

# New Journal of Physics

The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft 

IOP Institute of Physics

Published in partnership  
with: Deutsche Physikalische  
Gesellschaft and the Institute  
of Physics



## PAPER

### OPEN ACCESS

RECEIVED  
21 October 2021

REVISED  
13 January 2022

ACCEPTED FOR PUBLICATION  
28 January 2022

PUBLISHED  
18 February 2022

# Time-dependent POVM reconstruction for single-photon avalanche photo diodes using adaptive regularization

E Fitzke<sup>1</sup>, R Krebs, Th Haase<sup>1</sup>, M Mengler, G Alber<sup>1</sup> and Th Walther<sup>\*1</sup>

Institute for Applied Physics, Technische Universität Darmstadt, Schlossgartenstraße 7, 64289 Darmstadt, Germany

\* Author to whom any correspondence should be addressed.

E-mail: [thomas.walther@physik.tu-darmstadt.de](mailto:thomas.walther@physik.tu-darmstadt.de)

**Keywords:** quantum detector tomography, single-photon detector, avalanche photo diode, quantum communication, timing jitter, time-dependent POVM, SPAD

Original content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](#).

Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the  
title of the work, journal  
citation and DOI.



## Abstract

We report on the first realization of time-dependent quantum detector tomography for commercially available InGaAs avalanche photo detectors. For the construction of appropriate time-dependent POVMs from experimentally measured data, we introduce a novel scheme to calculate the weight of the regularization term based on the amount of measured data. We compare our POVM-based results with the theoretical predictions of the previously developed model by Gouzien *et al* (2018 *Phys. Rev. A* **98** 013833). In contrast to our measurement-based construction of a time-dependent POVM for photon detectors, this previous investigation extends a time-independent POVM to a time-dependent one by including effects of detector timing jitter and dead time on the basis of particular model assumptions concerning the inner physical mechanisms of a photon detector. Our experimental results demonstrate that this latter approach is not sufficient to completely describe the observable properties of our InGaAs avalanche photo detectors. Thus, constructing the time-dependent POVM of a detector by direct quantum tomographic measurements can reveal information about the detector's interior that may not easily be included in time-independent POVMs by *a priori* model assumptions.

## 1. Introduction

Many applications in quantum information science such as the Boson sampling approach to quantum computing [2] or the characterization of quantum states [3] can benefit from detailed knowledge about the performance of single-photon detectors. Furthermore, the minimal requirements on detectors necessary for loophole-free Bell tests can be estimated when detector efficiencies as well as dark count rates are included in the analysis [4]. Detailed knowledge about the detector's peculiarities can also be interesting in the context of quantum key distribution (QKD). QKD, proposed in 1984 by Bennett and Brassard, uses principles of quantum mechanics to distribute secure cryptographic keys [5–7]. While in principle QKD provides information-theoretic security, actual implementations of QKD systems contain imperfections that can dilute this perfect degree of security. Various attacks on single-photon detectors in QKD systems were demonstrated [8–10], stressing that detailed knowledge about the detectors is mandatory to maintain security. Alternatively, protocols immune to detector imperfections, known as measurement-device-independent QKD [11], with experimentally challenging requirements on the quality of the photon sources have to be employed [12].

There are two fundamentally different approaches to detector characterization: the first approach is to thoroughly investigate all relevant effects on the measurement that arise from the detector's components and their interplay and to develop a detailed model of the detection process. However, this approach can easily become impractical for complex detector systems. The second approach is quantum detector tomography, which aims to make as few assumptions as possible about the detector and instead reconstructs the measurement operator of a quantum detector from measurement results obtained from the detector itself [13–15]. Quantum detector tomography describes the detector by a positive operator-valued

measure (POVM) completely characterizing the device. The POVM can be reconstructed by analyzing the detection results obtained for a set of tomographically complete input states. Detector tomography can for example be used to characterize the qubit readout in quantum computers [16].

A tomographically complete basis has to span the Hilbert space of the detector input states [17]. So far, realizations of quantum detector tomography concentrate on single modes of the electromagnetic field, ignoring any time dependency [15]. Nevertheless, reduced tomographic analysis can yield valuable information about the figures of merit relevant in detector characterization, such as response time, dark count rate, efficiency, wavelength or photon-number resolution [18]. Examples are tomographic measurements and POVM reconstruction for phase-insensitive detectors based on avalanche photo diodes [14, 19], single photon superconducting nanostrip detectors [20], transition edge sensors [21], time-multiplexed superconducting detectors [22] or photon-number-discriminating nanostrip detectors [23–25] as well as the analysis of phase-sensitive detectors [26–28].

If not the complete characterization of the detector is the primary interest, there exist less data-intensive alternatives to a full quantum detector tomography such as self-testing [29], self-calibrating [30, 31], or data-pattern tomography [32, 33]. For example, the last technique has successfully been demonstrated experimentally for quantum state tomography [34–36]. In contrast to quantum detector tomography the detector's POVM is never revealed and responses for unknown signals are fitted to data patterns of known probe states.

Although a time dependence is immanent to every measurement process, time-dependent tomography is rarely discussed. However, the detailed knowledge of the detector timing jitter is relevant for time-bin quantum measurements in QKD or quantum state tomography [37], for example. A time-dependent theoretical model for POVMs of non-photon-number-resolving detectors including timing jitter and dead time was recently proposed by Gouzien *et al* [1]. Here, we extend the tomography of single-photon avalanche detectors to time-dependent POVMs and test the validity of the model proposed by Gouzien *et al* [1] for our detectors. To the best of our knowledge this is the first experimental implementation of time-dependent detector tomography.

This paper is organized as follows: first, we briefly review the theory of quantum detector tomography and introduce time-dependent POVMs in section 2. By focusing on a single pulse shape for the input states we reduce the dimensionality of the detector's input Hilbert space. Since the reconstruction of the POVM elements from measured data is a mathematically ill-posed problem, regularization is necessary [14]. Often, the weight of this regularization is chosen by trial and error. Instead, we propose a novel adaptive regularization in section 3, weighting the regularization based on the amount of measured data. We show benchmarking results of the proposed scheme in comparison with a fixed-weight regularization. Our experimental setup is presented in section 4. With results integrated over the measurement time window we reconstruct the time-independent POVM of seven detectors, compare them with the expectation for ideal detectors with finite efficiency and deduce the detection efficiencies for different detector settings in section 5. Subsequently, we make use of the time resolution of the same data to reconstruct the time-dependent POVMs and apply them to one of the detectors in section 6. Finally, we compare our results in section 7 with the model of Gouzien *et al* [1] in order to evaluate its relevance for the theoretical description of our photon detector.

## 2. Time-dependent quantum detector tomography in the photon number basis

In this section, we use a time-dependent detector POVM for describing a phase-independent click-or-no-click detector under the assumption that this detector is hit by non-entangled input states. This detector model is based on a model previously presented by Gouzien *et al* [1] and takes advantage of a temporal multimode formalism as used by Rohde *et al* [38], for example. We generalize the previous work of Gouzien *et al* [1] by not restricting ourselves to a specific model of the detector's inner working. We also briefly discuss the relationship between the POVM reconstruction of Feito *et al* [14, 15] and the maximum-likelihood estimation of the POVM elements used in the following.

The most general description of the measurement results of a quantum measurement process is given by a POVM  $\Pi$  [39], i.e. by a set of positive semi-definite measurement operators  $\Pi = \{\hat{\Pi}_i\}$  with  $\sum_i \hat{\Pi}_i = 1$ , where  $i$  labels the different possible measurement results. If a quantum state  $\hat{\rho}$  is prepared, the probability of obtaining measurement result  $i$  yields

$$p_i(\hat{\rho}) = \text{tr}(\hat{\rho}\hat{\Pi}_i) \quad (1)$$

and  $p_i \geq 0$  is ensured by the positive semi-definiteness of the operators  $\hat{\Pi}_i$  [13]. Quantum detector tomography is concerned with the reconstruction of these measurement operators from tomographic measurements [14].

Throughout this paper we consider detectors with two measurement results, i.e. ‘click’ and ‘no click’, in a time-dependent setting such that the probability density associated with a ‘click’ event at time  $t$  is given by  $p_{\text{click}}(t, \hat{\varrho}) = \text{tr}(\hat{\varrho} \hat{\pi}_{\text{click}}(t))$ . The corresponding time-independent POVM for a time interval  $I$  is given by  $\hat{\Pi}_{I,\text{click}} = \int_I \hat{\pi}_{\text{click}}(t) dt$  and  $\hat{\Pi}_{I,\text{no click}} = 1 - \hat{\Pi}_{I,\text{click}}$ . A complete time-dependent tomography is experimentally challenging, since it has to span the infinite dimensional space of all photon states at each instant of time [38]. Thus, detector tomography is often reduced to a single mode [14, 15, 22, 25] of the radiation field.

For a single mode, Fock states  $|k\rangle$  form a tomographically complete set of basis states. Here, we are interested in describing time-dependent phenomena. For this purpose we use the temporal multimode formalism from [38] which has already been used to formulate a model for time-dependent POVMs by Gouzien *et al* [1]. Thereby, we restrict the relevant Hilbert space to non-entangled time-localized states and assume that the detector dead time is much longer than the time interval considered so that at most one click can be registered in the time interval of interest.

Avalanche photo diodes do not have any external phase reference and are thus phase-insensitive detectors. POVMs of phase-insensitive detectors are diagonal in the Fock basis [17] and can thus be described by POVM operators of the form

$$\hat{\pi}_i = \sum_{k=0}^{\infty} \Theta_{k,i} |k\rangle \langle k|. \quad (2)$$

Including the time dependencies of the photons arriving at an avalanche photo diode we can represent the photon detector’s POVM in the form

$$\hat{\pi}_{\text{click}}(t) = \mathcal{T} \sum_{k=0}^{\infty} \int_{\mathbb{R}^k} p_{\text{click},k}(t, \tau_k) \hat{P}_k(\tau_k) d\tau_k. \quad (3)$$

Thereby,  $\tau_k = \{\tau_1, \dots, \tau_k\}$  denotes the times at which  $k$  time-localized photons arrive at the photon detector. The time-ordering operator is denoted by  $\mathcal{T}$  and ensures the ordering  $\tau_1 < \tau_2 < \dots < \tau_k$ . The quantum state of the  $k$  time-localized photons is given by the projection operator

$$\hat{P}_k(\tau_k) = |\tau_k\rangle \langle \tau_k| \quad \text{with} \quad |\tau_k\rangle = \bigotimes_{j=1}^k \hat{a}^\dagger(\tau_j) |0\rangle. \quad (4)$$

Consistent with the rotating wave approximation, the creation operator  $\hat{a}^\dagger(\tau_j) = (2\pi)^{-1/2} \int_{\mathbb{R}} \hat{a}^\dagger(\omega) \exp(i\omega\tau_j) d\omega$  describes the creation of a single time-localized photon at time  $\tau_j$  when the assumption is made that the bandwidth of the field excitation is much smaller than the optical center frequency [40, 41]. The corresponding annihilation operators fulfill the commutation relation  $[\hat{a}(\tau_i), \hat{a}^\dagger(\tau_j)] = \delta(\tau_i - \tau_j)$ . The probability density  $p_{\text{click},k}(t, \tau_k)$  describes the probability that a state with  $k$  photons localized at times  $\tau_1 \dots \tau_k$  causes a click of the photon detector in the time interval  $[t, t + dt]$ . The POVM of (3) is a generalization of the time-dependent POVM model proposed by Gouzien *et al* [1] as the probability density  $p_{\text{click},k}(t, \tau_k)$  is not restricted to a specific model of the detector’s interior. In order to obtain a finite set of measurement operators, in the following we split the integral in (3) into time bins of width  $\Delta t$  labeled from  $i = 1$  to  $i_{\max}$ . Thus, the POVM to be reconstructed has  $i_{\max} + 1$  different POVM operators, one for each time bin plus one for no click in any time bin.

Compared to Fock states the overcomplete basis of coherent states  $|\alpha\rangle$  is more convenient for describing experiments as coherent states are naturally produced by attenuated laser light. If a detector is exposed to  $N(\alpha_j)$  such attenuated light pulses of a coherent state  $|\alpha_j\rangle$  the number of clicks concerning measurement result  $i$ , i.e.  $n_i(\alpha_j)$ , can be recorded for all possible measurement results  $i = 1, \dots, i_{\max}$ . This can be repeated for the different coherent states with  $j = 1, \dots, j_{\max}$ . The resulting relative frequencies  $f_i(\alpha_j) = n_i(\alpha_j)/N(\alpha_j)$  can be compared to the probabilities  $p_i(\alpha_j)$  predicted by a given POVM. In practice only a finite number  $j_{\max}$  of different values of  $\alpha$  can be measured. As coherent states are linear superpositions of infinitely many photon number eigenstates, in practice also only a maximum number of photons  $k_{\max}$  can be measured. Consequently, in terms of POVM parameters measured probabilities are described theoretically by the relation  $p_i(\alpha_j, \Theta) = \sum_{k=0}^{k_{\max}} |\langle \alpha_j | k \rangle|^2 \Theta_{k,i}$ . Thus, for a tomographic reconstruction of the POVM describing the photon detector the parameters  $\Theta_{k,i}$  have to be inferred from the measured frequencies  $f_i(\alpha_j)$ .

For the reconstruction, the probabilities are approximated by the measured frequencies  $f_i(\alpha_j)$ , so that in matrix notation the relation between POVM elements and measured frequencies is given by

$$F_{j_{\max} \times i_{\max}} = C_{j_{\max} \times (k_{\max} + 1)} \Theta_{(k_{\max} + 1) \times i_{\max}} \quad (5)$$

with  $C_{jk} = |\langle \alpha_j | k \rangle|^2 = \exp(-\mu_j) \mu_j^k / k!$  and with the mean photon number  $\mu_j = |\alpha_j|^2$ . In general, the matrix  $\mathbf{C}$  is not invertible which complicates the determination of the POVM elements  $\Theta_{i,k}$  from the measured data in  $\mathbf{F}$  with  $F_{ij} = f_i(\alpha_j)$ . One possibility to solve this problem is to minimize  $\|\mathbf{F} - \mathbf{C}\Theta\|_F$  with the Frobenius norm  $\|M\|_F = (\sum_{i,j} |m_{ij}|^2)^{1/2}$  [14, 15]. In order to avoid unphysical solutions from this ill-conditioned optimization problem it is convenient to add a regularization term to the objective function [42]. For example, in [14, 15] a quadratic regularization term

$$r \sum_{k,i} (\Theta_{k+1,i} - \Theta_{k,i})^2 \quad (6)$$

with a regularization coefficient  $r$  was used. The coefficient was chosen in a range so that a smooth distribution of the POVM elements is obtained and the reconstructed results are stable [14, 15].

The necessary regularization comes at the price of biasing the optimization. Thus, it is important to estimate a value for the regularization coefficient in order to avoid unduly biasing the optimization. In our subsequent treatment we split our measurement data into time bins and evaluate the bins individually. The number of recorded events per time bin varies across the considered time interval. The more data are available for a time bin, the smaller the statistical measurement uncertainty, the smaller  $r$  may be chosen. Instead of choosing a new value of  $r$  for each time bin by trial and error we propose an adaptive estimation value  $r$  which depends on the amount of measured data.

### 3. Adaptive regularization

In this section, we show the relationship between regularization terms that were used for POVM reconstructions in [14, 15] and Bayesian prior distributions in the maximum-likelihood estimation of POVMs. We use this relationship to derive an estimation for the weighting coefficient  $r$  of the regularization term that depends on the amount of measured data and on the number of reconstructed elements. Subsequently, we benchmark the adaptive regularization scheme in comparison with regularization with a fixed coefficient.

In order to motivate an estimation value of  $r$  we consider the maximum-likelihood approach [43] for inversion of (5). The measured frequencies  $f_i(\alpha_j)$  may be viewed as the empirical mean values of a Bernoulli experiment with probabilities  $p_i(\alpha_j)$  and  $(1 - p_i(\alpha_j))$  which has finite variance  $\sigma_{ij}^2 = p_i(\alpha_j)(1 - p_i(\alpha_j))$ . The measured data are generated from statistically independent repetitions of Bernoulli experiments. For a sufficiently large number of repetitions,  $\sigma_{ij}^2 \approx f_i(\alpha_j)(1 - f_i(\alpha_j))$  holds and the central limit theorem ensures that the distribution of  $f_i(\alpha_j)$  around  $p_i(\alpha_j, \Theta)$  is a normal distribution with variance  $\sigma_{ij}^2/N(\alpha_j)$ . This means the likelihood function is given by

$$L(\Theta) = \frac{1}{(2\pi)^{i_{\max} j_{\max}/2}} \exp \left( -\frac{1}{2} \sum_{i,j} \left( \frac{f_i(\alpha_j) - p_i(\alpha_j, \Theta)}{N(\alpha_j)^{-1/2} \sigma_{ij}} \right)^2 \right) \prod_{i,j} \sigma_{ij}^{-1} N(\alpha_j)^{1/2}. \quad (7)$$

In order to obtain an estimation for the parameters  $\Theta$ , the likelihood function can be maximized or, equivalently, the negative log-likelihood  $l(\Theta) = -\ln(L(\Theta))$  can be minimized. The sums over  $i$  are independent of each other, so that they can be minimized separately. When all constant factors are omitted and it is assumed that the  $\sigma_{ij}$  are independent of  $i$  for the same value of  $j$ , minimizing the negative log-likelihood  $l(\Theta)$  is equivalent to solving the least-squares minimization problem  $S(\Theta) = \|\mathbf{F} - \mathbf{C}\Theta\|_F^2$ . The square root is a strictly monotone function, so that minimizing the norm  $\|\mathbf{F} - \mathbf{C}\Theta\|_F$ , as it was done in [14, 15], is also equivalent to the maximum-likelihood approach.

Adding regularization terms biases the optimization. The term of (6), for example, biases POVMs towards close-by values for adjacent Fock basis matrix elements. This bias can be interpreted as stemming from information predating the measurement in the form of a Bayesian prior distribution. Bayes' theorem allows to relate the likelihood  $P(\mathbf{F}|\Theta)$  of detecting results  $\mathbf{F}$ , given the parameters  $\Theta$ , and the prior distribution  $P(\Theta)$ , to the posterior probability  $P(\Theta|\mathbf{F})$  of  $\Theta$  being the parameter set if  $\mathbf{F}$  is measured:

$$P(\Theta|\mathbf{F}) = \frac{P(\mathbf{F}|\Theta)P(\Theta)}{P(\mathbf{F})}. \quad (8)$$

Therefore, the (additive) regularization term can be understood as the negative log-likelihood of the (multiplicative) prior, so that the negative log-posterior function becomes

$$l_{\text{posterior}}(\Theta) = l(\Theta) - \ln(P(\Theta)) + C \quad (9)$$

with the additive constant  $C$  arising from the probability  $P(\mathbf{F})$  of (8).

Regularization terms as in (6) can thus be interpreted as a Gaussian prior  $P(\Theta) = \exp(-\gamma \sum_{k,i} (\Theta_{k+1,i} - \Theta_{k,i})^2/2)$ . The coefficient  $\gamma$  can be interpreted as the inverse covariance of neighboring matrix elements of  $\Theta$ . Therefore,  $\gamma^{-1/2}$  is the expected characteristic distance of neighboring matrix elements. A prior can be constructed under the assumption that the matrix elements of  $\Theta$  are equidistantly spaced between 0 and  $k_{\max}$ . Consequently, it can be expected that the average distance between neighboring matrix elements is  $k_{\max}^{-1}$  along the  $k$ -axis. For the  $i_{\max}$  different measurement results it can be assumed that they are equally distributed. This yields the relation  $\gamma = k_{\max}^2 i_{\max}^2$ .

The coefficient  $r$  can also be related to the statistical measurement error which can be estimated by taking the maximum over the variances of the normal distributions in (7) as

$$\epsilon_i^2 = \max_j \frac{\sigma_j^2}{N(\alpha_j)} \approx \max_j \frac{f_i(\alpha_j)(1 - f_i(\alpha_j))}{N(\alpha_j)}. \quad (10)$$

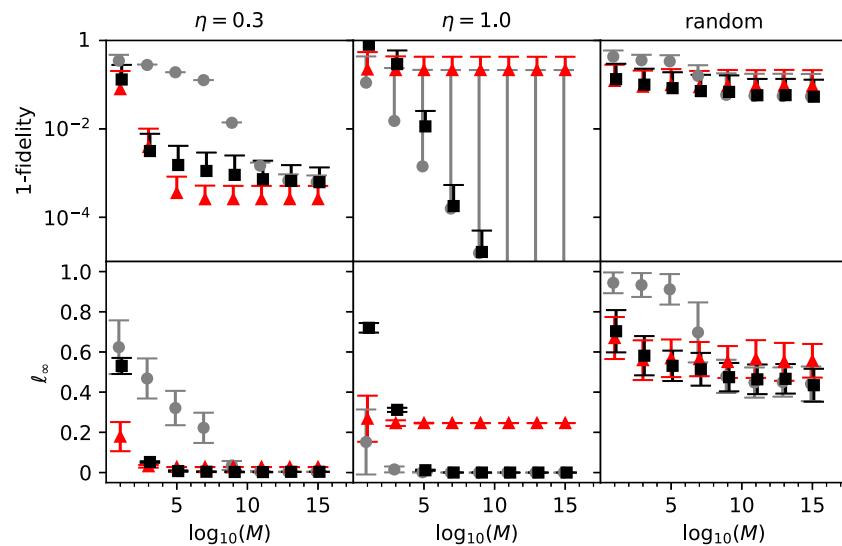
Here, the maximum is taken over the tomographic states  $j$  in order to obtain an upper bound on the uncertainty.

The minimum of the negative log-posterior remains unchanged under multiplication with a positive constant. Therefore, we multiply the objective function by  $\epsilon^2$  and separate the regularization coefficient according to the relation  $r = \epsilon^2 \gamma$ . Thus, for a fixed value of  $\gamma$ , the weight of the regularization term becomes smaller for more accurate measurements. The more data are available the smaller the measurement uncertainty  $\epsilon^2$  and the more the prior bias is discounted.

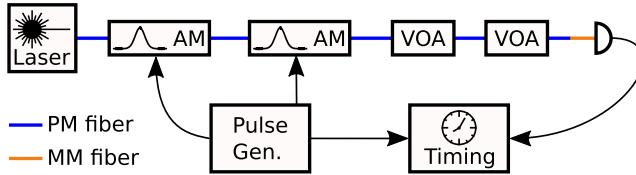
In order to show the importance of regularization and the disadvantages of a static regularization term independent of the statistical measurement uncertainty, especially if the amount of measured data becomes smaller, we now compare three different types of reconstruction schemes with least-squares minimization. These three schemes differ in the regularization used in the minimization. We compare results without regularization with static regularization, i.e. with a constant weight  $r$ , and with the adaptive scheme motivated above where the regularization weight is adjusted by the variance of the measured data. We used a value of  $r = 0.1$  for the static case and a value of  $\gamma = k_{\max}^2$  for the adaptive case, as we consider only one result, i.e.  $i_{\max} = 1$ . A cutoff of  $k_{\max} = 29$  was chosen for the maximum Fock state used in the tomography and 30 coherent tomographic states were chosen with mean photon numbers  $\mu = \{0, 1, \dots, 29\}$ . The data were recorded by randomly sampling measurement data for known POVMs of detectors with only two results. Three different types of POVMs are studied in the benchmarking: two ideal detectors with sensitivities  $\eta = 1$  and  $\eta = 0.3$  with ‘no-click’ POVM elements  $\Theta_{k,0} = (1 - \eta)^k$ , and thirdly POVMs with randomly sampled diagonal elements.

For the POVMs obtained from the reconstruction the maximum norm denoted by  $\ell_\infty = \max_k |\vartheta_{k,\text{true}} - \vartheta_{k,\text{reconstr.}}|$  and the fidelity  $f = \sum_k (\vartheta_{k,\text{true}} \vartheta_{k,\text{reconstr.}})^{1/2}$  with respect to the true POVM are compared. Within each simulated experiment all tomographic states were measured  $M$  times. The same statistics is performed over various values of  $M$  in order to quantify the performance for an increasing amount of data. The simulation was repeated  $N$  times for each value of  $M$  in order to estimate both the mean and the variance of the performance. The number of repetitions  $N$  was set to a value of 100. For each POVM and each tomographic state  $M$  measurements were sampled, a tomography was performed, and the results were compared with the true POVM. Among the studied POVMs the diagonal elements of the random POVMs were newly sampled from a uniform distribution in each of the  $N$  iterations. The other two ensembles used the same POVM across all iterations.

Results of this benchmarking are presented in figure 1 clearly showing the improvement gained by regularization for low values of  $M$  across all chosen underlying POVMs. For higher values of  $M$ , all methods show improvement with increasing  $M$ . For the random POVM, the data do not appear sufficient for high quality reconstruction, as both the error and its variance remain quite large, even for the highest values of  $M$ , compared to the other two columns. Furthermore, it appears that the statically regularized method does not converge to optimum values, but to levels of a close-by but distinct value in all benchmarks. Only in the case of the detector with  $\eta = 0.3$ , all three methods seem to have similar asymptotic performance. The asymptotics for the statically regularized method emphasize that the regularization parameter should not be chosen independently of the number of data points. The adaptively regularized method shows for large datasets an asymptotic performance similar to least-squares tomography without regularization. It improves the results when the datasets are small, similar to the static regularization, but avoids convergence to wrong POVMs when the amount of data is large enough. We thus use adaptive regularization for the reconstruction of POVMs in the following sections.



**Figure 1.** Benchmarking with detector type varying by column: the closeness of the POVM elements of the true detector and of the reconstructed elements are quantified by  $\ell_\infty = \max_k |\vartheta_{k,\text{true}} - \vartheta_{k,\text{reconstr.}}|$  and a metric proportional to the fidelity  $1 - f = 1 - \sum_k (\vartheta_{k,\text{true}} \vartheta_{k,\text{reconstr.}})^{1/2}$ . For the one-fidelity metric, each error bar that crosses zero is not plotted. The first two columns show idealized detectors of finite sensitivity  $\eta$  without dark counts. Grey circles represent unregularized least squares, red triangles represent statically regularized least squares and black squares represent adaptive least squares. The x-axes show the number of trials  $M$  per tomographic state.



**Figure 2.** Fiber-based setup for time-dependent and time-independent detector tomography. AM: LiNbO<sub>3</sub> amplitude modulator, VOA: variable optical attenuator, pulse gen.: electronic pulse generator, timing: time tagging electronics, PM fiber: polarization maintaining fiber, MM fiber: multimode fiber.

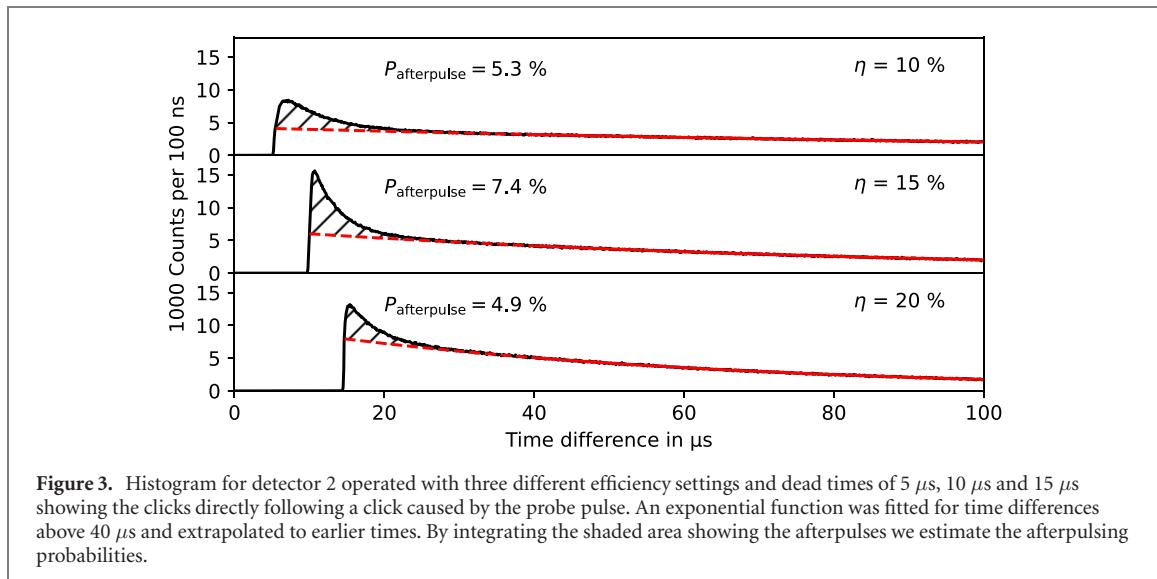
#### 4. Experimental setup

Tomography measurements were performed for seven free-running commercial InGaAs single-photon avalanche photo diodes (model ID220 with multimode fiber, ID Quantique). These detectors have three efficiency settings (10%, 15%, 20%) corresponding to different photo diode voltages. The dead time can be selected between 1  $\mu$ s and 20  $\mu$ s. In general, higher efficiencies and shorter dead times are preferable, but these settings come with a trade-off: the higher the efficiency is set, the higher the probability for detector afterpulsing and dark counts for the same dead time setting. Afterpulses can be suppressed by choosing higher values for the dead time when a high efficiency is set. Thus, three combinations of detector settings were chosen for the experiments: 5  $\mu$ s dead time for 10% set efficiency, 10  $\mu$ s for 15% and 15  $\mu$ s for 20%.

In order to perform both time-independent and time-dependent tomography, we set up the experiment shown in figure 2. Laser pulses with a defined mean photon number  $\mu$  were generated and detector clicks were registered correlated to the pulse emission.

The setup consists of a fiber-coupled cw DFB diode laser with a central wavelength of 1550.52 nm and 74 mW output power, two cascaded amplitude modulators, a manual and an electronic variable attenuator as well as electronics for timing acquisition and pulse generation.

The first modulator was used to shape pulses with an FWHM duration of 0.24 ns at a repetition rate of 10 kHz. The rate was chosen to be low enough so that a repetition cycle was much longer than the detector dead time. Hence, unwanted correlations between subsequent clicks introduced by afterpulses or variations in the dead time were prevented. The pulse duration for the second modulator was set to 10 ns and the delay was set according to the optical delay between both modulators, ensuring that the second modulator was completely opened when a pulse passed by. Therefore, the pulse shape was solely determined by the first modulator. Within the opening time window of the second modulator, the extinction ratio is determined by



the first modulator only. Outside of the time window, the extinction ratios of both modulators multiply, resulting in a sufficiently high suppression of uncorrelated detection events from photons leaking through the modulators during the time between pulses.

Both modulators were driven by a dual-channel pulse generator (HP 8131 A). The detector's electrical output as well as the trigger output of the pulse generator were connected to the timing acquisition electronics (ID Quantique ID900) with 13 ps resolution. Timestamps for detector clicks and the trigger pulses were recorded in different channels.

The first modulator was stabilized by a bias controller in order to prevent changes in the pulse shape and energy. In order to avoid bias drifts of the second modulator, a recalibration of the bias voltage was run before measuring each value of the mean photon number  $\mu$  by sweeping the voltage and setting the bias voltage that minimized the rate of uncorrelated counts.

For the tomography measurements, the average power was measured behind the second modulator and the pulse energy was calculated from the repetition rate. The mean photon number of the pulses  $\mu$  was scanned by adjusting the attenuation value of the variable attenuator.

## 5. Reconstruction of time-independent POVMs

In order to calculate time-independent POVMs for the detectors, histograms for the time difference between reference clicks from the pulse generator and detector clicks were calculated for a measurement series over the mean photon numbers  $\mu$  between 0 and  $\mu_{\max} = 50$  in steps of 2 with 10 min measurement time per value.

As we only want to consider the direct clicks of the detector, clicks in a time window of 8 ns covering the maximum in the arrival time histogram (cf figure 6 in section 6) were selected for the tomography. The clicks in this time window are directly caused by the probe pulse. Therefore, the influence of dark counts and afterpulses on the results is suppressed. In order to verify the successful suppression of afterpulsing, we calculated the histograms of arrival times for the clicks following next after a click in the time window as shown in figure 3. It can be compared to the click probability for the next click  $p_\lambda(t)$  of an ideal detector without dead time and afterpulsing, which is given by a homogeneous Poisson process with a constant click rate  $\lambda$  and  $p_\lambda(t) = \lambda \exp(-\lambda t)$  [44].

For small time differences, the click rate is zero and shows a steep edge when the detector recovers from the dead time. When a click happens roughly one dead time after the preceding click, the detector may have not recovered completely. Including such clicks in the POVM reconstruction would thus lead to incorrect results. Thus, we excluded clicks preceded by another click within a time frame of the set dead time plus two microseconds from further analysis. The maximum is caused by afterpulses. For longer time differences between clicks above approximately 30  $\mu\text{s}$ , the probability shows an exponential decay as it would be expected from the photon arrival statistics of uncorrelated counts. By fitting an exponential function to the tail for times above 40  $\mu\text{s}$  and extending this function to the afterpulsing region, we estimated the number of afterpulses as the area between the histogram and the fit and calculated the afterpulsing probability to 5.3%, 7.4% and 4.9% for the three efficiency settings in the order 10%, 15% and

20%, respectively. The afterpulses are restricted to time differences of less than 40  $\mu\text{s}$  for all settings. Thus, by choosing the repetition rate to 10 kHz, afterpulses do not occur in the next pulse cycle and can thus not cause unwanted correlations between cycles.

The detection probabilities for the POVM reconstruction were calculated by dividing the sum of detections within the 8 ns time window by the number of pulses. In this way, we obtained the time-independent POVMs from the detection probabilities by minimizing

$$\|\mathbf{f} - \mathbf{C}\vartheta\|_2^2 + \epsilon^2 \gamma \sum_{k=0}^{k_{\max}} (\vartheta_{k+1} - \vartheta_k)^2 \quad (11)$$

over  $\vartheta$ . Here, the matrices  $\mathbf{F}$  and  $\Theta$  became vectors  $\mathbf{f}$  and  $\vartheta$ , as only one single detection result, the no-click event in the whole interval, was considered.

For the photon numbers, a reasonable cutoff  $k_{\max}$  needs to be found. The coefficients  $C_{jk}$  decrease for higher  $k$  according to the Poissonian distribution  $\exp(-\mu)\mu^k/k!$  with the standard deviation of the photon number given by  $\mu^{1/2}$ . We decided to reconstruct the elements up to  $\mu_{\max}$  plus two standard deviations, i.e. chose  $k_{\max} \approx \mu_{\max} + 2\sqrt{\mu_{\max}} = 64$  resulting in  $C_{j_{\max} k_{\max}} < 1\%$ . Thus, the weighting coefficients of the regularization term are  $\gamma = 64^2$  and  $\epsilon^2 \approx 4.4 \times 10^{-8}$  which was calculated for detector 2 with 15% efficiency and 10  $\mu\text{s}$  dead time according to (10). Convergence of the minimization was facilitated by explicitly implementing the gradient of (11).

In order to compare the results, we consider the POVM for an ideal detector with efficiency  $\eta$ . When such a detector is exposed to a photon, the probability that it is not triggered is given by  $(1 - \eta)$ . Thus, the no-click probability for  $k$  photons is given by  $P_{\text{no click}}(k) = 1 - P_{\text{click}}(k) = (1 - \eta)^k$  and (2) becomes  $\hat{\pi}_{\text{no click}} = \sum_{k=0}^{\infty} (1 - \eta)^k |k\rangle \langle k|$ . Thus, the no-click probabilities for a number state  $|k\rangle$  and for a coherent state  $|\alpha\rangle$  of an ideal detector are given by

$$\langle k | \hat{\pi}_{\text{no click}} | k \rangle = (1 - \eta)^k \quad \text{and} \quad \langle \alpha | \hat{\pi}_{\text{no click}} | \alpha \rangle = \exp(-\eta\mu). \quad (12)$$

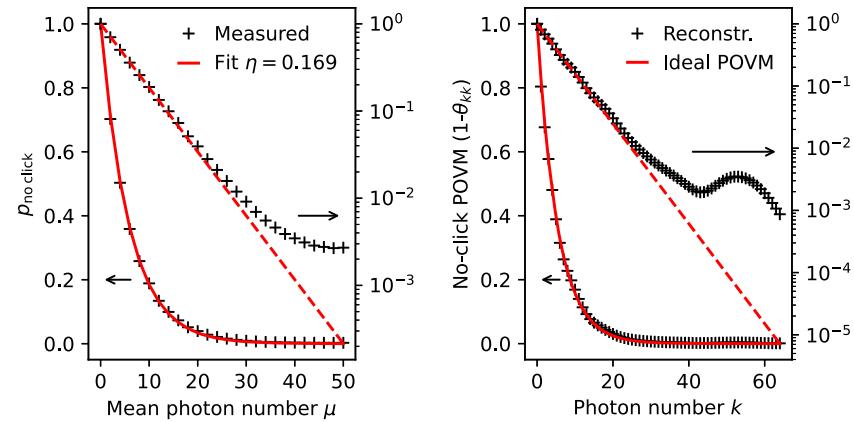
In order to obtain a value for the detector efficiency, the POVM element for  $k = 1$  can be considered or, alternatively, the efficiency can directly be derived from the measured data by fitting an exponential function to the number of clicks over the mean photon number  $\mu$  according to (12).

The time-independent POVMs were reconstructed for all seven detectors. Exemplary results for detection probabilities and time-independent POVMs of detector 2 (cf figure 5) with an efficiency setting of 15% and 10  $\mu\text{s}$  dead time are shown in figure 4 along with a fit for an ideal detector. The measured no-click probability well matches the exponential distribution of an ideal detector with  $\eta = 16.9\%$  according to (12). For values above  $\mu = 30$ , however, the logarithmic scale shows that the no-click probability is higher than for an ideal detector and approaches a value of  $2.7 \times 10^{-3}$ , irrespective of a further increase of  $\mu$ . In order to investigate this effect, the selection criterion for clicks was extended from 12  $\mu\text{s}$  to 99  $\mu\text{s}$  as the required distance to the preceding click. However, no change was observed. Thus, we conclude that this effect is independent of the time difference to the preceding click and is not related to the dead time. The POVM elements shown in figure 4 reflect this behavior: up to photon numbers around 30, they match the model of an ideal detector well, but for higher values of  $k$  they are larger than predicted by the model. We investigated the dip visible at  $k = 43$  in the logarithmic plot of POVM elements by varying  $k_{\max}$  and  $\mu_{\max}$ . We observed that a change of  $\mu_{\max}$  lead to significant changes in the shape of the dip or even multiple dips. The dip may thus be regarded as a small boundary effect caused by the finite value of  $\mu_{\max}$  in the experiment.

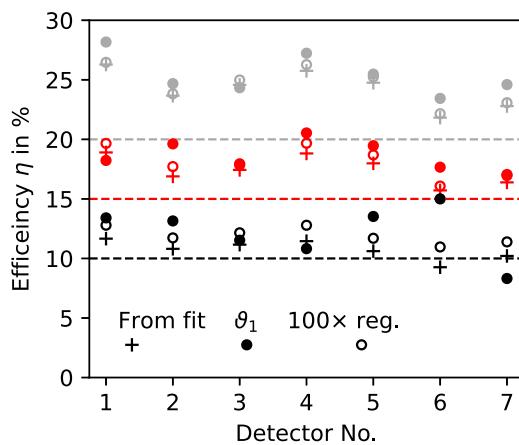
For all seven detectors the efficiencies obtained from the POVM reconstruction are shown in figure 5. Systematic relative measurement uncertainties for  $\mu$  are introduced by the accuracy of the photo detector used for attenuator calibration (5%) and by variations of losses in the fiber–fiber connections (10%). However, these values are constant for the measurements shown. In principle, these values can be further improved by using a tightly calibrated photo detector for the attenuation calibration and by using permanent, spliced fiber connections. Repeated measurements of the same detector yielded a variation of 8%.

We determined the values for the detection efficiencies in three different ways which are compared in the figure. First, the efficiency was extracted from a fit of the exponential distribution in (12) to the measured values of  $p_{\text{no click}}(\mu)$ . Second, the POVM elements were reconstructed with the proposed adaptive weight  $r = \epsilon^2 \gamma$  of the regularization term and the value  $(1 - \vartheta_1)$  was interpreted as efficiency. Third, the second procedure was repeated with a 100 times stronger regularization.

The determined efficiency values in figure 5 match the expected values stated as detector settings generally being slightly higher than these values. For most of the detectors the efficiency obtained from the strongly regularized reconstruction is in better agreement with the value determined from the fit than the value obtained with normal regularization. The fact that reasonable POVMs can be obtained although the



**Figure 4.** Time-independent POVM elements for detector 2 (cf figure 5) with 15% efficiency and 10  $\mu\text{s}$  dead time: from a fit of the no-click probabilities according to (12), the detector efficiency was estimated to  $\eta = 16.9\%$ . This value was used to plot the ideal no-click POVM according to (12) on the right-hand side. Each diagram shows the data once in a linear scale (left vertical axis) and in a logarithmic scale (right vertical axis, dashed line).

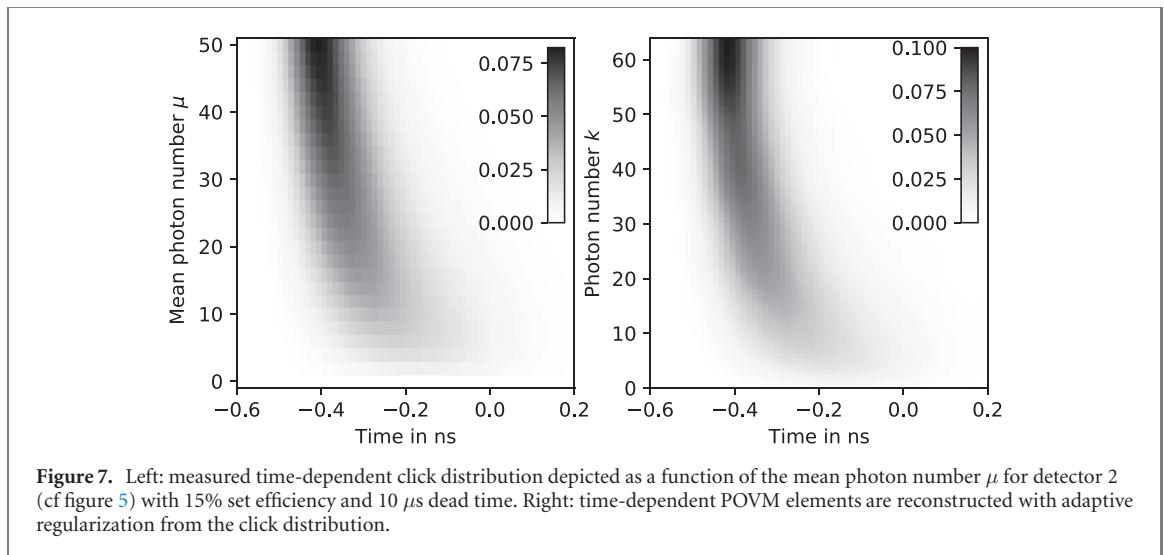
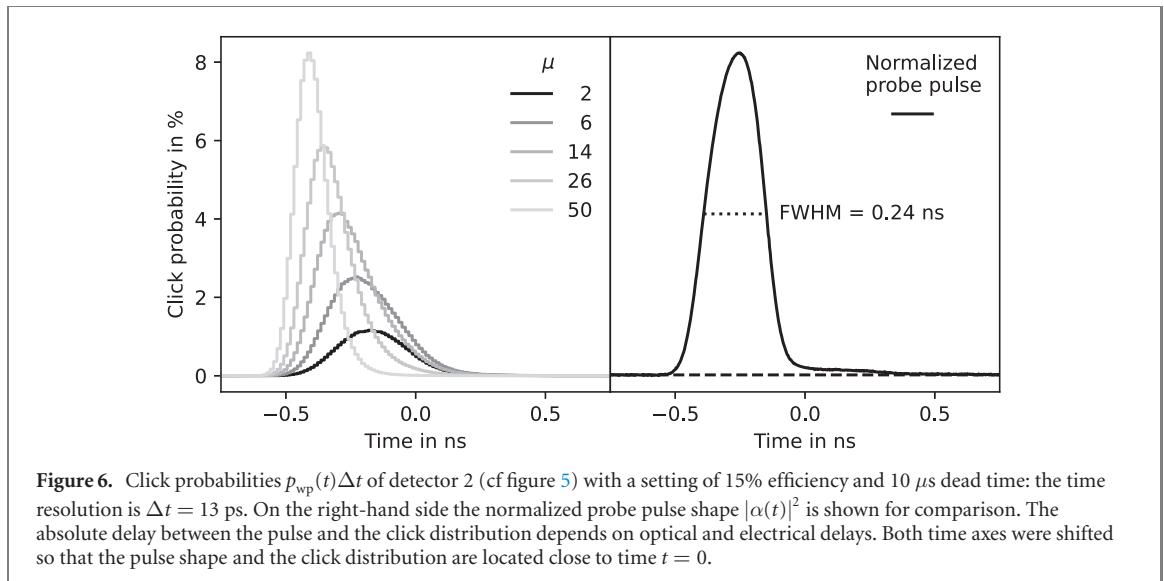


**Figure 5.** Detection efficiencies of the seven detectors tested, calculated from an exponential fit to  $p_{\text{no click}}(\mu)$  and from the first POVM element  $\vartheta_1$ , reconstructed with the proposed weighting factor  $r = \epsilon^2 \gamma$  of the regularization term: in addition, results with a 100 times stronger regularization are shown. The colors indicate the detector efficiency settings 10% (black), 15% (red) and 20% (gray).

regularization coefficient can be varied by more than two orders of magnitude has already been observed in [14]. In this previous investigation it has been concluded that the regularization is mainly necessary to ensure a well-conditioned optimization and that choosing the regularization coefficient in this range does not excessively distort the results. Thus, we conclude that the proposed adaptive regularization coefficient can be understood as a rule-of-thumb value to obtain reasonable results from the POVM reconstruction. Notably, it is not a strict value so that larger or smaller values may also be chosen, depending on the specific situation.

## 6. Time-dependent POVMs

The temporal resolution of the setup also allowed for a time-resolved measurement of detection probabilities. In general, the temporal distribution of the clicks depends on the temporal shape of the probe pulse and on the detector response. Exemplary results for detector 2 are shown in figure 6 along with the probe pulse shape. With increasing values of  $\mu$ , the maximum of the click distribution shifts to earlier times with respect to the input pulse and also becomes narrower. This effect can be understood intuitively: after the click the detector switches into the dead time, so that other photons in the pulse cannot cause subsequent clicks for the same pulse. The higher  $\mu$ , the higher the probability that a photon located early within the pulse causes a click. The deformation of the click distribution can be expected to become significant above  $\eta\mu \approx 1$ , as a detector without dead time would likely yield multiple clicks per pulse, but a detector with dead time only registers the first click per pulse.



From the resulting click probability distribution, a time-dependent POVM as in (2) was reconstructed, where clicks in a specific time bin correspond to a specific detection result  $i$ . Here, we minimized (11) for each time bin individually, with the variance  $\epsilon^2$  calculated from the data in this time bin only. The optimization was performed with bounds for the POVM elements to ensure physically reasonable values between 0 and 1. The measured click probabilities and the reconstructed POVM for detector 2 are shown in figure 7.

The shape of the POVM roughly matches that of the click probability distribution. This raises the question whether the description of the detector by a time-dependent POVM is necessary or if the temporal shape of the distribution can be explained by modeling the detector behavior. In the next section, we compare the reconstructed POVM to the POVM predicted by a detector model describing the click probability deformation explained above.

## 7. Test of a detector model including dead time and timing jitter

In the last section we argued that the changes of the shape of the click distribution with increasing mean photon number  $\mu$  qualitatively match the expectation for a detector which only responds to the first photon in each pulse. Gouzien *et al* [1] formalized this argument and developed a model describing time-dependent POVMs of click-or-no-click detectors. In this section, we investigate whether this model is sufficient in order to describe the observed temporal detector click probability for our detectors.

The model of Gouzien *et al* [1] assumes a detector having an intrinsic jitter distribution  $J(T)$ . The value  $J(t - \tau)dt$  is the probability that a photon hitting the detector at time  $\tau$  causes a click in the time interval

$[t, t + dt]$ . Causality requires that  $J(t < 0) = 0$ . Furthermore, it is assumed that the dead time is much longer than the probe pulse duration and that the detector switches immediately into dead time after the first click. This means that after a click the detector is inactive for the remaining pulse duration and all other photons in the pulse cannot cause subsequent clicks. The probabilities

$$p_1(t, \tau) = \eta J(t - \tau) \quad \text{and} \quad p_{1,\text{not}}(t, \tau) = 1 - \eta \int_{\tau}^t J(t' - \tau) dt' \quad (13)$$

describe the probability that a single photon hitting the detector at  $\tau$  causes a click at  $t$  and that a photon has not caused a click up to time  $t$ , respectively.

The probability  $p_{\text{click}}(t, \tau_k)$  of obtaining a click at time  $t$  from a pulse with  $k$  photons arriving at times  $\tau_k = (\tau_1, \dots, \tau_k)$  can now be written as the sum of the probabilities that one particular photon causes the click and the probability that all other  $k - 1$  photons did not yet cause a click:

$$p_{\text{click}}(t, \tau_k) = \sum_{j=1}^k p_1(t, \tau_j) \prod_{\substack{l=1 \\ l \neq j}}^k p_{1,\text{not}}(t, \tau_l). \quad (14)$$

Multiplying the probabilities is justified by the assumption that apart from the dead time effect the photons cause clicks according to the jitter distribution independently of each other. The POVM model proposed in [1] is a combination of the general time-dependent POVM from (3) and the specific form of  $p_{\text{click}}(t, \tau_k)$  from (14). In [1] the click distributions for single-photon and two-mode biphoton states are calculated for this specific POVM.

In order to compare the predictions of this model with our experimental results we calculate the click distribution for a coherent wavepacket. Therefore, we insert (14) into the general formula for the time-dependent POVMs (3) and apply it to a continuous multi-mode coherent wave packet

$$|\Psi_{\text{wp}}\rangle = \exp \left( \int_{\mathbb{R}} (\alpha(t) \hat{a}^\dagger(t) - \text{h.c.}) dt \right) |0\rangle. \quad (15)$$

It is assumed that the wave packet has a sufficiently narrow bandwidth and  $[\hat{a}(t), \hat{a}^\dagger(t')] = \delta(t - t')$  holds [40]. Thus,  $|\alpha(t)|^2$  describes the time-dependent photon flux of the probe pulse with mean photon number  $\mu = \int_{\mathbb{R}} |\alpha(t)|^2 dt$ . In appendix A we show that the probability of obtaining a click at time  $t$  for such a wave packet  $|\Psi_{\text{wp}}\rangle$  is given by

$$p_{\text{wp}}(t) = -\frac{\partial}{\partial t} \exp \left( -\eta \int_{-\infty}^t (J * |\alpha|^2)(t') dt' \right), \quad (16)$$

with  $*$  indicating convolution.

This equation has a structure known from Poisson processes. For an inhomogeneous Poisson process with time-dependent rate  $\lambda(t)$ , the probability for the first detection in the interval  $[t, t + dt]$  is  $p_\lambda(t)dt = \lambda(t) \exp \left( -\int_{-\infty}^t \lambda(t') dt' \right) dt$  [44]. Thus,  $p_{\text{wp}}(t)$  resembles the probability density for the time up to the first click of an inhomogeneous Poisson process with the click rate

$$\lambda(t) = \eta(J * |\alpha|^2)(t). \quad (17)$$

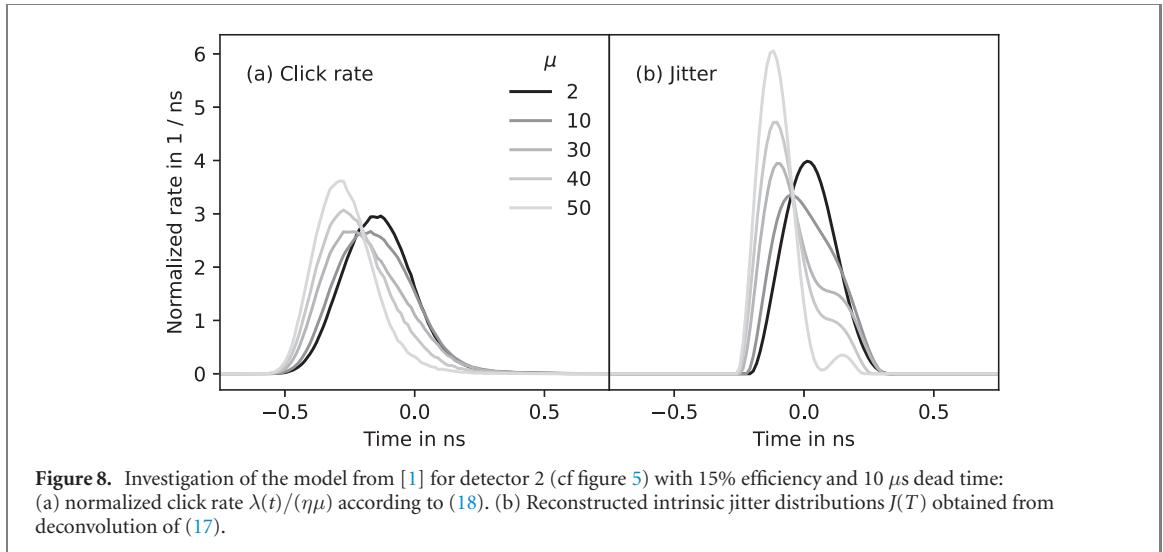
The structure of  $p_{\text{wp}}(t)$  can be understood by recalling that the detection of a coherent state yields Poissonian statistics. Here, the temporal shape of the wave packet is modified by convolution with the intrinsic detector jitter distribution. The Poissonian statistics reflects the fact that all photons are treated independently. As only the first click is registered due to the detector switching into dead time, the resulting distribution is the probability density of the first-click-time of this process.

In order to check whether this model is valid for our detectors, we investigated whether it is possible to reconstruct the jitter distribution  $J(T)$  from the measured click distribution  $p_{\text{wp}}(t)$  according to (16). In order to reconstruct  $J(T)$ , we define the cumulative rate  $\Lambda(t) = \int_{-\infty}^t \lambda(t') dt'$  and write  $\int_{-\infty}^t p_{\text{wp}}(t') dt' = 1 - \exp(-\Lambda(t))$ , where we used  $\Lambda(t \rightarrow -\infty) = 0$ . Solving for  $\lambda(t)$  yields

$$\lambda(t) = \frac{\partial \Lambda(t)}{\partial t} = -\frac{\partial}{\partial t} \ln \left( 1 - \int_{-\infty}^t p_{\text{wp}}(t') dt' \right) = p_{\text{wp}}(t) \left( 1 - \int_{-\infty}^t p_{\text{wp}}(t') dt' \right)^{-1}. \quad (18)$$

The right-hand side can be directly calculated from the measured data, without requiring a calculation of  $\eta$  or  $\mu_{\text{wp}}$ . The click rates  $\lambda(t)$  are shown in figure 8 for different values of  $\mu$ .

We also computed the jitter distribution  $J(T)$  by deconvolution from  $\lambda(t)$  according to (17) for different mean photon numbers  $\mu$ .



**Figure 8.** Investigation of the model from [1] for detector 2 (cf figure 5) with 15% efficiency and 10  $\mu$ s dead time: (a) normalized click rate  $\lambda(t)/(\eta\mu)$  according to (18). (b) Reconstructed intrinsic jitter distributions  $J(T)$  obtained from deconvolution of (17).

**Table 1.** Statistical metrics used for comparing the POVM model and the model with the jitter rate from [1].

Models	Jitter rate	POVM model
Negative log-likelihood	883 203	385
Bayes factor (POVM against jitter rate)	$10^{383403}$	
AIC score	1766 531	8705

For the numerical implementation of the deconvolution with discrete data, it is convenient to introduce the normalized discrete pulse shape  $\mathbf{I}$  with  $I_i = |\alpha(t_i)|^2 / \mu_{wp}$  and the normalized rate  $\boldsymbol{\lambda}$  with  $\lambda_i = \lambda(t_i) / (\mu_{wp}\eta)$ , with  $\sum_i \lambda_i = \sum_i I_i = 1$ . The discrete convolution with the jitter  $\mathbf{J}_i$  can then be written as  $\boldsymbol{\lambda} = (\mathbf{I} * \mathbf{J}) = \mathbf{T}\mathbf{J}$ , where  $\mathbf{T}$  is a Toeplitz matrix constructed from  $\mathbf{I}$ . The deconvolution can be performed by minimizing  $\|\boldsymbol{\lambda} - \mathbf{T}\mathbf{J}\|_2^2$  over  $\mathbf{J}$ . The formulation with the Toeplitz matrix enabled direct implementation of the gradient by using  $\nabla_{\mathbf{J}}\|\boldsymbol{\lambda} - \mathbf{T}\mathbf{J}\|_2^2 = 2\mathbf{T}^T(\mathbf{T}\mathbf{J} - \boldsymbol{\lambda})$ , which facilitated the convergence.

We optimized with the constraint  $J(T) \geq 0$ . In order to obtain smooth results, we penalized large variations in the first derivative of the jitter distribution by introducing the regularization term  $\sum_i (J_{i+1} - J_i)^2$ , multiplied by a weighting coefficient. The objective function is thus very similar to the objective function for the POVM reconstruction in (11). According to the model, the detector should have one distinct jitter distribution that explains the resulting click distributions for all values of  $\mu$  according to (16). Consequently, the normalized rate  $\lambda(t)$  and the deconvolved jitter distribution should be independent of  $\mu$ . However, with increasing  $\mu$ , the rate distribution becomes narrower and shifts towards earlier times, meaning that the model underestimates the previously discussed deformation of the click probabilities in figure 6. The effect is even more pronounced in the jitter distributions and appeared for all seven detectors and for all three detector settings.

For a more rigorous statistical evaluation of the jitter rate model, we compute the log-likelihood of both the jitter rate model and the POVM model. Using the log-likelihood, we determine the Bayes factor between the models. Additionally, the score for the Akaike information criterion (AIC) [45] is determined. The results are shown in table 1 and a detailed discussion of these calculations is described in appendix B.

As the Bayes factor is many (383 401) orders of magnitudes larger than the Bayes factor of  $10^2$  classified as ‘decisive’ in [46], the comparison gives overwhelming evidence that the model with the jitter rate should be rejected. The AIC additionally tests whether the larger number of parameters in the POVM model is sufficient grounds to still consider the jitter rate model. For each model a score is computed, and then the model with the minimum score is selected. In contrast to the Bayes factor, the score depends on the number of parameters, to compensate overfitting due to the higher number of parameters [45]. The jitter rate model has 62 parameters, while the POVM model has  $62 \times 64$  parameters, thus a considerable proportion of the aic-value (cf (B.4)) of the POVM model is constituted by the parameter term. Due to the extremely large difference between the Akaike scores, the jitter rate model can still be rejected with a high degree of certainty.

From these two observations we conclude that the jitter rate model is not sufficient to completely describe the time-dependent click distribution of our detectors. A possible explanation for the deviation can

be found in the detection mechanism by electron avalanches. When the amplitude of an electron avalanche reaches a threshold level, the detector emits an electric pulse to indicate a detection event. The main contribution of the timing jitter in InGaAs single-photon avalanche detectors comes from the distribution of transit times of charge carriers in the absorption region and by the distribution of the avalanche build-up time in the multiplication region [47]. When multiple photons hit the detector, the avalanches can add up. The threshold for a detection event is thus reached faster than expected for photons triggering independent events. It can be expected that the effect will cause a shift of the click distribution to earlier times that is stronger than the prediction of the model. The underlying general reason may thus be the violation of the model assumption that, except for the dead time effect, all photons can be treated independently. An option to include such effects would be to develop complex models of the detector behavior including more details of the detection mechanism such as in [48, 49]. However, this approach is contrary to the tomographic approach, which is to introduce as few as possible general assumptions about the detector.

## 8. Conclusion

We performed tomographic measurements on avalanche single-photon detectors that enabled the reconstruction of both time-independent and time-dependent POVMs. The time-independent tomography results were in agreement with a simple model of an ideal detector. By this method, we deduced and compared the detection efficiencies of seven detectors. For the time-dependent POVM reconstruction we derived an estimation for the weighting coefficient which adapts the regularization term based on the amount of available data in different time bins. Benchmarking the new method showed a superior performance in comparison with reconstructions based on a fixed coefficient. We then reconstructed time-dependent POVMs by using the adaptive regularization. Finally, we investigated whether the model for time-dependent POVMs proposed by Gouzien *et al* [1] can explain the measured POVMs and showed that the model is not able to explain the performance of our InGaAs single-photon avalanche detectors in a satisfactory way. This example demonstrates the strength of detector tomography in comparison with less flexible modeling approaches. Thus, measuring the time-dependent POVM of a detector with quantum tomographic methods can reveal information about the detector's interior that may not easily be included by *a priori* model assumptions.

## Acknowledgment

This research has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—SFB 1119—236615297.

## Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

## Appendix A. Derivation of the click probability $p_{\text{wp}}(t)$

Here, we calculate the click probability  $p_{\text{wp}}(t)$  for Gouziens's model [1] applied to a coherent wavepacket by inserting (14) into (3)

$$p_{\text{wp}}(t) = \langle \Psi_{\text{wp}} | \hat{\pi}_{\text{click}}(t) | \Psi_{\text{wp}} \rangle = \mathcal{T} \sum_{k=0}^{\infty} \int_{\mathbb{R}^k} p_{\text{click}}(t, \boldsymbol{\tau}_k) \langle \Psi_{\text{wp}} | \hat{P}_k(\boldsymbol{\tau}_k) | \Psi_{\text{wp}} \rangle d\boldsymbol{\tau}_k. \quad (\text{A.1})$$

The integral is subject to time ordering  $\tau_1 < \tau_2 < \dots < \tau_k$ . However, both  $p_{\text{click}}(t, \boldsymbol{\tau}_k)$  from (14) and  $\langle \Psi_{\text{wp}} | \hat{P}_k(\boldsymbol{\tau}_k) | \Psi_{\text{wp}} \rangle$  from (4) are symmetric under permutation of  $\tau_1 \dots \tau_k$ . The time ordering can thus be expressed by extending the integration range to the complete real line for each  $\tau$  and simultaneously dividing by the number of  $k!$  permutations

$$p_{\text{wp}}(t) = \sum_{k=0}^{\infty} \frac{1}{k!} \int_{\mathbb{R}^k} p_{\text{click}}(t, \boldsymbol{\tau}_k) \langle \Psi_{\text{wp}} | \hat{P}_k(\boldsymbol{\tau}_k) | \Psi_{\text{wp}} \rangle d\boldsymbol{\tau}_k. \quad (\text{A.2})$$

From the definition of the multimode coherent state

$$|\Psi_{wp}\rangle = \exp\left(-\frac{1}{2}\int_{\mathbb{R}} |\alpha(t)|^2 dt\right) \sum_{n=0}^{\infty} \frac{1}{n!} \left(\int_{\mathbb{R}} \alpha(t) \hat{a}^\dagger(t) dt\right)^n |0\rangle \quad (\text{A.3})$$

we can calculate  $\langle \Psi_{wp} | \hat{P}_k(\tau_k) | \Psi_{wp} \rangle = e^{-\mu_{wp}} \prod_{j=1}^k |\alpha(\tau_j)|^2$ . Inserting this expression into the expression for  $p_{wp}(t)$  yields the result

$$\begin{aligned} p_{wp}(t) &= e^{-\mu_{wp}} \sum_{k=0}^{\infty} \int_{\mathbb{R}^k} \frac{1}{k!} p_{\text{click}}(t, \tau_k) \prod_{j=1}^k |\alpha(\tau_j)|^2 d\tau_k \\ &= e^{-\mu_{wp}} \sum_{k=0}^{\infty} \int_{\mathbb{R}^k} \frac{1}{k!} \left( \sum_{j=1}^k \eta |\alpha(\tau_j)|^2 J(t - \tau_j) \prod_{l \neq j} |\alpha(\tau_l)|^2 \left(1 - \eta \int_{\tau_l}^t J(t' - \tau_l) dt'\right) \right) d\tau_k \\ &= -e^{-\mu_{wp}} \frac{\partial}{\partial t} \sum_{k=0}^{\infty} \frac{1}{k!} \left( \int_{\mathbb{R}} |\alpha(\tau)|^2 \left(1 - \eta \int_{\tau}^t J(t' - \tau) dt'\right) d\tau \right)^k \\ &= -\frac{\partial}{\partial t} \exp\left(-\eta \int_{-\infty}^t (J * |\alpha|^2)(t') dt'\right). \end{aligned} \quad (\text{A.4})$$

## Appendix B. Bayes factor, Akaike information criterion and model comparison

When comparing different models of the same phenomena, it is often interesting to have a criterion for selecting the better model from a set of models. Some criteria account for the number of parameters, thus privileging ‘simpler’ models in the selection. This helps avoid overfitting. In the following, two criteria are discussed: the Bayes factor and the AIC.

### B.1. Bayes Factor and Akaike information criterion

The Bayes factor can be used to quantitatively compare two models with respect to given data. Let  $P(\text{model}_1)$  and  $P(\text{model}_2)$  the prior probabilities for both models and  $P(\text{data}|\text{model}_1)$  and  $P(\text{data}|\text{model}_2)$  the probabilities to obtain the data when a model is given. The posterior probabilities for model  $i$  can be then calculated from (8) for a Bayesian update with the measured data:

$$P(\text{model}_i|\text{data}) = \frac{P(\text{data}|\text{model}_i)P(\text{model}_i)}{P(\text{data})}. \quad (\text{B.1})$$

The ratio of the posterior probabilities is given by

$$\frac{P(\text{model}_1|\text{data})}{P(\text{model}_2|\text{data})} = K \frac{P(\text{model}_1)}{P(\text{model}_2)} \quad (\text{B.2})$$

with the Bayes factor

$$K = \frac{P(\text{data}|\text{model}_1)}{P(\text{data}|\text{model}_2)} = \exp(l(\text{model}_2) - l(\text{model}_1)) \quad (\text{B.3})$$

and the negative log-likelihood function  $l$  which incorporates the data. Thus the Bayes factor directly quantifies by which factor the ratio of posterior probabilities differs from the ratio of prior probabilities after a Bayesian update using the data. A criterion for privileging one model over the other decisively is a Bayes factor over 100 [46].

For the AIC each model $_i$  with  $k_i$  parameters is assigned a score value of [45]

$$\text{aic}(\text{model}_i) = 2k_i + 2l(\text{model}_i) \quad (\text{B.4})$$

with the negative log-likelihood function  $l(\text{model})$ . It should be noted that the collection may encompass entirely different families of models with different parameter spaces [45]. The model with the lowest aic-value is then selected as optimal. Obviously  $k$  is constant for a family of models parametrized by the same parameters. Inside such a space the AIC selects the model with the lowest negative log-likelihood. Only when comparing models with different numbers of parameters, the AIC differs in any way from the max-likelihood selection. The first term counting the number of parameters penalizes models with many parameters and therefore helps avoid overfitting [45, 50].

## B.2. Model comparison

In this section we compare the model obtained by our tomographic algorithm and the timing jitter model analyzed in section 7. We assume that all measurement errors are normally distributed. For a single Gaussian with mean  $m$  and standard deviation  $\sigma$ , the negative log-likelihood in the variable  $x$  is

$$l_{\text{Gauss}}(m, \sigma) = -\ln(L_{\text{Gauss}}(m, \sigma)) = \frac{1}{2\sigma^2}(x - m)^2 + \frac{1}{2} \log 2\pi + \log \sigma. \quad (\text{B.5})$$

Note here, that the last two terms do not depend on  $x$  and remain constant as long as the same Gaussian is considered.

We aim to compare the model entailed by the reconstructed POVM, with the optimal version of the jitter rate model. We thus optimize the least square distance between the measured click probability  $p_{\text{wp}}$  and the jitter rate model click probability  $p_{\text{JR}}$  with one single click rate  $\lambda$ , so that the click probability according to this model is given by (16):

$$p_{\text{JR}}(t) = -\frac{\partial}{\partial t} \exp\left(-\eta\mu \int_{-\infty}^t \lambda(t') dt'\right) = \eta\mu\lambda(t) \exp\left(-\eta\mu \int_{-\infty}^t \lambda(t') dt'\right). \quad (\text{B.6})$$

According to (17), the click rate  $\lambda$  would be restricted to a convolution of the pulse shape with the jitter distribution. By optimizing for a common click rate  $\lambda$  explaining the probabilities for all values of  $\mu$  simultaneously instead of optimizing for a common jitter rate, we allow a broader class of solutions including those that cannot be represented by a convolution. Therefore, the model comparison is independent of the measured probe pulse shape.

We discretize  $\lambda(t)$  and  $p_{\text{JR}}(t)$  into the time bins as defined in section 2 such that  $\sum_i \lambda_i = 1$ . Each time bin is assigned a standard deviation (and thus a measurement error) by taking the maximum standard deviation as in equation (10). These measurement errors determine the log-likelihood function as

$$l(\lambda) = \sum_i \frac{1}{2\sigma_i^2} (f_i - p_{\text{JR},i})^2 + \sum_i \left( \frac{1}{2} \log 2\pi + \log \sigma_i \right), \quad (\text{B.7})$$

with measured click probabilities  $f_i$  as introduced in section 2. The second sum is independent of the parameters, and thus left out for the final objective function

$$U(\lambda_i) = \sum_i \frac{1}{2\sigma_i^2} (f_i - p_{\text{JR},i})^2. \quad (\text{B.8})$$

The likelihood used for the POVM model itself is an adapted version of (B.7):

$$p_{\text{POVM},i} = \sum_{k=0}^{k_{\max}} \Theta_{k,i} \frac{\mu^k}{k!} e^{-\mu} \quad (\text{B.9})$$

$$l_{\text{POVM},i}(\Theta_{n,i}) = \sum_i \frac{1}{2\sigma_i^2} (f_i - p_{\text{POVM},i})^2 - \sum_i \left( \frac{1}{2} \log 2\pi + \log \sigma_i \right). \quad (\text{B.10})$$

We then compare both, the optimal jitter rate model from (B.7) and the POVM model from section 6, using the Bayes factor and the AIC calculated from the negative log-likelihood functions in (B.7) and (B.10). The results are summarized in table 1.

## ORCID iDs

E Fitzke  <https://orcid.org/0000-0002-1046-6176>

Th Haase  <https://orcid.org/0000-0001-9516-6094>

G Alber  <https://orcid.org/0000-0003-4676-0024>

Th Walther  <https://orcid.org/0000-0001-8114-1785>

## References

- [1] Gouzien E, Fedrici B, Zavatta A, Tanzilli S and D'Auria V 2018 *Phys. Rev. A* **98** 013833
- [2] Zhong H-S *et al* 2020 *Science* **370** 1460–3
- [3] D'Auria V, Fornaro S, Porzio A, Solimeno S, Olivares S and Paris M G A 2009 *Phys. Rev. Lett.* **102** 020502
- [4] Sauer A and Alber G 2020 *Cryptography* **4** 2
- [5] Bennett C H and Brassard G 2014 *Theor. Comput. Sci.* **560** 7–11
- [6] Gisin N, Ribordy G, Tittel W and Zbinden H 2002 *Rev. Mod. Phys.* **74** 145–95

- [7] Scarani V, Bechmann-Pasquinucci H, Cerf N J, Dušek M, Lütkenhaus N and Peev M 2009 *Rev. Mod. Phys.* **81** 1301–50
- [8] Lydersen L, Wiechers C, Wittmann C, Elser D, Skaar J and Makarov V 2010 *Nat. Photon.* **4** 686–9
- [9] Jogenfors J, Elhassan A M, Ahrens J, Bourennane M and Larsson J-A 2015 *Sci. Adv.* **1** e1500793
- [10] Xu F, Ma X, Zhang Q, Lo H K and Pan J W 2020 *Rev. Mod. Phys.* **92** 025002
- [11] Lo H-K, Curty M and Qi B 2012 *Phys. Rev. Lett.* **108** 130503
- [12] Wang W, Tamaki K and Curty M 2021 *Sci. Rep.* **11** 1678
- [13] Luis A and Sánchez-Soto L L 1999 *Phys. Rev. Lett.* **83** 3573–6
- [14] Feito A, Lundeen J S, Coldenstrodt-Ronge H, Eisert J, Plenio M B and Walmsley I A 2009 *New J. Phys.* **11** 093038
- [15] Lundeen J S, Feito A, Coldenstrodt-Ronge H, Pregnell K L, Silberhorn C, Ralph T C, Eisert J, Plenio M B and Walmsley I A 2009 *Nat. Phys.* **5** 27–30
- [16] Chen Y, Farahzad M, Yoo S and Wei T C 2019 *Phys. Rev. A* **100** 052315
- [17] Coldenstrodt-Ronge H B et al 2009 *J. Mod. Opt.* **56** 432–41
- [18] van Enk S J 2017 *J. Phys. Commun.* **1** 045001
- [19] D'Auria V, Lee N, Amri T, Fabre C and Laurat J 2011 *Phys. Rev. Lett.* **107** 050504
- [20] Akhlaghi M K, Majedi A H and Lundeen J S 2011 *Opt. Express* **19** 021305
- [21] Brida G et al 2012 *New J. Phys.* **14** 085001
- [22] Natarajan C M, Zhang L, Coldenstrodt-Ronge H, Donati G, Dorenbos S N, Zwiller V, Walmsley I A and Hadfield R H 2013 *Opt. Express* **21** 893–902
- [23] Renema J J, Frucci G, Zhou Z, Mattioli F, Gaggero A, Leoni R, de Dood M J A, Fiore A and van Exter M P 2012 *Opt. Express* **20** 2806
- [24] Renema J J et al 2014 *Phys. Rev. Lett.* **112** 117604
- [25] Endo M, Sonoyama T, Matsuyama M, Okamoto F, Miki S, Yabuno M, China F, Terai H and Furusawa A 2021 *Opt. Express* **29** 11728–38
- [26] Zhang L, Coldenstrodt-Ronge H B, Datta A, Puentes G, Lundeen J S, Jin X-M, Smith B J, Plenio M B and Walmsley I A 2012 *Nat. Photon.* **6** 364–8
- [27] Grandi S, Zavatta A, Bellini M and Paris M G A 2017 *New J. Phys.* **19** 053015
- [28] Zhang L, Datta A, Coldenstrodt-Ronge H B, Jin X-M, Eisert J, Plenio M B and Walmsley I A 2012 *New J. Phys.* **14** 115005
- [29] Yang T H, Vértesi T, Bancal J D, Scarani V and Navascués M 2014 *Phys. Rev. Lett.* **113** 040401
- [30] Mogilevtsev D, Řeháček J and Hradil Z 2009 *Phys. Rev. A* **79** 020101
- [31] Mohammadi M and Brańczyk A M 2014 *Phys. Rev. A* **89** 012113
- [32] Řeháček J, Mogilevtsev D and Hradil Z 2010 *Phys. Rev. Lett.* **105** 010402
- [33] Mogilevtsev D, Ignatenko A, Maloshtan A, Stoklasa B, Rehacek J and Hradil Z 2013 *New J. Phys.* **15** 025038
- [34] Cooper M, Karpiński M and Smith B J 2014 *Nat. Commun.* **5** 4332
- [35] Harder G, Silberhorn C, Rehacek J, Hradil Z, Motka L, Stoklasa B and Sánchez-Soto L L 2014 *Phys. Rev. A* **90** 042105
- [36] Altorio M, Genoni M G, Somma F and Barbieri M 2016 *Phys. Rev. Lett.* **116** 100802
- [37] Czerwinski A, Sedziak-Kacprowicz K and Kolenderski P 2021 *Phys. Rev. A* **103** 042402
- [38] Rohde P P, Mauerer W and Silberhorn C 2007 *New J. Phys.* **9** 91
- [39] Nielsen M and Chuang I 2010 *Quantum Computation and Quantum Information* 10th anniversary edn (Cambridge: Cambridge University Press)
- [40] Loudon R 2000 *The Quantum Theory of Light* 3rd edn (Oxford: Oxford University Press)
- [41] Brecht B, Reddy D V, Silberhorn C and Raymer M G 2015 *Phys. Rev. X* **5** 041017
- [42] Pereyra M 2017 *SIAM J. Imaging Sci.* **10** 285–302
- [43] Fiurášek J 2001 *Phys. Rev. A* **64** 024102
- [44] Streit R L 2010 *The Poisson Point Process* (Berlin: Springer) pp 11–55
- [45] Akaike H 1974 *IEEE Trans. Autom. Control* **19** 716–23
- [46] Jeffreys H 1983 *The Theory of Probability (International Series of Monographs on Physics)* 3rd edn (Oxford: Oxford University Press)
- [47] Amri E, Boso G, Korzh B and Zbinden H 2016 *Opt. Lett.* **41** 5728–31
- [48] Tan S L, Ong D S and Yow H K 2007 *J. Appl. Phys.* **102** 044506
- [49] Petticrew J D, Dimler S J, Zhou X, Morrison A P, Tan C H and Ng J S 2018 *IEEE J. Sel. Top. Quantum Electron.* **24** 1–6
- [50] deLeeuw J 1992 Introduction to akaike (1973) information theory and an extension of the maximum likelihood principle *Breakthroughs in Statistics* (Berlin: Springer) pp 599–609