vineis, C.J.; Younger, R.D. *Appl. Phys. Lett*. **2006,** *88,* 133503.
[20] Trifonov, A.; Subacius, D.; Berzanskis, A.; Zavriyev, A. *J. Mod. Opt*. **2004,** *51,* 1399–1415.
[21] Liu, M.; Hu, C.; Bai, X.; Guo, X.; Campbell, J.C.; Pan, Z.; Tashima, M.M. *IEEE J. Sel. Top. Quantum Electron*. **2007,** *13,* 887–894.
[22] Restelli, A.; Bienfang, J. *J. Mod. Opt*., in press.
[23] Hu, C.; Zheng, X.; Campbell, J.C.; Onat, B.M.; Jiang, X.; Itzler, M.A. *J. Mod. Opt*. **2011,** *58,* 201–209.
[24] Namekata, N.; Inoue, S. *Proc. SPIE* **2011,** *7945,* 79452K.
[25] Anderson, W.A.; Jiao, K.L. In *Indium Phosphide and Related Materials: Processing, Technology, and Devices*; Katz, A., Ed.; Artech House: Boston, 1992; Chapter 3.
[26] Ghioni, M.; Gulinatti, A.; Rech, I.; Maccagnani, P.; Cova, S. *Proc. SPIE*. **2008,** *6900,* 69001D.
[27] Jonscher, A.K. *Solid-State Electron*. **1990,** *33,* 139–142.
[28] Ghioni, M.; Gulinatti, A.; Rech, I.; Zappa, F.; Cova, S. *IEEE J. Sel. Top. Quantum Electron*. **2007,** *13,* 852–862.