This is the accepted manuscript made available via CHORUS. The article has been published as:

Distillation of Indistinguishable Photons

Jeffrey Marshall

Phys. Rev. Lett. **129**, 213601 — Published 16 November 2022

DOI: 10.1103/PhysRevLett.129.213601

Distillation of Indistinguishable Photons

Jeffrey Marshall[1, 2, *]

[1] QuAIL, NASA Ames Research Center, Moffett Field, CA 94035, USA

[2] USRA Research Institute for Advanced Computer Science, Mountain View, CA 94043, USA

A reliable source of identical (indistinguishable) photons is a prerequisite for exploiting interference effects, which is a necessary component for linear optical based quantum computing, and applications thereof such as Boson sampling. Generally speaking, the degree of distinguishability will determine the efficacy of the particular approach, for example by limiting the fidelity of constructed resource states, or reducing the complexity of an optical circuits output distribution. It is therefore of great practical relevance to engineer heralded sources of highly pure and indistinguishable photons. Inspired by magic state distillation, we present a protocol using standard linear optics which can be used to increase the indistinguishability of a photon source, to arbitrary accuracy. In particular, in the asymptotic limit of small error ϵ, to reduce the error to $\epsilon' < \epsilon$ requires $O((\epsilon/\epsilon')^2)$ photons. We demonstrate the scheme is robust to detection and control errors in the optical components, and discuss the effect of other error sources.

Introduction– Linear optical quantum computing (LOQC) is an attractive paradigm for realizing fault-tolerance, since photons in free space have extremely long coherence times, and can be manipulated via high fidelity linear optics which may not require the same level of cooling as other approaches [1]. In LOQC, qubits are constructed out of photons which can exist in two modes, common choices being spatial modes, or using the polarization degrees of freedom. Fault tolerance can in principle be achieved via the KLM protocol with sufficient numbers of qubits and using error correction [2], or using cluster states in a measurement-based approach to quantum computing [1, 3–8].

In order to make use of photons for computational purposes requires a source of highly indistinguishable photons. The Hong-Ou-Mandel (HOM) effect [9] is the prototypical example which shows fundamental differences in which identical versus distinguishable photons interfere (or do not). In this conceptually simple experiment, two photons are incident upon a 50:50 beamsplitter, which results in a bunching of the two photons in the case they are indistinguishable. On the other hand, when the input photons are distinguishable, the signal from an HOM experiment (the HOM 'dip') is diminished by an amount related to the infidelity of the two photons [10].

The HOM effect is a crucial ingredient for realizing LOQC, for the interference between identical photons can be used to create entanglement over computational degrees of freedom [2, 11–13]. For example, fusion measurements can be used to create large cluster states out of primitive entangled states, such as Bell states or small GHZ states [14]. However, the presence of distinguishability will generally result in less entanglement generated over the computational degrees of freedom, compared to the ideal state [15, 16].

Similarly, for specific applications of LOQC, such as Boson sampling [17], multi-photon interference is the key ingredient to generate a computationally intractable distribution, which is reduced in complexity with distinguishability [18].

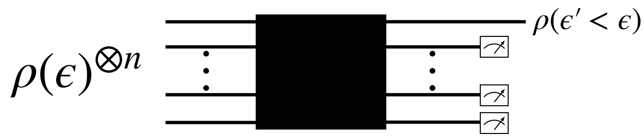

Figure 1. **Cartoon schematic of distillation scheme.** n copies of a noisy photon state with error rate ϵ (Eq. (3)), incident upon n spatial rails, are used to distill a single photon of lower error $\epsilon' < \epsilon$. This is achieved upon post-selection of a particular detection pattern of $n-1$ photons in the measured rails. The black box is at this point unspecified but will be an array of beamsplitters between the rails to enact interference.

It is therefore necessary to be able to generate photons with as high an overlap as possible. In this Letter, we present a technique inspired by magic state distillation [19], which is used to 'distill' indistinguishable photons from a photon source which outputs photons that are partly distinguishable (or in other words, a source with non-unit purity). This task can be phrased in a few equivalent ways, and is related to state purification [20, 21] and discrimination [22].

Commonly narrowband filters are used to generate heralded highly pure photons from pair sources, however in practice the photon yield becomes prohibitively small at high enough target purity [23]. Moreover, naive filtering of a single photon source, whilst yielding highly pure photons, will be unheralded. Our scheme instead works under a different paradigm, where independent single photons that are partly distinguishable are used to produce a source of heralded and pure photons, utilizing multi-photon interference.

A cartoon example of our general idea is shown in Fig. 1, whereby n copies of a noisy photon state are used to produce single photons, with a lower degree of distinguishability. Input photons to the circuit populate spatial modes (horizontal lines), which we will often refer to

* jmarshall@usra.edu

as 'rails', and can be implemented physically via optical fibres, for example. The black box is a circuit composed of beamsplitters (and possibly other linear optical components), and the output photon is conditioned on the post-selection of a particular measurement outcome, i.e. the detection of $n-1$ photons, in some configuration. The key observation behind our scheme is that identical photons interfere in a fundamentally different manner than partly distinguishable ones, which can be exploited using beamsplitters, and ultimately used to reduce the distinguishability of noisy photon sources. The scheme works so long as the initial purity is above around 60%.

Related Work– Whilst preparing this manuscript, we became aware of a morally similar scheme proposed by Sparrow and Birchall (SB) in Ref. [15], under the name 'HOM filtering'. In this scheme, $n \geq 2$ photons are incident upon n rails, which are post-selected upon bunching in a single rail. Photon subtraction is then used to output a single photon of a higher fidelity. This scheme is conceptually elegant, and results in asymptotic scaling of the error $\epsilon \to \epsilon/n$. However, it is apparent that the scheme becomes prohibitive for even modest n, as the probability to measure the desired outcome falls worse than exponentially in n [[**24**]]; we compute in Supplemental Information, SI A, the post-selection success probability to be asymptotically (i.e. at error approaching zero)

$$P_{p.s.}^{(SB)} \leq \frac{n}{2^n}\prod_{m=2}^{n}\frac{m}{2^m} = \frac{n^2(n-1)!}{\sqrt{2^{n^2+3n-2}}}, \tag{1}$$

meaning huge numbers of photons are required to distill a single purer one (for $n = 2, 3, 4, 5, 6$, one requires on average 8, 42, 341, 4369, 93206 photons respectively).

Our scheme overcomes two issues identified by SB in their protocol, namely we achieve higher success probabilities (and therefore use fewer photons), and do not require explicit multiple photon subtraction [[**25**]]. Eventually we believe a hybrid scheme can be invoked, as in regimes of higher error, the SB scheme can outperform the present approach, whereas at lower errors, our scheme is most efficient. We will discuss this in the Results section.

Theory– An arbitrary single photon state can be written as a sum over modes [26, 27]:

$$|\psi\rangle = \sum_{s\in\{h,v\}}\int d\omega\, c_{s,\omega}|s,\omega\rangle = \sum_{i=0}^{\infty} c_i\hat{a}_i^\dagger|\mathbf{0}\rangle = \sum_{i=0}^{\infty} c_i|\psi_i\rangle. \tag{2}$$

The term after the first equals sign represents the explicit representation over the polarization (s being e.g. horizontal h or vertical v) and frequency (ω) domains, and going to the second equals sign we have picked a countable orthonormal basis in the separable Hilbert space to represent the continuous degrees of freedom (and absorbed the s index into the new sum). The state $|\mathbf{0}\rangle$ is the vacuum state, and $\hat{a}_i^\dagger$ creates a photon in the i'th mode, where for now we use the explicit state representation $\hat{a}_i^\dagger|\mathbf{0}\rangle = |\psi_i\rangle$. By construction, these basis states are orthogonal $\langle\psi_i|\psi_j\rangle = \delta_{ij}$, and the amplitudes $c_i \in \mathbb{C}$ square sum to 1: $\sum_i |c_i|^2 = 1$.

We now describe the model of a noisy photon source which is used in this work. A non-ideal photon source will output photons according to Eq. (2), but with realization dependent coefficients c_i (that is, they are different for each generated photon). Without loss of generality we can pick the basis so that the desired mode to populate is the 0'th one, i.e. $|\psi_0\rangle$ is the state which would be generated each time by a perfect photon source. We consider fluctuations around this ideal by assuming the source can generate photons in the 0'th mode with probability $1-\epsilon$, i.e. $\langle|c_0|^2\rangle = 1-\epsilon$, where the angle brackets indicate the realization average. We will similarly define $p_i := \langle|c_i|^2\rangle$, where $\sum_{i>0} p_i = \epsilon$. We further make a random phase approximation so that $\langle c_i c_j^*\rangle = 0$ for $i \neq j$, which means the photon source can be equivalently described as a dephased mixture:

$$\rho(\epsilon) = (1-\epsilon)\,|\psi_0\rangle\langle\psi_0| + \sum_{i>0} p_i\,|\psi_i\rangle\langle\psi_i|. \tag{3}$$

This approximation amounts to the 'error amplitudes' $c_{j>0} = |c_j|e^{i\phi_j}$ receiving a random phase ϕ_j (independent of the norm) on each realization. With this, we can therefore interpret the photon source as generating a photon in the ideal state $|\psi_0\rangle$ with probability $1-\epsilon$, or with probability ϵ an orthogonal 'error mode' is populated (i.e. from one of the $\hat{a}_{i>0}^\dagger$). We will similarly call the $|\psi_{i>0}\rangle$ as an 'error state' (orthogonal to $|\psi_0\rangle$).

We define the indistinguishability within our model as the mean overlap of pure states generated by the source, i.e. $\mathcal{I} := \text{mean}(|\langle\phi|\psi\rangle|)$. Under our assumptions, this is equivalent to sampling pure states from ρ, from which it is easy to show $\mathcal{I} = \text{tr}(\rho^2)$, i.e. it is the purity. The aim of this work is to maximise the indistinguishability/purity by minimizing ϵ.

To simplify the analysis, we can consider the small error (small ϵ) limit. At sufficiently small ϵ it is unlikely to observe more than one error state according to the above statistical description; if we draw n samples from distribution ρ [[**28**]], we either get n copies of $|\psi_0\rangle$, or $n-1$ copies of $|\psi_0\rangle$, and one copy of some orthogonal error state $|\psi^\perp\rangle$ (i.e. $|\psi^\perp\rangle$ is one of the $|\psi_{i>0}\rangle$). Note, in our subsequent analysis we will still take into account the cases when more than one error mode is populated, but for now we can work in the limit of only single errors, for convenience. We can write the n photon state, to first order as (see SI B)

$$\rho^{\otimes n} = (1-\epsilon)^n|\Psi_0\rangle\langle\Psi_0| + \epsilon(1-\epsilon)^{n-1}\sum_{k=1}^{n}|\Psi_k\rangle\langle\Psi_k| + O(\epsilon^2), \tag{4}$$

where we have introduced notation $|\Psi_0\rangle = |\psi_0\rangle^{\otimes n}$ and $|\Psi_k\rangle = |\psi_0\rangle^{\otimes(k-1)}|\psi^\perp\rangle|\psi_0\rangle^{\otimes(n-k)}$. The error term $O(\epsilon^2)$ contains the states of n photons composed of $n-2$ copies of $|\psi_0\rangle$, and two error states $|\psi_{i>0}\rangle$. The tensor structure comes from the spatial mode representation, as in Fig. 1. For now we write the error state generically as $|\psi^{\perp}\rangle$, as we will later see at first order it is unimportant for our analysis which particular error mode $i > 0$ is populated in state $|\Psi_k\rangle$.

In order to enact interference between photons of the above form, we will utilise a beamsplitter. In our notation a beamsplitter is described by 4 parameters, and acts on (spatial) mode creation operators $\hat{a}^{\dagger}, \hat{b}^{\dagger}$ as follows:

$$\begin{aligned} \hat{a}^{\dagger} &\rightarrow e^{i(\phi_0+\phi_R)}\sin(\theta)\hat{a}^{\dagger} + e^{i(\phi_0+\phi_T)}\cos(\theta)\hat{b}^{\dagger} \\ \hat{b}^{\dagger} &\rightarrow e^{i(\phi_0-\phi_T)}\cos(\theta)\hat{a}^{\dagger} - e^{i(\phi_0-\phi_R)}\sin(\theta)\hat{b}^{\dagger}. \end{aligned} \tag{5}$$

We assume the parameters $\{\theta, \phi_{0,R,T}\}$ are agnostic to the impinging photons internal state [29], and therefore any single photon incident upon such a beamsplitter will be split in the same manner as any other. A 50:50 beamsplitter refers to the case $\theta = \pi/4$, where there is equal transmission to the other mode (T), or reflection to the same mode (R). Throughout we use the the convention for the phases $\phi_0 = \pi/2, \phi_R = -\pi/2, \phi_T = 0$.

Since we utilise optical components that are state-agnostic, and any single photon in state $|\psi_{i>0}\rangle$ will not interfere with the ideal state $|\psi_0\rangle$ (by orthogonality), it has no bearing on the output statistics of a circuit of form Fig. 1 which particular error mode $i > 0$ is actually populated when state $|\Psi_k\rangle$ is sampled from $\rho^{\otimes n}$. For this reason we can write the single error state simply as $|\psi^{\perp}\rangle$, as mentioned above.

Now that we have described the basic components in our construction, all that remains is to outline the post-selection over detection events. We will require access to photon number resolving detectors which we assume are ideal; it will always detect the exact number of photons present (though it will in fact be enough to distinguish between 0,1,2,3 photons, which will be clear later). The post-selection on a detection event of m photons can be described by taking the partial trace of the measured rail(s) after applying a measurement operator on the state [15, 30]. If before measurement the state is ρ, and we place a detector at the k'th rail to detect m photons, the post-selected state will be $\mathrm{Tr}_k[\Pi_k^{(m)}\rho\Pi_k^{(m)}]/N$, where $\Pi_k^{(m)}$ sums over all rank 1 projectors onto pure states which contain m photons in the k'th rail. N is for normalization.

Results– The central question we wish to answer is whether one can engineer the schematic diagram Fig. 1 with a suitable number n of photons, and linear optical components in the black box, so that the output state has less error than Eq. (3), upon a suitable post-selection. If one can do this, the process can be repeated indefinitely until arbitrary accuracy (i.e. ϵ is arbitrarily small).

From our studies, this in fact defines a large class of optical circuit of varying numbers of photons and linear optical components. We however will focus our attention on the 'best' performing that we found (where here best has a precise meaning, in terms of the number of photons required to distill a photon to some particular accuracy). Indeed, there is scope for the discovery of improved circuits. We will assume all components and detectors are perfect, so that the only source of error is in the photon generator, but discuss such errors in the SI.

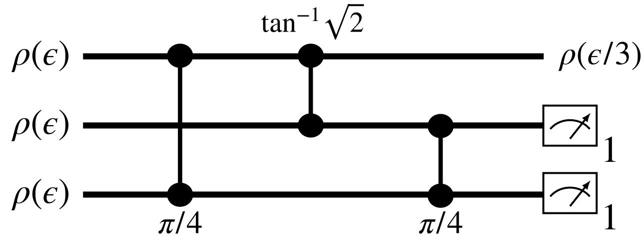

Figure 2. **Three photon distillation scheme.** A successful measurement corresponds to a single photon registered in each of the two measured rails (indicated by the '1' subscript on the detectors). The vertical lines with black circles represent beamsplitters between the rails on which the black circles intersect. The first and third beamsplitters are 50:50 ($\pi/4$ in the diagram), and the middle is asymmetric with $\theta = \tan^{-1}\sqrt{2} \approx 0.955$ (less likely to transmit). In the asymptotic limit of small ϵ, the error is reduced by a factor of 1/3, and post-selection succeeds with probability 1/3.

The circuit of present interest is shown in Fig. 2, composed of three rails (each taking one incident photon), and three beamsplitters, two which are symmetric, and one which is asymmetric, biased to higher reflectivity (to stay in same mode). Note permutations of this circuit also perform identically (keeping the angle of the middle beamsplitter $\tan^{-1}\sqrt{2}$).

First let us consider the ideal input of three identical photons in state $|\psi_0\rangle$ sampled from ρ, which we will denote using occupation number (Fock) representation over the rails as $|1,1,1\rangle$. This input occurs with probability $(1-\epsilon)^3$. The output of the circuit, before measurement is (up to a global phase)

$$\frac{1}{\sqrt{3}}|1,1,1\rangle - \frac{\sqrt{2}}{3}(i|3,0,0\rangle - |0,3,0\rangle + i|0,0,3\rangle), \tag{6}$$

which has probability of 1/3 to obtain the correct post-selected state [[**31**]].

If on the other hand a single error state is present, i.e. one of $\{|\Psi_k\rangle\}_{k=1}^3$ is sampled (each occurring with probability $\epsilon(1-\epsilon)^2$), the output in the relevant subspace before measurement, is $\frac{1}{\sqrt{27}}\sum_{k=1}^3|\Psi_k\rangle$, up to a phase. The post-selection therefore succeeds with total probability 1/9, and the outputted (unmeasured) photon is ideal $|\psi_0\rangle$ with post-selected probability 2/3 (see SI C for more information).

The key observation behind the scheme is that the ideal input is successfully post-selected upon three times as often than the case where an error is present (1/3 Vs 1/9), which allows the errors to be filtered out, approximately at a rate of 1/3 error reduction per round.

One can produce an upper bound on the error reduction (see SI C), $\epsilon \to \epsilon'$ under the scheme:

$$\epsilon' \leq \frac{\epsilon}{3}\frac{1+2\epsilon}{1-2\epsilon+3\epsilon^2-\epsilon^3} = \frac{\epsilon}{3} + \frac{4\epsilon^2}{3} + O(\epsilon^3). \quad (7)$$

The reason this is a bound, instead of equality, is that the error reduction depends on the specifics of the distribution of errors in Eq. (3). In SI C we also produce a lower bound on the error, $\epsilon' \geq \frac{\epsilon}{3} + \frac{2\epsilon^2}{3} + O(\epsilon^3)$. The scheme can be used to reduce errors ($\epsilon' < \epsilon$) so long as the initial error ϵ is below around 43%.

The error reduction capabilities of our scheme is shown in Fig. 3, where we also compare to the SB protocol for $n = 2$ which as we will see is the most efficient SB protocol, and $n = 3$ (same number of photons per round as the present approach). We see our scheme outperforms SB for $n = 2$ for errors less than around 15%, and that our scheme converges with SB $n = 3$ at around 5% error. Note, for the SB scheme we plot the best case error reduction, whereas in reality it may perform worse than this, depending on the distribution of error modes, see Ref. [15] (though for small ϵ the difference becomes negligible).

In SI C we compute the probability of obtaining a valid post-selection measurement outcome (i.e. detection of a single photon at each of the two detectors), which scales as $(1-2\epsilon)/3 + O(\epsilon)^2$. Fig. 2 of SI C compares this to the SB $n = 2, 3$ protocols which have a lower post-selection probability, leading to a greater resource requirement. Since our scheme consumes 3 photons per use, we require around 9 photons to distill a single purer one to 1/3 the error. In comparison to SB for $n = 2, 3$, around 8 and 42 photons are required respectively to obtain 1/2, 1/3 error respectively. In the asymptotic error limit (which practically is for $\epsilon \lesssim 0.05$), one can compute the number of photons required to distill a photon to target error ϵ' as $O((\frac{\epsilon}{\epsilon'})^2)$ [[32]]. In comparison to SB $n = 2, 3, 4$, the exponent is 3, 3.4, 4.2 respectively. This implies in the asymptotic limit our scheme is the most efficient.

Lastly, we wish to mention we also discovered an $n = 4$ photon circuit (see SI D), which is essentially a generalization of the presented $n = 3$ circuit (though with only 50:50 beamsplitters), which can reduce errors by $\epsilon/4$, at the expense of a lower success probability – asymptotically 1/4 – meaning around 16 photons are required on each iteration, and still $O((\frac{\epsilon}{\epsilon'})^2)$ photons to distill to error ϵ'.

Discussion– We briefly comment here that the scheme has some attractive properties for experimental implementation, which is discussed in more detail in SI E. In particular, there is a natural robustness to detection errors, as well as control errors. We also mention the protocol can also be trivially implemented in the case where the individual photons come from different physical sources [[28]]. For example, single photons of modest purity and reasonably high production rate could be generated from heralded filtered SPDC pairs [33], and then boosted to a high target fidelity via distillation, which crucially, are still heralded.

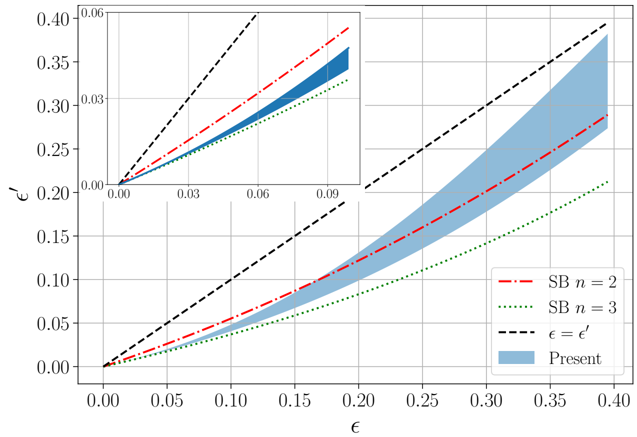

Figure 3. **Error reduction comparison of our scheme ('Present'), and those of SB for $n = 2, 3$.** Given a photon source with error ϵ (Eq. (3)), the post-selected output has error ϵ' (non-asymptotic Eq. (7) used for the upper-bound). The shaded region indicates the upper and lower bound on the error reduction of our scheme, as discussed in the main text (and SI C). For SB, we use the best case error reduction, Eq. (7.11) in Ref. [15] (also see Eq. (C6) in SI C). Inset: Zoom in on region $\epsilon < 0.1$.

Overall, in realistic scenarios, various errors will limit the upper bound on the indistinguishability that can be reached by our scheme, and a natural follow up can investigate robustness to these in practical settings. Additionally, the techniques presented here, we believe, have a diverse range of application, and can be utilized directly in resource state generation to construct circuits that are naturally resilient to distinguishability and loss errors, using similar mechanisms.

Acknowledgements– We thank Joseph Altepeter, Ryan Bennink, Patrick Birchall, Warren Grice, and Raphael Pooser for helpful discussions on the scheme presented, and for providing references to related literature.

We are also extremely grateful to fellow QuAIL team members Eleanor G. Rieffel, Shon Grabbe, Zhihui Wang and Salvatore Mandrá for support and various discussions on this work.

We are thankful for support from NASA Academic Mission Services, Contract No. NNA16BD14C.

We acknowledge support from DARPA, under DARPA-NASA IAA 8839, annex 129.

[1] P. Kok, W. J. Munro, K. Nemoto, T. C. Ralph, J. P. Dowling, and G. J. Milburn, "Linear optical quantum computing with photonic qubits," Rev. Mod. Phys. **79**,

135 (2007).

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

[3] R. Raussendorf and H. J. Briegel, "A One-Way Quantum Computer," Phys. Rev. Lett. **86**, 5188 (2001).

[4] N. Yoran and B. Reznik, "Deterministic Linear Optics Quantum Computation with Single Photon Qubits," Phys. Rev. Lett. **91**, 037903 (2003).

[5] M. A. Nielsen, "Optical Quantum Computation Using Cluster States," Phys. Rev. Lett. **93**, 040503 (2004).

[6] R. Raussendorf, J. Harrington, and K. Goyal, "A fault-tolerant one-way quantum computer," Annals of Physics **321**, 2242 (2006).

[7] H. J. Briegel, D. E. Browne, W. Dr, R. Raussendorf, and M. Van den Nest, "Measurement-based quantum computation," Nat. Phys. **5**, 19 (2009).

[8] S. Bartolucci, P. Birchall, H. Bombin, H. Cable, C. Dawson, M. Gimeno-Segovia, E. Johnston, K. Kieling, N. Nickerson, M. Pant, F. Pastawski, T. Rudolph, and C. Sparrow, "Fusion-based quantum computation," arXiv:2101.09310 (2021).

[9] C. K. Hong, Z. Y. Ou, and L. Mandel, "Measurement of subpicosecond time intervals between two photons by interference," Phys. Rev. Lett. **59**, 2044 (1987).

[10] A. M. Braczyk, "Hong-Ou-Mandel Interference," arXiv:1711.00080 (2017).

[11] M. Varnava, D. E. Browne, and T. Rudolph, "How Good Must Single Photon Sources and Detectors Be for Efficient Linear Optical Quantum Computation?" Phys. Rev. Lett. **100**, 060502 (2008).

[12] S. Bartolucci, P. M. Birchall, M. Gimeno-Segovia, E. Johnston, K. Kieling, M. Pant, T. Rudolph, J. Smith, C. Sparrow, and M. D. Vidrighin, "Creation of Entangled Photonic States Using Linear Optics," arXiv:2106.13825 (2021).

[13] T. C. Ralph and G. J. Pryde, "Chapter 4 - optical quantum computation," (Elsevier, 2010) pp. 209–269.

[14] D. E. Browne and T. Rudolph, "Resource-Efficient Linear Optical Quantum Computation," Phys. Rev. Lett. **95**, 010501 (2005).

[15] C. Sparrow, *Quantum interference in universal linear optical devices for quantum computation and simulation*, Ph.D. thesis, Imperial College London (2017).

[16] P. Birchall, *Fundamental advantages and practicalities of quantum-photonic metrology and computing*, Ph.D. thesis, University of Bristol (2018).

[17] S. Aaronson and A. Arkhipov, "The Computational Complexity of Linear Optics," in *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing*, STOC '11 (Association for Computing Machinery, New York, NY, USA, 2011) p. 333342.

[18] J. J. Renema, A. Menssen, W. R. Clements, G. Triginer, W. S. Kolthammer, and I. A. Walmsley, "Efficient Classical Algorithm for Boson Sampling with Partially Distinguishable Photons," Phys. Rev. Lett. **120**, 220502 (2018).

[19] S. Bravyi and A. Kitaev, "Universal quantum computation with ideal clifford gates and noisy ancillas," Phys. Rev. A **71**, 022316 (2005).

[20] J. I. Cirac, A. K. Ekert, and C. Macchiavello, "Optimal Purification of Single Qubits," Phys. Rev. Lett. **82**, 4344–4347 (1999).

[21] M. Ricci, F. De Martini, N. J. Cerf, R. Filip, J. Fiurášek, and C. Macchiavello, "Experimental Purification of Single Qubits," Phys. Rev. Lett. **93**, 170501 (2004).

[22] S. Stanisic and P. S. Turner, "Discriminating distinguishability," Phys. Rev. A **98**, 043839 (2018).

[23] P. J. Mosley, J. S. Lundeen, B. J. Smith, P. Wasylczyk, A. B. U'Ren, C. Silberhorn, and I. A. Walmsley, "Heralded Generation of Ultrafast Single Photons in Pure Quantum States," Phys. Rev. Lett. **100**, 133601 (2008).

[24] Numerically we find Eq. (1) can be approximated by $\exp(-0.29n^{1.96})$, see Fig. 1 of SI A.

[25] In principle photon bleeding can be invoked to achieve success probability approaching 1 for the photon subtraction phase [[12]] in SB, at the cost of additional circuit depth. The feasibility of this however depends on loss and detection errors which could accumulate.

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

[27] T. Opatrný, N. Korolkova, and G. Leuchs, "Mode structure and photon number correlations in squeezed quantum pulses," Phys. Rev. A **66**, 053813 (2002).

[28] Note that the presented protocol does not rely on the fact the photons come from the same source. The only requirement is that the density matrix description of the sources are of the general form Eq. (3) (where each source can have distinct probabilities ϵ, p_i, so long as $|\psi_0\rangle$ is the dominant mode).

[29] D. Gonzlez-Andrade, C. Lafforgue, E. Durn-Valdeiglesias, X. Le Roux, M. Berciano, E. Cassan, D. Marris-Morini, A. V. Velasco, P. Cheben, L. Vivien, and C. Alonso-Ramos, "Polarization- and wavelength-agnostic nanophotonic beam splitter," Sci. Rep. **9**, 3604 (2019).

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

[31] Note this scheme can also be used to prepare the superposition $|3,0\rangle + |0,3\rangle$, upon measuring zero photons in the middle rail of Eq. (6), which occurs with probability 4/9, which can be considered a 3 photon generalization of the standard HOM state.

[32] The calculation of $O((\epsilon/\epsilon')^2)$ photons comes simply by noting each iteration of the scheme requires asymptotically 9 photons, to reduce the error by 1/3. If we wish to obtain error ϵ', we require r iterations where $\epsilon' = \epsilon/3^r$, which consumes $9^r = (\epsilon/\epsilon')^2$ photons. The 'big O' notation captures the constant overhead when ϵ/ϵ' is not an exact power of 3.

[33] M. Massaro, E. Meyer-Scott, N. Montaut, H. Herrmann, and C. Silberhorn, "Improving SPDC single-photon sources via extended heralding and feed-forward control," New Journal of Physics **21**, 053038 (2019).