Tomography of photon-number resolving continuous-output detectors

To cite this article: Peter C Humphreys et al 2015 New J. Phys. 17 103044

View the article online for updates and enhancements.

Related content
- Integrated photonic sensing
  N Thomas-Peter, N K Langford, A Datta et al.
- Experimental quantum tomography of a homodyne detector
  Samuele Grandi, Alessandro Zavatta, Marco Bellini et al.
- Measuring measurement: theory and practice
  A Feito, J S Lundeen, H Coldenstrodt-Ronge et al.

Recent citations
- Detector-Independent Verification of Quantum Light
  J. Sperling et al
- Direct calibration of click-counting detectors
  M. Bohmann et al
- Photodetector figures of merit in terms of POVMs
  S J van Enk
Tomography of photon-number resolving continuous-output detectors

Peter C Humphreys\textsuperscript{1,4}, Benjamin J Metcalf\textsuperscript{3}, Thomas Gerrits\textsuperscript{2}, Thomas Hemstra\textsuperscript{1}, Adriana E Lita\textsuperscript{2}, Joshua Nunn\textsuperscript{1}, Sae Woo Nam\textsuperscript{1}, Animesh Datta\textsuperscript{5}, W Steven Kolkhammer\textsuperscript{1} and Ian A Walmsley\textsuperscript{1}

\textsuperscript{1} Clarendon Laboratory, Department of Physics, University of Oxford, OX1 3PU, UK
\textsuperscript{2} National Institute of Standards and Technology, Boulder, CO, 80305, USA
\textsuperscript{3} Department of Physics, University of Warwick, CV4 4AL, Coventry UK
\textsuperscript{4} Author to whom any correspondence should be addressed.
\textsuperscript{5} E-mail: peter.humphreys@physics.ox.ac.uk

Keywords: single photon detectors, detector tomography, quantum optics

Abstract

We report a comprehensive approach to analysing continuous-output photon detectors. We employ principal component analysis to maximize the information extracted from output signals, followed by a novel noise-tolerant parameterized approach to the tomography of photon-number resolving detectors. We further propose a measure for rigorously quantifying a detector’s photon-number-resolving capability. Our approach applies to all detectors with continuous-output signals. We illustrate our methods by applying them to experimental data obtained from a transition-edge sensor detector.

1. Introduction

The continuing development of highly efficient photon detectors has significant impact across a broad range of fields, from quantum information\cite{1} to astronomy\cite{2} and biomedical imaging\cite{3}. The physics underlying the operation of different photon detectors is rich and varied, but their outputs typically fall into two categories. Those such as photomultiplier tubes\cite{4}, avalanche photodiodes\cite{5} and superconducting nanowires\cite{6, 7} are often based on avalanche phenomena and naturally lead to the use of ‘click’ outcomes. For these detectors, the energy of the input state cannot be resolved. In contrast, detectors such as transition-edge sensors (TESs)\cite{8}, kinetic-inductance detectors\cite{2} and superconducting tunnel junctions\cite{9} rely on smooth transitions leading to continuous ‘trace’ outputs (avalanche photodiodes can also give continuous-valued outputs under appropriate conditions\cite{10}). This continuous output can allow the energy of the input pulse to be resolved, allowing for true photon-number sensitivity\cite{8}. Some of these, including TES detectors, are highly sensitive single-photon detectors with quantum efficiencies of up to 98%\cite{8, 11}. Others, such as microwave kinetic inductance detectors, allow unprecedented level of integration into large arrays\cite{2}. These advances over traditional discrete-output detectors will enable new applications in wide-ranging fields.

With these novel applications and regimes of performance come additional challenges in detector characterization. Unlike discrete-output detectors, many photon-number resolving detectors (PNRDs) produce a complex time-varying signal from which the input state must be inferred. Efficiently extracting information from these signals is therefore necessary to realize the full capability of such detectors\cite{12–16}.

The signal produced by a continuous-output detector is typically a time-dependent voltage with some dependence on photon number which may in general be nonlinear, as shown in figure 1(a). A set of such output signals $V = \{v_i(t)\}$, where $v_i(t)$ is a stochastic variable representing the $i$th detected signal, can be represented using a (deterministic) set of basis functions $\{w_j(t)\}$ such that

$$v_i(t) = \sum_j s_{ij} w_j(t),$$

(1)
where the stochastic variables $s_{ij}$ represent the weighting components for each signal $v_i(t)$. These $s_{ij}$ are each drawn from their respective associated sample spaces $S_j$. In general, this implies that, in order to capture the full output of the detector, it is necessary to determine $s_{ij}$ for all of the $n$ basis functions for each signal to be measured. For a truly continuous signal, $n$ is in principle infinite, but of course for any real experiment the upper limit to $n$ is set by the length of the sample recorded by the detector. This is set by the ratio of the sampling time window to the temporal resolution of the electronics. However, this finite signal still spans a space of high dimension; in our work a single trace from a TES detector consists of 1024 16 bit numbers. Directly analysing this signal is therefore impractical. This is particularly the case for detector tomography, which is necessary to rigorously characterize the relationship between input states and output signals [16–21]. Detector tomography requires a sufficiently small space of outputs that the probability of a given outcome can be estimated precisely from the measured data. For the full output space of our detector signal, we estimate the probability of the same trace occurring twice (to within the resolution of the analogue-to-digital converter) in a data set of $10^5$ traces to be on the order of $10^{-4}$, rendering tomography in this full space infeasible. This motivates the development of an approach to the characterization of continuous-output detectors that enables accurate and precise signal analysis and detector tomography.

Detector tomography has been previously carried out for continuous-output PNRDs with 5% quantum efficiencies [13], in which the continuous-output problem was circumvented by ‘binning’ the detector output based on the maximum amplitude of the signal. This approach does not make optimal use of the information available. Furthermore, as we will discuss, the numerical techniques for detector tomography used in the study are not effective in the high detection-efficiency regime, which is now accessible with TES detectors. Another recent work has explored algorithmic methods of interpreting the response of high detection-efficiency PNRDs based on cluster analysis [14]. Although this may prove useful for rapid characterization of a detector, it does not
unambiguously reconstruct the operators corresponding to the measurement and is therefore unable to provide a rigorous characterization of the detector response.

This paper is organized as follows. In section 2, we introduce the use of principal component analysis to compress the output signals from a TES detector. Section 3 contains a brief introduction to detector tomography, followed by discussion of our novel approach to tomography based on maximum likelihood estimation. Our results are used to estimate the system efficiency of our detector in section 4. Finally, in section 5 we consider other applications of this tomography data, and introduce a quantitative measure of photon number resolution.

2. Principal component analysis

We first consider the problem of efficiently extracting information from a high-dimensional detector signal data set. This is achieved by employing a standard technique from multi-variate statistics, namely principal component analysis [22]. For a given data set, this approach determines the optimal set of ‘principal component’ basis functions \( \{ w_j(t) \} \) such that each successive basis function captures the maximum amount of information possible from the data set (as measured by the variance of the ‘principal component scores’ \( \{ s_j \} \), given by \( \sigma_j = \langle s_j^2 \rangle - \langle s_j \rangle^2 \), while maintaining orthogonality with the previous components. Crucially, this implies that if the principal component basis is truncated to compress the data, the maximum amount of variance of the original data set will still be captured. In other words, the truncated principal component basis will provide the most faithful reconstruction of the data for a given number of components.

In an actual experiment, the signals \( v(t) \) and therefore the basis functions \( w_j(t) \) are necessarily discretized due to the finite sampling rate of the data acquisition system. In this case the set of signals \( V \) can be expressed as a matrix. It can be shown that the problem of determining \( \{ w_j(t) \} \) for \( V \) is equivalent to finding the eigenvectors of the experimentally determined covariance matrix \( \bar{V}^T \bar{V} \), where \( \bar{V} \) is the data set with the mean signal subtracted [22]. These eigenvectors can be efficiently determined using singular value decomposition. Once \( w_j(t) \) are known, \( s_j \) can be calculated from the detector signals \( v(t) \) by \( s_j = \int v(t) w_j(t) \, dt \).

We applied principal component analysis to a data set of TES traces. This was composed of subsets of 60 traces, each taken at 300 different coherent-state inputs with average photon numbers spanning linearly from 0 to approximately 15 photons per input signal pulse (the input signal pulse duration was 1 ns, much shorter than the 100 ns detector jitter; input pulses were also separated by 20 μs in order to avoid any signal overlap). We found that this size of data set was a reasonable balance between ensuring that enough traces were sampled while minimizing the time required to acquire and process the signals. In figures 1(a) and (b), example TES traces from this data set are plotted both in their original form, and in a reduced form using only the first two principal components \( w_1(t) \) and \( w_2(t) \). As can be seen, with just these two components, most of the structure of the traces has been reproduced. This can be shown more formally by comparing the variance \( \sigma_1 \) of \( \{ s_j \} \) for different principal component numbers \( j \), as plotted in figure 1(c). A larger variance shows that a given component is capturing more distinguishing information between the traces in the data set. The variance \( \sigma_1 \) is two orders of magnitude greater than \( \sigma_2 \), and this trend continues, with \( \sigma_1 \) rapidly decreasing as a function of \( j \). The reconstructed signals shown in figure 1(b) therefore capture 99.0% of the original signal covariance.

Interestingly, as figure 1(d) shows, \( w_1(t) \) is very close to the mean shape of the TES traces. This would be expected theoretically in the small-signal limit, in which the TES trace height simply scales linearly with the photon number [23]. This confirms that projecting onto the mean trace shape, as used by [14], is a useful approach for distinguishing TES signals in the few-photon limit using only a single parameter. Beyond providing a justification for this choice of processing method, the higher order principal components that are revealed by our analysis can provide additional data with which to characterize the response of a detector, particularly for higher photon numbers. For example, \( w_2(t) \) captures the increase in the pulse length with photon number (due to an increase in thermal recovery time), as will be further discussed below.

3. Detector tomography

Detector tomography is necessary to determine the correspondence between the reduced detector signals and the input number of photons [20]. The goal of detector tomography is to determine the positive-operator-valued measure (POVM) \( \{ \hat{\pi}(s) = \sum_{n,m=0}^{\infty} \theta_{n,m}(s) \ket{n} \bra{m} \} \) (written here in the Fock state basis) that fully characterizes the detector response; this is parameterized by the outcome \( s \) in the space of \((S_1, S_2, ..., S_n)\). Once the POVM is known, the probability density for detector outcome \( s \), given input state \( \rho \), is determined by the Born

\[5\text{ Details on the coherent state probe energy calibration method are given in appendix A.}\]
The standard approach to detector tomography consists of experimentally estimating the outcome probability densities \( p_s(\rho_k) \) for a set of known probe basis states \( \{ \rho_k \} \). Using these estimated probabilities, equation (2) can then, in principle, be inverted to find \( \hat{\pi}(s) \).

The set of probe states \( \{ \rho_k \} \) must provide a sufficient basis for the operator space of the POVM \( \{ \hat{\pi}(s) \} \); in other words, it must be tomographically complete. This constraint is satisfied by using a well established method [20] for tomography of PNRDs based on coherent-state probes \( |\alpha\rangle \langle \alpha| \). It is well known that coherent states form an over-complete basis for an optical mode. Coherent states are also straightforward to generate in the lab and retain Poissonian photon number statistics and the same functional form of Wigner distribution despite experimental losses during preparation, making them ideal probe states. Additionally, as TES detectors are phase insensitive, their response depends only on the magnitude of the coherent state parameter \( \alpha \), and not its phase. This significantly reduces the number of probe states needed to form a tomographically complete set of basis operators and removes the need for any phase reference in the experiment.

As for the principal component analysis, we measured the detector response to a set of 300 different probes \( \{ |\alpha_k\rangle \langle \alpha_k| \} \) with energies equally spaced between 0 and 15 photons on average per pulse. However, in this case, we acquired a larger data set of 49 152 traces for each probe energy, since the tomography procedure is more sensitive to statistical noise than principal component analysis. These measured signals were used to estimate the probability density functions \( p_s(|\alpha_k\rangle \langle \alpha_k|) \). Figure 1(e) shows an example measured probability density function for outcomes \( s \) in the space of \( S_1 \) given a probe state with a mean of 3.1 photons per pulse. In figure 2 this is extended to plot example probability density functions for outcomes in the two-dimensional output space of \( S_1 \) and \( S_2 \) given different coherent-state inputs. Additional structure along \( S_2 \) is visible, and could be incorporated into signal analysis to further distinguish input states. However, since the dominant contribution to the data variance is from \( w_1(t) \), particularly for the low photon numbers considered here, we choose to solely focus on \( S_1 \) for the remainder of the analysis. This slightly reduces the information available, but significantly reduces the computational demands of our numerical methods. It is anticipated that this technical shortcoming can be overcome in future extensions of this work.

These probability density functions \( p_s(|\alpha_k\rangle \langle \alpha_k|) \) are in fact a direct estimate of the Husimi \( Q \) function for a given \( |\alpha_k\rangle \), and are therefore sufficient to fully characterize the operation of the detector [20, 24]. Although these measurements are sufficient, for many quantum optics applications it is of interest to determine the response of the detector in the photon number basis, for example, when heralding states produced in spontaneous parametric down-conversion (SPDC) [25, 26]. A phase insensitive detector will have POVM elements diagonal

\[
p(s|\rho) = \text{Tr} \left[ \rho \hat{\pi}(s) \right].
\]

The probability density functions were calculated using Gaussian-kernel density estimation [32]. This technique is better suited to this problem than using histograms, as it is not necessary to choose an arbitrary binning of the data. Instead, this approach directly gives continuous-valued estimates of the functions.
in the photon-number basis [13]; these can therefore be expressed as
\[ \hat{\pi}(s) = \sum_{n=0}^{\infty} \theta_n(s) |n\rangle \langle n|. \]  
Inserting equation (3) into the Born rule (equation (2)), we find that the probability density for the outcome \( s \) given a coherent-state input \(|\alpha\rangle \langle \alpha|\) is
\[ p(s|\alpha) = \sum_{n=0}^{\infty} F_{n,n} \theta_n(s). \]  
where \( F_{n,n} = |\alpha|^4 \exp(-|\alpha|^2). \)

Using the set of probability density functions \( p(s|\alpha) \) associated with the input probe states \(|\alpha_1\rangle \langle \alpha_1|\), this relation can be numerically inverted to find the best solution for \( \theta_n(s) \) consistent with the physicality constraints
\[ \theta_n(s) \geq 0, \quad \text{and} \quad \int \theta_n(s) \, ds \leq 1. \]  

It is well known that the problem of inverting equation (4) to obtain \( \theta_n(s) \) is ill-conditioned [20]. We found that published methods of performing this numerical inversion based on constrained least squares techniques [13] did not give satisfactory results (figure 3). This may be in part due to the reduced overlap between the POVM elements for different photon numbers as compared to previous studies, because of our much higher system detection efficiency. This means that regularization techniques designed to promote this overlap [20, 21] do not work as effectively.

### 3.1. Maximum likelihood detector tomography routine

We used insights from the collected data to develop a novel detector tomography routine that is effective for high quantum efficiencies. We adopted a model in which the POVM elements for photon number \( n \) (in the space of \( S_1 \)) are assumed to be a sum of \( n+1 \) Gaussians, with widths, heights, and positions as free variables. This Gaussian-mixture model [27] is motivated by the observation that the dominant noise mechanisms (amplifier noise, thermal fluctuations etc) for many classes of detectors (including TES [28]) are expected to be Gaussian (due to the central limit theorem), which therefore leads to an overall Gaussian error on the detection signal. We note that our approach is readily extendable to higher order components (\( S_2, S_3, \ldots \)).

The Gaussian-mixture model gives the following expression for the POVM coefficients for photon number \( n \)
\[ \theta_n(s) = \sum_{j=1}^{n+1} \beta_{n,j} \mathcal{N}(s | \mu_{n,j}, \sigma_{n,j}), \]  
where \( \beta_{n,j} \) is a weighting factor for the Gaussian probability distribution \( \mathcal{N}(s | \mu_{n,j}, \sigma_{n,j}) \) in the outcome space \( s \), with mean \( \mu_{n,j} \) and standard deviation \( \sigma_{n,j} \).

We imposed the constraint that \( \mu_{n,j} = \mu_{n+1,j}^0 \) i.e. that the Gaussians from different photon numbers should be aligned. This is physically motivated by the fact that the detector cannot distinguish between cases where \( n \) photons were input and cases where \( n+1 \) photons were input and one photon was lost. Removing this constraint does not alter the solution significantly, beyond leading to a slight jitter in the location of the peaks for each photon number. However, this jitter complicates the additional analysis that we carried out, particularly

---

7 This assumes that the input state is a pure state—this can easily be extended to mixed state inputs [29]. Mixed states can result from classical uncertainties in \( \alpha \), however we found that the fluctuations in our laser power (<1% over the timescale of the entire experiment) did not lead to any significant differences in the results of our analyses, particularly when compared to the effects of the uncertainty in the calibration of the probe state energies.
with regard to compensating for the uncertainty in the probe state energies (as will be discussed in the next section). No constraint is placed on $\sigma_{\alpha_{ik}}$.

Substituting equation (6) into equation (4) gives the model outcome probabilities

$$p_m(s|\alpha, \chi) = \sum_{n_j} F_{n,n} \beta_{n,j} \mathcal{N}(s | \mu_{n,j}, \sigma_{n,j}),$$

where $\chi$ is used as shorthand to denote the set of all the parameters $\beta_{n,j}$, $\mu_{n,j}$, $\sigma_{n,j}$, in order to make the dependence on the model explicit.

This expression gives the posterior probability density for the TES detector producing an outcome $s$ in our model, given an input coherent-state probe $|\alpha\rangle$ $\langle\alpha|$. This posterior probability should be maximized for the measured data. Typically, maximum likelihood estimation [27] is carried out based on a set of observed outcomes $\{s_i\}$. In this case, the quantity to be maximized is the log-likelihood

$$\mathcal{L} = \log \left( \prod_i p_m(s_i | \alpha_i, \chi) \right) = \sum_i \log \left( p_m(s_i | \alpha_i, \chi) \right).$$

However, due to the large number of data points that we sample, evaluation of this sum becomes impractical. Instead, we can rewrite this equation in terms of the experimentally estimated outcome probability densities $p_s(s|\alpha_k)$ (see footnote 5) for each $|\alpha_k\rangle$ $\langle\alpha_k|$. (These are labelled $p_s$ to distinguish them from the model probability densities $p_m$). This gives

$$\mathcal{L} = \sum_k \int N_k p_s(s|\alpha_k) \log \left( p_m(s|\alpha_k, \chi) \right) ds,$$

where $N_k$ is the total number of samples measured at each value of $\alpha_k$. Since we measured the same number of samples per coherent state value, we neglect this constant factor that has no impact on the maximum likelihood estimation.

The full expression for the log-likelihood therefore becomes

$$\mathcal{L} = \sum_k \int ds p_s(s|\alpha_k) \log \left( \sum_{n_j} F_{n,n} \beta_{n,j} \mathcal{N}(s | \mu_{n,j}, \sigma_{n,j}) \right).$$

In order to maximize this log-likelihood, we follow the standard approach [27] of taking derivatives with respect to each parameter in the model. For example, differentiating with respect to $\mu_{n,j}$ gives

$$\frac{\partial \mathcal{L}}{\partial \mu_{n,j}} = \sum_k \int p_s(s|\alpha_k) \gamma_{k,n,j} \sigma_{n,j} (s - \mu_{n,j}) ds,$$

in which

$$\gamma_{k,n,j} \equiv \frac{F_{n,n} \beta_{n,j} \mathcal{N}(s | \mu_{n,j}, \sigma_{n,j})}{\sum_{n_j} F_{n,n} \beta_{n,j} \mathcal{N}(s | \mu_{n,j}, \sigma_{n,j}).$$

Rearranging leads to the following expression for $\mu_{n,j}$

$$\mu_{n,j} = \frac{1}{\kappa_{n,j}} \sum_k \int p_s(s|\alpha_k) \gamma_{k,n,j} s ds,$$

where $\kappa_{n,j} = \sum_k \int p_s(s|\alpha_k) \gamma_{k,n,j} ds$.  

Similarly we find that

$$\sigma_{n,j} = \frac{1}{\kappa_{n,j}} \int p_s(s|\alpha_k) \gamma_{k,n,j} (s - \mu_{n,j})^2 ds,$$

$$\beta_{n,j} = \frac{\kappa_{n,j}}{\kappa_n}, \quad \text{where} \quad \kappa_n = \sum_j \kappa_{n,j}. \quad (15)$$

Note that these expressions for the parameters are dependent on $\gamma_{k,n,j}$ and therefore do not form a closed-form solution. This means that the optimal solution cannot be found analytically. However, it can be shown that a simple routine consisting of the repeated application of two steps will converge to a solution [27]. In the first step, the current values of the parameters are used to calculate $\gamma_{k,n,j}$. This is then used in the second step to re-estimate the optimal values of the parameters using equations (13)–(15).

The results of this inversion are shown in figure 4(a). The efficacy of this model-based routine can be estimated by comparing $p_s(s|\alpha_k)$ and $p_m(s|\alpha_k)$. The average $L_1$ distance

$$\langle L_1 \rangle = \frac{1}{d} \sum_k \int \left| p_s(s|\alpha_k) - p_m(s|\alpha_k) \right| ds$$

between this reconstruction and the original data is 0.054 as compared...
to 0.047 for the unphysical reconstruction given by a least-squares approach (where \(L_1 = 0\) for perfectly overlapped distributions and \(L_1 = 1\) for completely distinguishable distributions). This shows that this model equally effectively captures the detector response (with only 0.7% lower overlap than the least squares approach) while providing a fit that is more consistent with our knowledge of the detector operation.

3.2. Incorporating calibration uncertainty

As a final step, it is necessary to incorporate the uncertainty in the calibration factor \(\eta_{\text{att}}\) used to determine the coherent-state probe energies. The POVM element coefficient \(\theta_n(s)\) gives the probability density \(p(s|n)\) that one will measure outcome \(s\) given \(n\) input photons. This probability is actually \(p(s|n, \eta_{\text{att}})\) since \(\eta_{\text{att}}\) is a variable in our tomography calculations. Our uncertainty in \(\eta_{\text{att}}\) must therefore be accounted for. Based on our error analysis (and assuming normally distributed errors), we can estimate the probability density \(p(\eta_{\text{att}})\) for \(\eta_{\text{att}}\).

Additionally, we can calculate \(p(s|n, \eta_{\text{att}})\) for different \(\eta_{\text{att}}\). Combining these, we can incorporate this statistical uncertainty into our POVM using

\[
p(s|n) = \int p(s|n, \eta_{\text{att}}) p(\eta_{\text{att}}) d\eta_{\text{att}}.
\]

The results of this analysis are shown in figure 4(b). The higher photon-number POVM elements are particularly sensitive to this uncertainty, and show correspondingly large deviations from their ideal values. This highlights the crucial importance of an accurately calibrated probe state source for detector tomography. Although our setup has a high calibration uncertainty of 8%, calibration uncertainties of less than 1% are achievable [29, 30].

4. Detector quantum efficiency

The POVM elements that we calculate using detector tomography completely characterize the detector response. Model-free detector tomography is obtained by treating the detector as a black box, and so in principle does not contain information on the system detection efficiency, i.e. the loss that occurs between the input and the detector.

However, the less general, but physically motivated model-based detector tomography approach that we have adopted can allow us to make an estimate of this efficiency. As noted above, we assume that the response of the detector to each photon number is composed of several Gaussian elements. We can make the further assumption that these different Gaussian elements occur due to the action of loss on an initial Fock state, leading to a statistical mixture of photon numbers at the detector, with one Gaussian element per photon number state detected. If this is the case, the areas of these elements should follow a binomial distribution within each of the Fock state POVM elements. For a given system detection efficiency, it is therefore possible to calculate the expected areas of these Gaussian elements and compare them to the actual tomography results. We used a

---

8 Details on the coherent state probe energy calibration method are given in appendix A.
numerical routine to find the loss level that minimized the $L_1$ norm between this predicted output and the tomography data; the results are shown in figure 5.

This analysis suggests that our system detection efficiency is $0.98 \pm 0.02 / \pm 0.08$. The asymmetric uncertainty arises as the efficiency is upper bounded at 1. The average difference between the predicted photon number distribution and the data is 2% (as measured by the $L_1$ norm of the two probability distributions). This suggests that our initial assumption that the Gaussian elements result from a statistical mixture of photon numbers at the detector is accurate.

Figure 5. Estimating the system detection efficiency. As discussed in the main text, we assume that loss before the detector leads to a binomial distribution of Gaussian elements when an initial Fock state is input to the detector. The relative weights of each of these Gaussian elements, as inferred from our maximum likelihood estimation protocol, are plotted here in gold. From this data, we can carry out a numerical routine to determine the system detection efficiency ($0.98 \pm 0.02 / \pm 0.08$) most consistent with our data. The predicted photon (and therefore Gaussian element) distribution resulting from this estimated detection efficiency is shown in blue for comparison.

9 The system detection efficiency is defined as the efficiency with which a photon in the fibre connected to the detector is detected [7]. The asymmetry in the system detection efficiency is enforced by hand to ensure physicality—the original errors are symmetric ($+0.08 / -0.08$).
5. Application of tomography data

The above tomography procedure gives the probability density $p(s|n)$ for a specific outcome $s$ given an $n$-photon input to the detector. However, in typical experiments, one is actually interested in the complementary probability density $p(n|s)$ that the input contained $n$ photons given that the detector measured outcome $s$.

As is the case with all detector characterization methods (which can only provide $p(s|n)$), inferring $p(n|s)$ requires Bayes’ theorem $p(n|s) = p(s|n)p(n)/p(s)$ and thus depends on the prior probability $p(n)$ of an $n$-photon input. Here, as an example we consider two distinct priors which might arise in applications. First, we consider a Poisson distribution $p(n) = e^{-\lambda} \frac{\lambda^n}{n!}$ which would result from a coherent-state input. We also consider a thermal distribution $p(n) = e^{-\lambda} (1 - e^{-\lambda}) \lambda^n$ which describes a thermal state input and, importantly, the single-mode marginal statistics of a SPDR source. If one mode of such a source is sent to a detector, $p(n|s, \lambda)$ represents the statistical mixture of photon numbers onto which the other mode is projected. Such information is extremely important for quantum information and metrology applications.

Two example probability distributions $p(n|s, \lambda)$ and $p(n|s, \lambda)$ are plotted in figures 6(a) and (b). As can be seen, the two priors lead to significant differences in the distributions. For the thermal distribution, the thermal prior suppresses the overlap between the outcomes associated with neighbouring photon numbers. This is because, for small $\lambda$, $n + 1$ input photons will occur much less frequently than $n$ photons. Therefore the predominant overlap contribution, due to an $n + 1$-photon input being detected in the space of outcomes most associated with $n$ input photons, occurs correspondingly less frequently than genuine $n$-photon inputs.

The Poissonian prior plotted in figure 6(b) has the opposite effect as the thermal prior, since for $n$ less than the mean photon number, an input of $n + 1$ photons is more probable than an input of $n$ photons, and therefore the overlap is promoted. It should be noted that in both cases, due to the truncation of our detector tomography at 17 input photons, the distributions $p(n|s)$ become inaccurate in regions in which significant contributions would be expected from photon numbers greater than this. In practice, this simply translates to an operational requirement that detector tomography must be extended to include all photon numbers that are expected to contribute in any given experiment.

5.1. Characterizing photon number resolution

Closely linked to determining $p(n|s)$ is the problem of finding a quantitative measure of the ‘photon-number resolution’ of the detector (as we discuss in appendix B, the system detection efficiency is not by itself sufficient to give such a quantitative measure for PNRD). Since $p(n|s)$ only gives information on the confidence with which a specific outcome $s$ can determine the photon-number input, we propose a measure that represents an average of this confidence, weighted by the probability density for $s$ given $n$ input photons.
Given an input of \( n \) photons, this confidence \( C_n \) represents the average probability ascribed to the \( n \) photon component of the inferred state \( \rho(s) = \sum n p(n) |n\rangle \langle n| \). More loosely, it represents the probability that the detector gives the correct photon number. Additionally, \( C_n \) is given by \( \int p(s) f^2(|s|) p(n) ds \), the average squared fidelity between the inferred detected state and an \( n \) photon number state \(|n\rangle\), weighted by the probability \( p(s) \). For the detection of a heralding state from a SPDC source \(^2\), this will therefore also be the fidelity of the heralded state with \(|n\rangle\). Note that the detector does not have information on the specific input photon-number \( n \); however, a prior distribution must be specified. This confidence is therefore a function of the distribution chosen. Figure 7(a) shows the confidence for different photon numbers as a function of the SPDC source thermal prior distribution parameter \( \lambda^2 \), where \( p(n) |n\rangle \langle n| = (1 - \lambda^2) \lambda^n \).

In order to facilitate comparison between different detectors, it may be useful to determine this confidence given a flat prior for the photon number

\[
C_n = \int_{-\infty}^{\infty} p(s) f(s) p(n) ds = \int_{-\infty}^{\infty} \frac{p(s)p(n)}{p(s)} ds = \int_{-\infty}^{\infty} \frac{p(s)p(n)}{\sum k p(s)k p(k)} ds.
\]

(17)

Figure 7. (a) Calculated confidence \( C_n \) for different photon numbers as a function of the thermal prior distribution parameter \( \lambda^2 \).

(b) Calculated confidence for our detector given a flat prior, as a function of photon number \( n \) (blue). The confidence assuming no uncertainty in the coherent-state probe energies is plotted for comparison (dashed red). The significant discrepancy between these curves highlights the importance of careful probe state calibration. The confidence is plotted for outcomes at the centres of the peaks in \( p(s) \) (dashed red), as explained in figure 8, (again assuming no probe state energy uncertainties). Finally, the confidence for a time-multiplexed pseudo-number-resolving detector (dashed green) is shown in order to demonstrate the applicability of this measure to PNRDs with different underlying modes of operation.

In order to determine this confidence given a flat prior for the photon number

\[
C_n = \int_{-\infty}^{\infty} p(s) f(s) p(n) ds = \int_{-\infty}^{\infty} \frac{p(s)n^2 p(n)}{\sum k p(s)k} ds.
\]

(18)

This is plotted in figure 7(b). As would be expected, our detector is extremely effective at resolving vacuum and lower photon numbers, while for higher photon numbers, the increasing effect of the detection inefficiency and gradual saturation of the detector leads to a reduced confidence in the outcomes. To demonstrate that this measure is widely applicable to different PNRDs, the confidence for the time-multiplexed pseudo-number-resolving detector with 8 time bins presented in \(^2\) is also shown.

For certain applications, such as heralding Fock states for use in quantum enhanced metrology \(^3\), it is important to maximize the fidelity of the inferred detected state with a photon number state \( C_n \). In these cases, the fidelity can be improved using post-selection strategies in which only a subset of outcomes are accepted. This

---

**Figure 7.** (a) Calculated confidence \( C_n \) for different photon numbers as a function of the thermal prior distribution parameter \( \lambda^2 \). (b) Calculated confidence for our detector given a flat prior, as a function of photon number \( n \) (blue). The confidence assuming no uncertainty in the coherent-state probe energies is plotted for comparison (dashed red). The significant discrepancy between these curves highlights the importance of careful probe state calibration. The confidence is plotted for outcomes at the centres of the peaks in \( p(s) \) (dashed red), as explained in figure 8, (again assuming no probe state energy uncertainties). Finally, the confidence for a time-multiplexed pseudo-number-resolving detector (dashed green) is shown in order to demonstrate the applicability of this measure to PNRDs with different underlying modes of operation.
is possible to explore using our detector tomography data since our treatment has explicitly avoided any binning of outcomes. One strategy is to only consider outcomes within windows centred on the peak maxima (figure 8). As would be expected, the highest confidence is obtained in the limit of the window width tending to zero, in which case the number of accepted outcomes would also tend to zero. This limit therefore upper bounds the performance of this strategy, and is plotted in figure 4. In order to demonstrate the utility of this analysis, we have neglected the probe state energy uncertainty incorporated into the other confidence measures, since our large uncertainty significantly masks the signal difference. Even with this simplification, the increase in confidence for our detector as compared to using the full space of outcomes is comparatively modest, since the overlap between different photon number POVM elements is dominated by the detection efficiency. However, as the detection efficiency of detectors improves, the intrinsic overlap between neighbouring Gaussian peaks is expected to become increasingly important. In this case, this post-selection strategy should become more effective.

6. Conclusions

We have presented a comprehensive approach to analysing the output of PNDRs. This centres on the use of principal component analysis to compress the output signals from detectors while maximizing the amount of retained information. In contrast to previous approaches, this output signal space compression allowed us to carry out detector tomography on a TES detector without arbitrarily binning the detection outcomes. This was further enabled by our novel approach to the numerical inversion of the Born rule using a Gaussian mixture model and maximum likelihood estimation. Based on these results, we considered the important role of the prior photon number distribution in interpreting detector outcomes, and introduced a quantitative measure of photon-number resolution that is able to account for the nonlinearity of the detector. We envisage such a measure being useful in directly comparing the efficacy of different detectors in resolving photon numbers as detector technologies continue to evolve.

Acknowledgments

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC EP/K034480/1), the European Office of Aerospace Research & Development (AFOSR EOARD; FA8655-09-1-3020) and the European Commission project SIQS. WSK is supported by EC Marie Curie fellowship (PIEF-GA-2012-331859). AD is supported by the EPSRC (EP/K04057X/1, EP/M01326X/1, EP/M013243/1). We thank Alvaro Feito for kindly helping us to obtain the data from his previous paper and Tim Bartley for his help in installing the TES detectors. Contribution of NIST, an agency of the US Government, not subject to copyright. This publication was supported by the Oxford RCUK Open Access Block Grant and accordingly can be accessed at doi:10.5287/bodleian:st74cq47v.

Appendix A. Calibrated light source

It is necessary to use a calibrated light source in order to produce coherent-state probes with known energies for detector tomography. Since we do not have access to a source calibrated to a radiometric standard, we built a calibrated source based on a Newport 918D-IG-OD3R power metre, which provides a specified calibration accuracy of 2% of absolute power and a linearity of better than 0.5%. This power metre was used to calibrate a series of fixed attenuators to reduce the output from a pulsed laser to the single-photon level with a known mean-photon number per pulse [30].

Our method uses a fibre beam splitter with a fixed fibre attenuator connected on one of the output ports, as shown in figure A1 (a). As long as this attenuation is well within the linear dynamic range of the power metre, we can obtain a calibration curve for the combined splitter-attenuator device relating the power measured at port 1A to the power at port 1B. In our case, the attenuation required to reach the single photon level is much greater.

![Figure 8](image-url) Post-selecting on outcomes within windows centred on the peak maxima (grey shaded regions) can be employed to boost the confidence of detected photon states.
than the dynamic range of the power metre. This forces us to use a second, calibrated splitter-attenuator device in series with the first figure A1(b). A weighted total least-squares algorithm [33] was used to find the total attenuation taking account of the absolute power errors in both variables. The total attenuation is given by the product of the two attenuators, but the errors in the measurements add linearly since they are not independent. Thus our final calibrated attenuation was found to be

\[ \eta_{\text{att}} = (2.10 \pm 0.16) \times 10^{-6}. \]  

(A1)

This relates the power measured at the monitor port 2A to the power at port 1B, such that \( P_{1B} = \eta P_{1A} \) (figure A1(c)). A variable attenuator is used to set the input power level before the calibrated attenuator so that we can probe our detector with a variety of coherent state amplitudes. We monitor the input power to the attenuator using port 2A and calculate the average photon number per pulse in port 1B which is coupled to the TES. The value of \( \eta_{\text{att}} \) also includes a correction to account for the Fresnel reflection from the unterminated fibre when plugged into the monitor power metre, which leads us to underestimate the total power that will be input when this fibre is instead directly coupled to the fibre leading to the TES. Fibre specifications give this loss at about 3.3% but there is a 1% uncertainty in this figure [34].

Appendix B. Comparison of confidence to system detection efficiency

When comparing the performance of different PNRDs, it is tempting to only consider the system detection efficiency. Although this metric is appealingly simple, here we demonstrate that it is not sufficient to fully quantify photon-number resolution. This is because this figure cannot account for the changing overlap between outcomes associated with different photon numbers. These changing overlaps often occur due to the nonlinearity of the detector response as a function of the input photon number. For TES detectors, this nonlinearity becomes increasingly apparent at higher photon numbers (as shown in figure 4). In figure B1 we compare the measured confidence for our detector with the calculated confidence for a detector with the same estimated system detection efficiency, but without overlapped detector outcomes for different detected photon numbers. In this case, the probability of detecting \( s \) photons, given an \( n \) photon input, is given by

\[ p(s|n) = \binom{n}{s} \eta^s (1 - \eta)^{n-s}. \]  

(B1)

Therefore the confidence for this detector is given by

\[ C_n = \sum_s \frac{p(s|n)\bar{p}(n)}{\sum_k p(k)\bar{p}(k)}. \]  

(B2)
As can be seen from the figure, for larger photon numbers this model increasingly overestimates the confidence for the detector. This mismatch will become progressively more important as the system detection efficiencies of detectors, highlighting the importance of a robust metric such as the confidence.

\[
\sum_{k} \sum_{s} \binom{n}{s} \eta^s (1 - \eta)^{n-s} p(n) = \sum_{k} \sum_{s} \binom{k}{s} \eta^s (1 - \eta)^{k-s} p(k). \tag{B3}
\]

As can be seen from the figure, the effective confidence from an analytic model based only on the system detection efficiency is plotted in green. As can be seen, as the photon number increases, this model increasingly overestimates the confidence, as it does not account for the nonlinearity in the detection system. This nonlinearity leads to an increasing overlap between the different photon number outcomes for higher photon numbers.

Figure B1. Reproduced from figure 7 is the calculated confidence \( C_n \) for our detector given a flat prior, assuming no uncertainty in the coherent-state probe energies (dashed red). For comparison, the effective confidence from an analytic model based only on the system detection efficiency is plotted in green. As can be seen, as the photon number increases, this model increasingly overestimates the confidence, as it does not account for the nonlinearity in the detection system. This nonlinearity leads to an increasing overlap between the different photon number outcomes for higher photon numbers.

References

[1] Calkins B et al 2013 Opt. Express 21 22657
[2] Day P K, Le Duc H G, Mazin B A, Vayonakis A and Zmuidzinas J 2003 Nature 425 817
[3] Art J 2006 Handbook of Biological Confocal Microscopy (Berlin: Springer) pp 251–64
[4] Lyashenko A, Nguyen T, Snyder A, Wang H and Arisaka K 2014 J. Instrum. 9 P11021
[5] Eisaman M D, Fan J, Migdall A and Poloyakov S V 2011 Rev. Sci. Instrum. 82 071101
[6] Natarajan C M, Tanner M G and Hadfield R H 2012 Supercond. Sci. Technol. 25 063001
[7] Marsili F et al 2013 Nat. Photonics 7 210
[8] Lita A E, Miller A J and Nam S W 2008 Opt. Express 16 3032
[9] Peacock A et al 1996 Nature 381 135
[10] Kardynal B E, Yuan Z L and Shields A J 2008 Nat. Photonics 2 425
[11] Fukuda D et al 2011 Opt. Express 19 870
[12] Avella A et al 2011 Opt. Express 19 23249
[13] Brida G, Ciavarella L, Degiovanni I F, Genovese M, Lolli M, Mingolla M G, Piacentini F, Rajteri M, Taralli E and Paris M G A 2012 New J. Phys. 14 085001
[14] Levine Z H, Gerrits T, Migdall A S, Samarov D V, Calkins B, Lita A E and Nam S W 2012 J. Opt. Soc. Am. B 29 2066
[15] Perina J, Haderka O, Allevi A and Bondani M 2014 Appl. Phys. Lett. 104 041113
[16] Brida G, Ciavarella L, Degiovanni I F, Genovese M, Migdall A, Mingolla M G, Paris M G A, Piacentini F and Poloyakov S 2012 Phys. Rev. Lett. 108 253601
[17] Luis A and Sánchez-Soto L L 1999 Phys. Rev. Lett. 83 3573
[18] D’Ariano G M, Maccone L and Presti P L 2004 Phys. Rev. Lett. 93 250407
[19] Mogilevtsev D 2010 Phys. Rev. A 82 1
[20] Lundeen J S, Feito A, Pregnell K L, Silberhorn C, Rath T C, Eisert J, Plenio M B and Walmsley I A 2008 Nat. Phys. 5 27
[21] Zhang L, Datta A, Coldenstrodt-Rongé H B, Jin X-M, Eisert J, Plenio M B and Walmsley I A 2012 New J. Phys. 14 115005
[22] Abdi H and Williams L J 2010 Wiley Interdiscip. Rev. Comput. Stat. 2 433
[23] Miller A J 2001 Development of a broadband optical spectrophotometer using superconducting transition-edge sensors PhD Thesis Stanford University
[24] Cooper M, Karpiński M and Smith B J 2014 Nat. Commun. 5 5332
[25] Ramelow S et al 2013 Opt. Express 21 6707
[26] Brida G et al 2012 Appl. Phys. Lett. 101 221112
[27] Bishop C M 2006 Pattern Recognition and Machine Learning (Berlin: Springer) p 738
[28] Burney J 2007 Transition-edge sensor imaging arrays for astrophysics applications PhD Thesis Stanford University
[29] Lunghi T, Kozrh B, Sanguinetti B and Zbinden H 2014 Opt. Express 22 18078
[30] Miller A J, Lita A E, Calkins B, Vayshenker I, Gruber S M and Nam S W 2011 Opt. Express 19 9102
[31] Datta A, Zhang L, Thomas-Peter N, Dorner U, Smith B J and Walmsley I A 2011 Phys. Rev. A 83 063836
[32] Botev Z I, Grotski J F and Kroese D P 2010 Ann. Stat. 38 2916–57
[33] Krystek M and Anton M 2007 Meas. Sci. Technol. 18 3438
[34] DeCusatis C 2013 Handbook of Fiber Optic Data Communication 4th edn (New York: Academic)