J. Phys. Photonics 2 (2020) 045008

Contrast resolution of few-photon detectors

Mattias Jönsson and Gunnar Björk

Department of Physics, KTH Royal Institute of Technology, AlbaNova University Center, SE 106 91 Stockholm, Sweden
E-mail: matjon4@kth.se and gbjork@kth.se

Keywords: contrast resolution, detector array, photon number resolution, single photon detector

Abstract

We investigate the minimum acquisition time, expressed as the number of image frames, and the minimum number of absorbed photons per pixel required to achieve a predefined contrast resolution in a monochromatic, pixelated image acquisition system at low light intensities (from well below one photon, to several hundred photons per pixel and frame). Primarily we compare systems based on the pixels of the photon-number-resolving (PNR) type of detectors and detectors that discriminate, in a binary fashion, between zero and non-zero photon numbers (so-called click detectors). We find that our model can seamlessly interpolate between the two. We also model detectors with intrinsic PNR capabilities and integrating detectors with a simple saturation model, derive the probability of errors in assigning the correct intensity (or 'gray level') and finally discuss how the estimated levels, which need to be based on threshold levels due to the stochastic nature of the detected photon number, should be assigned. Overall, we find that non-ideal PNR-detector-based systems offer advantages even over ideal click-detector-based systems when the incident mean photon number is sufficiently large, which is guaranteed to occur around ten photons per pixel and frame.

1. Introduction

In many fields of science, medicine and technology image acquisition plays an important role and imaging constitutes a large and wide research field. A common imaging problem to consider is the reconstruction of a gray-scale image from a possibly repeated measurement with an array of photon detectors. In an ordinary digital camera this is done after RGB color separation of the incoming image. Hence, even for color imaging the problem boils down to the resolution of narrow wavelength-band 'gray-scales', which necessitates the resolution of contrast between the different levels [1].

The imaging problem becomes more difficult when the illumination is constrained to the few-photon level, which is the case for a number of applications such as LIDAR [2, 3], biological spectroscopy [4], image-scanning microscopy [5] and low-dose x-rays [6]. In these cases the use of single-photon detectors such as single-photon avalanche photo-detectors, superconducting nanowire single-photon detectors (SNSPDs) or transition-edge sensors (TESs) is required to resolve faint signals. The construction of an imaging device with sufficient pixels to generate a high-resolution image is an area of ongoing research [7–12].

In recent developments it has been shown that some information about the incident photon number can be gained from avalanche photo-detectors and SNSPDs by analyzing the output pulses [13–15]. Alternatively, multiplexed structures [16–18] or inherent photon-number-resolving (PNR) detectors [19–22] can be used to gain photon-number resolution. Utilizing the fact that these detectors exhibit some PNR capabilities allows for more efficient information extraction of the input and this could improve reconstruction and make it more efficient.

In this paper we investigate the benefit of using detectors for faint-light image acquisition. The particular questions we address are the minimum acquisition time (expressed as the minimum number of image frames or accumulated number of measurements) and the minimum number of absorbed photons per pixel element that one needs to get a predefined contrast resolution in an image.
The contrast between two signal-producing structures in an image is conventionally defined as the difference in signal (e.g. intensity or detected photon number) between the two objects divided by a reference signal, not seldom the sum of the two signals [23]. However, in order to quantify the smallest contrast a detector can perceive, one also needs to consider the statistical variance (i.e. the noise) of the signals. The statistical variance is included in the contrast-to-noise ratio, which is defined as the (absolute) difference in signal between the two objects divided by the standard deviation of this difference (assuming no statistical correlation between the signals) [24]. However, particularly in the image recording of very weak signals, the signal’s standard deviation is a function of the strength of the signal and the detector type. Thus, for faint-light image acquisition, one needs a bit more elaborate analysis to determine the fundamental contrast resolution of an image recording system. This is the motivation behind this work.

In addition to determining the minimum number of frames or detected photon number needed to reach a certain contrast resolution, we also look at the error probability of correctly identifying one of many gray levels. Moreover, we ask if there is anything to gain by using PNR pixel detectors instead of so-called click detectors that can only resolve between zero and more than zero absorbed photons. We will not look into the imaging system itself at all, but only look at the detector array (consisting of individual few-photon detectors) at the image plane. We will also exclude details about the image acquisition such as what wavelengths are used, and what the acquisition time (or shutter speed) is, but have as our main parameter how many photons are detected. In this sense our study is generic, covering a wide range of different kinds of low-intensity image acquisitions.

The paper is organized as follows. In section 2 we define the resolution problem and derive the minimum number of absorbed photons to resolve a fixed number of gray levels in an image using a multiplexed, PNR detector. This model can, depending on the chosen parameters, describe either a so-called click detector (that is, a detector that can only resolve between zero photons and one or more photons) or a linear PNR detector and also cases in between these extremes.

In section 3.1 we analyze the performance of so-called intrinsic PNR detectors, such as TESs and superconducting tapered nanowire detectors (STaNDs). In section 3.2 we extend the analysis to integrating detectors, such as complementary metal–oxide–semiconductor (CMOS) detectors and charge-coupled device (CCD) detectors. In section 4 we look at the probability of making level estimation errors due to the statistical nature of the detected photon number. In section 5 we derive how the gray levels to be distinguished should optimally be distributed to benefit maximally from the detectors’ characteristics. Finally, in section 6, we summarize our findings and draw conclusions.

2. Contrast resolution of few-photon detectors

Suppose a semitransparent object is illuminated with a spatially uniform light source. The light that passes through the object is focused onto a detector array that, in an imaging scenario, would be called an array of pixels (see figure 1). In the image plane, the pixels are then illuminated by light of different intensities corresponding to the object’s transparency at the imaged point. (A very similar imaging scenario would ensue for any other illuminated, light-scattering object.) To get a good gray-scale image, one needs to resolve between different light intensities, each intensity coding a certain transparency (or reflectance) of the object.

Suppose we want to resolve $L$ levels of intensity. (In many contemporary imaging applications the intensity is coded into one byte, meaning that one discretizes the intensity to $2^8 = 256$ different levels, translating to transparencies in our case.) We will initially assume that these intensities are equally spaced, meaning that we would like to resolve between the average transmitted photon numbers:
\[
\mu(l) = \frac{\mu_0 l}{L-1}, \quad l \in \{0,1,2,\ldots,L-1\},
\]

where \(\mu_0\) is the mean illumination photon number per pixel and \(l/(L-1)\) is the transmittance of level \(l\).

To measure the transparency levels we consider each pixel in the array to be a multiplexed PNR detector consisting of \(n\) single-photon detector elements, where each detector element has a quantum efficiency \(\eta\) and a dark count probability \(p_d\) per single-shot measurement. The probability of each detector element clicking when \(m\) photons are incident to that detector element is given by

\[
Pr(\text{click} \mid m) = 1 - (1 - p_d)(1 - \eta)^m.
\]

Since imaging necessitates the detection of many modes and typically photon number fluctuations in these modes in one acquisition frame are uncorrelated, the photon number fluctuations detected by any pixel can be modeled by a Poisson distribution. Furthermore, the output from each pixel is assumed to be independent of preceding outputs, which is the case when the time between subsequent measurements is longer than the recovery time of the detectors. This can be achieved by using a pulsed light source or a sufficiently faint continuous light source such that the mean time between two measurements of the same detector element is longer than the recovery time. Under these assumptions, the probability of a pixel giving \(x \in \{0,1,\ldots,n\}\) clicks given the transparency level \(l\) is

\[
Pr(x \mid l) = \binom{n}{x} (1 - p_d)^n e^{-\mu(l)\eta} \left(\frac{e^{\mu(l)\eta/n}}{1 - p_d} - 1\right)^x,
\]

which is consistent with reference [15] when \(p_d = 0\). This model is rather general in the sense that it can describe the case when each pixel is a single-photon detector by setting \(n = 1\) and it can describe the case when each pixel is a linear PNR detector by letting \(n \to \infty\) and setting \(p_d = \nu/n\), where \(\nu\) is the average number of dark counts per pixel and measurement. To estimate the transparency levels we conduct \(k\) measurements (i.e. we take \(k\) image ‘frames’) with the detector array and assume that the light intensity is constant from frame to frame. The result of the measurements is a sequence of data \(\{X_i\}_{i=1}^k\) for each pixel which can be used to compute an estimate of \(l\). Using the maximum likelihood method to compute estimates yields

\[
\hat{l} = -\frac{n(L-1)}{\eta \mu_0} \ln \left(\frac{n - \langle X\rangle}{(1 - p_d)n}\right),
\]

where \(\langle X\rangle\) is the average taken over the measurement data:

\[
\langle X\rangle = \frac{1}{k} \sum_{i=1}^{k} X_i.
\]

The estimator \(\hat{l}\) is defined on the set \((0, \infty)\) while the level \(l\) is defined on the set \(\{0, 1, \ldots, L-1\}\), which means that we need to limit \(\hat{l}\) to the smaller set. This can be done in the maximum likelihood sense by defining a discretized estimator as

\[
\hat{l}' = \min \left\{L - 1, \arg\max_{l \in \{0,1,\ldots,L-1\}} I(l) \right\},
\]

where \(L\) is the likelihood function.

To ensure that the estimation is stable we require the separation to the next level to be \(d\) times the standard deviation of the estimator, i.e. \(1 \geq d \sqrt{\text{Var}(\hat{l})}\).

For large enough \(d\) we get levels that are well-separated and a reconstructed image that has the correct values with high probability. By using the Cramer–Rao bound we can get a lower bound on how many frames are required to generate an image with separation \(d\). We get

\[
1 \geq d \sqrt{\text{Var}(\hat{l})} \geq d \sqrt{\frac{1}{k I(l)}},
\]

where \(I(l)\) is the Fisher information. Computing the Fisher information from equation (3) from \(k\) independent measurements gives

\[
I(l) = \left(\frac{\mu_0 \eta}{L-1}\right)^2 \frac{1}{n(1 - p_d)^{-1}e^{\mu(l)\eta/n} - 1},
\]
The number of required measurements \( k \) plotted against the average mean photon number per pixel per frame \( \mu_0 \eta \) for different numbers of single-photon detector elements \( n \) in each multiplexed detector pixel and with no dark counts. When \( \mu_0 \eta \ll n \) there is no benefit to increasing the number of detector elements, but when \( \mu_0 \eta \gg n \) the number of measurements can be reduced significantly by increasing \( n \). As expected, \( k \) diverges when \( \mu_0 \eta \rightarrow 0 \) since the detector will always output \( X = 0 \) and the amount of information gained per measurement is zero. Similarly \( k \) diverges as \( \mu_0 \eta \gg n \) since the detector always outputs \( X = n \).

Combining the equations gives a bound on the minimum number of measurements required for any \( l \) to get \( d \) standard deviations of separation under ideal conditions:

\[
k \geq \frac{nd^2(L-1)^2}{\mu_0 \eta} \left[ (1-p_d)^{-1} \eta e^{\mu_0 \eta/n} - 1 \right].
\]

(9)

In figure 2 the number of required measurements \( k \) is plotted against the detected mean photon number \( \mu_0 \eta \) for different numbers of single-photon detector elements \( n \) in each multiplexed pixel. For low mean photon numbers \( \mu_0 \eta \ll n \) all detectors with at least \( n \) elements perform equally, while for high mean photon numbers \( \mu_0 \eta \gg n \) a large number of elements is beneficial. This suggests that detectors with the capability to resolve more than one photon are beneficial for estimating high mean photon numbers.

As seen in figure 2 there exists a mean photon number that minimizes the number of measurements required. This minimum is achieved when the mean photon number is selected as

\[
\mu_0 \eta = n \left[ 2 + W_0 \left( -2(1-p_d)e^{-2} \right) \right],
\]

(10)

where \( W_0 \) is the Lambert \( W \) function. When \( p_d = 0 \) the optimal mean photon number is given by \( \mu_0 \eta = 1.5936n \), while when \( p_d \rightarrow 1 \) the minimum is given by \( \mu_0 \eta \rightarrow 2n \) (see figure 3). Hence the interval for the optimal illumination is narrow and without any knowledge of the dark count rate it is difficult to select an appropriate illumination level.

### 3. Contrast resolution of linear PNR detectors

#### 3.1. Intrinsic PNR detectors

Intrinsic PNR detectors such as TESs or STaNDs can be seen as multiplexed detectors with a large number of elements. To a good approximation we can consider them to have an infinite number of elements, which together with equation (3) means that the probability of getting \( x \) as output given the transparency level \( l \) is a Poisson distribution given by

\[
Pr(x \mid l) = \exp \left\{ -\mu(l) \eta - \nu \right\} \frac{[\mu(l) \eta + \nu]^x}{x!},
\]

where \( \nu \) is the average number of dark counts per frame given by the relation \( \nu = p_d \eta n \). Similarly we can show that the required number of measurements to achieve \( d \) standard deviations of separation between levels is given by

\[
k \geq \frac{d^2(L-1)^2}{\mu_0 \eta} \left( 1 + \frac{\nu}{\mu_0 \eta} \right).
\]

(12)
Figure 3. The optimal illumination as a function of the dark count probability. The arrows mark which axis corresponds to which curve. The upper red curve shows the mean photon number that minimizes the number of measurements required. The lower blue curve shows the number of measurements required to achieve $\delta$ standard deviations of separation.

Unlike the case for finite $n$, the number of measurements $k$ does not have a minimum for a finite mean photon number and the optimal strategy according to the model is to use maximal illumination. However, in reality these PNR detectors are limited to how many photons they are capable of resolving due to overlapping output signals or non-linear output signals [14, 15, 25], which implies that there is an upper limit for when increased illumination is beneficial. To capture this behavior we need to consider a refined model where these effects are included.

A simple but somewhat crude model that describes this limitation can be achieved by introducing a hard cut-off on how many photons the detector is capable of measuring. This is done by assuming that inputs with up to $m$ photons are resolvable given that the inputs are sufficiently spaced in time such that the detector has time to recover, while photon numbers higher than $m$ produce outputs indistinguishable from an $m$ photon event. The qualitative behavior of the intrinsic PNR detector can then be investigated by setting an upper photon number limit for which the detector is not capable of resolving for some reason.

Using the hard cut-off model results in an output distribution which is given by equation (13) when $x < m$,

$$
Pr(x | l) = \sum_{y\geq m} \exp\{-\mu(l)\eta - \nu\} \frac{[\mu(l)\eta + \nu]^y}{y!}
$$

when $x = m$ and zero when $x > m$. Hence, the detector acts as a linear PNR detector up to $m$ photons and is then not capable of distinguishing between higher photon numbers due to the cut-off. It is therefore expected that the detector performs identically to a linear PNR detector small enough mean photon numbers.

Given $k$ data points from the intrinsic cut-off detector we can estimate the level $l$ using an estimator $\hat{l}$. Requiring that the estimator have a standard deviation $\delta$ times less than the separation to the next level gives the condition on the minimum number of measurements $k$ (see figure 4). The qualitative result is similar to the multiplexed case, but the gain of increasing $m$ is not as significant as that when increasing the number of detector elements in the multiplexed case. However, from figure 4 it is evident that the acquisition time can be significantly reduced by utilizing the extra information gained by measuring the photon numbers ($m > 1$) as compared to the case when only single photons are resolved ($m = 1$).

### 3.2 Integrating detectors

Integrating detectors, such as CMOS and CCD detectors, are similar to intrinsic PNR detectors in the sense that they are linear in some range of illumination strengths and we therefore expect these detectors to be described by equation (11) in this range. Like intrinsic PNR detectors, such detectors are limited to some upper number of photons that can be detected in one frame due to saturation. The saturation limit depends on the physical size of the pixel and it sets the upper bound for the validity of equation (11). As was done in the previous subsection, we can model this limit with a hard cut-off and we therefore expect the distribution at the saturation to be approximately given by equation (13).
CMOS and CCD detectors also exhibit a readout noise which is non-negligible for small photon numbers per frame, which sets a lower limit to the range where the detector is linear. The total noise (expressed as the standard deviation) $N$ below the saturation level can be modeled as

$$N = \sqrt{k[\mu(l)\eta + N_r^2]}$$

(14)

where $k$ is the number of frames summed over, $\mu(l)\eta$ is the average number of accumulated photo-electrons per pixel per frame and $N_r$ is the average pixel readout noise per frame. Here, we neglect the corresponding thermal noise per frame for simplicity, but for room-temperature operation and long frame integration times, for example, it should be added to the terms enclosed by square brackets in equation (14). Hence, it is required that $\mu(l)\eta/N_r^2 \gg 1$ for the detector to be in the linear range. For commercially available state-of-the-art CMOS and CCD cameras the readout noise standard deviation is around $N_r = 2$ [26, 27], which implies that such detectors behave linearly when at least 40, or so, photo-electrons are accumulated per pixel and frame. This suggests that the camera integration time relative to the impinging photon flux should be chosen such that the number of accumulated photo-electrons per pixel per frame is larger than this lower limit, but under the constraint that the frame rate must be reasonably high to be suitable for the application at hand and that the chosen integration time does not imply additional, non-negligible thermal noise.

Thus, integrating detectors are well described by the model considered in the previous subsection under the additional assumption that the photon number per frame is much larger than the read-out and thermal noise. This implies that the conclusions drawn from the previous subsection are also valid for CMOS and CCD detectors given the just mentioned restrictions, by setting $n \rightarrow \infty$ and provided that system-specific parameters, such as quantum efficiency, dark count rate and saturation intensity, are used as input in the model.

4. Estimation errors

Above we have used the standard deviation as a parameter to determine how well a detector is able to distinguish between different gray levels. Having a larger number of standard deviations between neighboring levels reduces the risk of incorrect classification, but we have so far not specified how the number of standard deviations translates to the error probability. Here, we want to quantify the error in the case where detectors described by equation (3) are used.

The probability for a correct classification is given by the probability that the discretized estimator $\hat{l}'$ is equal to the correct level $l$. To determine this probability we use a discrete estimator $\hat{l}'$ that is a function of the experimental mean $\langle X \rangle$, which gives the probability for correct classification as

$$Pr(\hat{l}' = a \mid l = a) = Pr(\langle X \rangle \in S(a) \mid l = a),$$

(15)
where \( S(a) \) is the set of values for which \( \langle X \rangle \) produces the discretized estimator \( l' = a \). This set is given by all \( \langle X \rangle \) that make the likelihood function maximal for the level \( l = a \) and it can be determined using the conditions

\[
\mathcal{L}(a) \geq \mathcal{L}(a + 1),
\]

\[
\mathcal{L}(a) \geq \mathcal{L}(a - 1).
\]

These conditions imply that for \( 0 < a < L - 1 \) the set is given by

\[
S(a) = \left( \frac{\eta_0 \eta (L - 1)^{-1}}{\ln[g(a)/g(a - 1)]}, \frac{\eta_0 \eta (L - 1)^{-1}}{\ln[g(a + 1)/g(a)]} \right),
\]

where the function

\[
g(a) = e^{\eta_0(a) a / n} - (1 - p_d).
\]

When \( a = 0 \) the lower limit for \( S(a) \) is given by 0 which is the minimal value for \( \langle X \rangle \). Hence

\[
S(0) = \left( 0, \frac{\eta_0 \eta (L - 1)^{-1}}{\ln[1/g(1)]} \right).
\]

Similarly, when \( a = L - 1 \) the upper limit for \( S(a) \) is given by \( n \) which is the maximal value for \( \langle X \rangle \). Hence,

\[
S(L - 1) = \left( \frac{\eta_0 \eta (L - 1)^{-1}}{\ln[g(L - 1)/g(L - 2)]}, n \right).
\]

The probability that the experimental average is in the set \( S(a) \) can be estimated using the central limit theorem, which states that the experimental average converges in distribution to a normal distribution as the number of acquisitions grows, i.e.

\[
\langle X \rangle \rightarrow \mathcal{N}\left( \frac{\mathbb{E}[X]}{k}, \frac{\text{Var}(X)}{k} \right), \text{as } k \rightarrow \infty,
\]

where \( \mathbb{E}[X] \) and \( \text{Var}(X) \) are the expected value and variance of \( X \), respectively. For sufficiently large \( k \) it holds approximately that

\[
\text{Pr}(\langle X \rangle \in S(a) | l = a) \approx \text{Pr}(\nu \in S(a) | l = a),
\]

where \( \nu \sim \mathcal{N}(\mathbb{E}[X], \text{Var}(X)/k) \) is a normally distributed variable with the same average as the random variable \( X \) and variance equal to the variance of \( X \) divided by the number of acquisitions.

In figure 5 the error probability as a function of the number of acquisitions \( k \) for the most difficult levels to resolve (levels \( L - 2 \) and \( L - 1 \)) is presented when the mean photon number has been chosen to minimize \( k \) (see equation (10)). The error probability decreases with increased \( k \) and an increased number of detector elements \( n \). In figure 6 the error probability as a function of the number of standard deviations \( d \) between two levels is presented when the mean photon number per frame \( \eta_0 \eta \) has been chosen to minimize \( k \). The curves are overlapping which shows that the error probability is uniquely determined by \( d \). At this optimum \( \eta_0 \eta \) the total exposure \( k \eta_0 \eta \) is equal to \(-d(L - 1)^2/2W_0(2(1 - p_d))e^{-2} \) for any \( n \).

5. Optimal gray-level spacing

In previous sections we have considered gray levels equally spaced, with the assumption that the level \( L - 1 \) corresponds to unit transmittance. When setting a requirement on how many measurements (or frames) \( k \) are needed to achieve \( d \) standard deviations of the separation of neighboring levels this is not the best scenario. With equal level spacing, darker levels are easier to resolve than lighter levels and it should therefore be possible to improve the contrast resolution by allowing non-linear level spacing.

The optimal level spacing occurs when every level is limiting for the number of acquisitions \( k \) given some requirement \( d \) on the number of standard deviations separating two neighboring levels. From the condition in equation (9) we get a differential equation:

\[
\left( \frac{\partial \mu(l)}{\partial l} \right)^2 \frac{n^2 \eta^2}{(1 - p_d)^{-1} e^{\eta_0(l) n / n} - 1} = \frac{d^2}{k},
\]
which holds for all levels \( l \in 0, 1, \ldots, L - 1 \). Let us assume that the first level is at zero intensity, i.e. \( \mu(l = 0) = 0 \); then the solution to the differential equation is given by

\[
\mu(l) = \frac{n}{\eta} \ln \left[ 1 + \tan^2 \left( \frac{ld}{2 \sqrt{kn}} + \arctan \left( \sqrt{\frac{pd}{1 - pd}} \right) \right) \right] + \frac{n}{\eta} \ln(1 - pd),
\]

given that the condition \( d^2/(4\eta) > 1 \) is met, which for most realistic situations should be the case. If not, some levels with a small index \( l \) will have \( k\mu(l) \leq 1 \) and since the measurement outcome is always an integer and for these levels the most likely outcome is zero or one, the levels cannot be distinguished.

Let the parameters \( \eta, n, d, k \) and \( pd \) be fixed; then there is a finite number of resolvable gray levels. Increasing the number of levels further would violate the condition that there should be \( d \) standard deviations of separation between levels. The maximal level is found by investigating where the solution \( \mu(l) \) stops being valid, which occurs when the solution diverges. The maximal level is therefore given by
Figure 7. Optimal level spacing when $n = 16, d = 1, k = 10^3$ and there are no dark counts. The mean photon number diverges above the maximal level $l_{\text{max}} = 397$ and the solution is therefore not valid above this level. For levels with a sufficiently small level index $l$, the function $\mu(l)$ can be approximated by $(ld)^2/4k$.

Figure 8. Transmittance as a function of the level when the maximal level is 1024 for optimal spacing and linear spacing. The optimal spacing grows quadratically for sufficiently small levels.

$$l_{\text{max}} = \max \{ l \in \mathbb{N} \mid |\mu(x)| < \infty, \forall x \in [0,l] \}$$

$$= \left\lfloor \frac{2\sqrt{kn}}{d} \left( \frac{\pi}{2} - \arctan \left( \frac{pd}{\sqrt{1-pd}} \right) \right) \right\rfloor.$$  \hspace{1cm} (26)

This equation shows the advantage of PNR detectors, since the maximum number of distinguishable gray levels increases as $n^{1/2}$, that is, essentially with the square root of the detector’s PNR capability. We also see that the detector’s quantum efficiency does not limit its resolution capability, whereas its dark count probability does. In figure 7 the mean photon number $\mu(l)$ is plotted as a function of the level $l$. As predicted the mean photon number diverges in the interval $l \in (l_{\text{max}}, l_{\text{max}} + 1]$ and the solution is no longer valid for levels higher than $l_{\text{max}}$.

Taking the ratio $\mu(l)/\mu(l_{\text{max}} - 1)$ we get the transmittance of level $l$. This is plotted in figure 8 assuming that $p_d = 0$. In the plot we have chosen $\sqrt{kn}/d \approx 326$ to get $\mu(l_{\text{max}} - 1) = 1024$. For sufficiently low transmittance (or level indices $l$) the transmittance grows quadratically with the level index. As a reference we have plotted 1024 equally spaced transmittances, with the highest being unity. Choosing levels that are equally spaced results in poorer performance, because as explained in section 2, the variance of the measured
signal is largest for the highest indices $l$ and therefore one should try to maximize the transmittance difference, translating to maximizing the mean average transmitted photon number difference, at the highest level indices. The discrepancy between the two spacing models can clearly be seen by comparing the derivatives of the two functions at $l_{\text{max}}$.

6. Summary

We have shown that in order to resolve contrast in a monochrome image and thereby extract information contained therein, it could be advantageous to use PNR detectors as detector elements in the image detector array. Provided that the object is illuminated so that the average number of detected photons per pixel (detector) per image frame is above unity and within the PNR range of the PNR detectors, non-ideal PNR detectors will provide a better contrast resolution than even an ideal, click-detector-based image acquisition system. The relative difference in performance between the two detector types diverges exponentially with an increasing number of detected photons per pixel per frame.

We have shown that the performance in terms of needed measurements (i.e. number of accumulated image frames) for a given acceptable error probability of a PNR detector decreases linearly with the average number of detected photons per pixel (detector) per image frame, given that the number is within the range resolvable by the detector. In contradistinction, a click-detector-based system has an optimal detected photon number per pixel (detector) per image frame. Below or above this optimum less information is extracted for a given number of measurements.

We have also quantified the probability of incorrect classification of the gray levels. It was shown that for a PNR detector this error probability is independent of the range of photons the detector can distinguish within, provided that the detected photons per pixel (detector) per image ‘frame’ are chosen optimally. The advantage of using a detector with greater PNR capability is that the number of needed measurements (or frames) decreases with increasing $n$. Since the statistical variation at the pixel level between frames is assumed to be uncorrelated, the accumulated needed transmitted photon number is independent of $n$, implying that under the assumptions just stated, the number of measurements needed $k$ scales as $\propto 1/n$.

Finally we have shown, not surprisingly, that to resolve maximally many gray levels with a given error probability (i.e. a certain $d$), the levels should not be equidistant in transmittance but should follow a non-linear level spacing. For the darkest shades, the transmittance should increase approximately quadratically with the level number $l$. This is in line with the human vision, where the eye is significantly better at distinguishing closely spaced levels of dark gray (faint light) than at distinguishing equally spaced levels of light gray (intense light) [28].

Acknowledgments

This work was supported by the Knut and Alice Wallenberg Foundation through a Quantum Sensing grant and by the Swedish Research Council (VR) through Grant 621-2014-5410.

Appendix A. Derivation of the Poisson photon counting distribution

Consider a multiplexed PNR detector with $n$ elements with quantum efficiency $\eta$ and dark count probability $p_d$. In a previous work [29] we showed that the probability to get $x \in \mathbb{N}$ clicks when a number state $|m\rangle$ is incident is given by

$$
Pr(x \mid m) = \frac{1}{n^m} \left( \begin{array}{c} n \\ x \end{array} \right) \sum_{l=0}^{x} (-1)^l (1 - p_d)^{n-x+l} \left( \begin{array}{c} x \\ l \end{array} \right) \times [n - (n - x + l)\eta]^m. \tag{27}
$$

Using this distribution we want to obtain the resulting derivation for when a coherent state with mean photon number $\mu$ is incident. Using a coherent state that has Poisson statistics gives the following probability to get $x \in \mathbb{N}$:
\[
Pr(x \mid \mu) = \sum_{m \in \mathbb{N}} \frac{e^{-\mu} \mu^m}{m!} Pr(x \mid m) \\
= e^{-\mu} \left( \sum_{l=0}^{x} \frac{(-1)^l (1 - p_d)^{n-x+l}}{l!} \right) \sum_{m \in \mathbb{N}} \frac{\mu^m}{m!} \frac{m!}{n} (n - (n - x + l))^{m}. \tag{28}
\]

Multiplying the inner sum by \(\exp -\mu [n - (n - x + l)]/n\) and using the fact that the Poisson distribution sums to one gives

\[
Pr(x \mid \mu) = \left( \frac{n^x}{x!} (1 - p_d)^{n-x} e^{-\mu} \right) \left( 1 - \frac{\mu}{n} \right)^{x} \tag{29}
\]

where the binomial theorem is used to compute the sum in the last equality.

The resulting probability distribution has the expected value

\[
\mathbb{E}[x] = n \left[ 1 - (1 - p_d) e^{-\mu/n} \right] \tag{30}
\]

and the variance

\[
\text{Var}(x) = \mathbb{E}[x^2] - \mathbb{E}[x]^2 \\
= n \left[ 1 - (1 - p_d) e^{-\mu/n} \right] \left[ 1 - p_d \right] e^{-\mu/n}. \tag{31}
\]

The Fisher information for the probability distribution given in equation \((29)\) is given by

\[
I(\mu) = \mathbb{E} \left( \left( \frac{\partial \ln Pr(x \mid \mu)}{\partial \mu} \right)^2 \right) \\
= \text{Var} \left( \frac{\partial \ln Pr(x \mid \mu)}{\partial \mu} \right), \tag{32}
\]

where the score has zero expected value. The score is given by

\[
\frac{\partial \ln Pr(x \mid \mu)}{\partial \mu} = \frac{\eta x}{n (1 - (1 - p_d) e^{-\mu/n})} - \eta, \tag{33}
\]

which together with equation \((31)\) gives the Fisher information

\[
I(\mu) = \frac{\eta^2 \text{Var}(x)}{n^2 (1 - (1 - p_d) e^{-\mu/n})^2} \\
= n \left[ (1 - p_d)^{-1} e^{\mu/n} - 1 \right], \tag{34}
\]

Appendix B. Estimators and variance

Here we derive an expression for the maximal likelihood estimator \(\hat{l}\) from a set of data \(\{X_i\}_{i=1}^k\) given that the illumination is given by equation \((1)\) and the output distribution is given by equation \((3)\). Using equation \((33)\) we get an estimator given by the condition

\[
0 = \sum_{i=1}^{k} \frac{\partial \ln Pr(X_i \mid \mu(\hat{l}))}{\partial \mu(\hat{l})} \frac{\partial \mu(\hat{l})}{\partial \hat{l}} \\
= \frac{\mu_0}{L - 1} \left[ \eta \sum_{i=1}^{k} X_i \left( 1 - (1 - p_d) e^{-\mu(\hat{l})n/\eta} \right) - k\eta \right]. \tag{35}
\]
Inverting the expression and introducing the average over the experimental data \( \langle X \rangle \) as equation (5) gives

\[
\hat{l} = - \frac{n(L - 1)}{\eta \mu_0} \ln \left( \frac{n - \langle X \rangle}{1 - p_d n} \right).
\]  

(36)

The maximal likelihood estimator \( \hat{l} \) is asymptotically unbiased [30] and the Cramer–Rao bound can therefore be used to give a lower bound on the variance:

\[
\text{Var}(\hat{l}) \geq \frac{1}{kI(l)} = \frac{1}{kl(\mu(\hat{l}))} \left( \frac{\partial \mu(\hat{l})}{\partial l} \right)^{-2} = \frac{n(L - 1)^2 [(1 - p_d)^{-1} e^{\mu(\hat{l})n/\eta} - 1]}{k(\mu_0 \eta)^2},
\]

(37)

where we use the variable substitution role for Fisher information and equation (34). Requiring \( d \) standard deviations of separation between peaks implies that every

\[
k \geq nd^2(L - 1)^2 \left( \frac{(1 - p_d)^{-1} e^{\mu(\hat{l})n/\eta} - 1}{(\mu_0 \eta)^2} \right)
\]

(38)

holds for \( l \in \{0, 1, \ldots, L - 1\} \). If the bound holds for \( l = L - 1 \) then it automatically holds for any other \( l \) and we can therefore simplify the expression by setting \( \mu(\hat{l}) = \mu_0 \), which gives equation (9). To minimize \( k \) with respect to \( \mu_0 \eta \) is equivalent to minimizing the function

\[
f(\xi) = \frac{\alpha e^{\xi/n} - 1}{\xi^2},
\]

(39)

where we set \( \xi = \mu_0 \eta \), and \( \alpha = (1 - p_d)^{-1} \). The minimum occurs when \( df/d\xi = 0 \) which corresponds to

\[
\alpha e^{\xi/n} = 2n\alpha e^{\xi/n} + 2n = 0,
\]

(40)

which under the constraint \( \xi \in [0, \infty) \) has the solution

\[
\xi = n \left[ 2 + W_0 \left( -\frac{2}{\alpha} e^{-2} \right) \right],
\]

(41)

where \( W_0 \) is the Lambert \( W \) function. Substituting back the variables gives the condition for the minimum in equation (10).

**ORCID iDs**

Mattias Jönsson  
https://orcid.org/0000-0001-8488-9035

Gunnar Björk  
https://orcid.org/0000-0002-2082-9583

**References**

[1] Aldaem H and Davidson R 2012 *Radiographer* 5946

[2] Zhou H, He Y, You L, Chen S, Zhang W, Wu J, Wang Z and Xie X 2015 *Opt. Express* 23 14603

[3] Shin D, Xu F, Venkatraman D, Lussana R, Villa F, Zappa F, Goyal V K, Wong F N C and Shapiro J H 2016 *Nat. Commun.* 7 12046

[4] Niwa K, Numata T, Hattori K and Fukuda D 2017 *Sci. Rep.* 7 45660

[5] Buttafava M et al 2020 arXiv:2002.11443

[6] Zhu Z and Pang S 2018 *Appl. Phys. Lett.* 113 231109

[7] Guerrieri F, Tisa S, Tosii A and Zappa F 2010 *IEEE Photonics J.* 2 759

[8] Bronzi D, Villa F, Tisa S, Tosii A, Zappa F, Durini D, Weyers S and Brockherde W 2014 *IEEE J. Sel. Top. Quantum Electron.* 20 354–63

[9] Allman M S et al 2015 *Appl. Phys. Lett.* 106 192601

[10] Gottardi L et al 2014 *Space Telescopes and Instrumentation 2014: Ultraviolet to Gamma Ray* vol 9144 ed Takahashi T, den Herder J-W A and Bautz M Int. Society for Optics and Photonics (SPIE) pp 792–8

[11] Gasparini L et al 2018 2018 *IEEE Int. Solid-State Circuits Conf. (ISSCC)* pp 98–100

[12] Wollman E E, Verma V B, Lita A E, Farr W H, Shaw M D, Mirin R P and Woo Nam S 2019 *Opt. Express* 27 35279

[13] Kardynal B E, Yuan Z L and Shields A J 2008 *Nat. Photonics* 2 425

[14] Nicolich K L, Cahall C, Islam N T, Lafayette G P, Kim J, Miller A J and Gauthier D J 2019 *Phys. Rev. Appl.* 12 034020

[15] Zhu D, Colangelo M, Chen C, Kozrh B A, Wong F N C, Shaw M D and Berggren K K 2019 arXiv:1911.09485

[16] Divoity A et al 2008 *Nat. Photonics* 2 302

[17] Mansili F, Bitauld D, Gaggero A, Jahamirnejad S, Leon R, Mattioli F and Fiore A 2009 *New J. Phys.* 11 045022

[18] Dauler E A, Kernan A J, Robinson B S, Yang J K, Voronov B, Goltsman G, Hamilton S A and Berggren K K 2009 *J. Mod. Opt.* 56 364

[19] Lita A E, Miller A J and Nam S W 2008 *Opt. Express* 16 3032
[20] Fukuda D et al 2011 Opt. Express 19 870
[21] Konno T, Takasu S, Hattori K and Fukuda D 2020 J. Low Temp. Phys. 199 27–33
[22] Young S M, Sarovar M and Leonard F 2020 ACS Photonics 7 821
[23] Hendrick R E and Raff U 1992 Magnetic Resonance Imaging ed D D Stark and W G Bradley (Maryland Heights, MI: Mosby) 2nd edn pp 109–44
[24] Timischl F 2015 Scanning 37 54
[25] Humphreys P C et al 2015 New J. Phys. 17 103044
[26] Instruments O Sona (https://andor.oxinst.com/assets/uploads/products/andor/documents/andor-sona-scmos-specifications.pdf)
[27] Photomics H ImagEM X2-1K EM-CCD (https://www.hamamatsu.com/resources/pdf/sys/SCAS0092E_IMAGEMX2s.pdf)
[28] Kimpe T and Tuyschaeuer T 2007 J. Digit. Imaging 20 422
[29] Jönsson M and Björk G 2020 Phys. Rev. A 101 013815
[30] Amari S 2016 Information Geometry and Its Applications Applied Mathematical Sciences (Berlin: Springer)