Implementing Non-Projective Measurements via Linear Optics: an Approach Based on Optimal Quantum State Discrimination

Peter van Loock[1], Kae Nemoto[1], William J. Munro, Philippe Raynal[2], Norbert Lütkenhaus[2]
Trusted Systems Laboratory
HP Laboratories Bristol
HPL-2005-210
November 28, 2005*

generalised measurements, linear optics

We discuss the problem of implementing generalized measurements (POVMs) with linear optics, either based upon a static linear array or including conditional dynamics. In our approach, a given POVM shall be identified as a solution to an optimization problem for a chosen cost function. We formulate a *general principle*: the implementation is only possible if a linear-optics circuit exists for which the quantum mechanical optimum (minimum) is still attainable after dephasing the corresponding quantum states. The general principle enables us, for instance, to derive a set of necessary conditions for the linear-optics implementation of the POVM that realizes the quantum mechanically optimal unambiguous discrimination of two pure *nonorthogonal* states. This extends our previous results on projection measurements and the exact discrimination of orthogonal states.

* Internal Accession Date Only
[1]National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
[2]Quantum Information Theory Group, Institute of Theoretical Physics I, Institute of Optics, Information and Photonics (Max-Planck Forschungsgruppe), Universität Erlangen-Nürnberg, Staudtstr.7, D-91058 Erlangen, Germany
Approved for External Publication

Implementing Non-Projective Measurements via Linear Optics: an Approach Based on Optimal Quantum State Discrimination

Peter van Loock[1], Kae Nemoto[1], William J. Munro[1,2], Philippe Raynal[3], and Norbert Lütkenhaus[3]
[1]*National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan*
[2]*Hewlett-Packard Laboratories, Filton Road, Stoke Gifford, Bristol BS34 8QZ, United Kingdom*
[3]*Quantum Information Theory Group, Institute of Theoretical Physics I, Institute of Optics, Information and Photonics (Max-Planck Forschungsgruppe), Universität Erlangen-Nürnberg, Staudtstr.7, D-91058 Erlangen, Germany*

We discuss the problem of implementing generalized measurements (POVMs) with linear optics, either based upon a static linear array or including conditional dynamics. In our approach, a given POVM shall be identified as a solution to an optimization problem for a chosen cost function. We formulate a *general principle*: the implementation is only possible if a linear-optics circuit exists for which the quantum mechanical optimum (minimum) is still attainable after dephasing the corresponding quantum states. The general principle enables us, for instance, to derive a set of necessary conditions for the linear-optics implementation of the POVM that realizes the quantum mechanically optimal unambiguous discrimination of two pure *nonorthogonal* states. This extends our previous results on projection measurements and the exact discrimination of orthogonal states.

PACS numbers: 03.67.Hk, 42.25.Hz, 42.50.Dv

I. INTRODUCTION

The implementation of *positive operator-valued measures* (POVMs) for photonic quantum state signals is an essential task in many quantum information protocols. In general, in order to implement such measurements, a nonlinear interaction of the signal states, described by a Hamiltonian at least cubic in the optical mode operators [1], is needed. With current technologies, however, these nonlinear effects are hard to obtain on the level of single photons. Apart from hybrid schemes based on weak nonlinearities and strong coherent probe pulses [2], an alternative approach for inducing a nonlinear element is to exploit the effective nonlinearity associated with a measurement. In particular, for photonic-qubit states, universal quantum gates and hence any POVM can be realized deterministically or asymptotically (near-deterministically), using linear optics, photon counting, entangled auxiliary photon states, and conditional dynamics (feedforward) [3–6]. Moreover, cheaper resources may suffice for the implementation of non-deterministic gates and POVMs, using feedforward [3, 7] or a static array of linear optics [3, 8–13]. Here we will focus on the implementation of POVMs using either static linear optics or feedforward and, in particular, photon counting. Although there are some specific results on this issue [14], a general and practical solution to the problem as to whether a given POVM can be implemented by linear optics is not known. Only for the special class of projective measurements, a set of simple criteria has been derived [15].

In the special case of a *projection measurement*, the "signal states" to be distinguished (i.e., the basis that spans the space to be projected on) are orthogonal. In this case, quantum mechanically, an exact discrimination with unit probability for a conclusive result is possible. However, if the implementation of the projection measurement is restricted to a limited class of transformations such as passive linear optics or Gaussian transformations, unit probability might be unattainable [15, 16]. The prime example to which such a no-go statement applies is the Bell measurement for polarization-encoded photonic qubit states [15–18]. Of course, such a no-go statement for *exact* state discrimination does not rule out the possibility for near-deterministic or non-deterministic implementations. For example, the simplest approximation to the single-photon qubit Bell measurement only requires a symmetric beam splitter and photon counting. This scheme achieves a success probability of one half, thus attaining the upper bound when using linear optics and photon counting, but neither auxiliary photons nor feedforward [19].

A hierarchy of simple criteria for the exact discrimination of orthogonal states can be derived via a *dephasing approach* [15]. The idea of this approach is to simulate the actual detection, for instance, in the photon number basis through a dephasing of the linearly transformed states, turning them into mixtures diagonal in the Fock basis. Any term in these mixtures represents a possible detection pattern for a given input state and a given linear-optics circuit. The requirement for an exact discrimination of the signal states is then that the overlap of the dephased density operators vanishes, corresponding to the non-existence of any coinciding patterns. Expressing the overlap in terms of the fidelity, this means that the fidelity of the orthogonal states must remain zero after the linear transformation and the dephasing operation have been applied to the states.

In order to extend the analysis of projection measurements [15] to *generalized measurements*, the first obvious approach is to consider von Neumann measurements in a larger Hilbert space. Suitably chosen, these are then equivalent to the POVM in the smaller signal space. In fact, any POVM can be expressed in such a way via the

Naimark extension. For signal states having only *one photon*, already the Naimark extension approach reveals that *any* POVM can be implemented with linear optics. A demonstration of this can be found in App. A. In general, for signal states with arbitrarily many photons, to decide whether a linear-optics implementation of a given POVM is possible is a non-trivial problem.

Here, in order to address this question, we refer to a *fundamental principle*, independent of the Naimark extension. In order to apply this principle, first, the POVM shall be identified as a solution of an optimization problem for some cost function. In terms of this cost function, the principle then states that the implementation is only possible if a linear-optics circuit exists for which the quantum mechanical optimum (minimum) is still attainable *after dephasing* the corresponding quantum states. Whether linear optics or more general linear transformations including multi-mode squeezing are sufficient to implement the corresponding POVM depends on the ability of these tools to obey the above general rule. Applying this rule to the fidelity of two nonorthogonal states will enable us to derive a set of necessary conditions for the implementation of the quantum mechanically optimal unambiguous state discimination (USD), extending our analysis of discriminating orthogonal states [15]. The USD of non-orthogonal states is a simple example for a non-projective POVM, where some measurement results are inconclusive, but the remaining results correctly identify the signal state.

The plan of the paper is as follows. First, in Sec. II, we are going to explain how the effect of the detection behind a linear-optics circuit can be described via dephasing. This enables us to present the main result of the paper, a general principle for the implementation of POVMs with linear optics. In Sec. III, we briefly review how the known criteria for linear-optics projection measurements follow from this general principle as a simple special case. Finally, we turn to the implementation of non-projective POVMs in Sec. IV, where our main focus will be on the unambiguous discrimination of two pure non-orthogonal states.

II. THE DEPHASING APPROACH TO POVMS

Given a general non-projective POVM, via the Naimark extension approach it is pretty hard to decide whether the POVM can be implemented with linear optics. Here we propose an alternative strategy independent of the Naimark extension, based upon a dephasing approach. The dephasing effect will be used to mimic the projection of the individual modes onto the detection basis. Let us first introduce the dephasing formalism.

The dephasing basis is determined by the detection mechanism of the implementation. This might be either the discrete photon number basis (photon counting) or the continuous quadrature eigenstate basis (homodyne detection). In the following, we will use the Fock basis as the dephasing basis. This basis can be easily substituted by other appropriate bases [15].

If the signal states $\hat{\rho}$ are linearly transformed into the states $\hat{\rho}_H$, and the photon number of the modes will be detected, the corresponding dephasing effect can be described as

$$\hat{\rho}_H \rightarrow \hat{\rho}'_H = \frac{1}{(2\pi)^N} \int d\phi^N e^{-i\vec{a}^\dagger D \vec{a}} \hat{\rho}_H e^{i\vec{a}^\dagger D \vec{a}} \ . \qquad (1)$$

Here, we used $d\phi^N \equiv d\phi_1 d\phi_2 ... d\phi_N$, the diagonal $N \times N$ matrix D, $(D)_{ij} = \delta_{ij}\phi_i$, and the vectors $\vec{a} = (\hat{a}_1, \hat{a}_2, ..., \hat{a}_N)^T$ and $\vec{a}^\dagger = (\hat{a}_1^\dagger, \hat{a}_2^\dagger, ..., \hat{a}_N^\dagger)$, representing the annihilation and creation operators of all the electromagnetic modes involved.

The effect of the dephasing is that it turns the linearly transformed states into a Fock-diagonal density matrix. This mixture contains all possible photon number patterns for a given input state and a given linear-optics circuit. The weights of the different terms in this mixture are determined by the probabilities for obtaining the corresponding pattern via photon detection. Thus, the dephasing formalism is an equivalent description for the effect of the detection after the linear-optics transformation (see Fig. 1). An example for a pure signal state $\hat{\rho} = |\chi\rangle\langle\chi|$, and hence a pure transformed state $\hat{\rho}_H \equiv |\chi_H\rangle\langle\chi_H|$, would be $|\chi_H\rangle = \alpha|110\rangle + \beta|101\rangle + \gamma|002\rangle$. In this case, the dephased state becomes $\hat{\rho}'_H = |\alpha|^2|110\rangle\langle 110| + |\beta|^2|101\rangle\langle 101| + |\gamma|^2|002\rangle\langle 002|$, corresponding to the possible detection patterns 110, 101, and 002.

The advantage of the dephasing formalism is that the effect of the detection is described on the level of the state transformations and the final states become as classical as they can get. These close-to-classical states can then be analyzed with respect to a given quantum information task. One may also consider only partially dephased states which are Fock-diagonal only with respect to the dephased modes. Partial dephasing mimics those protocols where only a subset of the modes is detected and a subsequent linear-optics transformation is applied to the remaining modes conditioned upon the measurement outcomes (conditional dynamics).

Using the dephasing formalism, we now propose the following strategy in order to decide whether a given POVM can be implemented via linear optics. First, the POVM shall be identified as a unique solution to an optimization problem. For the cost function to be optimized, we then refer to a general principle: the implementation is only possible if a linear-optics circuit exists for which the quantum mechanical optimum (minimum) is still attainable *after dephasing* the corresponding quantum states. Thus, optimizing the cost function for the dephased states must yield the same minimum as for the original signal states. The linear-optics circuit must be chosen such that

$$C^{\text{optimal}}_{\text{linear optics, dephasing}} \stackrel{!}{=} C^{\text{optimal}}_{\text{quantum mechanics}} \ , \qquad (2)$$

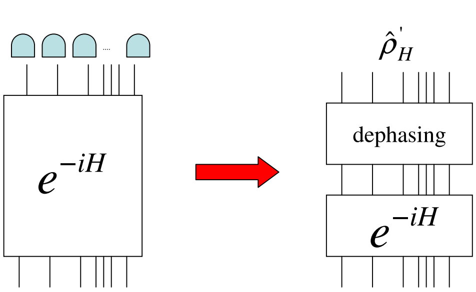

FIG. 1: An equivalent description of the detection mechanism after a unitary state transformation via dephasing. In the case of photon counting, the dephased density matrix is a mixture of all possible photon number patterns for a given input state and a given state transformation. Here, we are mainly concerned about linear-optics transformations.

where the symbol C denotes the corresponding cost functions [20]. This general criterion is a *necessary and sufficient* condition for the possibility of implementing the corresponding POVM. The sufficiency here is due to the close-to-classical character of the totally dephased output states which are directly linked to the click patterns of the implementation. In the case of only partially dephased states corresponding to a conditional-dynamics protocol, the statement in Eq. (2) is no longer sufficient but only necessary for the implementability of the POVM. Similarly, if the POVM is not a unique solution to the optimization problem, the condition in Eq. (2) is only necessary.

In general, it will be highly non-trivial to find the quantum mechanical optimum of the corresponding cost functions. In many cases, neither for pure states, as typically given before the dephasing, nor, in particular, for mixed states, as obtained after dephasing, a closed expression for the optimum exists.

However, for instance, for the non-projective POVM that is an optimal solution to the unambiguous discrimination of two pure non-orthogonal states, the corresponding cost function is the failure probability and its optimum/minimum before dephasing is simply the overlap (fidelity) of the states. For the mixed states after dephasing, in this case, at least a lower bound for the cost function can also be given in terms of the fidelity of the states. It is then possible to derive a relatively simple set of necessary conditions for the implementability of the corresponding POVM. Later we will discuss this example in detail. However, before applying the general principle in Eq. (2) to non-projective POVMs, let us first review how the known criteria for projection measurements follow from this principle as a simple special case.

III. PROJECTION MEASUREMENTS

Following the approach of the preceding section, given a projection measurement, we shall consider this measurement as the optimal solution to the discrimination of orthogonal states. A suitable cost function for an error-free state discrimination is the failure probability, i.e., the probability for obtaining an inconclusive result. Now the optimal strategy in order to discriminate states within an orthogonal set is to do a projection measurement on the space spanned by these orthogonal states. This strategy will always lead to a conclusive error-free result. Since this implies zero cost for discriminating orthogonal states, $C^{\text{optimal}}_{\text{quantum mechanics}} = 0$, a linear-optics implementation of exact state discrimination means that $C^{\text{optimal}}_{\text{linear optics, dephasing}} \stackrel{!}{=} 0$ according to Eq. (2).

In order to discriminate any two pure orthogonal states from the projection measurement basis, the quantum mechanically optimal/minimal failure probability is given by the overlap of the states to be discriminated. Expressing the overlap in terms of the fidelity, $F(\hat{\rho}_1, \hat{\rho}_2) \equiv \left(\text{Tr}\sqrt{\sqrt{\hat{\rho}_1}\hat{\rho}_2\sqrt{\hat{\rho}_1}}\right)^2$, for two pure orthogonal signal states, $+$ and $-$, of course, we have $F(\hat{\rho}_+, \hat{\rho}_-) = 0$. Hence after dephasing, the minimal failure probability must not become nonzero, in order to satisfy our principle in Eq. (2). Since in any mixed-state discrimination scheme, the squared failure probability is lower bounded by the fidelity of the mixed states [21], the condition for implementing the exact state discrimination becomes $F(\hat{\rho}'_{+,H}, \hat{\rho}'_{-,H}) \stackrel{!}{=} 0$. Thus, we have $\text{Tr}(\hat{\rho}'_{+,H}\,\hat{\rho}'_{-,H}) = 0$, since always $0 \leq \text{Tr}(\hat{\rho}_1\,\hat{\rho}_2) \leq F(\hat{\rho}_1, \hat{\rho}_2)$. Using the dephasing integral from Eq. (1), one can then derive a hierarchy of simple conditions for the exact discrimination of two or even more states [15]. These conditions are necessary and sufficient for the possibility of exactly implementing the corresponding projection measurement.

For a two-dimensional projection measurement, corresponding to the discrimination of two orthogonal states $|\chi_+\rangle$ and $|\chi_-\rangle$, the necessary and sufficient conditions for an exact implementation via linear optics and, for instance, photon counting, are given by [15],

$$\begin{aligned} \langle\chi_+|\hat{c}_j^\dagger\hat{c}_j|\chi_-\rangle &= 0\,, \quad \forall j\,, \\ \langle\chi_+|\hat{c}_j^\dagger\hat{c}_{j'}^\dagger\hat{c}_j\hat{c}_{j'}|\chi_-\rangle &= 0\,, \quad \forall j, j'\,, \\ \langle\chi_+|\hat{c}_j^\dagger\hat{c}_{j'}^\dagger\hat{c}_{j''}^\dagger\hat{c}_j\hat{c}_{j'}\hat{c}_{j''}|\chi_-\rangle &= 0\,, \quad \forall j, j', j''\,, \\ \vdots \quad &= \quad \vdots \end{aligned} \tag{3}$$

Here, the mode operators $\hat{c}_j = \hat{U}^\dagger\hat{a}_j\hat{U} = \sum_i U_{ji}\hat{a}_i$ are those corresponding to the output modes of the linear-optics circuit. In the remainder of this section, we will add some new and useful observations to the results of Ref. [15] on projection measurements.

Assuming signal states with a *fixed number of photons* (say N photons), there is an obvious interpretation for the highest order conditions (i.e., the Nth order conditions), because for these we have

$$\langle\chi_+|\hat{c}_j^\dagger\hat{c}_{j'}^\dagger\hat{c}_{j''}^\dagger\cdots\hat{c}_j\hat{c}_{j'}\hat{c}_{j''}\cdots|\chi_-\rangle =$$
$$\langle\chi_{+,H}|\hat{a}_j^\dagger\hat{a}_{j'}^\dagger\hat{a}_{j''}^\dagger\cdots\hat{a}_j\hat{a}_{j'}\hat{a}_{j''}\cdots|\chi_{-,H}\rangle \propto$$
$$\Psi^*(j,j',j'',...|+)\times\Psi(j,j',j'',...|-)\,, \qquad (4)$$

where $\Psi(j,j',j'',...|\pm)$ is the probability amplitude for detecting a photon in mode j and another photon in mode j', etc., when the input was the $+$ or $-$ state. Thus $\Psi(j,j',j'',...|\pm)$ represents the probability amplitude for any possible pattern to be detected at the output.

Now it becomes clear why any highest order must vanish for exact state discrimination. Only those patterns that do not occur at all and the successful patterns that can be triggered only by one of the two states lead to $\Psi^*(j,j',j'',...|+)\times\Psi(j,j',j'',...|-)=0$. In contrast, for any failure pattern, the product of the probability amplitudes becomes nonzero, $\Psi^*(j,j',j'',...|+)\times\Psi(j,j',j'',...|-)\neq 0$. As a result, the highest order conditions *alone* are *necessary and sufficient* for exact state discrimination. Fulfilling all the highest order conditions then implies that all lower order conditions are satisfied as well. However, note that the converse does not hold. The lower order conditions are only necessary, but not sufficient for exact state discrimination. Thus, if the lower order conditions are satisfied, the highest order conditions may well be violated. As the lower orders are easier to calculate than the higher orders, one would normally start by computing the lowest orders. In order to rule out the possibility of exact state discrimination, it is then sufficient to find a violation of any lower order condition ("no-go" statement). However, for a "go" statement, the lower orders alone do not suffice. In this case, for verifying that exact state discrimination is possible, one has to either calculate the higher orders as well or directly check a possible solution inferred from the lower orders. All these observations also indicate that for unambiguously discriminating two nonorthogonal signal states of fixed photon number, there must be at least one highest order condition that is violated (corresponding to the existence of at least one failure pattern and hence a nonzero failure probability).

Let us now consider non-projective POVMs including the optimal unambiguous discrimination of nonorthogonal states via linear optics.

IV. NON-PROJECTIVE POVMS

Our goal is now, similar to the criteria for projection measurements, to derive relatively simple conditions for the implementation of a given non-projective POVM. Our approach shall be based upon the general principle expressed in Eq. (2).

We have seen already that there are state estimation problems with trivial optimal POVM solutions. For instance, discriminating orthogonal states optimally means to perform the corresponding projection measurement. A very natural way to optimally discriminate quantum states drawn from a set of linearly independent states is to perform a POVM that minimizes the probabiltity of identifying the wrong states. This so-called minimum error discrimination (MED) can always be described by a projection measurement onto a suitably chosen basis in the signal Hilbert space [22]. Therefore, in order to decide whether for a given set of quantum states MED can be implemented via linear optics, we can also directly apply the conditions for projection measurements. An example for this is the MED of two symmetric coherent states $|\pm\alpha\rangle$ which cannot be accomplished via non-asymptotic linear-optics schemes [23].

Another trivial example is the optimal estimation of an unknown qubit state. In this case, the optimal mean fidelity $\bar{F}^{\rm optimal}_{\rm quantum\ mechanics}=2/3$ [20] can be attained by randomly choosing an arbitrary qubit basis, measuring in this basis, and estimating the state via the basis vector that corresponds to the outcome of the measurement. Thus, trivially, the optimal estimation of a completely unknown qubit state $\alpha|\bar{0}\rangle+\beta|\bar{1}\rangle$ in photonic dual-rail encoding, $|\bar{0}\rangle\equiv|10\rangle$, $|\bar{1}\rangle\equiv|01\rangle$, can be implemented by directly detecting the photons in the two modes. In fact, in order to satisfy our general principle in Eq. (2), we need to fulfil $1-\bar{F}^{\rm optimal}_{\rm linear\ optics,dephasing}=1-\bar{F}^{\rm optimal}_{\rm quantum\ mechanics}=1/3$; this can be accomplished by directly dephasing the input state [24].

An example that leads to highly non-trivial POVM solutions is the calculation of the accessible information in quantum communication, involving an extremely difficult optimization problem. Although our general principle expressed in Eq. (2) applies to this problem as well, here we are not going to attempt to treat a linear-optics implementation of the accessible information gain.

By contrast, a relatively simple optimization leads to the optimal unambiguous discrimination of quantum states, i.e., a scheme that either identifies the signal state correctly or it yields an inconclusive result with the smallest probability allowed by quantum theory. In this case, the cost function is the probability for obtaining an inconclusive result. Now it has been shown that in general, this failure probability squared has a lower bound determined by the fidelity of the signal states, $\mathrm{Prob}^2_{\rm fail}\geq F$ [21]. For two pure non-orthogonal signal states, the minimal failure probability squared exactly coincides with the overlap (fidelity) of the two states (assuming equal a priori probabilities [25–27]). Thus, as for implementing optimal unambiguous state discrimination (USD), we can directly apply our general principle to the fidelities of the states before and after dephasing. The corresponding optimal POVM solution is a non-trivial non-projective POVM, consisting of two POVM elements for the correct identification of the states and one that describes the inconclusive result [28–30]. Let us now consider the question whether this optimal USD of *two pure* non-orthogonal states can be implemented with linear optics.

A. Optimal unambiguous state discrimination

The optimal unambiguous state discrimination (USD) of two pure non-orthogonal states

$$\begin{aligned} |\chi_+\rangle &= \alpha|\bar{0}\rangle + \beta|\bar{1}\rangle \,, \\ |\chi_-\rangle &= \alpha|\bar{0}\rangle - \beta|\bar{1}\rangle \,, \end{aligned} \tag{5}$$

where $\alpha > \beta$ are assumed to be real and $\{|\bar{0}\rangle, |\bar{1}\rangle\}$ are two basis states, corresponds to a projection onto the orthogonal set (see also App. A),

$$|w_\mu\rangle = |u_\mu\rangle + |N_\mu\rangle \,, \tag{6}$$

in an extended Hilbert space. Here, the $\{|u_\mu\rangle\}$ are state vectors in a Hilbert space $\mathcal{K}$ such that

$$\hat{E}_\mu = |u_\mu\rangle\langle u_\mu| \tag{7}$$

are the POVM operators of a three-valued POVM, $\mu = 1, 2, 3$, with $\sum_\mu \hat{E}_\mu = \mathbb{1}$. The vectors $\{|N_\mu\rangle\}$ are defined in the complementary space $\mathcal{K}^\perp$ orthogonal to $\mathcal{K}$, with the total Hilbert space $\mathcal{H} = \mathcal{K} \oplus \mathcal{K}^\perp$. For the optimal USD, one can show that

$$\begin{aligned} |u_{1/2}\rangle &= \frac{1}{\sqrt{2}}\left(\frac{\beta}{\alpha}\,|\bar{0}\rangle \pm |\bar{1}\rangle\right) \,, \\ |N_{1/2}\rangle &= \frac{1}{\sqrt{2}}\sqrt{1-\frac{\beta^2}{\alpha^2}}\,|\bar{2}\rangle \,, \\ |u_3\rangle &= \sqrt{1-\frac{\beta^2}{\alpha^2}}\,|\bar{0}\rangle \,, \quad |N_3\rangle = -\frac{\beta}{\alpha}\,|\bar{2}\rangle \,, \end{aligned} \tag{8}$$

and $\langle\bar{2}|\bar{0}\rangle = \langle\bar{2}|\bar{1}\rangle = 0$. The first two POVM elements ($\mu = 1, 2$) here refer to the two signal states, whereas the third POVM element ($\mu = 3$) corresponds to the inconclusive result. To make the discrimination unambiguous, we have indeed $\mathrm{Tr}(\hat{E}_1|\chi_-\rangle\langle\chi_-|) = \mathrm{Tr}(\hat{E}_2|\chi_+\rangle\langle\chi_+|) = 0$ with $\hat{E}_\mu$ from Eq. (7) and Eq. (8). To make it optimal, we have

$$\begin{aligned} \mathrm{Prob}_{\mathrm{succ}} &= \mathrm{Tr}(\hat{E}_1|\chi_+\rangle\langle\chi_+|)/2 + \mathrm{Tr}(\hat{E}_2|\chi_-\rangle\langle\chi_-|)/2 \\ &= 1 - \mathrm{Prob}_{\mathrm{fail}} \\ &= 1 - \mathrm{Tr}(\hat{E}_3|\chi_+\rangle\langle\chi_+|)/2 - \mathrm{Tr}(\hat{E}_3|\chi_-\rangle\langle\chi_-|)/2 \\ &= 1 - |\langle\chi_+|\chi_-\rangle| = 1 - (\alpha^2 - \beta^2) = 2\,\beta^2 \,. \end{aligned} \tag{9}$$

As for the linear-optical implementation, using one-photon signal states and multiple-rail encoding, $|\bar{0}\rangle \equiv |100\rangle$, $|\bar{1}\rangle \equiv |010\rangle$, $|\bar{2}\rangle \equiv |001\rangle$, one can directly implement the corresponding POVM for the optimal USD, as described in App. A for general single-photon based POVMs [Eq. (A3) and Eq. (A4)]. In this case, the output states after the linear-optics circuit, $|100\rangle$, $|010\rangle$, and $|001\rangle$, uniquely refer to one of the three orthogonal states $|w_\mu\rangle$, and hence identify the signal states $|\chi_+\rangle$ and $|\chi_-\rangle$ with the best possible probability. However, in general, for arbitrary signal states, it turns out to be very hard to decide whether the optimal USD can be implemented, because of the infinite number of possible Naimark extensions. In the following, we will investigate the optimal USD of two pure states independent of the Naimark extension, using the general principle introduced in the preceding sections and expressed in Eq. (2).

A suitable cost function for the USD of two pure non-orthogonal states is the failure probability. When optimized over all possible POVMs, the minimal failure probability corresponds to the overlap of the states. Thus, according to Eq. (2) and since after dephasing the mixed-state USD failure probability is bounded from below by the fidelity [21], we obtain the condition

$$F(\hat{\rho}'_{+,H}, \hat{\rho}'_{-,H}) \stackrel{!}{=} F(\hat{\rho}_+, \hat{\rho}_-) \,, \tag{10}$$

where $F(\hat{\rho}_+, \hat{\rho}_-)$ is the fidelity of the input states and $F(\hat{\rho}'_{+,H}, \hat{\rho}'_{-,H})$ is the fidelity after linear optics and dephasing. Note that the fidelity of the dephased density matrices only yields a lower bound on the failure probability and the optimal failure probability may well exceed this bound. Thus, even for a fixed array of linear optics [31], described by totally dephased density matrices, the criterion in Eq. (10) is, in general, only a *necessary condition* for optimal USD. As a result, for the optimal USD of two pure states via linear optics and subsequent photon counting (of all modes after static linear optics or only one first mode in a conditional-dynamics scheme), we have the following rule: the linear-optics circuit must be chosen such that *the overlap of the two states in terms of the fidelity is the same before and after dephasing.* This statement, as expressed by Eq. (10), extends the exact discrimination of orthogonal states to the more general scenario for optimal discrimination of nonorthogonal states. Whether linear optics or, more generally, linear transformations including multi-mode squeezing (corresponding to arbitrary quadratic interactions) are sufficient to implement optimal USD depends on the ability of these tools to obey the above rule. When focusing on the special case of USD, a more direct derivation of the fidelity criterion in Eq. (10) is possible and given in App. B. Let us now examine the statement in Eq. (10) in more detail for a fixed array of linear optics.

B. Optimal USD via a fixed linear network

For a fixed array of linear optics, all output modes will be detected at once. Therefore, Eq. (10) refers to totally dephased density matrices. In order to check the criterion in Eq. (10), we find that the fidelity before and after linear optics becomes

$$F(\hat{\rho}_+, \hat{\rho}_-) = F(\hat{\rho}_{+,H}, \hat{\rho}_{-,H}) = \sum_{m,n} \alpha_m^* \alpha_n \beta_m \beta_n^* \,, \tag{11}$$

because after the linear-optics transformation, the output states will always take on the following form,

$$
\begin{aligned}
|\chi_{+,H}\rangle &= \sum_k \alpha_k |\{k\}\rangle + \sum_m \alpha_m |\{m\}\rangle \\
|\chi_{-,H}\rangle &= \sum_l \beta_l |\{l\}\rangle + \sum_m \beta_m |\{m\}\rangle \,, \qquad (12)
\end{aligned}
$$

where the coefficients depend on the linear-optics circuit chosen in a particular implementation. The indices k and l denote photon number patterns, i.e., N-mode Fock states, that exclusively occur in the expansion of $|\chi_{+,H}\rangle$ and $|\chi_{-,H}\rangle$, respectively. Hence these patterns unambiguously refer to the $+$ state or to the $-$ state. However, because of the finite overlap of the input states, we must include patterns that occur in the expansion of both states. These ambiguous patterns are denoted by the index m. In general, the amplitudes of the ambiguous N-mode Fock states in the expansions, and hence the probabilities for the corresponding patterns to be detected, may be different for the $+$ and the $-$ state.

After dephasing, the output states take on the following form

$$
\begin{aligned}
\hat{\rho}'_{+,H} &= \sum_k \mathrm{P}^+_k |\{k\}\rangle\langle\{k\}| + \sum_m \mathrm{P}^+_m |\{m\}\rangle\langle\{m\}| \\
\hat{\rho}'_{-,H} &= \sum_l \mathrm{P}^-_l |\{l\}\rangle\langle\{l\}| + \sum_m \mathrm{P}^-_m |\{m\}\rangle\langle\{m\}|, \qquad (13)
\end{aligned}
$$

corresponding to a dephasing of the states in Eq. (12) with the probabilities given by $\mathrm{P}^+_k = |\alpha_k|^2$, $\mathrm{P}^-_l = |\beta_l|^2$, $\mathrm{P}^+_m = |\alpha_m|^2$, and $\mathrm{P}^-_m = |\beta_m|^2$.

The fidelity after linear optics and dephasing is now given by

$$
\begin{aligned}
F(\hat{\rho}'_{+,H}, \hat{\rho}'_{-,H}) &= \left(\mathrm{Tr}\sqrt{\hat{\rho}'_{+,H}\, \hat{\rho}'_{-,H}}\right)^2 \\
&= \left(\sum_m \sqrt{\mathrm{P}^+_m \mathrm{P}^-_m}\right)^2 . \qquad (14)
\end{aligned}
$$

Thus, the fidelity criterion from Eq. (10) can be expressed by

$$
\sum_{m,n} \sqrt{\mathrm{P}^+_m \mathrm{P}^+_n \mathrm{P}^-_m \mathrm{P}^-_n} \stackrel{!}{=} \sum_{m,n} \alpha^*_m \alpha_n \beta_m \beta^*_n \,, \qquad (15)
$$

using Eq. (14) and Eq. (11). This, however, implies that

$$
\begin{aligned}
\sum_{m,n} |\alpha_m||\alpha_n||\beta_m||\beta_n| &\stackrel{!}{=} \sum_{m,n} |\alpha_m||\alpha_n||\beta_m||\beta_n| \\
&\quad \times\; e^{i(\phi^-_m - \phi^-_n + \phi^+_n - \phi^+_m)} \,, \qquad (16)
\end{aligned}
$$

where $\alpha_m = |\alpha_m| e^{i\phi^+_m}$ and $\beta_m = |\beta_m| e^{i\phi^-_m}$, etc. The only possible way to satisfy Eq. (16) is for $e^{i(\phi^-_m - \phi^-_n + \phi^+_n - \phi^+_m)} = 1$, $\forall m, n$. Thus, we have $\phi^-_m - \phi^+_m = \phi$, $\forall m$. A direct consequence of this result is that the overlap of the input states can be written as

$$
\begin{aligned}
|\langle\chi_+|\chi_-\rangle| &= |\langle\chi_{+,H}|\chi_{-,H}\rangle| \\
&= \left|\sum_m \alpha^*_m \beta_m\right| = \sum_m |\alpha_m||\beta_m|. \qquad (17)
\end{aligned}
$$

Let us now look at the first-order expression from the conditions in Eq. (3). We obtain

$$
\begin{aligned}
\langle\chi_+|\hat{c}^\dagger_j \hat{c}_j|\chi_-\rangle &= \langle\chi_{+,H}|\hat{a}^\dagger_j \hat{a}_j|\chi_{-,H}\rangle \qquad (18) \\
&= e^{i\phi} \sum_m |\alpha_m||\beta_m| \langle\{m\}|\hat{a}^\dagger_j \hat{a}_j|\{m\}\rangle \,,
\end{aligned}
$$

because annihilating a photon in the jth mode of both states only leads to nonzero contributions from coinciding patterns. Using Eq. (17) and Eq. (18), there are two observations we can make. First, the modulus of any first order expression $\langle\chi_+|\hat{c}^\dagger_j \hat{c}_j|\chi_-\rangle$ is bounded from above such that

$$
\left|\langle\chi_+|\hat{c}^\dagger_j \hat{c}_j|\chi_-\rangle\right| \leq N \; |\langle\chi_+|\chi_-\rangle| \;, \forall j \,, \qquad (19)
$$

where N is the maximum photon number in the states. In addition, we have

$$
\frac{\langle\chi_+|\hat{c}^\dagger_j \hat{c}_j|\chi_-\rangle}{\left|\langle\chi_+|\hat{c}^\dagger_j \hat{c}_j|\chi_-\rangle\right|} = \frac{\langle\chi_+|\hat{c}^\dagger_{j'} \hat{c}_{j'}|\chi_-\rangle}{\left|\langle\chi_+|\hat{c}^\dagger_{j'} \hat{c}_{j'}|\chi_-\rangle\right|} \,, \forall j, j' \,, \qquad (20)
$$

provided that $\langle\chi_+|\hat{c}^\dagger_j \hat{c}_j|\chi_-\rangle$ and $\langle\chi_+|\hat{c}^\dagger_{j'} \hat{c}_{j'}|\chi_-\rangle$ are both nonzero. Similarly, for the second-order expressions, we obtain

$$
\begin{aligned}
\langle\chi_+|\hat{c}^\dagger_j \hat{c}^\dagger_{j'} \hat{c}_j \hat{c}_{j'}|\chi_-\rangle &= \langle\chi_{+,H}|\hat{a}^\dagger_j \hat{a}^\dagger_{j'} \hat{a}_j \hat{a}_{j'}|\chi_{-,H}\rangle \qquad (21) \\
&= e^{i\phi} \sum_m |\alpha_m||\beta_m| \\
&\qquad \times \langle\{m\}|\hat{a}^\dagger_j \hat{a}^\dagger_{j'} \hat{a}_j \hat{a}_{j'}|\{m\}\rangle.
\end{aligned}
$$

This leads to

$$
\left|\langle\chi_+|\hat{c}^\dagger_j \hat{c}^\dagger_{j'} \hat{c}_j \hat{c}_{j'}|\chi_-\rangle\right| \leq N(N-1) \; |\langle\chi_+|\chi_-\rangle| \;, \forall j, j', \qquad (22)
$$

and

$$
\frac{\langle\chi_+|\hat{c}^\dagger_j \hat{c}^\dagger_i \hat{c}_j \hat{c}_i|\chi_-\rangle}{\left|\langle\chi_+|\hat{c}^\dagger_j \hat{c}^\dagger_i \hat{c}_j \hat{c}_i|\chi_-\rangle\right|} = \frac{\langle\chi_+|\hat{c}^\dagger_{j'} \hat{c}^\dagger_{i'} \hat{c}_{j'} \hat{c}_{i'}|\chi_-\rangle}{\left|\langle\chi_+|\hat{c}^\dagger_{j'} \hat{c}^\dagger_{i'} \hat{c}_{j'} \hat{c}_{i'}|\chi_-\rangle\right|} \,, \forall j, i, j', i' \,, \qquad (23)
$$

provided that $\langle\chi_+|\hat{c}^\dagger_j \hat{c}^\dagger_i \hat{c}_j \hat{c}_i|\chi_-\rangle$ and $\langle\chi_+|\hat{c}^\dagger_{j'} \hat{c}^\dagger_{i'} \hat{c}_{j'} \hat{c}_{i'}|\chi_-\rangle$ are both nonzero. Moreover, for all non-vanishing expressions, also the phases of different orders must coincide. As a result, we have proven the following theorem for the implementability of optimal USD of two pure nonorthogonal states with linear optics.

Theorem: it is a necessary (but, in general, not sufficient) criterion for the possibility of implementing the optimal USD of two pure nonorthogonal states $|\chi_\pm\rangle$ via static linear optics and photon counting that the hierarchies of conditions,

$$\begin{aligned}
&\frac{\langle\chi_+|\chi_-\rangle}{|\langle\chi_+|\chi_-\rangle|}\\
&= \frac{\langle\chi_+|\hat{c}_j^\dagger\hat{c}_j|\chi_-\rangle}{\left|\langle\chi_+|\hat{c}_j^\dagger\hat{c}_j|\chi_-\rangle\right|} = \frac{\langle\chi_+|\hat{c}_{j'}^\dagger\hat{c}_{j'}|\chi_-\rangle}{\left|\langle\chi_+|\hat{c}_{j'}^\dagger\hat{c}_{j'}|\chi_-\rangle\right|}\\
&= \frac{\langle\chi_+|\hat{c}_j^\dagger\hat{c}_i^\dagger\hat{c}_j\hat{c}_i|\chi_-\rangle}{\left|\langle\chi_+|\hat{c}_j^\dagger\hat{c}_i^\dagger\hat{c}_j\hat{c}_i|\chi_-\rangle\right|} = \frac{\langle\chi_+|\hat{c}_{j'}^\dagger\hat{c}_{i'}^\dagger\hat{c}_{j'}\hat{c}_{i'}|\chi_-\rangle}{\left|\langle\chi_+|\hat{c}_{j'}^\dagger\hat{c}_{i'}^\dagger\hat{c}_{j'}\hat{c}_{i'}|\chi_-\rangle\right|}\,, \text{etc.}\,,\\
&\qquad\qquad \forall j, i, j', i', \text{etc.}\,, \qquad (24)
\end{aligned}$$

for any non-vanishing orders, and

$$\begin{aligned}
\left|\langle\chi_+|\hat{c}_j^\dagger\hat{c}_j|\chi_-\rangle\right| &\leq N\ |\langle\chi_+|\chi_-\rangle|\ , \forall j\,,\\
\left|\langle\chi_+|\hat{c}_j^\dagger\hat{c}_{j'}^\dagger\hat{c}_j\hat{c}_{j'}|\chi_-\rangle\right| &\leq N(N-1)\ |\langle\chi_+|\chi_-\rangle|\ , \forall j, j',\\
\vdots \quad &\leq \quad \vdots \qquad (25)\\
\left|\langle\chi_+|\hat{c}_j^\dagger\hat{c}_{j'}^\dagger\hat{c}_{j''}^\dagger\cdots\hat{c}_j\hat{c}_{j'}\hat{c}_{j''}\cdots|\chi_-\rangle\right| &\leq\\
&N!\ |\langle\chi_+|\chi_-\rangle|\ , \forall j, j', j'', ...,
\end{aligned}$$

are satisfied, where N is the maximum photon number in the states. For the existence of a linear-optics solution to the POVM that realizes the optimal USD, output mode operators $\hat{c}_j$ must be found such that a unitary matrix U can be constructed with $\hat{c}_j = \sum_i U_{ji}\hat{a}_i$ and the hierarchies of conditions are satisfied for these operators.

Note that for fixed photon number, the first set of conditions of the theorem [that for the phases in Eq. (24)] implies the second one [that for the absolute values in Eq. (25)], however, the converse does not hold. The reason is that in *any* linear-optics scheme, when the input states have a fixed number of photons N, the total sum of a given order satisfies a relation similar to

$$\langle\chi_+|\sum_j \hat{c}_j^\dagger\hat{c}_j|\chi_-\rangle \;=\; N\,\langle\chi_+|\chi_-\rangle\,. \qquad (26)$$

Here, the sum of the first-order expressions leads to the total photon number operator for the output modes, and the relation in Eq. (26) follows from the photon number conservation property of linear optics. Analogous conditions can be found for the total sum of the higher order expressions. Now for the sum of the first orders, for instance, according to Eq. (26), we obtain

$$\left|\sum_j\langle\chi_+|\hat{c}_j^\dagger\hat{c}_j|\chi_-\rangle\right| = N\ |\langle\chi_+|\chi_-\rangle|\ , \qquad (27)$$

and, provided the phases of all non-vanishing first orders coincide as required to obey Eq. (24),

$$\left|\sum_j\langle\chi_+|\hat{c}_j^\dagger\hat{c}_j|\chi_-\rangle\right| = \sum_j\left|\langle\chi_+|\hat{c}_j^\dagger\hat{c}_j|\chi_-\rangle\right|\,. \qquad (28)$$

The two relations of Eq. (27) and Eq. (28) together imply that the first-order conditions in Eq. (25) are automatically satisfied by any linear-optics transformation that fulfils the first-order conditions in Eq. (24).

The hierarchies of conditions in Eq. (24) and Eq. (25) are necessary for optimal USD of two nonorthogonal states. In words, these criteria mean that for optimal USD, *the phases of all non-vanishing orders must coincide.* For example, any real orders must have the same sign in optimal USD. In the special case of orthogonal states, $\langle\chi_+|\chi_-\rangle = 0$, one can easily see in Eq. (25) that the hierarchy for exact state discrimination from Eq. (3) can be retrieved. An alternative derivation of the conditions in Eq. (24) and Eq. (25), independent of the fidelity criterion in Eq. (10), is given in App. C.

Let us now look at an example of two nonorthogonal states with only two photons (which is the simplest nontrivial extension to the trivial case of one-photon states). We are going to consider the two-photon toy model state $\alpha|20\rangle \pm \beta|11\rangle$, where, without loss of generality, α and β are assumed to be real. For the special case of an orthogonal pair $\alpha = \beta$, it is know that there is no linear-optics solution for optimally and hence exactly discriminating these two states, including feedforward and arbitrary auxiliary states [15]. Here we consider the nonorthogonal case without auxiliary photons, but arbitrarily many additional vacuum modes.

Defining $U_{j1} \equiv \nu_1$ and $U_{j2} \equiv \nu_2$ for the elements of the jth row of the unitary matrix in $\hat{c}_j = \sum_i U_{ji}\hat{a}_i$, the first-order expression for mode j then becomes

$$\begin{aligned}
\langle\chi_+|\hat{c}_j^\dagger\hat{c}_j|\chi_-\rangle \;&=\; 2\alpha^2|\nu_1|^2 - \beta^2(|\nu_1|^2 + |\nu_2|^2)\\
&\quad +\sqrt{2}\alpha\beta(\nu_1\nu_2^* - c.c.)\,. \qquad (29)
\end{aligned}$$

Now assuming a *fixed array* of linear optics, the conditions in Eq. (24) are necessary for optimal USD. In order to satisfy the first-order conditions there for any modes j, j', etc., the expression in Eq. (29) must become real for any j, j', etc., because the 0th order $\langle\chi_+|\chi_-\rangle = \alpha^2 - \beta^2$ is real. The same argument applies to the second-order expressions,

$$\begin{aligned}
\langle\chi_+|(\hat{c}_j^\dagger)^2\hat{c}_j^2|\chi_-\rangle \;&=\; 2|\nu_1|^2\big[\alpha^2|\nu_1|^2 - 2\beta^2|\nu_2|^2\\
&\quad +\sqrt{2}\alpha\beta(\nu_1\nu_2^* - c.c.)\big]\,. \qquad (30)
\end{aligned}$$

Knowing that all these expressions must be real, let us evaluate the first-order and second-order conditions for positive and negative 0th order $\langle\chi_+|\chi_-\rangle$, $\alpha^2 > \beta^2$ or $\alpha^2 < \beta^2$, respectively. According to Eq. (24), we obtain

$$\vec{\nu}\,\vec{a} \geq 0\,, \quad \text{and} \quad \vec{\nu}\,\vec{b} \geq 0\,, \qquad (31)$$

for $\alpha^2 > \beta^2$, and

$$\vec{\nu}\,\vec{a} \leq 0\,, \quad \text{and} \quad \vec{\nu}\,\vec{b} \leq 0\,, \tag{32}$$

for $\alpha^2 < \beta^2$, where

$$\vec{\nu} \equiv \begin{pmatrix} |\nu_1|^2 \\ |\nu_2|^2 \end{pmatrix}, \quad \vec{a} \equiv \begin{pmatrix} 2\alpha^2 - \beta^2 \\ -\beta^2 \end{pmatrix}, \quad \vec{b} \equiv \begin{pmatrix} \alpha^2 \\ -2\beta^2 \end{pmatrix}. \tag{33}$$

In Eq. (31), $\vec{\nu}\,\vec{b} \geq 0$ implies that $\alpha^2 \geq 2\beta^2$, because otherwise, for $\alpha^2 < 2\beta^2$, the only way to prevent $\vec{\nu}\,\vec{b}$ from becoming negative is to have $|\nu_1|^2 > |\nu_2|^2$ for any modes j, j', etc., according to Eq. (24). However, no unitary matrix can be constructed, where all the elements j, j', etc., in the first two columns satisfy $|\nu_1|^2 > |\nu_2|^2$. Similarly, in Eq. (32), $\vec{\nu}\,\vec{a} \leq 0$ leads to $\alpha^2 \leq \beta^2$ (which is also simply given by the 0th order). Thus, there is a regime, $\beta^2 \leq \alpha^2 < 2\beta^2$ (including the orthogonal case $\alpha^2 = \beta^2$), where optimal USD is impossible for a fixed array of linear optics and without auxiliary photons.

For $\alpha^2 = 2\beta^2$, the optimal solution is a simple 50/50 beam splitter, $|\nu_1|^2 = |\nu_2|^2 = 1/2$ for modes $j = 1, 2$. In this case, in agreement with Eq. (31), we obtain $\vec{\nu}\,\vec{a} = \beta^2 > 0$ and $\vec{\nu}\,\vec{b} = 0$. The orthogonal set of the corresponding von Neumann measurement becomes

$$\begin{aligned} |w_{1/2}\rangle &= \frac{1}{\sqrt{2}}\left[\frac{1}{\sqrt{2}}(|20\rangle + |02\rangle) \pm |11\rangle\right], \\ |w_3\rangle &= \frac{1}{\sqrt{2}}(|20\rangle - |02\rangle)\,, \end{aligned} \tag{34}$$

choosing for the Naimark extension $|\bar{2}\rangle \equiv |02\rangle$ [see Eqs. (5)-(8)]. A symmetric beam splitter turns these three states into $|20\rangle$, $|02\rangle$, and $|11\rangle$, respectively, which via photon counting uniquely refer to the three different POVM elements. Thus, optimal USD for $\alpha = \sqrt{2}\beta$ can be achieved with a simple beam splitter. The two signal states $\alpha|20\rangle \pm \beta|11\rangle$ are transformed by the symmetric beam splitter into $\sqrt{\frac{2}{3}}|20\rangle + \frac{1}{\sqrt{3}}|11\rangle$ and $\sqrt{\frac{2}{3}}|02\rangle + \frac{1}{\sqrt{3}}|11\rangle$, respectively. Indeed, we have $\text{Prob}_{\text{succ}} = 2/3 = 2\beta^2$.

Let us now apply the general fidelity criterion to a more sophisticated linear-optics implementation of state discrimination, namely one that includes conditional dynamics (feedforward): instead of detecting all output modes after the linear-optics circuit, one may select only one mode for detection. After this first measurement, one can then send the conditional state of the remaining modes through another linear-optics circuit which depends on the measurement outcome. In the most general approach, one can include as many feedforward steps as modes are in the signal states, or even more by adding auxiliary states.

C. Optimal USD via conditional dynamics

Let's assume, without loss of generality, that mode 1 is detected first, corresponding to a partial dephasing of the states only with respect to that mode. Now instead of writing the states after linear optics as in Eq. (12), we use the following expressions,

$$\begin{aligned} |\chi_{+,H}\rangle &= \sum_k \alpha_k |k\rangle_1 \otimes |\tilde{\gamma}_k^+\rangle + \sum_m \alpha_m |m\rangle_1 \otimes |\tilde{\gamma}_m^+\rangle \\ |\chi_{-,H}\rangle &= \sum_l \beta_l |l\rangle_1 \otimes |\tilde{\gamma}_l^-\rangle + \sum_m \beta_m |m\rangle_1 \otimes |\tilde{\gamma}_m^-\rangle\,, \end{aligned} \tag{35}$$

where this time, the states $|k\rangle_1$ and $|l\rangle_1$ represent those number states of mode 1 which only occur in the expansion of the + and the − state, respectively. The one-mode states $|m\rangle_1$ lead to the ambiguous detection events in mode 1. Finally, the states $|\tilde{\gamma}_k^+\rangle$, etc., refer to the corresponding conditional states of the remaining modes (after normalization). Similarly, for the partially dephased density operators, we obtain

$$\begin{aligned} \hat{\rho}'_{+,H} &= \sum_k \mathrm{P}_k^+ |k\rangle_1\langle k| \otimes |\tilde{\gamma}_k^+\rangle\langle\tilde{\gamma}_k^+| \\ &+ \sum_m \mathrm{P}_m^+ |m\rangle_1\langle m| \otimes |\tilde{\gamma}_m^+\rangle\langle\tilde{\gamma}_m^+| \\ \hat{\rho}'_{-,H} &= \sum_l \mathrm{P}_l^- |l\rangle_1\langle l| \otimes |\tilde{\gamma}_l^-\rangle\langle\tilde{\gamma}_l^-| \\ &+ \sum_m \mathrm{P}_m^- |m\rangle_1\langle m| \otimes |\tilde{\gamma}_m^-\rangle\langle\tilde{\gamma}_m^-|\,. \end{aligned} \tag{36}$$

Note that the partially dephased states are no longer diagonal in the Fock basis, i.e., the conditional density matrices may contain off-diagonal terms. The corresponding fidelities are now

$$\begin{aligned} F(\hat{\rho}_+, \hat{\rho}_-) &= F(\hat{\rho}_{+,H}, \hat{\rho}_{-,H}) \\ &= \sum_{m,n} \alpha_m^* \alpha_n \beta_m \beta_n^* \,\langle\tilde{\gamma}_m^+|\tilde{\gamma}_m^-\rangle\langle\tilde{\gamma}_n^+|\tilde{\gamma}_n^-\rangle^*\,, \end{aligned} \tag{37}$$

and

$$F(\hat{\rho}'_{+,H}, \hat{\rho}'_{-,H}) = \left(\sum_m \sqrt{\mathrm{P}_m^+ \mathrm{P}_m^-}\,\left|\langle\tilde{\gamma}_m^+|\tilde{\gamma}_m^-\rangle\right|\right)^2. \tag{38}$$

Finally, we end up having the following condition due to the fidelity criterion in Eq. (10),

$$\begin{aligned} &\sum_{m,n} \sqrt{\mathrm{P}_m^+ \mathrm{P}_n^+ \mathrm{P}_m^- \mathrm{P}_n^-}\,\left|\langle\tilde{\gamma}_m^+|\tilde{\gamma}_m^-\rangle\right|\left|\langle\tilde{\gamma}_n^+|\tilde{\gamma}_n^-\rangle\right| \stackrel{!}{=} \\ &\qquad \sum_{m,n} \alpha_m^* \alpha_n \beta_m \beta_n^* \langle\tilde{\gamma}_m^+|\tilde{\gamma}_m^-\rangle\langle\tilde{\gamma}_n^+|\tilde{\gamma}_n^-\rangle^*\,, \end{aligned} \tag{39}$$

or, in terms of the unnormalized conditional states,

$$\sum_{m,n} \left|\langle\gamma_m^+|\gamma_m^-\rangle\right|\left|\langle\gamma_n^+|\gamma_n^-\rangle\right| \stackrel{!}{=} \sum_{m,n} \langle\gamma_m^+|\gamma_m^-\rangle\langle\gamma_n^+|\gamma_n^-\rangle^*\,. \tag{40}$$

Now the only way to satisfy this condition is through

$$\frac{\langle\gamma_m^+|\gamma_m^-\rangle}{\left|\langle\gamma_m^+|\gamma_m^-\rangle\right|} \stackrel{!}{=} \frac{\langle\gamma_n^+|\gamma_n^-\rangle}{\left|\langle\gamma_n^+|\gamma_n^-\rangle\right|}\,, \tag{41}$$

for any nonzero overlaps labeled by m and n. In other words, for any two inconclusive one-mode detection events, any non-vanishing overlaps of the conditional states coming from the + signal and the − signal must have equal phases. Finally, we can now again examine the first-order condition of our criteria, however, here only for the detected mode 1,

$$\begin{aligned} \langle\chi_+|\hat{c}_1^\dagger\hat{c}_1|\chi_-\rangle &= \langle\chi_{+,H}|\hat{a}_1^\dagger\hat{a}_1|\chi_{-,H}\rangle \\ &= \sum_m \langle\gamma_m^+|\gamma_m^-\rangle\ {}_1\langle m|\hat{a}_1^\dagger\hat{a}_1|m\rangle_1 \,. \end{aligned} \tag{42}$$

Similar expressions hold for the higher orders in mode 1. Note that the different orders here are evaluated only for the first mode to be detected, corresponding to the first step in a conditional-dynamics scheme. Of course, in addition, one could calculate further expressions using the conditional states of modes 2 through N in order to derive more criteria for a conditional-dynamics protocol. Here, however, we only focus on the first step in any conditional-dynamics scheme, namely the detection of a first mode.

Using Eq. (42) and Eq. (41) (for any nonzero overlaps), it becomes clear now that the hierarchies of conditions necessary for optimal USD when detecting a first mode j are simply the subset of conditions in Eq. (24) and Eq. (25) referring to this one mode. Thus, we obtain

$$\begin{aligned} &\frac{\langle\chi_+|\chi_-\rangle}{|\langle\chi_+|\chi_-\rangle|} \\ &= \frac{\langle\chi_+|\hat{c}_j^\dagger\hat{c}_j|\chi_-\rangle}{\left|\langle\chi_+|\hat{c}_j^\dagger\hat{c}_j|\chi_-\rangle\right|} = \frac{\langle\chi_+|(\hat{c}_j^\dagger)^2\hat{c}_j^2|\chi_-\rangle}{\left|\langle\chi_+|(\hat{c}_j^\dagger)^2\hat{c}_j^2|\chi_-\rangle\right|} \\ &= \cdots = \frac{\langle\chi_+|(\hat{c}_j^\dagger)^n\hat{c}_j^n|\chi_-\rangle}{\left|\langle\chi_+|(\hat{c}_j^\dagger)^n\hat{c}_j^n|\chi_-\rangle\right|}\,,\text{etc.}\,, \end{aligned} \tag{43}$$

for any non-vanishing orders, and

$$\begin{aligned} \left|\langle\chi_+|\hat{c}_j^\dagger\hat{c}_j|\chi_-\rangle\right| &\leq N\ |\langle\chi_+|\chi_-\rangle|\ , \\ \left|\langle\chi_+|(\hat{c}_j^\dagger)^2\hat{c}_j^2|\chi_-\rangle\right| &\leq N(N-1)\ |\langle\chi_+|\chi_-\rangle|\ ,\text{etc.} \end{aligned} \tag{44}$$

In the next section, we discuss whether our new set of conditions enables us to make general statements about the use of auxiliary photons for the optimal USD of two nonorthogonal states.

D. Auxiliary photons for optimal USD

Let us investigate whether adding an auxiliary state can make optimal USD via linear optics possible when it is impossible without an ancilla state. The auxiliary state contains optical modes in addition to the signal modes, and these extra modes may be occupied by additional photons. Let us use the notation $|\chi_\pm\rangle = |s_\pm\rangle\otimes|\psi_{\rm aux}\rangle$, where now the states $|s_\pm\rangle$ represent the signal states and the state $|\psi_{\rm aux}\rangle$ is the auxiliary state.

Splitting the input modes into a set of signal and a set of auxiliary modes allows us to decompose the output mode operator $\hat{c}_j = \sum_i U_{ji}\hat{a}_i$ into two corresponding parts as (we drop the index j) $\hat{c} = b_{\rm s}\hat{c}_{\rm s} + b_{\rm aux}\hat{c}_{\rm aux}$, with real coefficients $b_{\rm s}$ and $b_{\rm aux}$, where $\hat{c}_{\rm s}$ acts only upon the signal modes and $\hat{c}_{\rm aux}$ only on the auxiliary modes. Now we find for the first-order expression

$$\begin{aligned} &\langle\chi_+|\hat{c}^\dagger\hat{c}|\chi_-\rangle \\ &= b_{\rm s}^2\langle s_+|\hat{c}_{\rm s}^\dagger\hat{c}_{\rm s}|s_-\rangle + b_{\rm s}b_{\rm aux}\langle s_+|\hat{c}_{\rm s}|s_-\rangle\langle\psi_{\rm aux}|\hat{c}_{\rm aux}^\dagger|\psi_{\rm aux}\rangle \\ &\quad + b_{\rm s}b_{\rm aux}\langle s_+|\hat{c}_{\rm s}^\dagger|s_-\rangle\langle\psi_{\rm aux}|\hat{c}_{\rm aux}|\psi_{\rm aux}\rangle \\ &\quad + b_{\rm aux}^2\langle s_+|s_-\rangle\langle\psi_{\rm aux}|\hat{c}_{\rm aux}^\dagger\hat{c}_{\rm aux}|\psi_{\rm aux}\rangle. \end{aligned} \tag{45}$$

For either signal or auxiliary states with a *fixed number of photons*, this becomes

$$\begin{aligned} \langle\chi_+|\hat{c}^\dagger\hat{c}|\chi_-\rangle &= b_{\rm s}^2\langle s_+|\hat{c}_{\rm s}^\dagger\hat{c}_{\rm s}|s_-\rangle \\ &\quad + b_{\rm aux}^2\langle s_+|s_-\rangle\langle\psi_{\rm aux}|\hat{c}_{\rm aux}^\dagger\hat{c}_{\rm aux}|\psi_{\rm aux}\rangle. \end{aligned} \tag{46}$$

Only when the signal states are orthogonal, the first-order expression without ancilla is proportional to that with ancilla, and hence the first-order condition does not change [15]. For the general case of nonorthogonal signal states, $\langle s_+|s_-\rangle \neq 0$, the first-order expression without ancilla is effectively displaced in phase space by adding an ancilla state. Let us look at the second-order for states with a fixed photon number,

$$\begin{aligned} \langle\chi_+|(\hat{c}^\dagger)^2\hat{c}^2|\chi_-\rangle &= b_{\rm s}^2\langle s_+|(\hat{c}_{\rm s}^\dagger)^2\hat{c}_{\rm s}^2|s_-\rangle \\ &\quad + b_{\rm aux}^2\langle s_+|s_-\rangle\langle\psi_{\rm aux}|(\hat{c}_{\rm aux}^\dagger)^2\hat{c}_{\rm aux}^2|\psi_{\rm aux}\rangle \\ &\quad + 4b_{\rm s}^2b_{\rm aux}^2\langle s_+|\hat{c}_{\rm s}^\dagger\hat{c}_{\rm s}|s_-\rangle \\ &\qquad \times\langle\psi_{\rm aux}|\hat{c}_{\rm aux}^\dagger\hat{c}_{\rm aux}|\psi_{\rm aux}\rangle. \end{aligned} \tag{47}$$

Here, in the general case of nonorthogonal states, we obtain again a phase-space displacement by adding the ancilla. It seems that, in general, we cannot rule out the possibility that adding an ancilla helps to satisfy the conditions for optimal USD when they cannot be fulfilled without ancilla. However, if the auxiliary state $|\psi_{\rm aux}\rangle$ contains no photons and is just the (multi-mode) vacuum state, we obtain $\langle\chi_+|(\hat{c}^\dagger)^2\hat{c}^2|\chi_-\rangle = b_{\rm s}^2\langle s_+|(\hat{c}_{\rm s}^\dagger)^2\hat{c}_{\rm s}^2|s_-\rangle$. As a result, adding auxiliary vacuum modes cannot change the sign or phase of the first order expression, since $b_{\rm s}^2$ is real and positive. However, adding auxiliary vacuum modes might be useful and necessary in order to build up a unitary matrix for the mode operators $\hat{c}_j = \sum_i U_{ji}\hat{a}_i$. In fact, in the one-photon example discussed after Eq. (9), adding an extra vacuum mode is essential in order to extend the two-dimensional signal Hilbert space to an at least three-dimensional space required for the POVM and to construct the corresponding unitary matrix.

There are also known examples, where adding extra photons makes the optimal USD of the two signal states via linear optics possible. One such example for the case of infinite-dimensional signal and auxiliary states, both

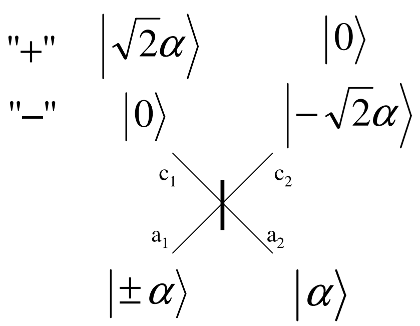

FIG. 2: Implementing the optimal unambiguous discrimination of two symmetric coherent states via a simple 50/50 beam splitter and an auxiliary coherent state of the same amplitude.

with *unfixed and unbounded photon number*, is the optimal USD of so-called binary coherent states. In this case, the optimal USD can be easily achieved using a 50/50 beam splitter and an ancilla coherent state (see Fig. 2). In our notation, one has $|s_\pm\rangle \equiv |\pm\alpha\rangle$ (α assumed to be real), $|\psi_{\rm aux}\rangle \equiv |\alpha\rangle$, and $|\chi_\pm\rangle = |s_\pm\rangle \otimes |\psi_{\rm aux}\rangle$. This two-mode state is now transformed by the 50/50 beam splitter into

$$\begin{aligned} |\chi_{+,H}\rangle &= |\sqrt{2}\alpha\rangle \otimes |0\rangle \\ |\chi_{-,H}\rangle &= |0\rangle \otimes |-\sqrt{2}\alpha\rangle \,. \end{aligned} \tag{48}$$

For these states, a detector click in mode 1 can only be triggered by the + state, whereas a click in mode 2 unambiguously refers to the − state. However, there are inconclusive "events" corresponding to the two-mode vacuum state, $|\psi_{\rm inconcl}\rangle = e^{-\alpha^2}|00\rangle$, using Eq. (C2) from App. C with $\phi = 0$. Since the failure probability is then given by $\mathrm{Prob}_{\rm fail}^{\rm lin.opt.} = (e^{-2\alpha^2} + e^{-2\alpha^2})/2 = e^{-2\alpha^2} = \langle +\alpha|-\alpha\rangle$, this scheme turns out to be optimal. Thus, we expect that the corresponding solution satisfies our criteria for optimal USD. For a particular mode j, again using $U_{j1} \equiv \nu_1$ and $U_{j2} \equiv \nu_2$ for the elements of the jth row of the unitary matrix in $\hat{c}_j = \sum_i U_{ji}\hat{a}_i$, we obtain the nth-order condition,

$$\begin{aligned} \langle\chi_+|(\hat{c}_j^\dagger)^n \hat{c}_j^n|\chi_-\rangle &= \langle +\alpha|-\alpha\rangle \Big(|\nu_1|^2\alpha^2 \\ &\quad -|\nu_2|^2\, {}_2\langle\psi_{\rm aux}|\hat{a}_2^\dagger\hat{a}_2|\psi_{\rm aux}\rangle_2\Big)^n . \end{aligned} \tag{49}$$

Apparently, for any mode $j = 1, 2$, any order $n \geq 1$ can be set to zero by choosing a 50/50 beam splitter, $|\nu_1|^2 = |\nu_2|^2 = 1/2$, and the appropriate ancilla state, $|\psi_{\rm aux}\rangle \equiv |\alpha\rangle$. This solution is indeed in agreement with the conditions that we derived for optimal USD. The obvious reason, why all nonzero orders vanish in this example, is that the only failure pattern here is $|00\rangle$ which always vanishes upon applying annihilation operators [see, e.g., Eq. (18)]. From this observation follows that also any cross orders for modes 1 and 2 will vanish with the above solution. Let us finally note that for the optimal USD of more than two coherent states, symmetrically distributed in phase space, the optimal USD [29] cannot be achieved as easily as for the binary case. However, there are asymptotic linear-optics solutions including the use of feedforward [32].

V. CONCLUSIONS

We considered the problem of implementing generalized measurements (POVMs) with linear optics. Such an implementation may either be based upon a static array of linear optics or it may include conditional dynamics (feedforward). Extending our previous results on projective measurements, we focused, in particular, on non-projective measurements. Our approach to this problem can be formulated as a general principle in the following way. We start by identifying a given POVM as a solution to an optimization problem for a chosen cost function. The implementation is then only possible if a linear-optics circuit exists for which the quantum mechanical optimum is still attainable after dephasing the corresponding quantum states. As an example for applying this principle to the problem of implementing a non-projective POVM, we discussed in detail the optimal unambiguous state discrimination (USD) of two pure nonorthogonal states. In order to implement the POVM that realizes the quantum mechanically optimal USD with linear optics, according to the general principle, the linear-optics circuit must be chosen such that the overlap of the states, in terms of the fidelity, is the same before and after dephasing. This statement extends the exact discrimination of orthogonal states to the more general scenario for optimal discrimination of nonorthogonal states. Using the fidelity criterion, we derived hierarchies of necessary conditions for the possibility of implementing the optimal USD of two pure nonorthogonal states via linear optics and photon counting. The resulting conditions are a generalization of our previous criteria for projection measurements and the exact discrimination of orthogonal states.

As for the detection mechanism, here we only studied the case of photon counting which leads to dephased states diagonal in the Fock basis. Potential extensions of our results may include different detection mechanisms such as homodyne detection, as we discussed previously already in the context of projective measurements. Moreover, apart from passive linear-optics circuits, our criteria can also be applied to arbitrary linear mode transformations, including multi-mode squeezing. When analyzing those POVMs that realize unambiguous state discrimination, one may also consider the USD of sets of three or more linearly independent states. Finally, let us emphasize that our approach of choosing suitable cost functions and applying them to the dephased quantum states might be as well useful for finding bounds on the efficiency of implementing POVMs with linear optics.

Acknowledgments

WM, PR, and NL acknowledge the support from the EU project RAMBOQ. PvL and KN acknowledge funding from MIC in Japan. PR and NL acknowledge the support of the DFG under the Emmy-Noether programme.

APPENDIX A: ONE-PHOTON SIGNAL STATES

Let us consider all those POVMs where the signal states contain only *one photon*. In this typical and important case, any unitary operation (gate) can be accomplished with linear optics [33]. This statement applies to arbitrary qu*d*it states, where each basis vector of the qu*d*it is described by one photon occupying one of d modes, $\hat{a}_i^\dagger|\mathbf{0}\rangle$, $i = 1...d$ ("multiple-rail encoding"). Similarly, any POVM can be implemented solely by means of linear optics for these one-photon signal states. This can be understood by looking at the corresponding Naimark extension of the POVM. The POVM is then described by a von Neumann measurement onto the orthogonal set

$$|w_\mu\rangle = |u_\mu\rangle + |N_\mu\rangle \ , \tag{A1}$$

in a Hilbert space larger than the original signal space. Here, the $\{|u_\mu\rangle\}$ are (unnormalized, potentially nonorthogonal) state vectors in a Hilbert space $\mathcal{K}$ such that

$$\hat{E}_\mu = |u_\mu\rangle\langle u_\mu| \tag{A2}$$

are the POVM operators of an N-valued POVM, $\mu = 1...N$, with $\sum_\mu \hat{E}_\mu = \mathbb{1}$. The vectors $\{|N_\mu\rangle\}$ are defined in the complementary space $\mathcal{K}^\perp$ orthogonal to $\mathcal{K}$, with the total Hilbert space $\mathcal{H} = \mathcal{K} \oplus \mathcal{K}^\perp$. If the dimension of the signal space is n, we have $|N_\mu\rangle = \sum_{i=n+1}^N b_{\mu i}|v_i\rangle$ with some complex coefficients $b_{\mu i}$ and $\{|v_i\rangle\}$ a basis in $\mathcal{K}^\perp$. In the multiple-rail encoding, this leads to an orthogonal set of vectors

$$|w_\mu\rangle = \sum_{j=1}^N U_{\mu j}\hat{a}_j^\dagger|\mathbf{0}\rangle \ , \tag{A3}$$

with a unitary $N \times N$ matrix U having elements $U_{\mu j}$. The application of a linear-optics transformation V to this set (in order to project onto it) can be written as

$$\begin{aligned} |w_\mu\rangle \longrightarrow |w'_\mu\rangle \ &= \ \sum_{j,k=1}^N U_{\mu j}V_{kj}^*\hat{a}_k^\dagger|\mathbf{0}\rangle \\ &= \ \sum_{k=1}^N \delta_{\mu k}\hat{a}_k^\dagger|\mathbf{0}\rangle = \hat{a}_\mu^\dagger|\mathbf{0}\rangle \ , \end{aligned} \tag{A4}$$

choosing $V \equiv U$. As a result, when detecting the outgoing state, for every one-photon click in mode μ, one can unambiguously identify the input state $|w_\mu\rangle$. This is why it is no surprise that any POVM for one-photon states can be implemented via linear optics with unit success probability [34].

For states other than one-photon states, it is a priori not clear whether a given POVM can be implemented with linear optics. A possible approach to deciding this would be to apply the criteria for projective measurements [15] to the orthogonal set in Eq. (6). The main difficulty then is that one must consider any possible Naimark extension vectors $\{|N_\mu\rangle\}$ in order to be able to decide whether the POVM can be implemented or not. In particular, the extension of the signal Hilbert space can be arbitrarily large. Therefore, in an optical implementation, arbitrary ancilla states must be taken into account, including arbitrarily many extra modes and photons. It seems that, in general, more complicated approaches are required to deal with the potentially infinite-dimensional problem of adding arbitrary auxiliary states [13] (however, see [8]). In this paper, we propose a dephasing approach to the problem of implementing POVMs via linear optics, independent of the Naimark extension.

APPENDIX B: ALTERNATIVE DERIVATION OF THE OPTIMAL-USD FIDELITY CRITERION

Without referring to the general principle in Eq. (2) for arbitrary cost functions and POVMs, here we directly derive the corresponding (necessary) criterion for the special case of optimal USD in terms of fidelities.

In general, for any state discrimination scheme based on *static* linear optics, we have the following fidelity bounds,

$$F(\hat{\rho}_+, \hat{\rho}_-) \leq F(\hat{\rho}'_{+,H}, \hat{\rho}'_{-,H}) \leq \left(\mathrm{Prob}_{\mathrm{fail}}^{\mathrm{lin.opt.}}\right)^2 . \tag{B1}$$

In words, the fidelity of the linearly transformed and dephased output states is lower bounded by the fidelity of the input states and upper bounded by the squared failure probability in the linear-optics implementation of unambiguous state discrimination. The lower bound here corresponds to the general rule that the fidelity of two density matrices cannot decrease under CPTP maps [35]. As for the upper bound, we may note that in any scheme, the linearly transformed and dephased output states take on the form of Eq. (13) corresponding to a total dephasing of the states in Eq. (12). Since the two density matrices in Eq. (13) are diagonal in the Fock basis and commute, we have the relation in Eq. (14). Now the failure probability is given by $\mathrm{Prob}_{\mathrm{fail}}^{\mathrm{lin.opt.}} = \sum_m (\mathrm{P}_m^+ + \mathrm{P}_m^-)/2$. However, we also have $(\mathrm{P}_m^+ + \mathrm{P}_m^-)/2 \geq \sqrt{\mathrm{P}_m^+\mathrm{P}_m^-}$, $\forall m$, thus proving the upper bound in Eq. (B1).

According to the fidelity bounds in Eq. (B1), we obtain Eq. (10) as a *necessary condition* for the optimal USD of two states via static linear optics and photon counting, because optimal USD requires $\left(\mathrm{Prob}_{\mathrm{fail}}^{\mathrm{lin.opt.}}\right)^2 = |\langle\chi_+|\chi_-\rangle|^2 = F(\hat{\rho}_+, \hat{\rho}_-)$.

One can now further exploit the fact that the bounds in Eq. (B1) also hold for partially dephased density matrices, corresponding to schemes that include *conditional dynamics*. In particular, the upper bound in Eq. (B1) holds for the partially dephased density matrices as well, because in any mixed-state discrimination scheme, the squared failure probability is lower bounded by the fidelity of the mixed states [21].

APPENDIX C: ALTERNATIVE DERIVATION OF THE USD CONDITIONS FOR A FIXED ARRAY

Let us consider the optimal USD of two pure nonorthogonal states using a *fixed array* of linear optics. We will give an alternative derivation of the conditions in Eq. (24) and Eq. (25), independent of the fidelity criterion in Eq. (10).

After the linear-optics transformation, the output states will always take on the form of Eq. (12), for convenience, written again here,

$$\begin{aligned} |\chi_{+,H}\rangle &= \sum_k \alpha_k |\{k\}\rangle + \sum_m \alpha_m |\{m\}\rangle \\ |\chi_{-,H}\rangle &= \sum_l \beta_l |\{l\}\rangle + \sum_m \beta_m |\{m\}\rangle \,. \end{aligned} \tag{C1}$$

The patterns labeled by k and l are those that unambiguously refer to the + state and to the − state, respectively. Because of the finite overlap of the input states, we must include patterns that occur in the expansion of both states. These ambiguous patterns are denoted by the index m. In general, the amplitudes of the ambiguous N-mode Fock states in the expansions, and hence the probabilities for the corresponding patterns to be detected, may be different for the + and the − state. In the following, we will first prove that in any *optimal* USD scheme, the modulus of the amplitudes of any failure pattern must indeed be equal for both states. Further, we will show that for optimal USD, any relative phases in the expansion of the failure patterns are reduced to a single global phase. As a result, the output states after linear optics in optimal USD must be describable in a three-dimensional vector space such that

$$\begin{aligned} |\chi_{+,H}\rangle &= |\psi^+_{\rm concl}\rangle + |\psi_{\rm inconcl}\rangle \\ |\chi_{-,H}\rangle &= |\psi^-_{\rm concl}\rangle + e^{i\phi}\,|\psi_{\rm inconcl}\rangle \,. \end{aligned} \tag{C2}$$

Here the states $|\psi^+_{\rm concl}\rangle$, $|\psi^-_{\rm concl}\rangle$, and $|\psi_{\rm inconcl}\rangle$ are all mutually orthogonal. They represent the vectors of all conclusive patterns for the + state, of those for the − state, and the vector of all inconclusive patterns, respectively.

As for the proof, we exploit the fact that in optimal USD, the failure probability equals the modulus of the overlap of the states to be discriminated, $\text{Prob}_{\rm fail} = |\langle\chi_+|\chi_-\rangle|$ (assuming equal a priori probabilities). This implies that a linear-optical implementation of optimal USD must satisfy

$$\begin{aligned} \text{Prob}^{\rm lin.opt.}_{\rm fail} &= \frac{1}{2}\sum_m (|\alpha_m|^2 + |\beta_m|^2) \stackrel{!}{=} |\langle\chi_+|\chi_-\rangle| \\ &= |\langle\chi_{+,H}|\chi_{-,H}\rangle| = \left|\sum_m \alpha_m^*\beta_m\right| \,, \end{aligned} \tag{C3}$$

using Eq. (C1). The factor 1/2 in the first line of Eq. (C3) corresponds to the a priori probabilities. Then, because of $|\sum_m \alpha_m^*\beta_m| \le \sum_m |\alpha_m^*\beta_m|$, we also have

$$\begin{aligned} \text{Prob}^{\rm lin.opt.}_{\rm fail} &= \frac{1}{2}\sum_m (|\alpha_m|^2 + |\beta_m|^2) \\ &\le \sum_m |\alpha_m^*\beta_m| \,, \end{aligned} \tag{C4}$$

or,

$$\sum_m (|\alpha_m| - |\beta_m|)^2 \le 0 \,. \tag{C5}$$

The last inequality proves that $|\alpha_m| = |\beta_m|$, $\forall m$. Moreover, it implies that $|\sum_m \alpha_m^*\beta_m| \stackrel{!}{=} \sum_m |\alpha_m^*\beta_m|$, and hence

$$\left|\sum_m |\alpha_m|^2 e^{i\phi_m}\right| \stackrel{!}{=} \sum_m |\alpha_m|^2 \,, \tag{C6}$$

using $\beta_m = \alpha_m e^{i\phi_m}$. However, Eq. (C6) can only be satisfied for $e^{i\phi_m} = e^{i\phi}$, $\forall m$. This concludes the proof of Eq. (C2). For the case of optimal USD, we can now replace Eq. (C1) by

$$\begin{aligned} |\chi_{+,H}\rangle &= \sum_k \alpha_k |\{k\}\rangle + \sum_m \alpha_m |\{m\}\rangle \\ |\chi_{-,H}\rangle &= \sum_l \beta_l |\{l\}\rangle + e^{i\phi}\sum_m \alpha_m |\{m\}\rangle \,. \end{aligned} \tag{C7}$$

Let us now use this result in order to calculate the first-order expression. Similar to Eq. (18), we obtain now

$$\begin{aligned} \langle\chi_+|\hat{c}_j^\dagger\hat{c}_j|\chi_-\rangle &= \langle\chi_{+,H}|\hat{a}_j^\dagger\hat{a}_j|\chi_{-,H}\rangle \\ &= e^{i\phi}\sum_m |\alpha_m|^2 \langle\{m\}|\hat{a}_j^\dagger\hat{a}_j|\{m\}\rangle \,. \end{aligned} \tag{C8}$$

Since analogous expressions can be found for all higher orders, the same arguments as those in the discussion after Eq. (18) apply here again. Thus, finally we obtain the same hierarchies of *necessary* conditions as in Eq. (24) and Eq. (25) for optimal USD using a fixed array of linear optics.

For the special case of exact discrimination of two orthogonal states $|\chi_+\rangle$ and $|\chi_-\rangle$ via photon counting, the linearly transformed states take on the form,

$$\begin{aligned} |\chi_{+,H}\rangle &= \sum_k \alpha_k |\{k\}\rangle \\ |\chi_{-,H}\rangle &= \sum_l \beta_l |\{l\}\rangle \,. \end{aligned} \tag{C9}$$

Now there are no ambiguous patterns in the expansions. Let us again examine the expression

$$\langle \chi_+|\hat{c}_j^\dagger \hat{c}_j|\chi_-\rangle = \langle \chi_{+,H}|\hat{a}_j^\dagger \hat{a}_j|\chi_{-,H}\rangle \,. \tag{C10}$$

According to Eq. (C9), the output states of the linear-optics transformation in exact state discrimination must satisfy $\langle \chi_{+,H}|\hat{a}_j^\dagger \hat{a}_j|\chi_{-,H}\rangle = 0$, because annihilating a photon in the jth mode of the two states only yields a nonzero overlap for coinciding patterns. Similarly, we have

$$\begin{aligned} \langle \chi_+|\hat{c}_j^\dagger \hat{c}_{j'}^\dagger \hat{c}_{j''}^\dagger \cdots \hat{c}_j \hat{c}_{j'} \hat{c}_{j''} \cdots |\chi_-\rangle &= \\ \langle \chi_{+,H}|\hat{a}_j^\dagger \hat{a}_{j'}^\dagger \hat{a}_{j''}^\dagger \cdots \hat{a}_j \hat{a}_{j'} \hat{a}_{j''} \cdots |\chi_{-,H}\rangle &= 0\,, \forall j, j', j'' \ldots , \end{aligned} \tag{C11}$$

because annihilating a photon in the jth, j'th, j''th, etc., mode of the two states also only yields a nonzero overlap for coinciding patterns. Thus, we end up having the following set of conditions for exact state discrimination

$$\begin{aligned} \langle \chi_{+,H}|\hat{a}_j^\dagger \hat{a}_j|\chi_{-,H}\rangle &= 0\,, \quad \forall j\,, \\ \langle \chi_{+,H}|\hat{a}_j^\dagger \hat{a}_{j'}^\dagger \hat{a}_j \hat{a}_{j'}|\chi_{-,H}\rangle &= 0\,, \quad \forall j, j'\,, \\ \langle \chi_{+,H}|\hat{a}_j^\dagger \hat{a}_{j'}^\dagger \hat{a}_{j''}^\dagger \hat{a}_j \hat{a}_{j'} \hat{a}_{j''}|\chi_{-,H}\rangle &= 0\,, \quad \forall j, j', j''\,, \\ \vdots \quad &= \quad \vdots \end{aligned} \tag{C12}$$

or, equivalently, as described in Eq. (3).

[1] S. Lloyd and S. L. Braunstein, Phys. Rev. Lett. **82**, 1784 (1999).
[2] K. Nemoto and W. J. Munro, Phys. Rev. Lett. **93**, 250502 (2004).
[3] E. Knill, R. Laflamme, and G. J. Milburn, Nature **409**, 46 (2001).
[4] J. D. Franson *et al.*, Phys. Rev. Lett. **89**, 137901 (2002).
[5] M. A. Nielsen, Phys. Rev. Lett. **93**, 040503 (2004).
[6] D. E. Browne and T. Rudolph, Phys. Rev. Lett. **95**, 010501 (2005).
[7] T. B. Pittman, B. C. Jacobs, and J. D. Franson, Phys. Rev. A **64**, 062311 (2001).
[8] E. Knill, Phys. Rev. A **68**, 064303 (2003).
[9] E. Knill, Phys. Rev. A **66**, 052306 (2002).
[10] T. C. Ralph *et al.*, Phys. Rev. A **65**, 012314 (2001).
[11] S. Scheel *et al.*, Phys. Rev. A **68**, 032310 (2003).
[12] S. Scheel and N. Lütkenhaus, New Journal of Physics **6**, 51 (2004).
[13] J. Eisert, Phys. Rev. Lett. **95**, 040502 (2005).
[14] J. Calsamiglia, Phys. Rev. A **65**, 030301(R) (2002).
[15] P. van Loock and N. Lütkenhaus, Phys. Rev. A **69**, 012302 (2004).
[16] P. van Loock, P. Raynal, and N. Lütkenhaus, unpublished.
[17] L. Vaidman and N. Yoran, Phys. Rev. A **59**, 116 (1999).
[18] N. Lütkenhaus, J. Calsamiglia, and K.A. Suominen, Phys. Rev. A **59**, 3295 (1999).
[19] J. Calsamiglia and N. Lütkenhaus, Appl. Phys. B **72**, 67 (2001).
[20] The statement in Eq. (2) can also be expressed in terms of "score functions" S, where $S = 1 - C$. An example for a possible score function is the mean fidelity in quantum state estimation. For the quantum mechanically optimal estimation of a given quantum state drawn from a continuous or discrete set of states, the "score" is then expressed in terms of the mean value of the fidelity $F(\hat{\rho}_\mu, \hat{\rho})$ between the actual input state $\hat{\rho}$ and the estimated state $\hat{\rho}_\mu$ depending on the result of a POVM $\hat{E}_\mu$. This mean value of $F(\hat{\rho}_\mu, \hat{\rho})$ can be calculated via $\bar{F} = \sum_{\mu=1}^N \left\langle \mathrm{Tr}(\hat{E}_\mu \hat{\rho}) F(\hat{\rho}_\mu, \hat{\rho}) \right\rangle$, where $\mathrm{Tr}(\hat{E}_\mu \hat{\rho})$ is the probability for obtaining the result μ when the actual input state was $\hat{\rho}$, and $\langle \cdots \rangle$ denotes the average over the entire alphabet of input states. An optimal estimation scheme achieves the maximum average fidelity $\bar{F}$, maximized with respect to the chosen strategy given by the set $\{\hat{E}_\mu, \hat{\rho}_\mu\}_{\mu=1}^N$.
[21] T. Rudolph, R. W. Spekkens, and P. S. Turner, Phys. Rev. A **68**, 010301(R) (2003); P. Raynal and N. Lütkenhaus, Phys. Rev. A **72**, 022342 (2005).
[22] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
[23] M. Takeoka, M. Sasaki, P. van Loock, and N. Lütkenhaus, Phys. Rev. A **71**, 022318 (2005).
[24] Another optimal POVM solution to estimating an unknown qubit state is the covariant POVM that consists of a continuous set of projectors on any possible qubit state (projecting on the "spin-coherent states"). Since this POVM solution is infinite, it is not very practical for an implementation. However, note that there are infinite POVMs that can be implemented. An example is the optimal estimation of an unknown coherent state $|\alpha\rangle$. The elements of the optimal infinite POVM solution in this case are the coherent-state projectors $|\beta\rangle\langle\beta|$. Its implementation only requires a symmetric beam splitter and two homodyne detectors, corresponding to a heterodyne detection or a measurement of the Q function of the input state. The Naimark extension here is due to the extra vacuum mode at the second input port of the beam splitter.
[25] I. D. Ivanovic, Phys. Lett. A **123**, 257 (1987).
[26] D. Dieks, Phys. Lett. A **126**, 303 (1988).
[27] A. Peres, Phys. Lett. A **128**, 19 (1988).
[28] Even for more than two input states, quantum mechanically, such a POVM always exists, i.e., there is a nonzero probability for a conclusive measurement result, provided the nonorthogonal input states are linearly independent [29]. Conversely, a given POVM is equivalent to the discrimination of linearly independent signal states, if the POVM elements can be written as rank one projectors onto the reciprocal vectors of the signal states [30].
[29] A. Chefles, Phys. Lett. A **239**, 339 (1998).
[30] A. Chefles, private communication.
[31] Obviously, for only partially dephased density matrices describing schemes that potentially include conditional dynamics where only a first mode is detected, the fidelity criterion in Eq. (10) cannot be sufficient for optimal USD.
[32] S. J. van Enk, Phys. Rev. A **66**, 042313 (2002).

[33] M. Reck *et al.*, Phys. Rev. Lett. **73**, 58 (1994).
[34] S. E. Ahnert and M. C. Payne, Phys. Rev. A **69**, 012312 (2004).
[35] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press (2000).