

Single-photon detection efficiency up to 50% at 1310nm with an InGaAs/InP avalanche diode gated at 1.25GHz

Alessandro Restelli, Joshua C. Bienfang, and Alan L. Migdall

Citation: *Appl. Phys. Lett.* **102**, 141104 (2013); doi: 10.1063/1.4801939

View online: <http://dx.doi.org/10.1063/1.4801939>

View Table of Contents: <http://apl.aip.org/resource/1/APPLAB/v102/i14>

Published by the [American Institute of Physics](http://www.aip.org).

Additional information on *Appl. Phys. Lett.*

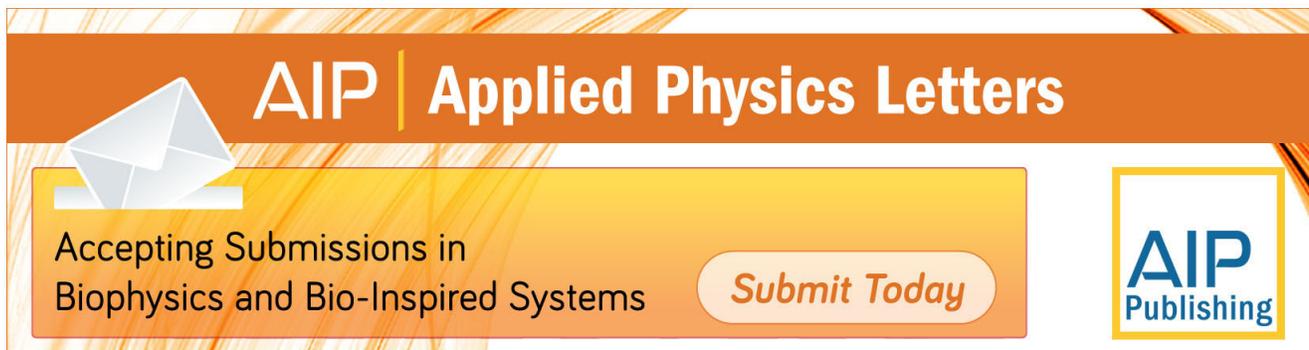
Journal Homepage: <http://apl.aip.org/>

Journal Information: http://apl.aip.org/about/about_the_journal

Top downloads: http://apl.aip.org/features/most_downloaded

Information for Authors: <http://apl.aip.org/authors>

ADVERTISEMENT



AIP | Applied Physics Letters

Accepting Submissions in
Biophysics and Bio-Inspired Systems

Submit Today

AIP
Publishing

Single-photon detection efficiency up to 50% at 1310 nm with an InGaAs/InP avalanche diode gated at 1.25 GHz

Alessandro Restelli,^{1,a)} Joshua C. Bienfang,^{1,2} and Alan L. Migdall^{1,2}

¹*Joint Quantum Institute, National Institute of Standards and Technology and University of Maryland, 100 Bureau Drive, Gaithersburg, Maryland 20899, USA*

²*National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, Maryland 20899, USA*

(Received 23 January 2013; accepted 1 April 2013; published online 10 April 2013)

We describe a gated Geiger-mode single-photon avalanche diode (SPAD) detection system in which both gating and avalanche discrimination are implemented by coherent addition of discrete harmonics of the fundamental gate frequency. With amplitude and phase control for each harmonic at the cathode, we form <340 ps bias gates, and with similar control at the anode we cancel the gate transient with >65 dB suppression, allowing avalanche-discrimination thresholds at the anode below 2 mV or <8 fC. The low threshold not only accurately discriminates diminutive avalanches but also achieves usable detection efficiencies with lower total charge, reducing the afterpulse probability and allowing the use of gate pulses that exceed the SPAD breakdown voltage by more than 10 V, both of which increase detection efficiency. With detection efficiency of 0.19 ± 0.01 , we measure per-gate afterpulse probability below 6.5×10^{-4} after 3.2 ns, and with detection efficiency of 0.51 ± 0.02 we measure per-gate afterpulse probability below 3.5×10^{-3} after 10 ns. © 2013 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4801939>]

Among the wide variety of applications of single-photon detectors, one of the most demanding is quantum communications, which puts a premium on low noise, high efficiency, and high speed. While many of these requirements can be met by cryocooled superconducting single-photon detectors,¹ InGaAs/InP single-photon avalanche diodes (SPADs) remain convenient and practical thermoelectrically cooled alternatives for the 950 nm to 1650 nm region. Unfortunately, InGaAs/InP SPADs are also notorious for significant afterpulsing due to charge traps,² an effect that generally limits them to gated-mode operation, and can require long recovery times (>1 μ s) and limit the detection rate in conventional detection systems,^{3,4} particularly when biased for high efficiency. Recently, sub-nanosecond periodic-gating techniques broadly classified as either sinusoidal gating^{5–8} or self-differencing^{9,10} have been developed for high-speed single-photon detection with InGaAs/InP SPADs, and have demonstrated dramatic improvements in detection rates. In these high-speed schemes, the periodicity of the gating signal facilitates suppression of the transient signal generated by the gate at the output of the SPAD, allowing the use of a low avalanche-discrimination threshold. The combination of low discrimination threshold and short gate duration significantly reduces the total avalanche charge, reducing afterpulsing. In this article, we describe a third approach to gate-transient suppression for high-speed periodic gating that is based on harmonic subtraction and that achieves low afterpulse probabilities and the highest detection efficiency ever reported in a detection system of this type.

The gate transient represents not only the capacitive response of the diode but also, due to the voltage dependence of the diode capacitance, a weak nonlinearity that mixes the gate spectrum. Sinusoidal gating schemes^{5–8} use a single

radio-frequency (RF) sine wave to gate the SPAD, and narrow band-stop filters at the 1st and 2nd harmonics (or alternatively a low-pass filter for the 2nd) to suppress the gate transient, in some cases below the floor set by thermal and amplifier noise.^{6,7} However, sinusoidal gating inextricably links the gate duration to the gate frequency and the over-voltage. For sub-nanosecond gates, it has been shown that the afterpulse probability reduces exponentially with the gate duration,¹¹ making it desirable to keep the gate as short as possible. Increasing the amplitude of the sinusoid and using ever sharper “peaks” as the gate can be done, but only to a limited practical extent. In addition, a band-stop filter at the gate frequency necessarily resides within the detection bandwidth and can distort the avalanche waveform and degrade signal discrimination.⁷

The self-differencing technique^{9,10} allows the use of essentially arbitrary waveforms, and can support short gate durations and high excess bias voltages (voltage above breakdown). However, for optimal cancellation of the gate transient, the attenuation and dispersion of the two delay lines used in this technique must match over the entire detection bandwidth.¹² Ultimately, the avalanche sensitivity is determined by the quality of this match, which becomes increasingly challenging as the gate duration is reduced and the bandwidth increases. Furthermore, this approach splits the avalanche signal between the two delay lines, effectively reducing the signal to noise ratio by 3 dB unless further efforts are made.

We use harmonic subtraction to suppress the gate transient. In this technique, the gate waveform is synthesized by summing harmonics of the gate frequency, resulting in a transient response composed of a larger set of harmonics that we cancel by destructive interference with reference sinusoids generated directly from the RF source, as illustrated in Fig. 1. The general idea of signal subtraction at the SPAD anode has been demonstrated in the context of a single

^{a)}Author to whom correspondence should be addressed. Electronic mail: alessandro.restelli@nist.gov

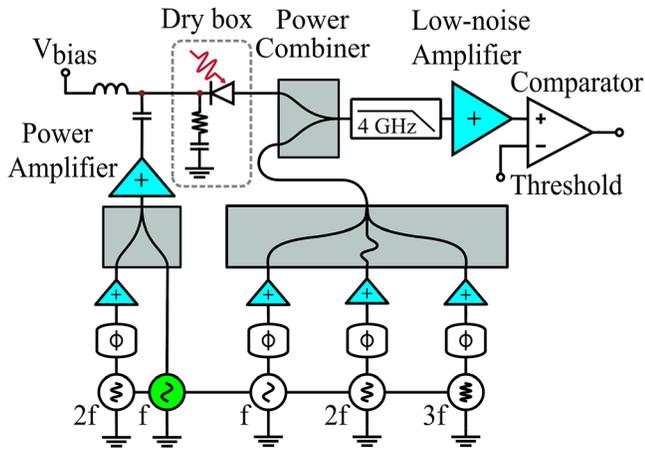


FIG. 1. Schematic of the detection system. The gate applied to the SPAD cathode is synthesized from the 1st and 2nd harmonics of the gate frequency ($f = 1.25$ GHz), and the gate transient is suppressed within the 4 GHz detection bandwidth by destructive interference with harmonics generated directly from the RF source.

sinusoid,^{7,13} and considered for more generalized scenarios.¹⁴ Forming the gate with a finite set of narrow-band spectral components allows high-quality transient cancellation (close to the thermal limit, as in some sinusoidal systems^{6,7}) even with relatively large-amplitude gate signals (e.g., 20 V), while simultaneously reducing the gate duration, as with self-differencing. With just two harmonics, the gate duration can be roughly 30% shorter than can be achieved with a purely sinusoidal gate of the same amplitude, resulting in a reduction of avalanche charge and afterpulsing. In addition, the signal subtraction is carried out in low-loss power combiners that efficiently preserve the avalanche waveform, maintaining the signal for discrimination.

Figure 1 illustrates the experimental setup. An oscillator generates the 1st harmonic at $f = 1.25$ GHz, and a passive frequency doubler generates a 2nd harmonic. These are combined with relative phase (ϕ) and amplitude (+) control and drive a 3 GHz power amplifier that produces the ≈ 20 V peak-to-peak gate waveform shown in Fig. 2(a). All measurements

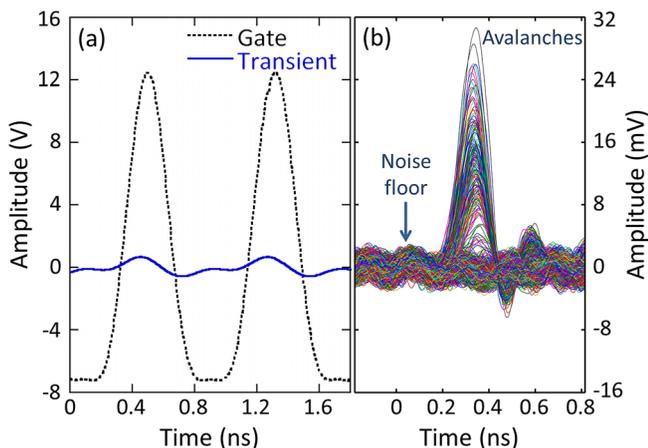


FIG. 2. (a) The gate waveform applied to the cathode, along with the resulting transient at the anode (with 50Ω termination). (b) Output signals after transient cancellation, referenced to the anode, illustrating the avalanche rise time and amplitude distribution. Here, higher harmonics are not suppressed and the noise floor is dominated by the oscilloscope and 10 GHz amplifier. $V_{\text{bias}} = 63.5$ V.

reported in this article are taken with this gate waveform. A bias-tee offsets the gate signal by V_{bias} to properly bias the SPAD for Geiger-mode operation. The SPAD is a commercially available fiber-pigtailed InGaAs/InP diode mounted inside a dry box with thermoelectric cooling.^{15,16} At our typical operating temperature of -22°C , the breakdown voltage is 65.1 V. When $V_{\text{bias}} = 63.5$ V, the voltage across the diode exceeds the breakdown voltage for 310 ps. Due to the slope of the gate edges, as V_{bias} is increased the gate duration increases by 20 ps/V. To reduce standing waves between the power amplifier and the high-impedance SPAD cathode ($\approx 500 \Omega$ at 1.25 GHz), we use a DC-blocked 50Ω load thermally isolated from the SPAD.

Figure 2(a) also shows the ≈ 1 V peak-to-peak gate transient at the anode that must be suppressed to discriminate small avalanche signals. The transient is dominated by the 1st harmonic, but has measurable contributions from many higher orders. To accurately discriminate avalanches, each spectral component in the gate transient is destructively interfered with a reference signal, as illustrated in Fig. 1. Fine control of the amplitude and phase of each harmonic component is necessary to achieve accurate cancellation, and when appropriately adjusted the first three harmonics in the gate transient can be suppressed by more than 65 dB. As an example of the resulting cancellation, and of the avalanche signal preservation, Fig. 2(b) shows 4092 waveforms acquired at the output of the power combiner. These measurements were acquired with a 10 GHz amplifier whose gain was divided out (note the vertical scale is in millivolts). In this measurement $V_{\text{bias}} = 63.5$ V, and the SPAD was illuminated at in every 16th gate by a < 30 ps, 1310 nm pulse with mean-photon number $\mu < 0.1$. Almost all avalanche events are distinguishable from the noise floor, which in this case is dominated by contributions from the amplifier and oscilloscope, though the 4th harmonic of the gate (5 GHz) is also discernible. The amplitude of the avalanche signals can be seen to vary widely, from the noise floor to several times larger. This distribution has direct bearing on detection efficiency and underscores the importance of a low discrimination threshold. Figure 2(b) also illustrates the benefit of preserving the avalanche signal. To be discriminated, an avalanche must trigger the comparator before the end of the gate; inhibiting high-frequency components in the avalanche signal may unduly suppress some signals, reducing the detection efficiency. We define our detection bandwidth to be < 4 GHz.

We achieve a lower noise floor and discrimination threshold by suppressing higher-order harmonics and noise outside the detection bandwidth with a 4 GHz low-pass filter and a low-noise amplifier (noise factor of 1.55), after the power combiner, as shown in Fig. 1. An additional asymmetric power splitter (not shown) after the low-noise amplifier monitors the transient suppression. In all, the output stage provides 16 dB of gain to the avalanche signal. When finely tuned, we measure the root-mean-square (RMS) of the noise floor at the output of the low-noise amplifier to be below 3.2 mV, or an apparent 0.51 mV RMS at the SPAD anode. For comparison, RMS Johnson noise across 50Ω at room-temperature is 0.057 mV in the same bandwidth. After the low-noise amplifier, an 8 GHz voltage comparator with

variable threshold detects avalanches for counting. Electronic noise makes no measurable contribution to the dark-count rate for discrimination thresholds down to 12.1 mV after the low-noise amplifier (1.92 mV at the anode), which we define as the minimum operative threshold. At this minimum threshold, and with the optical input blocked, the per-gate dark-count probability is $<10^{-5}$ when $V_{\text{bias}} \leq 62$ V, 2×10^{-5} at $V_{\text{bias}} = 63$ V, and rises to 7×10^{-5} at $V_{\text{bias}} = 65$ V. These dark-count probabilities are comparable to those of other high-speed periodically gated systems.

The detection efficiency is measured by illuminating every 16th gate ($f_L = 78$ MHz) with <30 ps, 1310 nm, $\mu = (0.1 \pm 0.005)$ attenuated pulses. The mean photon number is determined with an optical power meter with NIST-traceable calibration (2% uncertainty) and calibrated attenuators. The overall uncertainty is estimated to be 5%, mainly due to the reproducibility of the fiber connections. The delay of the optical pulse is adjusted to maximize the count rate in the illuminated gate, C_0 , measured with a gated counter (discussed below). The full width at half maximum of the count rate versus laser delay is 100 ps. The count rate in the gate just prior to the illuminated gate, C_{15} , is used as a measure of the contribution to C_0 from both dark counts and afterpulse events. We calculate the detection efficiency η as¹⁷

$$\eta = \mu^{-1} [\ln(1 - C_{15}f_L^{-1}) - \ln(1 - C_0f_L^{-1})]. \quad (1)$$

Figure 3 shows the detection efficiency as the comparator threshold is varied, for three different values of V_{bias} . It should be noted that InGaAs/InP SPADs are inherently more efficient at shorter wavelengths; at 1550 nm, the particular device we use is specified to have 80% of the quantum efficiency that it has at 1310 nm.¹³ The detection efficiency at 1310 nm rises above 0.5 when $V_{\text{bias}} = 65$ V. As can be seen in Fig. 2(a), at this bias, the peak of the gate pulse exceeds breakdown by approximately 12 V. As the threshold is raised through the distribution of avalanche amplitudes (cf. Fig. 2(b)), the detection efficiency traces a shape that can be accurately fit with the complementary error function (erfc; $\chi^2 < 0.17$ mV), as shown in Fig. 3, suggesting that the avalanche amplitudes have a Gaussian distribution.

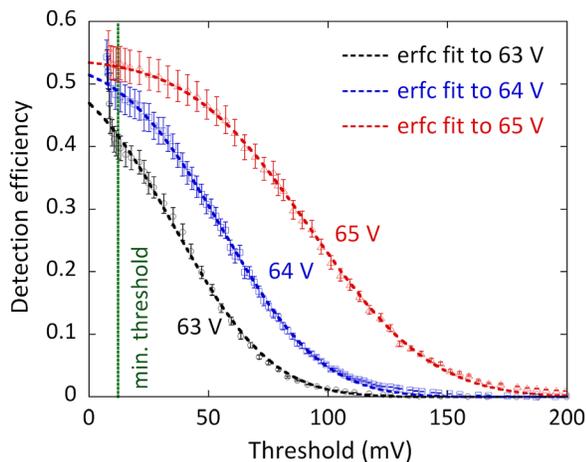


FIG. 3. Detection efficiency versus comparator threshold, and complementary-error-function fits, for three different bias voltages. The quality of the fit suggests that the avalanches have a Gaussian amplitude distribution. Error bars represent 5% measurement uncertainty.

To measure the minimum detectable avalanche charge in a manner that includes the response of the output stage and the comparator, we use the following approach. We measure the average avalanche charge by monitoring the average current through the SPAD with a picoammeter, and the total count rate in all gates. At the minimum threshold, and with $V_{\text{bias}} = 64$ V, the average avalanche charge is (36 ± 6) fC.

We then assume that the total charge in an avalanche scales linearly with the amplitude of the avalanche signal. Because the distribution of avalanche amplitudes appears to be symmetric, and, for $V_{\text{bias}} \geq 64$ V, almost all avalanches exceed the minimum threshold, we assume that the average avalanche amplitude is equal to the median avalanche amplitude. The median is determined by increasing the detection threshold until the total count rate has decreased by half, and is identified as the threshold required to detect an avalanche of average amplitude, and hence, of average charge. For $V_{\text{bias}} = 64$ V, the median threshold is 58.9 mV, which indicates that the 12.1 mV threshold detects avalanches with as little as (7.4 ± 1) fC, though this is likely an upper bound because undetected sub-threshold avalanches inflate the estimate.

We use the gated counter to measure the per-gate afterpulse probability as a function of time after an avalanche (up to 24.8 ns) by illuminating every 32nd gate with a <30 ps, 1310 nm attenuated pulse, and measuring the count rate in the illuminated gate and in each of the 31 non-illuminated gates in each 32-gate period. The gated counter, based on 10 GHz discrete logic, only counts events that occur within a temporal window defined by a pattern generator operating synchronously with the detector. In contrast to most time-interval analyzers and time-correlated single-photon counting systems, this gated counter imposes no hold-off or recovery time in which detection events may be missed, which is particularly important when making afterpulse measurements in high-speed detection systems. The afterpulse probability (AP_n) in each of the non-illuminated gates is calculated as $AP_n = (C_n - D)/C_0$, where C_n is the count rate in the n th gate, D is the average per-gate dark-count rate (i.e., the dark-count probability multiplied by 39 MHz), and C_0 is the count rate in the illuminated gate.

Figure 4 shows the per-gate afterpulse probability versus time for four different detection efficiencies (η), measured with the comparator threshold at the minimum. For all η , the per-gate afterpulse probability drops below 4×10^{-3} in less than 10 ns. For lower η , the afterpulse probability is significantly less; for $\eta = 0.19$, the per-gate afterpulse probability is in the 10^{-4} range less than 3.2 ns after an avalanche. For reference, the dark-count probability at this detection efficiency is $<5 \times 10^{-6}$. The afterpulsing rises sharply from $\eta = 0.43$ to $\eta = 0.51$, and reflects, among other things, the increase in average charge at higher bias voltages. As can be seen in Fig. 4, in most applications, a significant benefit would be gained by applying some degree of hold-off to the detection system to ignore gates with high afterpulse probability that occur shortly after an avalanche.

A preliminary investigation of high-speed counting shows that with the detector biased for 11% detection efficiency and illuminated in every 4th gate (312 MHz) by a pulse bright enough to induce counting in the illuminated

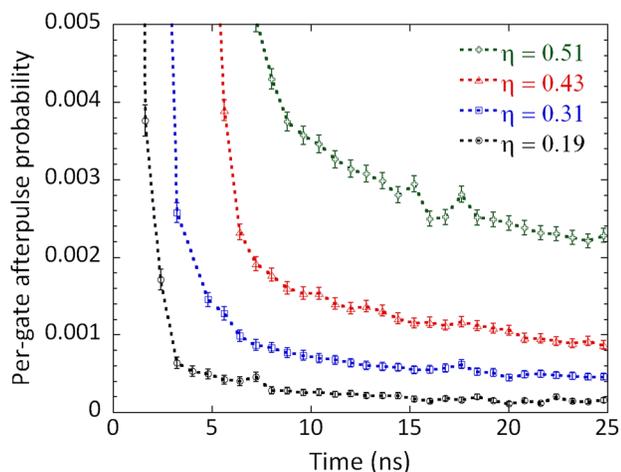


FIG. 4. (a) Per-gate afterpulse probability as a function of time from the illuminated gate, for four different detection efficiencies. Error bars represent the 3σ distribution.

gate at 100 MHz, the per-gate afterpulse probability measured in the non-illuminated gates is $<2 \times 10^{-3}$.

The approach described in this article achieves the high-quality gate-transient suppression observed in sinusoidal gating schemes, and allows the use of the short square-like gates used in self-differencing, while preserving the avalanche signal more efficiently than either approach. The increased detection efficiencies are a direct result of using higher excess bias voltages, which in turn are operable because of the reduced avalanche charge that results from the low discrimination threshold and short gate duration. All of these effects are the result of the system architecture based on narrow-band harmonics of the gate frequency, and underscore the importance of the complete bias and control system in the performance of a SPAD-based detector. We note that the ≈ 20 V gate signal brings the bias more than 7 V below breakdown; further investigation of the effects of this type of biasing is underway. It is also worthwhile to point out that

optimal use of the detection-system bandwidth would involve the same number of harmonics in both the gate synthesis and the transient cancellation; we estimate that the gate duration can be reduced by roughly half (<170 ps) by using up to the 4th harmonic at the cathode. Given the exponential relation between the gate duration and the total avalanche charge, such a reduction has the potential to result in significant benefits to high-speed single-photon counting.

This work was supported in part by the DARPA InPho program.

- ¹C. M. Natarajan, M. G. Tanner, and R. H. Hadfield, *Supercond. Sci. Technol.* **25**, 063001 (2012).
- ²M. A. Itzler, X. Jiang, and M. Entwistle, *J. Mod. Opt.* **59**, 1472 (2012).
- ³D. S. Bethune and W. P. Risk, *IEEE J. Quantum Electron.* **36**, 340 (2000).
- ⁴M. A. Itzler, X. Jiang, M. Entwistle, K. Slomkowski, A. Tosi, F. Acerbi, F. Zappa, and S. Cova, *J. Mod. Opt.* **58**, 174 (2011).
- ⁵N. Namekata, S. Sasamori, and S. Inoue, *Opt. Express* **14**, 10043 (2006).
- ⁶N. Namekata, S. Adachi, and S. Inoue, *IEEE Photonics Technol. Lett.* **22**, 529 (2010).
- ⁷Y. Liang, E. Wu, X. Chen, M. Ren, Y. Jian, G. Wu, and H. Zeng, *IEEE Photonics Technol. Lett.* **23**, 887 (2011).
- ⁸N. Walenta, T. Lunghi, O. Guinnard, R. Houlmann, H. Zbinden, and N. Gisin, *J. Appl. Phys.* **112**, 063106 (2012).
- ⁹Z. L. Yuan, B. E. Kardynal, A. W. Sharpe, and A. J. Shields, *Appl. Phys. Lett.* **91**, 041114 (2007).
- ¹⁰J. Zhang, R. Thew, C. Barreiro, and H. Zbinden, *Appl. Phys. Lett.* **95**, 091103 (2009).
- ¹¹A. Restelli, J. C. Bienfang, and A. L. Migdall, *J. Mod. Opt.* **59**, 1465 (2012).
- ¹²A. Restelli and J. C. Bienfang, *Proc. SPIE* **8375**, 83750Z (2012).
- ¹³J. C. Campbell, W. Sun, Z. Lu, M. A. Itzler, and X. Jiang, *IEEE J. Quantum Electron.* **48**, 1505 (2012).
- ¹⁴G. S. Kanter, U.S. patent 2011/0127415 A1 (2 January 2011).
- ¹⁵The identification of any commercial product or trade name does not imply endorsement or recommendation by the National Institute of Standards and Technology.
- ¹⁶Datasheet: Pigtailed Coaxial Single Photon Avalanche Diode, PGA-300, Princeton Lightwave, Inc., 2012.
- ¹⁷Y. Kang, H. X. Lu, Y. H. Lo, D. S. Bethune, and W. P. Risk, *Appl. Phys. Lett.* **83**, 2955 (2003).