Higher-Order Process Matrix Tomography of a Passively-Stable Quantum Switch

Michael Antesberger,[1,*] Marco Túlio Quintino,[2] Philip Walther,[1,3,4] and Lee A. Rozema[1]

[1] *Vienna Center for Quantum Science and Technology and Research Network Quantum Aspects of Space Time (TURIS), Faculty of Physics, University of Vienna, 1090 Vienna, Austria*
[2] *Sorbonne Université, CNRS, LIP6, 75005 Paris, France*
[3] *Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Boltzmanngasse 3, 1090 Vienna, Austria*
[4] *Christian Doppler Laboratory for Photonic Quantum Computer, Faculty of Physics, University of Vienna, 1090 Vienna, Austria*

(Received 1 June 2023; accepted 22 December 2023; published 9 February 2024)

The field of indefinite causal order (ICO) has seen a recent surge in interest. Much of this research has focused on the quantum switch, wherein multiple parties act in a superposition of different orders in a manner transcending the quantum circuit model. This results in a new resource for quantum protocols, and is exciting for its relation to issues in foundational physics. The quantum switch is also an example of a higher-order quantum operation, in that it transforms not only quantum states but also other quantum operations. To date, no quantum process without a definite causal order has been completely experimentally characterized. Indeed, past work on the quantum switch has confirmed its ICO by measuring causal witnesses or demonstrating resource advantages, but the complete process matrix has been described only theoretically. Here we report our performing higher-order quantum process tomography. However, doing so requires exponentially many measurements with a scaling worse than that of standard process tomography. We overcome this challenge by creating a new *passively stable* fiber-based quantum switch using active optical elements to deterministically generate and manipulate time-bin encoded qubits. Moreover, our new architecture for the quantum switch can be readily scaled to multiple parties. By reconstructing the process matrix, we estimate its fidelity and tailor different causal witnesses directly for our experiment. To achieve this, we measure a set of tomographically complete settings, which also spans the input operation space. Our tomography protocol allows the characterization and debugging of higher-order quantum operations with and without an ICO, while our experimental time-bin techniques could enable the creation of a new realm of higher-order quantum operations with an ICO.

DOI: 10.1103/PRXQuantum.5.010325

I. INTRODUCTION

The formalism of higher-order quantum operations (HOQOs) provides a framework to view quantum operations as objects that can be subjected to transformations [1–6]. This framework is particularly useful for analyzing causality in quantum mechanics. Since it was first realized that quantum mechanics allows for processes with an indefinite causal order (ICO) [6–8], the field of quantum causality has seen an increasing level of interest [9]. These processes are interesting for a variety of foundational topics [10–15], and also because it has been recognized that they can lead to ICO-based enhancements that go even beyond “normal quantum technology.” Examples of these advantages include applications in quantum computing [16–19], quantum communication [20–28], channel discrimination [29,30], metrology [31], reversing quantum dynamics [32,33], and even thermodynamics [34]. Experimentally, work has focused on either implementing various protocols [35–42] or verifying the ICO of a given experimental implementation [43–46].

In spite of this large body of work, there has not yet been a complete experimental characterization of processes with an ICO. Instead, previous work on ICOs focused mainly on designing and measuring witnesses to essentially provide a yes or no answer to the question “does this process have an indefinite causal order?” On the one hand, this is because no concrete protocol for a complete characterization has yet been presented. On the other hand, it is because the

---

[*] michael.antesberger@univie.ac.at

number of experimental settings required for a complete characterization was prohibitive in past experiments. Here we overcome both these hurdles, first presenting a protocol to perform "higher-order process matrix tomography," and then implementing a new experimental method to realize the quantum switch based on time qubits, based on the proposal in Ref. [47]. Our new passively stable implementation is based on active optical elements, and it allows us to acquire sufficient data (estimating almost 10 000 distinct probabilities) to fully reconstruct a process matrix demonstrating an ICO for the first time.

In the two-party quantum switch, Alice and Bob each act on a target system (typically taken to be a qubit) in their local laboratories. This target qubit is sent first to one party and then to the other. The order in which the target qubit is shared between the two is coherently controlled via a second, *control* qubit. If the control qubit is prepared in a superposition state, then the two parties act on the target qubit in a superposition of orders [Figs. 1(a)– 1(c)]. The quantum switch and the so-called quantum time flip [48–50] are the only processes that do not respect our standard notions of causality that have been experimentally implemented to date. The quantum switch is an example of a HOQO, in the sense that its inputs are not only the control and target qubits but are also Alice's and Bob's operations.

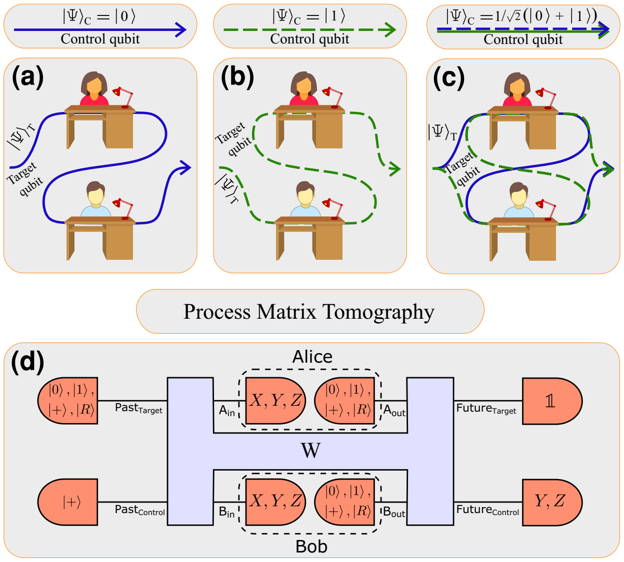

FIG. 1. The quantum switch. (a),(b)) The causally ordered process with the control in state $|0\rangle_C$ ($|1\rangle_C$), where Alice (Bob) acts before Bob (Alice) on the target system. (c) A superposition of orders with the control qubit in the state $\frac{1}{\sqrt{2}}\,|0\rangle_C + |1\rangle_C$. (d) The principle of the process tomography on the quantum switch. Alice and Bob perform projective measurements in three different bases and then prepare four different linearly independent states in their output. The same input states are prepared at the past target. The past control is fixed to the superposition state for generation of the indefinite order. Finally, the future control is measured in a different basis, while the future target is traced out (the two measurements shown are those that are implemented experimentally; to span the space, a third measurement is needed).

All experimental realizations of the quantum switch have been accomplished by encoding both the control system and the target system in a single photon. Typically, the control system is encoded in a path degree of freedom, which then determines the order in which the photon is routed between the two parties. In practice, this means that a photon is placed in a superposition of the two paths with use of a beam splitter, and these paths are then looped between two parties in a manner mimicking the paths in Fig. 1(c). The parties then act on a different degree of freedom, such as polarization [35], time bins [37], or orbital angular momentum [44]. The result of all these approaches is essentially a Mach-Zehnder interferometer (MZI), which must be phase stabilized [51]. Stabilizing the phase for long enough to acquire the required data for full higher-order quantum process tomography is a daunting experimental challenge.

We overcome this challenge by implementing a new passively stable quantum switch. In our experiment, the control system is encoded in a time-bin qubit, the target qubit is encoded in the polarization of the same photon, and active optical switches are used to route the photon between the two parties in superposition of both orders, as in the theoretical proposal in Ref. [47]. Two other experiments have achieved an intrinsically stable phase, using a Sagnac-like approach [37,52], and in the related studies, the propagation direction acts as the control system. Since there are only two propagation directions, the control system can only be a qubit, making it difficult to imagine scaling these methods to multiple parties. Moreover, the operations on the target must act identically, regardless of the propagation directions. As shown in Ref. [52], this is not straightforward to achieve even for unitary operations when the target system is a polarization qubit. Our approach, however, has straightforward generalization to multiple parties [47] and allows us to use standard polarization operations (both unitary and nonunitary), making it a promising new experimental method to create an ICO.

Recently, there has been some discussion in the community as to whether such photonic implementations of the quantum switch are simulations of an ICO [53,54], with some concluding that they may only have an ICO in a "weak sense" [55], while others have concluded that the experiments do have an ICO [56,57], or that they at least have a quantifiable resource advantage [58]. Here we do not address this debate, but we use the mathematical formalism of process matrices and HOQOs to describe our physical experiment.

One method to certify ICO is made via the violation of a so-called causal inequality [6]. This is a device-independent technique, similar to the use of a Bell violation to verify entanglement [59]. Unfortunately, it is not yet known if one can implement a quantum process that deterministically violates a causal inequality; moreover, it has been shown that the quantum switch cannot violate causal inequalities [17]. Instead, in the first implementation of the quantum switch [35], the ICO was indirectly proven by means of a game, where a player has to decide if two unitary gates either commute or anticommute (see Appendix B 1). By winning that game more often than one could with a definite causal order, the study authors concluded the experiment did not have a definite causal order. This method can be reframed in terms of a causal witness [13]. A causal witness is a measurement that can be used to verify if a process is causally nonseparable (i.e., if it has an ICO), and it has been experimentally implemented for the quantum switch [44,45]. Unlike a causal inequality, a causal witness is not device independent, requiring the assumption that the experimenter knows the correct quantum description of the experiment. Recently, progress has been made by relaxing the complete device-independent approach, allowing the certification of causal nonseparability under semi-device-independent assumptions [46,60,61], Bell-locality-like assumptions [15,45], and additional device-independent no-signaling assumptions [62,63].

In this work, we implement full experimental higher-order process tomography of the quantum switch. For this goal, we generalize the ideas from quantum state and quantum process tomography [2,4,64–70] to tackle tomography of arbitrary higher-order processes, including those without a definite causal order. In particular, we show that it is possible to construct tomographically complete measurement settings on arbitrary quantum processes by using a tomographically complete set of input states spanning the input state space; a tomographically complete set of measurement-repreparation channels spanning the input operation spaces; and a tomographically complete set of quantum measurements spanning the output state space [see Fig. 1(d)]. We then use these ideas to experimentally perform higher-order quantum process tomography on the quantum switch.

The rest of this paper is organized as follows. In Sec. II we introduce the theory of quantum process matrices, using the quantum switch as a paradigmatic example, and we present our causal tomography protocol. In Sec. III, we discuss our new passively stable architecture for the quantum switch. In Sec. IV we present our experimental results, and in Sec. V we discuss our findings.

II. THEORY

A. Process matrices and the quantum switch

The expression "quantum process" is a general term used to refer to the dynamics of quantum systems, and its precise meaning may depend on the context. For instance, when one is analyzing transformations between quantum states, "quantum process" refers to a quantum channel, which may be unitary (associated with closed quantum systems and mathematically described by unitary operators) or nonunitary (associated with open quantum systems and mathematically described by completely positive trace-preserving linear maps). In this scenario of transformations between states, one can experimentally determine the dynamics by means of what is known as quantum process tomography [65–67]. To do so, a complete set of known quantum states is fed into an unknown quantum process $\mathcal{E}$ and a complete set of measurements is performed on the output of the underlying process for each input state [70]. When performing standard process tomography, one reconstructs, for example, the χ matrix, which takes quantum sates as inputs and returns quantum states as outputs [71]. The χ matrix is often called a "process matrix"; however, we stress that this χ or process matrix is different from the process matrices discussed in the field of HOQO and ICO, the case that we address here.

An operation that transforms a quantum operation is referred to as a higher-order quantum operation. Causally ordered higher-order operations have appeared in different contexts under different names, i.e., quantum strategies [3], quantum combs [2], and non-Markovian quantum process [72], and have found applications in several branches of quantum information processing [4,32,73–75]. Interestingly, the formalism of higher-order operations also includes quantum processes that do not respect any definite causal order [6,8,29], such as the notorious quantum switch [8,29], the main object analyzed in this work. The quantum switch is a process that transforms a pair of unitary operators (U_A, U_B) into another operator that is a coherent superposition of the composition $U_A U_B$ and $U_B U_A$. In mathematical terms, the quantum switch is the transformation

$$(U_A, U_B) \mapsto |0\rangle\langle 0| \otimes U_A U_B + |1\rangle\langle 1| \otimes U_B U_A, \qquad (1)$$

where the first system on the rhs of Eq. (1) is referred to as the control system, since the order in which U_A and U_B will be performed may be controlled by setting the control qubit state. The second system is referred to as the target system, since it is the system on which the unitary operators act. The action of the quantum switch on general nonunitary quantum operations is uniquely determined by its action on unitary operations [76] and may be described via its actions on Kraus operators [71], or via the process matrix formalism discussed in the next section. Let Λ_A and Λ_B be quantum operations that act on density matrices as $\Lambda_A(\rho) = \sum_i K_i \rho K_i^\dagger$ and $\Lambda_B(\rho) = \sum_j L_j \rho L_j^\dagger$. The quantum switch then transforms the pair of operations Λ_A and Λ_B into another one with Kraus operators given by

$$W_{ij} = |0\rangle\langle 0| \otimes K_i L_j + |1\rangle\langle 1| \otimes L_j K_i. \qquad (2)$$

In this work, we perform experimental tomography of the quantum switch and analyze the robustness of its indefinite causality in a real experimental setup. For this goal, we build on known mathematical methods for causally ordered process tomography [65,68,69,72,77] that are used for higher-order causal process tomography experiments [78,79]) and present a simple but general tomographic procedure by explicitly presenting what we refer to as tomographically complete setting operators. Our setting operator approach is particularly convenient to analyze causal nonseparability witnesses with additional restrictions, the method we use to certify that the process we characterize does not respect any definite causal order, and that this noncausal property is robust with regard to white noise and general noise.

B. Choi-Jamiołkowski isomorphism and quantum operations

Higher-order transformations such as the quantum switch may be conveniently described by means of a *process matrix* [6], a formalism that is heavily based on the Choi-Jamiołkowski isomorphism [80,81], a method to represent linear maps as linear operators, and linear operators as vectors. Let $\mathcal{H}_{\text{in}}$ and $\mathcal{H}_{\text{out}}$ be finite linear (Hilbert) spaces associated with the input and the output. Let $U:\mathcal{H}_{\text{in}}\to\mathcal{H}_{\text{out}}$ be a linear operator. Its process vector $|U\rangle\!\rangle\in\mathcal{H}_{\text{in}}\otimes\mathcal{H}_{\text{out}}$ is then defined as [82]

$$|U\rangle\!\rangle^{\mathcal{H}_{\text{in}}\mathcal{H}_{\text{out}}}:=\sum_i |i\rangle\otimes U|i\rangle\,, \tag{3}$$

where $\{|i\rangle\}_i$ is the computational basis. The Choi vector of the identity operator is given by

$$|\mathbb{1}\rangle\!\rangle=\sum_i |i\rangle\otimes|i\rangle\,, \tag{4}$$

which is equivalent to a maximally entangled state up to normalization.

Let $\mathcal{L}(\mathcal{H})$ be the set of all linear operators acting on $\mathcal{H}$. Let $\mathcal{C}:\mathcal{L}(\mathcal{H}_{\text{in}})\to\mathcal{L}(\mathcal{H}_{\text{out}})$ be a linear map. Its Choi operator $C\in\mathcal{L}(\mathcal{H}_{\text{in}}\otimes\mathcal{H}_{\text{out}})$ is defined as

$$C^{\mathcal{H}_{\text{in}}\mathcal{H}_{\text{out}}}:=\sum_{ij}|i\rangle\langle j|\otimes\mathcal{C}(|i\rangle\langle j|). \tag{5}$$

The action of any linear map $\mathcal{C}$ on a state ρ can be written in terms of the Choi operator C as

$$\mathcal{C}(\rho)=\text{Tr}_{\text{in}}\left(\rho^{T^{\mathcal{H}_{\text{in}}}}\otimes\mathbb{1}^{\mathcal{H}_{\text{out}}}\,C^{\mathcal{H}_{\text{in}}\mathcal{H}_{\text{out}}}\right), \tag{6}$$

where ρ^T is the transpose of ρ in the computational basis and ρ is an arbitrary density operator acting on $\mathcal{H}_{\text{in}}$.

The Choi-Jamiołkowski isomorphism is very useful to represent quantum operations, because a linear map $\mathcal{C}:\mathcal{L}(\mathcal{H}_{\text{in}})\to\mathcal{L}(\mathcal{H}_{\text{out}})$ is completely positive if and only if its Choi operator $C\in\mathcal{L}(\mathcal{H}_{\text{in}}\otimes\mathcal{H}_{\text{out}})$ respects $C\geq 0$, and the map $\mathcal{C}$ is trace preserving if and only if $\text{Tr}_{\text{out}}(C)=\mathbb{1}^{\mathcal{H}_{\text{in}}}$. Since quantum channels are completely positive trace-preserving maps, all quantum channels have a simple and direct characterization in terms of their Choi operators. Before finishing this subsection, we also remark that if $\mathcal{C}$ is a unitary channel, that is, $\mathcal{C}(\rho)=U\rho U^\dagger$ for some unitary operator U, direct calculation shows that its Choi operator may be written as $C=|U\rangle\!\rangle\langle\!\langle U|$, where $|U\rangle\!\rangle$ is defined in Eq. (3).

A quantum instrument is a quantum operation that has a classical and a quantum output, and it formalizes the concept of a quantum measurement that has a postmeasurement quantum state. Mathematically, a set of linear maps $\{\mathcal{C}_i\}_i$, $\mathcal{C}_i:\mathcal{L}(\mathcal{H}_{\text{in}})\to\mathcal{L}(\mathcal{H}_{\text{out}})$ is a quantum instrument if all $\mathcal{C}_i$ are completely positive and $\mathcal{C}:=\sum_i\mathcal{C}_i$ is trace preserving. In the Choi operator picture, this is equivalent to having $C_i\geq 0$ and $\text{Tr}_{\text{out}}\left(\sum_i C_i\right)=\mathbb{1}^{\mathcal{H}_{\text{in}}}$. A simple and useful class of quantum instruments is the class of *measure-and-reprepare* instruments. In its most basic form, a measure and reprepare instrument simply performs a measurement described by the operators $\{M_i\}$ [83], and reprepares some fixed state σ. Its linear map is described by $\mathcal{R}_i(\rho):=\text{Tr}(\rho M_i)\sigma$, and its Choi operators are given by $R_i\in\mathcal{L}\left(\mathcal{H}_{\text{in}}\otimes\mathcal{H}_{\text{out}}\right)$, with $R_i=M_i^T\otimes\sigma$.

C. The quantum switch as a process matrix

We are now in position to present process matrices that describe transformations between the quantum channels of different parties. We start by presenting the process vector describing Fig. 1(a), which is simply a process where a system flows freely from a common past target space $\mathcal{H}_{P_{\text{T}}}$ to Alice's input space $\mathcal{H}_{A_{\text{in}}}$. Alice may perform an arbitrary operation as the state goes from $\mathcal{H}_{A_{\text{in}}}$ to $\mathcal{H}_{A_{\text{out}}}$. Later, the state goes freely from Alice's output space $\mathcal{H}_{A_{\text{out}}}$ to Bob's input space $\mathcal{H}_{B_{\text{in}}}$. Bob may then perform an arbitrary operation as the state goes from $\mathcal{H}_{B_{\text{in}}}$ to $\mathcal{H}_{B_{\text{out}}}$. Finally, the state goes freely from Bob's output space $\mathcal{H}_{B_{\text{out}}}$ to a common future target space $\mathcal{H}_{F_{\text{T}}}$. The process vector of this quantum process is

$$|A\to B\rangle:=|\mathbb{1}\rangle\!\rangle^{P_{\text{T}},A_{\text{in}}}\otimes|\mathbb{1}\rangle\!\rangle^{A_{\text{out}},B_{\text{in}}}\otimes|\mathbb{1}\rangle\!\rangle^{B_{\text{out}},F_{\text{T}}}. \tag{7}$$

Notice that although the vector $|A\to B\rangle$ represents a process, we do not use the double-ket notation, since $|A\to B\rangle$ is not the Choi vector of "$A\to B$." We should understand $|A\to B\rangle$ simply as convenient notation for $|\mathbb{1}\rangle\!\rangle^{P_{\text{T}},A_{\text{in}}}\otimes|\mathbb{1}\rangle\!\rangle^{A_{\text{out}},B_{\text{in}}}\otimes|\mathbb{1}\rangle\!\rangle^{B_{\text{out}},F_{\text{T}}}$.

By examining the superscripts, one can see that this process vector represents the flow of information (without any applied operations) in three steps: first, from the past target space to Alice's input space, then from Alice's output space to Bob's input space, and finally from Bob's output space to the future target space. As we show soon, the connections between Alice's (Bob's) input and output spaces are made by Choi operators acting on the respective space, which can be quantum channels or instruments. The process matrix in Eq. (7) is given by

$$W_{A\to B} := |A\to B\rangle\,\langle A\to B|\,, \tag{8}$$

where $A\to B$ indicates that Alice acts before Bob. We have not included Alice's or Bob's operations in this description; this is introduced in Sec. II E. Note that in the superscripts of the process vector in Eq. (7) (and in subsequent labeling of the Hilbert spaces), we have dropped $\mathcal{H}$. For example, A_{in} corresponds to the Hilbert space $\mathcal{H}_{A_{\text{in}}}$. Analogously, we may define the process where Bob acts before Alice, which will lead to a process vector

$$|B\to A\rangle := |\mathbb{1}\rangle\rangle^{P_{\text{T}},B_{\text{in}}} \otimes |\mathbb{1}\rangle\rangle^{B_{\text{out}},A_{\text{in}}} \otimes |\mathbb{1}\rangle\rangle^{A_{\text{out}},F_{\text{T}}}, \tag{9}$$

and its process matrix is given by

$$W_{B\to A} := |B\to A\rangle\,\langle B\to A|\,. \tag{10}$$

The quantum switch is a process that allows one to coherently alternate between $|A\to B\rangle$ and $|B\to A\rangle$. For that, we allow the common past and common future to have another system, denoted as a control system, which will be able to coherently alternate between the ordered processes. More formally, the common past space and common future space are now described by $\mathcal{H}_P = \mathcal{H}_{P_{\text{C}}} \otimes \mathcal{H}_{P_{\text{T}}}$ and $\mathcal{H}_F = \mathcal{H}_{F_{\text{C}}} \otimes \mathcal{H}_{F_{\text{T}}}$, respectively, and the Choi vector of the quantum switch is given by

$$\begin{aligned}|w_{\text{switch}}\rangle := {}& |0\rangle^{P_{\text{C}}} \otimes |A\to B\rangle \otimes |0\rangle^{F_{\text{C}}} \\ & + |1\rangle^{P_{\text{C}}} \otimes |B\to A\rangle \otimes |1\rangle^{F_{\text{C}}}\,,\end{aligned} \tag{11}$$

which corresponds to the process matrix

$$W_{\text{switch}} := |w_{\text{switch}}\rangle\,\langle w_{\text{switch}}|\,. \tag{12}$$

Almost all known applications of the quantum switch, e.g., computational advantages [17], channel discrimination [8], reducing communication complexity [21], semi-device-independent certification [60] and device-independent certification [62,84] of indefinite causality, do not require the general form of the quantum switch as presented in Eq. (12). Rather, in such applications, one starts with the control qubit in the $|+\rangle_{\text{C}} := (|0\rangle_{\text{C}} + |1\rangle_{\text{C}})/\sqrt{2}$ state, so that the process state corresponds to a coherent superposition of processes described by

$$\left|w^{+}_{\text{switch}}\right\rangle := \frac{|A\to B\rangle \otimes |0\rangle^{F_{\text{C}}} + |B\to A\rangle \otimes |1\rangle^{F_{\text{C}}}}{\sqrt{2}}. \tag{13}$$

Additionally, for all such applications, one does not use the future target system; hence, this qubit is often discarded. Mathematically, discarding a system corresponds to the partial trace. Hence, we construct the simplified version of the switch as

$$W^{+}_{s} := \text{Tr}_{F_{\text{T}}}\left(\left|w^{+}_{\text{switch}}\right\rangle\left\langle w^{+}_{\text{switch}}\right|\right). \tag{14}$$

In this work, we focus on the simplified quantum switch, and as is usual in the literature, we use "quantum switch" to refer to the process described in Eq. (14).

D. The general process matrix formalism

The process matrix formalism allows one to assign a matrix that perfectly describes transformations between arbitrary quantum objects, in particular, to transform quantum channels into quantum channels. The normalization constraints from quantum channels (or more general quantum objects) and a generalized notion of completely positive inputs lead to constraints on valid process matrices. In a nutshell, when focused on quantum channels, a matrix W is a process matrix if it is positive semidefinite and respects a set of affine constraints arising from the channel normalization conditions. These affine constraints are described in references such as Refs. [6,17] and may be viewed as causal constraints (for instance, they prevent local loops or the possibility of obtaining negative probabilities via Born's rule).

E. Measuring a process matrix

One of the main applications of the process matrix formalism is to provide mathematical methods to analyze the dynamics of a quantum process and to predict the outcomes of measurements performed on a quantum process. Below, we describe the scenario considered in this work, which is graphically represented in Fig. 2:

(1) $W \in \mathcal{L}\left(\mathcal{H}_P \otimes \mathcal{H}_{A_{\text{in}}} \otimes \mathcal{H}_{A_{\text{out}}} \otimes \mathcal{H}_{B_{\text{in}}} \otimes \mathcal{H}_{B_{\text{out}}} \otimes \mathcal{H}_F\right)$ is the process matrix that describes a bipartite scenario with a common past (target) and common future (control).

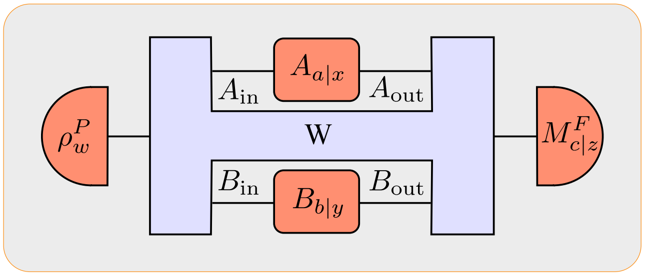

FIG. 2. Probing a quantum process W. Illustration on how to probe a bipartite quantum process W with a common past and a common future. Here ρ_w are quantum states, $\{A_{a|x}\}$ and $\{B_{b|y}\}$ are quantum instruments, and $\{M_{c|z}\}$ are quantum measurements.

(2) $\rho \in \mathcal{L}(\mathcal{H}_P)$ is a quantum state on the common past (target) space.
(3) $A_a \in \mathcal{L}(\mathcal{H}_{A_{\text{in}}} \otimes \mathcal{H}_{A_{\text{out}}})$ are the Choi operators of an instrument on Alice's space.
(4) $B_b \in \mathcal{L}(\mathcal{H}_{B_{\text{in}}} \otimes \mathcal{H}_{B_{\text{out}}})$ are the Choi operators of an instrument on Bob's space.
(5) $M_c \in \mathcal{L}(\mathcal{H}_F)$ are the measurement operators on the common future (control) space.

In the scenario described above, if W is the process matrix, one inputs the state ρ into the common past, Alice performs the instrument $\{A_a\}$, Bob performs the instrument $\{B_b\}$, the measurement $\{M_c\}$ is performed in the future, and the probability that Alice obtains the outcome a, Bob obtains the outcome b, and the future obtains the outcome c is given by

$$p(a,b,c) = \text{Tr}\left(W\left(\rho^P \otimes A_a^{A_{\text{in}}A_{\text{out}}} \otimes B_b^{B_{\text{in}}B_{\text{out}}}\right)^T \otimes M_c^F\right). \tag{15}$$

In practice, it is often convenient to have indices to label states, instruments, and measurements. In this work, we use $\{\rho_w\}$ to denote a set of states acting in the common past, $\{A_{a|x}\}$ for a set of instruments in Alice's space (a labels the classical outcome of the instrument and x labels the choice of instrument), $\{B_{b|y}\}$ for a set of instruments in Bob's space (b labels the classical outcome of the instrument and y labels the choice of instrument), and $\{M_{c|z}\}$ for a set of measurements in the future space (c labels the classical outcome and z labels the choice of measurement). We can then define the *setting operators* [85] as

$$S_{abc|xyzw} := \left(\rho_w^P \otimes A_{a|x}^{A_{\text{in}}A_{\text{out}}} \otimes B_{b|y}^{B_{\text{in}}B_{\text{out}}}\right)^T \otimes M_{c|z}^F, \tag{16}$$

which leads us to the so-called generalized Born's rule:

$$p\,(abc|xyzw) = \text{Tr}\left(W S_{abc|xyzw}\right). \tag{17}$$

The setting operator approach allows us to present the experimental setup in a compact and explicit manner, and it is convenient for data analysis. For instance, it allows us to easily include particular restrictions to general causal nonseparability witnesses, as we detail in Sec. IV B.

F. Process matrix tomography

The goal of quantum tomography is to completely characterize a quantum object by performing known measurements on it. Before discussing process matrix tomography, we revisit the standard case of quantum state tomography, where one aims to characterize an unknown state by analyzing the outcomes obtained after performing known measurements on it. If $M_{a|x} \in \mathcal{L}(\mathbb{C}_d)$ are known measurement operators, one can use the probabilities $p(a|x) = \text{Tr}(\rho\, M_{a|x})$ to uniquely reconstruct the unknown state ρ. When the set of operators $\{M_{a|x}\}$ spans the linear space of $\mathcal{L}(\mathbb{C}_d)$, the operator ρ may be obtained via $p(a|x)$ by standard linear inversion methods.

For qubit states, a standard set of tomographically complete measurements is formed by the three Pauli observables X, Y, and Z, which are associated with the measurement operators via their eigenprojectors: $\{\ket{+}\bra{+}, \ket{-}\bra{-}\}$, $\{\ket{y_+}\bra{y_+}, \ket{y_-}\bra{y_-}\}$, $\{\ket{0}\bra{0}, \ket{1}\bra{1}\}$, respectively, where $\ket{\pm} = (\ket{0} \pm \ket{1})/\sqrt{2}$ and $\ket{y_\pm} = (\ket{0} \pm i\ket{1})/\sqrt{2}$. In particular, the standard measurement operators from the set

$$\mathcal{S} := \{\ket{\psi_i}\bra{\psi_i}\}_{i=1}^4, \tag{18}$$

where

$$\ket{\psi_1}\bra{\psi_1} := \ket{0}\bra{0}, \quad \ket{\psi_2}\bra{\psi_2} := \ket{1}\bra{1}, \tag{19}$$
$$\ket{\psi_3}\bra{\psi_3} := \ket{+}\bra{+}, \ket{\psi_4}\bra{\psi_4} := \ket{y_+}\bra{y_+}. \tag{20}$$

These measurements are linearly independent, forming a (nonorthonormal) basis for $\mathcal{L}(\mathbb{C}_2)$.

We now consider the task of performing tomography of a qubit channel. As discussed in Sec. II B, every quantum channel $\mathcal{C} : \mathcal{L}(\mathcal{H}_{\text{in}}) \to \mathcal{L}(\mathcal{H}_{\text{out}})$, can be represented by its Choi operator $C \in \mathcal{L}(\mathcal{H}_{\text{in}} \otimes \mathcal{H}_{\text{out}})$. In this case, tomography can be performed by one preparing a set of states $\{\rho_w\}_w$, $\rho_w \in \mathcal{L}(\mathcal{H}_{\text{in}})$ and performing a complete set of measurements on each state. For qubits, the standard measurements are

$$\mathcal{M} := \{M_{i|j}\}_{i=1,j=1}^{i=2,j=3}, \tag{21}$$

where $M_{i|j}$ is a positive-operator-valued measure (POVM) element, the label i stands for the outcomes, and the label j stands for the choice of measurements. Hence, $\mathcal{M}$ is a set with three dichotomic measurements with POVM elements given by

$$M_{1|1} := \ket{0}\bra{0}, \quad M_{2|1} := \ket{1}\bra{1}, \tag{22}$$
$$M_{1|2} := \ket{+}\bra{+}, \quad M_{2|2} := \ket{-}\bra{-}, \tag{23}$$
$$M_{1|3} := \ket{y_+}\bra{y_+}, M_{2|3} := \ket{y_-}\bra{y_-}. \tag{24}$$

Note that, because of the normalization of probabilities, the measurements of some measurement operators are unnecessary. However, in practice, the orthogonal measurements $M_{2|j}$ are often measured to aid in the data normalization. We include them here, with a view to our experiment.

From these input states and measurements, one estimates the probabilities $p(a|x,w) = \text{Tr}\Big(\mathcal{C}(\rho_w)\, M_{a|x}\Big)$, which can also be written as

$$p\,(a|x,w) = \mathrm{Tr}\Big(C\,\left(\rho_w^T \otimes M_{a|x}\right)\Big) \tag{25}$$

$$= \mathrm{Tr}\left(C\,S_{a|xw}\right) \tag{26}$$

in the Choi formalism, where $S_{a|xw} := \rho_w^T \otimes M_{a|x}$, where $S_{a|xw}$ is a setting operator for standard quantum process tomography. One way to perform complete tomography is by ensuring that the setting operators $S_{a|xw}$ span the space $\mathcal{L}(\mathcal{H}_{\mathrm{in}} \otimes \mathcal{H}_{\mathrm{out}})$. Also, thanks to a property usually referred to as "local tomography" [86], if the set of operators $\{\rho_w^T\}_w$ spans $\mathcal{L}(\mathcal{H}_{\mathrm{in}})$ and the set $\{M_{a|x}\}_{a,x}$ spans $\mathcal{L}(\mathcal{H}_{\mathrm{out}})$, then the set of setting operators $\left\{\rho_w^T \otimes M_{a|x}\right\}_{w,a,x}$ spans $\mathcal{L}(\mathcal{H}_{\mathrm{in}} \otimes\ \mathcal{H}_{\mathrm{out}})$. In other words, full quantum channel tomography is always possible if one measures a set of characterized setting operators $\{S_{a|xw}\}_{a,x,w}$ that span the space $\mathcal{L}(\mathcal{H}_{\mathrm{in}} \otimes\ \mathcal{H}_{\mathrm{out}})$.

In principle, measuring a set of setting operators $\{S_{a|xw}\}_{a,x,w}$ that span $\mathcal{L}(\mathcal{H}_{\mathrm{in}} \otimes \mathcal{H}_{\mathrm{out}})$ is actually "overkill." More specifically, because of the normalization condition $\mathrm{Tr}_{\mathrm{out}}(C) = \mathbb{1}_{\mathrm{in}}$, respected by quantum channels, there are linear operators in $\mathcal{L}(\mathcal{H}_{\mathrm{in}} \otimes \mathcal{H}_{\mathrm{out}})$ that cannot be written as linear combinations of quantum channels, e.g., $|0\rangle\,\langle 0|_{\mathrm{in}} \otimes \mathbb{1}_{\mathrm{out}}$. One can then consider a set of operators $\{S_{a|xw}\}_{a,x,w}$ that span the set of quantum channels, a subspace with dimension strictly smaller than the dimension of $\mathcal{L}(\mathcal{H}_{\mathrm{in}} \otimes \mathcal{H}_{\mathrm{out}})$. In particular, the linear space $\mathcal{L}(\mathcal{H}_{\mathrm{in}} \otimes \mathcal{H}_{\mathrm{out}})$ has dimension of $d_{\mathrm{in}}^2\, d_{\mathrm{out}}^2$, and the linear span of quantum channels in $\mathcal{L}(\mathcal{H}_{\mathrm{in}} \otimes \mathcal{H}_{\mathrm{out}})$ has dimension of $d_{\mathrm{in}}^2(d_{\mathrm{out}}^2 - 1)$. We emphasize, however, that this does not represent a problem; in practice, use of an overcomplete measurement set is known to minimize the experimental errors in standard quantum tomography [87].

Finally, we now consider tomography of process matrices $W \in\ \mathcal{L}\left(\mathcal{H}_P \otimes \mathcal{H}_{A_{\mathrm{in}}} \otimes \mathcal{H}_{A_{\mathrm{out}}} \otimes \mathcal{H}_{B_{\mathrm{in}}} \otimes \mathcal{H}_{B_{\mathrm{out}}} \otimes \mathcal{H}_F\right)$, such as the quantum switch illustrated in Fig. 1(d). As discussed before, one way to perform tomography is to measure setting operators $S_{abc|xyzw}$ that span the linear space $\mathcal{L}\left(\mathcal{H}_P \otimes \mathcal{H}_{A_{\mathrm{in}}} \otimes \mathcal{H}_{A_{\mathrm{out}}} \otimes \mathcal{H}_{B_{\mathrm{in}}} \otimes \mathcal{H}_{B_{\mathrm{out}}} \otimes \mathcal{H}_F\right)$. Also, thanks to local tomography, we may consider sets of states and measurements that span the local space individually. We then consider the set of states given by Eq. (18), and the set of measurements is given by Eq. (21). For tomography of a higher-order process matrix, we then consider the set of measure-and-reprepare instruments (to be used as inputs for Alice's and Bob's channels) given by all combinations of the two sets above; that is,

$$\mathcal{R} := \{R_{i|(j,k)}\}_{i=1,j=1,k=1}^{i=2,j=3,k=4}, \tag{27}$$

where

$$R_{i|(j,k)} := M_{i|j} \otimes |\psi_k\rangle\,\langle\psi_k|^T. \tag{28}$$

The interpretation of Eq. (28) is as follows: First, the measurement j with POVM elements $\{M_{i|j}\}_i$ is performed. Then, the state $|\psi_k\rangle$ is prepared. Notice that in our measure-and-reprepare instruments, the prepared state $|\psi_k\rangle$ is independent of the measurement choice j and the obtained outcome i. One could also use measurement-dependent states, but even in this case we would need to estimate the probabilities of subsequent measurements for each reprepared state. Thus, the total number of estimated probabilities would not change.

It is worth mentioning explicitly that the linear space spanned by unitary channels is strictly contained in the linear space spanned by quantum channels (for an explicit characterization of such linear spaces, see Appendix A in Ref. [76]). Additionally, the linear space spanned by quantum channels is strictly contained in the linear space spanned by general instrument elements (which are arbitrary positive semidefinite operators that are not greater than or equal to general quantum channels). Hence, to perform full process tomography, one cannot simply use quantum channels (even if nonunitary channels are considered). Rather, full quantum instruments are required.

One can then perform full tomography with the setting operators

$$S_{abc|xyzw} := |\psi_w\rangle\,\langle\psi_w|^{T^{P_{\mathrm{T}}}} \otimes A_{a|x}^{A_{\mathrm{in}}A_{\mathrm{out}}} \otimes B_{b|y}^{B_{\mathrm{in}}B_{\mathrm{out}}} \otimes M_{c|z}^{F_{\mathrm{C}}}, \tag{29}$$

where $|\psi_w\rangle\,\langle\psi_w|$ are the four different quantum states in $\mathcal{S}$, $A_{a|x}$ and $B_{b|y}$ are each the $2 \times 3 \times 4 = 24$ instrument elements [88] of set $\mathcal{R}$ defined in Eq. (27), and $2 \times 3 = 6$ measurement elements for the future control space are the measurements $M_{c|z}$ of $\mathcal{M}$ [Eq. (21)]. In total, we then measure $4 \times 24 \times 24 \times 6 = 13\,824$ different settings. In this tomography approach, we do not apply any constraints on the process matrices. One could also reduce the number of required settings by imposing the assumption that valid process matrices necessarily belong to a particular linear subspace, as discussed in Sec. II D. Notice that, since the linear space spanned by general process matrices strictly contains the linear space spanned by causally ordered process matrices [13,89], the minimal number of required setting operators to perform full tomography for general process matrices is strictly larger than the required number for causally ordered processes.

In an ideal theoretical scenario, if we obtain the probabilities $p\,(abc|xyzw) = \mathrm{Tr}\left(W S_{abc|xyzw}\right)$ for a tomographically complete set of setting operators, standard linear inversion will uniquely identify the process W. However, because of finite statistics, we never obtain the exact probability $p\,(abc|xyzw)$, but obtain an approximation from measured frequencies. Also, because of measurement precision and other possible sources of errors, we cannot expect to obtain an exact reconstruction of the process matrix. Indeed, performing direct linear inversion often results in unphysical quantum states or processes. Instead, we aim to estimate a physical process matrix that agrees best with the experimental data.

To estimate our experimental process matrix W_{expt}, we perform a fitting routine to find the process matrix that best describes our measured data. We find that minimizing the least absolute residuals works quite well. To do this, we numerically search for a process matrix W_{expt} that minimizes the following expression:

$$r = \frac{1}{N_{\text{settings}}} \sum_{abcxyzw} \left| p_{\text{expt}}(abc|xyzw) - \text{Tr}(W_{\text{expt}}\, S_{abc|xyzw}) \right|, \tag{30}$$

where N_{settings} is the number of setting operators, and the minimization is further subject to the constraint that W_{expt} is a valid process matrix; that is, it is positive semidefinite, it has trace d^2, and it is constrained to the linear subspace described in Refs. [13,89]. This minimization can be performed by means of semidefinite programming and may be implemented with the help of numerical libraries such as YALMIP and MOSEK. Our MATLAB code implementing this is available from Ref. [90]. The first term in the absolute value is the experimentally measured probability p_{expt}, while the second term corresponds to what is predicted by quantum theory for the characterized settings $S_{abc|xyzw}$. Since W_{expt} is the only unknown quantity, the minimization of Eq. (30) delivers a process matrix that fits best to our experimental data, with our making no assumptions about the specific form of the process matrix.

It is worth mentioning that we also attempted to use the more standard maximum likelihood and least-squares fits. However, when attempting to solve these problems with YALMIP, our code was unable to define these objective functions after running for several days. On the other hand, our code [90] can solve the entire least absolute residuals problem in approximately 5 min. In numerical experiments on lower-dimensional systems, we found the fidelities between process matrices reconstructed with these different objective functions were slightly lower than our experimental error.

III. EXPERIMENT

A. Time-bin quantum switch

To date, most implementations of the quantum switch have been based on bulk optics. Since photonic quantum switches are essentially interferometers, inevitable phase drifts limit the measurement time or require active stabilization [35,45]. Furthermore, since adding more parties means that the dimension of the control system must be increased, scaling up previous architectures requires more and more spatial modes to be transmitted through the same optic, making it difficult to create a switch with more than two parties. Here we present a passively stable, fiber-based architecture for the quantum switch where the control system is encoded in a time degree of freedom of the photon. Thus, in our architecture, although the dimension of the control system must still be increased at the same rate, this can be done with additional time bins, but only one spatial mode must traverse each optical element. Furthermore, by our using the same interferometer to prepare and measure the control system, all phase fluctuations cancel out, making our setup passively stable. This is important for process tomography, as we must perform many measurements, and the experiment must remain stable during this time.

To create the time-bin qubit that we will use to control the order, we start by generating a photon pair, $\lambda = 1550$ nm, using spontaneous parametric down-conversion (SPDC). One photon of the pair is directly detected to herald the other photon, setting a time reference for the experiment. The second photon is sent to a 50:50 beam splitter—a fiber directional coupler (FDC)—that splits the incoming mode into two fibers of different lengths. We then deterministically recombine these two fiber paths using an ultrafast fiber optical switch (UFOS); see Fig. 3(a) and also the yellow section in Fig. 4(a) [91]. To do so, we generate an electronic pulse, triggered off the timing reference generated by detection of the first photon. This pulse is sent to the UFOS, which changes its state to first route the “photon component” from the short path, followed by the “photon component” in the long path, into the upper output mode of the UFOS-MZI [Figs. 3(a)(ii) and 3(a)(iii)]. The result is that the second photon is left in an equal superposition of two time bins in a single fiber [Fig. 3(a)(iv)]. Because the short time bin is transmitted through the FDC, while the long time bin is reflected, one mode picks up a reflection phase, while the other does not. Hence, in our experiment we prepare the control qubit in the state $|y-\rangle_{\text{C}} = (|S\rangle_{\text{C}} - i\,|L\rangle_{\text{C}})/\sqrt{2}$, where we have labeled the mode as the “short” (“long”) state $|S\rangle_{\text{C}}$ ($|L\rangle_{\text{C}}$) when it has taken the short (long) fiber path of the interferometer. The spacing between these two time bins is 150 ns, which is set by the response time of our UFOS.

The UFOSs we use to route the photon are BATi 2×2 Nanona fiber switches. In addition to creating the time-bin qubit, they allow us to route the photon in a controlled way through the quantum switch. Our UFOSs have a response time of 60 ns, with a maximal duty cycle of 1 MHz, and a cross-channel isolation greater than 20 dB for any polarization (see Refs. [92,93] for more details).

Having created the time-bin control qubit, we need to apply the quantum switch operation to the target system, which we encode in the polarization degree of freedom of the same photon. To route the photon, we use two additional UFOSs and follow the protocol illustrated in Figs. 4(b)(i)– 4(b)(vi). In particular, we send a voltage pulse train consisting of three low levels and two high

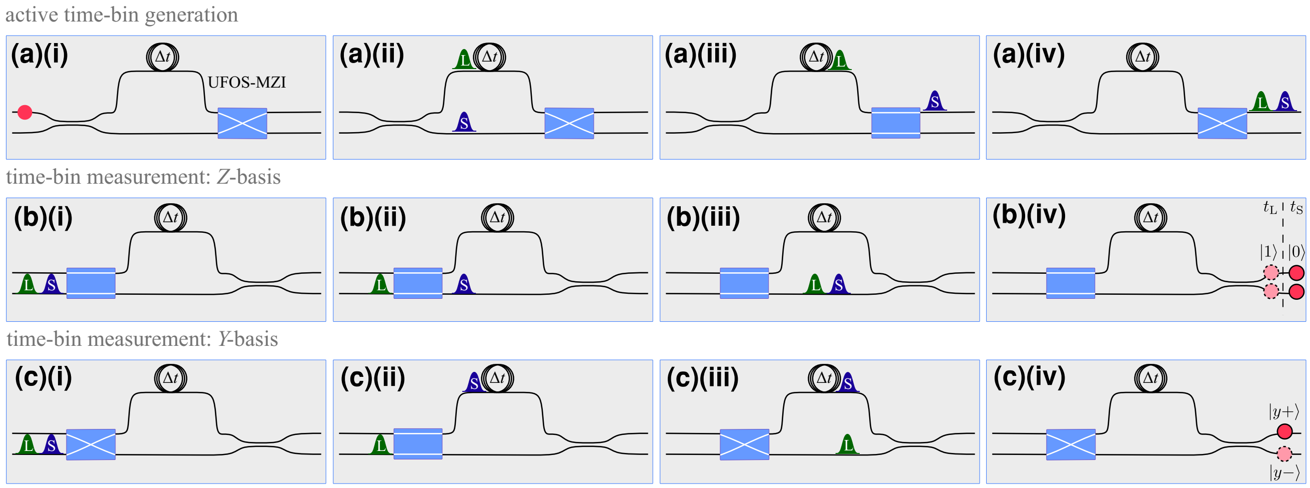

FIG. 3. Active generation and measurement of time-bin qubits. (a) An incident single photon is deterministically placed in a superposition of the “short” (S) and “long” (L) time bins, with use of an UFOS. This optical switch is UFOS-MZI in Fig. 4(a). However, for clarity, we have mirrored the device horizontally. (b) Our measurements in the Z basis. Here the UFOS remains in the “bar state.” After traversing the device, the photon remains in a superposition of the two incident time bins (but spread over two paths). In this situation, simply resolving the arrival time of the photon projects the control qubit into the Z basis. (c) A schematic of our deterministic measurement in the Y basis. Here the UFOSs alternate between the “cross” and bar states so that the short (long) time bin now takes the long (short) path. In this manner, the two time bins interfere on the beam splitter, so finding the photon in the upper (lower) path corresponds to projecting the time bin onto $|y+\rangle$ ($|y-\rangle$).

levels to the UFOSs. During each low level, the fiber switches are in a “cross state” (output modes are swapped with respect to the input), while during a high level the switch state is set to the “bar state” (input modes transmitted to output modes). As the time bins approach the quantum switch [Fig. 4(b)(i)] the UFOSs are initially in the cross state, which routes the short time bin $|S\rangle_C$ through Bob’s quantum channel [Fig. 4(b)(ii)]. Then the UFOSs change to the bar state [Fig. 4(b)(iii)], which sends $|L\rangle_C$ through Alice’s channel, while $|S\rangle_C$ travels over a fiber from the rhs UFOS to the lhs UFOS. Then the UFOSs see a low voltage level, and they are set to the cross state [Fig. 4(b)(iv)]. This sends $|S\rangle_C$ through Alice’s local laboratory, while $|L\rangle_C$ loops back to the lhs UFOS. In Fig. 4(b)(v) the UFOSs are in the bar state and hence $|L\rangle_C$ passes through Bob’s channel. At this point $|S\rangle_C$ exits the quantum switch. Finally, the fiber switches are set to the cross state [Fig. 4(b)(vi)] so that $|L\rangle_C$ leaves the quantum switch. At this point, depending on the control state, the target system has experienced a different order of Alice’s and Bob’s actions, which, as we describe shortly, act on the polarization state of the photon. Note that all the lengths of the fibers in the quantum switch are set to ensure the correct routing of time bins spaced by 150 ns.

The time-bin quantum switch from Fig. 4(b) is placed in the full fiber-based setup [Fig. 4(a)], in which the time bins are prepared and measured. The quantum switch itself is shown in the green section in Fig. 4(a). The type-II SPDC photon source [94] is shown in the red section [95]. Here, a polarizing beam splitter (PBS) reflects the heralding photon to a single-photon detector (blue area), while the other photon is transmitted to the Mach-Zehnder-like time-bin-generation interferometer described above (red section). Following this, we have a photon encoding a time-bin qubit in the state $(|S\rangle_C - i\,|L\rangle_C)/\sqrt{2}$ in the “upper” output of the interferometer (clockwise direction). The counterclockwise path of the loop (lower output mode of the UFOS-MZI) hosts an optical circulator with an empty port to filter out misguided photons, which can arise from the imperfect extinction ratio of our UFOSs. Next, the target system is encoded in the photon’s polarization. For this we use a PBS and a set of quarter-wave and a half-wave plates, as shown in the orange section in Fig. 4(a). Then we apply the two-switch operation to the target system described above [green sections in Figs. 4(a) and 4(b)], where we implement Alice’s and Bob’s instruments using short free-space sections containing wave plates and polarizers.

After exiting the switch, the photon follows the fiber loop in a clockwise direction and approaches the MZI used for time-bin generation (red section), but now from the opposite direction in the lower path. At this point, we can decide to measure the control qubit in the computational (Z) basis or a superposition (Y) basis. These measurements are illustrated in Figs. 3(b) and 3(c), respectively. For measurements in the computational basis, both time bins are routed by the UFOS along the lower path of the MZI, after which the two time bins split up at the FDC, and are then sent to detectors in the blue region. By measuring the arrival time, with respect to the herald detection, we can

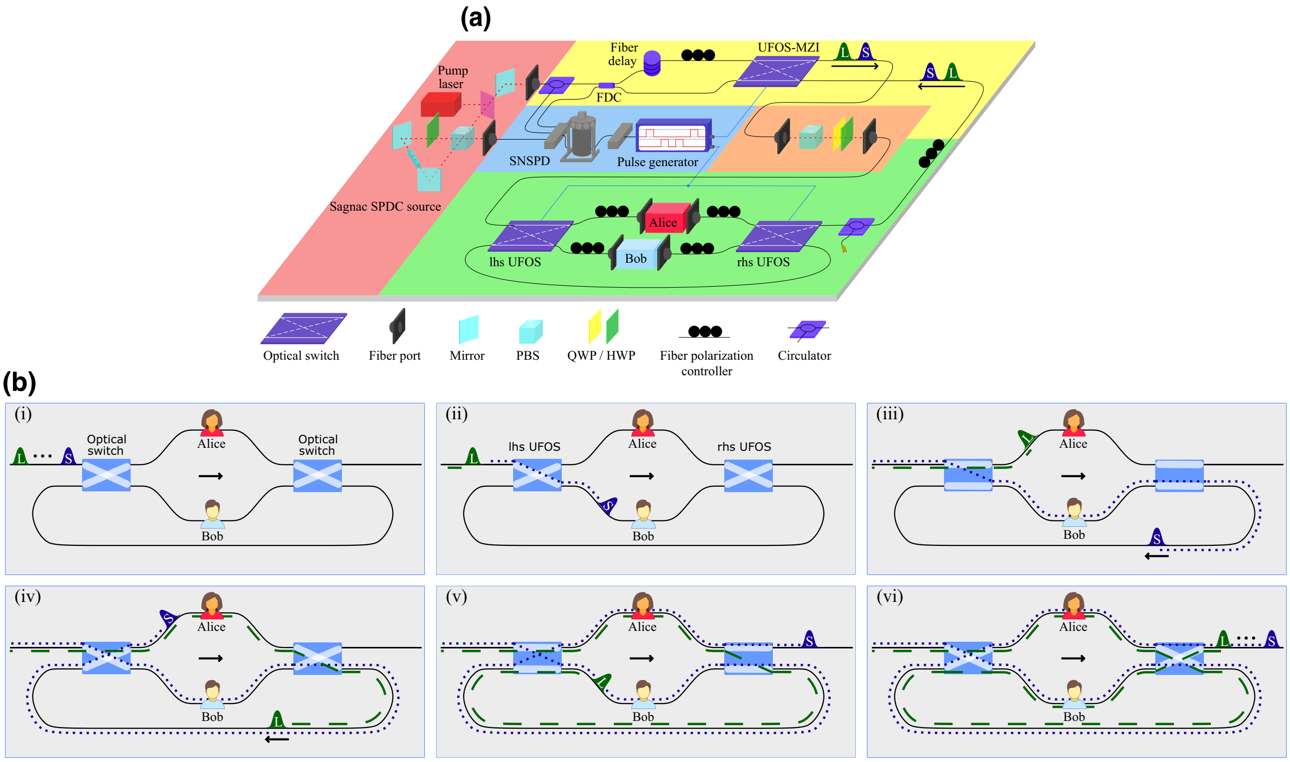

FIG. 4. Experimental setup. (a) The complete experiment. The individual sections are indicated with colors. The red section shows the Sagnac SPDC source, whuch generates heralded single photons. In the yellow section, we illustrate the asymmetric MZI used to generate and measure the time-bin control qubit. The orange section shows the target qubit preparation stage, which consists of a PBS and two wave plates. The green area hosts the fiber-based quantum switch. The heralded and heralding photons are detected wit use of superconducting-nanowire single-photon detectors (SNSPDs), shown in the blue area. By triggering off the detection event for a heralding photon, we use a pulse generator to control the optical switches in the setup. (b) The functionality of the quantum switch. By controlling the state of the optical switches, we route the two time bins in different orders through Alice's and Bob's quantum channels. After the switch operation, the target qubit has experienced the action of the quantum channels in a different order depending on the state of the time-bin qubit. Note that panels (i)–(vi) serve merely to illustrate the time-bin routing using the optical switches. The positions of the time bins in the diagram do not accurately represent their exact locations during the experiment. Experimentally, the time-bin spacing Δt between $|S\rangle_C$ and $|L\rangle_C$ stays unchanged. HWP, half-wave plate; L, long time bin; QWP, quarter-wave plate; S, short time bin.

distinguish between the short and long time bins. To measure in a superposition basis, we use the UFOS-MZI to send the time bins through the opposite paths of the interferometer ($|S\rangle_C$ takes the long path and $|L\rangle_C$ takes the short path) so that they arrive at the FDC at the same time. In this case, interference occurs at the FDC, and detecting a photon exiting the upper (lower) port corresponds to projecting the control qubit in $|y+\rangle_C$ ($|y-\rangle_C$).

With this in place, we collect the measurement statistics from different measurement settings by detecting coincidence events between the heralding photon and the FDC output or the circulator output. For each experimental configuration, we record approximately 1600 coincidence counts (approximately 21 000 total single-photon counts) over 10 s at the FDC output and the circulator output. The photon source generates approximately 1 480 000 single photons (approximately 116 000 coincidence events) in 10 s before the experiment. Thus, our entire quantum switch experiment has an overall insertion loss of approximately 18 dB. All our measurements are performed with superconducting-nanowire single-photon detectors from PhotonSpot Inc. The result, for a representative set of measurements, is shown in Fig. 5. Therein, the bars represent the theory for an ideal quantum switch with the control qubit in state $|y-\rangle_C$, described by the process matrix W_s^{y-}, while the points represent our experimentally measured data. One can observe good agreement between theory and experiment. Using unitary operations (rather than measure-and-reprepare instruments), we can also play the anticommuting-commuting gate discrimination game, as in Ref. [35]. We find a success probability of 0.974 ± 0.018, indicating high fidelity of our implementation (full details of this measurement are presented in Appendix B 1).

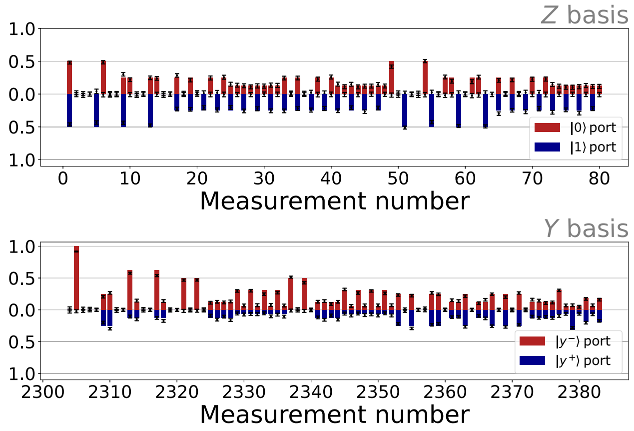

FIG. 5. Experimentally estimated probabilities. A small subset of the experimentally estimated probabilities. The bars represent the theory for the ideal process matrix W_s^{y-} and the points represent the experimental estimates. The upper (lower) panel shows measurements of the control qubit in the Z (Y) basis. The red bars are for outcomes $|0\rangle$ and $|y-\rangle$ and the blue bars are for outcomes $|1\rangle$ and $|y+\rangle$.

B. Experimental process matrix tomography

We now present our experimental reconstruction of the process matrix of our time-bin quantum switch. As discussed in Sec. II, to probe the underlying process, Alice and Bob must each implement a complete set of instruments. In our experiment, the target system is encoded in the polarization state of the photon, so Alice and Bob must act on this degree of freedom. Rather than the measurement-repreparation instruments defined in Eqs. (27) and (28), we use a slightly modified form $\tilde{R} := \tilde{R}_{i|(j,k)}$ presented in Eqs. (A9) and (A10) in Appendix A. In particular, Alice and Bob each have access to three different measurement bases $\tilde{M}_{i|j}$, where $j \in 1,2,3$ defines the measurement and i defines the outcome. For each j, $\tilde{M}_j := \tilde{M}_{1|j} - \tilde{M}_{2|j}$ is the observable associated with the POVM $\{\tilde{M}_{1|j}, \tilde{M}_{2|j}\}$. The specific operators we implement are defined in Eqs. (A5) and (A6). For example, $j = 1$ corresponds to the Z basis: $\tilde{M}_1 := \tilde{M}_{1|1} - \tilde{M}_{2|1} = Z$.

Experimentally, we implement these measurements using a polarizer fixed to transmit horizontally polarized light $|H\rangle$. We set the measurement basis using a quarter-wave plate and a half-wave plate before the polarizer set to the angles given in Eq. (A6). To implement the second part of the instrument—the repreparations—we must prepare one of four different states. We experimentally accomplish this using another quarter-wave plate and another half-wave plate to rotate the photon's polarization if it is transmitted through the polarizer. This allows us to prepare one of the four $\left|\tilde{\phi}_k\right\rangle$ states listed in Eq. (A8). Thus, overall, both parties can implement the 24 different measurement-repreparation operators defined in Eqs. (A9) and (A10) (six different measurement operators times four different repreparations).

In addition to Alice's and Bob's channels, we must send in a complete set of target states and perform measurements on the control qubit after the switch. To this end, we first prepare the target qubit in the four different input states $\left|\tilde{\psi}_w\right\rangle$ given in Eq. (A4). We set these states using the quarter-wave plate and half-wave plate pair mounted in the target preparation stage, shown in the orange area in Fig. 4(a) [the exact wave plate angles that we use are listed in Eq. (A4)].

Finally, at the output of the switch we must measure the state of the control qubit. This procedure is illustrated in Figs. 3(b) and 3(c). As discussed in Sec. III A, we use the same beam splitter to measure and prepare the time-bin qubit, but from opposite directions. As a result, the phase of this measurement basis is fixed to the Y basis. In our notation in Eqs. (A1) and (A2) in Appendix A, this corresponds to measurements $\tilde{C}_{1|2}$ and $\tilde{C}_{2|2}$. Experimentally, a $\tilde{C}_{1|2}$ versus $\tilde{C}_{2|2}$ result depends on from which port of the FDC the photon exits. As described above, we can additionally measure in the Z basis by fixing the UFOS-MZI to the bar state on the return trip such that the short and long time bins do not interfere at the FDC and observing the arrival time of the time bins. If we find the photon arrives earlier, this is associated with a $\tilde{C}_{1|1}$ detection event, while if it arrives later, it corresponds to $\tilde{C}_{2|1}$.

To be complete with regard to the future control space, we would require an additional measurement of the control qubit in the X basis, i.e., we need the measurements $M_{1|2}$ and $M_{2|2}$ from Eq. (23). In our experiment, this could be achieved with use of a fast phase modulator to apply the appropriate phase between the short path and the long path only in the reverse direction. However, we do not implement this here. Instead, we impose an additional constraint on our tomographic reconstruction. We require that $\mathrm{Tr}\left(W_{\mathrm{expt}} X^F\right) = 0$, where $X = |+\rangle\langle+| - |-\rangle\langle-|$. Given the passive phase stability of our experiment, this is a very good assumption. We verify this assumption by comparing reconstructions with and without this constraint. In particular, we find that the fidelity between the process matrices reconstructed with and without this constrain is 0.999982, well below our experimental error.

Overall, this results in $24 \times 24 \times 4 \times 4 = 9216$ setting operators of the form given in Eq. (A11) (number of Alice's settings times number of Bob's settings times number of target states times number of control measurements). However, for the control measurements, we have access to both ports of the FDC beam splitter simultaneously (i.e., there is a detector in each output port of the beam splitter), giving rise to 4608 different experimental configurations. From these data we can calculate the probabilities for each given setting operator. Experimentally, we measure count rates associated with each setting operator, which we must then normalize for conversion into the required probabilities. To do so, we use the normalization condition over the outcomes of all three measurements

$$\sum_{abc} p(abc|xyzw) = 1. \quad (31)$$

Thus, we define a normalization constant for every value of x, y, z, and w:

$$N_{xyzw} = \sum_{abc} C(abc|xyzw), \quad (32)$$

where $C(abc|xyzw)$ are the number of coincidences measured between the heralding detector and the detectors after the FDC, corresponding to the setting operator defined by a, b, c, x, y, z, and w. Then our experimentally estimated probabilities are defined as

$$p_{\text{expt}}(abc|xyzw) = \frac{C(abc|xyzw)}{N_{xyzw}}. \quad (33)$$

A small subset of the resulting probabilities is plotted in Fig. 5. By minimizing Eq. (30) [with the setting operators $S_{abc|xyzw}$ replaced by the experimental setting operators $\tilde{S}_{abc|xyzw}$ from Eq. (A11)], we can reconstruct the process matrix W_{expt}. Our MATLAB code implementing this minimization is available from Ref. [90].

IV. RESULTS

A. Fidelity

The experimentally obtained 64 × 64 process matrix and the ideal matrix are plotted in Fig. 6 as a 3D bar chart, where Fig. 6(a) shows the real part and Fig. 6(b) shows the imaginary part. The solid bars represent the experimentally reconstructed process matrix, while the transparent bars represent the theoretical process matrix W_s^{y-}. The x and y axes numerically label the basis elements. The relatively close agreement between the target process matrix and our experimental process matrix is already evident in Fig. 6.

To further assess the agreement between our experiment and theory, we estimate the fidelity of the measured process matrix W_{expt} to the ideal matrix W_s^{y-}. Since every valid process matrix normalized by its trace is a valid quantum state, we use the conventional expression for calculating the fidelity $F(\sigma, \rho) = \text{Tr}\left(\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right)$, with σ and ρ being different density matrices [71]. This results in fidelity

$$F(W_{\text{expt}}, W_s^{y-}) = 0.920 \pm 0.001, \quad (34)$$

where the error arising is estimated by a Monte Carlo simulation of the entire reconstruction procedure accounting for finite measurement statistics and small wave plate errors of 1°. Especially given the high dimension of our process matrix, this fidelity indicates that our experiment is quite close to theory.

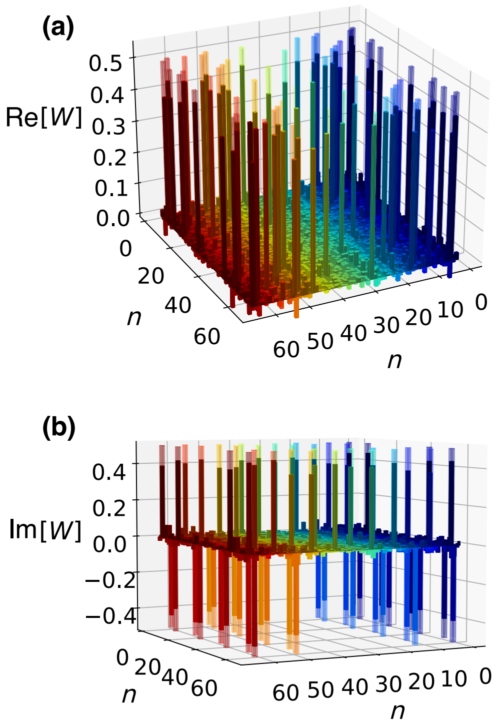

FIG. 6. Process tomography data. The experimentally recreated process matrix of the quantum switch. (a) The real part of W_s^{y-} and (b) the imaginary part. The color gradient along the x axis does not have a physical meaning; rather, it is color coded so as to identify the individual elements of this 64 × 64 matrix better. Additionally, the ideal process matrix is represented via the layer of semitransparent bars. Calculation of the fidelity between these two process matrices results in $F(W_{\text{expt}}, W_s^{y-}) = 0.920 \pm 0.001$.

To quantify the agreement between our experimental data and W_{expt}, we compare the residuals of our fit r [defined in Eq. (30)] with the average statistical error of our data η_{stat}. The residuals r can be interpreted as the disagreement between the outcome predicted by W_{expt} and the measured experimental outcome, averaged over all measurement settings. For our fit $r = 0.0089$, indicating an excellent match to our experimental data. We estimate our statistical errors as follows. First, we treat the probability to obtain an outcome abc as a binomial variable: either we detect a photon or we do not detect a photon. Then we estimate the variance of that setting as $N_{xyzw}\, p(abc|xyzw) \times (1 - p(abc|xyzw))$, where N_{xyzw} is the number of photons detected in all outcomes associated with $xyzw$ [defined in Eq. (32)]. Finally, we compute the average error per setting as

$$\eta_{\text{stat}} = \frac{1}{N_{\text{settings}}} \sum_{abc|xyzw} \frac{p(abc|xyzw)(1 - p(abc|xyzw))}{\sqrt{C}}. \quad (35)$$

This is simply the standard error of each setting operator averaged over all settings. Evaluating this for our data, we find $\eta_{\text{stat}} = 0.0056$. Given that $\eta_{\text{stat}} \approx r$, we conclude that our process matrix fits our data well.

B. Causal nonseparability

A bipartite process matrix without a common past is causally nonseparable when it cannot be written as a classical mixture of causally ordered processes [6,13]. When one is considering bipartite processes with a common past, such as the quantum switch considered in this work, there are different nonequivalent definitions of indefinite causality. In Refs. [13,43], a bipartite process matrix $W \in \mathcal{L}\left(\mathcal{H}_P \otimes \mathcal{H}_{A_{\text{in}}} \otimes \mathcal{H}_{A_{\text{out}}} \otimes \mathcal{H}_{B_{\text{in}}} \otimes \mathcal{H}_{B_{\text{out}}} \otimes \mathcal{H}_F\right)$ with common past and common future is said to be causally separable if it can be written as a convex sum of causally ordered process matrices. That is, W is causally ordered if we can write it as

$$W = pW^{A>B} + (1-p)W^{B>A}, \tag{36}$$

where $p \in [0, 1]$ and where $W^{A>B}$ and $W^{B>A}$ are causally ordered processes (objects also referred to as quantum combs [4,8], see Appendix C). Alternatively, in Ref. [11] the concept of extensible causally separable processes is proposed. This leads to a definition that differs from the one in Eq. (36), but is equivalent to the definition of causal separability presented in Ref. [96], which considers not only convex mixtures of causally ordered processes but also incoherent (hence, classical) control of causal order. The analysis and the numbers presented in this section and in the main part of this paper were obtained via the definition in Ref. [43], which is the one presented in Eq. (36). However, we stress that the results of our work are not qualitatively affected by the different definitions mentioned, in the sense that, in all cases, the process we obtain after tomography is not causally separable, and it is robust with regard to different kinds of noise. In Appendix C, we present a more detailed discussion of such definitions and how they result in small quantitative changes in the numbers presented here.

One method to quantify the degree to which our quantum process is causally nonseparable is by use of a causal witness. A causal witness is a Hermitian non-negative operator G such that $\text{Tr}\left(GW_{\text{sep}}\right) \geq 0$ for all causally separable processes. However, for all causally nonseparable processes (such as the quantum switch), one can always find a witness G such that $\text{Tr}\left(GW_{\text{s}}^{y-}\right) < 0$. Without additional constraints, the quantity $\text{Tr}\,(GW)$ does not have a physical meaning, and may be artificially inflated by multiplying the witness G by some constant. However, by setting additional normalization constraints on the witness G, one may identify the quantity $\text{Tr}\,(GW)$ with how much noise the process W can tolerate until it becomes causally separable. More concretely, let $\mathbb{1}_W := \mathbb{1}/d_{P_{\text{T}}} d_{A_{\text{out}}} d_{B_{\text{out}}}$ be the "white noise process," which simply consists of discarding everything and outputting white noise. The quantities in the denominator of $\mathbb{1}_W$ correspond to the dimensions of the Hilbert spaces. Since Alice and Bob implement qubit channels, their output space dimensions are both $d_{A_{\text{out}}} = d_{B_{\text{out}}} = 2$. Additionally, they perform operations on the target qubit system, making $d_{P_{\text{T}}} = 2$. Now, let W_{sep} be an arbitrary causally separable process matrix. In Ref. [13] it is shown that for any given process matrix W, the problem of minimizing the quantity $\text{tr}(GW)$ over all possible causal witnesses G given by

$$\min_{G} \text{Tr}(GW) \tag{37}$$

$$\text{subject to } \text{Tr}\left(GW_{\text{sep}}\right) \geq 0 \text{ for all } W_{\text{sep}}, \tag{38}$$

$$\text{Tr}(G) \leq \text{Tr}(\mathbb{1}_W) \tag{39}$$

is equivalent to its dual formulation, a problem that consists in finding how robust the given process W is with regard to white noise:

$$\min_{r} r \tag{40}$$

$$\text{subject to } \frac{W + r\mathbb{1}_W}{1+r} \text{ being causally separable.} \tag{41}$$

More specifically, in Ref. [13] it is shown that if G^* is the witness that solves the optimization problem presented in Eq. (37) and $r*$ is the solution of the optimization problem in Eq. (40), it holds that $\text{Tr}(G^*W) = -r^*$, which allows us to interpret $\text{Tr}(G^*W)$ as how robust W is with regard to white noise.

Alternatively, one may also consider the problem

$$\min_{G,\Omega} \text{Tr}(GW) \tag{42}$$

$$\text{subject to } \text{Tr}\left(GW_{\text{sep}}\right) \geq 0 \text{ for all } W_{\text{sep}}, \tag{43}$$

$$\text{Tr}(G\Omega) \leq 1, \tag{44}$$

where Ω is an arbitrary process matrix. In this case, the equivalent dual problem reads

$$\min_{r,\Omega} r \tag{45}$$

$$\text{subject to } \frac{W + r\Omega}{1+r} \text{ being causally separable,} \tag{46}$$

where Ω is an arbitrary process matrix. If G^* and r^* are the solutions of the optimization problems in Eqs. (42) and (45), respectively, we have $\text{Tr}(G^*W) = -r^*$, a quantity typically referred to as the "generalized robustness"; it may be viewed as the amount of noise one needs to add to W to make it causally separable in the worst-case scenario.

For the witness G to fit the setting operators implemented in our experiment, we impose an additional structure on the witness G that is given by

$$G = \sum_{abcxwyz} \alpha_{a,b,c,x,y,z,w} S_{abc|xyzw}, \tag{47}$$

where $\alpha_{a,b,c,x,y,z,w}$ are arbitrary real numbers and $S_{abc|xyzw}$ are the setting operators of our experiment (see Sec. II E). Additionally, for fixed setting operators, finding the maximal violation of a witness G with the normalization constraints related to white noise and generalized noise is a semidefinite program [13], and can be efficiently solved numerically [97]. Our code for doing so is available from Ref. [90].

With these tools, we can construct a variety of witnesses. First, we can construct witnesses using the complete measurement set [Eq. (29)] or our restricted measurement set [Eq. (A11)]. We can further design witnesses for two different process matrices W_{s}^{y-} or W_{expt}. This results in four witnesses: $G_{y-,\mathrm{all}}$, $G_{y-,\mathrm{res}}$, $G_{\mathrm{expt,all}}$, and $G_{\mathrm{expt,res}}$. Where the subscript “$y-$‘” (“expt”) indicates that the witness was optimized for the ideal (experimental) process matrix, and the subscript “all” (“res”) indicates that the witness was computed with use of the complete (restricted) measurement set. We can then further construct witnesses for either the generalized robustness or the white noise robustness.

The results for the generalized robustness witnesses are summarized in Table I. The first row in Table I shows the value of the four witnesses evaluated with W_{s}^{y-}, and the second row shows the experimental values, estimated with W_{expt}. In this case, we see that $\mathrm{Tr}(W_{\mathrm{s}}^{y-} G^{\mathrm{GR}}_{y-,\mathrm{res}}) \approx \mathrm{Tr}(W_{\mathrm{s}}^{y-} G^{\mathrm{GR}}_{y-,\mathrm{all}})$ and $\mathrm{Tr}(W^{\mathrm{expt}} G^{\mathrm{GR}}_{\mathrm{expt,res}}) \approx \mathrm{Tr}(W^{\mathrm{expt}} G^{\mathrm{GR}}_{\mathrm{expt,all}})$. In other words, the generalized robustness evaluated either with the complete setting operators or with the restricted setting operators is equal within experimental error. Evidently, the additional measurement on the future control system does not affect the generalized robustness. More interesting for the generalized robustness witnesses is the performance of the witnesses optimized for our experimental process matrix W_{expt}. Examining the performance of our experimentally estimated witnesses (the bottom row of Table I), we see that the witnesses designed specifically for our experimental process matrix increase the generalized robustness. In particular, $\mathrm{Tr}(W_{\mathrm{expt}} G^{\mathrm{GR}}_{\mathrm{expt,all}}) > \mathrm{Tr}(W_{\mathrm{expt}} G^{\mathrm{GR}}_{y-,\mathrm{all}})$ and $\mathrm{Tr}(W_{\mathrm{expt}} G^{\mathrm{GR}}_{\mathrm{expt,res}}) > \mathrm{Tr}(W_{\mathrm{expt}} G^{\mathrm{GR}}_{y-,\mathrm{res}})$. To quantify this, we note that the net improvement from the first entry in the experimental row to the last is approximately 14 standard deviations, which could prove very relevant to low-fidelity implementations of the quantum switch. This would not be readily possible without process matrix tomography being performed.

In Table II we summarize the results of our white noise witness analysis. In this case, we see a significant difference between the witnesses constructed with the restricted and complete measurement sets, with the complete measurements sets revealing greater white noise robustness in all cases. Furthermore, in the second row, we can see that each step progressively increases the experimental white noise robustness. The first entry $\mathrm{Tr}(G^{\mathrm{WN}}_{y-,\mathrm{res}} W_{\mathrm{expt}}) = -1.65 \pm 0.02$ shows the value that one would obtain for our setup without performing process matrix tomography. i.e., the witness was designed for the ideal process and uses only the experimentally implementable measurement settings. In the next column, $\mathrm{Tr}(G^{\mathrm{WN}}_{\mathrm{expt,res}} W_{\mathrm{expt}}) = -1.76 \pm 0.01$ is an improvement by our tailoring the witness for our experiment, however, still using only experimentally implementable measurements. In the next two columns, we improve both these values further by computing the witness assuming the complete measurement set. The final entry, $\mathrm{Tr}(G^{\mathrm{WN}}_{\mathrm{expt,all}} W_{\mathrm{expt}}) = -2.11 \pm 0.02$ is higher than the first entry by 23 standard deviations, clearly showing the power of full process matrix tomography. Process matrix tomography allows us to compute properties of the experimental process without having direct experimental access to them and we can precisely tailor our analysis to our experimental conditions. In Tables III and IV in Appendix C, we show the same analysis for the alternative definition of causal nonseparability. The trends observed therein are the same, although the absolute values of the robustnesses are smaller.

C. Worst-case process tomography

For the process tomography results presented in Sec. IV A, we found the process matrix that fit to our data best by minimizing Eq. (30). As discussed there, this resulted in a process matrix that describes our data

TABLE I. Generalized robustness witness analysis. A summary of the different generalized robustness witnesses constructed. The witnesses $G_{i,j}$ are labeled by two subscripts. The first indicates if the witness was designed for the ideal process matrix W_{s}^{y-} (subscript “$y-$”) or the experimental process matrix W_{expt} (subscript “expt”). The second subscript indicates if the restricted measurement set (subscript “res”) or the complete measurement set (subscript “all”) was used for the witness. The first row shows the value of the witness for W_{s}^{y-} and the second row shows the experimental values, which were evaluated as $\mathrm{Tr}[G_{i,j} W_{\mathrm{expt}}]$.

	$G^{\mathrm{GR}}_{y-,\mathrm{res}}$	$G^{\mathrm{GR}}_{\mathrm{expt,res}}$	$G^{\mathrm{GR}}_{y-,\mathrm{all}}$	$G^{\mathrm{GR}}_{\mathrm{expt,all}}$
W_{s}^{y-}	-0.5834	-0.5525	-0.5834	-0.5512
W_{expt}	-0.387 ± 0.003	-0.431 ± 0.003	-0.370 ± 0.003	-0.431 ± 0.002

TABLE II. White noise witness analysis. A summary of the different white noise witnesses constructed. The witnesses $G_{i,j}$ and process matrices are labeled by two subscripts described in the caption for Table I.

	$G^{\mathrm{WN}}_{y-,\mathrm{res}}$	$G^{\mathrm{WN}}_{\mathrm{expt,res}}$	$G^{\mathrm{WN}}_{y-,\mathrm{all}}$	$G^{\mathrm{WN}}_{\mathrm{expt,all}}$
W_{s}^{y-}	-2.296	-2.174	-2.767	-2.624
W_{expt}	-1.64 ± 0.02	-1.76 ± 0.01	-1.96 ± 0.02	-2.11 ± 0.02

very well. However, one could ask the following question: "Are there other causally separable process matrices that describe the data almost as well?" To answer this question, we perform a "worst-case" version of our process matrix tomography. To do so, rather than minimizing Eq. (30), we find the process matrix that maximizes the generalized robustness or the white noise robustness, $\mathrm{Tr}(W_{\mathrm{worst}}G^{\mathrm{GR}}_{j,\mathrm{all}})$ or $\mathrm{Tr}(W_{\mathrm{worst}}G^{\mathrm{WN}}_{j,\mathrm{all}})$, respectively. We do this using the witnesses designed for the ideal process matrix and the originally reconstructed experimental process matrix, but always with the complete measurement sets. This maximization is subject to the constraints that W_{worst} is a physical process matrix and that the predictions of W_{worst} match the experiment within some error ϵ:

$$\sum_{abcxyzw} |\mathrm{Tr}(S_{abc|xyzw}W) - p_{\mathrm{expt}}(abc|xyzw)| \leq \epsilon. \tag{48}$$

Here ϵ is closely related to the residuals defined in Eq. (30). In particular, if $\epsilon < r$, the maximization will fail, as there is no physical process matrix compatible with this constraint. Thus, we perform worst-case process tomography for the four witnesses discussed above starting from $\epsilon = r_{\mathrm{expt}} = 0.0089$ and going to $\epsilon = 0.015$. The results of this analysis are plotted in Figs. 7(a) and 7(b). We find that the generalized robustness witnesses are more tolerant ϵ, finding that our data are consistent with causally separable process matrices only for $\epsilon \lesssim 0.012$, while the white noise witnesses require $\epsilon \lesssim 0.0105$. Although this analysis suggests that the causal nonseparability is rather fragile, we stress the worst-case nature of this treatment: if a single causally separable process matrix is compatible with our data within ϵ, it will be returned, even if causally nonseparable data fit our data better. In any case, we see that for a range of experimentally relevant errors our data are compatible only with a causally nonseparable process matrix.

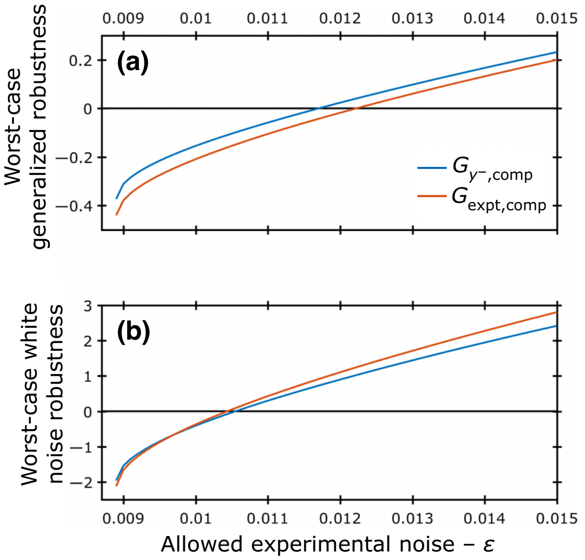

FIG. 7. Worst-case process tomography. (a) Worst-case generalized robustness and (b) worst-case white noise robustness causal witnesses versus the allowed deviation from the experimental data. For these plots, the tomography routine attempted to find a process matrix that maximized the witness (i.e., it searched for the "most causally separable" process matrix), while still agreeing with our experimental data with an average error of ϵ. For all witnesses considered, the process that minimizes ϵ is also the most causally nonseparable. For $\epsilon < 0.089$, no valid process matrix is found.

V. DISCUSSION

In this work, we have presented a protocol to perform process tomography on a higher-order quantum operation, the quantum switch. We discussed how to construct a complete set of measurements. The requirements for this go beyond standard quantum process tomography, wherein one must "only" send a complete set of input states through the process, and perform a complete set of measurements after the process. In particular, because HOQOs take quantum channels as inputs, we must also implement a complete set of quantum channels for each input channel. This can be achieved with measure-and-reprepare instruments. Since this procedure scales even worse than standard process tomography, we implement it using a new phase-stable architecture of the quantum switch, allowing long integration times.

Our photonic quantum switch uses a time-bin qubit as the control system. By recombining time-bin qubits with use of a passively stable interferometer, we were able to keep our experiment stable indefinitely. We believe this technique will be beneficial for various time-bin quantum information experiments and will be scalable to high-dimensional time-bin qudits. Furthermore, with a stable time-bin qudit source, there is straightforward expansion of the two-party quantum switch into a multiparty version [47]. Creating a multiparty quantum switch is an experimentally challenging task, with just one experiment to date superposing four parties [40]. However, this goal is well motivated both for fundamental studies of multiparty causal nonseparability [96,98,99] and for applications that require the ICO advantage to scale with the number of superposed parties [17,19,28,40,47].

The results of performing quantum process matrix tomography on our experiment show that we have indeed implemented a high-fidelity quantum switch, with fidelity $F = 0.920 \pm 0.001$. We then used our results to verify the causal nonseparability of our experiment, designing causal witnesses specifically for our experimental process matrix. Finally, we implemented worst-case process tomography, searching for a causally separable process that could also describe our measurement. To find such a process, we had to allow for an approximately 1.5 times larger disagreement between our measurements and our causally separable model. Our experimentally estimated process matrix yields a precise description of our physical implementation, and could prove an essential tool for experimentalists. Beyond estimating the process fidelity and evaluating the causal nonseparability, it could be used to debug experimental implementations and to simulate other experiments, including scenarios where Alice and Bob implement channels that were not used for process tomography or situations where the quantum switch makes up one part of a larger experiment.

Although our protocol was presented for the quantum switch, it could be adapted to general HOQOs in a straightforward manner. In the present work, we performed an overcomplete set of measurements, but it should be possible to implement a reduced set of measurements by taking into account the constraints on the space of physical process matrices. We also point out that many complexity-reducing techniques from standard state and process tomography, including compressed sensing [100], shadow tomography [101], and adaptive tomography [102], should apply to our protocol equally well. But we leave these as topics for future work.

All the data necessary to replicate, verify, falsify, and/or reuse this research are available online from Ref. [90].

ACKNOWLEDGMENTS

This project was funded in whole or in part by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 820474 (UNIQORN) and Grant Agreement No. 899368 (EPIQUS), Marie Skłodowska-Curie Grant Agreement No. 956071 (AppQInfo), and the QuantERA II Programme under Grant Agreement No. I 6002 - N (PhoMemtor); the Austrian Science Fund (FWF) through [10.55776/F71] (BeyondC), and [10.55776/FG5] (Research Group 5); the AFOSR via FA8655-20-1-7030 (PhoQuGraph) and FA9550-21-1-0355 (QTRUST); the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development, and the Christian Doppler Research Association.

APPENDIX A: EXPERIMENTAL TECHNIQUES

In this appendix, the SPDC photon pair source used is described and the technique implemented to compensate for unwanted polarization transformations induced by the birefringence of the optical fibers is explained. Moreover, the implemented experimental instruments and measurements are shown in detail.

1. Photon source

To generate photon pairs at wavelength $\lambda = 1550$ nm, we use a 775-nm, cw laser beam to pump a 30-mm-long Periodically poled potassium titanyl phosphate crystal. A dichroic mirror reflects 775-nm light and transmits photons at 1550 nm. The signal and idler photons are then separated on a PBS and coupled over optical fibers into the setup.

2. Polarization compensation

To ensure that the quantum switch performs the desired transformations on the photon's polarization, it is important to correct for the birefringent behavior of the fibers that connect Alice's and Bob's laboratories; i.e., to ensure that the fibers do not change the polarization state of the photon. Hence, each fiber link has to perform an identity operation. To this end, each fiber is equipped with a three-loop fiber polarization controller, which allows us to implement any unitary polarization transformation in the fiber. To implement a true identity operation, we must check that the correct transformation is applied in two different bases. We use the computational and diagonal bases. A convenient way to do this is to send classical light at the same wavelength as the single photons through the fibers and detect the polarization with a polarimeter at the fiber output. To ensure the identity transformation, we switch the light's polarization state at the target preparation stage (see Fig. 4) between horizontal $|H\rangle$ and diagonal $|D\rangle$, while adjusting the polarization controller until it converges to the correct setting in both bases. To correct the polarization for the second trip through the channels, we place a polarizer in one of the channels (without additional wave plates). This decouples the compensation from the previous fiber. Then we follow the same procedure and alternate the polarizer to transmit $|H\rangle$ and then $|V\rangle$. For this procedure to work properly, it is essential that the wavelengths of the classical light and the single photons are matched.

3. Experimental instruments and measurements

Here we explain in detail the measurements and instruments we implement in the laboratory, and how they relate the ideal settings discussed in the main text.

a. Future control measurement

Our control qubit is the time-bin qubit. Its past state is fixed to $|y-\rangle_C$. Ideally, we would measure the future control in three different bases described by Eqs. (22)–(24). However, because of experimental limitations, we measure the future control in only two bases:

$$\tilde{\mathcal{C}} := \{\tilde{C}_{c|z}\}_{c=1,z=1}^{c=2,z=2}, \tag{A1}$$

where

c	z	$\tilde{C}_{i\|j}$
1	1	$\|1\rangle\langle 1\|$
2	1	$\|0\rangle\langle 0\|$
1	2	$\|y-\rangle\langle y-\|$
2	2	$\|y+\rangle\langle y+\|$

(A2)

Experimentally, the measurement outcomes $c = 1$ and $c = 2$ correspond to finding the photon exiting different ports of the beam splitter (labeled “FDC” in Fig. 4). Note that the order of the output indices has swapped compared with Eqs. (22)–(24) so as to be consistent with our experimental convention.

b. Past target states

Our target system is encoded as in the polarization degree of freedom of the photon. We thus prepare its state by sending the photon to a polarizer set to transmit horizontal polarization, which we define to be the logical $|0\rangle_T$ state. We then set its state using a quarter-wave plate, followed by a half-wave plate. We can thus prepare the set of states given by

$$\tilde{\mathcal{S}} := \{\left|\tilde{\psi}_w\right\rangle\left\langle\tilde{\psi}_w\right|\}_{w=1}^{4}, \tag{A3}$$

where

QWP	HWP	w	$\|\tilde{\psi}_w\rangle$
0°	0°	1	$\|0\rangle$
0°	−45°	2	$\|1\rangle$
0°	−22.5°	3	$\|-\rangle$
−45°	0°	4	$\|y+\rangle$

(A4)

c. Alice’s and Bob’s instruments

As our target system is encoded in a polarization state, Alice and Bob must implement measure and reprepare channels on this degree of freedom. To do the measurement, they use a fixed polarizer to project the target qubit onto horizontal polarization. Using a quarter-wave plate and a half-wave plate before the polarizer, they can then set the measurement basis. Since this provides only one outcome (either the photon is transmitted or it is not transmitted, but they cannot detect the cases when a photon is absorbed by the polarizer) they must also explicitly set the wave plates to perform the orthogonal measurement so as to normalize the data to compute a probability. These measurements are defined by the set

$$\tilde{\mathcal{M}} := \{\tilde{M}_{i|j}\}_{i=1,j=1}^{i=2,j=3}, \tag{A5}$$

where the wave plate angles and resulting measurement operators are given by

QWP	HWP	i	j	$\tilde{M}_{i\|j}$
0°	0°	1	1	$\|0\rangle\langle 0\|$
0°	45°	2	1	$\|1\rangle\langle 1\|$
45°	22.5°	1	2	$\|+\rangle\langle +\|$
45°	67.5°	2	2	$\|-\rangle\langle -\|$
45°	0°	1	3	$\|y-\rangle\langle y-\|$
45°	45°	2	3	$\|y+\rangle\langle y+\|$

(A6)

Following their measurements, Alice and Bob reprepare the target state in one of four different options:

$$\tilde{\mathcal{P}} := \{\left|\tilde{\phi}_k\right\rangle\left\langle\tilde{\phi}_k\right|\}_{k=1}^{4}. \tag{A7}$$

This is again accomplished with a quarter-wave plate and a half-wave plate. Since the polarizers transmit horizontal polarization, the horizontally polarized postmeasurement photon is then rotated to one of the following states:

QWP	HWP	k	$\|\tilde{\phi}_k\rangle$
0°	0°	1	$\|0\rangle$
0°	45°	2	$\|1\rangle$
0°	22.5°	3	$\|+\rangle$
45°	0°	4	$\|y-\rangle$

(A8)

The net action of their instruments is then given by all combinations of the sets $\tilde{\mathcal{M}}$ and $\tilde{\mathcal{P}}$; that is,

$$\tilde{\mathcal{R}} := \{\tilde{R}_{i|(j,k)}\}_{i=1,j=1,k=1}^{i=2,j=3,k=4}, \tag{A9}$$

where

$$\tilde{R}_{i|(j,k)} := \tilde{M}_{i|j} \otimes \left|\tilde{\phi}_k\right\rangle\left\langle\tilde{\phi}_k\right|^T. \tag{A10}$$

We then have the following experimental setting operator:

$$\tilde{S}_{abc|xyzw} := \left|\tilde{\psi}_w\right\rangle\left\langle\tilde{\psi}_w\right|^{T^{P_\mathrm{T}}} \otimes \tilde{A}^{A_\mathrm{in}A_\mathrm{out}}_{a|x} \otimes \tilde{B}^{B_\mathrm{in}B_\mathrm{out}}_{b|y} \otimes \tilde{C}^{F_\mathrm{C}}_{c|z}, \tag{A11}$$

APPENDIX B: ADDITIONAL MEASUREMENTS

In this appendix, we describe additional measurements taken on this setup.

1. Commutation-anticommutation game

Shortly after it was proposed that the quantum switch offers a computational advantage over conventional quantum circuits for certain tasks [13,29], one such task—the commutation-anticommutation game—was experimentally realized with a two-party switch [35]. This task takes the form of a game, in which a referee provides a player with two unitary gates $\hat{A}$ and $\hat{B}$. The referee states that these two unitary gates either commute or anticommute, and the player's goal is to decide which statement is true. However, the player is permitted to use each gate only once. When the player has access to a quantum switch, that player can win the game perfectly. However, When the player is constrained to use a conventional (causally ordered) quantum circuit, the player would need to query one gate at least twice. The switch result can be easily understood by looking at the output state of the switch after application of a Hadamard operation on the control qubit. In the experiment, this is realized when the time bins are recombined on the beam splitter:

$$|\Psi\rangle = \frac{1}{2}\,|0\rangle_C\left\{\hat{A}\hat{B}\right\}|\Psi\rangle_T + \frac{1}{2}\,|1\rangle_C\left[\hat{A}\hat{B}\right]|\Psi\rangle_T. \tag{B1}$$

Since the player is sure that either the commutator or the anticommutator is zero, by measuring the control and finding it in $|0\rangle_C$, the player is sure the gates commute, while if the player finds it in $|1\rangle_C$, the player can conclude that the gates anticommute.

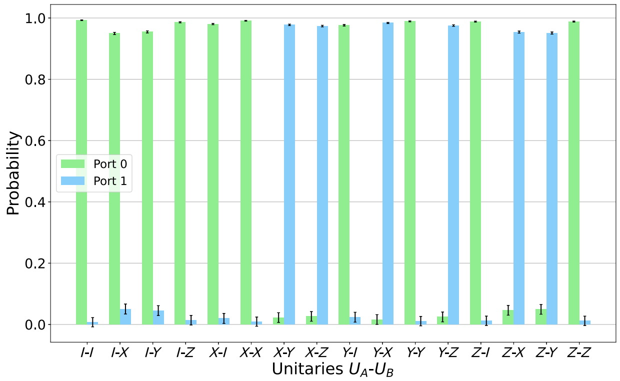

FIG. 8. Commutation game. The experimentally estimated probabilities for the commutation-anticommutation game. The green bars represent the probability that the photon leaves port 0, leading to the conclusion that the unitaries commute. The blue bars show the probability of the photon exiting port 1, and thus the conclusion that the gates anticommute. The x axis shows which gates have been implemented for $\hat{U}_A$-$\hat{U}_B$.

Before performing higher-order process tomography, we first verified the correct performance of the new architecture of our quantum switch by implementing this game. To do so, Alice and Bob could set the Pauli operators $X = \sigma_X$, $Y = \sigma_Y$, and $Z = \sigma_Z$ and the identity operator $I = \sigma_0$ with the well-known commutation relation $\left[\hat{\sigma}_i\hat{\sigma}_j\right] = 2i\epsilon_{ijk}\hat{\sigma}_k$ and anticommutation relation $\left\{\hat{\sigma}_i\hat{\sigma}_j\right\} = 2\delta_ij\,\mathbb{1}_{2x2}$ [71]. Figure 8 shows the probabilities for measuring the control either in state $|0\rangle_C$ or in state $|1\rangle_C$.

The probabilities were calculating by our measuring coincidence rates between the heralding detector and both output ports of the MZI with a coincidence window of 5 ns between them. Data were acquired for a total of 10 s for each setting. The visibility of the interferometer was measured as $v^2 = 0.97$. We calculated the success probability for the photon leaving the correct port as

$$\begin{aligned} p_\mathrm{succ} &= \frac{1}{2}\left(\bar{p}\left(0|[\cdot,\cdot]\right) + \bar{p}\left(1|\{\cdot,\cdot\}\right)\right) \\ &= 0.974 \pm 0.18. \end{aligned} \tag{B2}$$

TABLE III. Generalized robustness alternative definition witness analysis. A summary of the different generalized robustness witnesses constructed with use of the definition in Ref. [37], presented in Eq. (C8). The witnesses $G_{i,j}$ are labeled by two subscripts. The first indicates if the witness was designed for the ideal process matrix W_s^{y-} (subscript "$y-$") or the experimental process matrix W_expt (subscript "expt"). The second subscript indicates if the restricted measurement set (subscript "res") or the complete measurement set (subscript "all") was used for the witness. The first row shows the value of the witness for W_s^{y-} and the second row shows the experimental values, which were evaluated as $\mathrm{Tr}[G_{i,j}\,W_\mathrm{expt}]$.

	$G^\mathrm{GR}_{y-,\mathrm{res}}$	$G^\mathrm{GR}_{\mathrm{expt,res}}$	$G^\mathrm{GR}_{y-,\mathrm{all}}$	$G^\mathrm{GR}_{\mathrm{expt,all}}$
W_s^{y-}	−0.500	−0.483	−0.500	−0.484
W_expt	−0.346 ± 0.003	−0.361 ± 0.003	−0.323 ± 0.003	−0.363 ± 0.003

TABLE IV. White noise alternative definition witness analysis. A summary of the different white noise witnesses constructed with use of the definition in Ref. [37], presented in Eq. (C8). The witnesses $G_{i,j}$ and process matrices are labeled by two subscripts described in the caption for Table III.

	$G^{\mathrm{WN}}_{y-,\mathrm{res}}$	$G^{\mathrm{WN}}_{\mathrm{expt,res}}$	$G^{\mathrm{WN}}_{y-,\mathrm{all}}$	$G^{\mathrm{WN}}_{\mathrm{expt,all}}$
W_{s}^{y-}	-0.828	-0.800	-1.000	-0.965
W_{expt}	-0.572 ± 0.005	-0.614 ± 0.006	-0.689 ± 0.007	-0.739 ± 0.008

Thus, our new architecture achieves performance similar to that of previous implementations [35], but now in a phase-stable manner.

APPENDIX C: THE DEFINITION OF CAUSAL NONSEPARABILITY

Let $W \in \mathcal{L}\left(\mathcal{H}_P \otimes \mathcal{H}_{A_{\mathrm{in}}} \otimes \mathcal{H}_{A_{\mathrm{out}}} \otimes \mathcal{H}_{B_{\mathrm{in}}} \otimes \mathcal{H}_{B_{\mathrm{out}}} \otimes \mathcal{H}_F\right)$ be a bipartite process matrix with a common past and common future. A process $W^{A>B}$ is said to be causally ordered from A to B if it can be constructed by means of a sequential quantum circuit [6,13], also called a "quantum comb" [2,8]). More explicitly, a process matrix $W^{A>B} \in \mathcal{L}\left(\mathcal{H}_P \otimes \mathcal{H}_{A_{\mathrm{in}}} \otimes \mathcal{H}_{A_{\mathrm{out}}} \otimes \mathcal{H}_{B_{\mathrm{in}}} \otimes \mathcal{H}_{B_{\mathrm{out}}} \otimes \mathcal{H}_F\right)$ is causally ordered from A to B if it respects

$$\mathrm{tr}_F(W^{A>B}) = \mathrm{tr}_{B_{\mathrm{out}}F}(W^{A>B}) \otimes \frac{\mathbb{1}_{B_{\mathrm{out}}}}{d_{B_{\mathrm{out}}}}, \tag{C1}$$

$$\mathrm{tr}_{B_{\mathrm{in}}B_{\mathrm{out}}F}(W^{A>B}) = \mathrm{tr}_{A_{\mathrm{out}}B_{\mathrm{in}}B_{\mathrm{out}}F}(W^{A>B}) \otimes \frac{\mathbb{1}_{A_{\mathrm{out}}}}{d_{A_{\mathrm{out}}}}, \tag{C2}$$

$$\mathrm{tr}_{A_{\mathrm{in}}A_{\mathrm{out}}B_{\mathrm{in}}B_{\mathrm{out}}F}(W^{A>B}) = \mathrm{tr}_{PA_{\mathrm{in}}A_{\mathrm{out}}B_{\mathrm{in}}B_{\mathrm{out}}F}(W^{A>B}) \otimes \frac{\mathbb{1}_P}{d_P}. \tag{C3}$$

Analogously, a process matrix $W^{B>A} \in \mathcal{L}\left(\mathcal{H}_P \otimes \mathcal{H}_{A_{\mathrm{in}}} \otimes \mathcal{H}_{A_{\mathrm{out}}} \otimes \mathcal{H}_{B_{\mathrm{in}}} \otimes \mathcal{H}_{B_{\mathrm{out}}} \otimes \mathcal{H}_F\right)$ is causally ordered from B to A if it respects

$$\mathrm{tr}_F(W^{B>A}) = \mathrm{tr}_{A_{\mathrm{out}}F}(W^{B>A}) \otimes \frac{\mathbb{1}_{A_{\mathrm{out}}}}{d_{A_{\mathrm{out}}}}, \tag{C4}$$

$$\mathrm{tr}_{A_{\mathrm{in}}A_{\mathrm{out}}F}(W^{B>A}) = \mathrm{tr}_{B_{\mathrm{out}}A_{\mathrm{in}}A_{\mathrm{out}}F}(W^{B>A}) \otimes \frac{\mathbb{1}_{B_{\mathrm{out}}}}{d_{B_{\mathrm{out}}}}, \tag{C5}$$

$$\mathrm{tr}_{A_{\mathrm{in}}A_{\mathrm{out}}B_{\mathrm{in}}B_{\mathrm{out}}F}(W^{B>A}) = \mathrm{tr}_{PA_{\mathrm{in}}A_{\mathrm{out}}B_{\mathrm{in}}B_{\mathrm{out}}F}(W^{B>A}) \otimes \frac{\mathbb{1}_P}{d_P}. \tag{C6}$$

In the main part of this work, we follow the definition in Ref. [43], where a bipartite process matrix $W \in \mathcal{L}\left(\mathcal{H}_P \otimes \mathcal{H}_{A_{\mathrm{in}}} \otimes \mathcal{H}_{A_{\mathrm{out}}} \otimes \mathcal{H}_{B_{\mathrm{in}}} \otimes \mathcal{H}_{B_{\mathrm{out}}} \otimes \mathcal{H}_F\right)$ with common past and common future is said to be causally separable if it can be written as a convex sum of causally ordered process matrices; that is, if we can write

$$W = pW^{A>B} + (1-p)W^{B>A}, \tag{C7}$$

where $p \in [0,1]$ and where $W^{A>B}$ and $W^{B>A}$ are causally ordered processes. However, as discussed earlier, there exists a nonequivalent definition, which goes beyond a simple convex combination and allows incoherent classical control of causal orders. This definition is presented in Ref. [96], and it is proven to be equivalent to the notion of extensible causal, presented in Ref. [11]. In the bipartite scenario with a common past and common future, in Ref. [96] it is stated that a process matrix $W \in \mathcal{L}\left(\mathcal{H}_P \otimes \mathcal{H}_{A_{\mathrm{in}}} \otimes \mathcal{H}_{A_{\mathrm{out}}} \otimes \mathcal{H}_{B_{\mathrm{in}}} \otimes \mathcal{H}_{B_{\mathrm{out}}} \otimes \mathcal{H}_F\right)$ is causally separable when there exist causally ordered processes $W^{A>B}$ and $W^{B>A}$ such that

$$\frac{\mathbb{1}_P}{d_P} \otimes \mathrm{tr}_P(W) = pW^{A>B} + (1-p)W^{B>A}. \tag{C8}$$

Notice that the definition in Ref. [96] presented in Eq. (C8) is more relaxed than the one in Ref. [43] and presented in Eq. (C7). Indeed, as we show later, there are processes that are causally separable following the definition in Ref. [96] but are causally nonseparable following the definition in Ref. [43].

In the definition in Ref. [96], for a causal witness to be valid, one should include an extra constraint, which reads

$$G = \frac{\mathbb{1}_P}{d_P} \otimes \mathrm{tr}_P(G). \tag{C9}$$

We can then recalculate the numbers presented in Tables I and II, but following the definition in Ref. [96]. These results are presented in Tables III and IV, and we notice that although there are some differences in the numbers obtained, the qualitative result remains the same.

[1] P. Selinger and B. Valiron, in *Semantic Techniques in Quantum Computation*, edited by S. Gay and I. Mackie (Cambridge University Press, 2009), p. 135.

[2] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Quantum circuit architecture, Phys. Rev. Lett. **101**, 060401 (2008).

[3] G. Gutoski and J. Watrous, in *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing* (2007) p. 565.

[4] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Theoretical framework for quantum networks, Phys. Rev. A **80**, 022339 (2009).

[5] A. Bisio and P. Perinotti, Theoretical framework for higher-order quantum theory, Proc. R. Soc. Lond. Ser. A **475**, 20180706 (2019).

[6] O. Oreshkov, F. Costa, and Č. Brukner, Quantum correlations with no causal order, Nat. Commun. **3**, 1092 (2012).

[7] L. Hardy, Towards quantum gravity: A framework for probabilistic theories with non-fixed causal structure, J. Phys. A: Math. Gen. **40**, 3081 (2007).

[8] G. Chiribella, G. M. D'Ariano, P. Perinotti, and B. Valiron, Quantum computations without definite causal structure, Phys. Rev. A **88**, 022318 (2013).

[9] Č Brukner, Quantum causality, Nat. Phys. **10**, 259 (2014).

[10] Ä. Baumeler and S. Wolf, in *2014 IEEE International Symposium on Information Theory* (IEEE, Honolulu, HI, USA, 2014), p. 526.

[11] O. Oreshkov and C. Giarmatzi, Causal and causally separable processes, New J. Phys. **18**, 093020 (2016).

[12] Ä. Baumeler and S. Wolf, The space of logically consistent classical processes without causal order, New J. Phys. **18**, 013036 (2016).

[13] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi, and Č. Brukner, Witnessing causal nonseparability, New J. Phys. **17**, 102001 (2015).

[14] C. Branciard, Witnesses of causal nonseparability: an introduction and a few case studies, Sci. Rep. **6**, 26018 (2016).

[15] M. Zych, F. Costa, I. Pikovski, and Č. Brukner, Bell's theorem for temporal order, Nat. Commun. **10**, 1 (2019).

[16] L. Hardy, in *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle* (Springer, 2009), p. 379, https://link.springer.com/chapter/10.1007/978-1-4020-9107-0_21.

[17] M. Araújo, F. Costa, and Č. Brukner, Computational advantage from quantum-controlled ordering of gates, Phys. Rev. Lett. **113**, 250402 (2014).

[18] M. Araújo, P. A. Guérin, and ń. Baumeler, Quantum computation with indefinite causal structures, Phys. Rev. A **96**, 052315 (2017).

[19] M. J. Renner and Č. Brukner, Computational advantage from a quantum superposition of qubit gate orders, Phys. Rev. Lett. **128**, 230503 (2022).

[20] A. Feix, M. Araújo, and Č. Brukner, Quantum superposition of the order of parties as a communication resource, Phys. Rev. A **92**, 052326 (2015).

[21] P. A. Guérin, A. Feix, M. Araújo, and Č. Brukner, Exponential communication complexity advantage from quantum superposition of the direction of communication, Phys. Rev. Lett. **117**, 100502 (2016).

[22] D. Ebler, S. Salek, and G. Chiribella, Enhanced communication with the assistance of indefinite causal order, Phys. Rev. Lett. **120**, 120502 (2018).

[23] D. Jia and F. Costa, Causal order as a resource for quantum communication, Phys. Rev. A **100**, 052319 (2019).

[24] H. Kristjánsson, G. Chiribella, S. Salek, D. Ebler, and M. Wilson, Resource theories of communication, New J. Phys. **22**, 073014 (2020).

[25] G. Chiribella and H. Kristjánsson, Quantum Shannon theory with superpositions of trajectories, Proc. R. Soc. Lond. Ser. A **475**, 20180903 (2019).

[26] G. Chiribella, M. Banik, S. S. Bhattacharya, T. Guha, M. Alimuddin, A. Roy, S. Saha, S. Agrawal, and G. Kar, Indefinite causal order enables perfect quantum communication with zero capacity channels, New J. Phys. **23**, 033039 (2021).

[27] M. Caleffi and A. S. Cacciapuoti, Quantum switch for the quantum internet: Noiseless communications through noisy channels, IEEE J. Sel. Areas Commun. **38**, 575 (2020).

[28] L. M. Procopio, F. Delgado, M. Enríquez, N. Belabas, and J. A. Levenson, Sending classical information via three noisy channels in superposition of causal orders, Phys. Rev. A **101**, 012346 (2020).

[29] G. Chiribella, Perfect discrimination of no-signalling channels via quantum superposition of causal structures, Phys. Rev. A **86**, 040301 (2012).

[30] J. Bavaresco, M. Murao, and M. T. Quintino, Strict hierarchy between parallel, sequential, and indefinite-causal-order strategies for channel discrimination, Phys. Rev. Lett. **127**, 200504 (2021).

[31] X. Zhao, Y. Yang, and G. Chiribella, Quantum metrology with indefinite causal order, Phys. Rev. Lett. **124**, 190503 (2020).

[32] M. T. Quintino, Q. Dong, A. Shimbo, A. Soeda, and M. Murao, Reversing unknown quantum transformations: Universal quantum circuit for inverting general unitary operations, Phys. Rev. Lett. **123**, 210502 (2019).

[33] P. Schiansky, T. Strömberg, D. Trillo, V. Saggio, B. Dive, M. Navascués, and P. Walther, Demonstration of universal time-reversal for qubit processes, Optica **10**, 200 (2023).

[34] D. Felce and V. Vedral, Quantum refrigeration with indefinite causal order, Phys. Rev. Lett. **125**, 070603 (2020).

[35] L. M. Procopio, A. Moqanaki, M. Araújo, F. Costa, I. Alonso Calafell, E. G. Dowd, D. R. Hamel, L. A. Rozema, Č. Brukner, and P. Walther, Experimental superposition of orders of quantum gates, Nat. Commun. **6**, 7913 (2015).

[36] K. Goswami, Y. Cao, G. A. Paz-Silva, J. Romero, and A. G. White, Increasing communication capacity via superposition of order, Phys. Rev. Res. **2**, 033292 (2020).

[37] K. Wei, N. Tischler, S.-R. Zhao, Y.-H. Li, J. M. Arrazola, Y. Liu, W. Zhang, H. Li, L. You, Z. Wang, Y.-A. Chen, B. C. Sanders, Q. Zhang, G. J. Pryde, F. Xu, and J.-W. Pan, Experimental quantum switching for exponentially superior quantum communication complexity, Phys. Rev. Lett. **122**, 120504 (2019).

[38] Y. Guo, X.-M. Hu, Z.-B. Hou, H. Cao, J.-M. Cui, B.-H. Liu, Y.-F. Huang, C.-F. Li, G.-C. Guo, and G. Chiribella, Experimental transmission of quantum information using a superposition of causal orders, Phys. Rev. Lett. **124**, 030502 (2020).

[39] G. Rubino, L. A. Rozema, D. Ebler, H. Kristjánsson, S. Salek, P. Allard Guérin, A. A. Abbott, C. Branciard, Č. Brukner, G. Chiribella, and P. Walther, Experimental quantum communication enhancement by superposing trajectories, Phys. Rev. Res. **3**, 013093 (2021).

[40] M. M. Taddei, J. Cariñe, D. Martínez, T. García, N. Guerrero, A. A. Abbott, M. Araújo, C. Branciard, E. S.

Gómez, S. P. Walborn, L. Aolita, and G. Lima, Computational advantage from the quantum superposition of multiple temporal orders of photonic gates, PRX Quantum **2**, 010320 (2021).

[41] A. O. Pang, N. Lupu-Gladstein, H. Ferretti, Y. B. Yilmaz, A. Brodutch, and A. M. Steinberg, Experimental communication through superposition of quantum channels, arXiv preprint arXiv:2302.14820 (2023).

[42] W.-Q. Liu, Z. Meng, B.-W. Song, J. Li, Q.-Y. Wu, X.-X. Chen, J.-Y. Hong, A.-N. Zhang, and Z.-q. Yin, Experimentally demonstrating indefinite causal order algorithms to solve the generalized Deutsch's problem, arXiv e-prints arXiv:2305.05416 [quant-ph] (2023).

[43] G. Rubino, L. A. Rozema, A. Feix, M. Araújo, J. M. Zeuner, L. M. Procopio, Č. Brukner, and P. Walther, Experimental verification of an indefinite causal order, Sci. Adv. **3**, e1602589 (2017).

[44] K. Goswami, C. Giarmatzi, M. Kewming, F. Costa, C. Branciard, J. Romero, and A. G. White, Indefinite causal order in a quantum switch, Phys. Rev. Lett. **121**, 090503 (2018).

[45] G. Rubino, L. A. Rozema, F. Massa, M. Araújo, M. Zych, Č. Brukner, and P. Walther, Experimental entanglement of temporal order, Quantum **6**, 621 (2022).

[46] H. Cao, J. Bavaresco, N.-N. Wang, L. A. Rozema, C. Zhang, Y.-F. Huang, B.-H. Liu, C.-F. Li, G.-C. Guo, and P. Walther, Semi-device-independent certification of indefinite causal order in a photonic quantum switch, Optica **10**, 561 (2023).

[47] T. M. Rambo, J. B. Altepeter, P. Kumar, and G. M. D'Ariano, Functional quantum computing: An optical approach, Phys. Rev. A **93**, 052321 (2016).

[48] G. Chiribella and Z. Liu, Quantum operations with indefinite time direction, Commun. Phys. **5**, 190 (2022).

[49] Y. Guo, Z. Liu, H. Tang, X.-M. Hu, B.-H. Liu, Y.-F. Huang, C.-F. Li, G.-C. Guo, and G. Chiribella, Experimental demonstration of input-output indefiniteness in a single quantum device, arXiv e-prints arXiv:2210.17046 [quant-ph] (2022).

[50] T. Strömberg, P. Schiansky, M. Túlio Quintino, M. Antesberger, L. Rozema, I. Agresti, Č. Brukner, and P. Walther, Experimental superposition of time directions, arXiv e-prints arXiv:2211.01283 [quant-ph] (2022).

[51] Polarization has also been used [36,44] as a control system, but also in this case the polarization is used to route the photon into two paths in superposition, which also requires a stable phase.

[52] T. Strömberg, P. Schiansky, R. W. Peterson, M. Túlio Quintino, and P. Walther, Demonstration of a quantum SWITCH in a Sagnac configuration, arXiv e-prints, arXiv:2211.12540 [quant-ph] (2022).

[53] N. Paunković and M. Vojinović, Causal orders, quantum circuits and spacetime: Distinguishing between definite and superposed causal orders, Quantum **4**, 275 (2020).

[54] V. Vilasini and R. Renner, Embedding cyclic causal structures in acyclic spacetimes: No-go results for process matrices, arXiv preprint arXiv:2203.11245 [quant-ph] (2022).

[55] N. Ormrod, A. Vanrietvelde, and J. Barrett, Causal structure in the presence of sectorial constraints, with application to the quantum switch, arXiv preprint arXiv:2204.10273 [quant-ph] (2022).

[56] O. Oreshkov, Time-delocalized quantum subsystems and operations: On the existence of processes with indefinite causal structure in quantum mechanics, Quantum **3**, 206 (2019).

[57] A.-C. de la Hamette, V. Kabel, M. Christodoulou, and Č. Brukner, Quantum diffeomorphisms cannot make indefinite causal order definite, arXiv e-prints arXiv:2211.15685 [quant-ph] (2022).

[58] M. Fellous-Asiani, R. Mothe, L. Bresque, H. Dourdent, P. A. Camati, A. A. Abbott, A. Auffèves, and C. Branciard, Comparing the quantum switch and its simulations with energetically constrained operations, Phys. Rev. Res. **5**, 023111 (2023).

[59] O. Gühne and G. Tóth, Entanglement detection, Phys. Rep. **474**, 1 (2009).

[60] J. Bavaresco, M. Araújo, Č. Brukner, and M. T. Quintino, Semi-device-independent certification of indefinite causal order, Quantum **3**, 176 (2019).

[61] H. Dourdent, A. A. Abbott, N. Brunner, I. Šupić, and C. Branciard, Semi-device-independent certification of causal nonseparability with trusted quantum inputs, Phys. Rev. Lett. **129**, 090402 (2022).

[62] T. van der Lugt, J. Barrett, and G. Chiribella, Device-independent certification of indefinite causal order in the quantum switch, arXiv e-prints arXiv:2208.00719 [quant-ph] (2022).

[63] S. Gogioso and N. Pinzani, The topology and geometry of causality, arXiv e-prints arXiv:2206.08911 [quant-ph] (2022).

[64] M. Ziman, Process positive-operator-valued measure: A mathematical framework for the description of process tomography experiments, Phys. Rev. A **77**, 062112 (2008).

[65] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, J. Mod. Opt. **44**, 2455 (1997).

[66] J. Poyatos, J. I. Cirac, and P. Zoller, Complete characterization of a quantum process: The two-bit quantum gate, Phys. Rev. Lett. **78**, 390 (1997).

[67] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. James, N. K. Langford, T. C. Ralph, and A. G. White, Quantum process tomography of a controlled-NOT gate, Phys. Rev. Lett. **93**, 080502 (2004).

[68] A. Bisio, G. Chiribella, G. Mauro D'Ariano, S. Facchini, and P. Perinotti, Optimal quantum tomography, arXiv e-prints , arXiv:1702.08751 (2017).

[69] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Quantum-process tomography: Resource analysis of different strategies, Phys. Rev. A **77**, 032322 (2008).

[70] L. A. Rozema, D. H. Mahler, R. Blume-Kohout, and A. M. Steinberg, Optimizing the choice of spin-squeezed states for detecting and characterizing quantum processes, Phys. Rev. X **4**, 041025 (2014).

[71] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2005), 8th ed.

[72] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi, Non-Markovian quantum

processes: Complete framework and efficient characterization, Phys. Rev. A **97**, 012127 (2018).

[73] A. Bisio, G. Chiribella, G. M. D'Ariano, S. Facchini, and P. Perinotti, Optimal quantum learning of a unitary transformation, Phys. Rev. A **81**, 032324 (2010).

[74] M. Sedlák, A. Bisio, and M. Ziman, Optimal probabilistic storage and retrieval of unitary channels, Phys. Rev. Lett. **122**, 170502 (2019).

[75] J. Bavaresco, M. Murao, and M. T. Quintino, Unitary channel discrimination beyond group structures: Advantages of sequential and indefinite-causal-order strategies, J. Math. Phys. **63**, 042203 (2022).

[76] Q. Dong, M. T. Quintino, A. Soeda, and M. Murao, The quantum switch is uniquely defined by its action on unitary operations, Quantum **7**, 1169 (2023).

[77] G. A. L. White, F. A. Pollock, L. C. L. Hollenberg, K. Modi, and C. D. Hill, Non-Markovian quantum process tomography, PRX Quantum **3**, 020344 (2022).

[78] G. A. L. White, C. D. Hill, F. A. Pollock, L. C. L. Hollenberg, and K. Modi, Demonstration of non-Markovian process characterisation and control on a quantum processor, Nat. Commun. **11**, 6301 (2020).

[79] G. A. L. White, F. A. Pollock, L. C. L. Hollenberg, C. D. Hill, and K. Modi, From many-body to many-time physics, arXiv e-prints arXiv:2107.13934 [quant-ph] (2021).

[80] A. Jamiołkowski, Linear transformations which preserve trace and positive semidefiniteness of operators, Rep. Math. Phys. **3**, 275 (1972).

[81] M.-D. Choi, Completely positive linear maps on complex matrices, Linear Algebra Appl. **10**, 285 (1975).

[82] Upper indices on vectors and operators are used to indicate where these objects act. For instance, in $|U\rangle\!\rangle^{\mathcal{H}_{\text{in}}\mathcal{H}_{\text{out}}}$, the upper indices indicate that $|U\rangle\!\rangle \in \mathcal{H}_{\text{in}} \otimes \mathcal{H}_{\text{out}}$. Whenever it is clear from the context, we might drop the upper indices to avoid overly heavy notation.

[83] A set of operators $\{M_i\}$ represents a quantum measurement (i.e., it is a POVM) if $M_i \geq 0$ and $\sum_i M_i = \mathbb{1}$.

[84] S. Gogioso and N. Pinzani, The topology and geometry of causality, arXiv e-prints arXiv:2206.08911 (2022).

[85] We use the phrase "setting operator," since each of these operators will correspond to a single experimental setting.

[86] W. K. Wootters, Local accessibility of quantum states, Complexity, Entropy and the Physics of Information, edited by W. H. Zurek (Addison-Wesley, Reading, MA, 1990), pp. 39–46.

[87] M. D. de Burgh, N. K. Langford, A. C. Doherty, and A. Gilchrist, Choice of measurement sets in qubit tomography, Phys. Rev. A **78**, 052122 (2008).

[88] Note that here we identify $x = (j, k)$; that is, x can assume $12 = 3 \times 4$ different values.

[89] S. Milz and M. Túlio Quintino, Transformations between arbitrary (quantum) objects and the emergence of indefinite causality, arXiv e-prints arXiv:2305.01247 [quant-ph] (2023).

[90] M. Antesberger, M. T. Quintino, P. Walther, and L. A. Rozema, Raw data and analysis code for "Higher-order process matrix tomography of a passively-stable quantum SWITCH", (2023).

[91] F. Vedovato, C. Agnesi, M. Tomasin, M. Avesani, J.-Å. Larsson, G. Vallone, and P. Villoresi, Postselection-loophole-free Bell violation with genuine time-bin entanglement, Phys. Rev. Lett. **121**, 190401 (2018).

[92] G. L. Zanin, M. J. Jacquet, M. Spagnolo, P. Schiansky, I. A. Calafell, L. A. Rozema, and P. Walther, Fiber-compatible photonic feed-forward with 99% fidelity, Opt. Express **29**, 3425 (2021).

[93] G. L. Zanin, M. Antesberger, M. J. Jacquet, P. H. S. Ribeiro, L. A. Rozema, and P. Walther, Enhanced photonic Maxwell's demon with correlated baths, Quantum **6**, 810 (2022).

[94] T. Kim, M. Fiorentino, and F. N. C. Wong, Phase-stable source of polarization-entangled photons using a polarization Sagnac interferometer, Phys. Rev. A **73**, 012316 (2006).

[95] Our source is in a Sagnac configuration, although for this experiment we pump the Sagnac loop in only one direction so as not to generate polarization entanglement.

[96] J. Wechs, A. A. Abbott, and C. Branciard, On the definition and characterisation of multipartite causal (non)separability, New J. Phys. **21**, 013027 (2019).

[97] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, 2004).

[98] A. A. Abbott, C. Giarmatzi, F. Costa, and C. Branciard, Multipartite causal correlations: Polytopes and inequalities, Phys. Rev. A **94**, 032131 (2016).

[99] A. A. Abbott, J. Wechs, F. Costa, and C. Branciard, Genuinely multipartite noncausality, Quantum **1**, 39 (2017).

[100] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, Quantum tomography via compressed sensing: Error bounds, sample complexity and efficient estimators, New J. Phys. **14**, 095022 (2012).

[101] S. Aaronson, Shadow tomography of quantum states (2020), p. STOC18–368, arXiv:1711.01053 [quant-ph].

[102] D. H. Mahler, L. A. Rozema, A. Darabi, C. Ferrie, R. Blume-Kohout, and A. M. Steinberg, Adaptive quantum state tomography improves accuracy quadratically, Phys. Rev. Lett. **111**, 183601 (2013).