Calibrating photon-counting detectors to high accuracy: background and deadtime issues

Michael Ware* Alan Migdall†, Joshua C. Bienfang‡, and Sergey V. Polyakov†,
*Brigham Young University, Provo, UT
†Optical Technology Div., NIST, Gaithersburg, MD
‡Electron and Optical Physics Div., NIST, Gaithersburg, MD

May 4, 2006

Abstract

When photon-counting detectors are calibrated in the presence of a background signal, deadtime effects can be significant and must be carefully accounted for to achieve high accuracy. We present a method for separating the correlated signal from the background signal that appropriately handles deadtime effects. This method includes consideration of pulse timing and afterpulsing issues that arise in typical avalanche photo diode (APD) detectors. We illustrate how these effects should be accounted for in the calibration process. We also discuss detector timing issues that should be considered in detector calibration.

Keywords: Photon counting; Quantum metrology; Single photon detectors;

1 Introduction

The photon pairs produced in parametric down-conversion provide a fundamentally absolute way to calibrate single photon detectors [1–14]. Because the photons are produced in pairs, the detection of one photon heralds with certainty the existence of the other. To measure detection efficiency, a trigger detection system is placed to intercept some of the down-converted light. The detector under test (DUT) is then arranged to collect all the photons correlated to those seen by the trigger detector (and usually more). In the ideal case, the DUT channel detection efficiency is the ratio of the number of coincidence events to the number of trigger detection events in a given time interval. (By ideal case here, we mean that other than the two-photon source, there are no competing mechanisms causing the detectors to fire and by coincidence we mean that the two detectors fire due to the two photons of a pair.) If we specify the collection efficiency of the DUT and trigger channels by η_{DUT} and η_{trig}, respectively, then the total number of trigger counts is

$$N_{\mathrm{trig}} = \eta_{\mathrm{trig}} N_{\mathrm{p}} \tag{1}$$

and the total number of coincidence counts is

$$N_{\mathrm{c}} = \eta_{\mathrm{DUT}} \eta_{\mathrm{trig}} N_{\mathrm{p}}, \tag{2}$$

where N_{p} is the total number of down-converted photons emitted into the trigger channel during the counting period. The absolute detection efficiency of the DUT channel is then simply

$$\eta_{\mathrm{DUT}} = \frac{N_{\mathrm{c}}}{N_{\mathrm{trig}}}. \tag{3}$$

Note that this is the efficiency of the entire detection channel (including collection optics, etc.) and not just the efficiency of the DUT alone [15, 16].

A number of groups have pursued detector calibration using this method (see Ref. [15] for a more detailed history). We are currently pursing an experiment where we hope to achieve calibration uncertainties on the 0.1% level. To achieve this level of uncertainty, we must carefully consider how the idealized calibration setup described above is actually implemented in the laboratory. In this article we describe some subtle effects that arise from non-ideal measurement devices, and illustrate how they can be handled to achieve low uncertainties.

An accurate experimental determination of N_{trig} is usually straightforward: the electronic detector signals are summed during the counting period, and then darkcounts and counts due to background photons are estimated and subtracted off. Experimental techniques for making these estimates are detailed in Ref. [16].

Determining N_{C} accurately requires significantly more effort. In a typical calibration setup, N_{C} is determined by making a histogram of the delays between trigger and DUT detection events (see Figure 1a). In this setup the photons produced simultaneously are sent to two avalanche photodiode (APD) photon-counting modules, one designated as trigger and one as DUT. The trigger and DUT detector module outputs are sent to the start and stop inputs respectively of a time digitizing circuit that records the arrival time of each pulse. The electronic signal of the DUT is delayed with an appropriate length of cable to assure that that the DUT (stop) pulse arrives at the time digitizing circuit after the trigger (start) pulse.

The histogram in Fig. 1a records the delay between each DUT event and the most recent trigger event. The principal coincidence peak (A) is evident in the histogram at a delay of 55 ns, and to first approximation, the number of coincidences N_{C} can be estimated as the sum of events in this peak, with the background subtracted off. However, a more accurate determination of the number of coincidences requires an understanding of the smaller peaks (E) at 13 ns, (F) at 34 ns, and (D) at 98 ns. Moreover, it is not immediately clear what background to subtract because the levels are different on either side of the main peak. The difference in background levels is due to detector deadtime, and becomes more pronounced in situations where the DUT has high detection efficiency and the background count rate is high. In these situations, the simple background subtraction technique introduces small errors because it does not account for deadtime effects. In this paper we illustrate how to properly handle histogram data when calibrating high efficiency detectors in the presence of

high background rates. While the corrections to the simple background subtraction technique are small, they can be relevant when calibrating detectors to high precision. We also discuss timing and afterpulsing issues that arise when considering a histogram like the one in Fig. 1a.

2 Histogram Features

To correctly calibrate the detector, we must first consider all of the features (peaks, backgrounds, valleys, etc.) and determine their origins. Then we can separate 'heralded' photon events from background events. To illustrate this process, we consider the data in Fig. 1a. The main coincidence peak (A) is the most prominent feature of the histogram and obviously represents correlated events. The origin of the small 'shoulder' (labeled B) on the right the main coincidence peak is a little less obvious. In section 4 we will see that this shoulder represents valid, but delayed, coincidence counts, and we will discuss why they are delayed. The broad dip (labeled C) to the right of the peak is due to detector deadtime. Since the detector is most likely to receive a photon at a delay of ≈ 55 ns, it is also likely to be 'dead' for a finite duration at subsequent delays and thus is unable to register the arrival of background photons. This behavior complicates the process of separating background events from correlated events. In section 3 we discuss how to to properly separate the two types of events in the presence of deadtime.

Peak D is due to afterpulsing in the DUT. The APDs in our setup are actively quenched (i.e. the bias voltage is reduced to below the APD breakdown voltage) for ≈ 42 ns after an event is registered. When the quenching ends, there is a small probability that the detector will produce a false pulse (referred to as an afterpulse). The size of this peak indicates that an afterpulse is registered for $\approx 1\%$ of the pulses in the main coincidence peak. We will discuss the structure of the afterpulse in section 4.

The tiny peaks labeled E and F are specific to our particular setup. Peak E is produced by afterpulsing of the trigger detector. Because we have delayed the arrival of the DUT pulses ≈ 55 ns longer than the trigger pulses, the afterpulses of the trigger arrive before the correlated DUT pulses. Because we measure the delay between DUT pulses and the most recent trigger pulse, this results in a small coincidence peak 42 ns before the main peak. These are valid coincidences, although they make only a tiny correction to the final detection efficiency ($\approx 0.1\%$ of peak A) in the situation shown. This peak could have been eliminated by simply adjusting the delay lines so that the main coincidence peak occurred at a delay less than 42 ns (or ignoring trigger pulses that occur too close to the previous trigger). We chose to leave the setup as shown to help illustrate the other effects discussed below.

Peak F is also simply understood. In our setup, we couple trigger photons into a fiber and then couple the output of the fiber onto the trigger APD. There is a small back-reflection at either end of the fiber, and photons that experience a back-reflection at both ends of the fiber arrive at the APD after

traveling two extra fiber lengths. Our fiber is 2 m long and this extra double-pass corresponds to a delay of ≈ 20 ns. Again, the coincidences in this peak are valid, and account for $\approx 0.1\%$ of the total coincidence counts. Eliminating, or antireflection coating, the fiber in the trigger detection path would remove this peak, but we left it in our setup for convenience and to illustrate things that should be considered in a high accuracy detector calibration.

3 Separating Background from Signal

When high efficiency detectors are calibrated in situations with high background count rates, the simple background subtraction method discussed earlier introduces errors in the estimation of detector efficiency. For a more accurate treatment, we use a probabilistic treatment of the data to separate background events from events correlated to the trigger.

We begin with a measured histogram denoted by $H(\tau_i)$, where τ_i denotes the time delay for the i^{th} bin. As usual, the histogram records the cumulative number of detection events received in each bin. The number of histogram trigger pulses (i.e. histogram starts) received while accumulating the histogram is specified by T. Our goal is to separate the events represented by $H(\tau_i)$ into two categories: events correlated to the trigger pulse and background events. Formally, we write

$$H(\tau_i) = C_{\text{m}}(\tau_i) + B_{\text{m}}(\tau_i), \tag{4}$$

where $C_{\text{m}}(\tau_i)$ and $B_{\text{m}}(\tau_i)$ denote, respectively, the number of correlated and background events recorded in a given bin. The subscript 'm' reminds us that these quantities represent the number of correlated and background events that were actually *measured* in each bin. In a calibration, efficiency can be defined a number of ways. Here, we define it as how many correlated events *would have been measured* if the detector were 'live' (ready for a detection) each time a correlated photon was incident on the system. This number will be larger than the simple estimate of efficiency C_{m}/T, since the detector may be dead when the correlated photon arrives due to a previous background count. In addition, the background rate may not be a simple constant due to previous background or correlated events (recall that this is the origin of the dip labeled C in Fig. 1(a)).

We consider a simple model of detector deadtime where the detector is 'dead' for d time bins after registering an event (where d is an integer), and then immediately becomes 'live' again. For ease of notation, we will assume that the delay of the DUT channel has been chosen long enough that all of the correlated signal appears after the d^{th} histogram bin.[1] This requires that

$$C_{\text{m}}(\tau_i) = 0 \qquad (i \leq d). \tag{5}$$

[1] The specification that all correlated signal occurs after d bins is, of course, artificial and introduced only for notational convenience. In practice, it is usually easier just to record the correlated signal and then pad the leading portion of the histogram with some extra background rather than using extra long DUT delays.

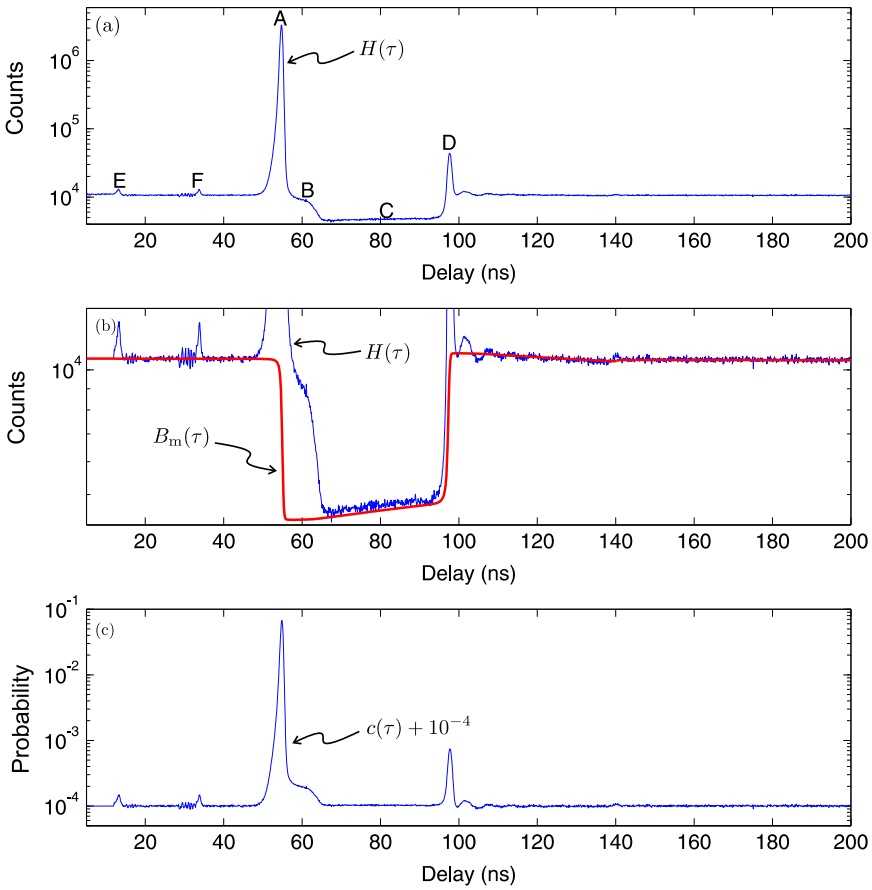

Figure 1: (a) Typical histogram $H(\tau)$. The main features are A–the main peak due to coincident photons, B–extended shoulder due to delayed output of APD, C–the region where the detector is dead after firing, and D–peak due to afterpulsing. E and F are minor effects discussed in the text. (b) The calculated background $B_{\rm m}(\tau)$ (heavy line) together with the measured histogram. (c) The probability of getting a correlated photon in each bin, $c(\tau)$, offset by 10^{-4} (so it can be plotted on a log scale).

We also assume that the mean number of background events in a given bin is constant before correlated events arrive, and define B_0 as the average number of background events/bin recorded in bins with $i < d$. Using B_0, we can calculate the probability b that a background event could be measured in a bin (during a given scan) assuming that the detector was alive at that moment. In the absence of correlated signal, this is simply

$$B_0 = b \times (T - dB_0). \tag{6}$$

The value in parenthesis of Eq. (6) gives the total number of times that it was possible to record an event in the bin (i.e. the total number of scans minus the number of scans where the detector was dead due to a background event in one of the previous d bins).

We can also write an expression for b in the presence of correlated signal. In this case we have

$$B_{\rm m}(\tau_i) = b \times \left[T - \sum_{j=i-d}^{i-1} H(\tau_j) - \Delta(\tau_i) \right] \qquad (i \geq d). \tag{7}$$

As before, the number in parenthesis is the number of scans where it was possible to record a background event in the bin. The sum gives the number of times that the detector was dead in the $i^{\rm th}$ bin due to a background or signal events in one of the previous d bins. The function $\Delta(\tau_i)$ accounts for the situations where a trigger pulses resets the delay before it reaches the $i^{\rm th}$ bin. (Recall that we histogram DUT events with respect to the most recent trigger event.) Since the trigger detector also experiences deadtime, $\Delta(\tau_i)$ is equal to zero for the first d bins of the histogram. After that, $\Delta(\tau_i)$ grows with τ_i (linearly to first approximation). It is usually a simple matter to extract $\Delta(\tau_i)$ from the measured histogram.

We can combine Eqs. (6) and (7) to write a convenient expression for $B_{\rm m}(\tau_i)$ in terms of measured quantities:

$$B_{\rm m}(\tau_i) = \frac{B_0}{T - dB_0} \times \left[T - \sum_{j=i-d}^{i-1} H(\tau_j) - \Delta(\tau_i) \right] \qquad (i > d). \tag{8}$$

To evaluate this equation, we need $B_{\rm m}(\tau_i)$ for $i \leq d$. Since we have specified that there is no correlated signal and $\Delta(\tau_i) = 0$ in this region of the histogram, we have $B_{\rm m}(\tau_i) = H(\tau_i) \approx B_0$, $(i \leq d)$. Figure 1b shows $B_{\rm m}(\tau)$ calculated for the histogram in Fig. 1a together with $H(\tau)$ for comparison. Note that our definition of b does not distinguish between a background event and an afterpulse of a background event, which results in slight underestimation of the background during the dead region in Fig. 1(b), however such underestimation results in $\approx 0.03\%$ error in the final overall detector efficiency.[2]

[2]To introduce a simple correction for afterpulses due to uncorrelated events, one could assume that the number of afterpulses in the histogram bin i is proportional to the number of events in the bin $(i - d)$ with a constant probability of an afterpulse β, i.e. by adding $\beta H(\tau_{i-d})$ to the righthand sides of Eqs. (6) and (7); β can be determined as a ratio of peaks D and A.

We can use the background found using Eq. (8) together with Eq. (4) to determine the number of correlated events measured in each bin (i.e. C_{m}). Then, we use C_{m} to calculate the probability $c(\tau_i)$ that a correlated photon would register in a given time bin, assuming that the DUT was live during that time bin:

$$c(\tau_i) = \frac{C_{\mathrm{m}}(\tau_i)}{T - \sum_{j=i-d}^{i} B_{\mathrm{m}}(\tau_j) - \Delta(\tau_i)}. \tag{9}$$

Since at most one correlated photon arrives in any given scan, the DUT will never be dead in a given bin due to a correlated signal in one of the previous d bins. (It is possible to have more than one correlated event per scan in cases where the DUT afterpulses, but the first correlated event will not cause the DUT to be dead for the second event.) Thus, the sum in the denominator only needs to account for the number of times the DUT was dead due to a background event received in one of the previous d bins. The function $\Delta(\tau_i)$ is included again since trigger pulses can reset the scan before a correlated event is recorded in a given bin. Figure 1c shows $c(\tau_i)$ for the histogram in Fig. 1a.

The function $c(\tau_i)$ gives a time resolved description of the DUT response to an input photon. A simpler measure of the DUT detector response is its detection efficiency η_{DUT}, which gives the probability that the detector will produce an output pulse from an input photon (without worrying about when the output pulse will occur). A first approximation to this measure is obtained by summing the probabilities $c(\tau_i)$ for each bin

$$\eta'_{\mathrm{DUT}} = \sum_i c(\tau_i), \tag{10}$$

where it is assumed that all time bins with any significant probability of registering the heralded pulse are included in the sum (e.g. for $10 < \tau < 70$ in our example) and after pulsing features are excluded. To get a more accurate representation of the detection efficiency, we must correct for background events in the trigger channel. To do this we need to calculate the probability that the trigger pulse was due to a heralding photon, as opposed to a background photon or dark count in the trigger channel. We denote this probability by t_{h}. The detector channel efficiency is then given by

$$\eta_{\mathrm{DUT}} = \frac{\eta'_{\mathrm{DUT}}}{t_{\mathrm{h}}} \tag{11}$$

To extract the efficiency of the DUT alone, the reflective and absorptive losses due to any optics in the channel path must be accounted for, as detailed in Ref. [16].

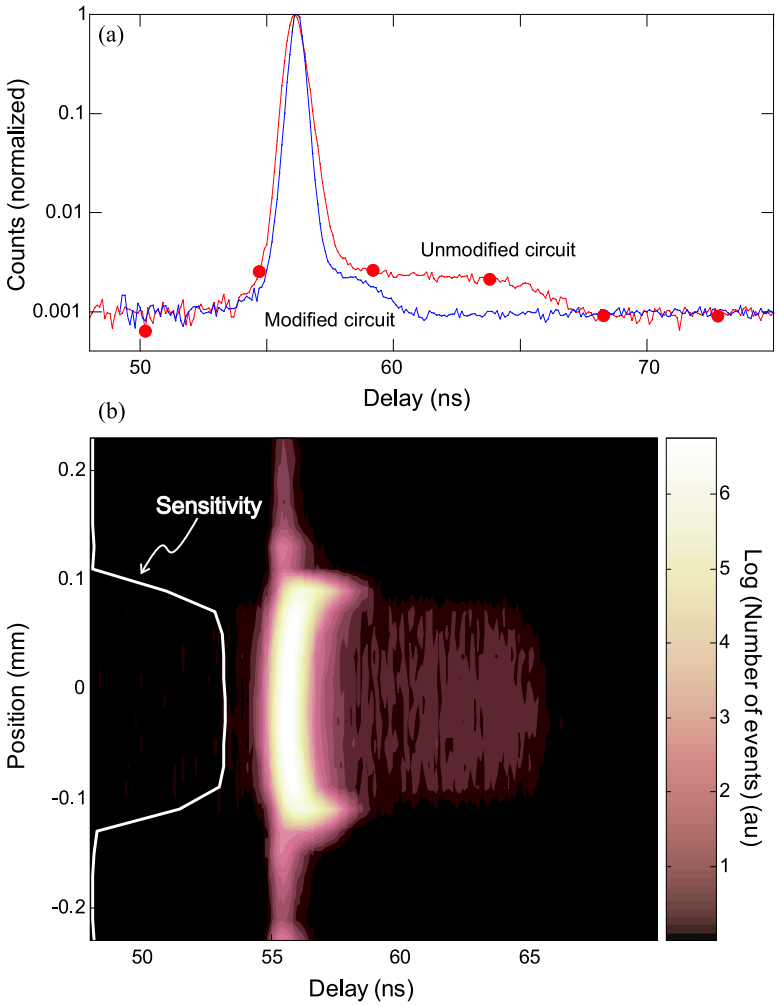

Figure 2: (a) Comparison of the histograms taken with standard and modified APD circuitry and (b) Histograms of APD with unmodified circuitry taken as the light spot is scanned across the diameter of the APD. Peak shifts of up to ≈2 ns are seen at the edges of the APD active region. The shoulder feature of (a) is also seen in (b) extending for ≈10 ns. The white curve (on left side) indicates the sensitivity of the detector versus position.

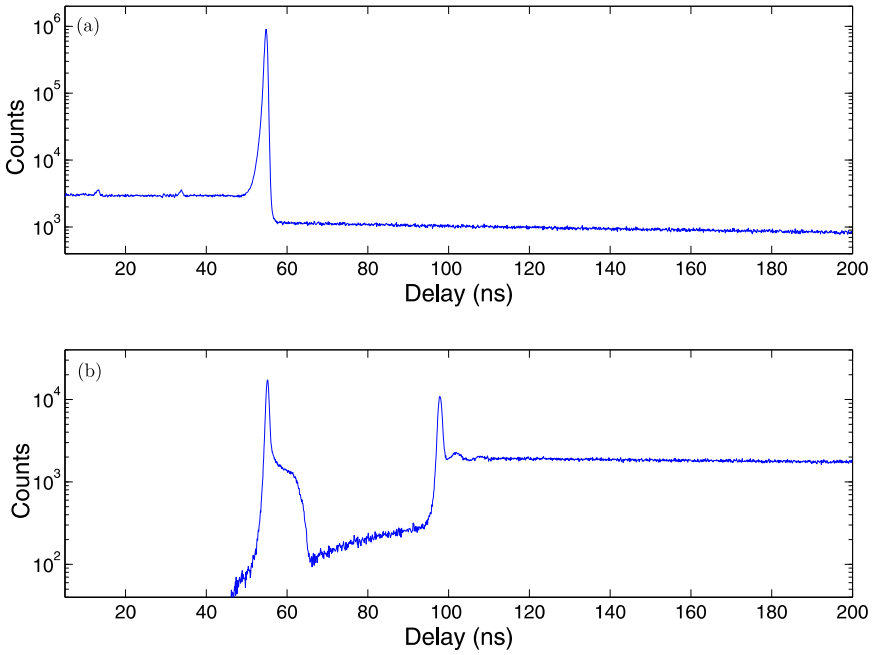

Figure 3: Histograms of first (a) and second (b) DUT events after trigger. These data sets are the two largest components making up the results of Fig. 1a. The reduced (and falling) background to the right of the peak in the upper panel results from the decreasing likelihood of a first event as earlier first events accumulate.

4 Tardy Detection Events and Increased Afterpulse Probability

Looking at the correlated response function, $c(\tau)$, in Fig. 1c, one might wonder about the origin of the small shoulder to the right of the main peak. Figure 2a shows a closeup of the shoulder. It turns out that this shoulder accounts for more than 1% of the total signal which is significant for a high-precision calibration. Therefore it is important to understand and treat these 'tardy' events correctly. A first guess might be that these events are due to photons that strike the outside rim of the active area of our APD where it is slower to respond, as edge effects often result in timing anomalies. To measure this effect, we focused the DUT light to a tight spot and scanned it across the APD (see Fig. 2b). We found that the 'rim' detections do indeed experience an extra delay. However, the detector response is delayed by no more than an additional ≈ 2 ns relative to the delays of events caused by photons closer to the center of the active area. Thus, this extra delay is not sufficiently long to produce the shoulder observed in Fig. 1c.

To get more information about these tardy pulses, we separated the histogram events by the order in which they were received after the trigger (i.e. the first event after a trigger, the second event after the trigger, etc.). Figure 3a shows the distribution of first events, and Fig. 3b shows the distribution of second events. As would be expected, the main coincidence peak consists primarily of first events, and the afterpulse peak is composed of second events. Note that the shoulder of tardy coincidence events (located at 57–62 ns) also consists of second DUT events. Since the DUT deadtime is ≈ 42 ns, this means that the tardy coincidence events were registered right after the detector had recovered from a previous avalanche recorded during the first 15–20 ns of the histogram. The first avalanche in this case is due to a background (as opposed to a correlated) event. This suggests that the timing of DUT pulses can be influenced by previous detection events (if they were recent enough). We now consider how the detector reacts when it receives a photon during its recovery period.

It turns out that the afterpulse probability is not a constant, but can depend on the background rate. This can be explained in terms of the active quenching circuitry of the detector. After a detection event, the active quench circuit of the DUT lowers the APD bias voltage to just below breakdown. After this quench period, the circuit raises the bias back to its original value. At this point the APD can avalanche again, and sometimes does so spontaneously causing an 'afterpulse'. However, if the detector receives a photon during the time of rising bias, it significantly enhances the probability that the detector will avalanche once the bias returns to its higher value. For higher background rates, it is more likely that a photon will strike the detector during this reset phase and thus increase the likelihood of afterpulses. This is why the probability of getting an afterpulse shown in Fig. 1(c) is significantly higher than expected from the APD specification of 0.3%. Interestingly, these 'twilight' pulses, which are triggered by photons arrived during the last moments of the deadtime, are delayed, sometimes significantly, as compared with 'normal' detections.

To demonstrate this effect, we sent pairs of optical pulses separated by a delay to an APD module and fed the output of the module to a time digitizing circuit. If the first pulse registered in the detector, it would start a histogram scan. If the detector produced a second pulse, it would stop the scan. In Fig. 4 we show a number of histograms for varying delay times between the two optical pulses. (The APD used for this measurement has a deadtime of 49.5 ns. This is different from the APD used in Figs. 1 to 3.) We find that if the photon comes during first ≈ 40 ns of the deadtime, i.e. while the active quench circuit has the bias below breakdown, it has no effect on the afterpulse. In such case, the afterpulse has a bimodal shape shown by the thick black line in Fig. 4a (previously observed in Ref. [17]). If the two light pulses are separated by more than ≈ 40 ns, so the second pulse arrives while the APD bias was rising back above breakdown, the probability of an afterpulse is increased and we see a distinct new peak in the afterpulse shape, which can be delayed by up to 10 ns from the time the APD would normally generate a TTL pulse. The data shows that the closer the photon arrival time is to the end of the dead zone, the shorter the observed delay (see the upper curve in Fig. 4). It is the spread in

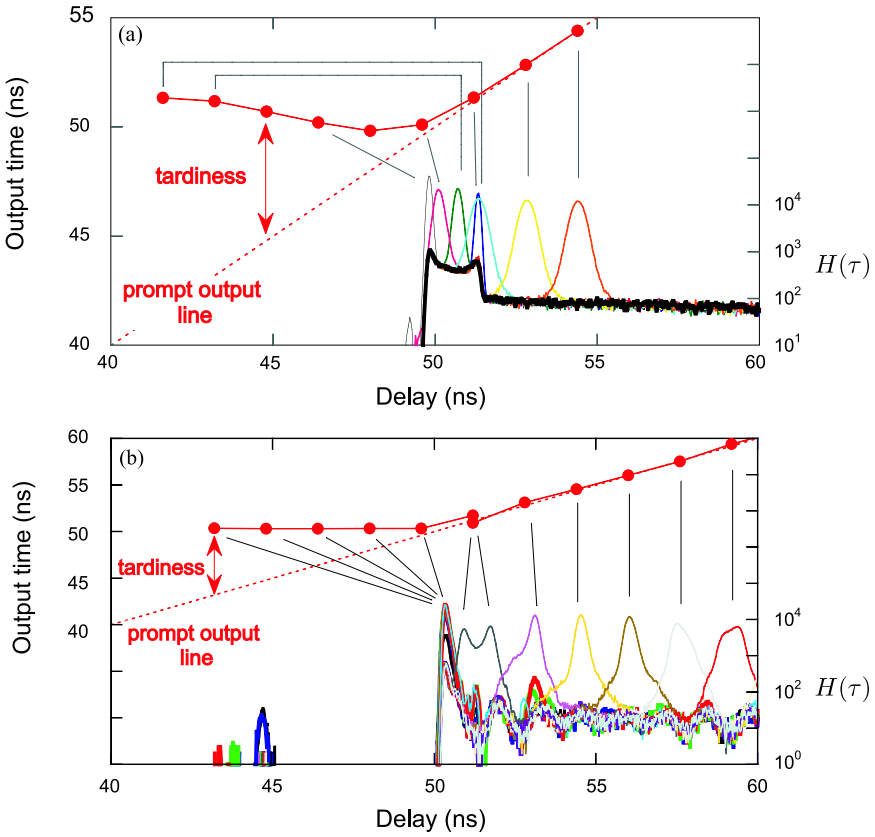

Figure 4: High time resolution histograms of afterpulse feature (cf. D of Fig. 1(a)) taken with photons arriving with a range of delays after registering a previous photon taken with two APD systems. One APD system with (a) unmodified and one with (b) modified circuitry. Thick line in (a) shows the afterpulse feature with no second photon received during the histogram. Peaks (thin lines and right axis) show histogram counts due to second photon arriving in the time region when the detector is recovering from previous pulse. Upper curves (dots and left axis) show the time of the pulse output versus the output time of a pulse when the detector is not recovering from a recent detection. Undelayed or prompt output would fall on the dashed line. Tardiness is indicated by the difference between the data and the dashed line. For the modified circuit (b) all 'twilight' pulses are emitted at the same time in the reset cycle, while with the unmodified circuit (a) the emission times vary nonmonotonically with when the photon was received. For this figure only, an APD with 49.5 ns deadtime was used.

these delays that sets the width of the shoulder feature of Figs. 1 to 3.

Interestingly, this feature of the detector can be eliminated with a different detection circuit. [18] We tested a detector with a modified avalanche detection circuit and found that it significantly reduced the late arrival features. With that circuit, in Fig. 4 (b) we see the bimodal afterpulse shape has collapsed to a single peak which is consistent with the flat output time line for those photons received during the twilight period. This results in the reduction of tardy events as can be seen in Figure 2 (a) which compares a histogram taken with the modified detector circuit, and one taken with the original active quench circuit.

5 Conclusion

We have illustrated how to accurately handle correlated histogram data in calibration situations with a high probability of getting a correlated photon in a scan, and also a non-negligible background rate in the presence of deadtime. We also illustrated how other phenomena, such as enhanced detector afterpulsing (and the associated tardy detector output) are due to a high background rate. While these effects are small, they are important to consider when doing high accuracy calibrations of single-photon-counting detectors. While one may define a detection efficiency as the ideal probability of a live detector producing a count (and only one count) due to a single incident photon, actual implementations are not so clean. In particular, we point out that due to deadtime, tardy, and afterpulse effects, the effective detection efficiency of a photon-counting detector is time and history dependent and the final result is an average over a specific set of conditions. So in cases where a high accuracy overall detection efficiency is required, it is important to understand exactly what those calibration conditions are and how to apply that result in other situations.

This work was supported in part by the Disruptive Technology Office (DTO) under Army Research Office (ARO) contract number DAAD19-03-1-0199, and DARPA/QUIST.

References

[1] Louisell, W. H., Yariv, A., and Siegman, A. E., 1961, *Physical Review*, **124**, 1646–1654.

[2] Zernike, F. and Midwinter, J. E., 1973, *Applied Nonlinear Optics*, (New York: Wiley).

[3] Burnham, D. C. and Weinberg, D. L., 1970, *Physical Review Letters*, **25**, 84–87.

[4] Klyshko, D. N., 1977, *Soviet Journal of Quantum Electronics*, **7**, 591–594.

[5] Klyshko, D. N., 1981, *Soviet Journal of Quantum Electronics*, **10**, 1112–1116.

[6] Malygin, A. A., Penin, A. N., and Sergienko, A. V., 1981, *Soviet Journal of Quantum Electronics* **11**, 939–941.

[7] Bowman, S. R., Shih, Y. H., and Alley, C. O., 1986, The use of geiger mode avalanche photodiodes for precise laser ranging at very low light levels: An experimental evaluation, Proceeding of SPIE, 663, 24–29.

[8] Rarity, J. G., Ridley, K. D., and Tapster, P. R., 1987, *Applied Optics*, **26**, 4616–4619.

[9] Penin, A. N. and Sergienko, A. V., 1991, *Applied Optics*, **30**, 3582–3588.

[10] Ginzburg, V. M., Keratishvili, N., Korzhenevich, E. L., Lunev, G. V., Penin, A. N., and Sapritsky, V., 1993, *Optical Engineering*, **32**, 2911–2916.

[11] Kwait, P. G., Steinberg, A. M., Chiao, R. Y., Eberhard, P. H., and Petroff, M. D., 1994, *Applied Optics*, **33**, 1844–1853.

[12] Migdall, A., Datla, R., Sergienko, A., Orszak, J. S., and Shih, Y. H., 1995/6, *Metrologia*, **32**, 479–483.

[13] Brida, G., Castelletto, S., Degiovanni, I. P., Novero, C., and Rastello, M. L., 2000, *Metrologia*, **37**, 625–628.

[14] Brida, G., Castelletto, S., Degiovanni, I. P., Genovese, M., Novero, C., and Rastello, M. L., 2000, *Metrologia*, **37**, 629–632.

[15] Ware, M. and Migdall, A. L., 2004, *Journal of Modern Optics*, **15**, 1549–1557.

[16] Migdall, A. L., 2001, *IEEE Transactions on Intrumentation and Measurement*, **50**, 478–481.

[17] Spinelli, A., Davis, L. M., and Dautet, H., 1996 "Actively quenched single-photon avalanche diode for high repetition rate time-gated photon counting" *Rev. Sci. Instrum.* **67**, 55–61.

[18] Cova, S., Ghioni, M., and Zappa, F. "Circuit for high precision detection of the time of arrival of photons falling on single photon avalanche diodes," US Patent 6384663.