

Afterpulse Reduction Through Prompt Quenching in Silicon Reach-Through Single-Photon Avalanche Diodes

Michael A. Wayne, Alessandro Restelli, Joshua C. Bienfang, and Paul G. Kwiat

Abstract—Reducing afterpulsing in single-photon avalanche diodes (SPADs) allows operation with shorter recovery times and higher detection rates. Afterpulsing in SPADs can be reduced by reducing the total avalanche charge. We use a periodic quenching system to arbitrarily vary the latency between the onset of an avalanche and the application of the quench, allowing us to characterize the afterpulsing behavior when the current flow is halted at time scales that are significantly shorter than can be achieved by standard active-quenching systems. Three different reach-through SPADs are characterized, and with prompt quenching we observe reductions in afterpulse probability of as much as a factor of 12. Beyond improving detection rates, reducing the total avalanche charge can also allow operation with higher excess bias voltages, which enables higher detection efficiency and more precise timing resolution.

Index Terms—Photodetectors, photodiodes, quantum communication, quantum detectors, semiconductor devices, single-photon detectors.

SINGLE-PHOTON avalanche diodes (SPADs), particularly in silicon, can provide low noise, high efficiency, and low jitter single-photon detection [1]. For many quantum-information systems, the performance characteristics of the single-photon detectors directly determine the performance of the overall experiment [2]. For example, in some quantum-random-number generators the maximum bit-generation rate is almost exclusively limited by the recovery time of the system's single-photon detectors [3], [4]. In many SPADs, the minimum resolvable inter-arrival time, as determined by the detector's timing jitter, and the minimum countable inter-arrival time, as determined by the recovery time, differ by several orders of magnitude. From this perspective, the recovery time imposes an inefficiency in the detector's resources; such detectors could,

in principle, resolve many more photon-arrival times than they can actually count.

For a SPAD operating above its breakdown voltage, a single photo-generated charge pair can result in the growth of a macroscopic current, or avalanche. During this process, it is possible for some proportion of the avalanche charge to become trapped in localized defects within the device. If one of these charge carriers is released while the device is above breakdown, it may initiate another avalanche, or "afterpulse" [5]. In many applications, spurious detections of this type are severely detrimental and a holdoff period is applied after each detection event by means of external circuitry. Such active-quenching systems lower the bias voltage and hold it at a value below breakdown for a pre-determined time, terminating the avalanche and allowing the trapped charge an opportunity to be released. After the holdoff, the SPAD is reset, and the bias restored to its initial value above breakdown to re-arm the device. The duration of the holdoff period is chosen to reduce the afterpulse probability to a tolerably low value, but also limits the maximum conceivable count rate to the inverse of the total recovery time [6].

For InGaAs/InP SPADs, notable increases in maximum count rate have been achieved by gating the device in a manner that strongly limits the total charge in each avalanche [1]. In this work we investigate the application of the same principle to Si reach-through SPADs. Due to their relatively thick absorption regions ($>15 \mu\text{m}$), reach-through SPADs have high detection efficiency in the red and near-infrared regions, but require large bias voltages that are more difficult to control at ultra-short time scales.

In this study we use an experimental setup that allows for a variable and arbitrarily short quench delay, the interval between the onset of an avalanche and its termination by the active change in bias voltage. This system allows the characterization of afterpulsing in three different types of reach-through SPADs under quenching conditions that heretofore have not been explored. We find that this regime is characterized by a sharp decrease in afterpulsing, by as much as a factor of 12, as the quench delay is shortened, indicating a significant reduction in the total avalanche charge by means of prompt active quenching. In addition, we present a simple numerical model, based on measured SPAD and circuit parameters, that gives excellent agreement with measured afterpulse probabilities. This model provides a useful basis for optimizing active-quenching circuits to achieve shorter recovery times and higher count rates.

Minimizing the total avalanche charge has long been a priority in the design of active-quenching systems [7], and a common and highly effective method uses a combination of both passive

Manuscript received May 5, 2014; revised June 30, 2014; accepted July 31, 2014. Date of publication August 13, 2014; date of current version September 17, 2014. This work was supported by the DARPA InPho Program through U.S. Army Research Office Award W911NF-10-1-0395.

M. A. Wayne is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA, with the Joint Quantum Institute, University of Maryland College Park, MD 20742 USA, and with the National Institute of Standards and Technology, Gaithersburg, MD 20899 USA (e-mail: mwayne@illinois.edu).

A. Restelli and J. C. Bienfang are with the Joint Quantum Institute, University of Maryland, College Park, MD 20742 USA, and also with the National Institute of Standards and Technology, Gaithersburg, MD 20899 USA (e-mail: alessandro.restelli@nist.gov; bienfang@nist.gov).

P. G. Kwiat is with the Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: kwiat@illinois.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JLT.2014.2346736

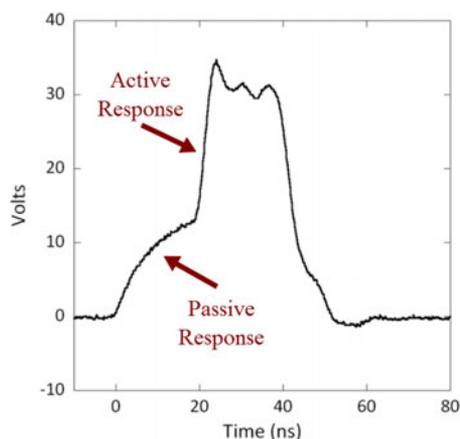


Fig. 1. Measured circuit response of an example SPAD module. Initially, the SPAD is slowly passively quenched until the active quenching response occurs at 20 ns, by which time the majority of the avalanche current has flowed through the device.

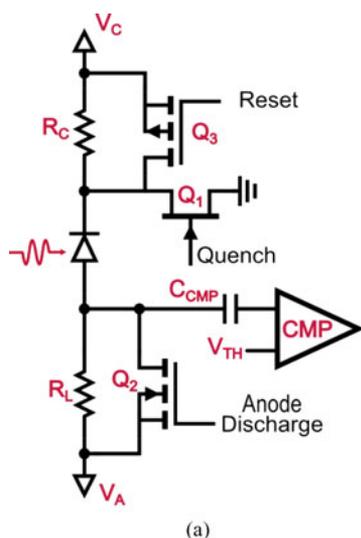


Fig. 2. Simplified schematic of periodic-quenching circuit.

and active quenching. In this approach, a large load resistance R_L , significantly greater than the series resistance of the SPAD R_d , causes the avalanche current to decay exponentially from its peak value. An active-quenching cycle then terminates any residual current, imposes a holdoff, and resets the SPAD to its active state in an orderly fashion [8]. For example, Fig. 1 shows the circuit response observed in one commercial SPAD module with a ≈ 50 ns recovery time and $< 1\%$ afterpulse probability. As can be seen, the time between the initiation of the avalanche and the active change in the bias voltage, the quench delay, is roughly 20 ns. In systems of this type, the peak current is determined primarily by R_d and the excess bias voltage V_E . The time constant for the current's exponential decay can be approximated by $\tau_c = R_d \times (C_d + C_S)$, where C_d is the series capacitance of the diode, and C_S accounts for any additional stray capacitance at the junction of the SPAD and the load resistance (cf. Fig. 3) [6]. For reach-through SPADs in

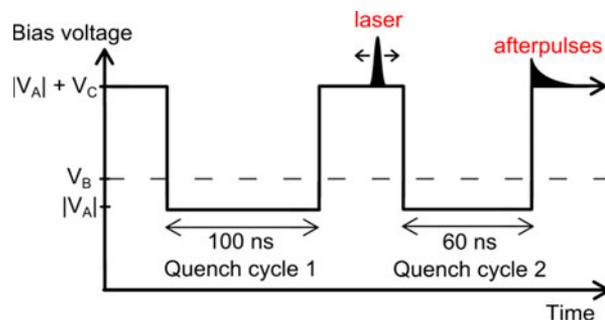


Fig. 3. Timing diagram of the afterpulse characterization measurement. During both quench cycles, the voltage across the SPAD is quenched to $|V_A|$, held off for a predetermined holdoff time, and then reset back to $|V_A| + V_C$. These quench cycles are repeated every $2 \mu\text{s}$.

well-designed circuits this time scale is short, on the order of a few nanoseconds, and imposes a natural minimum on the total charge flow with essentially passive methods. To have a significant impact on the total charge flow, the quench delay must therefore be shorter than τ_c , and this poses a significant challenge for the design of active quenching circuits, particularly when the excess bias voltage V_E is of the order of tens of volts, as for reach-through SPADs. Unless special effort is made, the combination of propagation delay and achievable slew rates for the quenching transition typically cause the quench delay to be greater than τ_c .

To realize significantly shorter quench delays we use the periodically quenched system illustrated in Fig. 2. This circuit uses a load resistance $R_L = 1 \text{ M}\Omega$, as in the passive-then-active approach described above, however, feedback that would normally condition the application of a quench cycle upon sensing an avalanche has been removed and is replaced by periodic control signals from synchronized pattern and function generators. The signal generators induce two quench cycles, quench/holdoff/reset, regardless of whether or not an avalanche occurred, as illustrated in Fig. 3.

The period of the system is $2 \mu\text{s}$, and for the majority of that time the SPAD is fully charged unless there is an avalanche. Avalanche events that occur far from a quench cycle, for example, dark counts or afterpulses, undergo passive quenching, after which the SPAD bias recharges slowly through R_L [6] with a time constant in the microsecond regime. To ensure that the diode is in a well-known state when an optical pulse is applied, an initial quench cycle (“Quench cycle 1” in Fig. 3, 100 ns holdoff) is applied to put the SPAD in its fully armed state and ready for single-photon detection under pre-determined bias conditions. The interval between the end of the first and the beginning of the second quench cycle forms a 50 ns window in which an attenuated laser pulse is applied at a time determined by the pattern generator. The temporal delay of the optical pulse can be adjusted arbitrarily, providing complete control of the interval between the onset of an avalanche and the application of the quench transition. After the end of the second quench cycle, afterpulses are recorded over the remaining $1.79 \mu\text{s}$, conditional on a detection event due to the laser. All detection events are recorded by a custom time-tagging system based on a 10 Gbit/s

demultiplexer,¹ [9], and a Virtex-6 FPGA (see footnote 1) [10] with 100 ps resolution and a dead time of 1.6 ns. Tagged events are buffered and passed over a PCIe bus interface for continuous analysis on a computer. Although the time-tagging system can record multiple afterpulse events, we find that secondary and tertiary afterpulse events are extremely rare ($< 0.01\%$), due to the slow recovery following passive quenching.

When fully armed, the voltage across the SPAD is $V_C - V_A$, where V_A is negative with magnitude a few volts smaller than the breakdown voltage V_B , and V_C is used to set the excess bias voltage. The value of V_C is such that the voltage across the SPAD is equal to $V_B + V_E$ when active, and is $|V_A|$ during the holdoff. To achieve a fast voltage transition for the quench, we use a wideband GaN transistor (see footnote 1) [11], denoted by Q1 in Fig. 2, to lower the cathode voltage to the ground potential and hold it there for the duration of the holdoff. This device provides a fast slew rate (>25 V/ns) that rapidly brings the SPAD below breakdown. However, its leakage current is significantly larger than more commonly used switching MOSFETs [5], which makes it unsuitable for acting on the high-impedance node. We therefore use resistor $R_C = 200 \Omega$ at the SPAD cathode, and a secondary lower-speed switching MOSFET, Q2, to discharge the anode shortly after the quench is applied to the cathode. The duration of the holdoff can be chosen by programming the pattern generator, and at the end of the holdoff another transistor, Q3, resets the cathode voltage to V_C . Finally, a high-speed comparator ac-coupled to the anode senses avalanches for counting, and latch signals from the pattern generator are used to mask the quench- and reset-transitions that couple through the diode.

We use this system to study prompt quenching in three different commercially available reach-through SPADs: SPAD A and B are available in a transistor-outline package, while SPAD C was removed from a module for use in the experimental setup. We operate the SPADs at a temperature of -10 °C. The breakdown voltage of each device, along with the excess bias voltage and background (non-illuminated) count rate, is listed in Table II of the Appendix. Each diode has a different design and their detection efficiencies depend differently on wavelength and excess bias voltage. For this reason it is not appropriate to choose a single detection efficiency at which to operate all three devices, as this may put some devices outside their normal operating regime. Rather, we chose an excess bias voltage in the upper range of where each device may normally be operated: 15 V for SPAD A and B, and 7.5 V for SPAD C. For example, the detection efficiencies of SPAD A and B are measured to be $(43 \pm 2)\%$ and $(34 \pm 2)\%$, respectively. The laser is an 850 nm gain-switched VCSEL producing < 30 ps pulses that are coupled to single-mode fiber, attenuated to a low mean photon number, and focused to a $< 10 \mu\text{m}$ spot on the SPAD's active area. An additional 10 nm band-pass filter in front of the SPAD suppresses background photons. For each SPAD, the

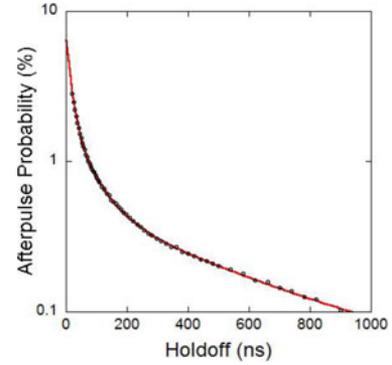


Fig. 4. Total afterpulse probability versus holdoff for SPAD B at -10 °C, $V_E = 15$ V, and a quench delay of ≈ 5 ns. By fitting multiple decaying exponentials (red) to experimental data (black), trap lifetimes and relative weights are extracted.

mean photon number of the optical pulse was adjusted to produce a count rate of approximately 50 kHz, and was always less than 0.3 photons per pulse.

To characterize afterpulsing as a function of quench delay, the arrival time of the optical pulse is moved incrementally towards the beginning of the second quench cycle, starting at 25 ns before the quench, and ending when avalanche events due to the optical pulse can no longer be detected. At each position, all counts after the end of the second quench cycle, conditional on a detection event having occurred within a 1.0 ns window at the arrival time of the optical pulse, are recorded. To distinguish afterpulses from coincidental background counts (e.g., dark counts, or light leakage), the background count probability is measured shortly after the first quench cycle, when the SPAD is fully armed and awaiting the arrival of an optical pulse. The background count rate is listed in Table II of the Appendix. This background-count probability is subtracted from the measured conditional count probability, giving the afterpulse probability. The resulting afterpulse probabilities versus quench delay for the three SPADs are shown in Figs. 5–7, along with curves from the numerical model described below.

We model the afterpulse probability as a function of quench delay based on the assumption that the observed afterpulse probability is proportional to the number of charge carriers N that are trapped in the device when it is reactivated at the end of the holdoff [12]. SPADs often exhibit various trap levels [13], and we model the total number of trapped charges as $N = \sum_i n_i$, where n_i is the population of the i th trap level. For each trap level we solve numerically the rate equation

$$\frac{dn_i}{dt} = a_i \frac{I(t, t_Q)}{q_e} - \frac{n_i}{\tau_i} \quad (1)$$

where $I(t, t_Q) \equiv I(t)(1 - H(t - t_Q))$ is the avalanche current flowing through the SPAD as a function of time, starting at $t = 0$ and terminating at the quench delay t_Q ($H(t)$ is the Heaviside function), q_e is the electron charge, and τ_i and a_i are the lifetime and relative weighting of each trap level i , respectively. Qualitatively, this rate equation expresses two competing terms: the population of traps due to current flow, and the depopulation

¹The identification of any commercial product or trade name does not imply endorsement or recommendation by the National Institute of Standards and Technology, nor does it imply that the materials or equipment implied are the best available for the purpose.

TABLE I
MEASURED TRAP LIFETIMES AND THEIR RELATIVE WEIGHTINGS
FOR EACH SPAD.

SPAD	τ_1 (ns)	a1	τ_2 (ns)	a2	τ_3 (ns)	a3
SPAD A	9.06 ± 0.10	0.99 ± 0.04	689.4 ± 77.5	0.01 ± 0.004		
SPAD B	16.39 ± 0.43	0.70 ± 0.01	71.63 ± 4.13	0.22 ± 0.03	485.3 ± 5.1	0.07 ± 0.05
SPAD C	58.88 ± 1.47	1				

Uncertainties indicate one standard deviation.

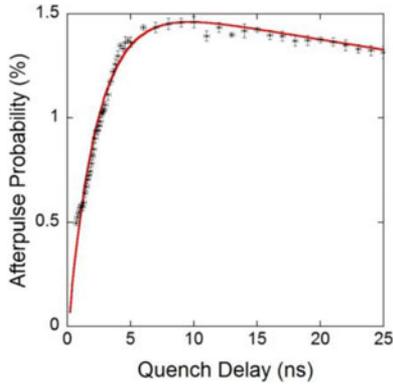


Fig. 5. Afterpulsing versus quench delay for SPAD B; model (red line) and data (black dots). The avalanche signal was sensed with $C_{CMP} = 5$ pF. Error bars represent one standard deviation.

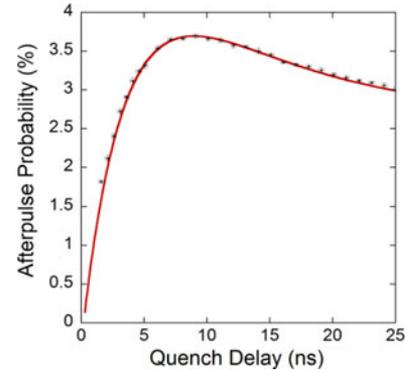


Fig. 6. Afterpulsing versus quench delay for SPAD A; model (red line) and data (black dots). The avalanche signal was with $C_{CMP} = 5$ pF. Error bars represent one standard deviation.

of traps according to their natural lifetime. The time-dependent current flowing through the SPAD, $I(t)$, is calculated from a numerical simulation of the circuit in Fig. 2, following the approach in ref. [14]. This model requires accurate measurements of the series resistance and capacitance of each SPAD. The methods and results of these measurements, along with the circuit simulation, are described in the Appendix.

Similar to time-correlated carrier counting [13], the trap lifetimes τ_i and relative weighting a_i for each device are extracted from measurements of the afterpulse probability versus holdoff. The relative weighting term is related to the trap density and the volume of the active region. To perform this measurement, an optical pulse is positioned ≈ 5 ns from the start of the second quench cycle, and the total afterpulse probability, conditioned on an avalanche from the optical pulse, is measured as the duration of the holdoff is varied from 10 ns to 1 μ s. The resulting afterpulse probability is fit with a multi-term exponential of the form $\sum_i a_i \exp(-t/\tau_i)$; Fig. 4 shows the measured afterpulse probability versus holdoff for SPAD B, along with the fit. Three trap levels are necessary to accurately fit these data, with the lifetimes and relative weightings listed in Table I. The shortest lifetime quantitatively agrees with the measurement reported in ref. [15]. The same procedure was carried out for SPAD A, for which only two trap levels were necessary for accurate fitting, and SPAD C, for which only a single trap level was required. Because we have only measured out to 1 μ s, the uncertainty of the longer trap lifetimes is higher.

Equation (1) is solved numerically for n_i using the parameters for each trap level, and the SPAD circuit model, for each value

of the quench delay t_Q . The resulting n_i are then summed to give the total number of trapped charges at the end of the holdoff N , as a function of t_Q . The relation between the total number of trapped charges and the afterpulse probability may depend on a variety of factors (for example, the avalanche initiation probability), we assume these factors are equal for each trap level and take the form of a multiplicative constant. With this single free parameter the model agrees very well with the observed afterpulse probability as a function of the quench delay t_Q , as can be seen in Figs. 5–7.

As can be seen in Figs. 5–7, each device exhibits a sharp decrease in afterpulsing at short quench delays, a behavior that is directly related to the current profile $I(t)$. As discussed in the Appendix, the avalanche current is assumed to rise at a time scale significantly shorter than the minimum achievable quench delays used here [16]. The minimum quench delay varies from one device to another, and is determined predominantly by the lowest operable comparator threshold, which in turn is determined by the ringing from the control signal transitions.

As shown in the Appendix, the product of the series resistance and capacitance for SPAD A and B are of comparable value, and both devices are operated with a 5 pF capacitor in front of the comparator. As a result, the time scale for the current decay for these two devices is similar (cf. Fig. 8), and hence the range of quench delays over which the afterpulse probability is sharply reduced (< 5 ns) (Figs. 5 and 6) is also similar. However, the magnitude of the afterpulse probability in these two devices differs by more than a factor of 2. This may be due to a difference in the trap density, or to a difference in the active volume of the devices [15]. In contrast, SPAD C (see Fig. 7)

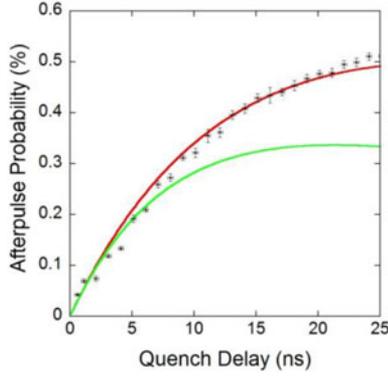


Fig. 7. Afterpulsing versus quench delay for SPAD C; model (red line) and data (black dots). Due to the small avalanche signal, the avalanche was sensed with $C_{CMP} = 10$ pF. A theoretical estimate of the afterpulsing behavior with a 5 pF capacitor is shown in green. Error bars represent one standard deviation.

exhibits a significantly lower slope in its afterpulse probability versus quench delay, which agrees with the significantly larger series resistance of this device. In addition, the avalanche signal was smaller for SPAD C and required the use of a larger (10 pF) capacitor at the comparator, which further increased the time constant of the current decay. Nonetheless, SPAD C exhibited the lowest overall afterpulse probability, and the largest relative reduction in afterpulsing with prompt quenching ($\approx 12\times$). The projected performance with a 5 pF capacitor at the comparator is also shown in Fig. 7.

Each device exhibits a maximum in its afterpulse probability as the quench delay is increased. This peak is formed by the competing processes represented in Eq. (1): as the quench delay increases, the total amount of charge that flows through the device increases by smaller increments, while the amount of time over which trapped charges may be released increases. The location of this peak is a function of the properties of the SPAD, the circuit in which it is operated, and the lifetimes of the various trap levels.

It should be noted that because afterpulses were counted conditional on the detection of a photon, the performance observed in these data are representative of what could be achieved if quenching were applied conditional on a detection event, as with a feedback circuit. These data show that through prompt quenching it is possible to lower the afterpulse probability by significant factors, at least 2, 3, and 12, for SPADs A, B, and C, respectively. To determine how much of a reduction in holdoff this improved afterpulse performance would allow we assume the afterpulse probability decays exponentially, and that a time t_{ho} is the holdoff necessary for the afterpulse probability to decay to some value AP (e.g., 1%):

$$AP = Ce^{-\frac{t_{ho}}{\tau}} \quad (2)$$

where C is some constant related to the total amount of charge. If, with prompt quenching, the afterpulse probability can be reduced by a factor of X , then the holdoff can be shortened to $(t_{ho} - t_{sh})$, at which point the afterpulse probability is once

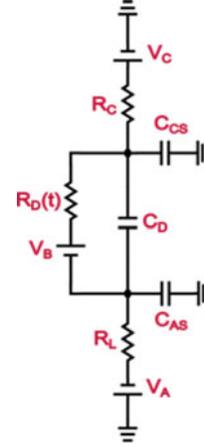


Fig. 8. Equivalent SPICE model of a SPAD circuit.

again AP, and the difference is given by

$$t_{ho} - t_{sh} = \tau \ln(X). \quad (3)$$

For example, with a holdoff of 60 ns, SPAD B displays its worst-case afterpulse probability ($\approx 1.45\%$) when the quench delay is 8 ns. The afterpulsing is reduced (see Fig. 5) by a factor of ≈ 3 with prompt quenching. According to Eq. (3), and considering only the dominant trap level, the holdoff could be reduced to 42 ns before the afterpulse probability would exceed the value it had with an 8 ns quench delay. A similar calculation using the reduction factor of ≈ 12 observed in SPAD C indicates that significant reductions in holdoff time are possible. It is worthwhile to note that at very short holdoff times, short-lived traps that were not detected or considered in this analysis may influence the minimum achievable recovery time.

In conclusion, we have demonstrated the potential benefit of applying prompt active quenching to a variety of reach-through SPADs. These studies show that significant reductions in the overall afterpulse probability are available if quenching is applied at time scales significantly shorter than the characteristic time scale of the current decay. This requirement is challenging, but may be achievable with new classes of wideband amplifiers. The resulting increase in maximum count rate is likely to be more modest, due to the logarithmic relationship between the afterpulsing reduction and the holdoff time. Nonetheless, these results show that there are benefits to be gained in the performance of reach-through SPADs that can be achieved with improved circuit design. It is also worthwhile to note that methods that reduce afterpulsing can, in some instances, allow operation of the SPAD under higher excess bias voltage, which may improve other performance characteristics, such as timing resolution and detection efficiency.

APPENDIX

To model the avalanche current flow as a function of time we use the simple circuit-level model given in Ref. [15]. In this model the SPAD is approximated as a parallel combination of its series capacitance and a time-dependent series resistance that

TABLE II
CIRCUIT PARAMETERS FOR EACH SPAD MEASURED AT $(-10 \pm 0.1)^\circ\text{C}$

SPAD	R_d (Ω)	C_d (pF)	V_B (V)	V_E (V)	Back (s^{-1})
SPAD A	525 ± 25	1.85 ± 0.2	216 ± 0.1	15	2000
SPAD B	315 ± 15	2.27 ± 0.5	398 ± 0.1	15	1000
SPAD C	1300 ± 65	1.66 ± 0.2	111 ± 0.1	7.5	200

V_E is the excess bias above breakdown and background is the number of counts measured when the laser is turned OFF; R_d , V_B , and C_d are defined in Fig. 8. Uncertainties in R_d and V_B represent one standard deviation, while for C_d the uncertainty encompasses the full-range of variation observed over all measurements.

changes from near infinite to some steady-state value upon photon detection. Accurate modelling requires measuring the series resistance and capacitance, as well as the additional capacitance at the anode and cathode due to the circuit in which the SPAD is operated.

The series resistance of a SPAD has been modeled previously with a complex time-dependent behavior [15], but for the after-pulsing measurements reported in this experiment we find that excellent agreement can be achieved by having the series resistance decay exponentially from near infinite to the measured steady-state series resistance at a time scale much faster than the RC time constant of the subsequent decay. To measure the steady-state series resistance of each SPAD we reduce the load resistance R_L to $\approx 1 \text{ k}\Omega$ (a value that will not cause passive quenching), and for device longevity we disable the reset at the end of quench cycle 2 so that the SPAD is only biased above breakdown during the interval between the two quench cycles. In this mode, when an avalanche occurs, the voltage at the cathode drops rapidly to a value determined by the voltage division of the excess bias voltage between R_C and R_d (cf., Fig. 8), where it remains until quench cycle 2 terminates the flow of current. We measure this steady-state cathode voltage with a high-impedance probe; the resulting values for R_d are listed in Table I.

The series capacitance of a reach-through SPAD is typically small ($< 5 \text{ pF}$), and must be measured when the device is reverse-biased near the breakdown voltage. To confirm accuracy, the series capacitance is measured using two different methods in separate test circuits. In the first, the diode capacitance is extracted from the RC time constant measured as the cathode recharges through a known resistance after the abrupt turn-off of a switching MOSFET. In this measurement the anode is biased negatively near breakdown, and the additional capacitance of the test circuit and device packaging is measured (using an empty package of the same type) and subtracted off to yield the capacitance of the diode. In the second method a sinusoidal signal at various frequencies from 100 to 300 MHz is ac-coupled to the negatively biased anode, and the cathode is connected to a spectrum analyzer. With knowledge of the sinusoidal amplitude at the anode, the capacitance can be determined from the signal strength measured by the analyzer. The results of both methods show good agreement, and the values obtained in the latter method are reported in Table II.

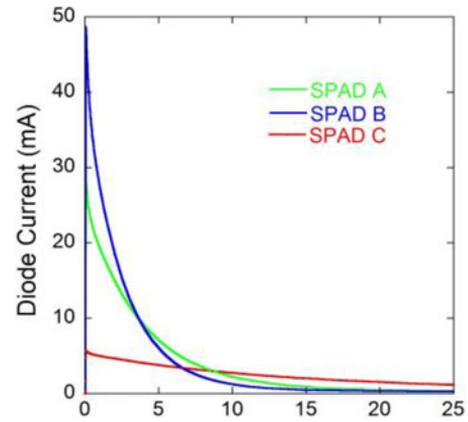


Fig. 9. Theoretical diode currents for SPAD A (green), SPAD B (blue), and SPAD C (red).

Finally, SPICE modeling of the circuit shown in Fig. 8 results in the diode currents of Fig. 9. The current profiles for SPAD A and B are similar, and are much faster than that of SPAD C. The time scale of the current flow for these two devices suggests that in order to achieve significant reductions in afterpulsing they must be promptly quenched in the few-nanosecond regime, whereas for SPAD C significant reductions in total charge can be achieved with quenching at more moderate time scales.

REFERENCES

- [1] A. Migdall, S. V. Polyakov, J. Fan, and J. C. Bienfang, Eds., *Single-Photon Generation and Detection: Physics and Applications*, vol. 45. New York, NY, USA: Academic, 2013.
- [2] P. J. Clarke, R. J. Collins, P. A. Hiskett, M.-J. García-Martínez, N. J. Krichel, A. McCarthy, M. G. Tanner, J. A. O'Connor, C. M. Natarajan, S. Miki, M. Sasaki, Z. Wang, M. Fujiwara, I. Rech, M. Ghioni, A. Gulinatti, R. H. Hadfield, P. D. Townsend, and G. S. Buller, "Analysis of detector performance in a gigahertz clock rate quantum key distribution system," *New J. Phys.*, vol. 13, no. 7, p. 075008, Jul. 2011.
- [3] M. A. Wayne and E. Jeffrey, "Photon arrival time quantum random number generation," *J. Mod. Opt.*, vol. 56, no. 4, pp. 516–522, 2009.
- [4] M. A. Wayne and P. G. Kwiat, "Low-bias high-speed quantum random number generator via shaped optical pulses," *Opt. Exp.*, vol. 18, no. 9, pp. 9351–9357, Apr. 2010.
- [5] M. Ghioni, S. Cova, F. Zappa, and C. Samori, "Compact active quenching circuit for fast photon counting with avalanche photodiodes," *Rev. Sci. Instrum.*, vol. 67, no. 10, pp. 3440–3448, Oct. 1996.
- [6] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, "Avalanche photodiodes and quenching circuits for single-photon detection," *Appl. Opt.*, vol. 35, no. 12, pp. 1956–1976, Apr. 1996.
- [7] H. Dautet, P. Deschamps, B. Dion, A. D. MacGregor, D. MacSween, R. J. McIntyre, C. Trotter, and P. P. Webb, "Photon counting techniques with silicon avalanche photodiodes," *Appl. Opt.*, vol. 32, no. 21, pp. 3894–3900, Jul. 1993.
- [8] S. Cova, A. Longoni, and A. Andreoni, "Towards picosecond resolution with single-photon avalanche diodes," *Rev. Sci. Instrum.*, vol. 52, no. 3, pp. 408–412, Mar. 1981.
- [9] Adantec ASNT2011-PQA Demultiplexer. (2014, Jan. 16). [Online]. Available: <http://www.asantec.com/26-asnt2011-pqa.html>
- [10] Xilinx Virtex-6 Family Overview. (2014, Jan. 16). [Online]. Available: <http://www.xilinx.com/products/silicon-devices/fpga/virtex-6/>
- [11] Nitronex NPTB00004 Datasheet. (2014, Mar. 24). [Online]. Available: <http://www.nitronex.com/pdfs/NPTB00004.pdf>
- [12] K. E. Jensen, P. I. Hopman, E. K. Duerr, E. A. Dauler, J. P. Donnelly, S. H. Groves, L. J. Mahoney, K. A. McIntosh, K. M. Molvar, A. Napoleone, D. C. Oakley, S. Verghese, C. J. Vineis, and R. D. Younger, "Afterpulsing in Geiger-mode avalanche photodiodes for 1.06 μm wavelength," *Appl. Phys. Lett.*, vol. 88, no. 13, pp. 133503–1–133501-3, Mar. 2006.

- [13] S. Cova, A. Lacaita, and G. Ripamonti, "Trapping phenomena in avalanche photodiodes on nanosecond scale," *IEEE Electron Device Lett.*, vol. 12, no. 12, pp. 685–687, Dec. 1991.
- [14] F. Zappa, A. Tosi, A. D. Mora, and S. Tisa, "SPICE modeling of single photon avalanche diodes," *Sens. Actuators Phys.*, vol. 153, no. 2, pp. 197–204, Aug. 2009.
- [15] M. Stipčević, D. Wang, and R. Ursin, "Characterization of a commercially available large area, high detection efficiency single-photon avalanche diode," *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3591–3596, Dec. 2013.
- [16] A. Ingargiola, M. Assanelli, I. Rech, A. Gulinatti, and M. Ghioni, "Avalanche current measurements in SPADs by means of hot-carrier luminescence," *IEEE Photon. Technol. Lett.*, vol. 23, no. 18, pp. 1319–1321, Sep. 2011.

Authors' biographies not available at the time of publication.