Statistics of Photon Arrival Time

Tom Murphy

February 2, 2001

In an experiment designed to measure the time of arrival of individual photons, it is of critical importance to understand the behavior of the system in the event that more than one "detectable" photon arrives at the detector element. This document explores some of the characteristics and relationships associated with the recorded photon arrival time. The analysis is in the context of discrete pulses of light impinging on a detector, with the overall aim being to assess a characteristic time of arrival (e.g., centroid) for the pulse.

1 Photon Multiplicity

The detectors to be used for APOLLO are avalanche photodiodes (APDs) arranged in an integrated array. Discussing for now the behavior of a single detector element, the APD is only sensitive to the first photon that triggers an avalanche. This photon first deposits energy into the detector material, creating an electron-hole (e-h) pair. The electron drifts toward the multiplication region of the APD, where it picks up speed and collisionally produces more e-h pairs, each of these producing yet more collisional e-h pairs, etc. Therefore, a photon is regarded as "detectable" if in a statistical sense it is likely to trigger this avalanche reaction. The detectability of a photon can be characterized by P_d, the photon detection efficiency, which can be broken into two parts: the quantum efficiency, Q, of the material, describing the likelihood of an incident photon producing that first electron-hole pair, and P_e, the photoelectron detection efficiency. The latter term describes the probability that a photoelectron (from the photo-generated e-h pair) succeeds in stimulating an avalanche. There exists a finite probability that the chain of ionization dies before it really gets going, due to a fluctuation to zero carriers. If the APD is flooded with photons all in the same pulse, only the first to trigger the avalanche is recorded, as the follow-ons at most participate in the same saturation event.

How many "detectable" photons are incident on the detector? A single event is incapable of answering this question, but a series of detections/non-detections can address this question statistically. Assuming the incident flux is steady from one pulse to the next, let us say that a small fraction, ε, of the pulses result in a positive detection, so that to rough approximation, the chances of detecting a single photon is $p_1 \approx \varepsilon$. If $\varepsilon \ll 1$, then the chances that two photons participated in the detection must be roughly $p_2 \approx \varepsilon^2$. The likelihood of no detection, by definition, is $p_0 \equiv 1 - \varepsilon$. Putting this together, and allowing for some arbitrary "multiplicity" factor, a_n, we must require that the sum of all possibilities equals one:

$$\sum_{n=0}^{\infty} p_n = (1-\varepsilon)(1 + \varepsilon + a_2\varepsilon^2 + a_3\varepsilon^3 + \ldots) = 1. \tag{1}$$

Where $a_0 = 1$ to match our definition of ε, and $a_1 = 1$ owing to the fact that the multiplicity for a single photon must be one, and also because with ε small, a single photon detection must make up almost all of the ε positive detections. The a_n coefficients are all obtained automatically upon the realization that

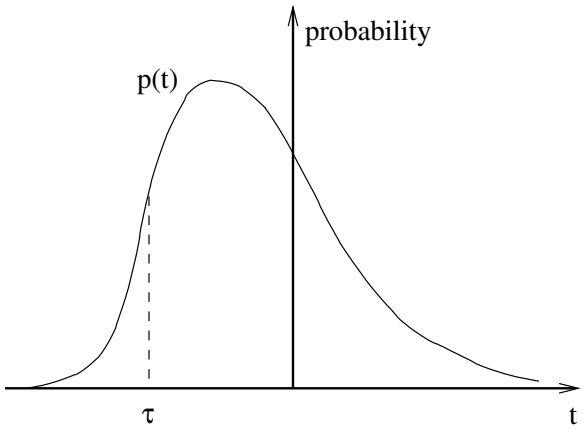

Figure 1: A pulse profile may take on the appearance above, which can be regarded as a probability function of arrival time of a random photon within the pulse.

the second term in parentheses above must equal $1/(1-\varepsilon)$, the Taylor expansion for which reveals that all $a_n = 1$. Therefore,

$$p_n = (1-\varepsilon)\varepsilon^n. \tag{2}$$

The expectation value for this probability distribution is found to be,

$$\langle n \rangle = \frac{\varepsilon}{1-\varepsilon}. \tag{3}$$

Thus if one expects 0.1 photons per pulse, a non-detection will result 90.9% of the time, a single-photon detection 8.26% of the time, a two-photon detection 0.75% of the time, etc. Easier numbers arise out of the inverse problem: positive detections are recorded 10% of the time ($\varepsilon = 0.1$). In this case, 9% are single-photons, 0.9% are doubles, 0.09% are triples, etc. Clearly, in order to work in the single-photon regime one wants to work at very low $\langle n \rangle$. If $\langle n \rangle = 1/3$, for example, $\varepsilon = 0.25$, and 25% of all positive detections are multiple-photon detections.

2 First-Photon Probability Distributions

An arbitrary pulse profile can be represented as a probability function with respect to time. The interpretation of this probability function in the limit of very low photon number ($N \ll 1$) is that $p(t)dt$ is the probability that when a photon *is* recorded, it appears at time between t and $t + dt$. Figure 1 illustrates the idea.

Working in the limit of very low photon detection rate (1%, say), the ensemble of reported arrival times for a large number of pulses would take on the shape of the temporal pulse profile, which is also the probability function, $p(t)$. If, however, each pulse contains many detectable photons, the reported time for each pulse arrival is skewed to early times. There is then a corresponding probability distribution describing the likely arrival time of the first photon. The probability density of a photon arriving at a time, τ, as indicated in Figure 1, must be proportional to $p(\tau)$, and also to the probability that $N-1$ photons are crammed into the pulse at a time later than τ. There is also a factor of multiplicity in

that any of the N photons within the pulse can be the first photon. Thus the probability distribution describing the arrival time of the first photon is

$$p_{\text{first}}(\tau) = p(\tau)N[1 - P(\tau)]^{N-1}, \tag{4}$$

where

$$P(\tau) \equiv \int_{-\infty}^{\tau} p(t)dt \tag{5}$$

is the cumulative probability of the pulse profile, with $P(-\infty) = 0$ and $P(\infty) = 1$. One sees that in the case of $N = 1$, the original probability distribution describing the pulse profile is recovered.

2.1 Example: Uniform Distribution

As an example of the way in which the probability distribution for the first photon becomes skewed, let's examine the effect on a square pulse pattern, defined as:

$$p(t) = \begin{cases} \frac{1}{T} & 0 \le t \le T \\ 0 & \text{otherwise} \end{cases},$$

where T is the pulse duration. $P(t)$ is simply t/T within the pulse, zero before, and one after. Thus the first photon arrival time distribution becomes:

$$p_{\text{first}}(t) = \begin{cases} \frac{1}{T}N(1 - \frac{t}{T})^{N-1} & 0 \le t \le T \\ 0 & \text{otherwise} \end{cases}.$$

For $N = 1$ this is just the original uniform distribution. For $N = 2$ it is a linearly descending function beginning at $2/T$ at $t = 0$, and ending at zero at $t = T$. In the general case, the function is a simple N^{th} order curve starting at N/T at time $t = 0$ and descending more and more rapidly to zero as the photon number is increased.

2.2 Example: Gaussian Distribution

A more universal example may be the Gaussian pulse profile, which is prevalent among short-pulse systems. Here, the pulse (probability) distribution is described by:

$$p(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}},$$

with the full-width at half-maximum (FWHM) equal to $\sigma \times \sqrt{\ln 256}$. The cumulative probability function can be put in terms of the "error function":

$$P(t) = \frac{1}{2} + \frac{1}{2}\text{erf}\left(\frac{t}{\sqrt{2}\sigma}\right),$$

with

$$\text{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du.$$

This leads to the first-photon probability distribution:

$$p_{\text{first}}(t) = \frac{N}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} \left[\frac{1}{2} - \frac{1}{2}\text{erf}\left(\frac{t}{\sqrt{2}\sigma}\right)\right]^{N-1}.$$

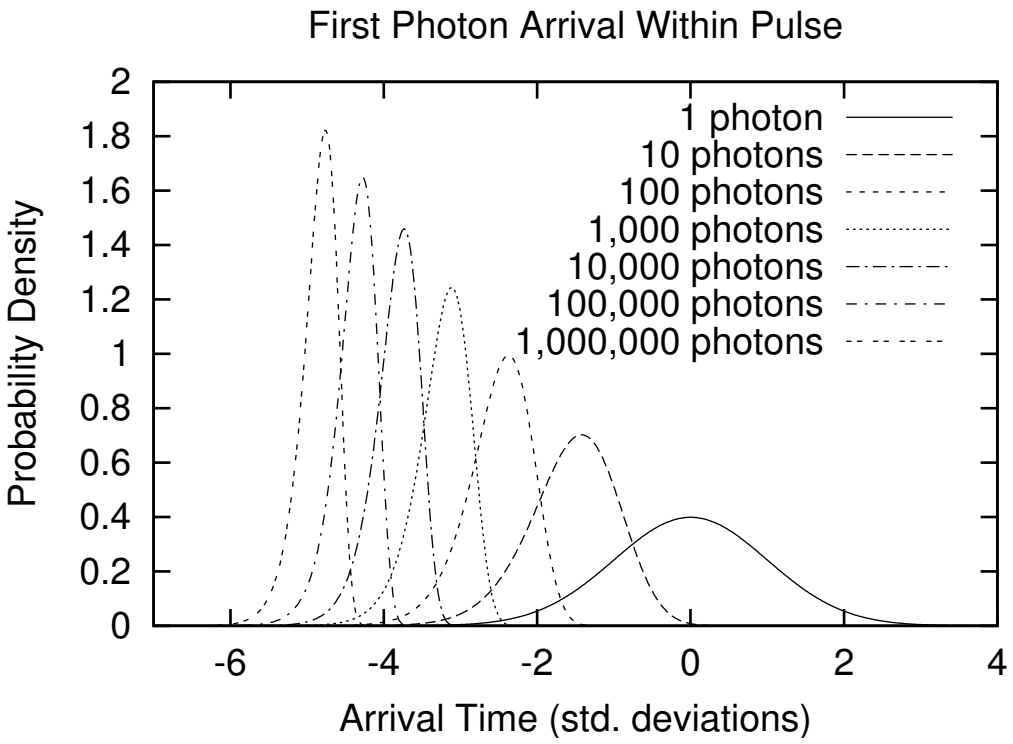

Figure 2: Probability distributions of the arrival time of the first photon within a Gaussian pulse (shown as 1-photon curve).

Figure 2 shows the appearance of this distribution for a wide range of detectable photon number. As is seen, the effective width of the pulse—as would be measured by an APD device only recording the first photon—gets narrower as the photon number is increased. At $N = 30$, the width is half that of the original distribution, ultimately narrowing to 25% at high photon numbers. Of course in a real system, the extent to which the pulse remains Gaussian in shape at low flux levels (pertinent to high photon numbers) becomes very important in understanding the behavior of the pulse width.

Given precise knowledge of the pulse shape (as may be ascertained with high-quality streak camera data), one may reconstruct the first photon probability distribution and use this to effectively "tune" the pulse width delivered to the APD—thus enabling greater understanding of the jitter inherent in the APDs and timing apparatus. An example plot I have seen of the pulse shape from the Hamamatsu PLP-02 laser pulser indicates that the leading edge of the pulse appears to be well approximated by a Gaussian curve to about 3–4σ before the peak. At times before this, the pulse intensity rapidly climbs over two orders-of-magnitude, but with a longer time-constant than is seen in the main Gaussian shape. The bottom line is that in a realistic case, varying the number of photons per pulse that arrive at the APD detector can effectively tune the pulse-width response by a factor of 3–4.

3 The Lunar Retro-reflector Arrays

The retro-reflector arrays left on the moon by the Apollo astronauts introduce a spread in photon arrival time simply as a consequence of being misaligned in angle due to lunar libration. At a maximum libration angle of 10°, the largest array (Apollo 15) can see a corner-to-corner misalignment of 20 cm in range, corresponding to 1350 ns in round-trip path delay. Imagining for a moment that our outgoing pulse were infinitely thin (instantaneous), the time-evolution of the pulse impinging on the retro-reflector arrays can be thought of as a planar sheet passing through the tilted array. Generally speaking, one corner of the array will be hit first, after which the locus of points defining the current pulse intersection is a line (let's call it an *isochrone*) across the array. This isochrone propagates across the array until it hits the far corner.

Treating the retro-reflector array as a rectangular block of uniform "retro-reflectivity"—rather than an array of discrete retro-reflectors—the pulse profile returned from the array would be composed of three segments, symmetrically distributed about the pulse center. The middle segment would have a uniform intensity as a function of time, owing to the fact that the surface area covered per unit time is constant once the isochrone's ends are running along parallel sides of the array (see Figure 3), and thus its length is constant. The beginning and end of the pulse will be linearly rising and descending segments, owing to the linear growth (and reduction) in the size of the isochrone as it sweeps out the corners of the array. Figure 4 shows the appearance of this pulse pattern. In terms of the array dimensions, the durations of the pulse segments are given by:

$$\left.\begin{array}{l} \sigma = \frac{2h\sin\theta}{v} \\ \sigma + \tau = \frac{2w\cos\theta}{v} \end{array}\right\} \text{ for } \tan\theta < \frac{w}{h},$$

$$\left.\begin{array}{l} \sigma + \tau = \frac{2h\sin\theta}{v} \\ \sigma = \frac{2w\cos\theta}{v} \end{array}\right\} \text{ for } \tan\theta > \frac{w}{h},$$

where $v = c/(\sin\psi)$ is the effective velocity of the isochrone across the surface (faster than light!!), with ψ representing the libration angle (misalignment of the array from the line-of-sight). The factor of two in the above expressions accommodates the round-trip nature of the pulse, effectively doubling the path difference across the array. The FWHM and RMS of the resulting pulse are given by

$$FWHM = 2\tau + \sigma, \text{ and}$$

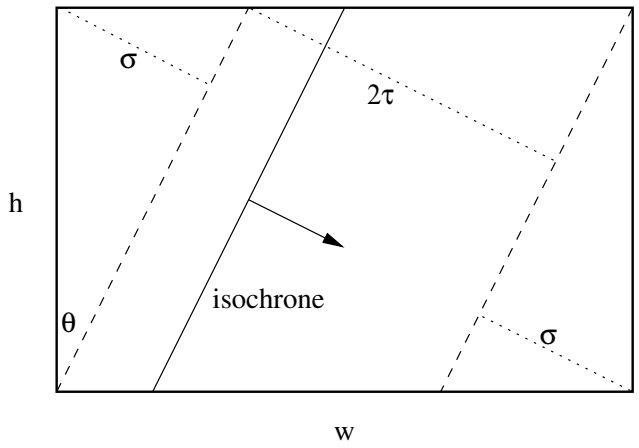

Figure 3: The retro-reflector array is traversed by the obliquely arriving pulse, intersecting the array surface on straight-line isocrones. The pulse return shape is then composed of three segments, of duration σ, 2τ, and σ.

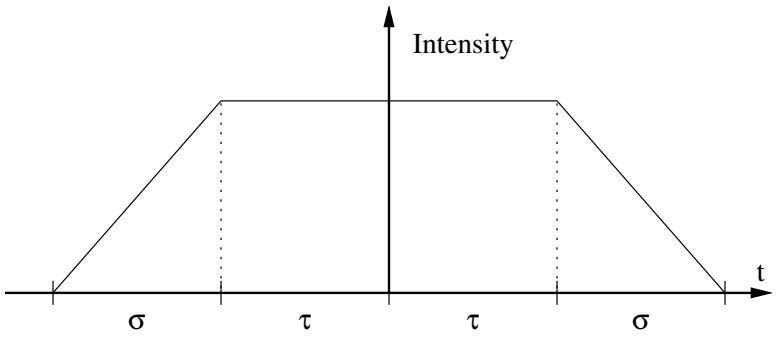

Figure 4: The retro-reflector arrays spread the pulse into a shape similar to that seen above, characterized by linearly growing/declining regions of duration σ and a flat-topped region of duration 2τ.

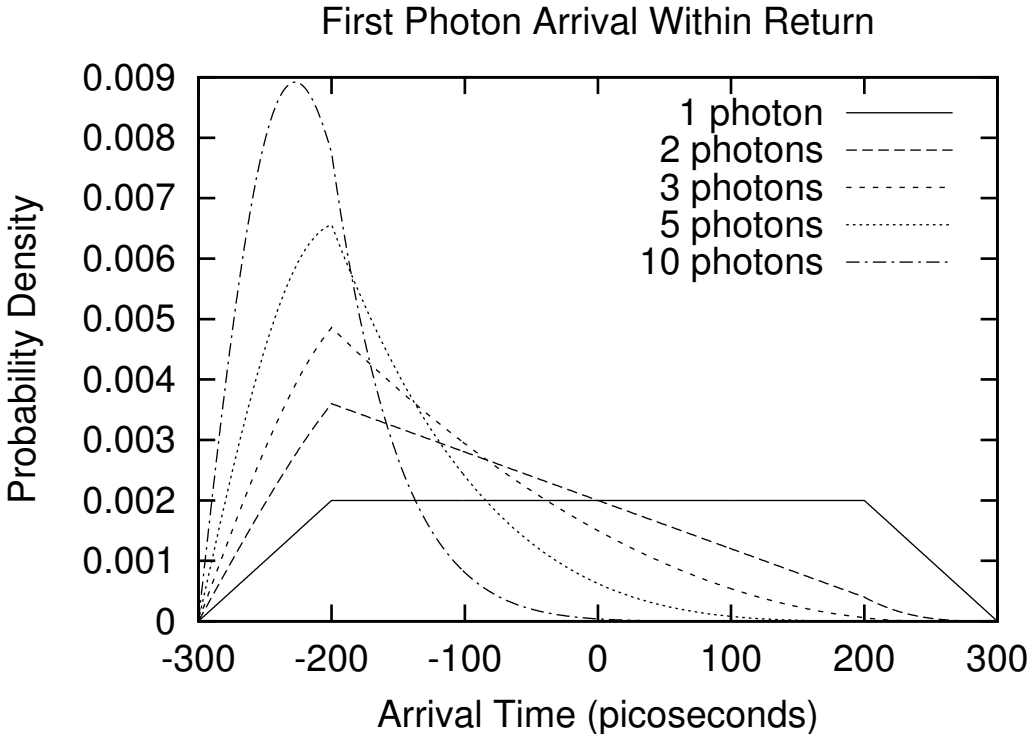

Figure 5: Arrival time of first photon from a typical Apollo array configuration.

$$RMS = \sqrt{\frac{\frac{2}{3}\tau^3 + \tau^2\sigma + \frac{2}{3}\tau\sigma^2 + \frac{1}{6}\sigma^3}{2\tau + \sigma}}.$$

The Apollo 11 & 14 arrays are each 0.46×0.46 square meters, and the Apollo 15 array is 0.61×1.04 m^2.

The probability function of the resulting pulse shape can be used to generate the probability distribution of the first photon arrival time in much the same way as in the examples above. Figure 5 shows the probability functions associated with 1, 2, 3, 5, and 10 photons per pulse. The functions used to generate the first-photon probability (via Equation 4) are included here for reference:

$$p(t) = \begin{cases} \frac{t+\tau+\sigma}{\sigma(2\tau+\sigma)} & -\tau-\sigma \le t < -\tau \\ \frac{1}{2\tau+\sigma} & -\tau \le t \le \tau \\ \frac{\tau+\sigma-t}{\sigma(2\tau+\sigma)} & \tau < t \le \tau+\sigma \end{cases},$$

$$P(t) = \begin{cases} \frac{t^2+2(\tau+\sigma)t+(\tau+\sigma)^2}{2\sigma(2\tau+\sigma)} & -\tau-\sigma \le t < -\tau \\ \frac{2\sigma t+2\tau\sigma+\sigma^2}{2\sigma(2\tau+\sigma)} & -\tau \le t \le \tau \\ \frac{-t^2+2(\tau+\sigma)t+2\tau\sigma-\tau^2+\sigma^2}{2\sigma(2\tau+\sigma)} & \tau < t \le \tau+\sigma \end{cases}.$$

The example shown in Figure 5 clearly shows the effect of the multi-photon early-range bias. A quantification of the probability curves appears in Table 1. The most important column of the table is the one for $\langle t \rangle$, which characterises how early the pulse will be perceived, on average, for the given

N	$\langle t \rangle$	$FWHM$	RMS	$t_{\max}$	$\Delta t_{-\frac{1}{2}}$	$\Delta t_{+\frac{1}{2}}$
1	0.0	500.0	147.2	—	−250	250
2	−84.9	279.0	120.2	−200.0	−54	225
3	−127.4	189.8	99.9	−200.0	−58	132
4	−153.1	154.7	85.4	−200.0	−62	93
5	−170.5	137.14	74.8	−200.0	−65	72
10	−211.6	103.4	48.4	−227.5	−49	54

Table 1: LLR Return Pulse Characteristics

number of photons. Keep in mind that this example applies to a particular array geometry, and that the numbers will often be two times smaller, and sometimes two times larger.

4 Satistics with the Array Detector

The use of an array detector in some ways makes the photon arrival time analysis easy, but in other ways complicates matters. If the number of detectable photons per pulse remained constant for a train of pulses, the rate of multiple photon hits could be easily computed from Equation 2. However, due to variations in intensity owing to speckle-induced variability of target illumination, the number of photons per pulse arriving at the detector may range over at least an order-of-magnitude. With a single-element detector operating in single photon mode, this variability would be indistinguishable from the purely statistical variation in received photon number, thus leading one to misrepresent the rate of occurrence of multiple photon hits, resulting in a bias to shorter computed ranges.

With an array detector, the strength of the return can be judged on a shot-by-shot basis, provided that the array is large enough to accommodate the entire return signal footprint. But this, too, is confused by particularly strong returns that trigger detections in most of the detector elements. If the parameters are well chosen, the number of triggered elements in a given pulse directly relates to the strength of that particular return. It is then possible to assess the likelihood of multiple returns in a given element based on the number of triggered elements relative to the number of elements available for the pulse to trigger.

In a simplified case, let's say that the return beam point spread function (PSF) is not a Gaussian, but a top-hat shaped distribution of uniform intensity over a well defined area. We further simplify the picture by pretending that the edges of the pattern are jagged like the pixel boundaries such that no elements are only partially illuminated. In such an idealized case, let us say that out of M illuminated elements, P are triggered. Then the value of ε in Equation 2 is simply $\varepsilon = P/M$, from which the fraction of elements with multiple photon detections may be ascertained.

The next step would be to understand the bias in estimated range given the likelihood of detections composed of N photons, as derived, for instance from Equation 4 and shown in Figure 5. A statistical correction could then be applied to the centroid of the reported arrival time values to represent the true centroid distance to the target array. For any given pulse, this corrected estimate may or may not be truly representative, as it is merely a guess as to how many photons were disenfracnhised by punching the same pre-punched chad. But on the whole, for a large number of pulses, this correction brings the reported range value closer to truth.

For example, if one deems that a particular array element registers a positive detection 10% of the time, then 90% of *these* detections require no bias adjustment, 9% require correction for two photons, 0.9% for three, etc. Using the retro-reflector array geometry represented in Figure 5 and Table 1, one

would apply a range bias of

$$bias = 0.09 \times (-84.9) + 0.009 \times (-127.4) + \ldots, \tag{6}$$

which comes out to about 9 ps, or 1.5 mm in one-way range. A 20% detection rate ($\varepsilon = 0.2$) results in a bias correction of 19 ps, or 3 mm in range.

Now abandoning the simplifications assumed above, let us treat the return beam PSF as Gaussian in shape, with a FWHM spanning w pixels. The Gaussian function described by this PSF is

$$n(r) = \frac{M \ln 256}{2\pi w^2} e^{-\ln 16 \frac{r^2}{w^2}}, \tag{7}$$

where the traditional σ^2 has been replaced with $w^2/\ln 256$, and $n(r)$ is expressed in terms of the number of photons contained in the entire pulse, M. Then the number of photons contained within a radius, R, is

$$n(< R) = \int_0^{2\pi} d\theta \int_0^R r n(r) dr, \tag{8}$$

which integrates to

$$n(< R) = M(1 - e^{-\ln 16 \frac{R^2}{w^2}}). \tag{9}$$

Exactly half of the total light is therefore contained within the FWHM circle ($R = w/2$). Thus the average photon intensity across the central region is half that encountered in the simplistic top-hat model, with all the photons distributed across this same area. The peak intensity in the center of the Gaussian function is $\ln 2 = 0.69$ times the average intensity of the top-hat version, or $2 \ln 2 = 1.39$ times the average intensity of the Gaussian over the central FWHM circle. 93.75% of the total energy is contained within a circle double the size of the FWHM circle.

The non-uniform distribution indicates that one wants to apply a different correction factor to each detector element depending on that element's illumination. The sparse regions in the skirts of the PSF are unlikely to need much correction, while the central elements will demand larger corrections. One appealing aspect of the experiment is that one may construct independent range estimates in bins of radial distance from the center of the PSF, whereby this multi-photon bias should show up very clearly in the absence of any correction.

An operational procedure may go as follows. Assume 1 arcsecond (FWHM) seeing, with 0.2 arcsecond pixels ($w = 5$). In the case of an average return containing 6 photons, with a laser repetition rate of 20 Hz, one expects 120 photons per second, on average. Based on expectations from speckle structure (with a well collimated beam), the number in any given shot will routinely (80% of shots) vary between 0.2 and 20 photons, with a median around 2.5. This puts the accumulated average photon count at 4.2 in the central pixel. This is enough signal to guide on, but longer integration (5–10 seconds) may be necessary to adequately determine the size and shape of the PSF. From this determined size, a per-shot expectation value can be mathematically determined from a Gaussian PSF profile (Equation 7), scaled in amplitude to match the actual photon count *for that particular pulse*. Then a range bias correction can be applied to *each* element in *each* shot. A second-order correction could account for multiplicity within a given pulse, adjusting the Gaussian amplitude upward accordingly. This is unlikely to be of significant impact for any but the strongest returns.

Figure 6 demonstrates the appearance of random pulses on the APD array, with full Poissonian statistics used to generate the examples. The underlying Gaussian function has 96.4% of its flux within the frame of the detector. As is seen, the size and shape of the PSF can be determined on time scales as short as 10 seconds, though more time would improve the quality of this assessment without badly affecting the application of this information to the individual pulses. An important thing to keep in mind is that the five panels at left in Figure 6 do not represent a typical time series of the lunar photon

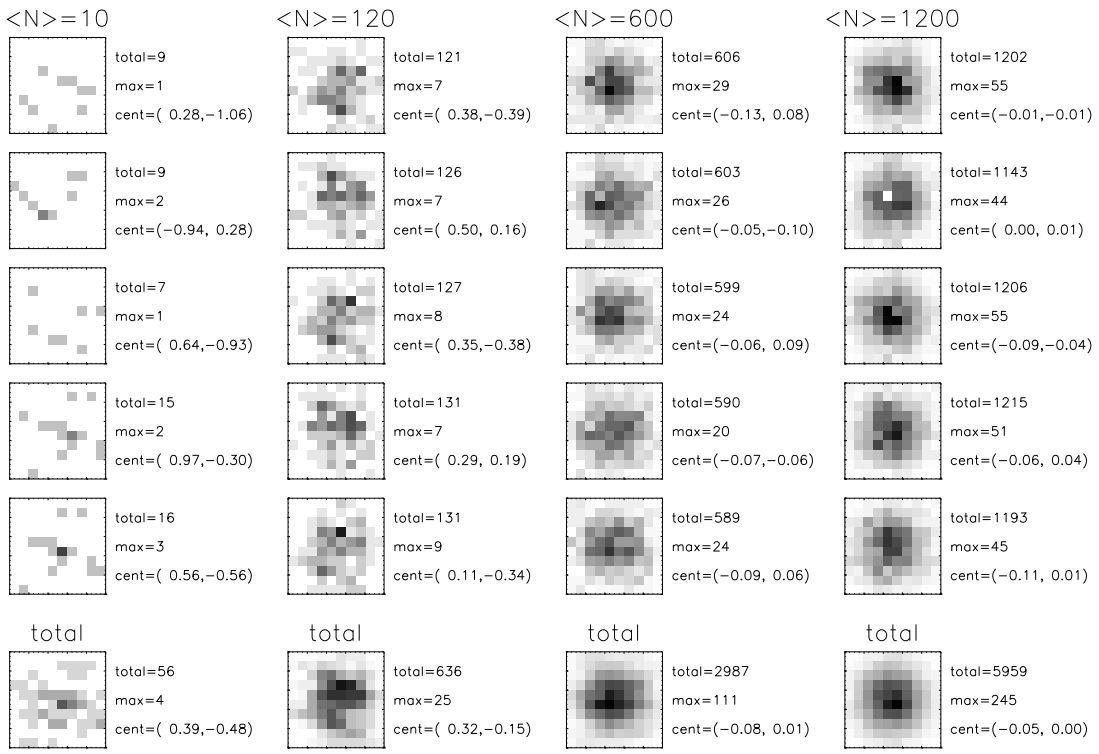

Figure 6: Example appearances of the detection patterns expected from a 10×10 APD array with 0.2 arcsecond pixels in 1.0 arcsecond seeing. Five samples are shown for each photon expectation number, accompanied by a total of the five samples at bottom. The leftmost column depicts the appearance of a typical strong return ($\langle n \rangle = 10$), and the three subsequent columns depict typical integrations for 1, 5, and 10 seconds, respectively.

return, as this varies wildly in terms of signal strength from pulse-to-pulse, as opposed to the stable $\langle n \rangle = 10$ scenario depicted in the figure.

Additional thought needs to be put into the effect of speckle structure in the return beam on the time bias per pixel. Because the pixel elements are large compared to the speckle scale (unlike on the transmit beam, where the Apollo array is miniscule compared to the scale of the speckle structure), this may not present a problem. It would be wise, however, to investigate whether the speckle structure dominates the sparse return patterns like those displayed in Figure 6. If so, the implication is that individual detections are more likely composed of more photons, as the spikes in the speckle pattern pack a meaner punch than would be the case for a smooth return pattern.

This analysis should be extended to arrive at an optimal array pixel scale for APOLLO. Smaller pixels are effective at distributing the returning photons among more elements, but are more adversely influenced by speckle structure. Moreover, the examples shown assume 1 arcsecond seeing, which is perhaps on the better side of typical—especially considering that many operations will occur in twilight conditions, which often involve rapidly changing conditions. And even with 1 arcsecond seeing, the full PSF is clearly extending to the edges of the array. Larger pixels provide greater field coverage for a given number of elements and undersample the speckle pattern more effectively such that more speckles are contained in each element. The downside is that multi-photon contamination is more typical, and can have a direct adverse effect on the one thing we care the most about—the range to the moon.