

## Measuring measurement: theory and practice

**A Feito<sup>1,2,5</sup>, J S Lundeen<sup>3</sup>, H Coldenstrodt-Ronge<sup>3</sup>,  
J Eisert<sup>4</sup>, M B Plenio<sup>1,2</sup> and I A Walmsley<sup>3</sup>**

<sup>1</sup> Institute for Mathematical Sciences, Imperial College London, SW7 2PG, UK

<sup>2</sup> QOLS, The Blackett Laboratory, Imperial College London,  
Prince Consort Road, SW7 2BW, UK

<sup>3</sup> Clarendon Laboratory, Oxford University, Parks Road, Oxford, OX1 3PU, UK

<sup>4</sup> Physics and Astronomy, University of Potsdam, 14476 Potsdam, Germany

E-mail: [ab1805@imperial.ac.uk](mailto:ab1805@imperial.ac.uk)

*New Journal of Physics* **11** (2009) 093038 (39pp)

Received 14 July 2009

Published 28 September 2009

Online at <http://www.njp.org/>

doi:10.1088/1367-2630/11/9/093038

**Abstract.** Recent efforts have applied quantum tomography techniques to the calibration and characterization of complex quantum detectors using minimal assumptions. In this work, we provide detail and insight concerning the formalism, the experimental and theoretical challenges and the scope of these tomographical tools. Our focus is on the detection of photons with avalanche photodiodes and photon-number resolving detectors and our approach is to fully characterize the quantum operators describing these detectors with a minimal set of well-specified assumptions. The formalism is completely general and can be applied to a wide range of detectors.

<sup>5</sup> Author to whom any correspondence should be addressed.

**Contents**

<b>1. Introduction</b>	<b>2</b>
<b>2. Detector tomography</b>	<b>3</b>
2.1. Definitions . . . . .	3
2.2. Assumptions . . . . .	3
2.3. Practical detector tomography . . . . .	4
2.4. Simple example: the perfect photcounter . . . . .	5
<b>3. Why detector tomography?</b>	<b>8</b>
3.1. The limits of calibration . . . . .	9
3.2. Quantitative entanglement verification . . . . .	9
<b>4. Modelling photodetectors</b>	<b>10</b>
4.1. Optical photon-number detectors . . . . .	10
4.2. Detector loss . . . . .	14
<b>5. Experimental reconstruction</b>	<b>14</b>
5.1. Results . . . . .	17
5.2. Detector Wigner functions . . . . .	20
<b>6. Ill conditioning and regularization</b>	<b>22</b>
6.1. Truncating the Hilbert space . . . . .	23
6.2. Pure versus mixed . . . . .	24
6.3. Smooth or not? . . . . .	29
6.4. Sharp and smooth . . . . .	31
<b>7. Conclusion</b>	<b>35</b>
<b>Acknowledgments</b>	<b>37</b>
<b>References</b>	<b>37</b>

**1. Introduction**

The quantum properties of nature reveal themselves only to carefully designed measurement techniques [1, 2]. In addition, most quantum information applications both computational and cryptographic, rely on a certain knowledge of the measurement apparatuses involved. Indeed for these protocols we often need to ensure that we associate detector outcomes with the correct quantum mechanical operation or quantum state. More critically, the assumption of a fully characterized detector completely underlies both quantum state tomography (QST) and quantum process tomography (QPT) [3]–[5]. State tomography has become an important tool for characterizing states, partially due to the realization that non-classical states are a resource for performing tasks such as enhanced precision metrology, quantum communication and quantum computation. Often taken for granted, measurement also plays a crucial role in these tasks and in some can even eliminate the need for entanglement. In enhanced precision metrology, appropriate measurements alone can give rise to super-resolution [2] and Heisenberg-limited sensitivity [1]. In communication, measurement allows entanglement swapping, and thus, is central to quantum repeaters. And for computing, measurement-based schemes enable quantum computation [6, 7]. It follows that measurement should also be considered a resource for quantum protocols. In QST a given number of measurements on many copies of an unknown

state reveal its density operator [3, 8, 9]. Characterizing the operators that govern an evolution or a channel—QPT—amounts to applying the process to a set of input states, and subsequently fully characterizing the output states [4, 5], [10]–[12]. In this paper, we study quantum detector tomography (QDT) [13]–[17], in which a detector’s outcome statistics in response to a set of input states determine the operators that describe that detector. State and detector tomography evidently exhibit a dual role: either the input is well known and the detector is to be characterized, or the detector is well known and the state is to be tomographically reconstructed.

Throughout the paper, we focus on examples from optics. However, QDT is a general concept, applicable to any quantum detector. Building upon previous theoretical descriptions [13]–[16], we will introduce the concept of detector tomography in the following section. Alongside, we will present examples detailing the reconstruction of simple detectors to introduce the key concepts. Subsequently, based on recent experiments [17] we will present the methods used in optical detector tomography and the convex optimization methods [18], which allow an efficient and simple numerical optimization. Finally, in the last section, we will discuss some of the theoretical and experimental challenges involved and how to address them.

## 2. Detector tomography

In this section, we will first define the concept of detector tomography. As it involves a number of fundamental assumptions, we will list and discuss them highlighting what constitutes a practical realization. Finally, we will describe a simple theoretical example.

### 2.1. Definitions

In quantum mechanics, the operator describing a measurement apparatus is, in its most general form, a positive operator-valued measure (POVM). The POVM elements  $\{\pi_n\}$  describe the possible outcomes, labelled here by  $n$ . Particularly, for a projective measurement, the POVM elements are orthogonal and simplify to the familiar form

$$\pi_n = |\psi_n\rangle\langle\psi_n|.$$

In quantum optics, an example of a POVM that consists of projectors is that of an eight-port homodyne,  $\{|\alpha\rangle\langle\alpha| : \alpha \in \mathbb{C}\}$ .

Now, given an input state  $\rho$ , the probability  $p_{\rho,n}$  of obtaining output  $n$  is

$$p_{\rho,n} = \text{Tr}(\rho\pi_n). \quad (1)$$

It follows that the POVM elements must be positive semi-definite,  $\pi_n \geq 0$ , while observing

$$\sum_n \pi_n = \mathbb{1}, \quad (2)$$

if we want probabilities adding up to one. Inverting equation (1) to extract  $\pi_n$  subject to the aforementioned conditions is the task behind detector tomography.

### 2.2. Assumptions

Detector tomography raises fundamental questions about the kind of information we can extract from Nature. It is reasonable to think that state tomography performed with poorly

characterized detectors can lead to unwanted errors. In addition, for quantum cryptography, a mischaracterization of states or detectors may lead to channels through which an eavesdropper may attack invalidating certain security proofs (see for example [19]). Indeed, if we misjudge the noise level, then we misjudge the information the eavesdropper possesses. Interestingly, quantum key distribution with totally uncharacterized detectors is possible in principle, based entirely on correlations violating Bell's inequalities, but at the expense of having much smaller rates [20, 21]. An in depth discussion of our assumptions is therefore necessary to avoid unwanted errors in our detector estimation.

Generally, there are going to be assumptions about the input states produced by the source. In the reconstruction, we will often need to assume we can truncate an infinite-dimensional Hilbert space, such as in the case of photon-number. On the detector side, a common assumption will be that the detector is memory-less: the previous measurements do not modify the result of future measurements. For example, this assumption fails when detector deadtimes are longer than the time between consecutive measurements. All of these assumptions need to be tested.

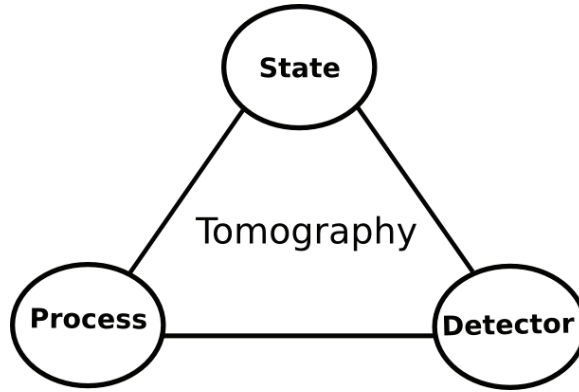
More generally, we can ask what the working minimal set of assumptions happens to be. An assumption-free tomography could use a complete black box approach: prepare a collection of unknown states, measure them and try to draw some conclusions about both the detector and the states. Specifically, we could have some classical controls to prepare quantum states  $\{\rho_\lambda\}$  characterized by the index  $\lambda$  and some classical pointer to indicate the possible outcomes, labelled  $n$ . Minimizing the set of assumptions would constrain us to draw our conclusions exclusively from the joint probability distribution  $\{p_{\lambda,n}\}$ .

To further constrain the problem we can add the standard assumptions: An underlying Hilbert space of fixed dimension, normalized positive density matrices and positive POVM elements  $\{\pi_n\}$  satisfying  $\sum_{n=1}^N \pi_n = \mathbb{1}$ . However, without further assumptions, the relationship between  $\lambda$  and  $\rho$ 's  $d^2 - 1$  parameters is completely unknown, as is that between  $n$  and  $\{\pi_n\}$ 's  $(N - 1)d^2$  parameters.

This discussion highlights the inherent difficulties that a fully general inference (or tomographic) scheme entails. Reasonable assumptions are thus needed but the question of general tomography remains an interesting one to be explored. In this direction, some progress has been made in self-testing maps. In this context, states are prepared with classical recipes and families of unitary gates are revealed with few assumptions about the quantum states (however, known measurements in the computational basis are required) [22]. As shown diagrammatically in figure 1, we can divide any experiment into preparation, evolution and measurement. If one of the elements of the triad is unknown or missing then we need to previously characterize the other two making assumptions in the process.

### 2.3. Practical detector tomography

In state tomography, one must perform a set of measurements  $\{\pi_n\}$  spanning the space of the density operator to be reconstructed. If the state is defined on a  $d$ -dimensional Hilbert space, then it will be fully specified by  $d^2 - 1$  real parameters (respecting the constraint of unit trace). To fully characterize a quantum detector, we need the data obtained from measuring input states from a well-characterized source. To recover all the POVM elements  $\{\pi_n\}$  from the measured statistics  $p_{\rho,n}$  the probe states or input states must also be chosen to form a set  $\{\rho_j\}$  that is tomographically complete: the span of the operators  $\{\rho_j\}$ —which are not necessarily linearly independent—must be the entire space from which  $\pi_n$  is taken. A spanning set forming an



**Figure 1.** It is generally possible to divide an experiment into preparation evolution and measurement. However, if one corner in the above diagram is unknown or missing, then we need assumptions about the other two.

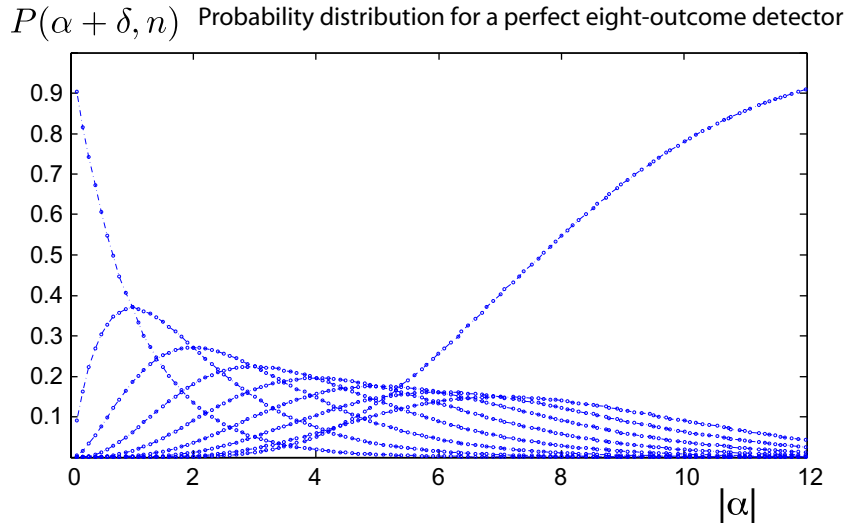
operator basis will have at least  $d^2(k-1)$  elements for a  $k$  outcome detector. In principle, this is sufficient to calculate the direct inversion of equation (1). However, experimental detector tomography carries additional requirements. Clearly, the probe states should be previously characterized, and large numbers of them should be easily and reliably generated. In the case of optical detectors, coherent states are ideal candidates since a laser can generate them directly and we can create a tomographically complete set by transforming their amplitude through attenuation (for example, with a beam splitter (BS)) and a phase shifter (an optical path delay). From a practical perspective we will need to generate coherent states with amplitudes within the order of magnitude of the detector so that detection events have reasonably high probabilities and the experiment can be completed in a short time. Using input states  $\{|\alpha\rangle\langle\alpha| : \alpha \in \mathbb{C}\}$  one can then reconstruct the  $Q$ -function of the detector [13] which is simply proportional to the measured statistics,

$$p_{\alpha,n} = \frac{1}{\pi^2} \langle \alpha | \pi_n | \alpha \rangle = \frac{1}{\pi} Q_n(\alpha). \quad (3)$$

Since  $Q_n(\alpha)$  of each POVM contains the same information as the element  $\pi_n$  itself, predictions of the detection probabilities for arbitrary input states can then be calculated directly from the  $Q$ -function. Note that underlying the use of single mode coherent states lies the assumption that our detector operates in the same optical mode as these states. A more general tomography would require a family of coherent states spanning any mode structure relevant to the detector (for example, with different spectral and momentum distributions). This is a challenging experimental feat that in our case is circumvented using detectors, which are coupled to the states with single-mode fibres. Let us then give an example of how to use the detector tomography framework with single-mode coherent states to characterize a simple idealized detector.

#### 2.4. Simple example: the perfect photocounter

Consider as a simple example the case of a perfect photon-number detector. This projective measurement is characterized by its POVM elements,  $\{\pi_n = |n\rangle\langle n| : n = 0, \dots, N\}$ . In the simplest of scenarios, using pure number states,  $\{\rho = |m\rangle\langle m| : m = 0, \dots, N\}$ , the



**Figure 2.** Outcome probability distributions for an eight-outcome detector. Each curve represents the probability of that outcome (0 clicks, 1 click, etc) happening versus the value of the intensity of the coherent state arriving at the detector.

characterization would be trivial, since the statistics

$$p_{\rho,n} = \delta_{n,m}$$

would immediately characterize our detector. From a more practical perspective, pure number states are very hard to generate, especially for high photon numbers. Using coherent states is therefore a more realistic approach. Assuming a collection  $\{|\alpha_i\rangle : i = 1, \dots, D\}$  of perfect coherent states our statistics would then become

$$\begin{aligned} p_{\alpha_i,n} &= \text{Tr}(|\alpha_i\rangle\langle\alpha_i|n\rangle\langle n|) \\ &= e^{-|\alpha_i|^2/2} \frac{|\alpha_i|^{2n}}{n!}. \end{aligned}$$

Of course in an experiment, we would not know the POVM elements in advance, and we would need to express them with a generic expression such as

$$\pi_n = \sum_{k,p} \theta_{k,p}^{(n)} |k\rangle\langle p|$$

suffering only the constraint  $0 \leq \pi_n \leq 1$ . Fortunately, these operators can be simplified: interposing a phase shifter in the coherent beam's path we can check if the statistics are independent of the phase. If they are we can infer the phase independence of the POVM. Our operators then become  $\pi_n = \sum_k \theta_k^{(n)} |k\rangle\langle k|$  and the statistics can be expressed as

$$p_{\alpha_i,n} = e^{-|\alpha_i|^2/2} \sum_{k=0}^{\infty} \frac{|\alpha_i|^{2k}}{k!} \theta_k^{(n)}. \quad (4)$$

For a perfect photon-number detector that can discriminate up to eight photons, the outcome probability distributions would be the one shown in figure 2.

We now consider how one would estimate the POVM elements from these probability distributions, which form a set of simple simulated data. Our goal is to invert equation (4). To do

so, we can write a matrix version of the equation. Given the set of coherent states  $\{\alpha_1, \dots, \alpha_D\}$ , and truncating the number states at a sufficiently large  $M$ , we can write

$$P = F \Pi. \quad (5)$$

Here, we have taken

$$\begin{aligned} p_{\alpha_i, n} &= \langle \alpha_i | \pi_n | \alpha_i \rangle \\ &= \langle \alpha_i | \left( \sum_{k=0}^M \theta_k^{(n)} |k\rangle \langle k| \right) | \alpha_i \rangle \\ &= e^{-|\alpha_i|^2} \sum_{k=0}^M \frac{|\alpha_i|^{2k}}{k!} \theta_k^{(n)} \\ &= \sum_{k=0}^M F_k(\alpha_i) \theta_k^{(n)} \\ &= \sum_{k=0}^M F_{i,k} \Pi_{k,n}. \end{aligned}$$

So the matrix  $P$  has entries

$$P_{i,n} = p_{\alpha_i, n},$$

$F$  entries  $F_{i,k} = F_k(\alpha_i)$ , and  $\Pi$  entries  $\Pi_{k,n} = \theta_k^{(n)}$ . For an  $N$ -outcome detector, this gives the matrix dimensions of  $P \in \mathbb{C}^{D \times N}$ ,  $F \in \mathbb{C}^{D \times M}$  and  $\Pi \in \mathbb{C}^{M \times N}$ .

Now, to obtain  $\Pi$ , we can simply solve the convex optimization problem

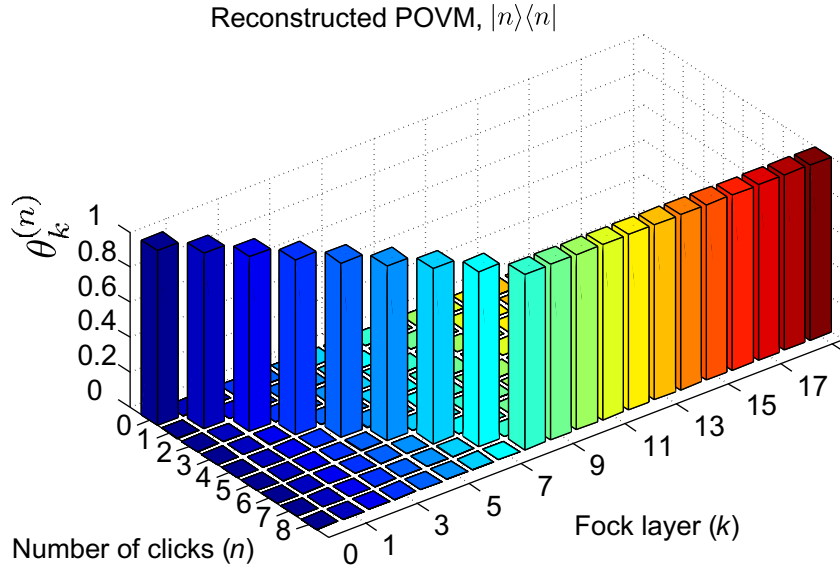
$$\begin{aligned} \min \{ \|P - F \Pi\|_2 \}, \\ \text{subject to } \pi_n \geq 0, \quad \sum_{n=1}^N \pi_n = \mathbb{1}. \end{aligned} \quad (6)$$

This is a convex problem, because the norm  $\|\cdot\|_2$ , defined as  $\|A\|_2 = (\sum_{i,j} |A_{i,j}|^2)^{1/2}$  for matrix  $A$  is a convex function and the positivity constraint  $\pi_n \geq 0$  is semi-definite. The result of a single such minimization is shown in figure 3, where we recover the expected POVM elements

$$\left\{ |0\rangle\langle 0|, |1\rangle\langle 1|, \dots, |7\rangle\langle 7|, \mathbb{1} - \sum_{k=0, \dots, 8} |k\rangle\langle k| \right\}. \quad (7)$$

It is remarkable how robust this method is in the face of statistical noise in the probe states. Indeed, if instead of using  $p_{\alpha_i, n} = \text{Tr}(|\alpha_i\rangle\langle\alpha_i| \pi_n)$  we use,  $p_{\alpha_i, n} = \text{Tr}(|\alpha_i + \delta_i\rangle\langle\alpha_i + \delta_i| \pi_n)$  where  $\{\delta_i\}$  represent a 2% random noise, then the results of the optimization are indistinguishable in the scale of figure 3. This will be discussed in more detail in later sections as it relates to the technical noise of the laser. With this example we have shown the power of detector tomography in a simple case. But such a detector might not need detector tomography altogether due to its simplicity. Let us then motivate further the use of detector tomography and its applications.





**Figure 3.** Reconstructed POVM of a theoretical photon-number detector with nine bins. The POVM elements,  $\pi_n = \sum_k \theta_k^{(n)} |k\rangle\langle k|$ , are the result of the optimization in equation (6), and show a perfect result even though a 2% random noise was added to the data (changing the value of the coherent state amplitudes).

### 3. Why detector tomography?

In the previous section, we have defined detector tomography and detailed the assumptions involved. In this section, we will further motivate this technique giving examples of its performance.

In spite of its first successful applications, one could question the need for detector tomography. After all, detectors have been calibrated in the past without tomographic techniques. However, as quantum technology makes striking advances, quantum detectors are becoming more complex and, thus, susceptible to imperfections that are not incorporated in the bottom-up approach of modelling. In contrast, with tomography one can design detectors with a top-down approach by fully characterizing the final detector operation. For example, photo-detection has seen the advent of single-carbon nanotube detectors [23], charge integration photon detectors (CIPD) [24], visible light photon counters (VLPC) [25], quantum dot arrays [26], superconducting edge and picosecond sensors [27, 28] or time-multiplexing detectors (TMD) based on commercial Si-APDs [29, 30]. Certainly, understanding and modelling in full detail the noise, loss and coherence characteristics of these technologies are far from trivial. Detector tomography is an answer to those challenges and, additionally, will allow us to benchmark competing detectors. Such characterized detectors also allow for the preparation of non-Gaussian states in a certified manner, such as photon subtracted states. Hence, such detectors are readily usable in protocols such as entanglement distillation based on non-Gaussian states [31]–[33], in schemes increasing entanglement by means of photon subtraction [34], enhancing the teleportation fidelity [35, 36] or in other applications of non-Gaussian states [37, 38], specifically in the context of quantum metrology.

Let us discuss more precisely how detector tomography can provide an advantage with respect to traditional calibration methods.



### 3.1. The limits of calibration

Standard calibration methods require a model of the detector. The parameters of this model are then estimated experimentally using states and standard assumptions. However, for specific applications, these models can become very complicated with a daunting number of parameters. For example, a detector as simple as a Yes/No avalanche photodiode (APD) becomes very hard to model when all noise sources are studied [39, 40]. Tomography sidesteps this by enabling the operational detector to be measured directly. Another advantage can be to avoid errors or pitfalls from standard calibration. An example is the use of Bell-state detectors that can appear to be working while frequency correlations in the input states obscure the results [41, 42]. To avoid such pitfalls detectors are often used in the ranges where their behaviour is easily calibrated. Detector tomography could extend the range of applicability of existing detectors and help design more complicated ones. Let us now look at other advantages of full detector tomography.

### 3.2. Quantitative entanglement verification

Once a detector is fully characterized, it can be used to characterize states in a certified fashion. In this respect, a detector is still useful if it is imperfect in the sense that its POVM elements are not just unit rank projections: one can use them in an estimation problem, or in a setting that unambiguously estimates properties of a state. This is one of the key applications of detector tomography, in that imperfect devices can be used to very reliably perform estimation. A specifically important example is the direct detection of entanglement: once detectors are characterized, one can perform measurements and then find lower bounds to entanglement measures, based on these measurements. This is an idea presented in [47] (see also [48]–[51]). Here, we discuss the specific issues related to infinite-dimensional systems and photon counters.

Assume that we have performed more than one type of measurements, labelled by  $k$ , for which we have completely characterized the POVM elements  $\{\pi_n^{(k)}\}$  to great accuracy. Using two such devices one can make local measurements on each part of a bipartite state. With the data from

$$d_{n,m}^{(k,l)} = \text{Tr}((\pi_n^{(k)} \otimes \pi_m^{(l)})\rho)$$

one can then ask for the minimal degree of entanglement consistent with it. This approach does not make any assumptions on the preparation of states, and works even for detectors with a low detection efficiency. This will already be incorporated in the stated bound. The unambiguously minimal degree of entanglement, say in terms of the negativity, is then given by the solution of the optimization problem [47]:

$$\begin{aligned} \min E_N(\rho) &= \|\rho^\Gamma\|_1 - 1, \\ \text{subject to } \text{Tr}((\pi_n^{(k)} \otimes \pi_m^{(l)})\rho) &= d_{n,m}^{(k,l)}, \\ \text{Tr}(\rho) &= 1, \quad \rho \geq 0. \end{aligned}$$

One easily finds lower bounds to the optimal solution to this problem by considering

$$P = \sum_{n,m,k,l} \alpha_{n,m}^{(k,l)} \pi_n^{(k)} \otimes (\pi_m^{(l)})^T + \beta \mathbb{1}$$

such that  $\|P\| \leq 1$ . Then a lower bound is readily given by [48]

$$\begin{aligned} \|\rho^\Gamma\|_1 &\geq \text{Tr}(\rho^\Gamma P) = \text{Tr}(\rho P^\Gamma) \\ &= \sum_{n,m,k,l} \alpha_{n,m}^{(k,l)} d_{n,m}^{(k,l)} + \beta. \end{aligned}$$

The optimal lower bound, based on such an approach, is in turn the solution of the convex optimization problem [18] in  $\alpha_{n,m}^{(k,l)}$  and  $\beta$  given by

$$\begin{aligned} \max \quad & \sum_{n,m,k,l} \alpha_{n,m}^{(k,l)} d_{n,m}^{(k,l)} + \beta, \\ \text{subject to} \quad & -\mathbb{1} \leq \sum_{n,m,k,l} \alpha_{n,m}^{(k,l)} \pi_n^{(k)} \otimes (\pi_m^{(l)})^T + \beta \mathbb{1} \leq \mathbb{1}. \end{aligned}$$

Some conditions need to be examined to make sure that the solution to this formulation (dual optimal) coincides with the solution of the original formulation (primal optimal). For linear programs or for semi-definite problems (SDP) as the one described these conditions are easily satisfied [18]. As we will see this type of optimization method is also useful for detector tomography itself.

Here, if one just makes use of POVM elements with a finite support, as one has in photon counting with an additional phase reference, then such bounds will provide very strong lower bounds. However, it will not provide good bounds for unbounded operators such as in homodyning. Hence, without having detectors of high efficiency, and without assumptions on the preparation of the entangled state, one can use a characterized detector to certify entanglement in quantitative terms.

#### 4. Modelling photodetectors

Before carrying out the full tomography of the photodetectors this section will show how to model them. These models will have to take into account the loss and finite number of bins from these detectors.

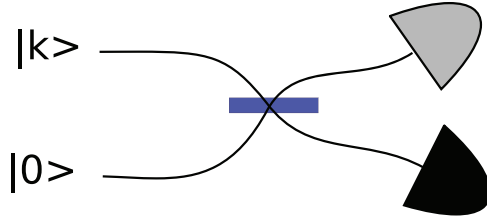
As we have seen, the aim of detector tomography is to identify the physical POVM closest to the experimental data with minimal assumptions on the functioning of the detectors. To compare this method with a more traditional calibration method let us study a photodetector modelling example. We will do so for an APD and a photon-number resolving detector able to detect up to eight photons [29]. In the next section, we will compare these models to the experimental results derived without any model.

##### 4.1. Optical photon-number detectors

An important detector in quantum optics is the single-photon counting module based on a silicon APD. It has two detection outcomes, either registering an electronic pulse (1 click) or not (0 clicks). A loss-free perfect version of it would implement the Kraus operators

$$\{|0\rangle\langle 0|, \mathbb{1} - |0\rangle\langle 0|\}, \quad (8)$$

distinguishing between the presence or absence of photons. However, some photons are absorbed without triggering a pulse. This loss can be modelled placing a BS in front of the



**Figure 4.** Diagram of a simplified two-bin multiplexing detector.

perfect detector [52]. The POVMs describing a detector with a BS of transmittivity  $\eta$  can then be written as,

$$\text{NO CLICK} : \pi_0 = \sum_{n=0}^{\infty} (1 - \eta)^n |n\rangle\langle n|, \quad (9)$$

$$\text{CLICK} : \pi_1 = \mathbb{1} - \sum_{n=0}^{\infty} (1 - \eta)^n |n\rangle\langle n|. \quad (10)$$

disregarding after pulsing or dark counts [53]. Having only two outcomes, this detector cannot distinguish the number of photons present.

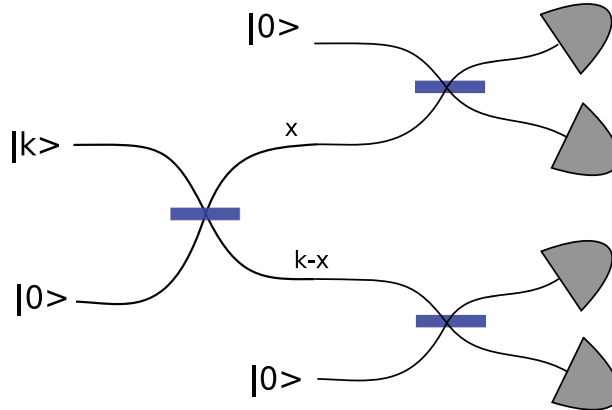
A more advanced detector called TMD does have certain photon-number resolution. It splits the incoming pulse into many temporal bins, making unlikely the presence of more than one photon per bin. All the time bins are then detected with two APDs. Summing the number of 1-click outcomes from all the bins one can then estimate the probability of having detected a number of incoming photons. This detector is not commercially available but one has been constructed by the Ultrafast Group in Oxford [29]. It has eight bins in total (four time bins in each of two output fibres) and thus nine outcomes—from zero to 8 clicks.

The theoretical description of this detector is a bit more involved since there is what we call the ‘binning problem’. Indeed, in addition to loss there is a certain probability that all photons will end up in a single time bin, or more generally that  $k$  incoming photons will result in less than  $k$  clicks. To account for the details describing these probabilities, we use a recursive relation [54, 55]. Our goal is to describe the following probability distribution:

$p^N(j/k)$ : Probability of having  $j$  clicks given that there were  $k$ -incident photons and that the detector has  $N$  bins (or modes). Let us start with the simplest possible TMD that consists of an input port, a BS (with reflectivity and transmittivity  $R$  and  $T$ ) and two YES/NO detectors. This detector is shown in figure 4 and has two bins.

For this simple example, the probability distribution we are after is  $p^2(j/k)$ . We will show how to calculate  $p^2(j/k)$ ,  $p^4(j/k)$  and then how to go from  $p^N(j/k)$  to  $p^{2N}(j/k)$ . For the two-bin case from figure 4 consider a BS with transmittivity  $T$  and reflectivity  $R$ . In that case:

- $p^2(j, 0) = \delta_{j,0}$  (if no photons are present we will only register 0 clicks).
- $p^2(1, k) = T^k + R^k$  (with probability  $T^k$ ,  $k$  photons end up in the lower bin and the same holds for the upper bin with  $R^k$ . The probability of a single click is the sum of these independent probabilities).
- $p^2(2, k) = 1 - T^k + R^k$  (if  $k \neq 0$  then only two events may happen: 1 click or 2. This complementary event has therefore  $P = 1 - (\text{probability of 1 click})$ .)



**Figure 5.** Diagram of a simplified four-bin multiplexing detector. The first BS distributes  $k$  photons according to a binomial distribution between the two two-bin multiplexing detectors of the second stage.

In the case of a four-bin detector shown in figure 5,  $k$  incoming photons are distributed to two two-bin detectors according to a binomial distribution.

Now let us evaluate the probability for the upper two-bin detector to register  $s$  counts if  $x$  photons entered while registering  $m$  clicks in the lower two-bin detector if  $k - x$  entered the lower port. This should be  $p^2(s, x)p^2(m, k - x)$  weighted by the probability that  $x$  photons enter the upper branch and  $k - x$  the lower one, which is

$$\binom{k}{x} T^{k-x} R^x.$$

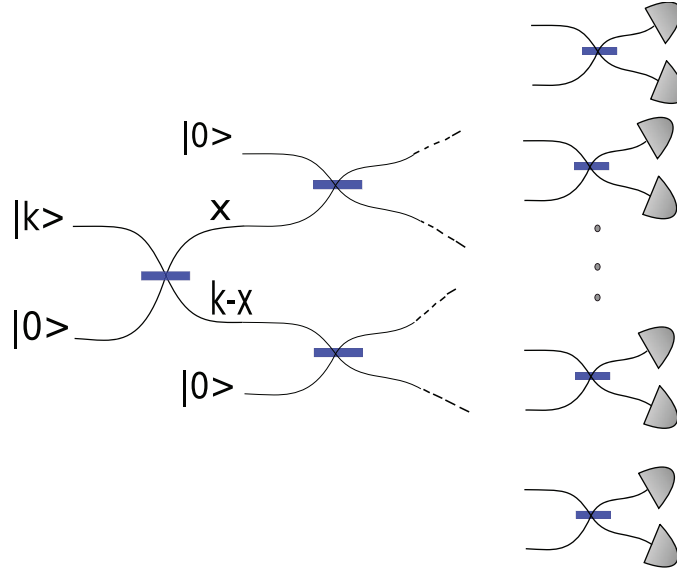
Now, the probability that  $j$  counts are found overall is found summing the weighted probability over all possible ways that the detectors can find  $j$  counts (i.e.  $m + s = j$ ) and summing over all possible ways of distributing  $k$  photons:

$$p^4(j/k) = \sum_{x=0}^k \sum_{m+s=j} \binom{k}{x} T^{k-x} R^x p^2(s, x) p^2(m, k - x).$$

We can extend the same argument to  $2N$ . Imagine we know  $p^N(j/k)$ . Now, for that detector to become a  $2N$ -bin detector all we need is to couple two of them to a BS, which will distribute the  $k$  photons as described above and as shown in figure 6. In that same fashion, we can then define the recursive relation

$$p^{2N}(j/k) = \sum_{x=0}^k \sum_{m+s=j} \binom{k}{x} T^{k-x} R^x p^N(s, x) p^N(m, k - x). \quad (11)$$

This of course can be generalized to accommodate a different BS reflectivity at each node (adding an index to  $T$  and  $R$  to account for its position). Based on this recursion and once we determine all  $T$  and  $R$ , we can write a simple and efficient program to generate the



**Figure 6.** Diagram of a  $2N$ -bin multiplexing detector. The first BS distributes  $k$  photons according to a binomial distribution between the two next  $N$ -bin stages.

corresponding theoretical POVMs. For example, a six-outcome detector would have a POVM, which can be captured in the following matrix:

	$\pi_0$	$\pi_1$	$\pi_2$	$\pi_3$	$\pi_4$	$\pi_5$
$ 0\rangle\langle 0 $	1	0	0	0	0	0
$ 1\rangle\langle 1 $	0	1	0	0	0	0
$ 2\rangle\langle 2 $	0	0.128	0.871	0	0	0
$ 3\rangle\langle 3 $	0	0.0168	0.334	0.648	0	0
$ 4\rangle\langle 4 $	0	0.00226	0.100	0.495	0.400	0
$ 5\rangle\langle 5 $	0	0.000309	0.0283	0.265	0.508	0.197
$ 6\rangle\langle 6 $	0	0.0000428	0.00772	0.123	0.421	0.447
$ 7\rangle\langle 7 $	0	0.00000601	0.00208	0.0536	0.291	0.653
$ 8\rangle\langle 8 $	0	0.000000852	0.000565	0.0224	0.181	0.794

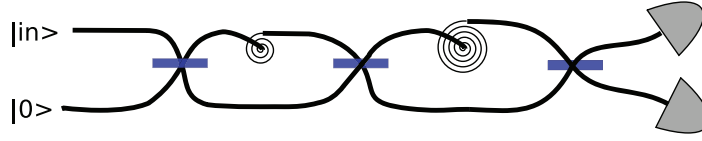
where  $B_{k,j} = p^5(j/k)$ ,  $j = 0, \dots, 5$  and  $k = 0, \dots, 8$ . For example, the 5-click event has a POVM element,

$$\pi_5 \simeq 0.2|5\rangle\langle 5| + 0.4|6\rangle\langle 6| + 0.6|7\rangle\langle 7| + 0.8|8\rangle\langle 8|,$$

etc. More generally, the measured statistics are related to the incoming photons by

$$p_j = \sum_k p^N(j/k)q_k,$$

where  $p_j$  is the probability of detecting  $j$  counts and  $q_k$  the probability that  $k$  photons arrived to the TMD [30]. The matrix  $\mathbf{B}$  introduced above then relates probabilities and density matrices through:  $\mathbf{p} = \mathbf{B} \cdot \rho$ .



**Figure 7.** Diagram of a simplified eight bin TMD. The spirals represent a delay in the optical fibre.

#### 4.2. Detector loss

TMD detectors have various sources of loss, meaning that photons are absorbed before triggering a detection event. The major sources of loss are the coupling to the fibres, the absorption and scattering in the delay fibres and the non-unit efficiency of the detectors [56]. A full description of the effect of losses is certainly complex, since loss occurs at many stages of the detector. One can model loss simply as a BS diverting photons out of an input state. However, one would have to include a BS before the detector (fibre coupling), a BS at each stage of fibre and a BS in front of each APD, altering equation (11) accordingly. Instead we give an effective description modelling loss with a single BS in front of the detector. The matrix capturing the losses has entries that are given by

$$L_{k',k} = \binom{k}{k'} \eta^{k'} (1 - \eta)^{k-k'},$$

being the binomial distribution accounting for loss since  $L_{k',k} = 0$  for all  $k < k'$ . Now combining both descriptions, we can decouple the loss from the binning, putting a BS coupled to the environment before the  $N$ -bin TMD resulting in

$$p_j = \sum_{k,k'} p^N(j/k') L_{k',k} q_k.$$

This relationship expresses how the incoming photons experience loss and then are distributed among the available modes. The model of the TMD described up to this point would be the one needed without detector tomography. One could for example try to measure independently the transmittivities of the inner BS, reconstruct the convolution (or binning) matrix  $\mathbf{B}$  and try to estimate the overall loss to include it in the matrix  $L$ . This would help us build a model of the TMD sketched in figure 7. By contrast, using detector tomography, we do not need to know anything about bins, BS, inner detectors or specific loss mechanisms. Moreover, anything left out of our detector model (e.g. dark counts, after pulsing, etc) would be included in a tomographic characterization.

### 5. Experimental reconstruction

We now turn to the description of the experimental realization, starting off with general considerations necessary for experimental detector tomography. We then describe the testbed used for the APDs and TMDs, which had been proposed in [16], in detail explaining how the experimental parameters were chosen and how potential sources of errors were taken into account.

As mentioned earlier, coherent states are ideal probe states with which to characterize optical detectors. This holds true for any optical detector, including polarization detectors, frequency detectors (e.g. spectrometers), and even detectors that discriminate inherently quantum states (e.g. a photonic Bell-state detector). In most of these cases, we are only interested in the detector's action at a particular input photon-number  $n$  (usually  $n = 1$ ). Still, we can reconstruct the detector POVM in the full photon-number basis and then restrict our attention to a particular subspace. It is interesting that an optical state which can be fully described in a classical theory of electromagnetism can be used to characterize uniquely quantum detectors. To characterize a completely unknown detector (i.e. a black box) one would need to vary all the available parameters of the probe set of coherent states: spatial-temporal mode, polarization, phase and amplitude (ensuring a tomographically complete set of states is constructed). However, given that frequency, time, position, momentum and amplitude have infinite range, this is impossible in practice. Consequently, one is required to make realistic assumptions about the range of operation and sensitivity of any unknown detector. One might additionally neglect some of these optical parameters if one is only interested in a particular aspect of the full characterization.

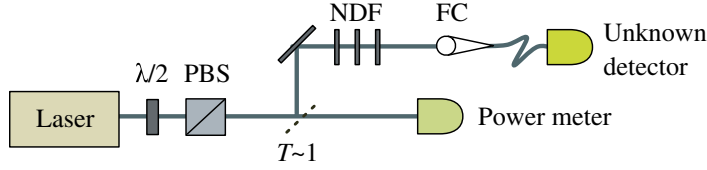
The subjects of our detector tomography, the APD and TMD, possess the unique features of single-photon sensitivity and photon-number resolution, respectively. To characterize these features we vary only the amplitude and phase of probe coherent states, while fixing the spatial-temporal mode. In particular, the input wavepackets have a time extent shorter than the time window of the electronics, and the centre wavelength is within range of the detectors. To fix the position or momentum of the light mode, light is coupled to both types of detector through single-mode optical fibre.

Critically, for the detectors to perform as intended and in order to ensure the detectors are memoryless, the wavepackets must be preceded and followed by time intervals in which there is no input light. The APD is known to have a deadtime of roughly 50 ns; a detection that occurs before the input wavepacket arrives at the detector will make the detector insensitive to the probe coherent state. The TMD splits the incoming wavepacket into time bins spread over 500 ns. The inverse of these two timescales then sets an upper limit on the rate at which we can send probe states to the detectors. We further limit the rate to ensure the detectors do not heat up, which would change their properties over time. These time variant features of the detectors could be illuminated with detector tomography but this would be unnecessarily complicated, as they are quickly evident without use of the full tomography procedure.

With these experimental considerations based on the examined detectors in mind, we need to produce the set of probe states.

We use a cavity dumper (APE Cavity Dumper Kit) to reduce the repetition rate of a pulsed mode-locked Ti:Sapph laser to  $R$ . Long-term drift of the laser power over 1 million pulses was  $<0.5\%$ . Our laser randomly varies in energy between subsequent pulses with a standard deviation of  $1.88 \pm 0.02\%$  of  $|\alpha|^2$ . We vary the amplitude  $\alpha$  of our probe coherent states by rotating their polarization with a half-waveplate ( $\lambda/2$ ) in front of a Glan–Thompson polarizer (P) as shown in figure 8. We attenuate the coherent states by reflecting them from a BS ( $T = 95\%$ ) and three neutral density filters (NDF) (i.e. spectrally flat attenuators). Note that  $R/T$  for the BS was measured with a relative deviation smaller than 1%. Along with some loss upon coupling into a single-mode fibre, we collect these attenuations together in an overall attenuation factor  $\gamma$ . We test for any variation in  $\gamma$  as a function of  $\alpha$  (e.g. which





**Figure 8.** Experimental setup: the amplitude of the probe coherent state is attenuated with a half-waveplate ( $\lambda/2$ ) and a Glan–Thompson polarizer (P). The light is then further attenuated by frequency independent filters or NDF and coupled into a fibre (FC).

might be caused by rotating the waveplate if it had a wedge) and find that the variation is less than 1%.

When operated with a gain high above their lasing threshold, lasers produce light well approximated by a coherent state. Coherent states (and statistical mixtures thereof) are unique amongst pure optical states in that, at a BS, the transmitted and reflected states are unentangled. Consequently, through attenuation we can vary the amplitude of a coherent state without changing its nature. In homodyne state tomography, one must use a balanced detection technique to negate technical noise in the laser. In contrast, by attenuating the laser light to the single-photon level in our detector tomography scheme, the resulting coherent state possesses an inherent amplitude uncertainty that renders the technical noise insignificant.

We choose a centre wavelength  $\lambda$  and a full width at half maximum (FWHM) bandwidth of  $\Delta\lambda$  that are appropriate for each detector. In the case of the APD detector (a Perkin Elmer SPCM-AQR-13-FC) we set  $\lambda = 780 \pm 1$  nm,  $\Delta\lambda = 20$  nm, and chose the appropriate rate  $R = 1.4975 \pm 0.0005$  kHz,  $J = 1\,472\,967$ , and  $\gamma = (5.66 \pm 0.08) \times 10^{-9}$ . For the TMD detector we set  $\lambda = 789 \pm 1$  nm,  $\Delta\lambda = 26$  nm,  $R = 76.169 \pm 0.001$  kHz,  $J = 38\,084$  and  $\gamma = (8.51 \pm 0.11) \times 10^{-9}$ .

Let us then discuss which errors do affect the experiment. Let us first note that, as of yet, there are no direct techniques to calibrate the power of light at the single-photon level [57]. In fact, there are no laboratory methods to make an absolute calibration of power at any intensity. Instead, a chain of photo-detectors are calibrated relative to each other. At the beginning of the chain is an absolute calibration system held at national standards institutes. In the case of our power meter (Coherent FieldMaxII-TO), the National Institute for Standards and Technology (NIST) uses a cryogenic bolometer as its absolute standard. This chain results in a 5% systematic uncertainty in our measurements of the laser power  $P$  (averaged over 0.2 s), measured at the transmitted port of the BS. The magnitude of  $\alpha$  for the probe state was found via  $|\alpha|^2 = \gamma P \lambda / (R h c)$ . For each value of  $\alpha$ , we recorded the number of times each detection outcome occurred in  $J$  trials (i.e. laser pulses), which provides an estimate of  $p_{n,\alpha}$ . The phase of  $\alpha$  was allowed to drift freely, during which no change in the  $p_{n,\alpha}$  was observed. Consequently, we did not actively vary the phase of our probe states.

The 5% uncertainty in  $P$  is the dominant error in our experiment. For a detector with over 95% efficiency, this error could result in estimates of  $\{p_{n,\alpha}\}$  that are incompatible with any physical detector. For example, this could result in more detector clicks on average than there were photons in the probe state on average. Gain in the detector could achieve this, but at the same time would necessarily introduce noise that would change the distribution of  $p_{n,\alpha}$ .

For detectors with lower efficiency, the systematic error in  $P$  will simply add or subtract from efficiency of the detector that results from the tomographic characterization.

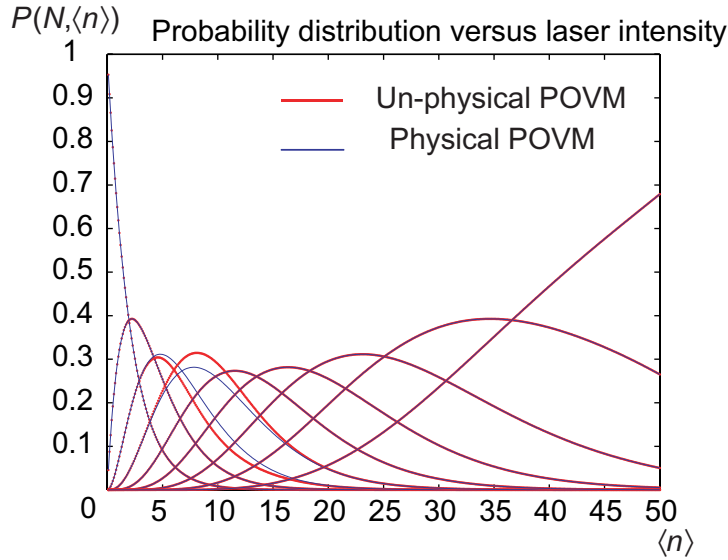
Since  $|\alpha|^2$  has an infinite range, one must set an upper limit on the magnitude of  $\alpha$  used in our set of probe states. A natural place for this is at the  $\alpha$  at which the detector behaviour saturates, i.e.  $p_{n,\alpha}$  stays constant as a function of  $\alpha$ . Since this occurs asymptotically, a somewhat arbitrary degree of constancy must be chosen; we set  $|\alpha|^2 = 120$  as our upper limit.

We expected the saturation behaviour of the TMD would be  $p_{8,\alpha} \approx 100\%$ , whereas we found that it was  $p_{8,\alpha} = 93.3\%$  and  $p_{7,\alpha} = 6.7\%$ . Already, the measured statistics (which are proportional to  $Q_n(\alpha)$ ) give a clear signature that our theoretical model of the detector is too simplistic. Moreover, as we increase the  $\alpha$  beyond our upper limit  $p_{8,\alpha}$  begins to decrease. The cause for this behaviour is a break down of our memoryless detector assumption and highlights how crucial it is to create probe states in the desired spatial-temporal mode. Along with the probe laser pulse, the cavity dumper also outcouples a small fraction of the preceding and following pulses in the Ti:Sapph cavity. These are separated in time from our probe pulse by 13 ns and each contain only 0.17% the energy of the probe pulse. Consequently, these extra pulses will have an insignificant effect on  $p_{n,\alpha}$  for most of the range of  $\alpha$ . However, at  $|\alpha|^2 = 120$ , the preceding pulse will be a coherent state with  $|\alpha|^2 = 0.2$ . Consequently, roughly 20% of the time bins in the TMD will be preceded by a photon. If the detector has, for instance, a 50% efficiency then 10% of the time bins will be preceded by a detection event that will not be counted as click by the electronics (due to their time window). More importantly, those bins will subsequently be unavailable to detect photons in our probe pulse, due to the 50 ns deadtime of the APDs inside the TMD. Thus, as we observe, roughly 10% of the bins will not result in a click. This behaviour was extraneous to the normal operation of the detector and so we limit  $|\alpha|^2$  to 30 in the tomographic reconstruction. However, it exemplifies the usefulness of even the basic detector tomography procedure, which results in approximate  $Q_n(\alpha)$ , for rough evaluation of the detector action. Some of these hypothesis or details could be further explored if we knew well some detectors or some states. For example, the response of BS and NDF to single photons could be explored (granted good single photons and reliable single-photon detectors). The time independence of the POVMs could also be studied with well-known states. However, we will see that the excellent fit of the data to the  $Q$ -function and of the reconstruction to the model suggests a sufficiently good understanding.

### 5.1. Results

We now turn to the tomographic reconstruction. To characterize our detector, we have measured the outcome probability distributions resulting from sending a tomographically complete set of input states (or probe states). The use of pure coherent states as probe states implies that the probability distribution is proportional to the  $Q$ -function, as seen in equation (3). In principle, the knowledge of the  $Q$ -function is then sufficient to predict the measurement probabilities for any incoming state since  $Q(\alpha, \alpha^*) = \langle \alpha | \rho | \alpha \rangle$  determine  $\rho$  completely.

However, a more useful and natural representation for photodetectors is the POVM expanded in the Fock basis. Another argument to find the POVM elements is that due to statistical noise, the  $Q$ -function could correspond to a non-physical POVM. Indeed, if we simply make a fit to the noisy measured probability distribution, this fit could correspond to negative ‘POVMs’. As an example consider figure 9 where two  $Q$ -functions are displayed together: even though they are very similar to each other, one corresponds to a detector with a non-positive



**Figure 9.** This plot presents the probability distributions corresponding to two detectors. One with negative ‘POVM’ elements and another one with positive ones. Both POVMs observe  $\sum \pi_n = \mathbb{1}$ .

‘POVM’ element (and thus negative probabilities) and one corresponds to a physical one. Our goal is therefore to reconstruct the POVM operators which most closely match the data and still observe quantum mechanics (and thus are positive).

Since we adopt a ‘black box’ approach we need not assume any of the properties described in the previous theoretical models. Only the accessible parts of the ‘black box’, i.e. number of outcomes and control (or lack of control) of the phase will determine the description of our detector. The lack of a phase reference simplifies the experimental procedure, allowing us to solely control the magnitude of  $\alpha$  (as has been done for tomography of a single photon [65]). A detector with no observed phase dependence will be described by POVM elements diagonal in the number basis,

$$\pi_n = \sum_{k=0}^{\infty} \theta_k^{(n)} |k\rangle \langle k|,$$

simplifying hence the reconstruction of  $\pi_n$ .

Again we can express the relationship between the statistics and our diagonal  $\pi_n$  as,

$$P = F \Pi,$$

if we measure  $D$  different values of  $\alpha, \alpha_1, \dots, \alpha_D$ , and truncate the number states at a sufficiently large  $M$ . For an  $N$ -outcome detector, the matrices will have dimensions  $P \in \mathbb{C}^{D \times N}$ ,  $F \in \mathbb{C}^{D \times M}$  and  $\Pi \in \mathbb{C}^{M \times N}$ . In addition

$$F_{i,k} = \frac{|\alpha_i|^{2k} \exp(-|\alpha_i|^2)}{k!}$$

can easily be rewritten when the input state is a mixed state. This was done indeed to account for the laser’s technical noise (as we will see in the next section) but gave similar results. For such

a detector, the physical POVM consistent with the data can be estimated through the following optimization problem:

$$\begin{aligned} & \min \{ \|P - F\Pi\|_2 + g(\Pi) \}, \\ & \text{subject to } \pi_n \geq 0, \quad \sum_{n=1}^N \pi_n = \mathbb{1}, \end{aligned} \quad (12)$$

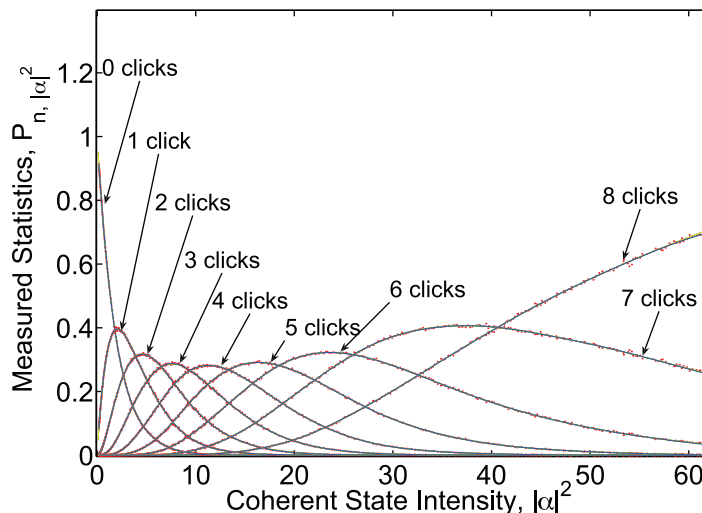
where the two-norm ensures it is a convex quadratic problem. Note that we also allow for convex quadratic filter functions  $g$ , which will be discussed in some detail later. These  $g$ , are related to the conditioning of the problem and must not depend on the type of detector. For example, no symmetry or knowledge of the typical POVM structures in photodetection can be assumed. If any, only general regularization functions that work for any POVM should be chosen. Now, for suitable filter functions (i.e. quadratic) the whole problem is a convex quadratic optimization problem, and hence also a SDP, which can be efficiently solved numerically [18]. Moreover, in this case, there exists a dual optimization problem whose solution coincides with the original problem. Thus, the dual problem provides a certificate of optimality since it provides a lower bound to the primal problem.

Care has to be taken that the optimization problem is well conditioned in order to find the true POVM of the detector. In finding their number basis representation, we are deconvolving a coherent state from our statistics, which is intrinsically badly conditioned due to the importance of the wings of the Gaussian. Similar issues of conditioning have been discussed in the context of state and process tomography, see e.g. [43, 60]. Due to a large ratio between the largest and smallest singular values of the matrices defining the quadratic problem, small fluctuations in the probability distribution can result in large variations for the reconstructed POVM. This can result in operators that approximate really well the outcome statistics and yet do not exhibit a smooth distribution in photon number. We will discuss how to treat this problem in the next section.

The measured probabilities for each outcome as a function of  $|\alpha|^2$  are displayed in figures 10 and 11. The probability distributions in the figures are equivalent modulo  $1/\pi$  to the  $Q$ -function of the detector. Interestingly the curves show very smooth profiles in spite of the statistical noise. It is also plain to see how the nonzero probabilities for each number of *clicks* correspond to different ranges of  $|\alpha|^2$ . Figure 12 shows the POVMs that result from the optimization in equation (12), which we will discuss later. A first remarkable property is that  $\pi_n$ , being the POVM for  $n$  clicks, shows zero amplitude for detecting less than  $n$  photons. That is, the detector shows essentially no dark counts. It should be noted that this was not assumed at the outset and is purely the result of the optimization. This sharp feature gives the detector its discriminatory power where  $n$  clicks means at least  $n$  photons in the input pulse.

To assess the performance of our method we compare it to the model described in the previous section. This time, however, the BS used in the model are not 50/50 but its reflectivities ( $R = [0.5018, 0.5060, 0.4192]$ ) were measured experimentally. This was done measuring the reflected and transmitted beams of a laser with a calibrated power meter. The yellow bars in figure 12, show the absolute value of the difference between the theoretical and the reconstructed POVM elements. The magnitude

$$\Delta_{\theta}^{(n,i)} = |\theta_i^{(n,\text{theo})} - \theta_i^{(n,\text{rec})}|$$



**Figure 10.** The measured probabilities for different ‘number-of-clicks’ are shown (red dots) as a function of  $|\alpha|^2 = \langle n \rangle$ . The plot shows the statistics for the TMD with nine time bins. The statistical error vertically is too small to be seen and the jitter of  $|\alpha|^2$  was estimated to be 2% of its value. An additional 5% systematic error in the calibration of the power meter is present but can be absorbed as loss. From the reconstructed POVM elements  $\{\pi_n\}$ , we generate the corresponding probability distributions  $\text{tr}(\rho_\alpha^{(\text{in})} \pi_n)$  (blue curves). These are generated for pure  $|\alpha\rangle\langle\alpha|$  or mixed  $\rho_\alpha^{(\text{in})}$  and for  $\pi_n$  reconstructed with the filter function or without it. For all these options, the probability distributions (blue lines) are so similar that they are indistinguishable on this scale.

is shown stacked on top of each coefficient of the POVM elements where (theo) stands for theoretical and (rec) for reconstructed. The small yellow bars reveal a good agreement with the model. We also calculate a form of fidelity finding that

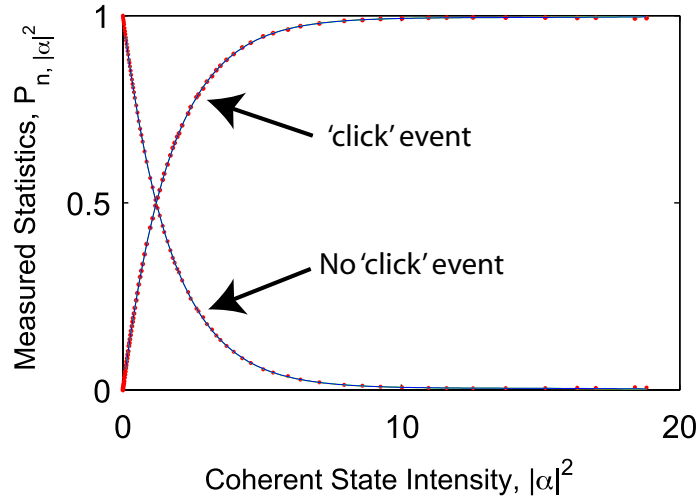
$$F = \text{Tr} \left( \left( (\pi_n^{(\text{theo})})^{1/2} \pi_n^{(\text{rec})} (\pi_n^{(\text{theo})})^{1/2} \right)^{1/2} \right)^2 \geq 98.7\%$$

holds for all  $n$ . Note that to calculate  $F$ , we normalized the POVM elements. This overlap indicates an excellent agreement between the two.

In addition, one can reconstruct a probability distribution: from the found POVMs to fit the data. The reconstruction is plotted as dark blue bars in figures 10 and 11. It is the equivalent of the  $Q$ -function had our probe states  $|\alpha\rangle\langle\alpha|$  with suitable complex  $\alpha$  been strictly pure. In fact, although formally distinct, the probability distribution associated with the reconstructed POVM using mixed or pure states are practically indistinguishable and are plotted together in figure 10 for comparison.

## 5.2. Detector Wigner functions

An alternative representation of the detectors, which can give us more insight about their structure comes from the quasi-probability distributions such as the Wigner function



**Figure 11.** The measured probabilities for the ‘click’ and ‘no-click’ events in the APD are shown as a function of  $|\alpha|^2 = \langle n \rangle$ . The statistical error vertically is too small to be seen and the jitter of  $|\alpha|^2$  was estimated to be 2% of its value. An additional 5% systematic error in the calibration of the power meter is present but can be absorbed as loss. From the reconstructed POVM elements  $\{\pi_n\}$  we generate the corresponding probability distributions  $\text{tr}(\rho_\alpha^{(\text{in})}\pi_n)$  (blue curves). These are generated for pure  $|\alpha\rangle\langle\alpha|$  or mixed  $\rho_\alpha^{(\text{in})}$  and for  $\pi_n$  reconstructed with the filter function or without it. For all these options, the probability distributions (blue lines) are so similar that they are indistinguishable on this scale.

[61, 62]. Since the POVM elements  $\pi_n$  are self-adjoint positive-semi-definite operators, a Wigner function  $W_n$  can be calculated in the standard way from the POVM element  $\pi_n$ :

$$W_n(x, p) = \frac{1}{\pi\hbar} \int_{-\infty}^{\infty} dy \langle x - y | \pi_n | x + y \rangle e^{2ipy/\hbar}, \quad (13)$$

where we have, as usual, now identified  $(x, p) \in \mathbb{R}^2$  as phase space coordinates of a single-mode with  $\alpha \in \mathbb{C}$ . However, since the POVMs do not have unit trace, this detector Wigner function will not be normalized,

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp W_n(x, p) < 1. \quad (14)$$

We should note that the marginals cannot be interpreted as probability distributions but we can still use  $W_n$  to calculate probabilities according to

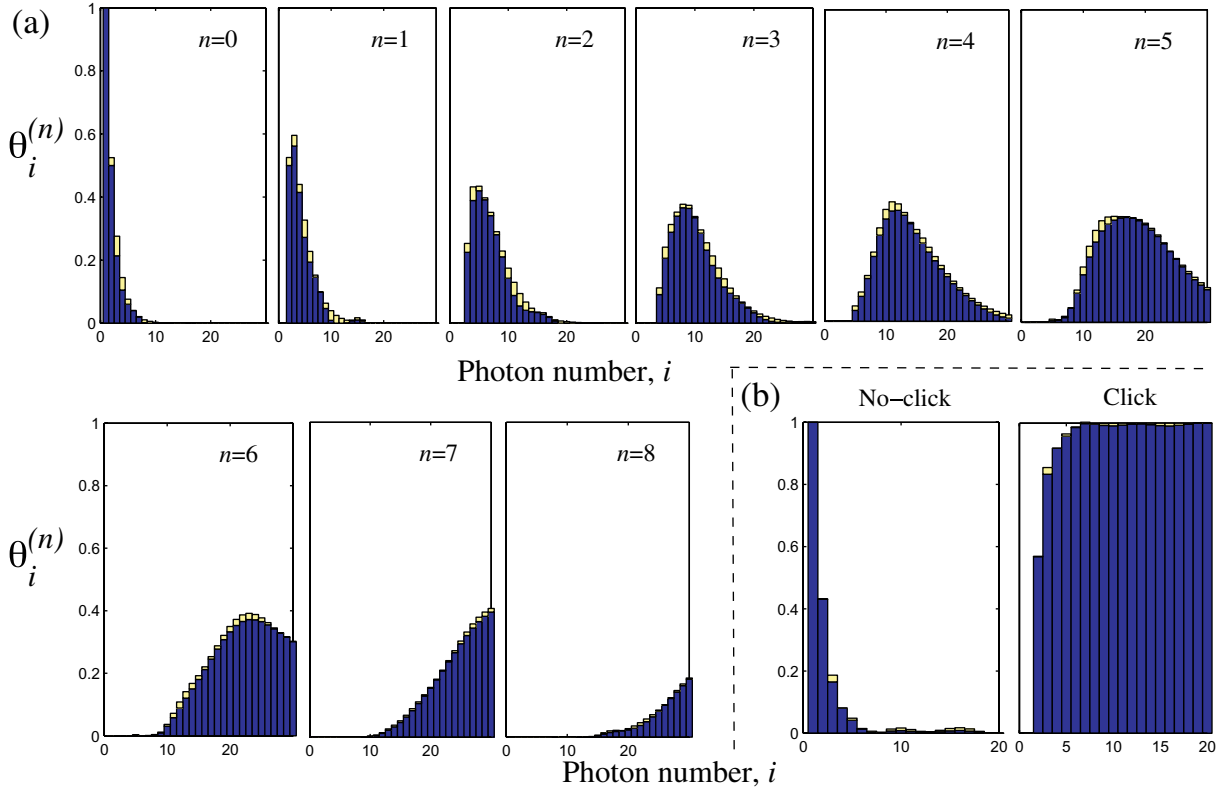
$$p_{\rho,n} = \text{Tr}(\rho\pi_n) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp W_\rho(x, p) W_n(x, p). \quad (15)$$

Since none of the detectors have phase sensitivity their Wigner functions are rotationally symmetric around the origin. In figure 13, we display a cut of the TMD Wigner function for the following POVM elements:

$$\{\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5\}.$$

Higher clicks are not displayed because their amplitude is too small to be compared with the rest. The interesting feature about the plot is the comparison with the theoretical TMD Wigner





**Figure 12.** Reconstructed POVMs for (a) the photon number resolving TMD and (b) the APD ‘yes/no’ detector. TMD POVMs were obtained up to element  $|60\rangle\langle 60|$  (therefore  $M = 60$ ), but are shown up to  $|30\rangle\langle 30|$  for display purposes. APD POVMs are shown in full. Stacked on top of each  $\theta_i^{(n)}$  where  $n$  is the number of clicks we show  $|\theta_i^{(n,\text{rec})} - \theta_i^{(n,\text{theo})}|$  in yellow. ‘rec’ stands for reconstructed and (theo) is the theoretical POVM expected from (a) a TMD modelled with 3 BS of measured reflectivities and 52.1% overall loss (b) a theoretical APD with 43.2% loss, respectively. Note that this result was obtained with a regularized optimization as explained in the next section.

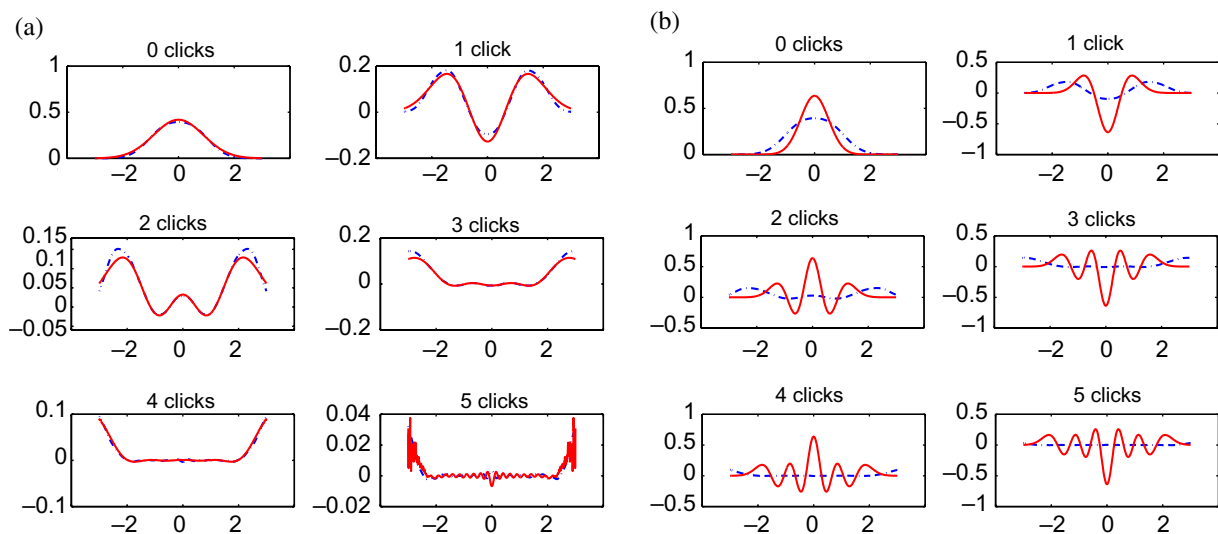
functions one can generate with the model. Indeed, comparing a theoretical loss-less TMD with the measured one we see how the amplitude of the Wigner function decreases rapidly for higher photon-numbers. On the other hand, comparison with the lossy theoretical model reveals a good agreement.

Let us now discuss in more detail the assumptions and details behind this reconstruction. As mentioned earlier in this section (see equation (12)) the inversion of the matrix describing the probe states can be an ill conditioned problem. The next section will show how to regularize such problems. In addition, the matrix describing the probe states should take into account their inherently mixed nature. The next section will also describe the precise role this mixedness plays.

## 6. III conditioning and regularization

One of the main problems encountered in the tomographic characterization of the detectors has to do with the numerical stability of the reconstruction. Such problems are common in





**Figure 13.** Wigner function of the first POVM elements of the TMD. Since the detectors have no phase reference, their Wigner functions are rotationally symmetric with respect to their centre and a cut contains all the information. The dotted blue curve represents the Wigner function of the reconstructed POVMs from 0 to 5 clicks. In red we can see the theoretical Wigner function for: (a) a theoretical TMD with 52% loss. (b) a theoretical TMD without loss. Paying attention to the scale we observe how dramatic the effect of loss is at damping the ripples in the Wigner function. It is also worth noting that the end ripples in (a) for the ‘5 click’ are just an edge effect due to number state cutoff.

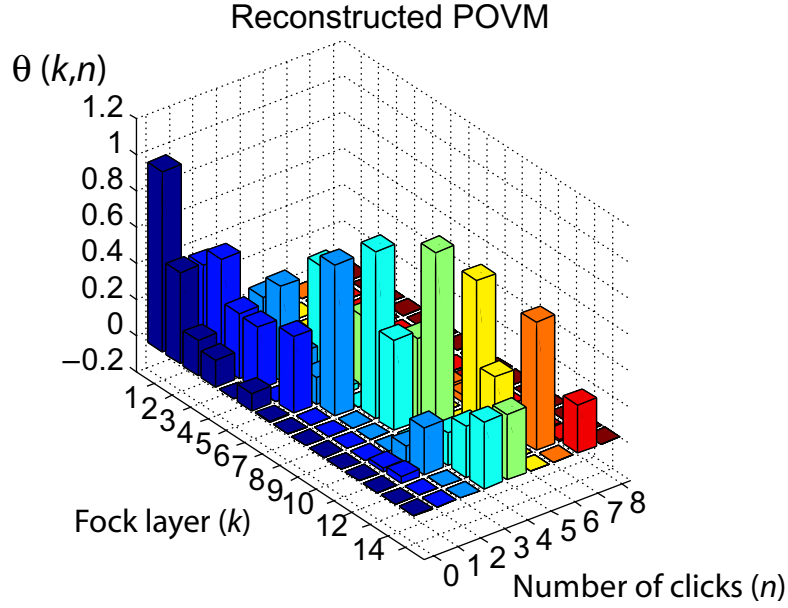
tomography [43, 60]. Consider, for example, the transformations involved in the inverse Radon transform and their inherent instabilities. Note also how going from the  $Q$ -function to the  $P$ -function is not always well defined [61]. Multiple tools exist to bridge the link between homodyne tomography and the density matrix description [66]. One of them involves the use of pattern functions [58, 64, 68]. That is, finding some functions  $G_k(\alpha)$  such that

$$\int Q^{(n)}(\alpha) G_k(\alpha) d^2\alpha = \theta_k^{(n)}.$$

However, finding the appropriate  $G_k$  involves the irregular wave functions (particular unbounded solutions of the Schrödinger equation) and proving them to be appropriate is typically as hard as estimating the error [63]. The use of maximum likelihood has also been explored and particularly for detector tomography [14, 15]. However, the speed of the convergence is not generally guaranteed to be high, becoming exponential for certain problems. Our approach, following the spirit of maximum-likelihood, translates the problem into a quadratic optimization one allowing for efficient semi-definite programming (SDP) (cf equation (12)). We discuss here the details, approximations and filters that lead to our solution of the problem.

### 6.1. Truncating the Hilbert space

The data were measured up to  $|\alpha|^2 = 150$  but was truncated at lower values in phase space. This was done to avoid noisy behaviour and the emergence of new regimes in the behaviour of



**Figure 14.** POVM reconstruction, using only minimization from equation (17). Dark blue:  $\pi_0 = \sum_{i=0}^{15} \theta_i^{(0)} |i\rangle\langle i|$ , lighter blue,  $\pi_1 = \sum_{i=0}^{15} \theta_i^{(1)} |i\rangle\langle i|$ , etc.

the detector. Memory effects requiring a larger POVM space were thus avoided as discussed in the experimental section. Notably, the effects related to the detector's deadtime, after pulsing or the dark counts from possible over heating were avoided staying in a low ( $|\alpha|^2 \simeq 70$ ) photon-number regime.

## 6.2. Pure versus mixed

The  $Q$ -function of our detector (directly measured) is proportional to

$$p_{\alpha,n} = \text{Tr}(|\alpha\rangle\langle\alpha|\pi_n). \quad (16)$$

From equations (5) and (6), for a diagonal POVM, we can write the problem as

$$\min \|P - F\Pi\|_2. \quad (17)$$

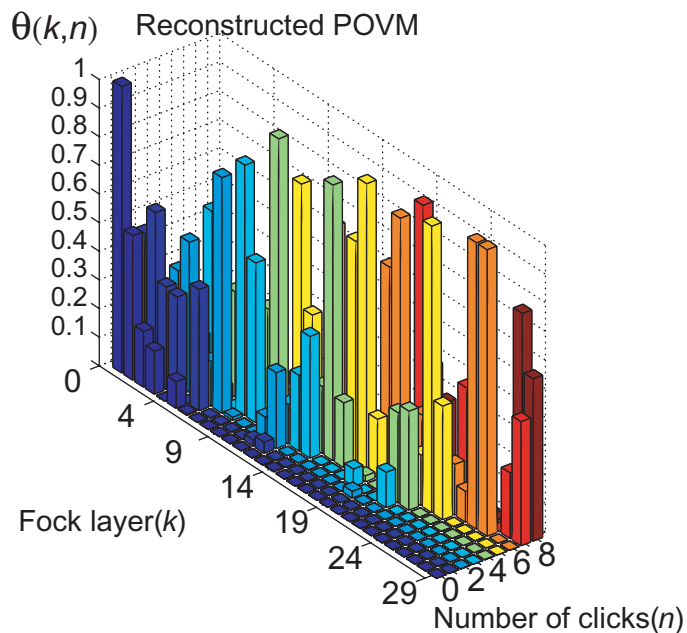
with the usual constraints  $\pi_n \geq 0$  and  $\sum_n \pi_n = \mathbb{1}$ . Using a semi-definite solver such as Yalmip, the obtained POVMs  $\{\pi_n\}$  shows irregular dips and a structure quite dissimilar from what a TMD is expected to do. The figure 14 shows a typical result, and figure 15 shows it for higher photon numbers revealing an even more irregular structure.

**6.2.1. Describing the laser's amplitude uncertainty.** A first meaningful observation is that some level of uncertainty existed in the amplitude  $x = |\alpha|^2$  of the coherent states. If  $D$  values of  $x$  were measured then the real

$$\bar{x} = (x_1, x_2, \dots, x_D)$$

might actually have been

$$\bar{x}_\delta = (x_1(1 + \delta_1), x_2(1 + \delta_2), \dots, x_D(1 + \delta_D)),$$



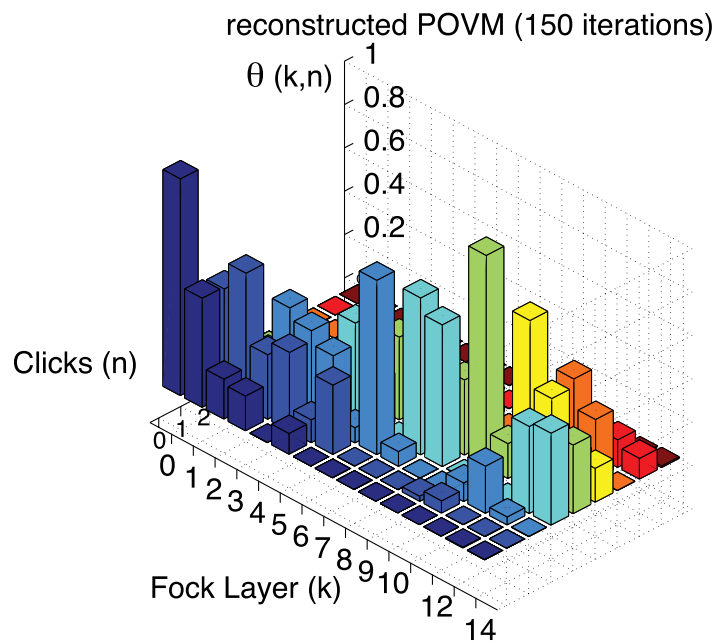
**Figure 15.** POVM reconstruction, using only minimization from equation (17). Dark blue:  $\pi_0 = \sum_{i=0}^{15} \theta_i^{(0)} |i\rangle\langle i|$ , lighter blue,  $\pi_1 = \sum_{i=0}^{15} \theta_i^{(1)} |i\rangle\langle i|$ , etc displayed up to number layer 30.

with some vector of errors  $(\delta_1, \dots, \delta_D)$ . To address the effect of this uncertainty on our minimization we can artificially introduce noise and then average over many runs of the optimization. In other words, since  $F$  in equation (17) depends on the measured values of  $|\alpha|^2$ , we can substitute  $\bar{x}$  with  $\bar{x}_\delta$  where  $\delta = (\delta_1, \dots, \delta_D)$  are independent and identically distributed random variables. Using  $\bar{x}_\delta$  we run the optimization and obtain a family of estimated POVM elements (each element of the family corresponds to a run of the optimization with a different  $\delta$ ) As a first approximation we may use a Gaussian probability distribution with zero mean and  $\sigma = 2\%|\alpha|^2$ . Note that 2% was the measured variance of the laser amplitude from pulse to pulse as shown in figure 18. Subsequently, we average over the POVMs obtained with different ‘jitters’  $\delta$  in  $N$  realizations, obtaining

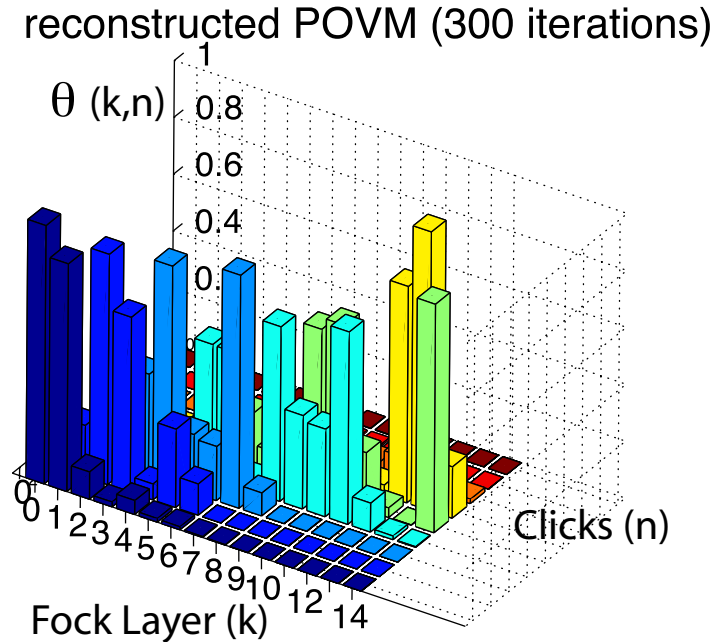
$$\pi_n^{(\text{average})} = \sum_j \pi_n^{(\delta_j)} / N.$$

Two hundred iterations of the optimization with subsequent averaging improves the appearance of the POVMs but barely solves the ‘dips’ observed. figures 16 and 17 are the examples of this approach, showing that this kind of averaging does not properly counteract the fluctuations in the reconstructed POVM.

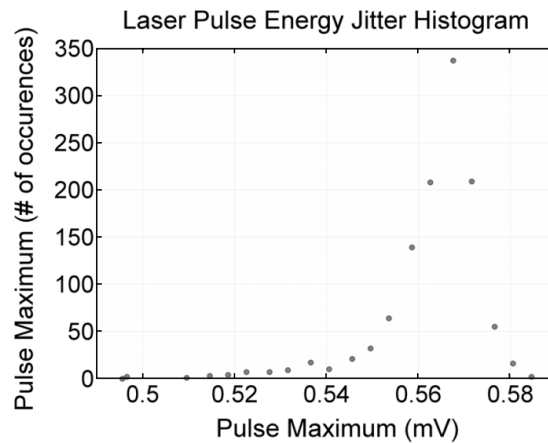
**6.2.2. Using mixed input states.** A key observation showing that the previous approach is not the appropriate treatment of uncertainty in  $x$  is that each probe state would be best described by



**Figure 16.** POVM reconstruction, using direct averaging (150 runs with 1% Gaussian noise).



**Figure 17.** POVM reconstruction, using direct averaging (300 runs with 2% Gaussian noise).



**Figure 18.** Measurement of the laser's amplitude variations from pulse to pulse.

a mixture of coherent states,

$$\rho_x = \int d^2\beta |\beta\rangle\langle\beta| f_x(\beta), \quad (18)$$

$$= \sum_{n,m=0}^{\infty} R_{n,m,x} |n\rangle\langle m|. \quad (19)$$

Here,  $f_x$  would be some distribution centred around  $x$  in phase space, leading to a mixed Gaussian state in case of a Gaussian classical probability distribution. We can integrate this state  $\rho_x$  over the complex phase since we have no phase reference available and focus solely on the amplitude of the coherent states or mixtures thereof. Measurements reveal that the intensity of the laser varies from pulse to pulse following a distribution that looks like a Lorentzian with a tail (see figure 18). A good approximation can however be made using a Gaussian distribution, leading to a Gaussian state, with standard deviation  $\sigma = 0.02|\alpha|^2$ , implying

$$R_{n,m,\alpha} = \frac{1}{\sigma\sqrt{2\pi}\sqrt{n!m!}} \int \beta^{n+m} e^{-\beta^2} f_x(\beta) d\beta.$$

with  $f_x(\beta) = e^{-(\beta^2-x)^2/(2\sigma^2)}$ . The probability to detect  $n$  clicks becomes then

$$p_n(\alpha) = \sum_{k=0}^{\infty} R_{k,k,\alpha} \theta_k^{(n)}. \quad (20)$$

To simplify these calculations we can write a distribution in  $\sqrt{x} = |\alpha|$ ,

$$\rho_{|\alpha|} = \int d^2\beta |\beta\rangle\langle\beta| g_{|\alpha|}(\beta) \quad (21)$$

with

$$g_{\alpha}(\beta) = e^{-(\beta-\alpha)^2/(2\Gamma^2)}.$$

In this case,  $\Gamma$  has been chosen such that the approximation  $f_x(\beta) \simeq g_{\alpha}(\beta)$  holds. These subtleties however do barely alter our results and POVMs are as irregular as previously.

To evaluate the difference introduced by the pure ( $|\alpha\rangle\langle\alpha|$ ) or mixed state ( $\rho_{|\alpha|}$ ) approach, we have studied their influence on the reconstructed POVMs. In the regularized optimization (i.e. for our final results), we have compared the POVMs obtained with each description finding that

$$\frac{\|\Pi_{\text{pure}} - \Pi_{\text{mixed}}\|_2}{\|\Pi_{\text{mixed}}\|_2} \leq 0.7\%. \quad (22)$$

In addition, we can look at each matrix element of the POVMs. If  $\theta_k^{(n)}$  are the matrix element reconstructed using a pure state, and  $\beta_k^{(n)}$  are the matrix elements reconstructed using a mixed state then the difference between each element will be small. In fact, we found that the relative difference  $|\theta_k^{(n)} - \beta_k^{(n)}|/|\theta_k^{(n)}|$  between a mixed and pure state derivation was always smaller than 1.3%. Also, from the reconstructed POVMs one can generate the probability distributions corresponding to a family of input states. In particular, when we generated the probability distributions for a mixed or a pure state (using respectively  $\beta_k^{(n)}$  and  $\theta_k^{(n)}$ ) they were indistinguishable from the data if plotted in figure 10. As expected, the effect of the laser's technical noise was negligible when using single-photon level coherent states. This of course contrasts with traditional homodyne tomography where the fluctuations in the coherent states can shift the strong local oscillator to a nearly orthogonal state.

However, since the problem of the irregular POVMs is not solved by the mixed state description we need to look further into the origin of these irregularities. One first remarkable (but expected) property is that large variations in the photon number degree of freedom of the POVMs result in minuscule differences in the probability distributions (see figure 12). Since one convolutes the photon number distribution with a Gaussian in  $\alpha$  to obtain the  $Q$ -function this behaviour has been expected. Conversely this means that small errors or statistical fluctuations in the  $Q$ -function can result in large errors in the POVM elements. Consider for example that if instead of

$$\min \|P - F\Pi\|_2,$$

we try to minimize

$$\min \|F^{-1}P - \Pi\|_2$$

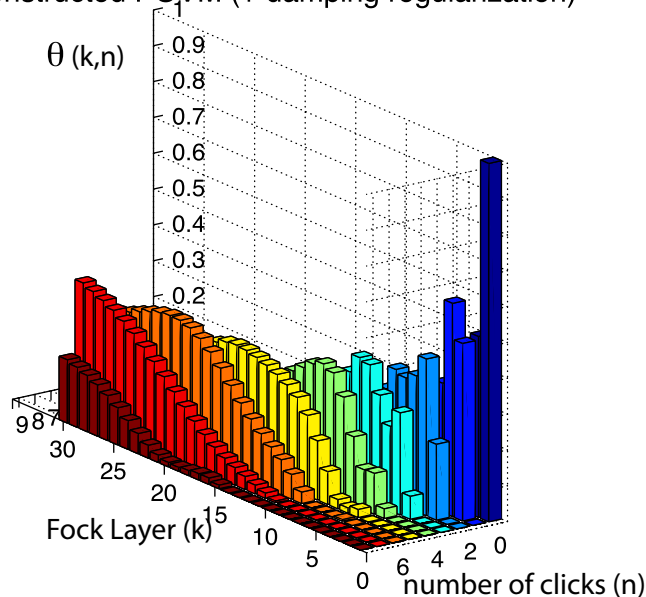
the SDP solver finds no sensible solution. This is because using the Moore–Penrose pseudo-inverse, we find  $F^{-1}F \neq \mathbb{1}$  due to its inherent ill conditioning, meaning that the ratio of largest and smallest singular values in  $F$  is large.

Various methods exist to try and regularize these problems. Whatever the chosen method it should assume as little knowledge as possible about the specific form of the sought POVM. Since  $F$  has very small values for high photon numbers one could enhance those values while preserving the minimization target. For example, we could run the optimization

$$\min \|PD - F\Pi D\|_2,$$

where  $D$  is a diagonal matrix aimed at regularizing the problem. This can be shown to introduce some improvement but does not solve the ‘dip problem’ completely. It is also hard to find the exact form of  $D$  that yields ‘good’ results without any prior knowledge about the expected POVMs. In addition (roughly speaking), it is hard to find a balance between having good results for low photon numbers and for high photon numbers.

reconstructed POVM (+ damping regularization)



**Figure 19.** Minimization using damping method on equation (23). Note that the point of view is opposite that of the previous plots. We see some dips around the 5th and 7th number layers.

Another approach is to introduce a sort of damping or specific penalization. For example, one could define a diagonal matrix  $M$  such that

$$M_{i,j} = \delta_{i,j}/j,$$

and use it to redistribute the weight of each POVM element, avoiding unreasonably large POVM element amplitudes (that compensate for low values in  $F$ ). The optimization could be recast as,

$$\min \{ \|P - F\Pi\|_2 + 0.03\|M\Pi\|_2 \}. \quad (23)$$

A result of this can be seen in figure 19. This method has the same shortcomings as the previous one: it is sensitive to the choice of parameters and the exact form of  $M$  is hard to determine without detailed prior assumptions.

A more reasonable method is to capture the relative smoothness of the POVM from a lossy detector. This method is also called smoothing regularization [18]. In this case, one single assumption needs to be made. The POVMs should exhibit a certain degree of ‘smoothness’.

### 6.3. Smooth or not?

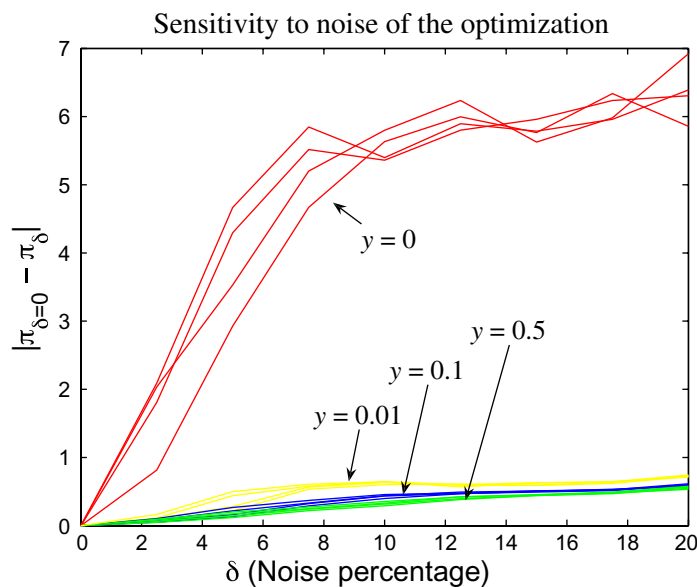
Let us first define what we mean by smooth. Smooth will mean in this context that the difference  $|\theta_k^{(n)} - \theta_{k+1}^{(n)}|$  is small for all  $k$  and  $n$ . In the optimization context, we will mean that our minimization is defined as follows:

$$\min \{ \|P - F\Pi\|_2 + yS \}, \quad (24)$$

with

$$S = \sum_{k,n} \left( \theta_k^{(n)} - \theta_{k+1}^{(n)} \right)^2$$





**Figure 20.** Illustration of the sensitivity to noise for two different minimization methods ( $y = 0$  corresponds to no regularization, and  $y \neq 0$  to an approach using a smoothing regularization). For each value of  $y$  and  $\delta$  we have run the optimization four times and displayed the results here to illustrate this variation.

for some fixed value of  $y$ . The smoothing function  $S$  will be independent of the detector, and will mildly penalize non-smooth POVM elements. This approach is further substantiated by the observation that the resulting POVMs are largely independent of the weight  $y$  that is given to the smoothness penalty.

As most quantum detectors, especially those discussed here, are lossy this is a particularly plausible feature. Indeed, if an optical detector has a POVM element with nonzero amplitude in  $|n\rangle\langle n|$ , then if it is lossy, it will have a positive amplitude in  $|n+1\rangle\langle n+1|$ ,  $|n+2\rangle\langle n+2|$ ,  $\dots$ ,  $|n+K\rangle\langle n+K|$ , decreasing with  $K$  but different from zero. In fact, in general, if the detector has a finite efficiency  $\eta$  which can be modelled with a BS, it will impose some smoothness on the distribution  $\theta_k^{(n)}$ . That is because if  $G(k)$  is the probability of registering  $k$  photons and  $H(k')$  is the probability that  $k'$  were present, then the loss process will impose [67]:

$$G(k) = \sum_{k'} \binom{k'}{k} \eta^k (1-\eta)^{k'-k} H(k').$$

Consequently, if  $\theta_k \neq 0$ , then  $\theta_{k+1}$ ,  $\theta_{k+2}$ , etc cannot be zero, but will have some relatively smooth distribution. This simple physical argument makes a certain smoothness plausible (but still should allow sharp transitions for  $m < n$ ).

For this detector (and for any photodiode based detector) assuming loss is reasonable and can make the ‘smoothness’ requirement appropriate. Let us however see if, without looking at the specific shape of our POVM, we can find an optimal smoothing coefficient  $y$  and justify further the use of the smoothing regularization.

**Table 1.** This table illustrates how sensitive the optimization is to the particular choice of  $y$  (the reference smoothing strength is  $y = 0.1$ ).

$y$	$y$ variation	$\Pi$ relative error (%)
0.0001	$x/1000$	27.3
0.001	$x/100$	12.2
0.01	$x/10$	4
0.05	$x/2$	1
0.5	$x \cdot 5$	3
1	$x \cdot 10$	5

One way to test this method is to quantify how resilient it is to noise in the data. To do so we introduce additional noise in  $x = |\alpha|^2$  to the measured data. For example, we can alter  $x$  in  $P_{i,n} = P(x_i(1 + \delta_i), n)$  where  $\delta = (\delta_1, \dots, \delta_D)$  is again a vector of random variables distributed with a Gaussian distribution with zero mean. This simulates a statistical uncertainty in the measurement of the coherent state. To see its effect on the reconstruction we use the figure of merit  $\|\Pi_\delta - \Pi_{\delta=0}\|_2$ . This quantity should evaluate how POVMs differ from the one without noise. It is seen that the additional smoothing penalty makes the optimization more robust, largely independent of the value of  $y$  (we can multiply  $y$  by a 100 and stay in the same regime). Using this smoothing regularization with noisy data seems therefore a good choice as seen in figure 20.

We have seen how smoothing makes the optimization more robust against noise but we should also ask how sensitive this optimization is to the exact choice of  $y$ . To do so we may use the following procedure: compare the POVM obtained using  $y = 0.1$  with that obtained varying  $y$  over four orders of magnitude. In figure 21, we plot the relative error  $100 * |\Pi_y - \Pi_{y=0.1}| / |\Pi_{y=0.1}|$ . Remarkably doubling the value of  $y$  results in an overall relative error in the POVM of less than 1%. Multiplying (or dividing)  $y$  by 10 gives a variation below 5% and 100fold variation results in a 12% variation. If we compare how this differs from the  $y = 0$  case, which is 110% different then we can conclude that the optimization is quite insensitive to the exact choice of the *smoothing parameter*  $y$ . Table 1 provides some values for reference.

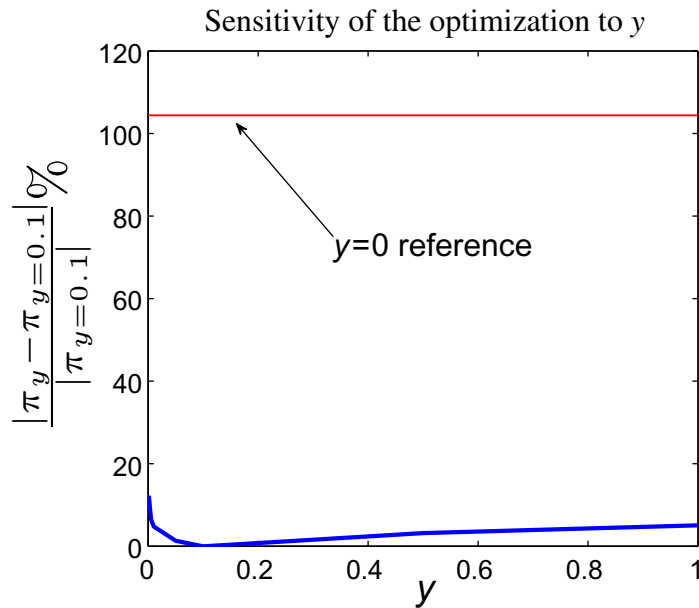
#### 6.4. Sharp and smooth

There is of course a limit to how much we can penalize non-smooth POVMs. Is it possible for the smoothing regularization to wash out all the sharp features of the POVM, thus smoothing in excess? This of course is a legitimate question that further restricts the reasonable range for  $y$ . To study that effect we analyse four cases:

1. A theoretical loss-less TMD, based on the model described in equation (11).
2. A lossy TMD, based on the above with added loss from an  $R = 52\%$  BS.
3. A perfect photon number detector, that is with  $\pi_n = |n\rangle\langle n|$ .
4. An artificial POVM with sharp variations, containing the POVM elements:

$$\pi_0 = |0\rangle\langle 0| + |2\rangle\langle 2|,$$

$$\pi_1 = |1\rangle\langle 1| + 1/2|3\rangle\langle 3|,$$



**Figure 21.** Illustration of how sensitive the optimization is to the specific choice of  $y$ . This plot shows the relative error with respect to the POVM elements obtained using  $y = 0.1$ , as a function of  $y$ . In red, and only for reference (since it does not change with  $y$ ), the value of the relative error for  $y = 0$  (no smoothing) is shown. We vary  $y$  in the range  $[0.001, 1]$ . It is remarkable that a 10 000% variation in  $y$  results in only a 12% variation. For  $y \in [0.05, 0.2]$  the relative error is less than 2% in  $\Pi$ .

$$\pi_2 = 1/2|3\rangle\langle 3| + |4\rangle\langle 4| + |5\rangle\langle 5|,$$

$$\pi_3 = |7\rangle\langle 7|,$$

$$\pi_4 = 1/4|6\rangle\langle 6| + 1/4|8\rangle\langle 8|,$$

$$\pi_5 = 1/4|6\rangle\langle 6| + 1/4|8\rangle\langle 8|,$$

$$\pi_6 = 1/2|6\rangle\langle 6|,$$

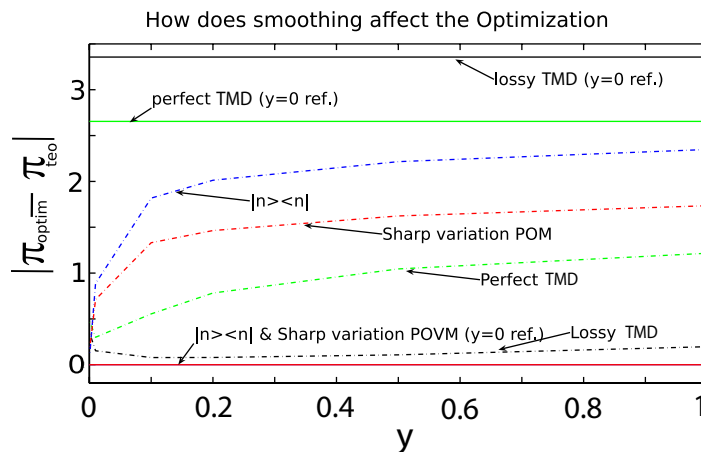
$$\pi_7 = 1/2|7\rangle\langle 7|,$$

$$\pi_8 = 1/2|8\rangle\langle 8| + \sum_{k=9}^{60} |k\rangle\langle k|$$

and observing

$$\sum_i \pi_i = \mathbb{1}.$$

To study the smoothing we generate the POVM elements  $\{\pi_n\}$  numerically, build a probability distribution  $\text{Tr}(\rho_\alpha \pi_n)$  and retrieve the  $\pi_n$  using the optimization from equation (24) for an increasing range of  $y$ . Then we compare these results with the theoretical POVMs we defined in order to generate the PD. All optimizations are done using the mixed-state approach

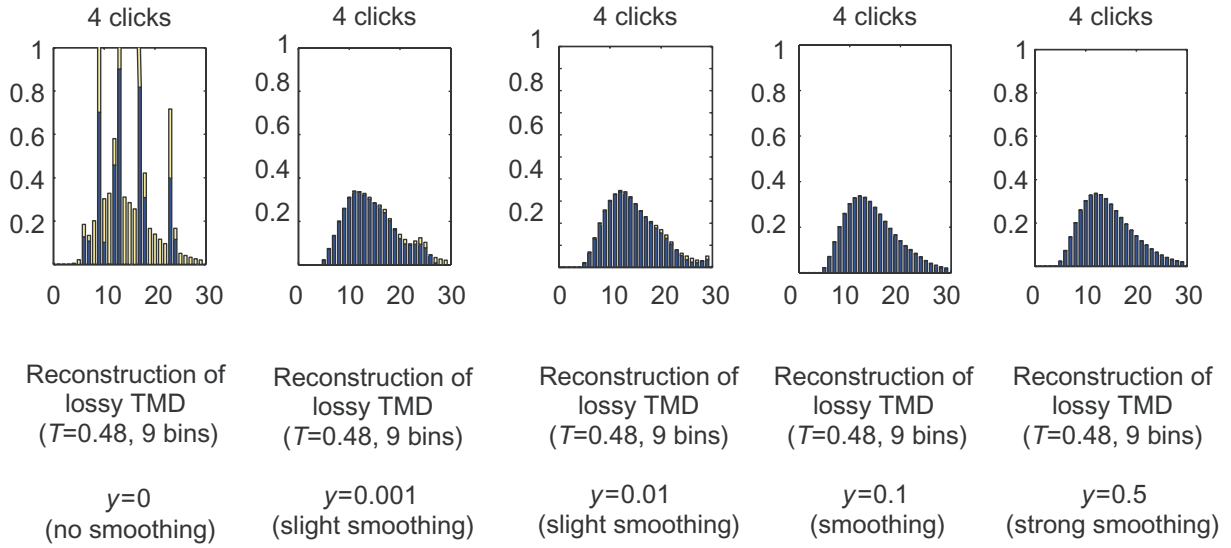


**Figure 22.** Illustration of how too much smoothing can fail to capture the sharp variations of a POVM. We define  $\Pi_{\text{theo}}$  as the matrix containing the POVM elements of the theoretical POVMs. From them we generate a probability distribution and reconstruct the POVMs  $\Pi_{\text{optim}}$  with the smoothing regularized optimization. The dotted lines represent  $\|\Pi_{\text{optim}} - \Pi_{\text{theo}}\|_2$  for different values of  $y$  and for a variety of POVMs (see following plots). The horizontal lines represent that same difference for  $y = 0$  and are plotted for reference.

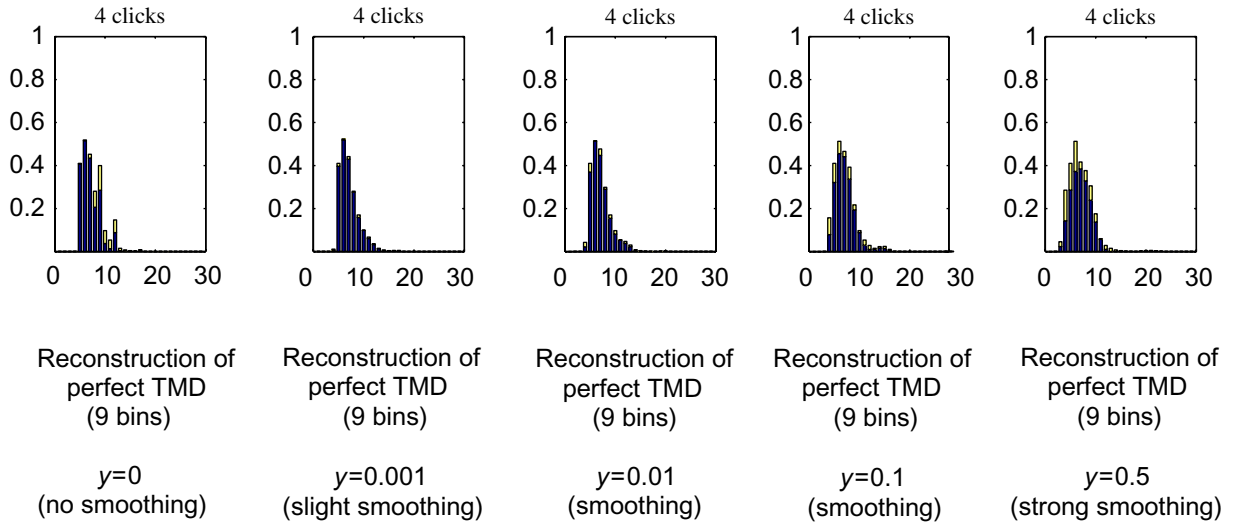
from equation (21). Broadly speaking, we find two distinct behaviours: POVMs with terms that decay slowly in photon number need regularization and are quite insensitive to the precise  $y$ . For sharp POVMs (without loss) the range  $0 < y < 0.01$  preserves their shape quite well, but further smoothing hides their true shape. These properties can be seen in figure 22 and are further illustrated in the figures that follow.

**6.4.1. Lossy TMD.** Figure 23 presents the evolution of the ‘4 click’ POVM element as we add more smoothing (or increase  $y$  in equation (24)). This element is chosen as an illustrative example but more details can be found in [59]. The figure shows in blue the coefficients  $\theta_i^{(4)}$  in  $\pi_4^{(\text{rec})} = \sum_{i=0}^{60} \theta_i^{(4)} |i\rangle\langle i|$ , where rec means reconstructed. In yellow, stacked on top of  $\theta_i^{(4)}$  we display  $|\theta_i^{(4,\text{rec})} - \theta_i^{(4,\text{theo})}|$ , where theo refers to the original POVM we used to generate the probability distribution. Clearly the smoothing improves the result and the exact value of  $y$  is rather unimportant. A sharp feature that is preserved however is  $\theta_i^{(4)} = 0$  for  $i < 4$  proving a good agreement with the model.

**6.4.2. Loss-less TMD.** Figure 24 shows also the ‘4 click’ event and the error associated with the reconstruction (yellow). This TMD shows in its distribution the finite number of bins as we described earlier. The distribution is not as broad as that of the lossy TMD and the smoothing is therefore not so effective. The raw SDP, with  $y = 0$ , performs quite well, and the POVM is quite insensitive to the smoothing, although, when given 1/2 of the weight in the optimization ( $y = 0.05$ ) the smoothing starts to become harmful.

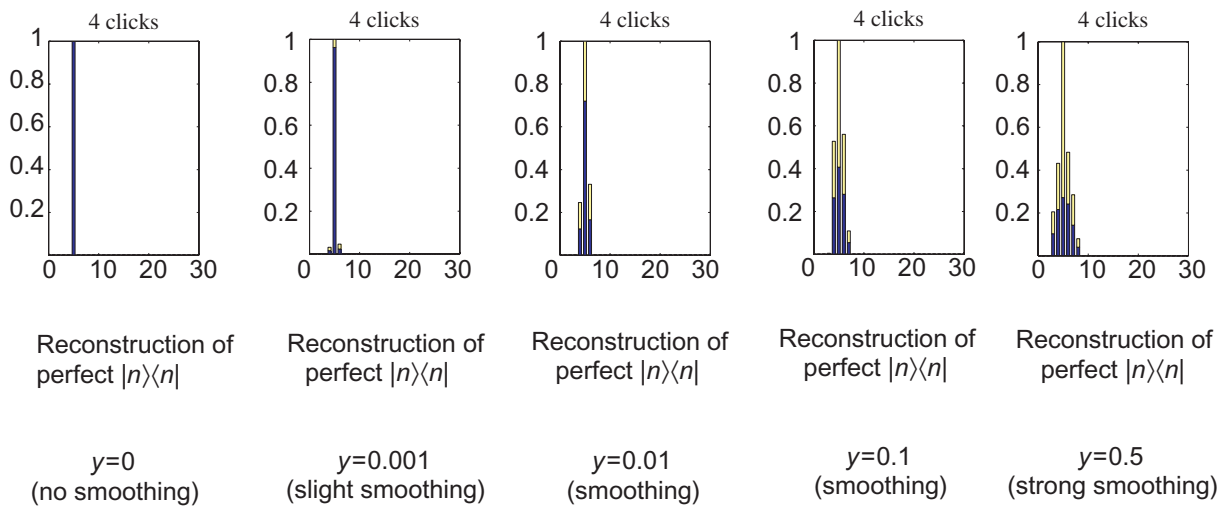


**Figure 23.** Smoothing evolution for a lossy TMD detector (loss = 52%). We show as an example the evolution of the POVM element  $\pi_4 = \sum_{i=0}^{60} \theta_i^{(4, \text{rec})} |i\rangle\langle i|$  as we increase the amount of smoothing (in  $y$ ). The yellow bars display  $|\theta_i^{(4, \text{theo})} - \theta_i^{(4, \text{rec})}|$  stacked on top of  $\theta_i^{(4, \text{rec})}$ .



**Figure 24.** Smoothing evolution for a perfect TMD detector (no loss). We show as an example the evolution of the POVM element  $\pi_4 = \sum_{i=0}^{60} \theta_i^{(4, \text{rec})} |i\rangle\langle i|$  as we increase the amount of smoothing (in  $y$ ). The yellow bars display  $|\theta_i^{(4, \text{theo})} - \theta_i^{(4, \text{rec})}|$  stacked on top of  $\theta_i^{(4, \text{rec})}$ .

**6.4.3. Perfect number detector.** Figure 25 shows also the ‘4 click’ event which in this case is simply  $\pi_4 = |4\rangle\langle 4|$ . A very interesting feature is that the simple SDP with  $y = 0$  achieves a perfect result. This happens in spite of using a mixed state as a probe state (mixture of amplitudes  $|\alpha\rangle$  around  $|\alpha\rangle\langle\alpha|$ ). The reconstruction is then robust for very well defined and sharp features, where the higher decaying coefficients do not introduce instabilities.



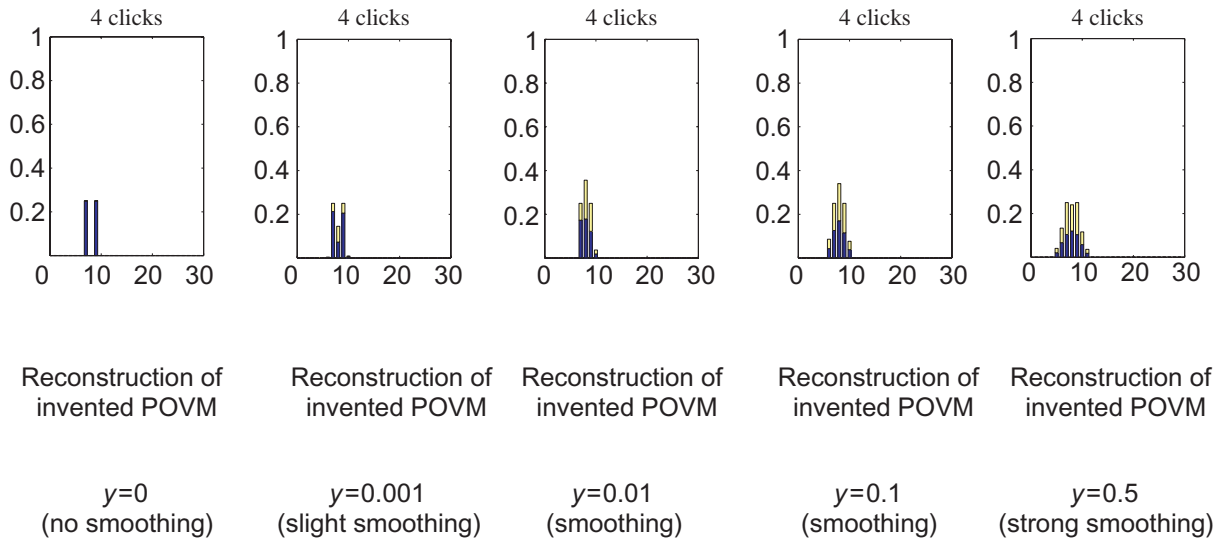
**Figure 25.** Smoothing evolution for a perfect photon-number detector, that is one with  $\pi_n = |n\rangle\langle n|$ . We show as an example the evolution of the POVM element  $\pi_4 = \sum_{i=0}^{60} \theta_i^{(4,\text{rec})} |i\rangle\langle i|$  as we increase the amount of smoothing (in  $y$ ). The yellow bars display  $|\theta_i^{(4,\text{theo})} - \theta_i^{(4,\text{rec})}|$  stacked on top of  $\theta_i^{(4,\text{rec})}$ .

**6.4.4. Sharp POVM.** We now discuss the situation of a POVM element that is not related to an experiment, but has been artificially generated to identify the limit of the smoothing regularization. The element displayed in figure 26 is  $\pi_4 = |7\rangle\langle 7| + |9\rangle\langle 9|$  and we can see that  $y = 0.1$  is already too much smoothing. Certainly to reconstruct a completely loss-less detector with such a structure smoothing is not an appropriate strategy. We must remember however that all current photon-number detectors that count particles do exhibit loss, and have therefore some degree of smoothness in them.

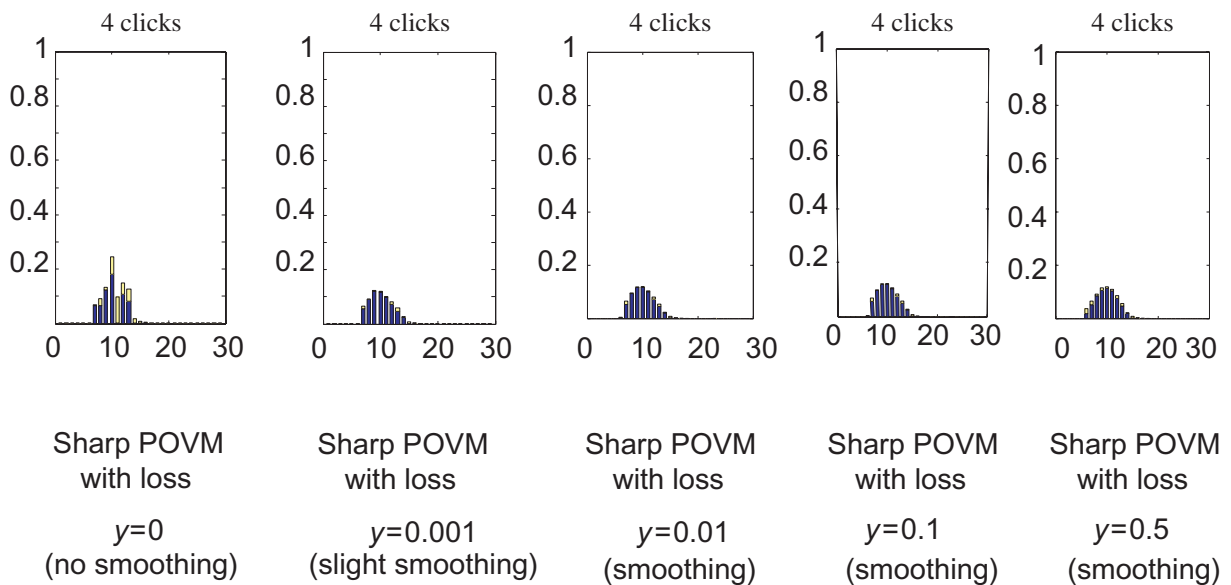
**6.4.5. Sharp POVM with loss.** The previous case could have given the impression that the reconstruction fails for a sharp POVM. However, one has to stress that smoothing (or a regularization for the optimization) is necessary when there is loss and an ill-conditioned matrix (which is a generic case in quantum optics using coherent states for detector tomography). Therefore, it is worth considering what happens when an invented POVM (as the previous one) is made more realistic adding some loss. Figure 27 illustrates this, showing the reconstruction of the previous sharp POVM element, which has suffered a 20% loss. Indeed in this case we can see a clear improvement as the smoothing helps regularize the optimization.

## 7. Conclusion

As quantum information and computation implementations evolve, detectors are becoming more complex. In addition, crypto security requires a careful statement of assumptions ideally kept to a minimum. This, as we have seen, calls for a black-box characterization of the operators they implement. We have seen the first implementation of this type of tomography. We discussed in detail the first experimental realization of QDT completing the triad of experimental state [3, 8, 9], process [4, 5, 10, 11] and detector tomography [17]. This detector characterization



**Figure 26.** Smoothing evolution for an invented POVM with sharp variations. Displayed is  $\pi_4 = |7\rangle\langle 7| + |9\rangle\langle 9|$ . We show as an example the evolution of the POVM element  $\pi_4 = \sum_{i=0}^{60} \theta_i^{(4,\text{rec})} |i\rangle\langle i|$  as we increase the amount of smoothing (in  $y$ ). The yellow bars display  $|\theta_i^{(4,\text{theo})} - \theta_i^{(4,\text{rec})}|$  stacked on top of  $\theta_i^{(4,\text{rec})}$ .



**Figure 27.** Smoothing evolution for an invented POVM with sharp variations, which has suffered a 20% loss (modelled by interposing a BS). Displayed is the reconstruction of the element  $\pi_4 = |7\rangle\langle 7| + |9\rangle\langle 9|$  when it suffers the mentioned loss. The reconstruction shows  $\pi_4 = \sum_{i=0}^{60} \theta_i^{(4,\text{rec})} |i\rangle\langle i|$  as we increase the amount of smoothing (in  $y$ ). The yellow bars display  $|\theta_i^{(4,\text{theo})} - \theta_i^{(4,\text{rec})}|$  stacked on top of  $\theta_i^{(4,\text{rec})}$ .



opens up more flexible and complex ways of detecting quantum states and accurately preparing non-classical light.

The reconstruction methods are simple and efficient. However, one has to pay close attention to the subtleties behind the ill-conditioning of such reconstructions whether its state or detector tomography. Fully characterizing a detector with this method can help get rid of complex or erroneous assumptions in the modelling. Furthermore, once they are fully characterized, one can redesign or alter the detectors with a direct feedback on their performance.

Detector tomography significantly benefits state tomography or metrology, as well as state preparation and the implementation of protocols in quantum information requiring detectors in state manipulation. Importantly, it enables the use of detectors that are noisy, nonlinear or that operate outside their intended range. One conclusion is that lossy detectors are often just as useful as perfect ones, as long as we know the exact POVMs and one can describe the rest of our experiment accordingly. This method will also allow the benchmarking of similar detectors making performance comparisons possible. Indeed one can also ask question concerning the power each detector has for preparing non-classical states. This opens a path for the experimental study of novel concepts such as the non-classicality of detectors. Another promising avenue is to translate homodyne tomography techniques to optical detector tomography. For example defining the detector tomography equivalents of balanced noise-reduction, direct measurement of the Wigner function or pattern functions. Naturally an immediate next step would involve characterizing detectors with off diagonal terms and phase sensitivity.

## Acknowledgments

This work was supported by the EU (COMPAS, HIP, MINOS, Marie Curie and QAP), the EPSRC, the Heinz-Durr Stipendienprogram of the SDV, the QIP-IRC, Microsoft Research, the EURYI Award Scheme and the Royal Society.

## References

- [1] Higgins B L, Berry D W, Bartlett S D, Wiseman H M and Pryde G J 2007 *Nature* **450** 393
- [2] Pagnell K L, Resch K J, Prevedel R, Gilchrist A, Pryde G J, O'Brien J L and White A G 2007 *Phys. Rev. Lett.* **98** 223601
- [3] Smithey D T, Beck M, Raymer M G and Faridani A 1993 *Phys. Rev. Lett.* **70** 1244
- [4] Poyatos J F, Cirac J I and Zoller P 1997 *Phys. Rev. Lett.* **78** 390
- [5] Altepeter J B, Branning D, Jeffrey E, Wei T C, Kwiat P G, Thew R T, O'Brien J L, Nielsen M A and White A G 2003 *Phys. Rev. Lett.* **90** 193601
- [6] Raussendorf R and Briegel H 2002 *Quantum Inf. Comp.* **6** 433
- [7] Raussendorf R, Browne D E and Briegel H J 2003 *Phys. Rev. A* **68** 022312
- [8] Vogel K and Risken H 1989 *Phys. Rev. A* **40**
- [9] Banaszek K, Radzewicz C, Wódkiewicz K and Krasinski J S 1999 *Phys. Rev. A* **60** 674
- [10] Chuang I L and Nielsen M A 1996 arXiv:quant-ph/9610001
- [11] D'Ariano G M and Maccone L 1998 *Phys. Rev. Lett.* **80** 5465
- [12] Lobino M, Korystov D, Kupchak C, Figueroa E, Sanders B C and Lvovsky A I 2008 *Science* **322** 563
- [13] Luis A and Sánchez-Soto L L 1999 *Phys. Rev. Lett.* **83** 3573
- [14] D'Ariano G M, Maccone L and Presti P L 2004 *Phys. Rev. Lett.* **93** 250407

- [15] Fiurášek J 2001 *Phys. Rev. A* **64** 024102
- [16] Coldenstrodt-Ronge H B, Lundeen J S, Pagnell K L, Feito A, Smith B J, Maurer W, Silberhorn C, Eisert J, Plenio M B and Walmsley I A 2008 *J. Mod. Opt.* **56** 432
- [17] Lundeen J S, Feito A, Coldenstrodt-Ronge H, Pagnell K L, Silberhorn Ch, Ralph T C, Eisert J, Plenio M B and Walmsley I A 2009 *Nat. Phys.* **5** 27
- [18] Boyd S and Vandenberghe L 2004 *Convex Optimization* (Cambridge: Cambridge University Press)
- [19] Makarov V, Vadim A, Anisimov A and Skaar J 2006 *Phys. Rev. A* **74** 022313
- [20] Masanes L I 2009 *Phys. Rev. Lett.* **102** 140501
- [21] Masanes L I, Renner R, Winter A, Barrett J and Christandl M 2006 arXiv:quant-ph/0606049
- [22] van Dam W, Magniez F, Mosca M and Santha M 2007 *SIAM J. Comput.* **37** 611
- [23] Freitag M, Martin Y, Misewich J A, Martel R and Avouris Ph 2003 *Nano Lett.* **3** 1067
- [24] Sasaki M, Wakui K, Mizuno J, Fujiwara M and Akiba M 2004 Quantum communication, measurement and computing *AIP Conf. Proc.* **734** 44
- [25] Kim J, Takeuchi S, Yamamoto Y and Hogue H H 1999 *Appl. Phys. Lett.* **74** 902
- [26] Shields A J, O'Sullivan M P, Farrer I, Ritchie D A, Hogg R A, Leadbeater M L, Norman C E and Pepper M 2000 *Appl. Phys. Lett.* **76** 3673
- [27] Gol'tsman G N, Okunev O, Chulkova G, Lipatov A, Semenov A, Smirnov K, Voronov B, Dzardanov A, Williams C and Sobolewski R 2001 *Appl. Phys. Lett.* **79** 705
- [28] Miller A J, Woo Nam S, Martinis J M and Sergienko A V 2003 *Appl. Phys. Lett.* **83** 791
- [29] Achilles D, Silberhorn C, Sliwa C, Banaszek K and Walmsley I A 2003 *Opt. Lett.* **28** 2387
- [30] Achilles D, Silberhorn C, Sliwa C, Banaszek K, Walmsley I A, Fitch M J, Jacobs B C, Pittman T B and Franson J D 2004 *J. Mod. Opt.* **51** 1499
- [31] Browne D E, Eisert J, Scheel S and Plenio M B 2003 *Phys. Rev. A* **67** 062320
- [32] Eisert J, Browne D E, Scheel S and Plenio M B 2004 *Ann. Phys., NY* **311** 431
- [33] Eisert J, Plenio M B, Browne D E, Scheel S and Feito A 2007 *Opt. Spectrosc.* **103** 181
- [34] Ourjoumtsev A, Dantan A, Tualle-Brouiri R and Grangier P 2007 *Phys. Rev. Lett.* **98** 030502
- [35] Olivares S, Paris M G A and Bonifacio R 2003 *Phys. Rev. A* **67** 032314
- [36] Opatrny T, Kurizki G and Welsch D-G 2000 *Phys. Rev. A* **61** 032302
- [37] Garcia-Patron Sanchez R, Fiurasek J, Cerf N J, Wenger J, Tualle-Brouiri R and Grangier Ph 2004 *Phys. Rev. Lett.* **93** 130409
- [38] Adesso G, Dell'Anno F, De Siena S, Illuminati F and Souza L A M 2009 *Phys. Rev. A* **79** 040305
- [39] Prochazka I *et al* 2007 *Proc. SPIE* **6583** 658302
- [40] Brida G, Degiovanni I P, Schettini V, Polyakov S V and Migdall A 2009 *J. Mod. Opt.* **56** 405
- [41] Kim Y H and Grice W P 2003 *Phys. Rev. A* **68** 062305
- [42] U'ren A B, Mukamel E, Banaszek K and Walmsley I A 2003 *Phil. Trans. R. Soc. A* **361** 1493
- [43] Boulant N, Havel T F, Pravia M A and Cory D G 2003 *Phys. Rev. A* **67** 042322
- [44] Klyshko D N 1980 *Sov. J. Quantum Electron.* **10** 1112
- [45] Rarity J D, Ridley K D and Tapster P R 1987 *Appl. Opt.* **26** 4616
- [46] Czitrovszky A, Sergienko A, Jani P and Nagy A 2000 *Metrologia* **37** 617
- [47] Audenaert K M R and Plenio M B 2006 *New J. Phys.* **8** 266
- [48] Eisert J, Brandão F G S L and Audenaert K M R 2007 *New J. Phys.* **9** 46
- [49] Guehne O, Reimpell M and Werner R F 2007 *Phys. Rev. Lett.* **98** 110502
- [50] Guehne O and Toth G 2009 *Phys. Rep.* **474** 1
- [51] Wunderlich H and Plenio M B 2009 arXiv:0902.2093
- [52] Leonhardt U 2003 *Rep. Prog. Phys.* **66** 1207
- [53] Coldenstrodt-Ronge H B and Silberhorn C 2007 *J. Phys.B: At. Mol. Opt. Phys.* **40** 3909
- [54] Plenio M B 2007 unpublished
- [55] Achilles D *et al* 2004 *J. Mod. Opt.* **51** 1499
- [56] Achilles D, Silberhorn C and Walmsley I A 2006 *Phys. Rev. Lett.* **97** 043602

- [57] Worsley A P *et al* 2009 *Opt. Express* **17** 4397
- [58] D'Ariano G M, Leonhardt U and Paul H 1995 *Phys. Rev. A* **52** R1801
- [59] Feito A 2008 Tools and methods for the distillation of entanglement in continuous variable quantum optics *PhD Thesis* Imperial College London
- [60] Ježek M, Fiurášek J and Hradil Z 2003 *Phys. Rev. A* **68** 012305
- [61] Schleich W P 2001 *Quantum Optics in Phase Space* (Berlin: Wiley)
- [62] Leonhardt U 1997 *Measuring the Quantum State of Light* (Cambridge: Cambridge University Press)
- [63] Leonhardt U, Munroe M, Kiss T, Richter T and Raymer M G 1996 *Opt. Commun.* **127** 144
- [64] Leonhardt U, Paul H and D'Ariano G M 1995 *Phys. Rev. A* **52** 4899
- [65] Lvovsky A I, Hansen H, Aichele T, Benson O, Mlynek J and Schiller S 2001 *Phys. Rev. Lett.* **87** 050402
- [66] Lvovsky A I and Raymer M G 2005 arXiv:quant-ph/0511044
- [67] Pregnell K L 2005 Retrodictive quantum state engineering, *PhD Thesis* Center for Quantum Computer Technology, Griffith University
- [68] Wünsche A 1997 *J. Mod. Opt.* **44** 2293