Heralded Multiplexed High-Efficiency Cascaded Source of Dual-Rail Entangled Photon Pairs Using Spontaneous Parametric Down-Conversion

Prajit Dhara[1,2], Spencer J. Johnson[3,4], Christos N. Gagatsos[1,2], Paul G. Kwiat[3,4], and Saikat Guha[1,2,*]

[1]*Wyant College of Optical Sciences, The University of Arizona, Tucson, Arizona 85721, USA*
[2]*NSF-ERC Center for Quantum Networks, The University of Arizona, Tucson, Arizona 85721, USA*
[3]*Department of Physics, University of Illinois Urbana-Champaign, Urbana, Illinois 61801, USA*
[4]*Illinois Quantum Information Science and Technology Center, University of Illinois Urbana-Champaign, Urbana, Illinois 61801, USA*

Deterministic sources of high-fidelity entangled qubit pairs encoded in the dual-rail photonic basis, i.e., presence of a single photon in one of two orthogonal modes, are a key enabling technology of many applications of quantum information processing, including high-rate, high-fidelity quantum communications over long distances. The most popular and mature sources of such photonic entanglement, e.g., those that leverage spontaneous parametric down-conversion (SPDC) or spontaneous four-wave mixing, generate an entangled (so-called continuous-variable) quantum state that contains contributions from high-order photon terms that lie outside the span of the dual-rail basis, which is detrimental to most applications. One often uses low pump power to mitigate the effects of those high-order terms. However, that reduces the pair generation rate, and the source becomes inherently probabilistic. We investigate a *cascaded* source that performs a linear-optical entanglement swap between two SPDC sources, to generate a *heralded* photonic entangled state that has a higher fidelity (to the ideal Bell state) compared to a free-running SPDC source. Furthermore, with the Bell swap providing a heralding trigger, we show how to build a multiplexed source, which despite reasonable switching losses and detector loss and noise, yields a fidelity versus success probability trade-off of a high-efficiency source of high-fidelity dual-rail photonic entanglement. We find, however, that there is a threshold of 1.5 dB of loss per switch, beyond which multiplexing hurts the fidelity versus success probability trade-off.

I. INTRODUCTION

Distributed high-fidelity entanglement will become a commodity as its demand stemming from a variety of promising applications increases. As the world makes progress towards realizing the vision of a *quantum internet* [1] to generate entanglement among many user groups at high rates, some of the biggest remaining enabling-technology challenges are: (1) scalable sources of high-fidelity on-demand photonic entanglement, (2) high-efficiency, high-bandwidth, high-coherence-time universal-quantum-logic-capable quantum memories, and (3) high-efficiency converters between various qubit forms native to the leading quantum-memory contenders and optical-frequency photonic qubits.

While there are no viable alternatives to optical-frequency qubits for long-distance transmission, there are many ways to encode a qubit in the photon [2]. Two of

those most commonly studied are: (a) the Knill-Laflamme-Milburn (KLM) dual-rail photonic qubit [3] where the presence of a single photon in one of two orthogonal (spatial, spectral, temporal, or polarization) modes encodes the two logical quantum states of a qubit; and (b) the Gottesman-Kitaev-Preskill (GKP) encoding [4], which encodes the qubit in a single bosonic mode excited in one of two coherent superpositions of displaced quadrature-squeezed states that are shifted with respect to one another in the phase space. Dual-rail qubits are conceptually easy to produce and manipulate using passive linear optics, but they need high-fidelity single-photon sources and single-photon detectors. Quantum logic on dual-rail qubits using passive linear optics and single-photon detectors is simple to build. Despite the gates being inherently probabilistic in the original KLM scheme [3], recent advancements on single-photon ancilla-assisted boosted linear-optical quantum logic [5] has ushered linear optical quantum computing using dual-rail qubits into highly scalable architectures [6–8]. Alternatively, the GKP qubit is known to be

*saikat@arizona.edu

the most loss-resilient photonic qubit encoding [2,9], and Clifford quantum logic is deterministically implementable using squeezers and linear optics [4]. However, not only are they hard to produce [10,11], there is no known way to store GKP qubits and GKP-basis entangled states in heralded quantum memories. In this paper, we therefore focus on the dual-rail qubit. The multiplexed heralded entanglement generation ideas we present here, however, are applicable to other photonic qubit encodings and to other (e.g., multiqubit) entangled states.

We focus in this work on the first challenge mentioned in the first paragraph above: that of designing an on-demand photonic entanglement source that produces high-fidelity two-qubit entangled Bell states with the qubits encoded in the dual-rail photonic basis. There have been many calculations of quantum repeater protocols [12,13] and quantum network routing algorithms [14,15] that assume the availability of unit-fidelity sources of dual-rail photonic entanglement. This results in these analyses predicting, despite inclusion of linear losses everywhere in the system, pristine dual-rail Bell states, i.e., entangled bits (ebits), being delivered to the communicating parties Alice and Bob. In reality, sources that deterministically generate dual-rail Bell states suitable for communications are quite challenging to build. Subunity entanglement fidelity has been incorporated in recent work on entanglement routing [16], but those have restricted their analyses to ideal Werner-like entangled states. Quantum dot sources [17,18] and other forms of quantum emitters [19] can theoretically generate single photons and entangled photon pairs on demand. Recently, these sources have achieved high (polarization) entanglement fidelity [20] and over 60% coupling efficiency into a single-mode optical fiber [21], though not in the same experiment. Moreover, the photon frequencies from individual emitters can vary slightly, and the emitted photon frequency is usually not compatible with telecommunications hardware. The most common and reliable sources of dual-rail entanglement used in practice rely on spontaneous parametric down-conversion (SPDC) [22–24], wherein single photons from a strong pump laser impinging on a carefully phase matched (possibly periodically poled) $\chi^{(2)}$ crystal splits into entangled photon pairs at two frequencies. Alternatively, one can employ the process of spontaneous four-wave mixing (FWM), in which a pair of pump photons give rise to an entangled pair [25,26]. For further analyses, we refer to SPDC but the conclusions would be equally valid for FWM sources.

There are many variants of SPDC-based entangled photon pair generation methods. However, a detailed physical analysis of these sources has shown that the complete quantum state generated, described by two copies of the so-called two-mode squeezed vacuum, contains contributions from vacuum and higher-order terms, e.g., two-photon-pair terms, in addition to the desired dual-rail Bell state, which can adversely affect both the distribution rates

and the fidelity of the distributed entanglement [27–29]. In fact, the pump power must be carefully optimized to maximize the entanglement rate, while adhering to a desired fidelity threshold. One common strategy is to turn down the pump power so low that the probability that the source produces two-pair (and higher-order) states becomes negligible. Of course, this increases the contribution of vacuum to the emitted state and reduces the rate at which the desired Bell states are produced. The vacuum term often does not affect the usability of the source in an application, either because it gets filtered out by the "click" of a detector, e.g., in a quantum key distribution (QKD) experiment that provides a *postselection trigger* to consider only those times slots that *had* a photon in it; or because a quantum memory provides a *heralding trigger* declaring that it successfully loaded a dual-rail photonic qubit into its native qubit domain (hence filtering out the vacuum).

Other than the reduced pair-production rate of the above strategy of turning down the pump power, another inherent problem with such a "free-running" standalone SPDC-based entanglement source is that it is probabilistic, and does not have a heralding trigger. In other words, we cannot in principle know in which time slot the source actually produced an entangled photon pair, a major detriment in many applications. One method to increase the probability of emitting a single photon into a particular time slot is to use a heralded single-photon source, e.g., from SPDC, combined with spatial, temporal [30,31], or spectral multiplexing [32,33]. However, for this to work with entangled pairs would require one of the photons to be detected immediately, undesirable for many applications. A second option is to use (four) single photons possibly from a multiplexed heralded single-photon source as inputs to a quantum circuit that probabilistically produces heralded entangled pairs [34–36]; combined with multiplexing this could enable "entanglement on demand." Schemes to simultaneously generate multiple indistinguishable single photons in different spatial modes in "parallel states" have been studied by Ref. [37]. These parallel states can be utilized for resource state generation by the probabilistic entangled pair generation circuitry.

In this paper, we propose a source design that alleviates all of the abovelisted problems, yielding a *high-rate*, *high-fidelity*, *near-deterministic* source of dual-rail entangled photonic qubit pairs, at the cost of high levels of multiplexing. The concept is inspired from prior work on heralded [38] and multiplexed [31,39] SPDC-based single-photon sources: we first create what we call a *cascaded SPDC source*, which employs two SPDC-based entanglement sources, and performs a linear-optical Bell state measurement (BSM), commonly called an "entanglement swap," to yield an entangled state on the "outer" undetected mode pairs. This state has much lower vacuum and high-order-photon contributions compared to a standalone SPDC source. Thereafter, we leverage the heralding
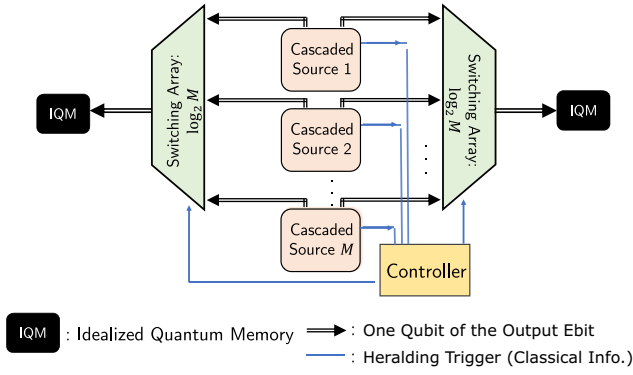
: IQM : Idealized Quantum Memory ⟹ : One Qubit of the Output Ebit
— : Heralding Trigger (Classical Info.)

FIG. 1. Multiplexing scheme for a near-deterministic source of near-unity-fidelity dual-rail Bell states. We multiplex M cascaded sources (each with one PNR-based BSM within it), whose outputs are connected into two M-to-1 switching arrays of $\log_2 M$ switches, each of $-\log_{10} \eta_s$ dB of loss. The switching array outputs the state of one of the successful cascaded sources in any time slot, assuming that one or more succeeds in that time slot. This output photonic entangled state is then loaded into a pair of ideal heralded quantum memories (shown as black boxes marked IQM).

trigger from the BSM to multiplex $M > 1$ such cascaded sources, using two switch arrays (each consisting of $\log_2 M$ switches) and a controller that lets out entangled photon pairs from the "successful" source, in order to improve the pair-production rate. See Fig. 1 for the overall concept. The switching losses, which scale up logarithmically in M, and photon-number-resolving (PNR) detector imperfections within the BSM, i.e., efficiency and dark counts, are incorporated into the engineering design study of the fidelity versus entanglement generation rate of the overall heralded-multiplexed source.

The article is organized as follows. Section II discusses the full quantum state description of a SPDC-based entanglement source, and our proposal for the cascaded source with a heralding trigger. Section III describes an idealized model of a heralded quantum memory that we use in the remainder of our analysis. Section IV presents a detailed performance evaluation of the ideal cascaded source. The modification of functionality with device impairments is examined in Sec. V. Finally, we present our analysis of the multiplexed cascaded source in Sec. VI, including switching losses in addition to the device imperfections considered in the previous section. We evaluate a parametric trade-off of fidelity versus success probability (of producing entangled pairs), while optimizing over the pump power of the individual SPDC sources and the number of cascaded sources in the multiplexed source. This trade-off shows that one can achieve a near-deterministic source of dual-rail Bell states in principle at high rates, despite reasonably nonideal devices. Section VII concludes the paper with thoughts on future work and applications of this study.

II. CASCADED SPDC SOURCE

A. Dual-rail entangled SPDC source: a review

Dual-rail encoded photonic qubits are defined by the presence of a single photon in one of two orthogonal modes [3]. In what follows, we use the language of polarization entangled states, but the arguments all apply equally well to any dual-rail encoding, utilizing any degree of freedom, i.e., polarization, time bin, spatial path, transverse spatial mode, frequency or mixed temporal spectral [40]. We denote the logical qubit states as $|\bar{0}\rangle$ and $|\bar{1}\rangle$; in the two-mode representation they can be expressed as $|0, 1\rangle$ and $|1, 0\rangle$, respectively. We use the following notation for two (of the four) mutually orthogonal dual-rail two-qubit Bell states:

$$|\Psi^{\pm}\rangle \equiv \frac{1}{\sqrt{2}} \left(|\bar{1}, \bar{0}\rangle + |\bar{0}, \bar{1}\rangle \right) = \frac{|1, 0; 0, 1\rangle \pm |0, 1; 1, 0\rangle}{\sqrt{2}}. \tag{1}$$

The "$\pm$" signs in Eq. (1), and Eq. (2) below, refer to the possibility of an additional π phase that could be applied to one of the polarization modes of one output pulse, e.g., using a half-wave plate, depending upon whether the desired Bell state for the application is $|\Psi^{+}\rangle$ or $|\Psi^{-}\rangle$.

A complete quantum-theoretic modeling of the polarization-dual-rail SPDC-based pulsed entanglement source was presented in Refs. [28,29]. The physical model of this entanglement source can be seen as two copies of two-mode squeezed vacuum (TMSV) states with one mode of each TMSV swapped. See Fig. 2 for a schematic representation. The output is described by four modes: two orthogonal polarization modes of each of the (spatiotemporal) modes of a pair of pulses emitted by the source. Hence, a two-qubit entangled Bell state requires four orthogonal modes to encode. The quantum state of this four-mode
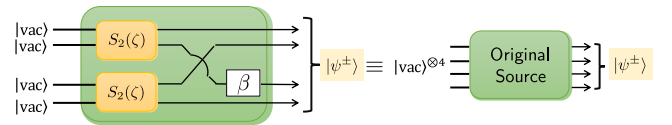

FIG. 2. A block model of the individual SPDC-based source of dual-rail entangled qubit pairs as presented in Ref. [28]. The output quantum state is approximated by Eq. (4). Here $S_2(\zeta)$ denotes a two-mode squeezing unitary, $|\text{vac}\rangle$ denotes a single-mode vacuum state, and the squeezing amplitude ζ is related to the mean photon number per mode, N_s, as $\sinh^2 |\zeta| = N_s$. One beam from each two-mode squeezed source is swapped and a potential optical phase shift $\beta (= 0$ or $\pi)$ is introduced. Appendix A includes a more detailed description and derivation of the quantum state description of this source.

output is given by

$$|\psi^{\pm}\rangle = \sum_{n=0}^{\infty} \sum_{k=0}^{n} (\pm 1)^k \sqrt{\frac{p(n)}{n+1}} |n-k, k; k, n-k\rangle, \quad (2)$$

where $p(n)$ is used to denote the probability of generating the n-photon term in each pair of mode and is given by

$$p(n) = (n+1) \frac{N_s{}^n}{(N_s + 1)^{n+2}} \quad (3)$$

with N_s the mean photon number per mode. Note that, hence, the mean photon number per (dual-rail) qubit is

$$|\psi^{\pm}\rangle = N_0 \left[\sqrt{p(0)} |0,0;0,0\rangle + \sqrt{\frac{p(1)}{2}} (|1,0;0,1\rangle \pm |0,1;1,0\rangle) + \sqrt{\frac{p(2)}{3}} (|2,0;0,2\rangle \pm |1,1;1,1\rangle + |0,2;2,0\rangle) \right]. \quad (4)$$

Here we have introduced $N_0 \equiv 1/\sqrt{p(0) + p(1) + p(2)} = (N_s + 1)^2/\sqrt{6N_s^2 + 4N_s + 1}$ as a normalization factor that we choose for convenience to ensure that $|\Psi^{\pm}\rangle$, despite the Fock truncation, is a unit-norm quantum state. In Appendix A, we show that, for $N_s \gtrsim 0.2$, the above truncated state leads to a bad approximation. Since all the results in this paper will use $N_s \ll 0.2$, this does not apply to the results reported herein. In Appendix G, we show how one would do a full exact analysis of everything reported in this paper, while employing the complete Gaussian-state description of $|\psi^{\pm}\rangle$.

The four-mode output state $|\psi^{\pm}\rangle$ in Eq. (2) is a superposition of vacuum, one of the two desired dual-rail-basis Bell states $|\Psi^{\pm}\rangle$, and additional states corresponding to $n \geq 2$ pairs of photons (in each of the two output pulses), with a geometrically distributed probability amplitude $p(n)$ as given in Eq. (3). Without the aid of auxiliary highly nonlinear operations such as a quantum memory or a nondemolition measurement we discuss later in this paper, these higher-order n-photon-pair terms cannot be eliminated from the source output, as the very nature of the underlying TMSV model determines the proportion of these "spurious" terms. Reducing the mean photon number per mode N_s by turning down the pump power does reduce the proportion of two-pair terms $p(2) = 3N_s^2/(N_s + 1)^4$, but, this happens at the expense of reducing $p(1)$ as well, and hence increasing $p(0)$; in the $N_s \ll 1$ regime, the vacuum term dominates. A quantum nondemolition (QND) measurement that performs a *vacuum or not* projection [41] on the source output could aid in eliminating the vacuum component. However, existing experimental proposals for implementing such a QND measurement involve nonlinear atom-photon interactions [42], and are currently infeasible to realize efficiently on traveling modes of an optical-frequency field.

$2N_s$. It should also be noted that this four-mode state is a Gaussian state, i.e., its Wigner function is an eight-variate Gaussian function of the field quadratures of the concerned modes. Hence, it is essentially a tensor product of two TMSV states with a pair of mode labels flipped.

Since we will be concerned with the $N_s \ll 1$ regime in this paper, we henceforth truncate the quantum state of the source up to the photon-number (Fock) support of two photon pairs [27]:

B. Cascaded SPDC source using two polarization-entangled sources and a BSM

We now describe our design of a *cascaded source* whose output state's fidelity with the desired Bell state can be higher than that of a standalone SPDC source's output $|\psi^{\pm}\rangle$ described in the previous subsection. The proposed design is shown in Fig. 3. We take two copies of $|\psi^{\pm}\rangle$ (labeled "original source" in a green box), and perform a local linear optical BSM using beam splitters and four PNR detectors. For the purposes of this section, we assume that all the components are ideal. This means that we consider no coupling losses from the SPDC sources into the BSM, perfect mode matching for the output photons, and
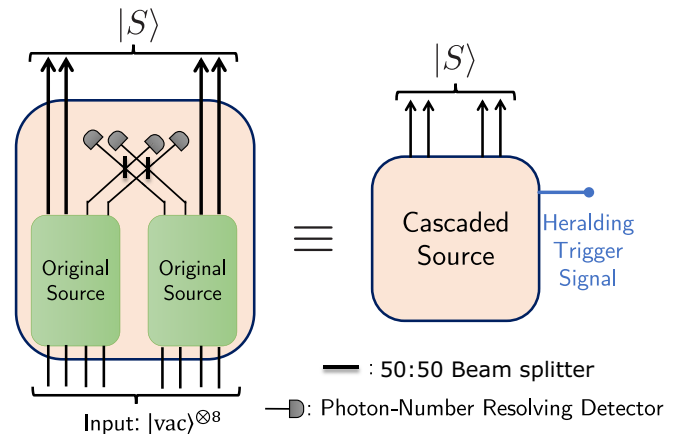

FIG. 3. Schematic of the proposed cascaded source that generates the entangled state given in Eq. (5) conditioned on the linear-optical BSM in the middle producing one of the two "desirable" click patterns. This cascaded source therefore has a heralding trigger, telling us in which time slot a copy of $|S\rangle$ is produced successfully; this feature is missing in the original source.

ideal PNR detectors in the circuit. We relax these assumptions in the more detailed analysis in Sec. V. If the states fed into the BSM were ideal Bell states, i.e., with no multipair contributions, the resulting state of the unmeasured four outer modes (shown as black arrows in Fig. 3) would be ideal Bell states as well. However, since the outputs (4) of the original sources $|\psi^\pm\rangle$ are not ideal Bell states $|\Psi^\pm\rangle$, despite observing a BSM "success," we might

generate spurious states on those outer modes that are not Bell states, i.e., arising from events where both pairs of photons originated in the same source.

If both the "original" sources produce the state $|\psi^+\rangle$, the heralded output state of the undetected outer modes, upon the occurrence of a *desirable* click pattern, is given by

$$|S\rangle = N_0'\left[\frac{p(1)}{2}(|1,0;0,1\rangle + (-1)^{m_1}|0,1;1,0\rangle) + (-1)^{m_2}\sqrt{\frac{p(0)p(2)}{3}}(|0,0;1,1\rangle + (-1)^{m_1}|1,1;0,0\rangle)\right], \quad (5)$$

where $p(0), p(1), p(2)$ are as defined in Eq. (3), and the normalization constant N_0' is given by

$$N_0' = \left(\frac{p(1)^2}{2} + \frac{2p(0)p(2)}{3}\right)^{-1/2}. \quad (6)$$

By *desirable*, we here signify the four click patterns (out of a possible eight) that are necessary but not sufficient to herald an entanglement swap between two dual-rail photonic modes on a linear optical BSM circuit. The reason the patterns are not sufficient is that these same patterns can also be produced by the undesirable event that both photons detected in the BSM came from only one of the SPDC sources, instead of one from each; unfortunately, the likelihood of these two processes are equal for SPDC. Therefore, to exclude the undesirable photon pair contribution from the same source, we must rely on postselection of a photon in each outer mode—either via direct detection or via a heralded quantum memory as discussed below.

Depending upon which of the four "desirable" click patterns occur on the four PNR detectors (e.g., 0011 implies no-click, no-click, 1-click, 1-click), the values of m_1 and m_2 in the heralded state in Eq. (5) are given as follows.

Click pattern	m_1	m_2
0011	0	0
1100	0	1
1001	1	1
0110	1	0

If both the "original" sources produce the state $|\psi^-\rangle$, the heralded output state of the undetected outer modes is the same as given in Eq. (5), except that the values of m_2 in the above table are flipped.

Henceforth, we drop the "$\pm$" superscript in the state $|S^\pm\rangle$, since we always assume to be working with the state $|S^+\rangle$. Furthermore, we say that the source was

"successful" in producing an entangled state when one of the first two desirable click patterns above (0011 or 1100) occur (i.e., $m_1 = 0$, though to be clear these click patterns can also arise when a single source produces both pairs). The reason for this is that we want the output state to be (close to) the $|S^+\rangle$ state. We use $|S\rangle$ to denote the desirable output state of the cascaded source, and not carry the m_2 index. This is because our results in this paper do not depend upon the value of m_2. Furthermore, if the memories (that load the distributed qubits) have faithful native quantum logic gates, it is easy to apply a local single-qubit unitary operation to turn the Bell state $|\Psi^-\rangle$ into $|\Psi^+\rangle$, and vice versa. So, if one wishes to be inclusive of the output state produced to be (close to) the $|\Psi^-\rangle$ state as well, our expression for the probability of success, in Eq. (10) for instance, can be multiplied by 2. Appendix C describes the hybrid modeling strategy used to derive the above results.

In what follows, we represent the cascaded source as the orange box shown in Fig. 3. The cascaded source has a *heralding trigger*, telling us in which time slot a copy of $|S\rangle$ is produced successfully, a feature missing in the original SPDC-based entanglement source.

III. IDEALIZED MODEL OF A HERALDED QUANTUM MEMORY

Quantum memories (QMs) are an essential component of entanglement distribution protocols; especially so for building quantum repeaters for long-distance entanglement distribution, and in distilling high-fidelity entanglement from low-fidelity entangled qubit pairs. Memories that can efficiently load one qubit of a photonic entangled state are necessary to store the quantum state for a time duration appropriate for the end application, or when it is ready to be interfaced to a larger quantum processor system, e.g., for performing teleported gates for distributed quantum computing.

Although various proposals for quantum memories exist in the literature, for the purposes of the performance evaluation of the source we propose in this paper, we want to

distill two important characteristics pertinent to our analysis: the QM can selectively load one dual-rail qubit (i.e., two orthogonal optical modes), and it has a heralding trigger. In other words, when the memory is successful in loading the qubit, it raises a (classical) binary-valued flag declaring a success or a *failure*.

We consider a rather idealized model for such a memory: one that performs a vacuum-or-not (VON) measurement, in a QND way. The QND measurement performed by this QM on the two incident modes can be expressed by the following positive-operator-valued measure (POVM) operators:

$$\hat{\Pi}_0 = |0,0\rangle\langle 0,0|, \quad \hat{\Pi}_1 = \hat{I}_2 - |0,0\rangle\langle 0,0| \quad (7)$$

with $\hat{I}_2 = \sum_{m,n} |n,m\rangle\langle n,m|$ the identity operator of the two-mode bosonic Hilbert space. If a two-mode optical quantum state $|\psi\rangle = \sum_{m,n} c_{m,n}|m,n\rangle$ is incident on this QM, with probability $p_{vac} = |c_{0,0}|^2$, the memory would raise a *failure* flag, and the postmeasurement state will be vacuum $|\psi_{vac}\rangle = |0,0\rangle$, i.e., nothing would be loaded into the quantum memory. However, with probability $p_{not\text{-}vac} = 1 - p_{vac}$, the memory would raise the *success* flag, and the postmeasurement state would be $|\psi_{not\text{-}vac}\rangle = N(|\psi\rangle - c_{0,0}|0,0\rangle)$, where $N = 1/\sqrt{1 - |c_{0,0}|^2}$ is a normalization constant. Note that if a two-photon state is incident on this memory, it *is* stored; however, in that case, for the incident state (5), the other ideal quantum memory would receive no photons, and this case would be rejected.

An experimental proposal for this VON measurement, with a photonic-domain postmeasurement state was conceived by Oi *et al.* in [42], using a reversible V-STIRAP atom-photon interaction. A QND-VON measurement, with the postmeasurement state being stored in a spin-based qubit, is implicit in a recently published experiment on measurement-device-independent (MDI) QKD to beat the repeaterless rate bound, using an asynchronous BSM based on a silicon-vacancy color center in a diamond nanophotonic chip [43]. A closely related platform for the processing and storage of polarization encoded dual-rail qubits has been proposed in Ref. [44].

The requirement we impose on a QM to have a heralding trigger is crucial to almost all quantum communication protocols. Practically, one way to achieve this is by using memories that entangle the incoming photonic state with the quantum state of the memory's internal qubit, for example, as in Refs. [43,44]. The heralding trigger consists of measuring the reflected photonic quantum state in the optical domain. The measurement outcome projects the quantum state of the qubit held by the QM into a local unitary equivalent of the photonic quantum state.

IV. PERFORMANCE EVALUATION OF THE IDEAL CASCADED SPDC SOURCE

In this section, we present a detailed analysis of the cascaded source when the associated hardware works ideally, comparing the performance of the cascaded source with that of the original noncascaded (SPDC-based entanglement) source.

A. Fidelity, assuming ideal devices

First, we compare the fidelity of the generated states with the target Bell state $|\Psi^+\rangle$. For the original source, this is given by

$$F(|\psi^+\rangle, |\Psi^+\rangle) = |\langle\psi^+|\Psi^+\rangle|^2 = \left|N_0\sqrt{p(1)}\right|^2. \quad (8)$$

Similarly, the fidelity of $|S\rangle$ with $|\Psi^+\rangle$ is given by

$$F(|S\rangle, |\Psi^+\rangle) = |\langle S|\Psi^+\rangle|^2 = \left|N_0'\frac{p(1)}{\sqrt{2}}\right|^2 = \frac{1}{2}, \quad (9)$$

where the factor of $1/2$ arises from the unwanted cases where both detected photons came from the same source.

Next, let us consider the fidelities with $|\Psi^+\rangle$ but when both mode pairs of the respective entangled states $|\psi\rangle$ (original source) or $|S\rangle$ (cascaded source) are loaded into a pair of idealized heralded quantum memories as described in Sec. III. It is simple to see by inspection of Eq. (5) that, after successful loading onto the ideal quantum memory, the cascaded source with ideal elements will yield a unit-fidelity Bell state loaded onto the two QMs. All four of these fidelities with $|\Psi^+\rangle$ (the original and the cascaded sources, with and without a QM) are plotted as a function of N_s in Fig. 4. We see that the cascaded source, assuming ideal elements for the BSM, has a superior fidelity
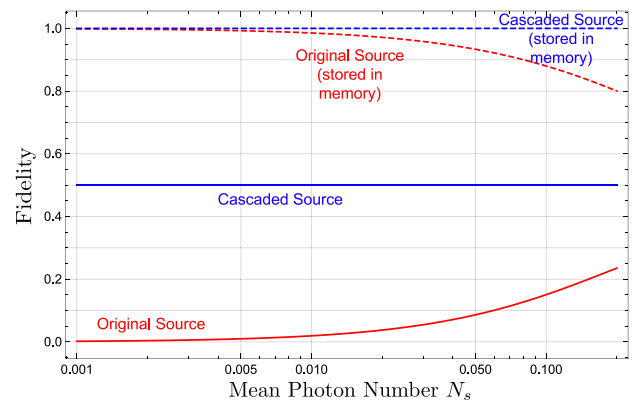

FIG. 4. Fidelity of the respective entangled states from the original SPDC source (red) and the cascaded source (blue) with the ideal Bell state $|\Psi^+\rangle$, plotted as a function of the mean photon number per mode (N_s). The corresponding state fidelities after loading each mode pair (qubit) into an idealized quantum memory are plotted using dashed lines.
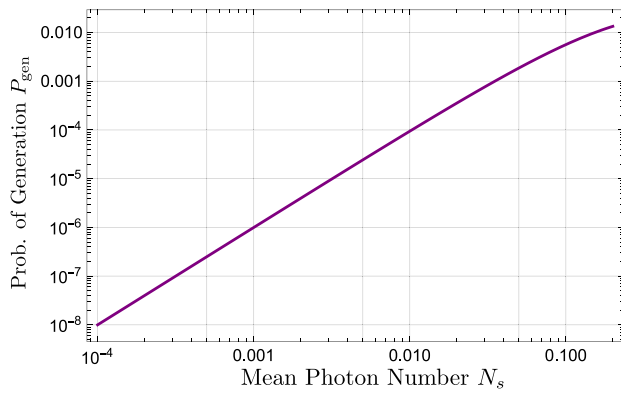
FIG. 5. Probability of generation of $|S\rangle$ as a function of the mean photon number per mode (N_s), assuming ideal devices.

compared with the original source. Note that the preceding analysis assumes that the state generated by each "original" source is the pure state given by Eq. (4). Realistic SPDC sources have an additional degree of freedom with respect to the temporal location of the emitted photons. This effect is commonly termed as a *timing walk-off*, which affects the quantum description of the emitted SPDC state, which in turn sets an upper bound on the fidelity (or, equivalently, manifests as a minimum infidelity) of the emitted state from the cascaded source, in comparison to the target state, $|\Psi^+\rangle$. We analyze this effect in detail in Appendix B.

B. Probability of entangled state generation, assuming ideal devices

Since the cascaded source is a heralded-state generation scheme, there is a generation probability, which we label P_{gen}, associated with it. This corresponds to the probability of the detection of a desirable click pattern in the BSM circuit, which is a function of N_s. Assuming ideal PNR detectors for the BSM and no other losses, this quantity is given by

$$P_{\text{gen}} = \frac{1}{2}\left(\frac{p(1)^2}{4} + \frac{p(0)p(2)}{3}\right) = \frac{N_s^2}{(N_s+1)^6}. \quad (10)$$

We plot P_{gen} as a function of N_s in Fig. 5. Here we note that the approximation made to simplify the state description is only valid up to $N_s \approx 0.2$, as plotted. Note that the first term in Eq. (10) describes the desired case where each source contributed one photon to the BSM, while the second term describes the undesirable case where one source produced two pairs and the other produced none.

V. PERFORMANCE EVALUATION OF THE CASCADED SOURCE WITH DEVICE NONIDEALITIES

In this section, we include the effects of device nonidealities into the analysis of the entangled state produced by
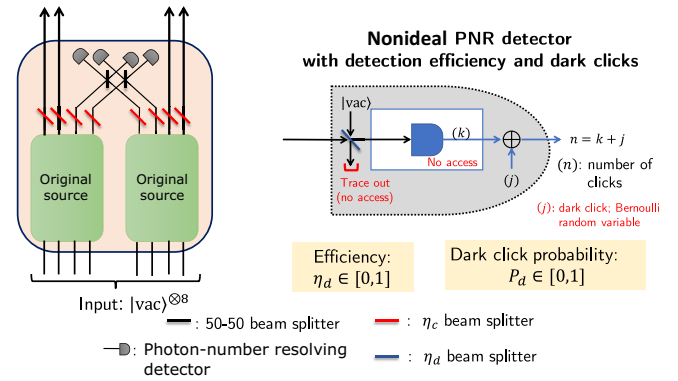

FIG. 6. Detailed schematic of the proposed cascaded source with all nonidealities marked. This includes the coupling efficiency η_c (red beam splitter), the nonideal detector efficiency η_d (blue beam splitter), and the dark click probability P_d (Bernoulli random variable j).

the cascaded source. The specific nonidealities we consider in this section are as follows.

1. *Detector efficiency and dark clicks.*—The leading candidates for PNR detection are the transition edge sensor (TES) detectors [17,45] and superconducting nanowire single-photon detectors (SNSPDs) [46,47]. Each kind is influenced by multiple effects that degrade their performance [45,48]. In our present analysis, we abstract off the nonideality of a PNR detector into two parameters: a subunity detector efficiency ($\eta_d \leq 1$) and a nonzero dark click probability ($P_d \geq 0$) per detection gate (which will be assumed to be the length of a pulse slot for our calculations). A pictorial schematic of our model of this nonideal PNR detector is given in Fig. 6. A detailed mathematical model of this two-parameter nonideal PNR detector is discussed in Appendix D.

2. *Coupling efficiency.*—We also account for losses in coupling the "inner" output modes of the SPDC sources (those that go into the BSM) into single-mode fiber. We label the effective efficiency of this coupling, per mode, as η_c. Our analysis of the derived formulae shows (see Appendices C and F) that we can combine the two efficiency parameters into one efficiency parameter, $\eta \equiv \eta_c \eta_d$, i.e., the output state will be identical for any given value of η, regardless of the actual values of η_c and η_d, as long as their product equals η.

3. *Photon indistinguishability.*—The required BSM that heralds the generation of the final state relies on two-photon or Hong-Ou-Mandel (HOM) interference to properly function. The interfering photons from each source must therefore be completely indistinguishable to achieve maximum fidelity. Any mismatches in the sources' spectral, temporal, or spatial modes will introduce distinguishability and reduce visibility for the interference measurement. This reduced visibility in turn leads to errors in the BSM and degrades the final output state fidelity. We

present a simple model to study this behavior. If the "inner" output modes of the SPDC sources have HOM visibility V (see Appendix B for a more detailed explanation), then whenever the two photons arrive at the central beam splitter, this corresponds to a probability of V that the photons interfere perfectly, along with a probability $1 - V$ that no interference occurs. Assuming that the "perfect visibility" state has fidelity F_0 with our target, then the final state fidelity with visibility V is

$$F(V) = F_0 V + 1/2 \times (1 - V) = 1/2 + (F_0 - 1/2)V. \tag{11}$$

The scaling with visibility is a linear effect, and can be mitigated by improved modal matching in the constituent SPDC sources that make up the cascaded source. In practice, care must be taken that the improved mode overlap is not achieved at the expense of reduced efficiency.

Figure 6 depicts the complete model with all nonidealities accounted for. When $P_d > 0$, the output of the cascaded source is a mixed state—a statistical mixture of pure states corresponding to various *true* click patterns in the detectors (several of which may not be one of the "desirable" patterns). These terms arise because, with some probability, noisy detectors could signal the presence of a photon (when in fact there is not one), leading the BSM to conclude the occurrence of a desirable pattern, and hence falsely declare a success. The mixed state, derived in full detail as per the techniques in Appendix C, contains the ideal-device pure state as in Eq. (5) along with other pure states that are generated when the apparent-desirable click pattern actually includes one or more dark counts.

Describing the effects of detector efficiency η_d is trickier. In general, the effect of $\eta_d < 1$ is equivalent to the detected modes being transmitted through a pure-loss bosonic channel of transmissivity η_d prior to being incident on a unity-efficiency detector. Therefore, the predetection state is a mixed state, and the final density matrix for the present analysis is not compactly expressible. Appendix C provides the detailed procedure for the calculations of this mixed state and Appendix F includes analytic formulae for the source performance metrics. Below, we present plots that show the trend of the metrics under consideration, e.g., fidelity and probability of success, as a function of N_s for various values of η_d and P_d. Note that, experimentally, TES detectors and SNSPDs have now demonstrated efficiencies above 98% [45,49]. Finally, the effect of η_c does not need to be discussed separately, since it can be subsumed into η_d, as discussed above.

A. Analysis of the raw photonic entangled state produced by the cascaded source

In Fig. 7, we plot the fidelity of the entangled state generated by the cascaded source with the ideal Bell state
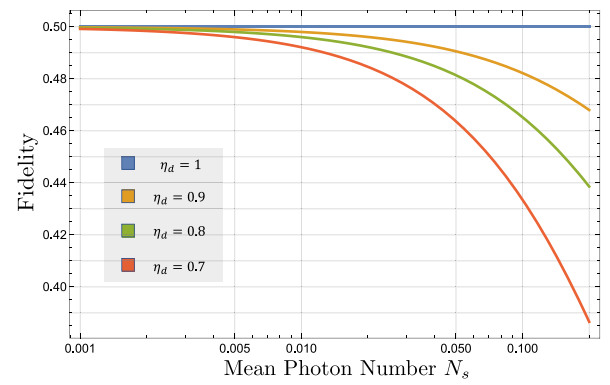

FIG. 7. Fidelity of the entangled state generated by the cascaded source with the ideal Bell state $|\Psi^+\rangle$ at various values of detector efficiency η_d, plotted as a function of N_s. We assume that $P_d = 0$.

$|\Psi^+\rangle$ as a function of N_s, at various values of detector efficiency $\eta_d \leq 1$, for $P_d = 0$. We note that the fidelity decreases monotonically from 0.5 (the maximum fidelity attained by the cascaded source, as shown in Sec. IV, Fig. 4) with *increasing* N_s for subunity efficiency, as expected.

In Fig. 8, we see that, as the dark click probability P_d increases above zero, the fidelity monotonically decreases with *decreasing* N_s dropping to zero as $N_S \to 0$. This is expected since, for very low N_s, dark clicks account for most of the purported BSM "success" events.

In Fig. 9, we combine the effect of nonzero dark clicks ($P_d > 0$) and subunity detection efficiency ($\eta_d < 1$) in the PNR detectors used for the BSM. The fidelity plots are exactly as expected—the two aforementioned forms of detector impairment pull the fidelity down from the maximum possible value of 0.5 at low N_s and high N_s, respectively. The fact that the different line styles (solid, dashed, dotted) overlap for a significant range of N_s
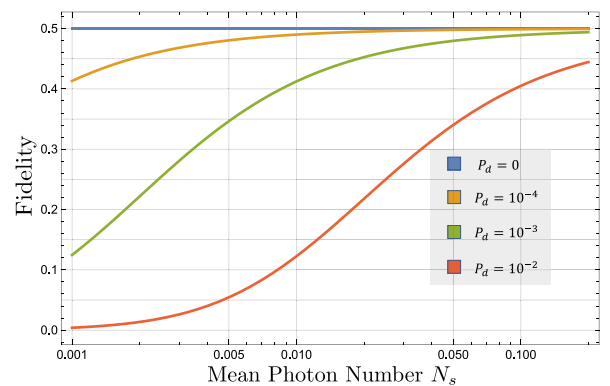

FIG. 8. Fidelity of the states from the cascaded source with the $|\Psi^+\rangle$ Bell state for various values of detector dark click probability (P_d). We assume that $\eta_d = 1$.
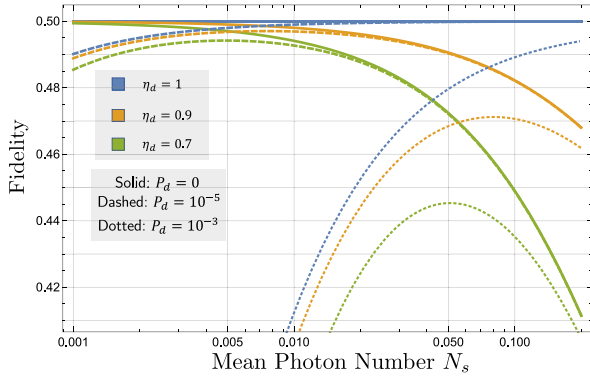
FIG. 9. Fidelity of the states from the cascaded source with the $|\Psi^+\rangle$ Bell state for various values of detector dark click probability (P_d) and detector efficiency (η_d). The colors signify different η_d values and the line styles signify different P_d values. The different plots for a fixed value of η_d overlap for a significant range of N_s.

indicates the insensitivity of the source to excess noise in the detectors.

B. Analysis of the entangled state produced by the cascaded source, after successfully loading the qubits into a pair of ideal heralded quantum memories

In Fig. 10, we plot the fidelity of the entangled state with the $|\Psi^+\rangle$ Bell state, *after* successfully loading the qubits into a pair of idealized heralded quantum memories. We remind the reader here of the conclusion in Sec. IV, Fig. 4—after successfully loading the entangled qubit pairs into idealized QMs, the fidelity of the entangled state with the $|\Psi^+\rangle$ Bell state, with an ideal BSM, is 1 for the cascaded source, regardless of the value of N_s.
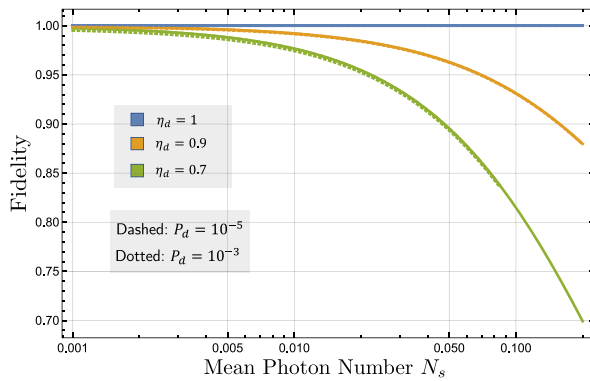

FIG. 10. Fidelity of the entangled states from the cascaded source with the $|\Psi^+\rangle$ Bell state, after successfully loading both qubits into idealized heralded quantum memories. The colors signify different η_d values and the line styles signify different P_d values. Plots with fixed η_d majorly overlap over all N_s for both values of P_d. This indicates minimal degradation in fidelity in the presence of dark clicks, when we load the final state into an IQM.
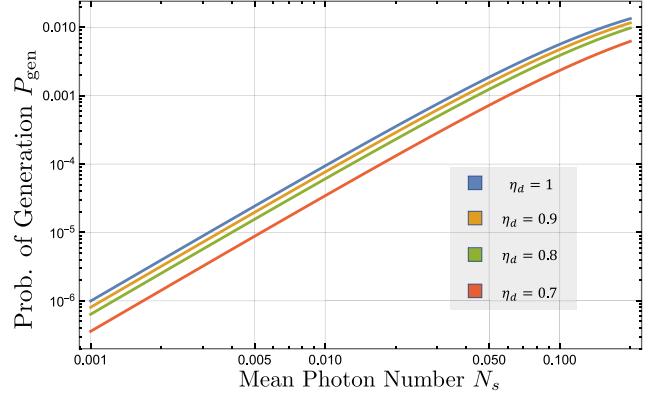

FIG. 11. Probability of generation of $|S\rangle$ as a function of the mean photon number per mode (N_s) for various values of detector efficiency (η_d). We assume that $P_d = 0$.

The plots in Fig. 10 show that the low-N_s reduction of the fidelity (due to nonzero P_d), as seen in Fig. 8, is almost completely suppressed by the quantum memories. Here we have limited the discussion to the cascaded source, but the same arguments apply to the original source: the QND detection of the "outer" mode photons, implied by the ideal heralding quantum memories, lifts the state fidelity at low N_s to 1 (see Fig. 4), even in the presence of moderate BSM detector dark counts.

The aforementioned point is important, and is key to us being able to construct a near-deterministic, near-unity-fidelity source of entanglement. We do so by multiplexing several cascaded sources, as shown in the next section.

C. Analysis of the probabilities of success of generating the raw photonic entangled state by a cascaded source

In Fig. 11, we plot P_{gen}, the probability of generation of $|S\rangle$ by the cascaded source, as a function of N_s for a few different values of η_d keeping $P_d = 0$. As expected, the entire
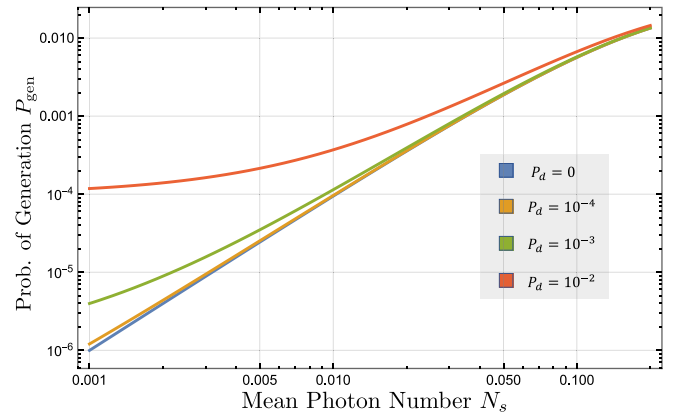

FIG. 12. Probability of generation of $|S\rangle$ as a function of the mean photon number per mode (N_s) for various values of detector dark click probability (P_d). We assume that $\eta_d = 1$.
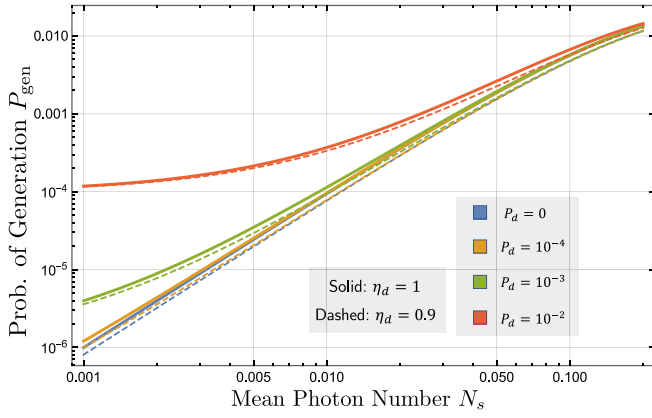
FIG. 13. Probability of generation of $|S\rangle$ as a function of the mean photon number (N_s) for various values of detector dark click probability (P_d) and detector efficiency (η_d). The colors signify different P_d values and the line styles signify different η_d values.

P_{gen} versus N_s plot shifts downwards with decreasing η_d as compared to the ideal scenario.

In Fig. 12, we plot P_{gen} versus N_s with $\eta_d = 1$ held fixed, but for a few nonzero values of P_d. Here, we see that, as P_d increases, the probability of success increases, but most of those purported BSM "desirable" patterns occur due to dark clicks, and as we already know from the low-N_s regime of the fidelity plots in Fig. 8, those spurious success events give very low fidelity, i.e., close-to-useless output states. Interestingly, however, for higher N_s, the effect of nonzero dark clicks is almost completely washed away by the PNR-detection-based BSM.

In Fig. 13, we plot P_{gen} versus N_s including the effects of both detector impairments: $P_d > 0$ and $\eta_d < 1$. The results are self-explanatory, and exactly as expected by combining the qualitative effects from Fig. 11 ($\eta_d < 1$, $P_d = 0$) and Fig. 12 ($\eta_d = 1$, $P_d > 0$), respectively.

VI. MULTIPLEXED CASCADED SOURCE OF ON-DEMAND HIGH-FIDELITY BELL STATES

The greatest advantage afforded by the heralding trigger is that it enables multiplexing (heralded) cascaded sources using an array of photonic switches. The design releases an entangled state based on which the constituent sources was successful in a given time slot. Such multiplexing has been shown to enable large enhancements in the success probability of HSPS [31]. This construction, using a bank of $M > 1$ cascaded sources, with both output mode pairs of each of the M sources fed into M-to-1 optical switch arrays (each built out of $\log_2 M$ switches), assisted by electronic controllers, is shown in Fig. 1. The switching arrays output the state of one of the successful cascaded sources in any time slot, assuming that one or more succeeds in

that time slot. If none succeed, the multiplexed source produces nothing. But the user of the source knows when such a *failure* event happens.

If there were no additional losses in switching, we could generate the entangled state in Eq. (5) with as high a success probability as we please, by increasing M indefinitely. The probability that an ideal multiplexed cascaded source generates an entangled pair is $P_{\text{success}} = 1 - (1 - P_{\text{gen}})^M$. To make the source near on-demand, we would pick $M \approx 1/P_{\text{gen}}$, which would ensure that on average at least one of the cascaded sources in the bank would have their internal BSM declare a success. Unfortunately, this simple-minded seemingly indefinite increase of P_{success} toward 1 by increasing M does not work when device nonidealities, especially the switching losses, are accounted for. Our modeling and analysis considers four device impairments: detection efficiency (η_d), coupling efficiency (η_c), switching efficiency (η_s), and dark-click probability (P_d). There are two design choices for the implementation of the heralded-multiplexed source: M (number of cascaded sources) and N_s (determined by pump power). The performance of the source is quantified by the trade-off of fidelity versus probability of success (i.e., rate of entangled pair production).

A. Performance evaluation of the heralded-multiplexed source

We consider a multiplexing scheme as described in the paragraph above (shown in Fig. 1), where multiple cascaded sources make parallel attempts to generate the target photonic state. This output photonic entangled state is then loaded into a pair of ideal heralded quantum memories (shown as black boxes marked IQM). The device metrics in our model are quantified by: (1) the coupling efficiency from the outputs of the cascaded source (η_c), (2) efficiencies of all the PNR detectors within the BSM (η_d), (3) the dark click probability (per qubit slot) of each detector in the BSM (P_d), and (4) switching losses per switch in the switch array, expressed as an effective transmissivity (η_s) (hence, the overall effective transmissivity being $\eta_s^{\log_2 M}$). The performance of the heralded-cascaded source is quantified by the probability of success P_{success} of the multiplexed source producing an entangled pair of dual-rail qubits, and the fidelity of that state produced with respect to the $|\Psi^+\rangle$ Bell state.

In Fig. 14, we plot the trade-off between the fidelity and P_{success} for chosen values of $\eta = \eta_c \eta_d$ and η_s in the form of a scatter plot, as M and N_s are both varied. Each point in this scatter plot represents a unique choice of M and N_s, while the solid lines indicate the best trade-off the heralded-multiplexed source can achieve, given the device metrics η_c, η_d, and η_s, for $N_s \in [10^{-4}, 0.2]$ and $M \in [1, 10^6]$. Furthermore, we plot the performance trade-off for the standalone SPDC source (black dotted line)
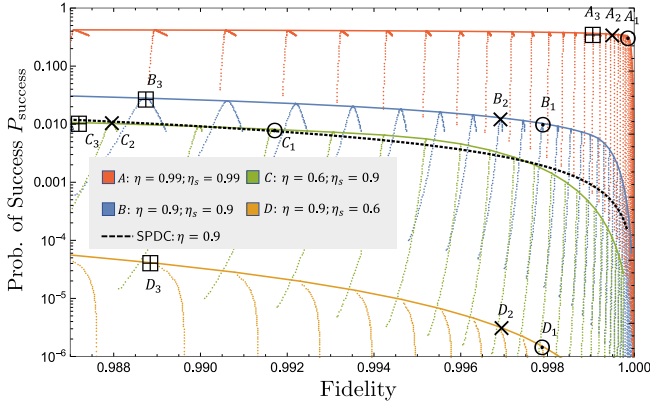
FIG. 14. Scatter plot of P_{success}, the probability of successfully generating an entangled state by the heralded-multiplexed source, and its fidelity (with the ideal Bell state) for $M \in [1, 10^6]$ and $N_s \in [10^{-4}, 0.2]$, with $\eta = \eta_c \eta_d$ and η_s held fixed. The solid lines are envelopes corresponding to the highest probability of success that can be achieved at a given fidelity target. We see that, for a given fidelity target, there is an optimal value of M that maximizes P_{success}. The N_s, M values corresponding to the marked points are summarized in Table I. For comparison, the success probability of using an "original" source (dotted black curve) is shown, assuming that its photons are also directed into the same sort of IQM (with a coupling efficiency $\eta = 0.9$).

when loaded onto a quantum memory (with a coupling efficiency $\eta = 0.9$) that we have assumed for the multiplexed cascaded approach. We note that the performance is actually quite comparable to the much more resource-intensive approach discussed here.

Table I summarizes some sample M (columns) and N_s (bottom row) values for the marked points in Fig. 14. The points give us a sense of the required multiplexing size and mean photon number combinations that yield the output state of desired fidelity with a maximal probability of generation. In Figs. 15 and 16, we plot the envelope of this aforementioned trade-off for fixed values of M (while N_s is varied). We note that, for a given set of device metrics,

TABLE I. Table of M and N_s values for the marked points with coupling η and switching efficiencies η_s indicated in Fig. 14. Note that the D points in Fig. 14 correspond to $\eta_s < 0.707$, and hence multiplexing ($M > 1$) does not improve P_{success}. This is why the optimum $M = 1$ for this row.

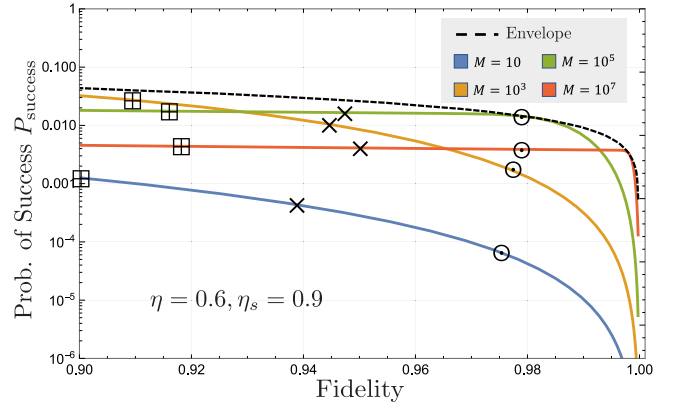| Pt. label	η	η_s	Subscript		
			1	2	3
A	0.99	0.99	4.6×10^5	2.1×10^5	1.7×10^4
B	0.9	0.9	2.1×10^5	1.1×10^5	8×10^3
C	0.6	0.9	5.4×10^5	2.4×10^5	2.0×10^4
D	0.9	0.6	1	1	1
N_s value			0.003	0.005	0.018

FIG. 15. Envelopes of P_{success}, the probability of successfully generating an entangled state by the heralded-multiplexed source, and its fidelity (with the ideal Bell state) for varying values of M and $N_s \in [10^{-4}, 0.2]$, with $\eta = 0.6$ and $\eta_s = 0.9$. For low switching loss (i.e., higher η_s), in the high-fidelity regime, the envelopes tend to go higher as M increases. But the trend is opposite in the low-fidelity regime. The envelope marked by the black dashed line is identical to the green solid plot in Fig. 14, i.e., that corresponding to $\eta = 0.6$ and $\eta_s = 0.9$. The marked points correspond to the following N_s requirements: $N_s = 0.01$ ($\odot$); $N_s = 0.019$ ($\times$); $N_s = 0.04$ ($\boxplus$). We assume that $P_d = 0$.

there is an optimal value of M beyond which P_{success} cannot be increased any further. Moreover, there are various limitations on the performance of the cascaded source, if the original SPDC sources are imperfect, as discussed earlier in Sec. IV.
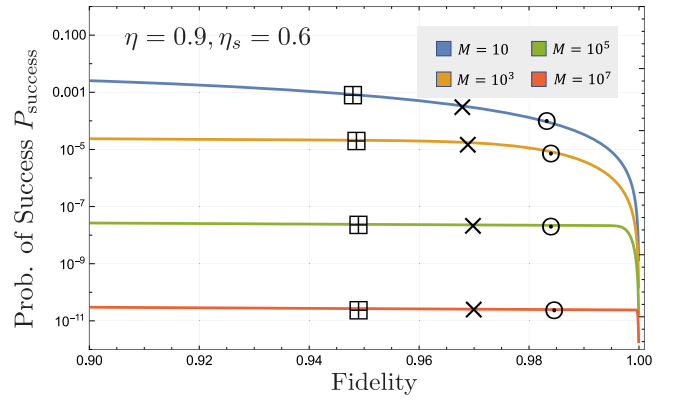

FIG. 16. Envelopes of P_{success}, the probability of successfully generating an entangled state by the heralded-multiplexed source, and its fidelity (with the ideal Bell state) for varying values of M and $N_s \in [10^{-4}, 0.2]$, with $\eta = 0.9$ and $\eta_s = 0.6$. For this high switching loss, the envelopes monotonically go lower as M increases. The marked points correspond to the following N_s requirements: $N_s = 0.028$ ($\odot$); $N_s = 0.05$ ($\times$); $N_s = 0.09$ ($\boxplus$). We assume that $P_d = 0$.

B. Effect of switching loss on system performance

We note an interesting reversal in behavior when the switching loss per switch (quantified by η_s) increases beyond a threshold value of 1.5 dB (which corresponds to $\eta_s = 1/\sqrt{2} \approx 0.707$). When the loss per switch is below this threshold, the envelope is seen to attain its maximal value for an optimum choice of M (see Fig. 15). However, when the loss per switch is higher than 29.3%, the trend reverses and increasing M is detrimental to the performance of the scheme (see Fig. 16). This places a hard limit on the per-switch loss and the number of cascaded sources in a viable and useful implementation of the cascaded-multiplexed source.

The intuitive reason for this reversal in the trend is as follows. The size of the switching array scales as $\log_2(M)$. Assuming switching efficiency per switch of η_s [i.e., $\log_{10}(1/\eta_s)$ dB of switching loss per switch], the output modes from a successful cascaded source undergo an additional loss corresponding to an effective transmission of $\eta_s^{\log_2(M)}$. Now, unlike the case with lossless switches, increasing the number of cascaded sources M still increases the probability of success as $1 - (1 - P_{\text{gen}})^M$, while simultaneously decreasing the probability that a successful output from one of the cascaded sources would be successfully loaded into the memory (because $\eta_s^{\log_2(M)}$ decreases as M increases). Given M and N_s, the probability of success for a multiplexed source to successfully generate an entangled state is given by

$$P_{\text{success}} = [1 - (1 - P_{\text{gen}})^M](1 - p_{|00\rangle})^2, \quad (12)$$

where P_{gen} is the success probability of an individual cascaded source and $p_{|00\rangle}$ is the probability the idealized
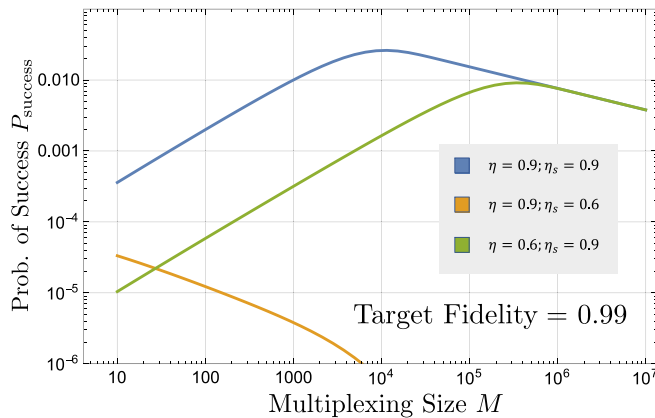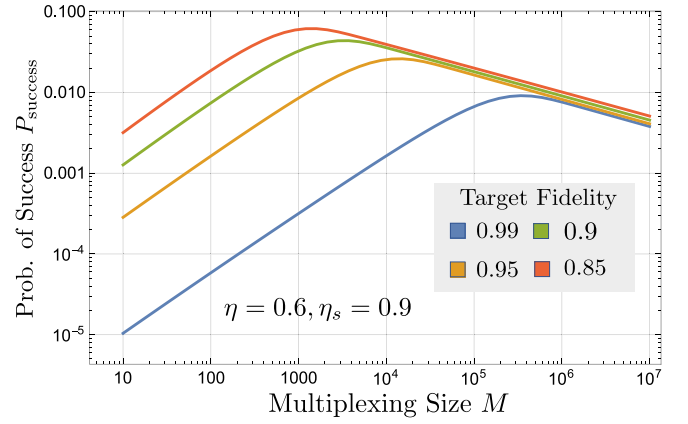

FIG. 18.   The probability P_{success} as a function of the multiplexing size M for different target fidelities. The pump power is optimized to achieve the given fidelity target with $\eta = 0.9$ and $\eta_s = 0.6$. The value of M required to achieve the highest value of P_{success} increases as the target fidelity is increased.

quantum memory (IQM) on one side of the heralded-multiplexed source shown in Fig. 1 *fails* to load the photonic qubit into the memory. This $p_{|00\rangle}$ term increases as M increases due to compounding switching losses. The M-dependent portion of this second term $(1 - p_{|00\rangle})^2$ is a multiplicative term: $[\eta_s^{\log_2(M)}]^2$. When P_{gen} is small, the first term in Eq. (12), $1 - (1 - P_{\text{gen}})^M \approx MP_{\text{gen}}$. It is simple to see that at $\eta_s = 1/\sqrt{2} \approx 0.707$, $[\eta_s^{\log_2(M)}]^2 = 1/M$, which in turn means that P_{success} becomes equal to $MP_{\text{gen}} \times 1/M$, hence becoming insensitive to M. Therefore, when $\eta_s > 1/\sqrt{2}$, P_{success} increases as M increases, while for $\eta_s \leq 1/\sqrt{2}$, P_{success} decreases as M increases.

FIG. 17.   The probability P_{success} as a function of the multiplexing size M for a target fidelity of 0.99. The pump power is optimized to achieve the given fidelity target at the specified values of η and η_s. For high switching loss (i.e., lower η_s; see the orange plot), we see P_{success} decreasing (as opposed to increasing) as M increases.
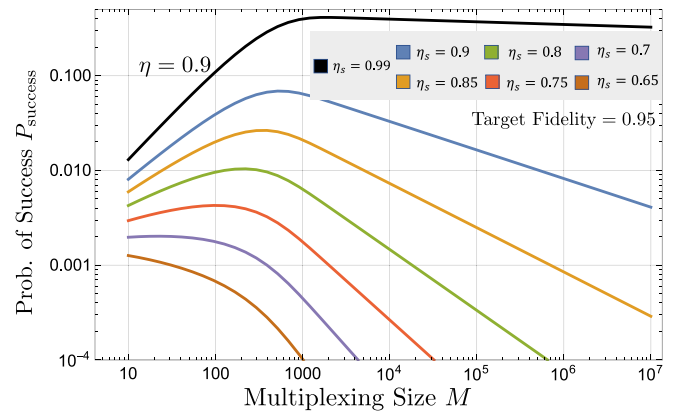

FIG. 19.   The probability P_{success} as a function of the multiplexing size M for values of $\eta_s \in [0.65, 0.9]$ at a fixed $\eta = 0.9$. The pump power is optimized to achieve the target fidelity = 0.95. We see that the value of M required to achieve the maximum value of P_{success} increases as η_s decreases, until at $\eta_s \approx 0.7$ (purple line) when we see a complete turnaround and multiplexing is detrimental for the source's P_{success}.
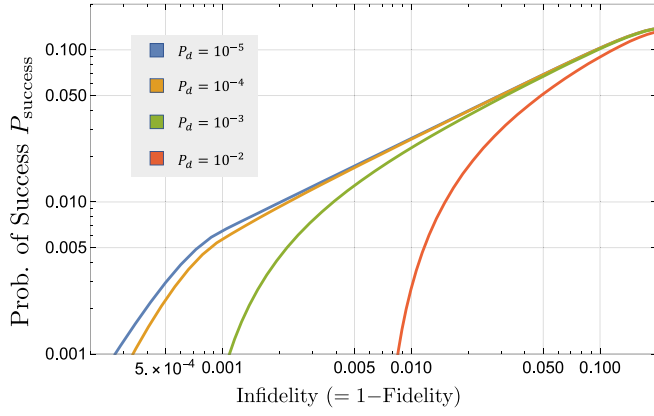
FIG. 20. The probability P_{success} and infidelity trade-off at various values of M and N_s. Here we assume that $\eta = 0.9$ and $\eta_s = 0.9$. The solid lines correspond to the probability of success for a given infidelity target. Note that, with higher dark count probability, the achievable fidelity becomes more restricted.

We plot P_{success} as a function of M for a given fidelity target ($= 0.99$) in Fig. 17. We see that, for one of the plots, for which a lower η_s is chosen, P_{success} decreases as M increases, as discussed above. Furthermore, we examine how P_{success} behaves in the regime of high η_s (for various values of M) in Fig. 18. We observe that, for every (η, η_s) combination, where $\eta_s > 1/\sqrt{2}$, P_{success} is maximized for an optimal value of M. This optimal value of M increases as we increase the target fidelity. In the plots in Fig. 19, we numerically extract the per-switch loss (value of η_s) where the P_{success} versus M trend reverses for a given value of target fidelity and η. We find that this turnaround happens at around $\eta_s \approx 0.7$ (corresponding to 1.5 dB of loss
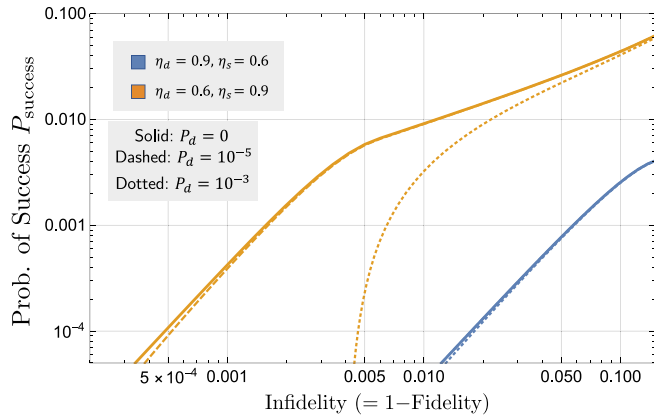

FIG. 21. The probability P_{success} and infidelity trade-off at various values of M and N_s. Here $\eta = 0.9$, $\eta_s = 0.6$ (blue) and $\eta = 0.6$, $\eta_s = 0.9$ (orange). We vary P_d between 0 (solid), 10^{-5} (dashed), and 10^{-3} (dotted). Note that, with higher dark count probability, the achievable fidelity becomes more restricted. The solid and dashed lines are largely indiscernible because they mostly overlap with one another.
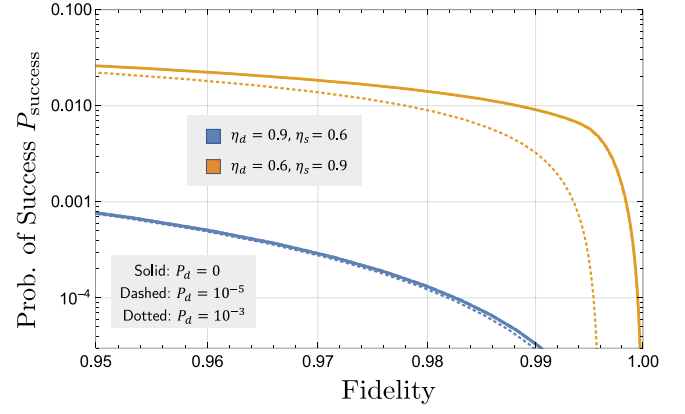

FIG. 22. The outer envelopes of the scatter plot of P_{success} versus fidelity when M and N_s are varied: $M \in [1, 10^6]$ and $N_s \in [10^{-4}, 1]$ with $\eta = \eta_c \eta_d$ and η_s held fixed. The envelopes correspond to the highest probability of success that can be achieved at a given fidelity target for an optimal value of M. Note that, with higher dark count probability, the achievable fidelity becomes more restricted. The solid, dashed, and dotted lines are largely indiscernible because they mostly overlap with one another.

per switch), as expected, and is not affected by the other losses in the system (i.e., $\eta = \eta_c \eta_d$) or the fidelity target we impose on the cascaded-multiplexed source.

C. The effect of detector dark clicks

The above analysis does not account for nonzero P_d. We observe that the inclusion of detector dark clicks ($P_d > 0$) restricts the maximum achievable fidelity and decreases the probability of success. To illustrate this, we plot P_{success} versus infidelity ($1-$ fidelity), achieved by the heralded-multiplexed source, in Fig. 20. In these plots, we assume that $\eta = 0.9$ and $\eta_s = 0.9$. In Fig. 21, we show the trade-off of P_{success} versus infidelity for two sets of values of losses. In these plots, we assume that $M = 10^6$, $\eta = 0.9$, $\eta_s = 0.6$ (blue lines) and $\eta = 0.6$, $\eta_s = 0.9$ (orange lines), with P_d varying between 0 (solid), 10^{-5} (dashed), and 10^{-3} (dotted). Finally, in Fig. 22, we plot the P_{success} versus fidelity trade-offs as in Fig. 14, but with $P_d > 0$. The main difference, as expected from the plots in Figs. 20 and 21, is that a nonzero P_d sets an upper limit on the state fidelity. However, for $P_d < 10^{-5}$, the reduction in the fidelity upper bound below unity is negligible. This level of dark click probability is easily available with state-of-the-art superconducting nanowire single-photon detectors [46] for a detection gate corresponding to gigahertz-scale repetition rates, which are readily achieved with SPDC-based entanglement sources.

VII. DISCUSSION AND FUTURE WORK

The primary pieces of intuition that drove the main results of this paper are that (a) cascading two SPDC-based polarization-entangled sources with a linear-optical BSM
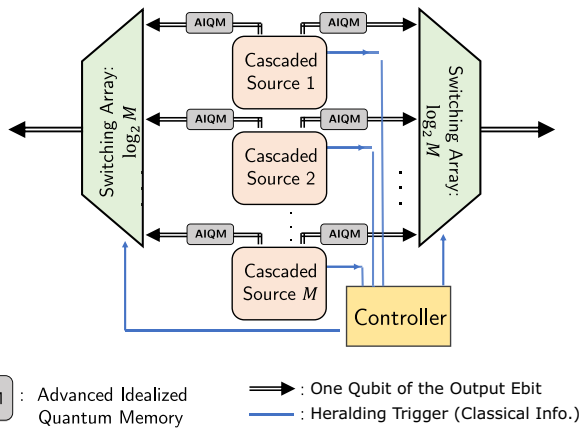
FIG. 23. Alternate multiplexing scheme for a near-deterministic source of the target Bell states. Similar to the scheme in Fig. 1 we multiplex M sources with the VON projective but *nondestructive* measurements (marked as the advanced ideal quantum memory) at each output. This memory is modeled as a black box that can perform the projective measurement on a photonic qubit while preserving it in its photonic modes. The heralding trigger additionally accounts for the vacuum output from the VON black box.

built using PNR detectors in the middle produces an entangled state whose fidelity can be pushed close to unity if there were a heralded quantum memory available that can filter out the vacuum contribution. Additionally, (b) the BSM provides a heralding trigger (again, not available in a free-running standalone SPDC source, unless one post-selects the photons) that lets us multiplex many cascaded sources with a photonic switch array.

One limitation of the cascaded source, indicated in Fig. 4, is that the maximum fidelity it can attain, for the raw photonic-domain entangled state it emits, is 0.5. Obviously, the fidelity of the standalone SPDC source is even worse; however, we should note (cf. Fig. 14) that if we allow the standalone source access to the same ideal quantum memory, its performance becomes comparable, with many fewer required resources. So, neither of these photonic entangled sources is of use to produce high-fidelity entanglement unless a heralded quantum memory were available that can faithfully filter out the vacuum contribution, or it were used in an application where such vacuum filtering would occur naturally in a postselected fashion as a result of photon detection, e.g., in QKD. It might be possible to further improve the quality of the output entangled state produced by the heralded-multiplexed source if we had an advanced version of the IQM, wherein along with the stated characteristics of the IQM in Sec. III, the IQM is additionally able to emit the stored qubit into the photonic domain, encoded in the dual-rail basis. This alternative design is depicted in Fig. 23. This advanced memory would likely come with an additional efficiency

cost (due to inefficiency in that storage qubit-to-photon readout process).

The multiplexed source we analyze in this paper may find application in satellite-based entanglement distribution, quantum repeaters, resource-efficient generation of more complex multiphoton entangled states for fault-tolerant quantum computing, and quantum sensors, among others. We leave the performance analysis of this source for specific applications open for future research.

APPENDIX A: DETAILED SOURCE ANALYSIS

The original source proposed in Refs. [28,29] uses spontaneous parametric down-conversion (SPDC) and additional linear optical elements to generate the state given in Eq. (4). Analysis of the complete interaction picture of the Hamiltonian that governs the dynamics of the quantum source (i.e., the weak parametric down-conversion) is given in Section IA of the original work and also in Ref. [27]. We present a high-level analysis of the same source using our choice of notation. The source (as depicted in Fig. 1 of Ref. [28]) can be "unfolded" as shown in Fig. 24.

After the generation of the two TMSV states, which are given by

$$|\psi_1\rangle = \sum_{n_1=0}^{\infty} \sqrt{\frac{N_s^{n_1}}{(N_s+1)^{n_1+1}}} \, |n_1^H\rangle \, |n_1^V\rangle, \quad \text{(A1a)}$$

$$|\psi_2\rangle = \sum_{n_2=0}^{\infty} \sqrt{\frac{N_s^{n_2}}{(N_s+1)^{n_2+1}}} \, |n_2^H\rangle \, |n_2^V\rangle, \quad \text{(A1b)}$$

the linear optical circuitry swaps the similarly polarized beams. The unitary U induces the swap of the photonic states in the modes labeled by $|n_1^V\rangle$ and $|n_2^V\rangle$ terms, which

can be compactly described by the transformation

$$|\psi_1\rangle \otimes |\psi_2\rangle = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sqrt{\frac{N_s^{n_1+n_2}}{(N_s+1)^{n_1+n_2+2}}} |n_1^H, n_1^V; n_2^V, n_2^H\rangle$$

$$\xrightarrow{\text{swap by } U} \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sqrt{\frac{N_s^{n_1+n_2}}{(N_s+1)^{n_1+n_2+2}}} (-1)^{n_2}$$

$$\times |n_1^H, n_2^V; n_1^V, n_2^H\rangle$$

$$= |\psi_{src.}\rangle. \tag{A2}$$

We define $g(m) \equiv \sqrt{N_s^m/(N_s+1)^{m+1}}$ and, for the sake of brevity, we use

$$\sqrt{p(0)} = g(0)^2, \quad \sqrt{\frac{p(1)}{2}} = g(0)g(1),$$
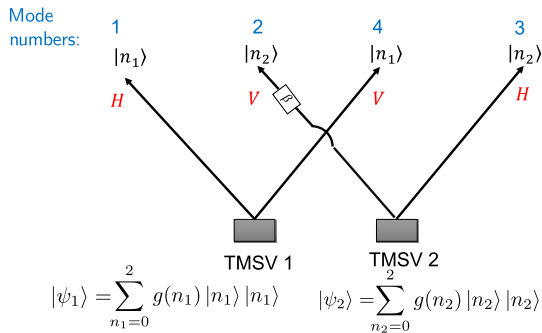
$$\sqrt{\frac{p(2)}{3}} = g(0)g(2) = g(1)^2. \tag{A3}$$

Therefore, the final state is given by

$$|\psi^{\pm}\rangle = N_0 \left[\sqrt{p(0)} |0,0;0,0\rangle \right.$$

$$+ \sqrt{\frac{p(1)}{2}} (|1,0;0,1\rangle \pm |0,1;1,0\rangle)$$

$$\left. + \sqrt{\frac{p(2)}{3}} (|2,0;0,2\rangle \pm |1,1;1,1\rangle + |0,2;2,0\rangle) \right],$$

$$\tag{A4}$$

FIG. 24.　Unfolded version of the source in Fig. 2 from Refs. [28,29]. The two passes through the nonlinear crystal generate two two-mode squeezed vacuum states ($|\psi_1\rangle$ and $|\psi_2\rangle$). The effect of the arm swap is explicitly shown with the choice of internal phase β.
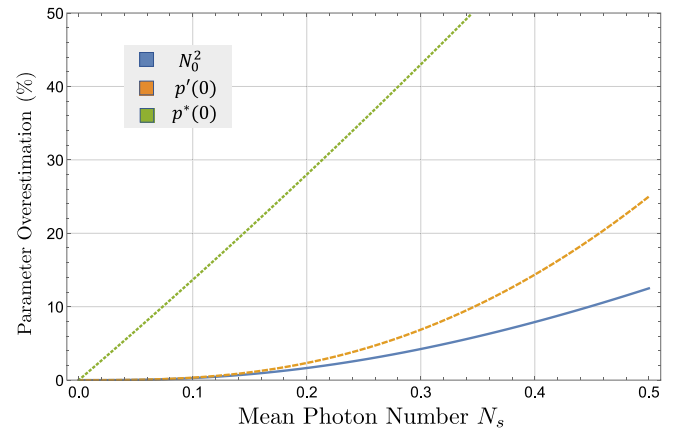
FIG. 25.　Parameter overestimation caused by various normalization schemes for state $|\psi\rangle$.

where $N_0 = 1/\sqrt{p(0)+p(1)+p(2)}$ is used to normalize the state. This method preserves the probability ratio $p(1)^2/[p(0)p(2)]$ even as N_s increases, at the cost of overestimating each probability individually [since Eq. (A4) omits higher-order terms that will become more relevant as N_s increases]. In Fig. 25 we plot this error, as well as the errors contributed by two additional methods of normalizing the state:

(a) define $p'(0) \equiv 1 - p(1) - p(2)$;
(b) define $p^*(2) \equiv 1 - p(0) - p(1)$.

The p' normalization preserves the one- and two-photon pair probabilities of $|\psi\rangle$, but overestimates vacuum contributions. Similarly, p^* normalization allows for accurate representation of the vacuum and one-photon pair terms, but all multipair events are treated as having two pairs. These alternate normalization schemes have some advantages since they do not overestimate the one-photon pair probability. However, this comes at the cost of faster divergence and less convenience when performing analytic calculations.

APPENDIX B: EFFECTS OF IMPERFECT PHOTON INDISTINGUISHABILITY

In our present analysis, we assume that the down-conversion process that generates the states given by Eqs. (A1a)–(A1b) emits photons in well-defined modes. Thus, a cascaded source that combines two original SPDC sources using a BSM is devoid of imperfections that may influence the heralded output state. In practical implementations of the SPDC source, this assumption breaks down. The emitted photons can occupy different spectral, spatial, or temporal modes. Hence, they are inherently distinguishable, which reduce the interference visibility factor from unity to some value V [50,51]. The calculation of V for a specific setup can be done by performing mode profile
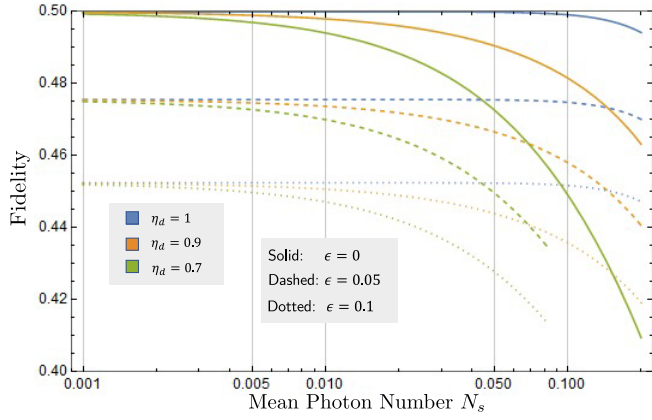
FIG. 26. Fidelity of the entangled state generated by the cascaded source with the ideal Bell state $|\Psi^+\rangle$ at various values of state decoherence ϵ and detector efficiency η_d, plotted as a function of N_s. We assume that $P_d = 0$.
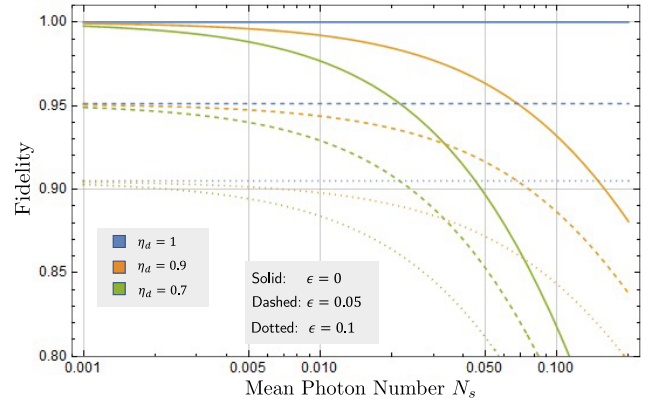

FIG. 27. Fidelity of the entangled state generated by the cascaded source with the ideal Bell state $|\Psi^+\rangle$ after successfully loading into idealized heralded quantum memories. The different line styles and colors signify various state decoherence (ϵ) and detector efficiency (η_d) values, respectively, plotted as a function of N_s. We assume that $P_d = 0$.

tomography on the individual down-conversion processes and evaluating a mode overlap integral. The present work does not consider a specific system setup for the down-conversion process; hence, we do not include actual mode overlap integral calculations in our work. Given the V for a specific setup, we use a simple model to study the effects. If the "inner" output modes of the SPDC sources have a HOM visibility V, assuming that the "perfect visibility" state has fidelity F_0 with our target, then the final state fidelity with visibility V is

$$F(V) = F_0 V + 1/2 \times (1 - V) = 1/2 + (F_0 - 1/2)V. \tag{B1}$$

Let us specifically consider the temporal degree of freedom for the emitted photons. A complete and rigorous analysis of this effect is beyond the scope of this work. However, as the dual-rail Bell state is the target state for the present article, we examine the single-pair emission terms in the complete photonic state emitted by the SPDC [i.e., $n_1 = 1$ in Eq. (A1a); $n_2 = 1$ in Eq. (A1b)].

It is well understood that timing walk-off between the emitted photons induces a partial decoherence in the output state; this can also be understood as a frequency-dependent relative phase between the two terms in the entangled state—tracing over that phase leads to effective loss of coherence. In the basis of the single-pair emission term, the density matrix of the SPDC source resembles

$$\hat{\rho}_{\text{single}} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1-\epsilon & 0 \\ 0 & 1-\epsilon & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{B2}$$

The parameter ϵ accounts for the decoherence of the state, where $\epsilon = 0$ denotes the absence of any decoherence due

to the timing walk-off effect. We observe that the introduction of decoherence in the state description limits the maximum fidelity achievable by the state. This can be explained by the presence of the nonzero decoherence parameter (in the cross terms of the density matrix), which decreases the overlap with the target state $|\Psi^+\rangle$. This is depicted in Fig. 26, where we plot the fidelity of the entangled state from a *cascaded* source with respect to the ideal Bell state $|\Psi^+\rangle$ as a function of the mean photon number (N_s), when the underlying SPDC source state has decoherence values (ϵ) as marked in the legend. In Fig. 27 we plot the state fidelity as a function of N_s, after the output photonic state has been loaded into the IQM.

Based on the results of the figure, we can roughly derive the following empirical relations between the maximum achievable fidelity (both pre- and postloading into the IQM):

$$\max_{\{N_s, \eta_d\}} F(\hat{\sigma}, |\Psi^+\rangle) = 0.5(1 - \epsilon), \tag{B3a}$$

$$\max_{\{N_s, \eta_d\}} F^{\text{VON}}(\hat{\sigma}, |\Psi^+\rangle) = (1 - \epsilon). \tag{B3b}$$

Here $\hat{\sigma}$ denotes the density operator of the output state (we omit the complete description for brevity). Therefore, it is clear that, given the value of ϵ for the underlying SPDC sources, the maximum fidelity target for the cascaded source (and in extension the multiplexed source from Sec. VI) is limited to $1 - \epsilon$.

APPENDIX C: HYBRID FOCK-COHERENT SYSTEM MODELING

Although complete system modeling in the Fock basis is exact and complete when we truncate the infinite basis

sufficiently, it poses a variety of computational difficulties. Inclusion of component efficiency involves the consideration of additional environment modes that need to be traced out. Treating the effects of pure loss using the Kraus operator formalism is one way to circumvent this difficulty. The Fock-basis representation of the output is not well suited for this treatment as account for the environment modes causes the system size to scale quickly. In our calculations we adopt a hybrid Fock-coherent approach.

This approach treats any linear optical component that manipulates the quantum state, as well as loss in the bosonic modes, in the coherent basis. The final state can then be projected onto the Fock basis to generate the complete density matrix description. We highlight the key steps to this approach in the subsequent paragraphs. The complete density matrix description of the final state is omitted for brevity in this appendix.

We begin by identifying that the original state from Eq. (4) is a pair of TMSV states. Hence, it is conceptually much simpler to treat the complete polarization entangled source as two different TMSVs that are connected, as shown in Fig. 24. The whole state is a tensor product of the two component states, with the mode labels suitably rearranged as described in Appendix A. Experimentally, this is analogous to the two down-conversion processes corresponding to the double pass through the nonlinear $\chi^{(2)}$ crystal with the idler modes being swapped using a polarizing beam splitter [28].

Given a single TMSV state $|\psi\rangle = \sum_m c(m) |m; m\rangle$, the corresponding density matrix is

$$\rho = \sum_{m,m'=0}^{n,n'} c_m c_{m'}^* |m; m\rangle \langle m'; m'|. \tag{C1}$$

We now take two such sources to model the complete SPDC source, changing the labels of their Fock state to keep the indices distinct. The final state is then represented as

$$
\begin{aligned}
\rho_A \otimes \rho_B &= \sum_{n_A, n'_A, n_B, n'_B} c_{A,m_A} c_{A,m'_A}^* c_{B,m_B} c_{B,m'_B}^* \sum_{\text{all m and m'}} |m_A, m_A, m_B, m_B\rangle \langle m'_A, m'_A, m'_B, m'_B| \\
&= \frac{1}{\pi^4} \int_{\alpha_i, \gamma_j} \sum_{\text{all } n \text{ and } m} C_{m_A, m_B, m'_A, m'_B} |m_A\rangle |\alpha_1, \alpha_2\rangle \langle \alpha_1, \alpha_2 | m_A, m_B\rangle |m_B\rangle \langle m'_A| \langle m'_A, m'_B | \gamma_1, \gamma_2\rangle \langle \gamma_1, \gamma_2 | \langle m'_B| \\
&= \frac{1}{\pi^4} \int \sum_{\text{all sums}} C_{m_A, m_B, m'_A, m'_B} |m_A\rangle |\alpha_1, \alpha_2\rangle |m_B\rangle \langle m'_A| \langle \gamma_1, \gamma_2 | \langle m'_B| \exp\left[-(|\alpha_1|^2 + |\alpha_2|^2 + |\gamma_1|^2 + |\gamma_2|^2)/2\right] \\
&\quad \times \frac{\alpha_1^{m_A}}{\sqrt{m_A!}} \frac{\alpha_2^{m_B}}{\sqrt{m_B!}} \frac{\gamma_1^{m'_A}}{\sqrt{m'_A!}} \frac{\gamma_2^{m'_B}}{\sqrt{m'_B!}}.
\end{aligned}
\tag{C2}
$$

In the second and third equalities of Eq (C2), we project a pair of the modes on the coherent basis elements $|\alpha\rangle_i, |\gamma\rangle_j$ where $i, j = 1, 2$. These two modes are those that get mixed and detected in the final cascaded source. To account for all losses and the beam splitter interaction, one can take a pair of modes whose coherent basis kets are $|\alpha_i\rangle$ and $|\alpha_j\rangle$, and carry out the necessary manipulations on the coherent basis elements as

$$
\begin{aligned}
|\alpha_i, \alpha_j\rangle &\xrightarrow{\eta_c} |\alpha_i \sqrt{\eta_c}, \alpha_j \sqrt{\eta_c}\rangle \\
&\xrightarrow{\text{50:50 BS}} \left| \frac{\alpha_i \sqrt{\eta_c} + \alpha_j \sqrt{\eta_c}}{\sqrt{2}}, \frac{-\alpha_i \sqrt{\eta_A} + \alpha_j \sqrt{\eta_c}}{\sqrt{2}} \right\rangle \\
&\xrightarrow{\eta_d} \left| \sqrt{\eta_d} \frac{\alpha_i \sqrt{\eta_c} + \alpha_j \sqrt{\eta_c}}{\sqrt{2}}, \sqrt{\eta_d} \frac{-\alpha_i \sqrt{\eta_A} + \alpha_j \sqrt{\eta_c}}{\sqrt{2}} \right\rangle.
\end{aligned}
\tag{C3}
$$

Here, η_c and η_d account for the coupling and detection efficiencies, respectively. Hence, the complete interaction that generates the heralded state from the cascaded source can be summarized as

$$(\rho_A \otimes \rho_B)^{\otimes 2} = \bar{\rho} \xrightarrow{\text{loss+BSM}} \tilde{\rho} \xrightarrow{\text{det.}} \tilde{\sigma}_i, \tag{C4}$$

where $\tilde{\sigma}_i$ is the output state conditioned on a certain click pattern as given in the main text. Note that $\tilde{\sigma}_i$ is expressed in a mixed basis (part Fock and part coherent). In order to get back to the Fock-basis density matrix, we adopt the techniques developed in Ref. [52] to perform a Fock projection on the coherent basis description of the quantum state of a bosonic mode.

Performing the preceding mathematical operations on the final density matrix lets us evaluate the various metrics of interest:

1. calculating $\text{Tr}(\tilde{\sigma}_i)$ yields the generation probability (P_{gen});

2. the fidelity can be determined as

$$F = \frac{\langle \Psi^+ | \tilde{\sigma}_i | \Psi^+ \rangle}{\mathrm{Tr}(\tilde{\sigma}_i)}. \tag{C5}$$

APPENDIX D: MODELING A PRACTICAL PHOTON-NUMBER RESOLVING DETECTOR

PNR detectors [45,47,48] are key components of the linear optical BSM circuit, heralding the outcome of an entanglement swap attempt. When swapping entanglement between two ideal dual-rail basis Bell pairs, only specific click patterns across the multiple detectors indicates a possible "success." In our proposal for the improved source, this "success" information heralds the generation of the state given in Eq. (5).

Theoretically, the ideal PNR measurement is a projective measurement given by the action of POVM elements $\Pi_n = |n\rangle\langle n|, n \in \mathbb{Z}^+$, on the input state, say $|\psi\rangle$. A measurement result of k clicks would collapse the input state onto one of the Fock basis elements, in this case $|k\rangle$. One can use ideal PNR detectors repeatedly to determine the photon statistics of the input state and "reconstruct" the state (this is the whole focus of quantum state tomography). This simple model is shown in Fig. 28.

Since we consider the use of imperfect PNR detectors, there are multiple factors that must now be taken into account. Presently, we only focus on the effects that affect the quantum state being detected. Factors that influence the physical operability of the detector are not treated by this model, i.e., timing jitter, postdetection dead time, afterpulsing, and count saturation are not explicitly considered here (though they may be included implicitly, e.g., dead time results in reduced efficiency, and afterpulsing leads to increased noise counts). We assume that the PNR detector has a detection efficiency η_d and dark click probability P_d (uncorrelated to the input state), that the detector is synchronized with the rest of the circuit, and that the pulse profile, bandwidth, and frequency are optimized *a priori*. The detector efficiency can be interpreted as the input state being transmitted through a bosonic pure loss channel of transmissivity η_d before the actual detection happens, which yields an (inaccessible) outcome k. The detector dark clicks can be treated as a Bernoulli random variable with probability of success P_d. The outcome of this random variable is convoluted with the actual output from the ideal PNR measurement after loss. Thus, an observation of $l \geq 1$ clicks may signify one of two cases.

(a) There were l clicks and no dark clicks.
(b) There were $l - 1$ clicks and a single dark click.

Thus, the probability of l clicks in the detector is given by

$$P[l] = P[k = l - 1] \times P_d + P[k = l] \times (1 - P_d). \tag{D1}$$

The complete model of the nonideal PNR measurement is depicted in Fig. 29. We note that observation of a correct click pattern indicates a successful entanglement swap only for a small subset of cases. For example, an observed pattern of $\vec{n}_{\mathrm{observed}} = (1, 1, 0, 0)$ may correspond to any of the cases

$$\begin{aligned}
P_{(1,1,0,0)}^{\mathrm{observed}} &= (1 - P_d)^4 \times P_{(1,1,0,0)} \\
&+ P_d(1 - P_d)^3 \times [P_{(1,0,0,0)} + P_{(0,1,0,0)}] \\
&+ P_d^2(1 - P_d)^2 \times P_{(0,0,0,0)}, \tag{D2}
\end{aligned}$$

where $P_{(\vec{n})}$ is the probability that in reality the detection pattern was $(\vec{n})$ prior to the dark clicks.

In the proposed cascaded source, if we consider only detector dark clicks (no detector loss, i.e., $\eta_d = 1$), we would obtain the following mixed state at the output:

$$\bar{\rho} = (1 - P_d)^4 \hat{\rho}_1 + (1 - P_d)^3 P_d \, \hat{\rho}_2 + (1 - P_d)^2 P_d^2 \hat{\rho}_3 \tag{D3}$$

with

$$\begin{aligned}
\hat{\rho}_1 = c_1 (|1,0;0,1\rangle + |0,1;1,0\rangle + |1,1;0,0\rangle + |0,0;1,1\rangle) \\
\times (\langle 1,0;0,1| + \langle 0,1;1,0| + \langle 1,1;0,0| + \langle 0,0;1,1|), \tag{D4a}
\end{aligned}$$

$$\begin{aligned}
\hat{\rho}_2 = c_2 (|1,0;0,0\rangle + |0,0;1,0\rangle + |0,0;0,1\rangle + |0,1;0,0\rangle) \\
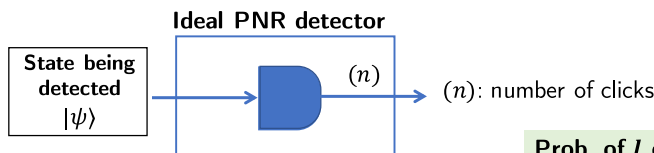\times (\langle 1,0;0,0| + \langle 0,0;1,0| + \langle 0,0;0,1| + \langle 0,1;0,0|), \tag{D4b}
\end{aligned}$$

$$\hat{\rho}_3 = c_3 |0,0;0,0\rangle\langle 0;0;0;0|. \tag{D4c}$$

Here

$$c_1 = \frac{p(1)^2}{16}, \quad c_2 = \frac{p(0)p(1)}{4}, \quad c_3 = p(0)^2. \tag{D5}$$

APPENDIX E: "VACUUM OR NOT" QUANTUM NONDEMOLITION MEASUREMENT

One essential processing step required to attain a high fidelity for the generated quantum state with respect to

Ideal PNR detector

State being detected $|\psi\rangle$ — (n) → (n): number of clicks

Prob. of l clicks: $|\langle l|\psi\rangle|^2$

Measurement projectors: $\{|n\rangle\langle n|\}; n \in 0, 1, 2, \ldots, \infty$

FIG. 28. Model of the ideal PNR detector, which generates clicks depending upon the photon statistics of the input state.
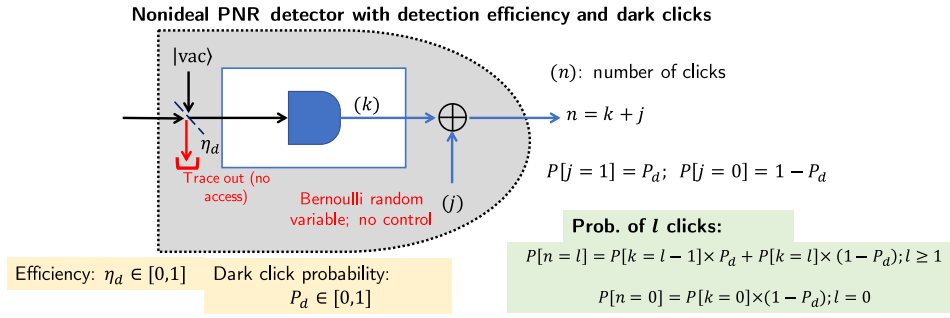
FIG. 29. Model of the nonideal PNR detector. The nonideal PNR detector has two parameters that affect the operation, namely the detector efficiency η_d and dark click probability P_d.

the target entangled state is a VON quantum nondemolition measurement. Such a measurement is a theoretical tool essential to filter out the vacuum component of the quantum state, a major component that drives down the state fidelity. Given an N-mode quantum state $|\psi\rangle$, the VON measurement can be ideally modeled as a black box producing one of two outcomes.

(a) The N-mode vacuum state: $|\psi_0\rangle = |0\rangle^{\otimes N}$ with a probability of $p_{|0\rangle^{\otimes N}}$.

(b) The vacuum-subtracted quantum state from $|\psi\rangle$,

$$|\psi_1\rangle = \frac{|\psi\rangle - \sqrt{p_{|0\rangle^{\otimes N}}}\,|0\rangle^{\otimes N}}{\sqrt{1 - |p_{|0\rangle^{\otimes N}}|^2}}$$

with a probability of $1 - p_{|0\rangle^{\otimes N}}$. The VON measurement and it outcomes are summarized in Fig. 30.

The probability of the vacuum outcome is given by $|p_{|0\rangle^{\otimes N}}|^2 = |\langle 00\cdots 0|\psi\rangle|^2$. We note that, while a photonic implementation of the VON measurement is still an open problem, there are preliminary proposals to implement the same in atomic systems coupled to optical cavities [42].

APPENDIX F: ANALYTIC EXPRESSIONS FOR THE FIDELITY AND P_{gen}

The most general formulae for the fidelity and probability of generation P_{gen} of the output quantum state from the cascaded source (considering inefficiencies in coupling and nonideal detectors) are given by

$$F(|S\rangle, |\Psi^+\rangle) = \frac{\mathcal{A}_1}{\mathcal{B}_1}, \tag{F1a}$$

$$P_{\text{gen}} = \mathcal{B}_1, \tag{F1b}$$

where

$$\mathcal{A}_1 = \frac{2N_s^2[(5P_d^2 - 4P_d + 1)\eta_c^2\eta_d^2 + 2(1 - 3P_d)P_d\eta_c\eta_d + 2P_d^2]}{(N_s + 1)^6}, \tag{F2}$$

$$\begin{aligned}
\mathcal{B}_1 = \frac{4}{(N_s + 1)^8}\big(&3\eta_c^4\eta_d^4[P_d(11P_d - 10) + 2]N_s^4 - 2\eta_c^3\eta_d^3[P_d(17P_d - 13) + 2]N_s^3(4N_s + 1) \\
&+ \eta_c^2\eta_d^2 N_s^2\{P_d^2[N_s(211N_s + 118) + 21] - 2P_d[N_s(61N_s + 34) + 6] + 11N_s^2 + 6N_s + 1\} \\
&- 2\eta_c\eta_d P_d(3P_d - 1)N_s(4N_s + 1)(6N_s^2 + 4N_s + 1) + P_d^2(6N_s^2 + 4N_s + 1)^2\big).
\end{aligned} \tag{F3}$$

The fidelity and probability of generating the output quantum state after the IQM's VON filtering (considering inefficiencies in coupling and nonideal detectors) are given by

$$F^{\text{VON}}(|S\rangle, |\Psi^+\rangle) = \frac{\mathcal{A}_2}{\mathcal{B}_2}, \tag{F4a}$$

$$P_{\text{gen}}^{\text{VON}} = \mathcal{B}_2, \tag{F4b}$$

where

$$\mathcal{A}_2 = \frac{\eta_c^2\eta_d^2 + P_d^2[-3\eta_c\eta_d + \eta_c\eta_d(5\eta_c\eta_d - 3) + 2] + \eta_d P_d(2\eta_c - 4\eta_c^2\eta_d)}{2(N_s + 1)^6}, \tag{F5}$$
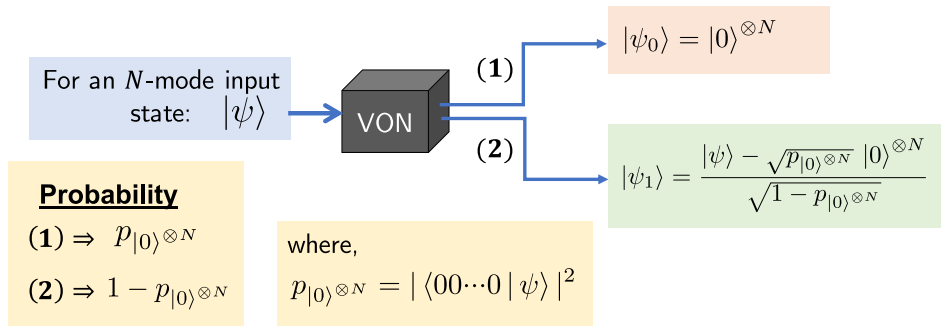
FIG. 30. Representation of the "vacuum or not" quantum non-demolition measurement. Given an input N-mode quantum state, the VON black box separates the N-mode vacuum, from the desired nonvacuum component, with a probability $p_{|0\rangle^{\otimes N}}$. With the complementary probability it separates the nonvacuum component from the input state.

$$
\begin{aligned}
\mathcal{B}_2 = {} & \frac{1}{2(1+N_s)^8} \{ P_d^2 [17\eta_c^2\eta_d^2 - 24\eta_c\eta_d + N_s^2 (66\eta_c^4\eta_d^4 - 272\eta_c^3\eta_d^3 + 397\eta_c^2\eta_d^2 - 240\eta_c\eta_d + 50) \\
& - 2N_s(34\eta_c^3\eta_d^3 - 93\eta_c^2\eta_d^2 + 78\eta_c\eta_d - 20) + 8] + 2\eta_c\eta_d P_d [-5\eta_c\eta_d + N_s^2(-30\eta_c^3\eta_d^3 + 104\eta_c^2\eta_d^2 - 115\eta_c\eta_d + 40) \\
& + N_s(26\eta_c^2\eta_d^2 - 54\eta_c\eta_d + 26) + 4] + \eta_c^2\eta_d^2[N_s^2(12\eta_c^2\eta_d^2 - 32\eta_c\eta_d + 21) + N_s(10 - 8\eta_c\eta_d) + 1]\}.
\end{aligned} \tag{F6}
$$

APPENDIX G: GAUSSIAN MODELING OF THE CASCADED SOURCE

Despite the use of a hybrid Fock-coherent modeling (as discussed in Appendix C) for the source analysis, we note that the complete system can be modeled as a Gaussian boson sampling circuit [53,54]. This observation stems from the fact that all input states (i.e., the vacuum state) and quantum operations (i.e., two-mode squeezing, beam splitters, and necessary pure-loss channels) are Gaussian in nature. PNR-detection-based heralding of the final state is the only step that introduces non-Gaussianity. This means that, prior to detection, quadrature displacements and the covariance matrix provide a complete description of the quantum state. Ref. [52] covers the analysis of a Gaussian boson sampling circuit using the coherent basis K function.

Let us begin with the Gaussian pure state after the unitary $U(\zeta)$ as $|\psi_0\rangle$ (marked by the red dashed line in Fig. 31). Since the internal beam splitters of the circuit are balanced, we can commute all our coupling losses to manifest just before detection. This is an operational trick to keep the state description clear of difficulties with swapping mixed states. This is justified because coupling losses are uniform and the beam splitters are balanced. We can write the K function of state $|\psi_0\rangle$ as $K(\vec{\alpha})$ [52] as

$$
|\psi_0\rangle = \frac{1}{(2\pi)^8} \int d^{16}\alpha \, K(\vec{\alpha}) \, |\vec{\alpha}\rangle, \tag{G1}
$$

where $|\vec{\alpha}\rangle = |\alpha_1, \ldots, \alpha_8\rangle$ is the coherent basis vector and $\alpha_i = (q_{\alpha_i} + ip_{\alpha_i})/\sqrt{2}$. Here $q_{\alpha_i}, p_{\alpha_i}$ are the quadrature variables of the ith mode as marked in Fig. 31. Similarly, the density operator description of state $|\psi_0\rangle$ can be expressed in terms of the K function as

$$
\hat{\rho}_0 = |\psi_0\rangle\langle=| \frac{1}{(2\pi)^{16}} \int d^{16}\alpha \, d^{16}\beta \, K(\vec{\alpha})K^*(\vec{\beta}) \, |\vec{\alpha}\rangle\langle\vec{\beta}| . \tag{G2}
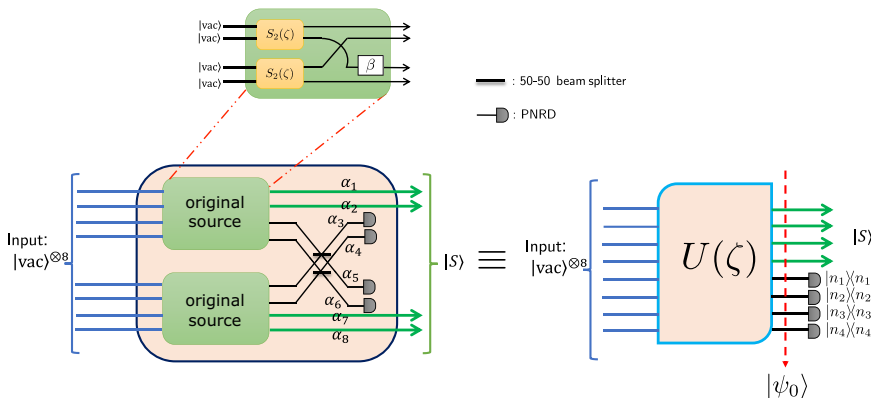$$

FIG. 31. The complete abstraction of the *cascaded* source as a Gaussian boson sampling circuit setup for state preparation. We start with eight input vacuum modes and perform a unitary operation $U(\zeta)$ (which includes the unitary for the two-mode squeezers and intermediate beam splitters) parameterized by ζ. This corresponds to the mean photon number per mode N_s of the underlying two-mode squeezed states.

We adopt a Kraus operator-based approach (similar to Appendix C) to account for the pure loss (due to coupling and detection efficiency). The Kraus operators for the action of a channel of transmissivity η are given by

$$A_k = \sqrt{\frac{(1-\eta)}{k!}} \eta^{\hat{n}/2} \hat{a}^k. \tag{G3}$$

The action of these operators on a general coherent basis term $|\gamma\rangle\langle\delta|$, where $\gamma, \delta \in \mathbb{C}$, is given by

$$\sum_{k=0}^{\infty} \hat{A}_k |\gamma\rangle\langle\delta| \hat{A}_k^\dagger = \sum_k \frac{(1-\eta)^k}{k!} (\sqrt{\eta})^{\hat{n}} \hat{a}^k |\gamma\rangle\langle\delta| \hat{a}^{\dagger k} (\sqrt{\eta})^{\hat{n}}$$

$$= \sum_k \frac{(1-\eta)^k}{k!} (\gamma\delta^*)^k \exp\left(-\frac{(|\gamma|^2 + |\delta|^2)(1-\eta)}{2}\right) |\gamma\sqrt{\eta}\rangle\langle\delta\sqrt{\eta}|$$

$$= \exp\left((\gamma\delta^*)(1-\eta) - \frac{(|\gamma|^2 + |\beta|^2)(1-\eta)}{2}\right) |\gamma\sqrt{\eta}\rangle\langle\delta\sqrt{\eta}|. \tag{G4}$$

Hence, we observe that the basis elements are modified as

$$|\alpha_1; \alpha_2; \alpha_3; \alpha_4; \alpha_5; \alpha_6; \alpha_7; \alpha_8\rangle\langle\beta_1; \beta_2; \beta_3; \beta_4; \beta_5; \beta_6; \beta_7; \beta_8|$$
$$\Rightarrow g(\vec{\alpha}, \vec{\beta}, \vec{\eta}) \times |\alpha_1\sqrt{\eta_c}; \alpha_2\sqrt{\eta_c}; \alpha_3\sqrt{\eta_c\eta_d}; \alpha_4\sqrt{\eta_c\eta_d}; \alpha_5\sqrt{\eta_c\eta_d}; \alpha_6\sqrt{\eta_c\eta_d}; \alpha_7\sqrt{\eta_c}; \alpha_8\sqrt{\eta_c}\rangle$$
$$\times \langle\beta_1\sqrt{\eta_c}; \beta_2\sqrt{\eta_c}; \beta_3\sqrt{\eta_c\eta_d}; \beta_4\sqrt{\eta_c\eta_d}; \beta_5\sqrt{\eta_c\eta_d}; \beta_6\sqrt{\eta_c\eta_d}; \beta_7\sqrt{\eta_c}; \beta_8\sqrt{\eta_c}|. \tag{G5}$$

The function $g(\vec{\alpha}, \vec{\beta}, \vec{\eta})$ accounts for the mixed nature of the final state after loss, and is compactly expressed as

$$g(\vec{\alpha}, \vec{\beta}, \vec{\eta}) = \prod_{i=1}^{8} \exp\left((\alpha_i\beta_i^*)(1-\eta_i) - \frac{(|\alpha_i|^2 + |\beta_i|^2)(1-\eta_i)}{2}\right), \tag{G6}$$

where

$$\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3; \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8), \tag{G7a}$$

$$\vec{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8), \tag{G7b}$$

$$\vec{\eta} = (\sqrt{\eta_c}, \sqrt{\eta_c}, \sqrt{\eta_c\eta_d}, \sqrt{\eta_c\eta_d}, \sqrt{\eta_c\eta_d}, \sqrt{\eta_c\eta_d}, \sqrt{\eta_c}, \sqrt{\eta_c}). \tag{G7c}$$

Hence, we can write the density matrix of the state after loss as

$$\hat{\rho}_1 = \frac{1}{(2\pi)^{16}} \int d^{16}\alpha \, d^{16}\beta \, K(\vec{\alpha}) \, K^*(\vec{\beta}) \, g(\vec{\alpha}, \vec{\beta}, \vec{\eta}) \, |\vec{\alpha}\vec{\eta}\rangle\langle\vec{\beta}\vec{\eta}|. \tag{G8}$$

In the general approach to analyze Gaussian boson sampling circuits, we perform photon-number projection on the modes to be detected [11,52], which yields a non-Gaussian state (pure if there are no losses in any of the modes and mixed otherwise) that can be characterized. In the current analysis,

1. we have a specific requirement for the quantum state in the undetected modes, i.e., $|\Psi^+\rangle$;

2. we know the "correct" detection patterns from our preliminary analysis.

With this knowledge, we may subsume the photon-number detection step into our fidelity calculation. Since we know that the target Bell state exists in the undetected "outer" modes (along with spurious terms), given the measurement outcomes from the PNRDs (say we get n_1, n_2, n_3, and n_4 clicks, respectively), we effectively know the eight-mode state that would yield a Bell state should the intermediate four modes be detected. Therefore, using this insight, our simplified technique for calculating the fidelity is equivalent to evaluating the overlap:

$$F(|\xi\rangle, \hat{\rho}_1) = \langle\xi|\hat{\rho}_1|\xi\rangle. \tag{G9}$$

Here

$$|\xi\rangle = \frac{1}{\sqrt{2}}(|1, 0, n_1, n_2, n_3, n_4, 0, 1\rangle$$
$$+ (-1)^{m_1} |0, 1, n_1, n_2, n_3, n_4, 1, 0\rangle) \tag{G10}$$

and n_1, n_2, n_3, n_4 can be any of the patterns from the table that determine the internal phase terms (see Sec. II). The probability P_{gen} to generate the state from Eq. (5) is given by $\text{Tr}(\Pi_e \hat{\rho}_1)$, where

$$\Pi_e = \mathbb{1}^{\otimes 2} \otimes |n_1, n_2, n_3, n_4\rangle\langle\otimes| \, \mathbb{1}^{\otimes 2}. \tag{G11}$$

This simplifies to

$$
\begin{aligned}
\text{Tr}(\Pi_e \hat{\rho}_1) = \frac{1}{(2\pi)^{16}} \int d^{16}\alpha \, d^{16}\beta \; & K(\vec{\alpha}) \, K^*(\vec{\beta}) \, g(\vec{\alpha}, \vec{\beta}, \vec{\eta}) \\
\times \; & \langle\beta_1\sqrt{\eta_c}|\alpha_1\sqrt{\eta_c}\rangle \langle\beta_2\sqrt{\eta_c}|\alpha_2\sqrt{\eta_c}\rangle \langle\beta_7\sqrt{\eta_c}|\alpha_7\sqrt{\eta_c}\rangle \langle\beta_8\sqrt{\eta_c}|\alpha_8\sqrt{\eta_c}\rangle \\
\times \; & \langle n_1|\alpha_3\sqrt{\eta_c\eta_d}\rangle \langle n_2|\alpha_4\sqrt{\eta_c\eta_d}\rangle \langle n_3|\alpha_5\sqrt{\eta_c\eta_d}\rangle \langle n_4|\alpha_6\sqrt{\eta_c\eta_d}\rangle \\
\times \; & \langle\beta_3\sqrt{\eta_c\eta_d}|n_1\rangle \langle\beta_4\sqrt{\eta_c\eta_d}|n_2\rangle \langle\beta_5\sqrt{\eta_c\eta_d}|n_3\rangle \langle\beta_6\sqrt{\eta_c\eta_d}|n_4\rangle .
\end{aligned}
\tag{G12}
$$

We identify two distinct types of overlap terms in the integral that simplify as

$$\langle\beta_i\eta_i|\alpha_i\eta_i\rangle = \exp\left[-\frac{\eta_i^2}{2}(|\alpha_i|^2 + |\beta_i|^2 - 2\alpha_i\beta_i^*)\right], \tag{G13}$$

$$\langle\beta_i\eta_i|n_j\rangle\langle n_j|\alpha_i\eta_i\rangle = \eta_i^{n_j}\frac{(\alpha_i\beta_i^*)^{n_j}}{n_j!}\exp\left[-\frac{\eta_i^2}{2}(|\alpha_i|^2 + |\beta_i|^2)\right]. \tag{G14}$$

In the present manuscript, we have sufficient evidence from the hybrid analysis in Appendix C to claim that the cascaded source can be efficiently operated for $N_s < 0.2$. The reader may solve the detailed integrals in Eqs. (G9) and (G12) if they wish to analyze the performance of the source in the regime of $N_s > 0.2$.

[1] S. Wehner, D. Elkouss, and R. Hanson, Quantum internet: A vision for the road ahead, Science **362**, 6412 (2018).

[2] V. V. Albert, K. Noh, K. Duivenvoorden, D. J. Young, R. T. Brierley, P. Reinhold, C. Vuillot, L. Li, C. Shen, S. M. Girvin, B. M. Terhal, and L. Jiang, Performance and structure of single-mode bosonic codes, Phys. Rev. A **97**, 032346 (2018).

[3] E. Knill, R. Laflamme, and G. J. Milburn, A scheme for efficient quantum computation with linear optics, Nature **409**, 46 (2001).

[4] D. Gottesman, A. Kitaev, and J. Preskill, Encoding a qubit in an oscillator, Phys. Rev. A **64**, 012310 (2001).

[5] F. Ewert and P. van Loock, 3/4-Efficient Bell Measurement with Passive Linear Optics and Unentangled Ancillae, Phys. Rev. Lett. **113**, 140403 (2014).

[6] M. Gimeno-Segovia, P. Shadbolt, D. E. Browne, and T. Rudolph, From Three-Photon Greenberger-Horne-Zeilinger States to Ballistic Universal Quantum Computation, Phys. Rev. Lett. **115**, 020502 (2015).

[7] M. Pant, D. Towsley, D. Englund, and S. Guha, Percolation thresholds for photonic quantum computing, Nat. Commun. **10**, 1070 (2019).

[8] S. Bartolucci, P. Birchall, H. Bombin, H. Cable, C. Dawson, M. Gimeno-Segovia, K. Johnston, E. Kieling,

[9] K. Noh, V. V. Albert, and L. Jiang, Quantum capacity bounds of gaussian thermal loss channels and achievable rates with Gottesman-Kitaev-Preskill codes, IEEE Trans. Inf. Theory **65**, 2563 (2019).

N. Nickerson, M. Pant, F. Pastawski, T. Rudolph, and C. Sparrow, Fusion-based quantum computation, arXiv:2101.09310 [quant-ph] (2021).

[10] M. Eaton, R. Nehra, and O. Pfister, Non-gaussian and Gottesman–Kitaev–Preskill state preparation by photon catalysis, New J. Phys. **21**, 113034 (2019).

[11] D. Su, C. R. Myers, and K. K. Sabapathy, Conversion of gaussian states to non-gaussian states using photon-number-resolving detectors, Phys. Rev. A **100**, 052301 (2019).

[12] S. Guha, H. Krovi, C. A. Fuchs, Z. Dutton, J. A. Slater, C. Simon, and W. Tittel, Rate-loss analysis of an efficient quantum repeater architecture, Phys. Rev. A **92**, 022357 (2015).

[13] M. Pant, H. Krovi, D. Englund, and S. Guha, Rate-distance tradeoff and resource costs for all-optical quantum repeaters, Phys. Rev. A **95**, 012304 (2017).

[14] M. Pant, H. Krovi, D. Towsley, L. Tassiulas, L. Jiang, P. Basu, D. Englund, and S. Guha, Routing entanglement in the quantum internet, Npj Quantum Inf. **5**, 1 (2019).

[15] P. Nain, G. Vardoyan, S. Guha, and D. Towsley, On the analysis of a multipartite entanglement distribution switch, SIGMETRICS Perform. Eval. Rev. **48**, 49 (2020).

[16] K. Goodenough, D. Elkouss, and S. Wehner, Optimizing repeater schemes for the quantum internet, Phys. Rev. A **103**, 032610 (2021).

[17] Y. Arakawa and M. J. Holmes, Progress in quantum-dot single photon sources for quantum information technologies: A broad spectrum overview, Appl. Phys. Rev. **7**, 021309 (2020).

[18] R. Uppu, F. T. Pedersen, Y. Wang, C. T. Olesen, C. Papon, X. Zhou, L. Midolo, S. Scholz, A. D. Wieck, A. Ludwig, and P. Lodahl, Scalable integrated single-photon source, Sci. Adv. **6,** 50 (2020).

[19] J. Lee, V. Leong, D. Kalashnikov, J. Dai, A. Gandhi, and L. A. Krivitsky, Integrated single photon emitters, AVS Quantum Sci. **2,** 031701 (2020).

[20] R. N. Patel, T. Schröder, N. Wan, L. Li, S. L. Mouradian, E. H. Chen, and D. R. Englund, Efficient photon coupling from a diamond nitrogen vacancy center by integration with silica fiber, Light Sci. Appl. **5,** e16032 (2016).

[21] Y. Yonezu, K. Wakui, K. Furusawa, M. Takeoka, K. Semba, and T. Aoki, Efficient single-photon coupling from a nitrogen-vacancy center embedded in a diamond nanowire utilizing an optical nanofiber, Sci. Rep. **7,** 12985 (2017).

[22] P. G. Kwiat, K. Mattle, H. Weinfurter, A. Zeilinger, A. V. Sergienko, and Y. Shih, New High-Intensity Source of Polarization-Entangled Photon Pairs, Phys. Rev. Lett. **75,** 4337 (1995).

[23] P. G. Kwiat and A. G. White, in *International Quantum Electronics Conference (1998), Paper QWL1* (Optical Society of America), p. QWL1.

[24] P. G. Kwiat, E. Waks, A. G. White, I. Appelbaum, and P. H. Eberhard, Ultrabright source of polarization-entangled photons, Phys. Rev. A **60,** R773 (1999).

[25] J. A. Armstrong, N. Bloembergen, J. Ducuing, and P. S. Pershan, Interactions between light waves in a nonlinear dielectric, Phys. Rev. **127,** 1918 (1962).

[26] M. M. Fejer, G. A. Magel, D. H. Jundt, and R. L. Byer, Quasi-phase-matched second harmonic generation: Tuning and tolerances, IEEE J. Quantum Electron. **28,** 2631 (1992).

[27] H. Krovi, S. Guha, Z. Dutton, J. A. Slater, C. Simon, and W. Tittel, Practical quantum repeaters with parametric down-conversion sources, Appl. Phys. B **122,** 52 (2016).

[28] P. Kok and S. L. Braunstein, Postselected versus non-postselected quantum teleportation using parametric down-conversion, Phys. Rev. A **61,** 042304 (2000).

[29] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger, Experimental quantum teleportation, Nature **390,** 575 (1997).

[30] M. Gimeno-Segovia, H. Cable, G. J. Mendoza, P. Shadbolt, J. W. Silverstone, J. Carolan, M. G. Thompson, J. L. O'Brien, and T. Rudolph, Relative multiplexing for minimising switching in linear-optical quantum computing, New J. Phys. **19,** 063013 (2017).

[31] F. Kaneda and P. G. Kwiat, High-efficiency single-photon generation via large-scale active time multiplexing, Sci Adv **5,** eaaw8586 (2019).

[32] J. Pseiner, L. Achatz, L. Bulla, M. Bohmann, and R. Ursin, Experimental wavelength-multiplexed entanglement-based quantum cryptography, Quantum Sci. Technol. **6,** 035013 (2021).

[33] E. Meyer-Scott, C. Silberhorn, and A. Migdall, Single-photon sources: Approaching the ideal through multiplexing, Rev. Sci. Instrum. **91,** 041101 (2020).

[34] Q. Zhang, X. H. Bao, C. Y. Lu, X. Q. Zhou, T. Yang, T. Rudolph, and J. W. Pan, Demonstration of a scheme for the generation of "event-ready" entangled photon pairs from a single-photon source, Phys. Rev. A **77,** 062316 (2008).

[35] S. Stanisic, N. Linden, A. Montanaro, and P. S. Turner, Generating entanglement with linear optics, Phys. Rev. A **96,** 043861 (2017).

[36] S. A. Fldzhyan, M. Y. Saygin, and S. P. Kulik, Compact linear optical scheme for bell state generation, arXiv:2105.06306 [quant-ph] (2021).

[37] T. Kiyohara, R. Okamoto, and S. Takeuchi, Unified integration scheme using an $N \times N$ active switch for efficient generation of a multi-photon parallel state, Opt. Express **28,** 17490 (2020).

[38] F. Kaneda, K. Garay-Palmett, A. B. U'Ren, and P. G. Kwiat, Heralded single-photon source utilizing highly nondegenerate, spectrally factorable spontaneous parametric downconversion, Opt. Express **24,** 10733 (2016).

[39] T. Hiemstra, T. Parker, P. Humphreys, J. Tiedau, M. Beck, M. Karpiński, B. J. Smith, A. Eckstein, W. S. Kolthammer, and I. Walmsley, Pure Single Photons From Scalable Frequency Multiplexing, Phys. Rev. Appl. **14,** 014052 (2020).

[40] B. Brecht, D. V. Reddy, C. Silberhorn, and M. G. Raymer, Photon Temporal Modes: A Complete Framework for Quantum Information Science, Phys. Rev. X **5,** 041017 (2015).

[41] M. M. Wilde, S. Guha, S. Tan, and S. Lloyd, in *2012 IEEE International Symposium on Information Theory Proceedings* (2012), p. 551.

[42] D. K. L. Oi, V. Potoček, and J. Jeffers, Nondemolition Measurement of the Vacuum State or its Complement, Phys. Rev. Lett. **110,** 210504 (2013).

[43] M. K. Bhaskar, R. Riedinger, B. Machielse, D. S. Levonian, C. T. Nguyen, E. N. Knall, H. Park, D. Englund, M. Lončar, D. D. Sukachev, and M. D. Lukin, Experimental demonstration of memory-enhanced quantum communication, Nature **580,** 60 (2020).

[44] K. C. Chen, E. Bersin, and D. Englund, A polarization encoded photon-to-spin interface, Npj Quantum Inf. **7,** 1 (2021).

[45] L. A. Morais, T. Weinhold, M. P. de Almeida, A. Lita, T. Gerrits, S. W. Nam, A. G. White, and G. Gillett, Precisely determining photon-number in real-time, arXiv:2012.10158 [quant-ph] (2020).

[46] R. Baghdadi, E. Schmidt, S. Jahani, I. Charaev, M. G. W. Müller, M. Colangelo, D. Zhu, K. Ilin, A. D. Semenov, Z. Jacob, M. Siegel, and K. K. Berggren, Enhancing the performance of superconducting nanowire-based detectors with high-filling factor by using variable thickness, Supercond. Sci. Technol. **34,** 035010 (2021).

[47] C. Cahall, K. L. Nicolich, N. T. Islam, G. P. Lafyatis, A. J. Miller, D. J. Gauthier, and J. Kim, Multi-photon detection using a conventional superconducting nanowire single-photon detector, Optica **4,** 1534 (2017).

[48] S. H. Tan, L. A. Krivitsky, and B.-G. Englert, in *Quantum Communications and Quantum Imaging XIV* (International Society for Optics and Photonics, 2016), Vol. 9980, p. 99800E.

[49] D. Fukuda, G. Fujii, T. Numata, K. Amemiya, A. Yoshizawa, H. Tsuchida, H. Fujino, H. Ishii, T. Itatani, S. Inoue, and T. Zama, Titanium-based transition-edge photon number resolving detector with 98% detection efficiency with index-matched small-gap fiber coupling, Opt. Express **19,** 870 (2011).

[50] R.-B. Jin, M. Takeoka, U. Takagi, R. Shimizu, and M. Sasaki, Highly efficient entanglement swapping and teleportation at telecom wavelength, Sci. Rep. **5**, 9333 (2015).

[51] S. Wang, C.-X. Liu, J. Li, and Q. Wang, Research on the Hong-Ou-Mandel interference with two independent sources, Sci. Rep. **9**, 3854 (2019).

[52] C. N. Gagatsos and S. Guha, Efficient representation of gaussian states for multimode non-gaussian quantum state engineering via subtraction of arbitrary number of photons, Phys. Rev. A **99**, 053816 (2019).

[53] C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, Gaussian Boson Sampling, Phys. Rev. Lett. **119**, 170501 (2017).

[54] A. P. Lund, A. Laing, S. Rahimi-Keshari, T. Rudolph, J. L. O'Brien, and T. C. Ralph, Boson Sampling from a Gaussian State, Phys. Rev. Lett. **113**, 100502 (2014).