


General protocols for the efficient distillation of indistinguishable photons

Jason Saied^{1,*}, Jeffrey Marshall^{1,2}, Namit Anand^{1,3} and Eleanor G. Rieffel¹¹*Quantum Artificial Intelligence Laboratory (QuAIL), NASA Ames Research Center, Moffett Field, California 94035, USA*²*USRA Research Institute for Advanced Computer Science, Mountain View, California 94043, USA*³*KBR, Inc., 601 Jefferson St., Houston, Texas 77002, USA* (Received 13 May 2024; revised 6 December 2024; accepted 20 February 2025; published 26 March 2025)

We introduce state-of-the-art protocols to distill indistinguishable photons, reducing distinguishability error rates by a factor of n , while using a modest amount of resources scaling only linearly in n . Our resource requirements are significantly lower and the protocols have fewer hardware requirements than in previous works, making large-scale distillation experimentally feasible for the first time. This efficient reduction of distinguishability error rates has direct applications to fault-tolerant linear optical quantum computation, potentially leading to improved thresholds for photon loss errors and allowing smaller code distances, thus reducing overall resource costs. Our protocols are based on Fourier transforms on finite Abelian groups, special cases of which include the discrete Fourier transform and Hadamard matrices. This general perspective allows us to unify previous results on distillation protocols and introduce a large family of efficient schemes. We utilize the rich mathematical structure of Fourier transforms, including symmetries and related suppression laws, to quantify the performance of these distillation protocols both analytically and numerically. Finally, our work resolves an open question concerning suppression laws for the n -photon discrete Fourier transform: the suppression laws are exactly characterized by the well-known zero transmission law if and only if n is a prime power.

DOI: [10.1103/PhysRevApplied.23.034079](https://doi.org/10.1103/PhysRevApplied.23.034079)

I. INTRODUCTION

Highly pure and indistinguishable photons are a prerequisite for use in quantum information processing. The effect of distinguishability has been famously demonstrated by the Hong-Ou-Mandel (HOM) experiment, which shows fundamentally different statistics in the cases when photons are identical or not. The HOM effect (and its generalizations) is a crucial ingredient for realizing linear optical quantum computation (QC) [1], where the interference between identical photons can be used to create entanglement over computational degrees of freedom (aided by postselective measurements). For example, fusion measurements can be used to create large cluster states out of primitive entangled states, such as Bell states or small Greenberger-Horne-Zeilinger (GHZ) states [2]. These states can then be used to realize fault-tolerant quantum computation, in paradigms such as fusion-based QC [3]. However, the presence of distinguishability will generally result in less entanglement generated over the computational degrees of freedom [4,5], meaning their computational “resource” is reduced. In fact, the degree of distinguishability has been shown to reduce

the classical computational complexity (and thus potential quantum advantage) of boson sampling [6]. Distinguishability between photons can also be unheralded (since the correct number of photons are eventually detected), making such errors potentially harder to deal with in linear optical QC than photon loss, the dominant source of noise in linear optics, which is generally heralded. We sketch the potential impact of distinguishability errors on fusion-based QC in Sec. II C, where we argue that reducing distinguishability may greatly improve error thresholds for photon loss and reduce the overall resource requirements of the fault-tolerant quantum computation. In particular, due to the benefit of significantly reduced distinguishability error rates and the efficiency of the distillation protocols we present in this work, we argue in the Appendix that our protocols may be worthwhile as a complement (or cheaper alternative) to state-of-the-art single-photon sources to enable more scalable and resource-efficient fault-tolerant linear optical quantum computation.

The conventional approach to achieving highly pure photons relies upon spectral filtering. While spectral filtering can produce very uniform photons, it has drawbacks. First, such methods are typically unheralded, meaning if a single photon impinges upon a filter, it is unknown if it exits the filter. Second, as the target fidelity is increased, the probability a photon passes the filter decreases [7].

*Contact author: jason.saied@nasa.gov, jasonsaied@gmail.com

To overcome these issues, in 2017 a fully linear optical scheme (with Fock basis measurements) to distill indistinguishable photons from partly distinguishable ones was presented [4], which was subsequently improved upon in Ref. [8]. An experimental demonstration of the distillation scheme of Ref. [4] was recently given in Ref. [9].

Current experiments typically measure the degree of photon-photon distinguishability in linear optics via a HOM dip, which counts coincidence events, allowing one to compute the *visibility* $V = \text{Tr}[\rho_a \rho_b]$, where, e.g., “ a ”, “ b ” label two different photon sources. Values for the visibility vary quite a bit, depending on the type of single-photon source used, photon encoding, spectral filtering, etc., with an approximate range found in the literature to be $V \in [0.74, 0.99]$ [10–16]. In this work we characterize the error via a parameter ϵ we call the *distinguishability error rate*, discussed in Sec. II B. We have $\epsilon \approx 1 - \sqrt{V}$, which for reference gives ϵ approximately in the range $[0.005, 0.15]$, based on the quoted visibility values.

The general idea of these distillation protocols is shown in Fig. 1, whereby n (noisy) single photons are evolved under a linear optical unitary, with postselection performed by photon-number-resolving detectors (PNRDs). It can be shown that by carefully performing the postselection based on the interference generated by the unitary evolution, one can guarantee that the output photons will have a higher expected overlap with the target state. In this work we generalize the state-of-the-art three- and four-photon protocols of Ref. [8] to allow the scheme to work with any number n of photons (as in Fig. 1). We prove in Theorem III.2 that for input photons with distinguishability error rate ϵ , our n -photon distillation protocols herald the output of single photons with distinguishability error rate $e_n(\epsilon) \approx \epsilon/n$. In particular, large enough values of n lead to arbitrary reduction of the distinguishability error rate. The n -photon protocols use approximately $4n$ photons to obtain this reduction (see Theorem III.9). As discussed in more detail below and in Remark III.10, the previous state of the art [8] requires $O(n^2)$ photons to achieve the same amount of error reduction, implying (for example) that our protocols are over 20 times more efficient when $n = 81$. Further, in order to reduce the error rate by more than a factor of 4 (as in Marshall’s four-photon protocol [8]) *without* exponentially

scaling resource costs, the previous techniques require iteration of several smaller distillation protocols, along with active feedforward and quantum memory (e.g., optical delay lines) so that successfully distilled photons can be used in later rounds of distillation. The additional hardware requirements of these previous protocols may ultimately *increase* the error rate: for example, the delay lines may cause the photons to experience different amounts of dispersion. For the main protocols presented in this work, however, there is no need for iteration, feedforward, or quantum memory during distillation. Thus the alternative distillation protocols are significantly more resource efficient and require less complex hardware, making them far more feasible for experimental implementation.

We now briefly discuss our general framework for distillation and its utility. The n -photon protocols depend on an $n \times n$ unitary matrix U , determining an n -mode linear optical unitary. We consider protocols with U coming from the discrete Fourier transform F_n (for $n \geq 3$) and Hadamard matrices $H_n = H^{\otimes r}$, where H is the 2×2 Hadamard matrix and $n = 2^r \geq 4$. We note that the previous state-of-the-art three- and four-photon protocols of Ref. [8], corresponding to $U = F_3$ and $U = H_4$ respectively, were discovered numerically. Reference [8] gives rigorous proofs regarding the performance of these protocols, but these proofs are by direct computation: there was no theoretical understanding of why those unitaries would be particularly useful for distillation, nor any indication of how one might generalize them. In this work, we put both protocols into a larger theoretical context, generalizing them to the families F_n and H_n . We further unify these families by observing that both correspond to Fourier transforms on finite Abelian groups of order n (\mathbb{Z}_n for Fourier and $\mathbb{Z}_2^{\times r}$ for Hadamard), and in fact any such Fourier transform (with $n \geq 3$) leads to a similarly efficient distillation protocol. (The main text focuses on the F_n and H_n cases, with the generalization discussed in detail in Appendix C.) Questions about distillation then become questions about suppression laws for general Fourier transforms. This generality allows us to combine powerful results on suppression laws in the Fourier [17] and Hadamard [18] settings into a single framework and apply it to characterize the performance of our distillation protocols. In particular, as discussed further below, suppression laws are often determined by symmetry properties of the relevant unitary [19]; for the Fourier transform on the finite Abelian group G , the relevant group of symmetries is simply G . This is the perspective motivating the proofs and conjectures in this work. We also note that, due to the theoretical connections we establish between distillation and suppression laws, our results have consequences beyond distillation. In particular, we resolve an open problem due to the 2010 paper, Ref. [17], by proving precisely when the known suppression laws for F_n are both necessary and sufficient conditions for an output pattern to be



FIG. 1. Sketch of n -photon distillation scheme, where n photons are injected into a linear optical unitary U . PNRDs are used to postselect on certain $n - 1$ photon outcomes, resulting in a single photon out (in the absence of losses, dark counts, etc.). It is possible to engineer the unitary and postselection criteria so that the output single-photon error is reduced.

suppressed (see Theorem III.11). We return to this subject below.

As with the earlier works [4,8], the n -photon distillation protocols presented here are nondeterministic with some *heralding rate* $h_n(\epsilon) < 1$, the probability of heralding the output of a distilled single photon. Large heralding rates are desirable in order to minimize overhead. In Theorem III.4, we give a formula for the heralding rate $h_n(\epsilon)$ up to first order in ϵ , depending only on the heralding rate in the error-free case, $h_n(0)$. In Theorem III.5 and the following discussion, we give a formula for $h_n(0)$ in terms of hypergeometric functions and observe that it quickly approaches $1/4$, with $h_n(0) - \frac{1}{4} \sim 1/16n$. In particular, for small error rates ϵ (and $n \geq 4$), the n -photon distillation protocol succeeds with probability near $1/4$, and therefore we expect to use $4n$ input photons to obtain an output photon with reduced error rate approximately ϵ/n . This is a significant improvement over the protocol of Sparrow and Birchall [4], which requires exponentially many photons to achieve the same reduced error rate ϵ/n [8]. The protocols of Refs. [4,8] may be iterated, with successfully distilled photons stored and used as input to later rounds of distillation; this iterative approach allows for similar error rates with cubic or quadratic resources, respectively, but requires active feedforward and quantum memory (optical delay lines) during distillation. Our protocols, on the other hand, use only linearly scaling resources *without* requiring active feedforward or memory during distillation. In Sec. III C, we discuss the resource requirements of the protocols in Refs. [4,8] and the present work in more detail. Generally, we see that for small ϵ , our protocols achieve comparable reduction in error at significantly lower cost. We give a concrete example in Remark III.10: to reduce the error rate ϵ by a factor of $n = 81$, the Fourier protocols require 20 times fewer photons than the previous state-of-the-art protocol of Ref. [8]. In particular, the F_{81} protocol would require an average of only four runs using 81 photons each, with no interaction between subsequent runs. We note, however, that these resource estimates require ϵ to be sufficiently small. Each protocol has a *threshold* value of ϵ above which it fails to reduce the error rate; as n increases, this threshold decreases. (See Figure 3.) Further, even below threshold, it can be advantageous to concatenate multiple distillation protocols, beginning with smaller values of n and feeding the less distinguishable output photons into larger n protocols in subsequent stages. We discuss such iterated distillation schemes in Sec. III C.

For larger values of ϵ , the first-order approximations of $h_n(\epsilon)$ discussed above no longer suffice. In Sec. III D we conjecture a lower bound on $h_n(\epsilon)$ as a function of ϵ . This lower bound may be used to give explicit guarantees on the heralding rate for large n , as long as ϵ is sufficiently small relative to n . For example, if $\epsilon \leq 1/n$, we conjecture that the n -photon protocol's heralding rate is bounded below by $1/4e \approx 0.092$. These conjectures are

supported by numerical evidence and are motivated by certain symmetry properties of the distillation protocols, discussed in Sec. III D. As discussed above, these symmetry properties are related to the suppression laws of Refs. [18–20] that underlie our main results and motivate the use of the Fourier and Hadamard unitaries. These suppression laws have also been applied to the related problem of verifying boson sampling devices [21,22]. For the Hadamard case, the suppression is entirely governed by certain parity-preservation conditions determined by the symmetries [18,19]. For the Fourier transform, the corresponding *zero transmission law* (ZTL) was only known to determine some of the suppression, with cases (the smallest being F_6) in which there are more suppressed patterns than dictated by the ZTL [17]. In Theorem III.11, we prove that the suppressed patterns are exactly determined by the ZTL if and only if n is a prime power.

In Sec. III E, we give some numerics of interest. We provide detailed analysis of the case $n = 8$ as a representative example. This is the smallest protocol beyond Refs. [4,8] with n a power of 2, which means that we may consider both Fourier and Hadamard distillation protocols on the same number of photons and compare their performance. We also study the following problem: given an approximately known input error rate ϵ , which distillation protocol gives the greatest reduction in error if we are free to choose the value of n ? This problem is numerically answered for $n \leq 16$ in Figs. 6 and 7. We observe several interesting trends in these results, related to the symmetry properties of Sec. III D. In particular, the optimal protocols seem to be Fourier transform protocols in which there are *more* suppressed patterns than explained by the ZTL (i.e., by Theorem III.11, with n not a prime power).

In Sec. III F, we consider the effect of photon loss on our protocols. Since photon loss will be detected by a distillation protocol with high probability (especially for large n), we are most concerned with the effect on the heralding rates (and thus the resource requirements). We give numerics and prove lower bounds on the lossy heralding rates as a function of $h_n(\epsilon)$, thus obtaining upper bounds on the required resources. In the $\epsilon = 0$ limit, the lower bound is exactly attained.

We further demonstrate in Appendix D that the ability to distill photons is in fact typical, as shown by a randomized scheme based on Haar sampled unitaries. However, these versions are significantly less resource efficient (requiring a far greater number of photons) compared to those based on the Fourier or Hadamard matrices, due to a greater degree of constructive interference as a result of symmetries present in these matrices (as discussed in Sec. III D).

While readers familiar with linear optics should be able to skip most of the preliminaries in Sec. II, we recommend reviewing the notation used in Sec. II B, especially the error models and Eq. (15).

We also note that a paper with similar results on distillation protocols, restricting to the case F_n [23], was released simultaneously with the first version of the present work.

A. Summary of contributions

We now briefly summarize the main results of this work. For each positive integer $n \geq 3$, we introduce several n -photon distillation protocols of the form in Fig. 1, with the unitaries U corresponding to Fourier transforms on finite Abelian groups G of order n . We assume that we have distinguishability error rate ϵ (defined in Sec. II B below). We then prove the following results:

(1) (Theorem III.4) Upon successful heralding, the n -photon protocols output a single photon with reduced distinguishability error rate

$$e_n(\epsilon) = \epsilon/n + O(\epsilon^2).$$

(2) (Theorem III.2) The probability of successful heralding, called the *heralding rate* $h_n(\epsilon)$, may be expressed in terms of the error-free heralding rate $h_n(0)$ as follows:

$$h_n(\epsilon) = h_n(0) - (n-1)h_n(0)\epsilon + O(\epsilon^2).$$

For $n \geq 4$, we have $h_n(0) \approx 1/4$, and therefore for small ϵ , we obtain heralding rate

$$h_n(\epsilon) \approx \frac{1}{4} - \frac{n-1}{4}\epsilon.$$

(3) (Theorem III.9) The expected number of photons required to distill a single output photon using an n -photon protocol is $n/h_n(\epsilon)$. For $n \geq 4$ and small ϵ , the number of photons required is then approximately $4n$. This is in contrast to the previous state-of-the-art protocols [8], for which the cost is $O(n^2)$ (see Remark III.10).

(4) (Theorem III.17) If each beam splitter in the linear optical circuit has loss rate λ , then the lossy heralding rate is

$$h_n(\epsilon; \lambda) \geq (1 - \lambda)^{(n-1) \log n} h_n(\epsilon).$$

The lower bound becomes an equality when $\epsilon = 0$. In this setting, the number of photons required for successful distillation is, on average, approximately

$$\frac{4n}{(1 - \lambda)^{(n-1) \log n}}, \quad (1)$$

as shown in Eq. (46). For $\lambda = 0.01$, $n = 16$, distillation to reduce the error rate by a factor of 16 then requires an average of 7.3 runs of 16 photons at a time, with a total expected cost of approximately 117 photons. We note that this outperforms even the *lossless* version of the previous state-of-the-art protocols [8], which would require

approximately 256 photons to achieve a similar reduction in error.

(5) [Eq. (50)] Let $\Lambda := 1 - (1 - \lambda)^{\log n}$ be the average loss probability per photon. Given successful heralding, the probability of obtaining a single ideal photon in the output mode (i.e., no loss and no distinguishability errors) is

$$(1 - \Lambda)(1 - e_n(\epsilon)) + O(n\Lambda\epsilon),$$

where $e_n(\epsilon)$ is the error rate in the lossless case as above. In other words, up to *second-order* corrections involving both a photon loss and a distinguishability error, the output fidelity is simply the product of the output fidelity in the lossless case and the probability that the output photon is not lost.

(6) (Theorem III.11) We prove that for the discrete Fourier transform F_n , the suppression laws are precisely characterized by the zero transmission law of Ref. [17] if and only if n is a prime power. The connection between suppression laws and the performance of the corresponding distillation protocols is discussed in Sec. III D and Appendix A.

(7) (Sec. III E) Given a distinguishability error rate ϵ , we provide numerical simulations to study which distillation protocols (i.e., which choice of n and Abelian group of order n) lead to the smallest output error rates.

II. PRELIMINARIES

A. Linear optics

In what follows, we will consider the space of n photons (not necessarily indistinguishable) in m (external) modes. We begin with the *first quantization*, following the notation of Ref. [24]. In this framework, each photon has state space $\mathcal{H} = \mathcal{H}_{\text{ext}} \otimes \mathcal{H}_{\text{int}}$, where \mathcal{H}_{ext} corresponds to the “external” modes—those manipulated by the experiment—and \mathcal{H}_{int} corresponds to the “internal” modes—those not manipulated by the experiment. We identify $\mathcal{H}_{\text{ext}} = \mathbb{C}^m$ with basis $|0\rangle, \dots, |m-1\rangle$, where $|k\rangle$ describes a photon in mode k . The state space corresponding to n such photons is

$$\mathcal{H}^{\otimes n} = \mathcal{H}_{\text{ext}}^{\otimes n} \otimes \mathcal{H}_{\text{int}}^{\otimes n}. \quad (2)$$

Given a $|\psi\rangle \in \mathcal{H}^{\otimes n}$, it has a corresponding *external density matrix* in $\mathcal{H}_{\text{ext}}^{\otimes n}$ obtained by taking the partial trace over $\mathcal{H}_{\text{int}}^{\otimes n}$. For experiments involving only linear optics on the external modes, the behavior of the state is fully characterized by its external density matrix [24]. For our purposes, however, we will often need to consider both the external and internal degrees of freedom.

The standard basis for $\mathcal{H}_{\text{ext}}^{\otimes n}$ in the first quantization picture is $|m_1, \dots, m_n\rangle = |m_1\rangle \otimes \dots \otimes |m_n\rangle$, where the i th

photon is in mode $m_i \in \{0, \dots, m-1\}$. When it is necessary to clarify that a state is using first quantization notation, we will instead write $|m_1, \dots, m_n\rangle_{1Q}$.

In this work, we will consider many different representations of permutation groups on $\mathcal{H}^{\otimes n}$ and the related subspaces. In the *photon permutation representation*, the symmetric group S_n acts on $\mathcal{H}^{\otimes n}$ by permuting the tensor factors, thus permuting the particles. The group $U_{n,m}$ of *linear optical unitaries* is the group of unitary operators on $\mathcal{H}^{\otimes n} = \mathcal{H}_{\text{ext}}^{\otimes n} \otimes \mathcal{H}_{\text{int}}^{\otimes n}$ of the form $U^{\otimes n} \otimes I^{\otimes n}$, where $U \in U(m)$ is an m -mode unitary operator on \mathcal{H}_{ext} . In particular, linear optical unitaries operate only on the external degrees of freedom and are symmetric with respect to permutation of photons. We will often write \hat{U} to denote the linear optical unitary corresponding to $U \in U(m)$; when the context is clear, we often simply write U . We also extend the notation to arbitrary operators T on \mathcal{H}_{ext} (not necessarily unitary), writing \hat{T} for the operator $T^{\otimes n} \otimes I^{\otimes n}$ on $\mathcal{H}^{\otimes n} = \mathcal{H}_{\text{ext}}^{\otimes n} \otimes \mathcal{H}_{\text{int}}^{\otimes n}$.

We will also use the *second quantization* representation of photonic states. We begin with the case in which only external degrees of freedom are considered. In this setting, we use the *Fock basis* states $|s_0, \dots, s_{m-1}\rangle = |s_0, \dots, s_{m-1}\rangle_{2Q}$, which describe an m -mode state with s_0 photons in mode 0, s_1 photons in mode 1, and so on. For any mode i , we consider the *creation operator* acting on that mode,

$$a_i^\dagger = \sum_{n \geq 0} \sqrt{n} |n\rangle \langle n-1|_i. \quad (3)$$

We also have $a_i = (a_i^\dagger)^\dagger$, the corresponding annihilation operator. Representing the *vacuum* state by $|\vec{0}\rangle$, we may express all Fock basis states (up to normalization) by applying appropriate creation operators:

$$\begin{aligned} |s_0, \dots, s_{m-1}\rangle_{2Q} \\ = \frac{1}{\sqrt{s_0! s_1! \dots s_{m-1}!}} (a_0^\dagger)^{s_0} \dots (a_{m-1}^\dagger)^{s_{m-1}} |\vec{0}\rangle. \end{aligned} \quad (4)$$

Linear optical unitary evolution in the second quantization is characterized by the same $U \in U(m)$ matrix mentioned above, which independently evolves each creation operator: $a_j^\dagger \rightarrow \sum_{i=0}^{m-1} U_{ij} a_i^\dagger$. Note that the Fock basis states are symmetric with respect to permutation of photons: for example,

$$|1, 1\rangle_{2Q} = a_0^\dagger a_1^\dagger |\vec{0}\rangle = \frac{1}{\sqrt{2}} (|0, 1\rangle_{1Q} + |1, 0\rangle_{1Q}). \quad (5)$$

Thus the Fock basis is a basis for the *symmetric subspace* of $\mathcal{H}_{\text{ext}}^{\otimes n}$.

Remark II.1. We may convert between first and second quantization as follows. We let F be the symmetrization map from $\mathcal{H}_{\text{ext}}^{\otimes n}$ to its symmetric subspace, with (for example)

$$F(|0, 1\rangle_{1Q}) = \frac{1}{\sqrt{2}} (|0, 1\rangle_{1Q} + |1, 0\rangle_{1Q}) = |1, 1\rangle_{2Q}. \quad (6)$$

In particular, we have $F(|m_0, \dots, m_{n-1}\rangle_{1Q}) = |s_0, \dots, s_{m-1}\rangle_{2Q}$, where s_i is the number of photons in mode i (the number of indices j with $m_j = i$). The map F is not invertible, but given $|s_0, \dots, s_{m-1}\rangle_{2Q}$, there is a unique *weakly increasing* $|m_0, \dots, m_{n-1}\rangle_{1Q}$ with $F(|m_0, \dots, m_{n-1}\rangle_{1Q}) = |s_0, \dots, s_{m-1}\rangle_{2Q}$.

When there are internal degrees of freedom under consideration, we use the notation $a_i^\dagger[\xi_j]$ to indicate the creation of a particle with internal state $|\xi_j\rangle$ in external mode i . For a fixed orthonormal basis $\{|\xi_0\rangle, |\xi_1\rangle, \dots\}$ of \mathcal{H}_{int} , we will consider (appropriately normalized) states of the form

$$a_{m_0}^\dagger[\xi_{i_0}] \dots a_{m_{n-1}}^\dagger[\xi_{i_{n-1}}] |\vec{0}\rangle. \quad (7)$$

These states are symmetric under permutation of photons, as the notation $a_k^\dagger[\xi_{ij}]$ only indicates the mode and internal state, not the particular photon (tensor factor in $\mathcal{H}^{\otimes n}$); in fact, they form a basis for the photon-permutation-symmetric subspace of $\mathcal{H}^{\otimes n}$. This is in alignment with Ref. [24].

We will consider two main types of measurements. The first is *photon-number-resolving detection* (PNRD), formalizing the notion of a measurement that counts the number of photons in each mode. (We may also perform PNRD on a subset of modes in the obvious way.) This is a projective measurement on $\mathcal{H}_{\text{ext}}^{\otimes n}$, projecting onto the subspaces $V_{s_0, \dots, s_{m-1}}$ where

$$\begin{aligned} V_{s_0, \dots, s_{m-1}} \\ = \text{span}\{|m_0, \dots, m_{n-1}\rangle_{1Q} : F(|m_0, \dots, m_{n-1}\rangle_{1Q}) \\ = |s_0, \dots, s_{m-1}\rangle_{2Q}\} \end{aligned} \quad (8)$$

(recalling the map F of Remark II.1). This is just the space spanned by all states with s_0 photons in mode 0, s_1 photons in mode 1, and so on. We call (s_0, \dots, s_{m-1}) the obtained *measurement pattern*. For symmetric states in $\mathcal{H}_{\text{ext}}^{\otimes n}$, PNRD is simply projection onto the Fock basis; the above version allows for nonsymmetric states and internal degrees of freedom. Note that the internal degrees of freedom may be traced out for PNRD.

For certain calculations, we will be interested in measuring the internal modes as well. Specifically, we consider *internal-external measurement*, a projective measurement

with respect to the states (7) (appropriately normalized). We note, however, that this will only be used mathematically, to calculate output error rates of distillation protocols; the protocols themselves do not require any measurements beyond PNRD.

We now define perfect indistinguishability of photons, following Ref. [24]. Let ρ be a density matrix, and let ρ_{ext} be the corresponding external state, obtained by tracing out the internal Hilbert space. Let $P^{(n)}$ be the symmetrizer on $\mathcal{H}_{\text{ext}}^{\otimes n}$, projecting to states that are invariant under permutation of photons. Then ρ is *perfectly indistinguishable* if $P^{(n)}\rho_{\text{ext}} = \rho_{\text{ext}}$. The main examples in this paper correspond to states $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle = a_{m_0}^\dagger[\xi_0] \cdots a_{m_{n-1}}^\dagger[\xi_0]|\vec{0}\rangle$. More generally, any state $|\psi\rangle$ that can be expressed as a pure tensor $|\psi_{\text{ext}}\rangle \otimes |\psi_{\text{int}}\rangle$, with $|\psi_{\text{ext}}\rangle$ and $|\psi_{\text{int}}\rangle$ symmetric under permutation of photons, is perfectly indistinguishable. We discuss a nontrivial such example in Appendix B 3. On the other hand, any state of the form (7) involving two distinct (orthogonal) internal states $|\xi_i\rangle, |\xi_j\rangle$ is *not* perfectly indistinguishable.

B. Models for distinguishability

Following Refs. [4,8,25], we will consider variants of the *random source* (RS) model for photon distinguishability, in which photon sources independently produce photons with average internal state $\rho = \sum_{j \geq 0} p_j |\xi_j\rangle\langle\xi_j|$, where the $|\xi_j\rangle$ range over an orthonormal basis for \mathcal{H}_{int} . We fix this basis $\{|\xi_0\rangle, |\xi_1\rangle, \dots\}$ for \mathcal{H}_{int} going forward. Note that this model is equivalent, under the assumption that the internal degrees of freedom cannot be resolved or manipulated by the experiment, to a pure state description $\sum_j e^{i\phi_j} \sqrt{p_j} |\xi_j\rangle$ [4,25]. Without loss of generality, we assume that p_0 is the dominant eigenvalue and write $\epsilon = 1 - p_0 = \sum_{j \geq 1} p_j$, so that

$$\rho = \rho(\epsilon) = (1 - \epsilon) |\xi_0\rangle\langle\xi_0| + \epsilon \sum_{j \geq 1} p'_j |\xi_j\rangle\langle\xi_j| \quad (9)$$

in the RS model, with $p'_j = p_j/\epsilon$. We write $\rho = \rho(\epsilon)$ to emphasize the dependence on the parameter ϵ (suppressing the p_i from the notation). We view the state as a probabilistic mixture: with large probability $1 - \epsilon$ we obtain the “ideal” state $|\xi_0\rangle$, and with small probability ϵ we obtain one of the “distinguishable” error states $|\xi_j\rangle$, orthogonal to $|\xi_0\rangle$. In terms of creation operators, we may represent $\rho(\epsilon)$ by the appropriate probabilistic mixture of the creation operators $a^\dagger[\xi_j]$.

In this work, we often restrict our attention to the *uniform RS* (URS) model, in which we take $p'_1 = p'_2 = \dots = p'_R = 1/R$ and $p_j = 0$ for $j > R$, so that

$$\rho(\epsilon) = (1 - \epsilon) |\xi_0\rangle\langle\xi_0| + \frac{\epsilon}{R} \sum_{j=1}^R |\xi_j\rangle\langle\xi_j|. \quad (10)$$

For ease of exposition and simulation, we will often consider two extreme cases that are useful for restricting the possible internal states. First, we have the same bad bits (SBB) model [25], where we take $R = 1$ in the URS model, so that each photon has average internal state

$$\rho(\epsilon) = (1 - \epsilon) |\xi_0\rangle\langle\xi_0| + \epsilon |\xi_1\rangle\langle\xi_1|. \quad (11)$$

This model captures a situation with systematic errors in which photons are prone to exhibit the same type of distinguishability error. This may be viewed as the special case in which the internal state space \mathcal{H}_{int} is two dimensional. On the other hand, we consider the orthogonal bad bits (OBB) model, where the i th photon has internal state

$$\rho_i(\epsilon) = (1 - \epsilon) |\xi_0\rangle\langle\xi_0| + \epsilon |\xi_{i+1}\rangle\langle\xi_{i+1}|. \quad (12)$$

Recall that $\langle\xi_i|\xi_j\rangle = \delta_{ij}$. As discussed in Ref. [4], this corresponds to the limit of the URS model as $R \rightarrow \infty$. This model captures a situation in which error states are uniformly random with many degrees of freedom, so that it is vanishingly unlikely for the same error to occur twice in the same experiment. Since the OBB and SBB models may be viewed as opposite extremes of the URS model, we choose them as a particular focus for numerical simulation and examples.

Finally, we compare our parameter ϵ to the *visibility*. In the URS model, one can compute the visibility between two states drawn from the distribution as $V = \text{Tr}[\rho(\epsilon)^2] = (1 - \epsilon)^2 + \epsilon^2/R$. For the OBB limit, this gives visibility $V = (1 - \epsilon)^2$. In any case we have $V = 1 - 2\epsilon + O(\epsilon^2)$, so when V is large there is little difference between the corresponding values of ϵ in different error models. Using the above, visibility $V \geq 0.74$ (a rough lower bound on experimental visibility values in some recent literature [10–16]) approximately translates to $\epsilon \leq 0.15$. We are therefore primarily interested in effective distillation protocols for these relatively small values of ϵ .

1. Initial states

We briefly return to the general RS model. We will take $m = n$, considering experiments in which n photons are injected into n different (external) modes of an interferometer, with the photon in mode i having internal state $\rho_i(\epsilon)$. When $\epsilon = 0$ and internal degrees of freedom are neglected, this corresponds to the idealized Fock state $|1, \dots, 1\rangle_{2Q}$. In general, we call the initial state $\rho^{(n)} = \rho^{(n)}(\epsilon)$. We may decompose $\rho^{(n)}$ as follows:

$$\begin{aligned} \rho^{(n)} = & \sum_{j_0, \dots, j_{n-1}} (p_{j_0} \cdots p_{j_{n-1}}) a_0^\dagger[\xi_{j_0}] \cdots \\ & \times a_{n-1}^\dagger[\xi_{j_{n-1}}] |\vec{0}\rangle\langle\vec{0}| a_0[\xi_{j_0}] \cdots a_{n-1}[\xi_{j_{n-1}}]. \end{aligned} \quad (13)$$

This is a probabilistic mixture of many terms; by abuse of notation, we often write

$$a_0^\dagger[\xi_{j_0}] \cdots a_{n-1}^\dagger[\xi_{j_{n-1}}] |\vec{0}\rangle = |1, \dots, 1\rangle_{2Q} \otimes |\xi_{j_0}, \dots, \xi_{j_{n-1}}\rangle. \quad (14)$$

Note that in general, the left-hand side is not actually a pure tensor. For example, $a_0^\dagger[\xi_0]a_1^\dagger[\xi_1] |\vec{0}\rangle = |0, 1\rangle_{1Q} \otimes |\xi_0, \xi_1\rangle + |1, 0\rangle_{1Q} \otimes |\xi_1, \xi_0\rangle$ (up to normalization).

We say that a state (14) has k *distinguishability errors*, where k is the number of indices i with $j_i \neq 0$. We will often consider small values of ϵ , so that the expression is dominated by the terms that are first order or less in ϵ ; in other words, those with 0 or 1 distinguishability errors. We note that, for any RS model, the weight of the terms with exactly 1 distinguishability error is $n(1 - \epsilon)^{n-1} \sum_{j \geq 1} p_j = n\epsilon(1 - \epsilon)^{n-1}$. For the calculations of interest, we will see that these first-order terms exhibit essentially the same behavior regardless of the particular model.

In the URS model (or its OBB limit), we write the input state as

$$\rho^{(n)}(\epsilon) = \sum_{k=0}^n \binom{n}{k} \epsilon^k (1 - \epsilon)^{n-k} \Phi_k, \quad (15)$$

where Φ_k is the mixed state involving all terms with k distinguishability errors, normalized to have $\text{Tr}(\Phi_k) = 1$. Going forward, we will exclusively discuss the URS framework, but the results involving only terms with 0 or 1 distinguishability errors, such as Theorems III.2, III.4, III.5, III.9, III.12 in fact hold for any RS model.

C. Distillation and error correction

Before proceeding to the main results on distillation of distinguishability errors, we give a brief overview of a major motivation for this work, namely the potential relevance of distillation to photonic quantum error correction. As discussed later in this section, distinguishability errors can lead to logical errors in fault-tolerant quantum computation. As a result, reducing distinguishability error rates can lead to higher thresholds for other errors (such as photon loss), which may be harder to engineer away. This can therefore allow for smaller code distances and smaller system sizes. Of course, this could in principle be achieved by either distillation or source engineering. In practice however, engineering a source to achieve (for example) an order of magnitude reduction in error, historically, appears to be a challenging task. Distillation, on the other hand, is a general, tunable, and source-agnostic approach that can be used in a modular fashion as part of single photon generation. The protocols discussed in the present work are highly efficient (with an order of magnitude error reduction requiring around 40 photons) and can be implemented

with existing technologies [9], making them a promising supplement to source engineering. We therefore believe both source engineering and distillation to be key aspects to consider for realizing highly pure, indistinguishable and heralded single-photon sources.

For our discussion of quantum error correction, we consider *fusion-based quantum computation* (FBQC), a paradigm for universal fault-tolerant quantum computation starting from single-photon sources [3]. The details of FBQC are beyond the scope of this work, but the main idea is to use probabilistic entangling measurements and multiplexing to generate many copies of constant-sized entangled *resource states*, then carry out (destructive) probabilistic Bell measurements, called *Type-II fusions*, between appropriate pairs of qubits (photons). This is similar in spirit to other measurement-based approaches to quantum error correction, but with the state generation happening in a modular fashion to allow for multiplexing. We view Type-II fusion as measuring two observables, $X \otimes X$ and $Z \otimes Z$. One may show [4,26] that fusion between an ideal photon and another with internal state $\rho(\epsilon)$ results in flipping the value of the $X \otimes X$ observable with probability $\epsilon/2$. Then in FBQC, we may approximate distinguishability as leading to errors in fusion measurements. In particular, since the roles of the $X \otimes X$ and $Z \otimes Z$ observables are often swapped, we roughly approximate a distinguishability error rate ϵ by measurement errors with probability $\epsilon/4$.

To illustrate the effect of distinguishability errors in this setting, we consider the (already very small) error rate $\epsilon \approx 5 \times 10^{-3}$, roughly corresponding to the visibility achieved in Ref. [16]. This translates to an approximate measurement error rate of $p_m = \epsilon/4 \approx 1.25 \times 10^{-3}$. For simplicity, we consider the most basic FBQC architecture, using the 6-ring resource state to construct an analog of the surface code [3,27]. In that setting, our error rate p_m is below the measurement error threshold and yields a photon-loss threshold of approximately 6×10^{-3} , as seen in Fig. 9 of Ref. [28]. However, we will see that there is significant benefit to decreasing p_m further. As calculated in Sec. III C, applying our H_{16} distillation protocol reduces the distinguishability error rate to $\epsilon' \approx 3.6 \times 10^{-4}$, with a corresponding approximate FBQC measurement error rate of $p'_m \approx 8.9 \times 10^{-5}$ and photon-loss threshold very near the maximum threshold of 1.6×10^{-2} . Thus we nearly *triple* the photon-loss threshold by improving the distinguishability error rate even beyond the values obtained by state-of-the-art single-photon sources. Due to the high photon-loss rates in linear optics, increasing this threshold is a major priority.

We now consider the effect of reduced error rates on the code distance and resource requirements. Let us consider a simple (unrealistic) model with no photon loss, only measurement error due to distinguishability. The logical error rate p_L roughly scales as follows, where p_m is

the measurement error rate, p_{th} is the measurement error threshold, and d is the code distance [27,29,30]:

$$p_L \sim C \left(\frac{p_m}{p_{\text{th}}} \right)^{d/2}. \quad (16)$$

Taking $C = 1$ for convenience, we consider $p_m = 1.25 \times 10^{-3}$ as in the above example, and the corresponding lossless measurement error threshold $p_{\text{th}} = 2.1 \times 10^{-3}$ (from Fig. 9, Ref. [28]). This leads to

$$p_L \sim (0.77)^d. \quad (17)$$

If we target logical error rate $p_L \leq 10^{-6}$, this requires a code distance of $d \geq 53$. On the other hand, if we use H_{16} distillation as above to reduce the measurement error rate to $p'_m \approx 8.9 \times 10^{-5}$, we obtain

$$p'_L \sim (0.21)^d, \quad (18)$$

requiring a code distance of $d \geq 9$ to obtain the same desired logical error rate. Since this FBQC architecture is carried out in “logical blocks” of d^3 resource states [27], distillation would reduce the required number of resource states per logical block from 148 877 to 729. Of course, in practice, one will need d much larger than 9 in order to handle photon loss, which is the dominant source of noise. But as argued above, reduced distinguishability also leads to increased tolerance of photon loss. This in turn will reduce the required code distance via a similar argument.

In summary, we see that reducing distinguishability error rates far below the error threshold can improve thresholds for other types of errors, such as photon loss, and can greatly reduce the large-scale resource requirements of a fault-tolerant linear optical quantum computation. We may view distillation and source engineering as complementary means toward achieving this reduction. Since distillation requires only standard linear optical hardware, it is likely far cheaper to use distillation rather than purchasing or engineering a state-of-the-art photon source. Further, as seen in the example above, even the best known photon sources may not have low enough distinguishability error rates for efficient fault tolerance. In such a setting, one may apply distillation to the output of state-of-the-art single-photon sources to obtain even smaller distinguishability error rates and therefore more scalable and efficient linear optical quantum computation.

III. MAIN RESULTS

A. Distillation protocols

In this work, we will consider *distillation protocols* generalizing the work of Ref. [8], described in Protocol III.1 and illustrated in Fig. 1. Unless stated otherwise, we assume that the only errors are distinguishability errors,

and we model distinguishability with the URS error model (or the OBB limit) with error rate ϵ . (Photon loss will be addressed in Sec. III F.)

Consider an $n \times n$ matrix U . Define the set of *ideal patterns* corresponding to U to be the set of all (s_0, \dots, s_{n-1}) such that

$$s_0 = 1 \text{ and } \langle s_0, \dots, s_{n-1} | \hat{U} | 1, \dots, 1 \rangle \neq 0, \quad (19)$$

where both states are Fock basis states (with no internal degrees of freedom).

Protocol III.1. Given a number n of modes and an $n \times n$ unitary matrix U (corresponding to a linear optical unitary), we describe the corresponding n -photon *distillation protocol*.

(1) Input a state σ of n photons in n distinct modes. We will typically consider $\sigma = \rho^{(n)}(\epsilon)$.

(2) Apply \hat{U} , mapping $\sigma \mapsto \mu = \hat{U}\sigma\hat{U}^\dagger$.

(3) Perform PNRD on the final $n-1$ modes of μ , obtaining a *measurement pattern* (s_1, \dots, s_{n-1}) with s_i being the number of photons found in mode i . Label mode 0, the unmeasured mode, as the *output mode*; we call the single-mode postmeasurement state the *output state*. Letting $s_0 = n - \sum_{j \geq 1} s_j$, the output state has s_0 photons. We call (s_0, \dots, s_{n-1}) the corresponding *completed measurement pattern*.

(4) *Postselect* for ideal patterns, as defined in Eq. (19). In other words, if the pattern is not ideal, the photon is rejected.

The purpose of the distillation protocol is to herald the output of single photons with reduced distinguishability error rate, enabling more efficient linear optical quantum computation. By postselecting for patterns with $s_0 = 1$, we guarantee that the output state has a single photon. (Assuming distinguishability is the only error.) By postselecting for ideal patterns in particular, we *herald* the output of a single photon with (hopefully) a smaller rate of distinguishability error. This is simply an application of conditional probability: one needs to show that, given the knowledge that the completed measurement pattern is ideal, the output photon is more likely to be ideal.

Given such a distillation protocol with input state $\sigma = \rho^{(n)}(\epsilon)$, there will be two main metrics of interest, functions of the input error rate ϵ . First is the *heralding rate* $h_n(\epsilon)$, the probability of obtaining an ideal pattern during the measurement step. Second is the *output error rate* $e_n(\epsilon)$. If distinguishability is the only error, this is the probability that, given the heralding of an ideal pattern, the output photon has a “distinguishable” internal state $|\xi_i\rangle$, $i > 0$. (This is made rigorous via an internal-external measurement, as discussed in Sec. II A. We discuss the case in which there are loss errors in addition to distinguishability

in Sec. III F.) For an arbitrary density matrix σ , we write $h_n(\sigma)$, $e_n(\sigma)$ to be the corresponding heralding and output error rates for the distillation protocol with input state σ .

We will be particularly interested in protocols using the quantum Fourier transform and Hadamard matrices. We briefly discuss these matrices and the corresponding ideal patterns now: for a more detailed analysis, we refer to Appendix A. We also compare to the case of Haar random matrices in Appendix D.

The Hadamard matrices under consideration are defined for $n = 2^r$ by $H_n = H^{\otimes r}$, where H is the standard 2×2 Hadamard matrix. (We note that this family is typically called the family of *Sylvester matrices*.) A recursive linear optical circuit for H_n is given in Fig. 9. For H_n , we always take n to be a power of 2 with $n \geq 4$, even if not explicitly stated. The case $n = 4$ is equivalent to the four-photon distillation protocol in the Appendix of Ref. [8], where there is only one ideal pattern, $(1, 1, 1, 1)$. To describe the ideal patterns for H_n , we take a completed measurement pattern $p = (s_0, \dots, s_{n-1})$ and “convert to first quantization,” obtaining integers $0 \leq g_0 \leq g_1 \leq \dots \leq g_{n-1} \leq m - 1$ such that s_i is the number of indices j with $g_j = i$. For example, $p = (1, 2, 0, 1)$ corresponds to $g = (0, 1, 1, 3)$. With this notation, the ideal patterns for H_n are exactly those satisfying $s_0 = 1$ and $g_0 \oplus \dots \oplus g_{n-1} = 0$, where \oplus is binary XOR without carrying [18].

Letting $\omega = \exp(2\pi i/n)$, the n -mode Fourier transform matrix is given by

$$F_n = \frac{1}{\sqrt{n}}(\omega^{ij})_{0 \leq i, j \leq n-1}. \quad (20)$$

The distillation protocol for F_3 is equivalent to the three-photon protocol in the main text of Ref. [8], where there is only one ideal pattern, $(1, 1, 1)$. With the same “first

quantization” notation as above, the ideal patterns for F_n satisfy the *zero transmission law* (ZTL), $g_0 + \dots + g_{n-1} \equiv 0 \pmod n$ [17,19]. In Theorem III.11, we prove that whenever n is a prime power, then the ideal patterns are exactly characterized by the ZTL (and the condition $s_0 = 1$). Further, for n not a prime power, we show that the set of ideal patterns is a proper subset of those satisfying this condition (and $s_0 = 1$). We further discuss these properties in Secs. III D and Appendix A 2 below.

More generally, we also consider

$$F_{(n_1, \dots, n_\ell)} = F_{n_1} \otimes \dots \otimes F_{n_\ell}, \quad (21)$$

where $n = n_1 \dots n_\ell$ and the n_i are integers satisfying $n_i \geq 2$. Noting that $H = F_2$, we have $F_n = F_{(n)}$ and $H_n = F_{(2, 2, \dots, 2)}$. We discuss this case in far greater detail in Appendix C. Our discussion in the main text will focus primarily on the F_n ($n \geq 3$) and H_n ($n \geq 4$) cases, with mentions of the “Fourier” case generally referring to F_n unless stated otherwise. (Note, even if not explicitly stated, we will always consider $n > 2$ because $H_2 = F_2 = H$ has no ideal patterns, by the standard HOM effect.) We have the following result.

Theorem III.2. Consider an n -photon distillation protocol with unitary $U = F_{(n_1, \dots, n_\ell)}$, $n = n_1 \dots n_\ell > 2$, and input state $\rho^{(n)}(\epsilon)$. (Note this encompasses the cases F_n and H_n .) The output error rate satisfies

$$e_n(\epsilon) = \frac{\epsilon}{n} + O(\epsilon^2). \quad (22)$$

This is proven in Appendix B 3. As a consequence of the theorem, we see that for sufficiently small ϵ , $e_n(\epsilon) \approx \epsilon/n$. Then, when ϵ is small enough, the protocols with larger n are better at reducing distinguishability errors. This is

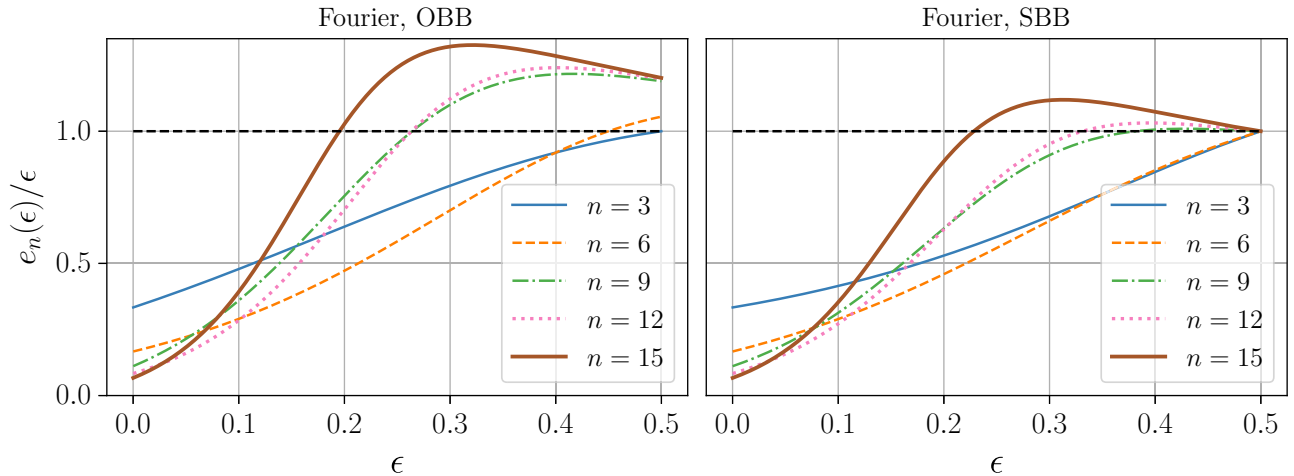


FIG. 2. Plot of the error rate reduction with OBB (left) and SBB (right) error models, for the Fourier (F_n) distillation protocol. For $\epsilon \rightarrow 0$, $e_n(\epsilon)/\epsilon \rightarrow 1/n$. Distillation is successful at reducing the error rate when the curve is less than 1, marked by the horizontal dashed line. (Also see Fig. 3.)

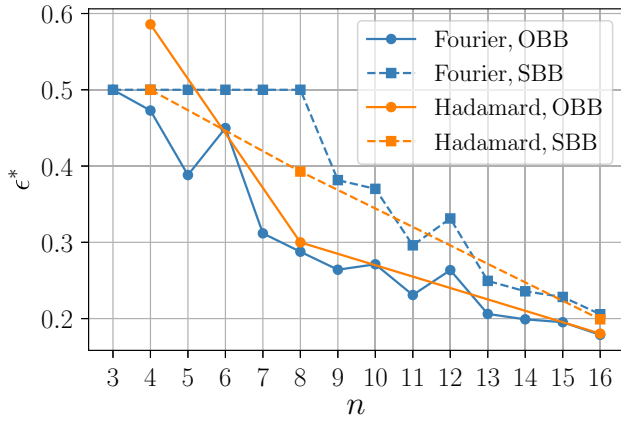


FIG. 3. Plot of the distillation thresholds, the smallest value $\epsilon^* > 0$ such that $e_n(\epsilon^*) = \epsilon^*$. In particular, distillation is advantageous for the given protocol, as long as the initial error ϵ satisfies $\epsilon < \epsilon^*$. We plot both SBB and OBB noise models, for Hadamard and Fourier (F_n) matrices. Numerically we observe the Fourier OBB case to scale approximately as $\epsilon^* \sim 1/n^{0.61}$.

visible in Fig. 2. However, we note that as ϵ grows, the protocols with large n tend to perform worse, due to the higher order terms not considered in Theorem III.2. In particular, each protocol has a *distillation threshold*, a value of ϵ after which it does not necessarily reduce the error rate. We numerically plot some representative threshold values in Fig. 3; these values appear to decrease as n increases, likely approaching 0 in the limit $n \rightarrow \infty$. For practical values of n , however, the thresholds are reasonably large. For example, the 16-photon protocols improve the error rate as long as $\epsilon < 0.179$ (for OBB and SBB error models).

Remark III.3. We note that Theorem III.2 still holds even if, in Protocol III.1, we postselect for only a *subset* of the ideal patterns. In particular, Ref. [23] (released simultaneously with the first version of the present work) considers the output error rates obtained by postselecting on only a *single* ideal pattern at a time. Each pattern yields the same reduction in error up to first order; however, Ref. [23] observes numerically that, when considering *higher-order* terms in ϵ , different patterns may lead to significantly different output error rates. Thus one might consider pruning the set of ideal patterns and keeping only those with the smallest output error rates. When ϵ is small relative to the distillation threshold, so that higher order terms are negligible, this will not be beneficial: a smaller set of ideal patterns leads to a smaller heralding rate and thus larger resource costs, without significantly improving the output error rate. (See Sec. III C for a discussion of resource costs.) However, as seen in Fig. 2, higher-order terms become increasingly nontrivial as ϵ grows, and it is worth investigating whether such methods could lead to improved performance for large ϵ .

In order to understand the utility of a distillation protocol, we must also consider the heralding rate, which determines the resource cost. We investigate this problem in Sec. III B.

B. Heralding rates

For a distillation protocol with heralding probability $h_n(\epsilon)$, we expect to repeat it $1/h_n(\epsilon)$ times in order to successfully herald a distilled output photon. Since each iteration involves n photons, the expected number of photons required for a single distilled output photon is then $n/h_n(\epsilon)$. Thus, we are interested in estimating the heralding rate, and especially obtaining lower bounds, in order to obtain upper bounds on the required resources.

Recalling the decomposition (15) for $\rho^{(n)}(\epsilon)$, we have

$$h_n(\epsilon) = \sum_{k=0}^n \binom{n}{k} \epsilon^k (1-\epsilon)^{n-k} h_n(\Phi_k) \quad (23)$$

$$= (1-\epsilon)^n h_n(\Phi_0) + n\epsilon(1-\epsilon)^{n-1} h_n(\Phi_1) + O(\epsilon^2) \quad (24)$$

$$= h_n(\Phi_0) + \epsilon n(h_n(\Phi_1) - h_n(\Phi_0)) + O(\epsilon^2). \quad (25)$$

Thus, for sufficiently small ϵ , the heralding rate is determined by $h_n(\Phi_0) = h_n(0)$ and $h_n(\Phi_1)$, corresponding to heralding rates for input states with 0 or 1 errors. We have the following:

Theorem III.4. Consider an n -photon distillation protocol with unitary $U = F_{(n_1, \dots, n_\ell)}$, $n = n_1 \cdots n_\ell > 2$, and input state $\rho^{(n)}(\epsilon)$. (In particular, we may take $U = F_n, H_n$.)

- (1) We have $h_n(\Phi_1) = 1/n h_n(0)$.
- (2) The heralding rate is given by

$$h_n(\epsilon) = h_n(0) - (n-1)h_n(0)\epsilon + O(\epsilon^2). \quad (26)$$

Proof. The first claim is proven in Appendix B 3. The second claim follows from the first and Eqs. (23)–(25). ■

Then up to first order in ϵ , the heralding rate is entirely governed by the ideal heralding rate $h_n(0)$. The ideal heralding rate in the cases of interest is characterized as follows. In Appendix D, we compare these results to the case of Haar random linear optical unitaries.

Theorem III.5. Let $U = F_{(n_1, \dots, n_\ell)}$, where $n = n_1 \cdots n_\ell > 2$. More generally, we may take U to be any $n \times n$ unitary with entries in the first row identically equal to $1/\sqrt{n}$.

(1) The ideal heralding rate has the following equivalent expressions:

$$h_n(0) = \left(\frac{-1}{n}\right)^{n-1} (n-1)! \sum_{t=0}^{n-1} (n-t) \frac{(-n)^t}{t!} \quad (27)$$

$$= {}_2F_0(-(n-1), 2; 1/n), \quad (28)$$

where ${}_2F_0(a, b; z) = \sum_{j \geq 0} ((a)_j (b)_j / j!) z^j$ is the generalized hypergeometric function, with $(a)_j = a(a+1) \cdots (a+j-1)$.

(2) $\lim_{n \rightarrow \infty} h_n(0) = 1/4$.

The proof of the first equality in Part 1 is given in Appendix B 1 a; the second follows from reindexing $t \mapsto n-1-j$ and rewriting the terms. Part 2 follows from a result of Vaclav Kotesovec [31], proven using the Maple library *gdev* [32]. We adapt this proof to our setting in Appendix B 1 b.

Remark III.6. We note that $\{n^{n-1} h_n(0)\}_{n \geq 1}$ is visibly a sequence of integers. Kotesovec entered the same sequence into The On-Line Encyclopedia of Integer Sequences as sequence A277458 in 2016 [31], as the coefficients of an exponential generating function related to the Lambert W function. The asymptotic above is given in the encyclopedia listing, up to the factor of n^{n-1} . We note that the W function appears in many physical problems [33], and certain variants were recently used in proposed protocols for verification of Gaussian boson sampling [34].

The hypergeometric expression allows for quick numerical estimation of the ideal heralding rate. The first few values are $h_3(0) = 1/3$, $h_4(0) = 1/4$, $h_5(0) \approx 0.264$. In Fig. 4, we plot the difference $h_n(0) - 1/4$ and numerically observe that $h_n(0)$ quickly approaches $1/4$, with asymptotic $h_n(0) - 1/4 \sim 1/16n$. In particular, we conjecture the following.

Conjecture III.7. For $n \geq 5$, the sequence $h_n(0)$ monotonically decreases to $1/4$.

As an immediate consequence of Theorems III.4 and III.5, we have

Proposition III.8. For $n \geq 4$ and sufficiently small ϵ ,

$$h_n(\epsilon) \approx \frac{1}{4} - \left(\frac{n-1}{4}\right) \epsilon. \quad (29)$$

In Fig. 5, we numerically plot heralding rates as a function of ϵ for the Fourier case. This plot, and others in this work, are obtained by numerically calculating the terms $h_n(\Phi_k)$, $0 \leq k \leq n$, and substituting into Eqs. (23)–(25). Thus we obtain $h_n(\epsilon)$ as a polynomial in ϵ with explicitly

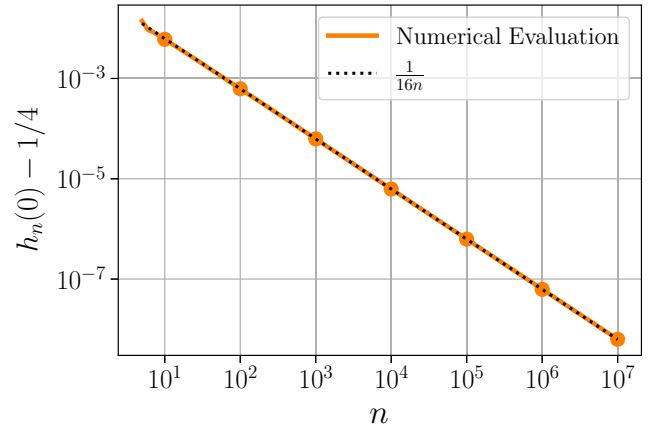


FIG. 4. Numerical evaluation of $h_n(0) - 1/4$ for $n = 5$ up to 10^7 . The data (solid line) contains all values of $n \in [5, 10^5]$, as well as multiples of 10^5 up to 10^6 , and multiples of 10^6 up to 10^7 (we include circular data points at powers of 10 to guide the eye). This was computed using the hypergeometric function, Eq. (28). We observed numerically that the deviation of $h_n(0)$ from $1/4$ falls as $1/(16n)$, plotted as the dashed line.

known coefficients, up to rounding errors in the calculations for $h_n(\Phi_k)$. No sampling or fitting with respect to particular values of ϵ is required. In Appendix F, we explicitly give a list of computed heralding rates and output error rates as functions of ϵ . We see from Fig. 5 that the protocol with $n = 3$ has a significantly better heralding rate than the larger protocols; however, as observed in Theorem III.2, the larger protocols have better output error rates for small ϵ . Further, since the heralding rate $h_n(0)$ stabilizes at $1/4$, it seems that for small ϵ it is advantageous to increase the value of n .

We use a similar expansion to the above, discussed in Appendix B, to calculate the output error rates $e_n(\epsilon)$, plotted in Fig. 2. We note that as ϵ grows larger, it is not necessarily better to increase n ; this is especially clear in Figs. 3 and 6. We further discuss the trade-offs between different protocols in Sec. III C.

C. Comparison of resource costs

We now compare the resource costs of our distillation protocols with previous work. In particular, we consider the expected number of photons required to obtain output error rate $\epsilon' \approx \epsilon/n$ in the small ϵ regime.

The first photon distillation protocol, due to Sparrow and Birchall and called HOM filtering [4], requires n photons per distillation attempt, with heralding probability decaying exponentially to 0 [8]. Thus the expected number of photons required is exponential in n . If instead one iterates the two-photon version of this scheme with active feedforward, it is possible to have a scheme that scales cubically $O(n^3)$, as pointed out in Ref. [8].

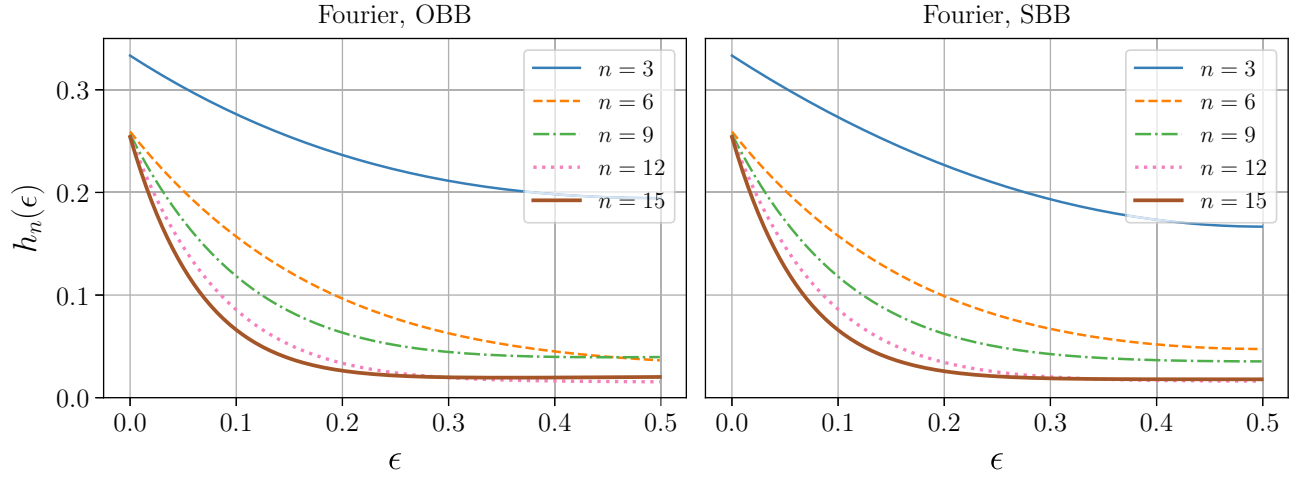


FIG. 5. Plot of the heralding rate $h_n(\epsilon)$ with OBB (left) and SBB (right) error models for the Fourier distillation protocol. The zero-error heralding rate $h_n(0)$ is given analytically by Theorem III.5.

Marshall [8] proposed the three-photon Fourier distillation scheme and suggested iterating the scheme to arbitrarily reduce the error rate at the cost of additional photons. In particular, many batches of photons with error rate ϵ are put through a three-photon distillation protocol, giving output photons with error rate near $\epsilon/3$ upon successful heralding; these then undergo further rounds of distillation to obtain error rate approximately $\epsilon/3^2$, then $\epsilon/3^3$, and so on. Without active feedforward, it is easy to see that the expected number of photons to obtain output error rate $\epsilon' \approx \epsilon/n$ (where n is a power of 3) is exponential in n , since it is highly unlikely for many distillation protocols to succeed simultaneously. With active feedforward and the ability to perfectly store successfully heralded photons in memory, Marshall's iterated distillation protocol can be made more efficient: the output photons from one round

of distillation can be stored until enough are available to use as input for subsequent rounds of distillation. Such an iterated protocol reduces the error rate by a factor of $n = 3^r$ with an expected cost of n^2 photons [8]. In practice, however, these additional requirements may introduce more errors, for example, due to dispersion introduced by optical delay lines.

In our present setting, an n -photon protocol $F_{(n_1, \dots, n_\ell)}$ obtains output error rate $e_n(\epsilon) \approx \epsilon/n$ upon successful heralding in the small ϵ regime. The heralding rate satisfies $h_n(0) \approx 1/4$ for $n \geq 4$ (see Theorem III.5 and the following discussion). Then we have the following:

Theorem III.9. For $n \geq 4$ and ϵ sufficiently small relative to n , the expected number of photons required to distill

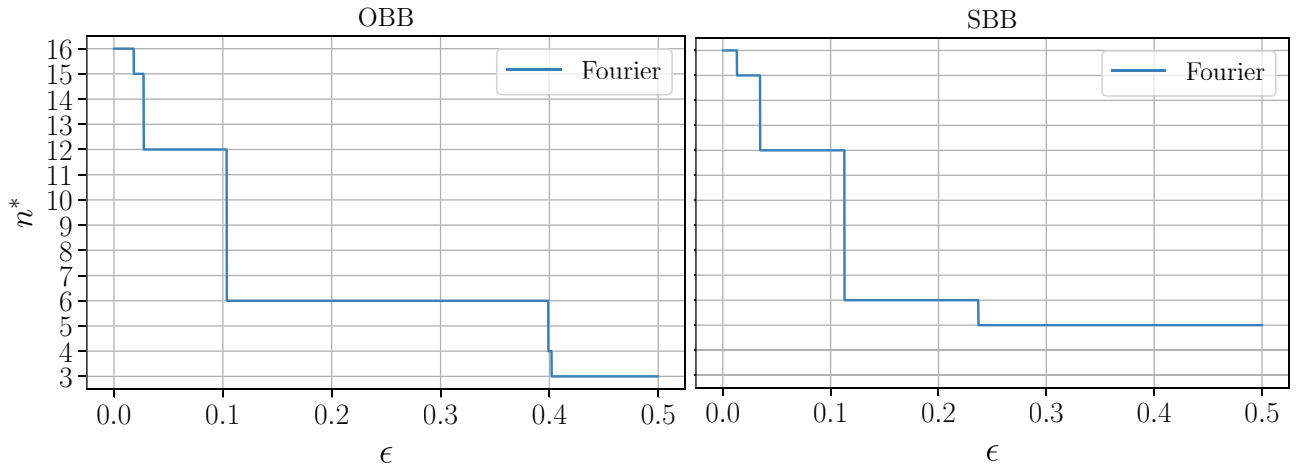


FIG. 6. Given an initial error rate ϵ , we plot the value n^* of n for which the protocol with $U = F_n$ gives the greatest error reduction (this does not take into account heralding rates or expected resource counts). We considered protocols up to size $n = 16$. As expected, for small ϵ we have $n^* = 16$, since $e_n(\epsilon) \approx \epsilon/n$.

a single output photon with error rate $e_n(\epsilon) \approx \epsilon/n$ is

$$\frac{n}{h_n(\epsilon)} \approx \frac{n}{h_n(0)} \approx 4n. \quad (30)$$

In particular, our n -photon distillation protocol requires only *linearly scaling resources* to obtain the target error in the low ϵ regime, *without* requiring feedforward or memory. Thus our protocols are the most resource-efficient known method for distilling the distinguishability error rate to an arbitrarily small value.

Remark III.10. We now more closely consider the feasibility of the F_n distillation protocols relative to previous work, specifically the iterated F_3 protocol of [8] discussed above. For $n = 3^r$, the iterated F_3 protocol requires r concatenated rounds of F_3 distillation protocols, with an expected cost of n^2 photons, in order to reduce the error rate by a factor of n . Further, the iterated protocol of Ref. [8] (for $r > 1$ rounds) requires active feed-forward and storing photons in memory, which in fact can *increase* the distinguishability error rate, as discussed above. On the other hand, the F_n protocol instead requires around $4n$ photons for the same amount of error reduction, *without* requiring memory or active feed-forward. For example, to reduce the error rate by a factor of $n = 81$, the protocol of Ref. [8] requires 4 rounds of the F_3 distillation scheme, resulting in the use of $n^2 = 6561$ photons. The present work requires approximately $4n = 324$ photons, using a single F_{81} protocol. This is more than 20 times more efficient than the iterated F_3 protocol. In fact, if one is willing to use 6561 photons, the number used by the iterated F_3 protocol, they may instead use a single iteration of F_{1640} , reducing the error rate by a factor of approximately 1640 without increasing the cost. (And note that the difficulty of implementation is potentially *decreased*, since memory and feed-forward are not required.) Thus, given a desired level of performance, the F_n protocols are both dramatically more efficient and simpler to implement than the work of Ref. [8], which in turn was far more efficient than Ref. [4].

We recall, however, that our protocols for large n have an ever-decreasing *threshold* after which the protocol fails to reduce the distinguishability error rate. (See Fig. 3.) In particular, in the discussion above, one must take ϵ sufficiently small so that it is below the threshold for all relevant values of n . This limitation does not apply to the iterated protocol of Ref. [8], as the relevant threshold is that of the three-photon protocol. We note that these thresholds are not a major concern for moderately sized protocols: as discussed above, we expect realistic experimental protocols to have $\epsilon \leq 0.15$ or so, below threshold for all $n \leq 16$, as seen in Fig. 3.

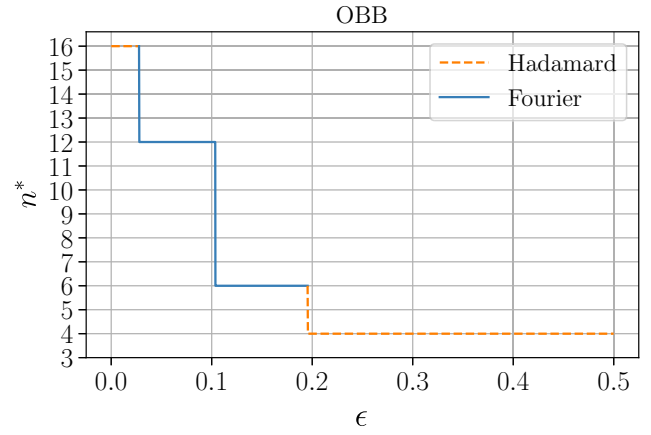


FIG. 7. We extend the plot of Fig. 6 (OBB) to include the Hadamard protocols $U = H_4, H_8, H_{16}$. The different line styles indicate whether the protocol of interest is Fourier or Hadamard. Note that in the SBB case, the plot is unchanged from the one in Fig. 6 (i.e., for the SBB error model, the Fourier protocol always has the greatest error reduction).

However, even below the threshold, we may not always want to choose the largest value of n available. This is discussed in Sec. III E, especially Figs. 6 and 7, where, for each ϵ , we numerically identify the value of n with smallest output error rate $e_n(\epsilon)$. If the initial error rate ϵ is large, we note that one may iterate distillation protocols with gradually increasing numbers of photons. The first rounds will have small n , chosen according to the figures, allowing for the reduction of the distinguishability error rate. Once the error rate is sufficiently small, one may use larger protocols in later rounds, further decreasing the error rate at a lower cost. (However, as discussed above, iterative protocols come at the cost of requiring active feedforwarding and quantum memory during distillation.)

We briefly consider some practical examples. Recall that, from visibility estimates in the literature [10–16], we approximate that the initial error rate ϵ_0 is likely in the range [0.005, 0.15]. We first consider the high end of this range. Given initial $\epsilon_0 = 0.15$, applying the protocol with $U = F_6$ will give output error rate $\epsilon'_0 = e_6(\epsilon_0) \approx 0.056$ in the OBB model. [Calculated using the exact formula for $e_6(\epsilon)$ in Appendix F.] If a smaller error rate is required, we may then insert the output photons with error rate ϵ'_0 into the $U = F_{12}$ protocol, giving output error rate $\epsilon''_0 = e_{12}(\epsilon'_0) \approx 0.0097$. We now consider the low end of the range, corresponding to very small error rate $\epsilon_0 = 0.005$, roughly corresponding to the visibility found in Ref. [16]. Of the protocols we considered, the optimal one in the OBB model would be $U = H_{16}$. After one iteration, we obtain output error rate $\epsilon'_0 \approx 0.0003565$, an improvement by a factor of around 14. Thus distillation protocols can have a significant impact in both the low and high error regimes.

D. Symmetries and conjectures

In this section, we briefly discuss the significance of certain symmetry properties to our results. We will return to this subject in much greater detail in Appendix A. Further, in Appendix E, we discuss how these and other symmetries can be leveraged to greatly reduce the complexity of calculating the $h_n(\Phi_k)$ and related quantities via simulation.

We consider a distillation protocol determined by $n \times n$ matrix U . We refer to a *mode symmetry* of U to be an $n \times n$ permutation matrix P such that there exists an $n \times n$ diagonal matrix D with $UP = DU$. In a boson sampling experiment, this corresponds to the observation that permuting the *modes* (not the photons) of the input state is equivalent to applying phase shifts on the output state. Let G be an Abelian group of such mode symmetries for U . In other words, G is an Abelian permutation group that is simultaneously diagonalized by U . One can show that such symmetries lead to suppression laws [19]; in other words, we obtain a set \mathcal{S}_G of patterns (s_0, \dots, s_{n-1}) , determined by the symmetry group G , such that

$$\langle s_0, \dots, s_{n-1} | \hat{U} | 1, \dots, 1 \rangle \neq 0 \implies (s_0, \dots, s_{n-1}) \in \mathcal{S}_G. \quad (31)$$

In a distillation setting, recalling Eq. (19), we see that the *ideal patterns* must be a subset of \mathcal{S}_G . As discussed above, in the Hadamard case, the ideal patterns may be exactly identified as the elements of \mathcal{S}_G (satisfying our additional restriction $s_0 = 1$) [18, 19]. In the Fourier case, we say that patterns in \mathcal{S}_G satisfy the *zero transmission law* (ZTL), and the ideal patterns are in general a proper subset of the ZTL patterns [17]. The first nonideal pattern satisfying the ZTL occurs for $n = 6$ (see Appendix A 2). In general, not all suppression laws may be explained by symmetry conditions such as the ZTL [35]: we expect that the extra suppression for $U = F_6$ is such a case. In Appendix A 2 a, we prove the following theorem.

Theorem III.11. If n is a prime power, the suppression laws for $U = F_n$ are exactly characterized by the zero transmission law. In particular, the ideal patterns (s_0, \dots, s_{n-1}) for $U = F_n$ are precisely the elements of \mathcal{S}_G with $s_0 = 1$. If $n > 1$ is not a prime power, we may explicitly exhibit a pattern in \mathcal{S}_G with $s_0 = 1$ that is *not* an ideal pattern.

We also discuss analogs of the ZTL for general $F_{(n_1, \dots, n_\ell)}$ in Appendix C.

Beyond constraining the ideal patterns, these symmetries may be used to simplify calculations for heralding and output error rates. In particular, we have the following, proven in Appendix A.

Theorem III.12. Let U be an $n \times n$ unitary with corresponding symmetry group G and set \mathcal{S}_G of symmetry-preserving patterns. Let $|\eta\rangle \in \mathcal{H}^{\otimes n}$ be an arbitrary pure state, with normalized projection $|\eta_+\rangle$ onto the space of G -symmetric states. We have the following results.

(1) Overlap with states in \mathcal{S}_G may be calculated in terms of the symmetrized state $|\eta_+\rangle$:

$$\langle s_0, \dots, s_{n-1} | \hat{U} | \eta \rangle = \langle \eta_+ | \eta \rangle \langle s_0, \dots, s_{n-1} | \hat{U} | \eta_+ \rangle.$$

(2) The heralding and error rates may be calculated in terms of the symmetrization:

$$h_n(|\eta\rangle\langle\eta|) = |\langle \eta_+ | \eta \rangle|^2 h_n(|\eta_+\rangle\langle\eta_+|), \quad (32)$$

$$e_n(|\eta\rangle\langle\eta|) = e_n(|\eta_+\rangle\langle\eta_+|). \quad (33)$$

(3) Further assume that $U = H_n$, where n is a power of 2, or $U = F_n$, where n is a prime power. Then

$$h_n(|\eta_+\rangle\langle\eta_+|) = |(\langle 1 | \otimes I) \hat{U} | \eta_+ \rangle|^2. \quad (34)$$

In particular, this gives a method for calculating heralding rates $h_n(|\eta\rangle\langle\eta|)$ while only checking whether output patterns have 1 photon in the output mode, rather than verifying whether the entire pattern is ideal. In Appendix B 2, we express $e_n(|\eta\rangle\langle\eta|)$ in a similar form involving only measurements on the 0th mode.

Remark III.13. We briefly comment on the assumption on U in part 3 above. More generally, we can replace this with the assumption that all patterns (s_0, \dots, s_{n-1}) in \mathcal{S}_G with $s_0 = 1$ and $\langle s_0, \dots, s_{n-1} | \hat{U} | \eta \rangle \neq 0$ are ideal. As discussed above, this is automatically satisfied for U as in the theorem, as all patterns in \mathcal{S}_G with $s_0 = 1$ are ideal. For $|\eta\rangle = |1, \dots, 1\rangle$ perfectly indistinguishable, Eq. (34) follows from the definition of ideal pattern regardless of the unitary U , allowing for the straightforward computation of $h_n(0)$ as in Appendix B 1 a. Further, we show in Appendix B 3 that we may use the indistinguishable case to calculate $h_n(\Phi_1)$ for any $U = F_{(n_1, \dots, n_\ell)}$ regardless of the value of $n = n_1 \cdots n_\ell > 2$. In the Fourier case with n not a prime power, one may modify Protocol III.1 to postselect for any patterns in \mathcal{S}_G with $s_0 = 1$ rather than the smaller set of ideal patterns, so that the general assumption given above holds by definition. This intentionally increases the output error rate, but allows for the application of Theorem III.12 Part 3 and makes the behavior of the protocol more predictable for all n . Noting Fig. 6, however, we advise against doing this, as the Fourier protocols with n not a prime power generally seem to be optimal without this modification.

These symmetry considerations are essential in the proofs given in the Appendix. Further, they lead to the following conjectures.

Conjecture III.14. As above, assume that $U = H_n$, where n is a power of 2, or $U = F_n$, where n is a prime power. Let $|\eta\rangle, |\Delta\rangle$ be states of the form (14), where $|\eta\rangle$ has k distinguishability errors and $|\Delta\rangle$ has n (so that in the latter case, all photons are mutually fully distinguishable). Let $|\eta_+\rangle, |\Delta_+\rangle$ be the G symmetrizations, as above. Then

$$h_n(\Phi_0) \leq h_n(|\eta_+\rangle\langle\eta_+|) \leq h_n(|\Delta_+\rangle\langle\Delta_+|). \quad (35)$$

In particular,

$$\frac{1}{n}h_n(\Phi_0) \leq h_n(\Phi_k) \leq \frac{\gcd(k, n)}{n}h_n(|\Delta_+\rangle\langle\Delta_+|). \quad (36)$$

The heart of this conjecture is that for G -symmetrized states, more mode symmetry seems to *decrease* the heralding rate. Thus the lower bound comes from the fully mode-symmetric state $\Phi_0 = |1, \dots, 1\rangle\langle 1, \dots, 1|$ and the upper bound from a state $|\Delta\rangle$ with as little symmetry as possible. The coefficients come from Theorem III.12 Part 2 and trivial bounds on $|\langle\eta_+|\eta\rangle|^2$ when $|\eta\rangle$ has k distinguishability errors.

For F_n where n is not a prime power, Conjecture III.14 fails, with such protocols tending to have lower heralding rate than predicted: for example, in the OBB model we have $h_6(\Phi_2)/h_6(\Phi_0) \approx 0.132 < 1/6$. However, the conjecture seems to hold for general F_n if we modify the protocol as suggested in Remark III.13. With this modification, we would, for example, have $h_6(\Phi_2)/h_6(\Phi_0) \approx 0.216 > 1/6$, and the $n = 6$ case would then fit into the same patterns as the prime-power cases. [But this would negatively impact the error rate, e.g., increasing $e_6(\Phi_2)$ from 0.017 to 0.026.]

As a corollary of the previous conjecture, we obtain the following.

Conjecture III.15. For the n -photon Fourier or Hadamard protocols, with n a prime power as above, we have

$$h_n(\epsilon) \geq h_n(0) \left((1 - \epsilon)^n \left(1 - \frac{1}{n} \right) + \frac{1}{n} \right). \quad (37)$$

In particular, if $\epsilon \leq 1/n$, by Theorem III.5 we have

$$\begin{aligned} \lim_{n \rightarrow \infty} h_n(\epsilon) &\geq \lim_{n \rightarrow \infty} h_n(0) \left(\left(1 + \frac{1}{n} \right)^{n+1} + \frac{1}{n} \right) \\ &= \frac{1}{4e} \approx 0.092. \end{aligned} \quad (38)$$

Then assuming the conjecture holds, we see that as long as ϵ is small relative to n , the heralding rates cannot get too small even for large n . Further, if Conjecture III.7 holds,

we have for all $n \geq 3$

$$h_n(\epsilon) \geq \frac{1}{4e}. \quad (39)$$

Recall from our earlier discussion of error thresholds that choosing n and ϵ appropriately is also necessary to ensure that the distillation protocols reduce error rates. Here we note that the numerical threshold values obtained in Fig. 3 are all far greater than $1/n$. Thus it appears that if $\epsilon < 1/n$, we obtain both a large heralding rate and a distillation protocol that improves the output error rate.

E. Numerics

In this section, we give further numerics for our protocols. The details of how our numerical simulations are performed is discussed in Appendix E, with data provided in Appendix F.

First, we consider the question of optimal distillation protocols given an error rate ϵ . In particular, as discussed in Sec. III C, we expect the heralding rates of our protocols to be reasonably large, with linearly growing resource requirements. The limitation is instead the growth of the error rate: protocols with large n give the smallest output error rate when ϵ is small, but perform worse for large ϵ . Thus, in Fig. 6, we consider all protocols with $U = F_n$, $3 \leq n \leq 16$, and for each ϵ determine the value of n with $e_n(\epsilon)$ as small as possible. We are most interested in $\epsilon \leq 0.15$, as discussed above. As expected, we find that the largest protocol is optimal for small ϵ , and as ϵ grows the optimal protocol becomes smaller and smaller. Further, in the OBB model, there is a large plateau for ϵ roughly between 0.1 and 0.4 in which the six-photon protocol is optimal. We expect this is due to the fact for $n = 6$, the set of ideal patterns is strictly smaller than the set of symmetry-preserving patterns, as discussed in Sec. III D and Appendix A 2. In fact, for both error models and reasonably small values of ϵ , the dominant protocols seem to have $n = 6, 12, 15$, all of which are not prime powers and have a set of ideal patterns strictly smaller than the set of symmetry-preserving patterns [20] (recall Theorem III.11). (Note we exclude $n = 16$ from the discussion here, as by Theorem III.2 it is guaranteed to be optimal for sufficiently small ϵ .) The strong performance of $n = 6, 12, 15$ is unsurprising, since a smaller set of ideal patterns leads to a smaller heralding rate: see the discussion below Conjecture III.14. As explained in more detail below, this should generally reduce the error rate as well. We also note that these seemingly optimal protocols have n divisible by 3. Understanding this divisibility condition may explain why there is no point at which $n = 14$ is preferable to $n = 12$ or $n = 15$, even though all three have fewer ideal patterns than symmetry-preserving patterns. In Fig. 7, we repeat this analysis while including the Hadamard protocols as well. We find that for reasonably small values of ϵ , the

same Fourier protocols (n not a prime power and divisible by 3) are generally preferred. (The fact that H_{16} is preferred over F_{16} in the OBB model seems to be related to the discussion of the $n = 8$ case below.) We conjecture that if we included all $n \leq 18$ in the analysis, we would see the optimal protocol transition directly from F_{18} to F_{15} or F_{12} .

In the preceding analysis, we compared protocols with different values of n . Next, we fix $n = 8$ and numerically investigate the heralding and error rates of the distillation protocols using different unitaries and error models. We choose to study $n = 8$ here because it is a power of 2, allowing for comparison of corresponding Fourier and Hadamard distillation protocols. (We include a third protocol corresponding to $F_{(4,2)}$ in Appendix C.) The $n = 4$ case was originally presented in Ref. [8], so we focus on the $n = 8$ case. We note that similar results to the below hold for the $n = 4$ case. For this discussion, we note the following formula for the output error rate, given in Appendix B 2:

$$e_n(\epsilon) = \frac{\bar{e}_n(\epsilon)}{h_n(\epsilon)} = \frac{\sum_{k=1}^n \binom{n}{k} \epsilon^k (1 - \epsilon)^{n-k} e_n(\Phi_k) h_n(\Phi_k)}{\sum_{k=0}^n \binom{n}{k} \epsilon^k (1 - \epsilon)^{n-k} h_n(\Phi_k)}. \quad (40)$$

We recall that $e_n(\Phi_k)$ is a conditional probability: the probability that, if a state with k distinguishability errors leads to an ideal pattern, it outputs a nonideal photon. Recalling Theorems III.2, III.4, and III.5, the quantities $h_8(\Phi_0)$, $h_8(\Phi_1)$, $e_8(\Phi_1)$ are fixed and we have the first-order approximations

$$\begin{aligned} e_8(\epsilon) &= e_8(\Phi_1)\epsilon + O(\epsilon^2), \\ h_8(\epsilon) &= h_8(\Phi_0)(1 - 7\epsilon) + O(\epsilon^2), \end{aligned} \quad (41)$$

regardless of the error model or the choice of $U \in \{F_8, H_8\}$. Then we expect much of the difference between protocols and models to be explainable via the second-order terms, which are determined by $h_8(\Phi_2)$, $e_8(\Phi_2)$ in addition to the

TABLE I. We plot the values of $h_8(\Phi_2)$ and $e_8(\Phi_2)$ for the variant protocols and error models considered here. These are the dominant values that differ between cases, and they determine the behavior of $h_8(\epsilon)$, $e_8(\epsilon)$ up to second order.

	$h_8(\Phi_2)$	$e_8(\Phi_2)$
Fourier, OBB	0.040	0.377
Fourier, SBB	0.042	0.328
Hadamard, OBB	0.038	0.338
Hadamard, SBB	0.076	0.338

above fixed quantities. In particular, $e_8(\Phi_2)$ and $h_8(\Phi_2)$ affect the numerator of Eq. (40) to second order; $h_8(\Phi_2)$ also affects the denominator, but only to third order. We give values for these second-order quantities in Table I. One may think of Φ_2 in the OBB model as an average of all states with two “different” errors; in the SBB model, Φ_2 is an average of states with two of the “same” error.

In Fig. 8, we plot the heralding rates and output error rates in the Fourier and Hadamard cases, using both the OBB and SBB error models. We are particularly interested in the small ϵ regime. We note from the plot that three of the four cases considered have very similar heralding rates $h_8(\epsilon)$. The Hadamard case with SBB error model, however, has a significantly larger heralding rate. This is clearly reflected in the values of $h_8(\Phi_2)$ in Table I, with the Hadamard SBB case nearly doubling the other values. Since the only second-order effect of $h_8(\Phi_2)$ on Eq. (40) is in the numerator, an increase in this quantity should increase the overall error rate. This explains why this case has significantly worse error rate than the others, as seen in Fig. 8. Intuitively, the Hadamard protocol does not notice two of the “same” (SBB) errors as often, so we are more likely to have distinguishable output photons.

Remark III.16. We reiterate the above observation: even though a large heralding rate reduces the overall resource requirements, it can also lead to larger error rates. This

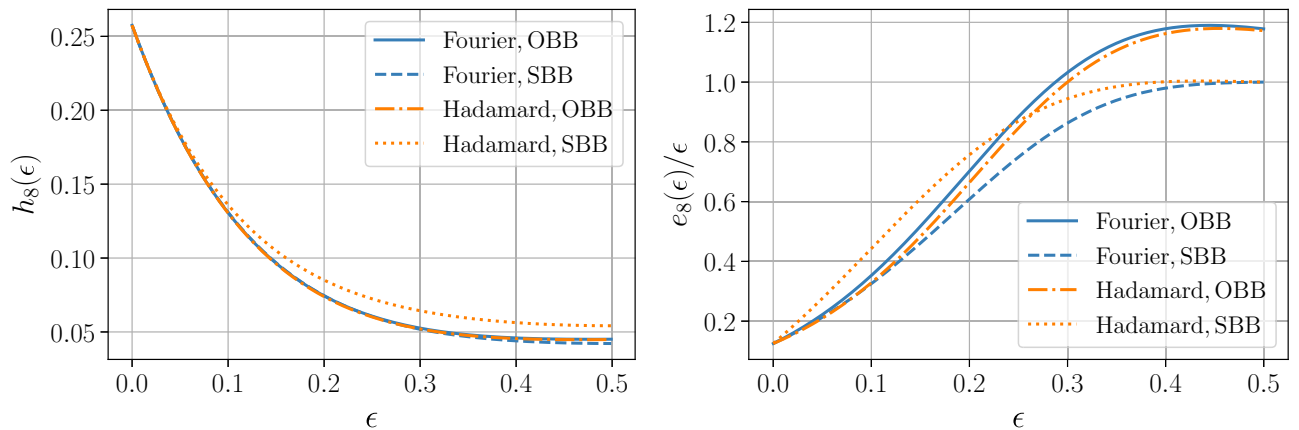


FIG. 8. Heralding rate (left) and output error rate (right) for $n = 8$ Fourier and Hadamard protocols, for both OBB and SBB noise models.

is what makes our protocols notable, especially as n increases: they maintain a nontrivial heralding rate while arbitrarily decreasing the error rate.

Next, we fix an error model and compare error rates $e_n(\epsilon)$ for F_n and H_n . For SBB, the Fourier case $U = F_8$ has smaller output error rate, translating to better performance. This is explained by the analysis above. When we consider the OBB error model, however, we see that the Hadamard protocol $U = H_8$ has the smaller error rate and better performance, with both $h_8(\Phi_2)$ and $e_8(\Phi_2)$ decreasing in this case. Then for a *fixed* $n = 2^r$, we cannot conclude whether Fourier or Hadamard unitaries are preferable in general: this depends strongly on the error model relevant to the experiment in question. Note that this is in contrast with the discussion of Figs. 6 and 7 above, in which we found that Fourier protocols often perform better when we are free to choose the value of n .

Similarly, we may fix the unitary (Fourier or Hadamard) and compare the error rates $e_n(\epsilon)$ under the two error models. From Fig. 8, we see that the Hadamard protocol performs better with OBB errors, whereas the Fourier protocol performs better with SBB errors. In the Hadamard case, this may be explained by the increased heralding rate under the SBB model, as discussed above. For $U = F_8$, however, the heralding rate does not significantly differ between the two error models. [In fact, the protocol with smaller error rate $e_n(\epsilon)$ has slightly *larger* heralding rate!] Then the discrepancy is caused by the change in $e_8(\Phi_2)$, which is proportionally more significant. Intuitively, for the F_8 protocol, states with two errors give ideal output patterns at roughly the same rate regardless of error model, but when heralding occurs and the errors are “different,” the error photons are more likely to end up in the output mode.

In summary, we note that Fourier protocols, especially those with n divisible by 3 and not a prime power, seem to be the most effective at reducing the error rate. For a fixed $n = 2^r$, however, whether F_n or H_n performs better seems to be highly dependent on the error model. Finally, the Hadamard protocols seem to struggle with filtering out SBB errors, while the Fourier protocols are equally capable of filtering out both types but, when heralding occurs, are more likely to direct OBB errors to the output mode.

F. Incorporation of photon loss

We now discuss the incorporation of photon-loss models into our analysis of the Fourier and Hadamard distillation protocols. We begin with a brief discussion of loss models.

We initially consider the *uniform beamsplitter loss model*, in which our linear optical circuit is made up of lossy beam splitters, and each photon has probability λ of loss at each beam splitter. We model this as a photon-loss channel with loss probability λ , occurring independently on each of the two relevant modes, just before the beam

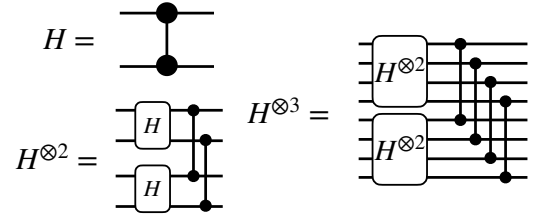


FIG. 9. Illustration of construction of the circuit implementing the $n \times n$ Hadamard unitaries $H_n = H^{\otimes r}$, for $r = 1, 2, 3$, where $n = 2^r$. It generalizes in the obvious way. The circuit components are 50:50 beam splitters. One can see directly that each mode interacts with r beam splitters.

splitter. We recall, however, that symmetric loss commutes through linear optics [36]. For the unitary $U = H_n$, we give an explicit recursive circuit decomposition in Fig. 9, and note that the described loss channels may be commuted to the end of the circuit. In particular, since each path from an input mode to an output mode involves exactly $\log_2 n$ beam splitters, the loss probability on each mode is $\Lambda = 1 - (1 - \lambda)^{\log_2 n}$. For simplicity, we consider the same loss model in the Fourier case. This may likely be derived as above; for example, when n is a power of 2, a circuit decomposition similar to that in Fig. 9 is given in Ref. [37], and the same analysis holds. Then, going forward, we will use the *simplified loss model* depicted in Fig. 10, in which partially distinguishable photons are given as input to the circuit, the linear optical unitary \hat{U} is applied, each photon independently undergoes a loss channel with probability $\Lambda = 1 - (1 - \lambda)^{\log_2 n}$, and the final $n - 1$ modes are measured. We note that in this model, the probability that no photons are lost is $(1 - \Lambda)^n = (1 - \lambda)^{n \log_2 n}$.

We begin by discussing the heralding rate, which we now write as $h_n(\epsilon; \lambda)$ to allow for dependence on the loss rate, with $h_n(\epsilon) = h_n(\epsilon; 0)$. Heralding requires the detection of $n - 1$ photons in the latter $n - 1$ modes; in particular, heralding is only possible if at most one photon

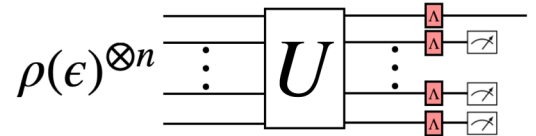


FIG. 10. Illustration of our error model. As in the rest of the paper, n partially distinguishable photons, each with internal state $\rho(\epsilon)$, are input into n different external modes, and we apply the linear optical unitary \hat{U} corresponding to an $n \times n$ unitary U . Unique to this section, we then apply a photon-loss channel, with loss probability $\Lambda = 1 - (1 - \lambda)^{\log_2 n}$, to each mode. We then perform PNRD on the final $n - 1$ modes and postselect as usual. For the Hadamard circuits of Fig. 9, this is equivalent to the uniform beam-splitter loss model with probability λ of loss at each beam splitter.

has been lost. We have the corresponding decomposition

$$h_n(\epsilon; \lambda) = (1 - \Lambda)^n h_n(\epsilon) + n\Lambda(1 - \Lambda)^{n-1} h_n^{(1)}(\epsilon), \quad (42)$$

where $h_n(\epsilon) = h_n(\epsilon; 0)$ encapsulates the heralding probability when no photons are lost and $h_n^{(1)}(\epsilon)$ is the heralding probability when exactly one photon is lost [as mentioned above, $h_n^{(k)}(\epsilon) = 0$ for $k > 1$]. Note that we have two possibilities for preloss output patterns contributing to $h_n^{(1)}(\epsilon)$: either we have an ideal pattern and the single photon in mode 0 is lost, which occurs with probability $1/nh_n(\epsilon)$; or we have a *nonideal* pattern (with no photons in mode 0), and a photon in a nonzero mode is lost, and the resulting pattern on the final $n - 1$ modes matches an ideal pattern. We write the latter probability as $c_n(\epsilon)$. In either case, no photons exit mode 0. From this description, we decompose $h_n^{(1)}(\epsilon) = 1/nh_n(\epsilon) + c_n(\epsilon)$, and substituting into Eq. (42) gives

$$h_n(\epsilon; \lambda) = (1 - \Lambda)^{n-1} h_n(\epsilon) + n\Lambda(1 - \Lambda)^{n-1} c_n(\epsilon). \quad (43)$$

Theorem III.17. Let all notation be as above. For the protocol corresponding to $U = F_{(n_1, \dots, n_\ell)}$, $n = n_1 \cdots n_\ell$, we have

$$h_n(\epsilon; \lambda) \geq (1 - \Lambda)^{n-1} h_n(\epsilon) = (1 - \lambda)^{(n-1) \log n} h_n(\epsilon). \quad (44)$$

If $\epsilon = 0$, the bound is tight:

$$h_n(0; \lambda) = (1 - \Lambda)^{n-1} h_n(0) = (1 - \lambda)^{(n-1) \log n} h_n(0). \quad (45)$$

Proof. The first claim follows by bounding $c_n(\epsilon) \geq 0$ in Eq. (43). For the second claim, we argue that $c_n(0) = 0$.

We recall that when $\epsilon = 0$, the (preloss) output patterns satisfy certain symmetry constraints, given in Sec. III A. In particular, in the notation used above, we have $\sum_i g_i \equiv 0 \pmod n$ in the Fourier case and $\bigoplus_i g_i = 0$ in the Hadamard case. (We have similar restrictions for general $F_{(n_1, \dots, n_\ell)}$, discussed in Appendix C.) For a nonideal pattern to contribute to c_n , there must exist two symmetry-preserving patterns of n photons in n modes, one ideal and one non-ideal, both of which may give the same $(n - 1)$ -photon pattern after a loss. The symmetry conditions given above make this impossible: given a pattern of $n - 1$ photons in n modes, there is a unique mode j such that adding a photon to mode j results in a symmetry-preserving pattern. This is clear for the Fourier and Hadamard cases given above and is discussed for the general case in Appendix C. ■

This may be used to obtain an upper bound on the resources required for distillation in the presence of loss, namely

$$\frac{n}{h_n(\epsilon; \lambda)} \leq \frac{n}{(1 - \lambda)^{(n-1) \log n} h_n(\epsilon)} \approx \frac{4n}{(1 - \lambda)^{(n-1) \log n}}, \quad (46)$$

where the approximation follows from Theorem III.5, as in Theorem III.9. With a nonzero loss rate λ per beam splitter, the heralding rate is exponentially suppressed in $n \log n$, leading to exponentially growing resource requirements. Given a fixed amount of resources and error rates ϵ, λ , this effectively puts an upper bound on the maximum feasible size n of a distillation protocol. For moderate values of n , however, the effect is not so drastic. We show numerical scaling of the heralding probability in Fig. 11 for $n = 8$,

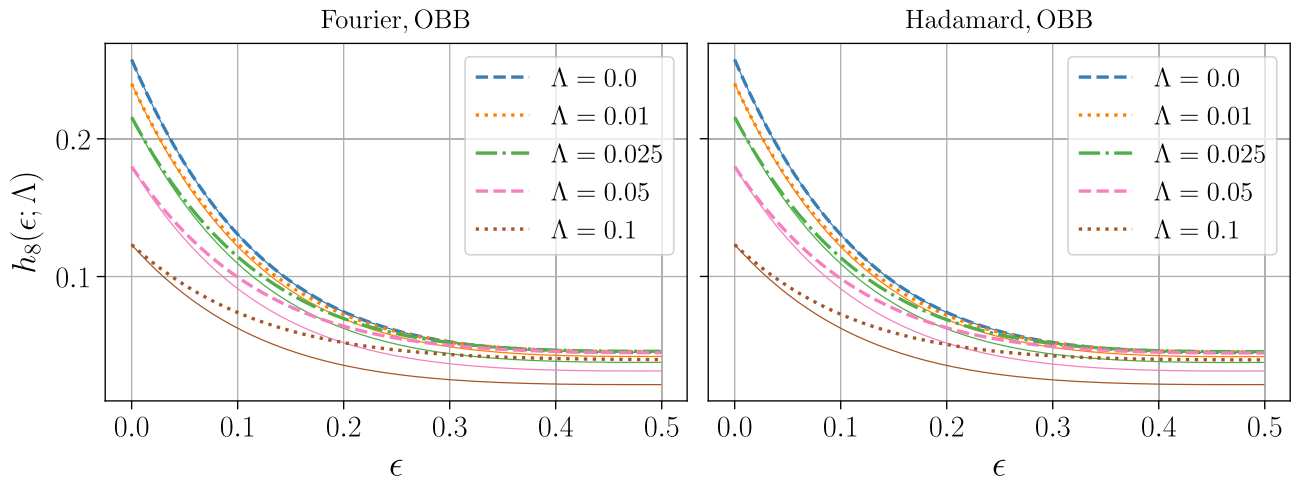


FIG. 11. Heralding probability with photon loss in the Fourier (left) and Hadamard (right) protocols, for $n = 8$, with the OBB error model. The SBB curves look similar. Note that the parameter in the legend is $\Lambda = 1 - (1 - \lambda)^{\log n}$. The thin solid lines represent the lower bound, Eq. (44), and notice they intersect the data curves at $\epsilon = 0$ as per Eq. (45). We see the lower bound is tighter for smaller Λ .

where we see that even a high loss rate of $\Lambda = 0.1$ only cuts the heralding probability in half. As a further example, we consider distillation with $n = 16$. Without loss, distillation would require four runs on average, with a total expected cost of approximately 64 photons. With a loss rate of $\lambda = 0.01$ ($\Lambda \approx 0.04$), the above implies that distillation would require an average of 7.3 runs, with a total expected cost of approximately 117 photons.

We now briefly discuss the output error rate. In the presence of photon loss, the definition of output error rate given in Sec. III A is no longer well-defined, as it assumes there is always one output photon. In the more general setting, it is more natural to consider the *output fidelity* $f_n(\epsilon; \lambda)$ to be the probability that, given the heralding of an ideal pattern, there is a single output photon with internal state $|\xi_0\rangle$. (In other words, we count the probability of outputs with no photon loss *and* an ideal output photon.) The output error rate is then $e_n(\epsilon; \lambda) = 1 - f_n(\epsilon; \lambda)$. We note that $f_n(\epsilon; \lambda)$ is straightforwardly computable in terms of the quantities we have already discussed. In particular, since successful output requires no photon loss,

$$f_n(\epsilon; \lambda) = P(\text{no loss}|\text{herald})f_n(\epsilon; 0) \quad (47)$$

$$= \frac{(1 - \Lambda)^n h_n(\epsilon; 0)}{h_n(\epsilon; \lambda)} f_n(\epsilon; 0) \quad (48)$$

$$= \frac{(1 - \Lambda)f_n(\epsilon; 0)}{1 + n\Lambda c_n(\epsilon)/h_n(\epsilon)}, \quad (49)$$

where the second equality follows from Bayes' theorem and the third follows from expanding $h_n(\epsilon; \lambda)$ according to Eq. (43). To estimate the output fidelity, we recall that $h_n(0) \approx 1/4 \neq 0$ (see Theorem III.5) and $c_n(0) = 0$ (see the proof of Theorem III.17), and therefore $c_n(\epsilon)/h_n(\epsilon) = O(\epsilon)$. Expanding Eq. (49) as a Maclaurin series in ϵ , we obtain

$$\begin{aligned} f_n(\epsilon; \lambda) &= \frac{(1 - \Lambda)f_n(\epsilon; 0)}{1 + n\Lambda c_n(\epsilon)/h_n(\epsilon)} \\ &= (1 - \Lambda)f_n(\epsilon; 0) + O(n\Lambda\epsilon). \end{aligned} \quad (50)$$

Then up to *second-order* corrections involving both a photon loss and a distinguishability error, the output fidelity is simply the output fidelity in the lossless case multiplied by the probability that the output photon is not lost.

We may also phrase Eq. (47) as follows. Given successful heralding, at most one photon can be lost. If there is one loss, then there is no photon in the output mode. If there is no loss, then we reduce to the lossless case, Theorem III.2.

IV. DISCUSSION

We have provided a set of protocols for reducing photonic distinguishability errors, using n photons to reduce the error rate by a factor of n . This generalizes the work

of Refs. [4,8]. Our protocols, based on the Hadamard and Fourier unitaries (or more generally the Fourier transform on any finite Abelian group, as discussed in Appendix C), can scale to arbitrarily many photons (in contrast to Ref. [8]) and retain a high heralding probability near 1/4 even for large n (in contrast to Ref. [4]). These protocols are efficient in the sense that they require approximately $4n$ photons to reduce the error rate by a factor of n . This is a major improvement over the previous state-of-the-art protocol of Ref. [8], which requires $O(n^2)$ photons to obtain similar error reduction. We consider a detailed example in Remark III.10, demonstrating the greatly reduced resource costs in our setting. Further, to reduce the error rate by more than a factor of 4 without exponentially scaling costs, the protocols of Ref. [8] require an iterative scheme involving active feedforwarding and quantum memory during distillation, which can potentially introduce more errors and negate the benefits of distillation. These additional assumptions are not required for our protocols, which are therefore significantly more resource efficient *and* potentially simpler to implement experimentally.

We note that a major potential application of distillation protocols is to fault-tolerant linear optical quantum computation, for example, fusion-based quantum computation (FBQC). Distinguishability errors may often lead to measurement errors in FBQC and similar paradigms, and therefore reducing the distinguishability error rates can improve the thresholds for other errors. Increasing the photon-loss threshold is of particular relevance, since loss tends to be the dominant source of noise. In turn, these improved loss thresholds may result in significantly decreased resource requirements for FBQC. Since our distillation protocols are efficient, source-agnostic, and require only standard linear optical hardware, this motivates the use of distillation to improve distinguishability error rates beyond even those obtained by state-of-the-art single-photon sources. We discuss these considerations in greater detail in Sec. II C.

In this work, we established the theory for these efficient distillation protocols. We proved and conjectured various results, including exact formulas and lower bounds for heralding rates in Secs. III B and III D, characterization of the effect of photon loss in Sec. III F, and discussion of resource costs and optimal choices of distillation protocols in Secs. III C and III E. Much of this analysis was enabled by the symmetries present in the Fourier and Hadamard matrices [19] and, more generally, the Fourier transform associated with a finite Abelian group. This theoretical perspective explains earlier work on distillation [8], relates it to the study of suppression laws, and provides insight into the efficiency of the distillation protocols.

We also show in Appendix D that while distillation is *possible* using almost any unitary (in particular, for Haar random unitaries), it does not yield an efficiently

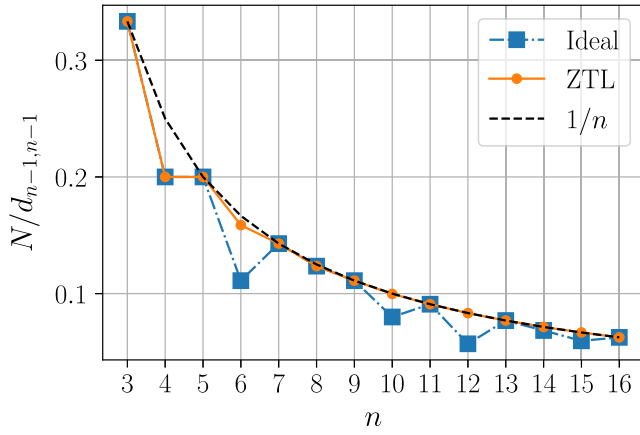


FIG. 12. For the Fourier matrices F_n , we plot the ratio of the total number N of ideal patterns (respectively, ZTL patterns with one photon in mode 0) to the total number of possible patterns $d_{n-1,n-1}$ (dimension of the space of $n-1$ photons in $n-1$ modes). As expected from Ref. [17], it scales like $1/n$. We can also see here that for prime powers, the ZTL patterns match the ideal ones (see Theorem III.11), but otherwise there is additional suppression.

scaling protocol. First we observe that in the Haar case (as with Fourier and Hadamard), the heralding rate for the ideal patterns asymptotically approaches $1/4$ (the “ideal” patterns in the Haar case are typically just all possible patterns, since there is no symmetry). However, in the Haar case, postselecting on all ideal patterns increases the error (approximately doubling it). Instead, Haar random unitaries can be used for distillation only when one chooses the postselection patterns as a small subset of ideal patterns with a high amount of constructive interference (high probability). Although the Haar case does not scale efficiently, these observations help to highlight how the Hadamard and Fourier protocols achieve efficiency; the symmetry causes a concentration on a relatively small number of possible patterns, as demonstrated in Fig. 12. This figure shows that in the Fourier case, the total proportion of ideal patterns (out of all possible ones) scales like $1/n$, as expected from analysis in Ref. [17]. Remarkably, the weight of these patterns remains approximately constant, at $1/4$, as proven in Theorem III.5.

We proved in Theorem III.11 that for n a prime power, the Fourier protocols (like the Hadamard protocols [18]) have ideal patterns exactly characterized by symmetry, resolving an open problem due to Ref. [17]. However, not all suppression laws come from symmetry conditions [35]. In particular, our numerical analysis in Sec. III E indicates that the optimal distillation protocols likely correspond to F_n where n is not a prime power. (See Fig. 6.) In those cases, again by Theorem III.11, the ideal patterns are a strict subset of those determined by the known symmetry conditions. Further, there seem to be additional properties determining which $n \neq p^r$ lead to optimal distillation

protocols: for example, F_n with n divisible by 3 seem to outperform other non-prime-power n . Thus we believe that future work on distillation should focus on understanding and taking advantage of the extra suppression occurring in these cases.

ACKNOWLEDGMENTS

We are grateful for support from the NASA SCan program, from DARPA under IAA 8839, Annex 130, and from NASA Ames Research Center. J.M. is thankful for support from NASA Academic Mission Services, Contract No. NNA16BD14C. N.A. is a KBR employee working under the Prime Contract No. 80ARC020D0010 with the NASA Ames Research Center. We also thank Vaclav Kotesovec for pointing out the proof of Theorem III.5 Part 2 discussed in Appendix B 1 b and Bojko Bakalov for suggesting the study of general Fourier transforms on finite Abelian groups, as discussed in Appendix C. The United States Government retains, and by accepting the article for publication, the publisher acknowledges that the United States Government retains, a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for United States Government purposes.

APPENDIX A: SYMMETRY

1. General case

In this section, we review an argument due to Ref. [19] to show that certain types of symmetries of a unitary U determine the transition probabilities of the associated linear optical unitary \hat{U} . Suppose we have an $n \times n$ unitary U . As above, we write $\hat{U}|\psi\rangle$ to denote the action of the corresponding linear optical unitary on the state $|\psi\rangle \in \mathcal{H}^{\otimes n}$. In this section, we only consider PNRD. As shorthand, we write $\langle s_0, \dots, s_{n-1} |$ for the projector onto the subspace of states with s_j photons in mode j . (Regardless of internal states.) Thus we will be interested in the transition probabilities $\langle s_0, \dots, s_{n-1} | \hat{U} | \psi \rangle$.

Suppose there is some $n \times n$ permutation matrix P such that $UP = DU$, where D is a diagonal matrix with diagonal entries d_0, \dots, d_{n-1} . (By unitarity, all $|d_i| = 1$.)

Remark A.1. Before continuing, we emphasize that the linear optical unitary \hat{P} is a permutation of the *external modes* of the input state, acting by $P^{\otimes n} \otimes I^{\otimes n}$ on $\mathcal{H}_{\text{ext}}^{\otimes n} \otimes \mathcal{H}_{\text{int}}^{\otimes n}$. These should not be confused with the permutations of the *photons* discussed in Sec. II A.

For simplicity, we begin with a special case in which the input state is fully indistinguishable and there are no internal modes. We have

$$\hat{U}|1, \dots, 1\rangle = \hat{U}\hat{P}\hat{P}^\dagger|1, \dots, 1\rangle = \hat{D}\hat{U}|1, \dots, 1\rangle, \quad (\text{A1})$$

since $|1, \dots, 1\rangle$ is unaffected by permutations of the modes. Then for any s_0, \dots, s_{n-1} , we have

$$\begin{aligned} \langle s_0 \cdots s_{n-1} | \hat{U} | 1, \dots, 1 \rangle \\ = \langle s_0 \cdots s_{n-1} | \hat{D} \hat{U} | 1, \dots, 1 \rangle \\ = d_0^{s_0} \cdots d_{n-1}^{s_{n-1}} \langle s_0 \cdots s_{n-1} | \hat{U} | 1, \dots, 1 \rangle. \end{aligned} \quad (\text{A2})$$

Since all d_i are nonzero, we conclude that either

$$d^s = 1 \text{ or } \langle s_0 \cdots s_{n-1} | \hat{U} | 1, \dots, 1 \rangle = 0, \quad (\text{A3})$$

where $d^s := d_0^{s_0} \cdots d_{n-1}^{s_{n-1}}$. Thus, the output patterns arising with nonzero probability must satisfy $d^s = 1$. These precisely correspond to the known suppression laws for the Fourier and Hadamard cases [17, 18], discussed below.

Note that the only property of $|1, \dots, 1\rangle$ we needed was its invariance under certain permutations of its modes. More generally, suppose that $|\psi\rangle \in \mathcal{H}^{\otimes n}$ is a general pure state of n photons in n external modes, potentially distinguishable. Assume that $P|\psi\rangle = e^{i\phi}|\psi\rangle$. Then the same argument gives

$$\begin{aligned} \langle s_0 \cdots s_{n-1} | \hat{U} | \psi \rangle \\ = \langle s_0 \cdots s_{n-1} | \hat{U} \hat{P} \hat{P}^\dagger | \psi \rangle = e^{-i\phi} \langle s_0 \cdots s_{n-1} | \hat{D} \hat{U} | \psi \rangle \\ = e^{-i\phi} d^s \langle s_0 \cdots s_{n-1} | \hat{U} | \psi \rangle. \end{aligned} \quad (\text{A4})$$

We conclude that either $d^s = e^{i\phi}$ or $\langle s_0 \cdots s_{n-1} | \hat{U} | \psi \rangle = 0$. We summarize in the following lemma.

Lemma A.2 ([19]). Let U , P , and D be $n \times n$ unitary matrices, with P a permutation matrix and D diagonal with diagonal entries d_0, \dots, d_{n-1} . Suppose $UP = DU$. Let $|\psi\rangle$ be a (pure) state of n photons in n modes satisfying $\hat{P}|\psi\rangle = e^{i\phi}|\psi\rangle$. If $\langle s_0, \dots, s_{n-1} | \hat{U} | \psi \rangle \neq 0$, we have

$$d^s = e^{i\phi}. \quad (\text{A5})$$

We now apply this result to the calculation of certain transition probabilities. Let

$$\mathcal{S} = \{|s_0, \dots, s_{n-1}\rangle : \sum s_j = n, d^s = 1\} \quad (\text{A6})$$

be the set of output patterns arising with nonzero probability from P -symmetric input. As before, we have that $d^s := d_0^{s_0} \cdots d_{n-1}^{s_{n-1}}$. We decompose the state space into eigenspaces for P with eigenvalues $e^{i\phi}$. For any $|\psi'\rangle$ with eigenvalue $e^{i\phi} \neq 1$ and $|s\rangle \in \mathcal{S}$, the lemma implies $\langle s | \hat{U} | \psi' \rangle = 0$. Then, for arbitrary $|\eta\rangle$, decompose

$$|\eta\rangle = \langle \eta_+ | \eta \rangle |\eta_+\rangle + |\eta_\perp\rangle, \quad (\text{A7})$$

where $|\eta_+\rangle$ is the unit vector projection of $|\eta\rangle$ into the $+1$ eigenspace and $|\eta_\perp\rangle$ collects the terms from other

eigenspaces. We have, for $|s\rangle \in \mathcal{S}$,

$$\langle s | \hat{U} | \eta \rangle = \langle s | \hat{U} (\langle \eta_+ | \eta \rangle |\eta_+\rangle + |\eta_\perp\rangle) = \langle \eta_+ | \eta \rangle \langle s | \hat{U} | \eta_+\rangle. \quad (\text{A8})$$

More generally, consider an Abelian group G of $n \times n$ permutation matrices, where each $P \in G$ has corresponding diagonal matrix D_P with $UP = D_P U$. (In other words, writing $UPU^\dagger = D_P$, we see that G is a group of simultaneously diagonalizable permutation matrices, and the change of basis is given by U .) Let $d_{P,i}$ index the diagonal entries of D_P , with d_P^s extending the d^s notation above. We consider the set of output patterns that occur with nonzero probability for all $P \in G$:

$$\mathcal{S}_G = \{|s\rangle : \forall P \in G, d_P^s = 1\}. \quad (\text{A9})$$

We may generalize the above argument to obtain the following, Part 1 of Theorem III.12:

Lemma A.3. Let all notation be as above. Let $|\eta\rangle$ be an arbitrary pure state, and let $|\eta_+\rangle$ be its projection into the joint $+1$ eigenspace of all $P \in G$, normalized to be a unit vector (if nonzero). For $|s\rangle \in \mathcal{S}_G$, we have

$$\langle s | \hat{U} | \eta \rangle = \langle \eta_+ | \eta \rangle \langle s | \hat{U} | \eta_+\rangle. \quad (\text{A10})$$

Further, if $|s\rangle \notin \mathcal{S}_G$ and $\langle \eta_+ | \eta \rangle \neq 0$, then $\langle s | \hat{U} | \eta \rangle = 0$.

Proof. Since G is Abelian, we may decompose $\mathcal{H}^{\otimes n}$ into a direct sum of joint eigenspaces for G . We index these eigenspaces V_α by functions $\alpha : G \rightarrow \mathbb{C}^*$, such that for $P \in G$ and $|\eta_\alpha\rangle \in V_\alpha$, we have

$$P |\eta_\alpha\rangle = \alpha(P) |\eta_\alpha\rangle. \quad (\text{A11})$$

Note that if $\alpha \neq \beta$, then V_α is orthogonal to V_β . In particular, given $|\eta\rangle$ as above, we may decompose

$$|\eta\rangle = \sum_\alpha c_\alpha |\eta_\alpha\rangle, \quad (\text{A12})$$

where the set of all nonzero $|\eta_\alpha\rangle$ is an orthonormal set and $c_\alpha = \langle \eta_\alpha | \eta \rangle$. The term $|\eta_+\rangle$ above corresponds to α trivial, $\alpha(P) = 1$ for all $P \in G$. Now consider $|s\rangle \in \mathcal{S}_G$, so that all $d_P^s = 1$. If α is nontrivial, find P such that $\alpha(P) \neq 1$. We have $d_P^s = 1 \neq \alpha(P)$, so by Lemma A.2, $\langle s | \hat{U} | \eta_\alpha \rangle = 0$. This gives the first claim; the second is a direct consequence of Lemma A.2. ■

In other words, to calculate the output probabilities of patterns in \mathcal{S}_G , it suffices to consider the projection of $|\eta\rangle$ into the space of G -symmetric vectors. Since $|1, \dots, 1\rangle$ is fully symmetric, we observe the following.

Lemma A.4. Let all notation be as above. Let $|s\rangle$ satisfy $\langle s|\hat{U}|1, \dots, 1\rangle \neq 0$. Then $|s\rangle \in \mathcal{S}_G$.

In particular, the *ideal patterns* discussed in Sec. III A satisfy $|s\rangle = |s_0, \dots, s_{n-1}\rangle$, with $s_0 = 1$ and $\langle s|\hat{U}|1, \dots, 1\rangle \neq 0$. Then the ideal patterns are a subset of \mathcal{S}_G ; to calculate $\langle s|\hat{U}|\eta\rangle$ for an ideal pattern $|s\rangle$, it suffices to consider $\langle s|\hat{U}|\eta_+\rangle$.

For later use, we consider a special case. Assume that for all $P \in G$, we have $\langle \eta|P|\eta\rangle \in \{0, 1\}$, so that each permutation either fixes $|\eta\rangle$ or takes it to an orthogonal state. Let K be the *stabilizer* of $|\eta\rangle$, the subgroup of G with $P|\eta\rangle = |\eta\rangle$ for $P \in K$. Letting G/K be the set of left cosets of K in G , we have

$$|\eta_+\rangle = \frac{1}{\sqrt{|G/K|}} \sum_{g \in G/K} g|\eta\rangle. \quad (\text{A13})$$

Note that

$$\langle \eta_+|\eta\rangle = \frac{1}{\sqrt{|G/K|}} = \sqrt{|K|/|G|} \geq \frac{1}{\sqrt{|G|}}. \quad (\text{A14})$$

2. Fourier case

One of the central examples we consider is the Fourier case, $U = F_n$. We use the notation $\omega = e^{2\pi i/n}$. Take the permutation matrix P to correspond to the cyclic permutation sending column j to column $j + 1$ (modulo n). We get $UP = DU$, where $d_i = \omega^i$. The group G of symmetries is the cyclic group generated by P . Then the set $\mathcal{S} = \mathcal{S}_G$ consists of $|s_0, \dots, s_{n-1}\rangle$ with

$$1 = \prod_i \omega^{-is_i} = \omega^{-\sum_i is_i}, \quad (\text{A15})$$

or equivalently $\sum_i is_i \equiv 0 \pmod{n}$. This is the zero transmission law (ZTL) for the Fourier transform [17]. We call the patterns in \mathcal{S} *ZTL patterns*. This is typically rephrased as in Sec. III A above, by letting g_0, \dots, g_{n-1} be the weakly decreasing sequence for which s_i counts the number of j with $g_j = i$. The ZTL then translates to $\sum_i g_i \equiv 0 \pmod{n}$.

As an example, consider $n = 2$, so that $U = F_2 = H_2 = H$. Only the patterns $s = (2, 0), (0, 2)$ satisfy the ZTL: this is precisely the HOM effect. (Although note that neither of these are ideal patterns, since neither has $s_0 = 1$.) Further, Lemma A.2 explains why \hat{H} must map the symmetric states $|1, 1\rangle$ and $\frac{1}{\sqrt{2}}(|2, 0\rangle + |0, 2\rangle)$ to $\text{span}\{|2, 0\rangle, |0, 2\rangle\}$, while mapping the antisymmetric state $\frac{1}{\sqrt{2}}(|2, 0\rangle - |0, 2\rangle)$ to $|1, 1\rangle$ (the only non-ZTL pattern of two photons in two modes). Thus the ZTL may be viewed as a direct extension of the HOM effect.

By Lemma A.4, the ideal patterns must be a subset of \mathcal{S} . However, as discussed in Sec. III A above, there exist ZTL patterns $|s\rangle$ with $s_0 = 1$ that are *not* ideal patterns. (This is discussed in the original paper [17] without

the $s_0 = 1$ restriction.) In other words, these patterns satisfy the required symmetry conditions but still cannot be obtained from the input state $|1, \dots, 1\rangle$. The first example arises for $n = 6$, where there are six ZTL patterns $|s\rangle$ with $s_0 = 1$ that are not ideal patterns. These are

$$(1, 0, 1, 1, 2, 1), (1, 0, 2, 0, 1, 2), (1, 1, 0, 1, 1, 2), \\ (1, 1, 2, 1, 1, 0), (1, 2, 1, 0, 2, 0), (1, 2, 1, 1, 0, 1). \quad (\text{A16})$$

In the following section, we characterize exactly when this occurs.

a. Fourier case for prime powers

In this section, we prove the following theorem.

Theorem A.5. Let $n \geq 2$ be a positive integer. Let g_0, \dots, g_{n-1} be a list of integers in $0, \dots, n-1$, and let $|s\rangle = |s_0, \dots, s_{n-1}\rangle$ be the corresponding Fock state, with s_i counting the number of j with $g_j = i$. Let A be the $n \times n$ matrix whose i th row is the g_i th row of $\sqrt{n}F_n$.

(1) Further assume $n = p^r$, where p is a prime. Then $\text{perm}(A) \neq 0$ if and only if $\sum_i g_i \equiv 0 \pmod{n}$.

(2) Instead assume that n is *not* a prime power. Then there exists a choice of g_0, \dots, g_{n-1} satisfying $\sum_i g_i \equiv 0 \pmod{n}$ and $s_0 = 1$ such that $\text{perm}(A) = 0$.

By the permanent formulation of boson sampling [38], we see that for n a prime power and $|s\rangle$ satisfying the zero transmission law, we have $\langle s|U|1, \dots, 1\rangle \neq 0$. We immediately obtain the following.

Corollary A.6. If $n = p^r$, where p is a prime, then for the distillation protocol corresponding to $U = F_n$, all ZTL patterns with $s_0 = 1$ are ideal patterns. If n is not a prime power, then there exist ZTL patterns with $s_0 = 1$ that are *not* ideal patterns.

This gives Theorem III.11. Before the proof, we briefly introduce some additional notation. Let x_0, \dots, x_{n-1} be indeterminates; for an n -tuple $s = (s_0, \dots, s_{n-1})$, let $x^s = x_0^{s_0} x_1^{s_1} \dots x_{n-1}^{s_{n-1}}$. We will use the *generic circulant matrix* C , the $n \times n$ matrix with (i, j) entry equal to x_{i+j} (where the indices are taken modulo n). We now prove Theorem A.5, a direct application of results of Refs. [39, 40]. We note that these works make no explicit reference to boson sampling or the matrix A defined above: instead, their results are related to our setting by Eq. (A17) below.

Proof. Let all notation be as in the theorem. The determinant of the circulant matrix, $\det(C)$, is a polynomial in

x_0, \dots, x_{n-1} , and the coefficient of x^s in $\det(C)$ is [39]

$$[x^s] \det(C) = \sum_{\sigma \in S_n} \omega^{\sum_{j=0}^{n-1} j g_{\sigma(j)}} = \text{perm}(A). \quad (\text{A17})$$

We note that only the first equality is observed in Ref. [39], in the proof of the main theorem; the second equality of Eq. (A17) is immediate from the definition of the permanent. Now, Ref. [39] shows that for n a prime power, $\sum_i g_i \equiv 0 \pmod n$ if and only if Eq. (A17) is nonzero. This proves Part 1 of the theorem. For Part 2, suppose that n is not a prime power. Then Theorem 3.5 of Ref. [40] explicitly constructs a list of non-negative integers g_0, \dots, g_{n-1} satisfying

- (1) $\sum_i g_i \equiv 0 \pmod n$,
- (2) the coefficient of x^s in $\det(C)$ is 0.

The first condition is the ZTL; the second shows that the corresponding permanent vanishes, by Eq. (A17). Then we must only show that there is some g_i with multiplicity 1, i.e., the corresponding s_j is equal to 1. Once this is proven, then by the cyclic symmetry of the Fourier transform we obtain an example with $s_0 = 1$. [More explicitly, we may note that if we perform the transformation $g_i \mapsto g_i + 1$ for all i , the condition $\sum_i g_i \equiv 0 \pmod n$ is unchanged, and Eq. (A17) will differ only by a factor of $\omega^{\sum_{j=0}^{n-1} j} = \pm 1$. This operation on the g_i corresponds to cyclically permuting the s_j ; thus, we may do this repeatedly, without changing $|\text{perm}(A)|$, until we obtain an example with $s_0 = 1$.] It remains to show the relevant properties of the g_i . For this, we must give more details of the construction of Ref. [40]. We leave the proof that the state we construct has vanishing permanent to Ref. [40], which uses Corollary 6 of Ref. [41]. Suppose that $n = q_1 q_2$, where $q_1, q_2 \geq 2$ are relatively prime. By Bezout's theorem (exchanging the roles of q_1, q_2 if necessary), we find positive integers c_1, c_2 with $c_1 q_1 = 1 + c_2 q_2$, where $1 \leq c_1 < q_2$ and $1 \leq c_2 < q_1$. We then define

$$\begin{aligned} M_1 &= c_2 q_2 - 1 \\ M_0 &= n - c_2 q_2 - 2 \\ A_1 &= c_2 q_2 - c_2 c_1 + 1 \\ A_2 &= n - c_2 q_2 \\ A_3 &= n - c_2 q_2 + c_1 c_2. \end{aligned}$$

We may easily verify that $A_1 > 1$ and

$$\begin{aligned} A_3 - A_2 &= c_1 c_2 > 0, \\ A_2 - A_1 &= (q_2 - c_1)(q_1 - c_2) > 0, \end{aligned}$$

so we have $1 < A_1 < A_2 < A_3$. The state is then constructed by

$$\begin{aligned} g_0 &= g_1 = \dots = g_{M_0-1} = 0, \\ g_{M_0} &= g_{M_0+1} = \dots = g_{M_0+M_1-1} = 1, \\ g_{n-3} &= A_1, \quad g_{n-2} = A_2, \quad g_{n-1} = A_3, \end{aligned}$$

where we note $M_0 + M_1 = n - 3$. In particular, there are M_0 photons in mode 0, M_1 photons in mode 1, and additional photons in the three distinct modes $A_1, A_2, A_3 > 1$. Then we have all $s_{A_j} = 1, j = 1, 2, 3$. This completes the proof. ■

We now give the $n = 6$ example of the construction of Ref. [40]. We write $q_1 = 3, q_2 = 2$, so that $6 = q_1 q_2$. Note that with $c_1 = c_2 = 1$, we have $c_1 q_1 = 1 + c_2 q_2 = 3$. We obtain $M_0 = 2, M_1 = 1, A_1 = 2, A_2 = 4, A_3 = 5$. This gives

$$g_0 = g_1 = 0, \quad g_2 = 1, \quad g_3 = 2, \quad g_4 = 4, \quad g_5 = 5, \quad (\text{A18})$$

with corresponding Fock state $|s\rangle = |2, 1, 1, 0, 1, 1\rangle$. We cyclically permute this state to see that $|1, 1, 0, 1, 1, 2\rangle$ is another such example, now with one photon in the 0th mode. We note that this is one of the examples given above. As another example, for $n = 18$ we obtain $|s\rangle = |1, 0, 0, 0, 7, 8, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0\rangle$.

Remark A.7. We note that the proof of Theorem 3.5 of Ref. [40] has a typographical error in its first paragraph. In particular, the a_i there (equivalent to our g_i) are said to satisfy $\sum_i i a_i \equiv 0 \pmod n$ and $\sum_i a_i = n$, which are in fact the conditions on our s_i . These conditions are not used in the proof, and from context it is clear that this was a simple typographical error: elsewhere in *ibid.*, the notation a_i is used for the g_i and they are said to satisfy the usual ZTL $\sum_i a_i \equiv 0 \pmod n$; and the notation α_i or M_i is used for our s_i . [See the discussion between Eq. (3) and Proposition 3.2 of *ibid.*] The proof of the theorem reduces to checking the conditions given in Corollary 6 of Ref. [41], which is in fact stated in terms of our g_i .

To obtain Eq. (A17), one simply notes that the eigenvalues of the generic circulant matrix C have the form $\sum_j \omega^{ij} x_j$, so that

$$\det(C) = \prod_i \sum_j \omega^{ij} x_j. \quad (\text{A19})$$

For terms involving x^s , one needs to sum over all possible products of the form

$$\omega^{0j_0} x_{j_0} \omega^{1j_1} x_{j_1} \dots \omega^{(n-1)j_{n-1}} x_{j_{n-1}} \quad (\text{A20})$$

in which there are s_0 copies of x_0 , s_1 copies of x_1 , and so on. The possible values of the sequence j_k are precisely the permutations of g_0, \dots, g_{n-1} , giving Eq. (A17).

We now note further consequences of Eq. (A17). Since $\det(C)$ must be a polynomial in the x_i with integer coefficients, we immediately see that $\text{perm}(A)$ is an integer. Then the weights $\langle s | F_n | 1, \dots, 1 \rangle$ are all integers up to a known normalization factor of $n^n (s_0!) \cdots (s_{n-1}!)$. (Recall that we defined A in terms of $\sqrt{n} F_n$, explaining the factor of n^n . The factorials come from the permanent formulation of boson sampling [38].) Further, we note that for a fixed n , Eq. (A17) allows one to compute *all* $\langle s | U | 1, \dots, 1 \rangle$, generally expressed in terms of permanents of $n \times n$ matrices, through the calculation of the determinant of a single $n \times n$ matrix (with symbolic entries). Equation (A17) also allows for the connection of boson sampling with the literature on the coefficients of $\det(C)$, which offers various combinatorial formulas such as those in Ref. [41], which were instrumental in proving Theorem A.5. We note that the relevant permanents $\text{perm}(A)$ were previously related with $\det(C)$ in Ref. [42], in the context of the three-dimensional Ising model.

3. Hadamard case

We now consider the Hadamard case. We let $n = 2^r$ and identify $\mathcal{H}_{\text{ext}} = \mathbb{C}^n$ with $(\mathbb{C}^2)^{\otimes n}$ by representing kets by binary strings. We take $U = H_n = H^{\otimes r}$, where H is the 2×2 Hadamard matrix. Let X and Z be the 2×2 Pauli matrices. Of course, H satisfies $HX = ZH$. This is, in fact, a special case of the framework above, since X is a permutation matrix and Z is diagonal. Following Eq. (A6), we find that in the case $n = 2$, \mathcal{S} consists of patterns $|s_0, s_1\rangle$ with $s_0 + s_1 = 2$ and $(-1)^{s_1} = 1$. In other words, $s_0, s_1 \in \{0, 2\}$, recovering the classic Hong-Ou-Mandel effect.

We may extend this analysis to obtain symmetries of H_n by taking tensor products. We consider the following symmetry group G of U :

$$G = \{X^{i_1} \otimes X^{i_2} \otimes \cdots \otimes X^{i_r} : i_j \in \{0, 1\}\}. \quad (\text{A21})$$

These satisfy

$$H_n(X^{i_1} \otimes X^{i_2} \otimes \cdots \otimes X^{i_r}) = (Z^{i_1} \otimes \cdots \otimes Z^{i_r})H_n. \quad (\text{A22})$$

We will compute S_G , defined as above. Letting X_i and Z_i be the actions of X and Z on the i th tensor factor, we note that G is generated by X_1, \dots, X_r and $H_n X_i = Z_i H_n$. To proceed, we will write down the map $\mathbb{H}_{\text{ext}} = \mathbb{C}^n \rightarrow (\mathbb{C}^2)^{\otimes n}$ explicitly: for $k = \sum_{j=0}^{n-1} c_j 2^j$, we map $|k\rangle \mapsto |c_0, \dots, c_{n-1}\rangle$. Then $Z_i |k\rangle = (-1)^{c_i} |k\rangle$. We now consider what the X_i symmetry tells us about $|s\rangle = |s_0, \dots, s_n\rangle \in \mathcal{H}_{\text{ext}}^{\otimes n}$. As discussed in Sec. III A, we convert to “first quantization,” letting g_0, \dots, g_{n-1} be the weakly increasing sequence with s_0 copies of 0, s_1 copies of 1, etc. The g_j can be viewed

as the list of modes occupied by the n photons. Let $(g_j)_i$ be the coefficient of 2^i in the binary expansion of s_j . By Lemma A.2, we see that $|s\rangle \in S_G$ only if $\sum_j (g_j)_i \equiv 0 \pmod{2}$. [In other words, $(g_0)_i \oplus \cdots \oplus (g_{n-1})_i = 0$, where \oplus represents the XOR of two bits.] Since we may consider symmetries arising from all X_i , this must hold for all i . Then we recover the characterization of S_G discussed in Ref. [18] and Sec. III A above: extending \oplus to represent the bitwise XOR operation, $|s\rangle \in S_G$ if and only if $g_0 \oplus \cdots \oplus g_{n-1} = 0$.

Unlike the case of the Fourier transform, in the Hadamard case the symmetries given fully characterize the ideal patterns. In other words, the ideal patterns for $U = H_n$ are the elements of S_G with $s_0 = 1$. (This is proven in Ref. [18] without the additional condition $s_0 = 1$.)

APPENDIX B: DETAILED CALCULATIONS

Here we discuss the calculations for heralding probabilities and error rates in detail. Let U be the n -mode Fourier or Hadamard matrix, with corresponding symmetry group G . Consider the URS model with parameter R , or the limiting OBB model. Recall the decomposition Eq. (15) of the n -photon state with error rate ϵ :

$$\rho^{(n)}(\epsilon) = \sum_{k=0}^n \binom{n}{k} \epsilon^k (1 - \epsilon)^{n-k} \Phi_k, \quad (\text{B1})$$

where Φ_k is the (normalized) uniform mixture of all states

$$a_0^\dagger[\xi_{j_0}] \cdots a_{n-1}^\dagger[\xi_{j_{n-1}}] |\vec{0}\rangle = |1, \dots, 1\rangle_{2Q} \otimes |\xi_{j_0}, \dots, \xi_{j_{n-1}}\rangle \quad (\text{B2})$$

[recalling Eq. (14)] with k distinguishability errors, meaning that k of the indices j_0, \dots, j_{n-1} are nonzero. For example, in the $R = 1$ (SBB) case, Φ_k has $\binom{n}{k}$ terms, corresponding to a choice of k distinguishable photons. For general R , Φ_k has $\binom{n}{k} k^R$ terms; first we choose the k distinguishable photons, then each one has R choices of internal state. In the OBB limit, since each photon has a unique error state, Φ_k again has $\binom{n}{k}$ terms to consider, corresponding to a choice of k distinguishable photons.

In the following subsections, we will consider detailed computations with the heralding and error rates, of both theoretical and computational significance. For practical purposes, we will see that in order to calculate these rates as an *exact* function of ϵ , one must only compute, for all $|\eta\rangle$ of the form (B2), $h_n(|\eta\rangle\langle\eta|)$ and $e_n(|\eta\rangle\langle\eta|)$.

1. Heralding rate

First we discuss the heralding rate. As discussed in Sec. III B, we can calculate the heralding rate via

$$h_n(\epsilon) = \sum_{k=0}^n \binom{n}{k} \epsilon^k (1-\epsilon)^{n-k} h_n(\Phi_k). \quad (\text{B3})$$

Since Φ_k is a uniform mixture, $h_n(\Phi_k)$ is simply the average heralding rate for states of the form Eq. (B2) with k distinguishability errors. We discuss the efficient calculation of the $h_n(\Phi_k)$ via simulation in Appendix E.

As observed in Lemma A.4, the ideal patterns are a subset of those in \mathcal{S}_G with $s_0 = 1$. By Lemma A.3, then, we have

$$\begin{aligned} h_n(|\eta\rangle\langle\eta|) &= \sum_{|s\rangle \text{ ideal}} \left| \langle s | \hat{U} |\eta\rangle \right|^2 \\ &= |\langle \eta_+ | \eta \rangle|^2 \sum_{|s\rangle \text{ ideal}} \left| \langle s | \hat{U} |\eta_+\rangle \right|^2 = h_n(|\eta_+\rangle\langle\eta_+|). \end{aligned} \quad (\text{B4})$$

This gives the heralding rate statement in Part 2 of Theorem III.12. To prove the corresponding statement in Part 3, we consider the sum over ideal patterns. In the Hadamard case, the ideal patterns are exactly characterized by the conditions $|s\rangle \in \mathcal{S}_G$, $s_0 = 1$. In the Fourier case, this no longer holds in general, although in many cases, the patterns satisfying $\langle s | \hat{U} |\eta\rangle \neq 0$ and $s_0 = 1$ are guaranteed to be ideal. This includes when $|\eta\rangle$ has 0 or 1 distinguishability errors, as in Appendix B 3 below. This also holds when n is a prime power, as proven in Appendix A 2 a. (Also see the discussion of Remark III.13.) Thus, going forward, we will assume that for all patterns $|s\rangle \in \mathcal{S}_G$ with $s_0 = 1$ and $\langle s | \hat{U} |\eta\rangle \neq 0$, we have $|s\rangle$ ideal. Then from this assumption and Lemma A.2, we may include extra vanishing terms to extend the sum to all $|s\rangle$ with $s_0 = 1$:

$$h_n(|\eta\rangle\langle\eta|) = |\langle \eta_+ | \eta \rangle|^2 \sum_{|s\rangle: s_0=1} \left| \langle s | \hat{U} |\eta_+\rangle \right|^2 \quad (\text{B5})$$

$$= |\langle \eta_+ | \eta \rangle|^2 \left| \langle (1| \otimes I) \hat{U} |\eta_+\rangle \right|^2 \quad (\text{B6})$$

$$= |\langle \eta_+ | \eta \rangle|^2 \langle \eta_+ | \hat{U}^\dagger (|1\rangle\langle 1| \otimes I) \hat{U} |\eta_+\rangle. \quad (\text{B7})$$

This allows for the computation of heralding probabilities without needing to check for ideal patterns. This gives the heralding rate statement in Part 3 of Theorem III.12.

We may further simplify as follows. We have $\langle 1| = \langle \vec{0}| a_0$, where $a_0 = \sum_i a_0[\xi_i]$ is the external annihilation operator (which is summed over all ξ_i so that it annihilates photons regardless of distinguishability). Then $|1\rangle\langle 1| \otimes$

$I = a_0^\dagger (|0\rangle\langle 0| \otimes I) a_0$. Since the entries in the first row of U are uniformly equal to $1/\sqrt{n}$, we have

$$\hat{U}^\dagger a_0^\dagger = \frac{1}{\sqrt{n}} \sum_i a_i^\dagger \hat{U}^\dagger \quad (\text{B8})$$

and thus

$$\begin{aligned} h_n(|\eta\rangle\langle\eta|) &= \frac{1}{n} |\langle \eta_+ | \eta \rangle|^2 \sum_{i,j} \langle \eta_+ | a_i^\dagger \hat{U}^\dagger (|0\rangle\langle 0| \otimes I) \hat{U} a_j |\eta_+\rangle. \end{aligned} \quad (\text{B9})$$

We note that $\hat{U}^\dagger (|0\rangle\langle 0| \otimes I) \hat{U}$ is a product of three operators, each of which acts in a “linear optical” way, transforming each photon independently (and all acting only on external modes). For \hat{U} , \hat{U}^\dagger this is by definition; meanwhile, $|0\rangle\langle 0| \otimes I$ acts on each creation operator by $a_i^\dagger \mapsto \delta_{i \neq 0} a_i^\dagger$. Then the product has the same linear optical form. By the unitarity of U and (again) the fact that the entries in its first row are uniformly equal to $1/\sqrt{n}$, we may write

$$\hat{U}^\dagger (|0\rangle\langle 0| \otimes I) \hat{U} = \hat{\Pi}, \quad (\text{B10})$$

recalling the notation \hat{T} for an operator (not necessarily unitary) acting independently on each photon, as introduced in Sec. II A. Here, the operator Π on \mathcal{H} is defined by $\Pi = I - 1/n \mathbf{1}$, where $\mathbf{1}$ satisfies $\mathbf{1} a_i^\dagger = \left(\sum_{j=0}^{n-1} a_j^\dagger \right) \mathbf{1}$. In particular, $1/n \mathbf{1}$ is the projection into the mode-symmetric subspace of \mathcal{H} , and Π is the complementary projection. We have

$$h_n(|\eta\rangle\langle\eta|) = \frac{1}{n} |\langle \eta_+ | \eta \rangle|^2 \sum_{i,j} \langle \eta_+ | a_i^\dagger \hat{\Pi} a_j |\eta_+\rangle. \quad (\text{B11})$$

a. Ideal heralding

We now compute $h_n(0)$ by considering the special case of Eq. (B11) with $|\eta\rangle = |\eta_+\rangle = |1, \dots, 1\rangle_{2Q}$, a perfectly indistinguishable Fock state. More generally, these results apply to any pure tensor of $|1, \dots, 1\rangle_{2Q}$ with an internal state. This will be used in Appendix B 3 below. We also note that the above assumptions involving the relationship between ideal and symmetry-preserving patterns are not needed here. In fact, we will only assume that U is an $n \times n$ unitary matrix with all entries in its first row equal to $1/\sqrt{n}$. By definition, the ideal patterns are those with $s_0 = 1$ and $\langle s | \hat{U} |\eta\rangle \neq 0$, so we may immediately express the heralding rate in the form Eq. (B7), then simplify as above to obtain Eq. (B11).

Now, we write

$$a_j |1, \dots, 1\rangle = \prod_{r \neq j} a_r^\dagger |\vec{0}\rangle. \quad (\text{B12})$$

Applying $\hat{\Pi}$, we obtain

$$\hat{\Pi} a_j |1, \dots, 1\rangle = \prod_{r \neq j} \left(a_r^\dagger - \frac{1}{n} \sum_v a_v^\dagger \right) |\vec{0}\rangle. \quad (\text{B13})$$

Then the heralding rate becomes

$$h_n(0) = \frac{1}{n} \sum_{i,j} \langle \vec{0} | \prod_{r' \neq i} a_{r'} \prod_{r \neq j} \left(a_r^\dagger - \frac{1}{n} \sum_v a_v^\dagger \right) |\vec{0}\rangle. \quad (\text{B14})$$

We note that the terms in this sum are invariant up to relabeling of the modes. Then each term may be reduced to one of two cases, $i = j = 0$ or $i = 1, j = 0$. The former occurs n times and the latter $n^2 - n = n(n-1)$ times. We obtain

$$h_n(0) = \langle \vec{0} | \prod_{r' \neq 0} a_{r'} \prod_{r \neq 0} \left(a_r^\dagger - \frac{1}{n} \sum_v a_v^\dagger \right) |\vec{0}\rangle \quad (\text{B15})$$

$$+ (n-1) \langle \vec{0} | \prod_{r' \neq 1} a_{r'} \prod_{r \neq 0} \left(a_r^\dagger - \frac{1}{n} \sum_v a_v^\dagger \right) |\vec{0}\rangle. \quad (\text{B16})$$

Labeling the terms in Eqs. (B15) and (B16) as B_n and L_n , respectively, we have the following.

Lemma B.1.

$$B_n = \left(\frac{-1}{n} \right)^{n-1} (n-1)! \sum_{t=0}^{n-1} \frac{(-n)^t}{t!} \quad (\text{B17})$$

$$L_n = \left(\frac{-1}{n} \right)^{n-1} (n-1)! \sum_{t=0}^{n-1} (n-t-1) \frac{(-n)^t}{t!}. \quad (\text{B18})$$

Note that the lemma immediately gives

$$h_n(0) = B_n + L_n = \left(\frac{-1}{n} \right)^{n-1} (n-1)! \sum_{t=0}^{n-1} (n-t) \frac{(-n)^t}{t!}, \quad (\text{B19})$$

which proves Theorem III.5. In the remainder of this section, we will prove the lemma, which is a straightforward counting argument. We focus on B_n , with L_n being similar.

We view B_n as the inner product of $a_1^\dagger \cdots a_{n-1}^\dagger |\vec{0}\rangle$ and $\prod_{r>0} \left(a_r^\dagger - \frac{1}{n} \sum_v a_v^\dagger \right) |\vec{0}\rangle$. In particular, we expand the

latter and sum the coefficients involving exactly one $a_{r'}^\dagger$ for all $r' > 0$ (and no a_0^\dagger). Each term in the expansion is a product of some number t of the a_r^\dagger and $n-1-t$ other a_v^\dagger chosen from some $1/n \sum_v a_v^\dagger$. These terms vanish in the inner product unless the a_r^\dagger all have $r > 0$ and the a_v^\dagger involve the remaining $n-1-t$ nonzero indices. Then we expand as follows, where t is the number of a_r^\dagger used, as above, and c_t is the number of such nonvanishing terms:

$$B_n = \langle \vec{0} | a_1 \cdots a_{n-1} \prod_{r>0} \left(a_r^\dagger - \frac{1}{n} \sum_v a_v^\dagger \right) |\vec{0}\rangle \quad (\text{B20})$$

$$= \langle \vec{0} | a_1 \cdots a_{n-1} \sum_{t=0}^{n-1} \left(\frac{-1}{n} \right)^{n-1-t} c_t a_1^\dagger \cdots a_{n-1}^\dagger |\vec{0}\rangle \quad (\text{B21})$$

$$= \sum_{t=0}^{n-1} \left(\frac{-1}{n} \right)^{n-1-t} c_t. \quad (\text{B22})$$

We note there are $\binom{n-1}{t}$ ways to choose t distinct values of a_r^\dagger , then $(n-1-t)!$ ways to choose which factors the remaining a_v^\dagger come from. This gives

$$\begin{aligned} \left(\frac{-1}{n} \right)^{n-1-t} c_t &= \left(\frac{-1}{n} \right)^{n-1-t} \binom{n-1}{t} (n-1-t)! \\ &= \left(\frac{-1}{n} \right)^{n-1-t} \frac{(n-1)!}{t!} \\ &= \left(\frac{-1}{n} \right)^{n-1} (n-1)! \frac{(-n)^t}{t!}, \end{aligned} \quad (\text{B23})$$

proving the result for B_n .

b. Ideal heralding asymptotics

We now prove Theorem III.5 Part 2 giving the asymptotic formula for $h_n(0)$, following the proof of Kotesovec. We let $W(z)$ be the Lambert W function, the principal branch of the multivalued solution to the equation $W(z)e^{W(z)} = z$.

First, by Theorem III.5 Part 1, we have

$$n^{n-1} h_n(0) = (-1)^{n-1} (n-1)! \sum_{t=0}^{n-1} (n-t) \frac{(-n)^t}{t!}. \quad (\text{B24})$$

For $n \geq 1$, elementary algebraic manipulation recovers the formula for the exponential generating function of $-1/(1-W(-z))$ given on the OEIS [31]. In other words,

expanding

$$\frac{-1}{1 - W(-z)} = \sum_{n \geq 0} w_n z^n, \quad (\text{B25})$$

we have

$$h_n(0) = \frac{n!}{n^{n-1}} w_n. \quad (\text{B26})$$

By Stirling's formula, this gives

$$h_n(0) \sim \frac{\sqrt{2\pi} n^{3/2}}{e^n} w_n. \quad (\text{B27})$$

Now, following Kotesovec, we use the Maple library *gdev*. Specifically, we use the procedure *equivalent*, which applies saddle-point methods to determine the asymptotic behavior of coefficients of generating functions [32]. From the command

$$\text{equivalent}(-1/(1 - \text{Lambert}W(-z)), z, n, 1), \quad (\text{B28})$$

we find

$$w_n \sim \frac{e^n}{4\sqrt{2\pi} n^{3/2}}. \quad (\text{B29})$$

Then by Eq. (B27),

$$h_n(0) \sim \frac{\sqrt{2\pi} n^{3/2}}{e^n} \cdot \frac{e^n}{4\sqrt{2\pi} n^{3/2}} = \frac{1}{4}, \quad (\text{B30})$$

giving Theorem III.5 Part 2.

2. Error rate

We now discuss the calculation of error rates. The error rate is a conditional probability: the probability that the output photon is distinguishable, given successful heralding. In our setting, this is calculated as follows. First (extending notation to arbitrary input states as usual), we may write the conditional probability $e_n(\rho)$ as a quotient of $\bar{e}_n(\rho)$, the probability of successful heralding *and* distinguishable output, divided by $h_n(\rho)$, the probability of successful heralding. In particular, we have

$$e_n(\epsilon) = \frac{\bar{e}_n(\epsilon)}{h_n(\epsilon)}. \quad (\text{B31})$$

The denominator may be calculated as in Appendix B 1, so we focus on $\bar{e}_n(\epsilon)$ here. As with the other notation, we extend the notation \bar{e}_n to arbitrary input states and write

$$\bar{e}_n(\epsilon) = \sum_{k=1}^n \binom{n}{k} \epsilon^k (1 - \epsilon)^{n-k} \bar{e}_n(\Phi_k) \quad (\text{B32})$$

[noting that $\bar{e}_n(\Phi_0) = 0$]. Of course, $\bar{e}_n(\Phi_k)$ is the average of $\bar{e}_n(|\eta\rangle\langle\eta|)$ over all $|\eta\rangle$ of the form Eq. (B2) with k distinguishability errors. Let $|\eta\rangle$ be such a term. To calculate

$\bar{e}_n(|\eta\rangle\langle\eta|)$, we perform an internal-external measurement on $\hat{U}|\eta\rangle$ (see Sec. II A). That is, we project onto the states (7) and calculate the probability that the external Fock state is ideal and the photon in the output mode 0 has internal state $|\xi_i\rangle$, $i > 0$.

As above, by Lemma A.3 (or rather, its slight extension to internal-external measurements) we may write

$$\bar{e}_n(|\eta\rangle\langle\eta|) = |\langle\eta_+|\eta\rangle|^2 \bar{e}_n(|\eta_+\rangle\langle\eta_+|). \quad (\text{B33})$$

Dividing both sides by $h_n(|\eta\rangle\langle\eta|)$, applying Eq. (B4), and rewriting $\bar{e}_n(\rho) = e_n(\rho)h_n(\rho)$, we obtain

$$e_n(|\eta\rangle\langle\eta|) = e_n(|\eta_+\rangle\langle\eta_+|). \quad (\text{B34})$$

This completes Part 2 of Theorem III.12.

Similarly to the heralding rate, we may rewrite the expression for $\bar{e}_n(|\eta\rangle\langle\eta|)$ in a way that removes the burden of checking that the output pattern satisfies the appropriate symmetries:

$$\bar{e}_n(|\eta\rangle\langle\eta|) = |\langle\eta_+|\eta\rangle|^2 \sum_{i \geq 1} |(\langle 0| \otimes I) a_0[\xi_i] \hat{U}|\eta_+\rangle|^2. \quad (\text{B35})$$

This requires the assumption that the relevant patterns in \mathcal{S}_G are ideal, as discussed in Remark III.13 and Appendix B 1. This completes the proof of Part 3 of Theorem III.12.

Again writing $\bar{e}_n(\Phi_k) = e_n(\Phi_k)h_n(\Phi_k)$, we note that we have the following expression for the error rate:

$$\begin{aligned} e_n(\epsilon) &= \frac{\sum_{k=1}^n \binom{n}{k} \epsilon^k (1 - \epsilon)^{n-k} e_n(\Phi_k) h_n(\Phi_k)}{\sum_{k=0}^n \binom{n}{k} \epsilon^k (1 - \epsilon)^{n-k} h_n(\Phi_k)} \\ &= \frac{\sum_{k=1}^n \binom{n}{k} \epsilon^k (1 - \epsilon)^{n-k} e_n(\Phi_k) \tilde{h}_n(\Phi_k)}{1 + \sum_{k=1}^n \binom{n}{k} \epsilon^k (1 - \epsilon)^{n-k} \tilde{h}_n(\Phi_k)}, \end{aligned} \quad (\text{B36})$$

where $\tilde{h}_n(\Phi_k) = h_n(\Phi_k)/h_n(\Phi_0)$. These expressions $\tilde{h}_n(\Phi_k)$ are given conjectured bounds in Conjecture III.14. Thus the error rate can be described in terms of these relatively well understood ratios of heralding rates and the quantities $e_n(\Phi_k)$. We discuss the efficient calculation of these quantities via simulation in Appendix E.

3. First-order approximations

We now return to the setting of the URS (or OBB) model of photon distinguishability. We assume that ϵ is independent of n and consider approximations of $h_n(\epsilon)$ and $e_n(\epsilon)$ up to first order in ϵ . As discussed in Sec. III B, we have

$$h_n(\epsilon) = h_n(0) + \epsilon n(h_n(\Phi_1) - h_n(0)) + O(\epsilon^2), \quad (\text{B37})$$

where Φ_1 is the evenly weighted probabilistic mixture of all terms of the form Eq. (14) involving exactly one

distinguishability error. Further, from Eq. (B36), we see that

$$e_n(\epsilon) = \epsilon n e_n(\Phi_1) \tilde{h}_n(\Phi_1) + O(\epsilon^2). \quad (\text{B38})$$

Then the first-order approximations require only knowledge of $h_n(0)$, discussed above, and $h_n(\Phi_1)$, $e_n(\Phi_1)$. Further, we will find that all terms of Φ_1 exhibit the same behavior with regard to distillation protocols. Without loss of generality, we will consider

$$|\eta\rangle = a_0^\dagger[\xi_1] a_1^\dagger[\xi_0] \cdots a_{n-1}^\dagger[\xi_0] |\vec{0}\rangle. \quad (\text{B39})$$

We will consider $U = F_n$ or $U = H_n$. Note that we will *not* place any additional assumptions on the relationship between ideal patterns and those in \mathcal{S}_G , so we allow all $n = 2^r$ in the Hadamard case and all $n \geq 3$ in the Fourier case. More generally, we may consider any $n \times n$ unitary U with first row identically equal to $1/\sqrt{n}$ (so that Appendix B 1 a applies) and with a symmetry group G of order n acting transitively on the n modes. In particular, as discussed in Appendix C, these results hold when U is the Fourier transform corresponding to any Abelian group G of order n . (Note that F_n and H_n are special cases.) In particular, for $P \in G$ and $|\eta\rangle$ as above, all $P|\eta\rangle$ are mutually orthogonal, corresponding to permuting the single “distinguishable” photon to each of the modes $0, \dots, n-1$. By the discussion in Sec. A 1, we obtain the associated G symmetrization

$$|\eta_+\rangle = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} a_0^\dagger[\xi_0] \cdots a_{i-1}^\dagger[\xi_0] a_i^\dagger[\xi_1] a_{i+1}^\dagger[\xi_0] \cdots a_{n-1}^\dagger[\xi_0] |\vec{0}\rangle, \quad \text{with } \langle \eta_+ | \eta \rangle = \frac{1}{\sqrt{n}}. \quad (\text{B40})$$

We note that $|\eta_+\rangle$ is invariant under *arbitrary* permutations of the external modes, not just those in G . In fact, by direct computation we see that

$$|\eta_+\rangle = \text{symm}(|0, 1, \dots, n-1\rangle_{1Q}) \otimes \text{symm}(|\xi_1, \xi_0, \dots, \xi_0\rangle) \quad (\text{B41})$$

$$= |1, \dots, 1\rangle_{2Q} \otimes (|\xi_1, \xi_0, \dots, \xi_0\rangle + |\xi_0, \xi_1, \xi_0, \dots, \xi_0\rangle + \cdots + |\xi_0, \dots, \xi_0, \xi_1\rangle). \quad (\text{B42})$$

In other words, $|\eta_+\rangle$ is a perfectly indistinguishable state, a pure tensor of the Fock state $|1, \dots, 1\rangle_{2Q}$ and some symmetric internal state. Then the results of Appendix B 1 a

apply here, and we have

$$h_n(|\eta\rangle\langle\eta|) = \frac{1}{n} h_n(|\eta_+\rangle\langle\eta_+|) = \frac{1}{n} h_n(0). \quad (\text{B43})$$

Further, since all terms of Φ_1 have the same symmetrization (up to relabeling of the $|\xi_i\rangle$, $i > 0$), we have

$$h_n(\Phi_1) = \frac{1}{n} h_n(0), \quad (\text{B44})$$

or equivalently $\tilde{h}_n(\Phi_1) = 1/n$. This gives Theorem III.4.

Next we consider the output error rate. In this setting, we may characterize $e_n(|\eta_+\rangle\langle\eta_+|)$ as the probability that, given successful heralding of a single output photon, the output photon has internal state $|\xi_1\rangle$. Since $|\eta_+\rangle$ is a pure tensor as described above, and linear optics and PNRD only act on the external part of the state, the postheralding state is still a pure tensor with the same internal part given in Eq. (B41). This internal part is fully symmetric, so the probability of any given photon having internal state $|\xi_1\rangle$ is exactly $1/n$. Then by Eq. (B34),

$$e_n(\Phi_1) = e_n(|\eta_+\rangle\langle\eta_+|) = \frac{1}{n}. \quad (\text{B45})$$

Combined with Eqs. (B38) and (B44), this proves Theorem III.2.

APPENDIX C: FOURIER TRANSFORM FOR FINITE ABELIAN GROUPS

In this section, we consider distillation protocols for a more general family of unitaries than the Fourier and Hadamard matrices above. In particular, we consider

$$F_{(n_1, \dots, n_\ell)} = F_{n_1} \otimes \cdots \otimes F_{n_\ell}, \quad (\text{C1})$$

where $n = n_1 \cdots n_\ell$ and the $n_i \geq 2$ are arbitrary. We have $F_n = F_{(n)}$, by definition. Further, since $F_2 = H$, we have $H_{2^r} = F_{(2, 2, \dots, 2)}$. More generally, the unitary (C1) corresponds to the Fourier transform of a finite Abelian group G of order $n = n_1 \cdots n_\ell$, namely

$$G \cong \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_\ell}. \quad (\text{C2})$$

Here \times is the direct product of groups and \mathbb{Z}_m is the cyclic group of order m . In particular, F_n corresponds to \mathbb{Z}_n and H_{2^r} corresponds to $\mathbb{Z}_2 \times \cdots \times \mathbb{Z}_2$. We observe that as written, there are some redundancies for different choices of n_1, \dots, n_ℓ ; we discuss this further below.

Note that in Appendices B 1 a and B 3, where we compute the 0th- and 1st-order terms of $h_n(\epsilon)$ and $e_n(\epsilon)$, very few properties of the unitaries F_n and H_n are used. In particular, to calculate $h_n(0)$, we observe in Appendix B 1 a that we only need U to be an $n \times n$ unitary matrix with

entries in the first row identically equal to $1/\sqrt{n}$. To calculate $h_n(\Phi_1), e_n(\Phi_1)$ as in Appendix B 3, we only need U to have a symmetry group G of order n (in the sense of Appendix A 1) that acts transitively on the n modes. Then the following theorem implies that these results apply to general $F_{(n_1, \dots, n_\ell)}$ as well.

Theorem C.1. Let $n \geq 2$ be a positive integer with $n_1 \cdots n_\ell = n$. We have the following:

(1) With a suitable labeling of the modes (i.e., a change of basis via a permutation matrix), the first row of $U = F_{(n_1, \dots, n_\ell)}$ is identically equal to $1/\sqrt{n}$.

(2) $F_{(n_1, \dots, n_\ell)}$ has a symmetry group G isomorphic to $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_\ell}$.

(3) G transitively permutes the modes.

In particular, up to first order in ϵ , the heralding and error rates of the distillation protocol for $U = F_{(n_1, \dots, n_\ell)}$ depend only on n .

Proof. For each i with $1 \leq i \leq \ell$, we view F_{n_i} as an operator on \mathbb{C}^{n_i} . Then $F_{(n_1, \dots, n_\ell)}$ is naturally an operator on $\mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_\ell} \cong \mathbb{C}^n$, which has basis $|m_1\rangle \otimes \cdots \otimes |m_\ell\rangle$ ($0 \leq m_i < n_i$). This may be identified with the usual basis for \mathbb{C}^n by

$$\begin{aligned} |m_1\rangle \otimes \cdots \otimes |m_\ell\rangle \\ \mapsto |m_1 + n_1 m_2 + n_1 n_2 m_3 + \cdots + (n_1 \cdots n_{\ell-1}) m_\ell\rangle, \end{aligned} \quad (\text{C3})$$

where we interpret the ket on the right-hand side modulo n . In this basis, $|0\rangle$ corresponds to $|0\rangle^{\otimes \ell}$, so for $g = m_1 + n_1 m_2 + n_1 n_2 m_3 + \cdots + (n_1 \cdots n_{\ell-1}) m_\ell$, it is clear that

$$\langle 0 | F_{(n_1, \dots, n_\ell)} | g \rangle = \prod_i \langle 0 | F_{n_i} | m_i \rangle = \prod_i \frac{1}{\sqrt{n_i}} = \frac{1}{\sqrt{n}}, \quad (\text{C4})$$

proving the first claim. For the second, we note that by Appendix A, each F_{n_i} factor has a symmetry group $G_i \cong \mathbb{Z}_{n_i}$ with generating permutation matrices $P_i \in G_i$ and diagonal matrices D_i satisfying $F_{n_i} P_i = D_i F_{n_i}$. As with the Hadamard case, this directly extends to the tensor product, with

$$\begin{aligned} (F_{n_1} \otimes \cdots \otimes F_{n_\ell}) (P_1^{q_1} \otimes \cdots \otimes P_\ell^{q_\ell}) \\ = (D_1^{q_1} \otimes \cdots \otimes D_\ell^{q_\ell}) (F_{n_1} \otimes \cdots \otimes F_{n_\ell}). \end{aligned} \quad (\text{C5})$$

This proves the second claim, with $G = G_1 \times \cdots \times G_\ell$. (Note this same reasoning gives a version of the ZTL for the general case, discussed below.) Finally, we observe that the action of G on the modes is transitive. This is clear from the description $|m_1\rangle \otimes \cdots \otimes |m_\ell\rangle$ of the basis for

\mathbb{C}^n . In particular, G_i freely permutes the i th factor without affecting any of the others; then for arbitrary basis vectors $|j_1\rangle, |j_2\rangle$, one may find a suitable element $P \in G$ such that $P |j_1\rangle = |j_2\rangle$. ■

We now give the analog of the ZTL for general $F_{(n_1, \dots, n_\ell)}$. Let $g_0, \dots, g_{n-1} \in \{0, \dots, n-1\}$ describe a Fock state $|s\rangle$ as usual, and by Eq. (C3) identify each g_j with a tuple $(m_1^{(j)}, m_2^{(j)}, \dots, m_\ell^{(j)})$. Consider Eq. (C5) with $q_i = 1$ and all other $q_j = 0$. Let ω_{n_i} be our usual choice of primitive n_i th root of unity. Recall from Appendix A 2 that D_i is diagonal with k th diagonal entry $\omega_{n_i}^k$. Then

$$\begin{aligned} \langle s | (F_{n_1} \otimes \cdots \otimes F_{n_\ell}) | 1, \dots, 1 \rangle \\ = \langle s | (F_{n_1} \otimes \cdots \otimes F_{n_\ell}) (I^{\otimes(i-1)} \otimes P_i \otimes I^{\ell-i}) | 1, \dots, 1 \rangle \end{aligned} \quad (\text{C6})$$

$$= \langle s | (I^{\otimes(i-1)} \otimes D_i \otimes I^{\ell-i}) (F_{n_1} \otimes \cdots \otimes F_{n_\ell}) | 1, \dots, 1 \rangle \quad (\text{C7})$$

$$= \omega_{n_i}^{-\sum_{j=0}^{n-1} m_i^{(j)}} \langle s | (F_{n_1} \otimes \cdots \otimes F_{n_\ell}) | 1, \dots, 1 \rangle, \quad (\text{C8})$$

so as before, for this quantity to be nonvanishing we require, for all i ,

$$\sum_{j=0}^{n-1} m_i^{(j)} \equiv 0 \pmod{n_i}. \quad (\text{C9})$$

(Note that the sum has n terms, not n_i terms.) This is the generalized zero transmission law. Note that this is very natural in the tensor-product basis, but less so in terms of the standard basis for \mathbb{C}^n constructed in Eq. (C3). Later in this section we will discuss the compatibility of this result with the ZTL for F_n , using a different basis for \mathbb{C}^n .

We note that in the case with all $n_i = 2$, we precisely recover the description of the ZTL in terms of decompositions into binary strings for the Hadamard unitaries, as discussed in Ref. [18] and Appendix A 3 above. In particular, the decomposition of an arbitrary mode g_j into the form $m_1^{(j)} + 2m_2^{(j)} + 2^2 m_3^{(j)} + \cdots + 2^{\ell-1} m_\ell^{(j)}$, where all $m_i^{(j)} \in \{0, 1\}$, is precisely its representation in base 2, and Eq. (C9) demands that all $\sum_j m_i^{(j)} \equiv 0 \pmod{2}$.

Note that for any $F_{(n_1, \dots, n_\ell)}$, if we are given the locations g_0, \dots, g_{n-2} of only $n-1$ photons, there is a unique mode g_{n-1} that completes the pattern to one satisfying the generalized ZTL. In particular, expressing each g_j for $0 \leq j < n-1$ in terms of the $m_i^{(j)}$ as above, Eq. (C9) uniquely determines $m_i^{(n-1)}$ for all i . This then uniquely determines g_{n-1} by Eq. (C3). This is used in the proof of Theorem III.17.

We now discuss different ways of expressing the unitaries of Eq. (C1). We introduce an additional bit of notation: for c relatively prime to n_j , define $F_{n_j}^{(c)}$ to have (a, b)

entry $\omega_{n_j}^{abc}$; in other words, we make a different choice of primitive n_j th root of unity, replacing $\omega_{n_j} = \exp(2\pi i/n_j)$ with $\omega_{n_j}^c = \exp(2\pi ic/n_j)$. These variants are not similar in general (in the technical sense that they are not equivalent up to a change of basis); however, we have $F_{n_j}^{(c)} Q = F_{n_j}$ for some permutation matrix Q . (The permutation taking column c to column 1, column $2c$ to column 2, etc., with indices taken modulo n_j .) In terms of the corresponding linear optical unitary, as long as we only apply $\hat{F}_{n_j}^{(c)}$ to mode-symmetric states such as $|1, \dots, 1\rangle$, we see that the possible output patterns (and their amplitudes) are unchanged. In the context of Theorem C.1, for the cyclic permutation P_j generating G_i , we have $F_{n_j}^{(c)} Q^\dagger P_j Q = D_j F_{n_j}^{(c)}$. (Note D_j is unchanged.) Then the corresponding symmetry group is still isomorphic to G_j , with the isomorphism given by $P \mapsto Q^\dagger P Q$, and we get exactly the same ZTL and suppression laws. For our purposes, then, we may view the $F_{n_j}^{(c)}$ as interchangeable as long as c is relatively prime to n_j .

Earlier, we commented that F_{n_j} corresponded to the Fourier transform on \mathbb{Z}_{n_j} , which is a map between two vector spaces of dimension n_j . These spaces are isomorphic, but not canonically so; to write F_{n_j} as a matrix over \mathbb{C}^{n_j} requires one to arbitrarily choose a primitive n_j th root of unity. Different choices give the various $F_{n_j}^{(c)}$. Thus when discussing the correspondence between finite abelian groups and matrices for their discrete Fourier transforms, we will need to be flexible about the choice of root of unity. In particular, the unitary in Eq. (C1) may be described by different choices of n_1, \dots, n_ℓ , as long as we are comfortable with additional permutation matrices Q as above. By the fundamental theorem of finite Abelian groups, we may always express the finite Abelian group G in the form $G \cong \mathbb{Z}_{p_1^{r_1}} \otimes \dots \otimes \mathbb{Z}_{p_\ell^{r_\ell}}$, where $p_1^{r_1} \geq \dots \geq p_\ell^{r_\ell}$ and the p_i are (not necessarily distinct) primes. Correspondingly, up

to a relabeling of the modes we may express $F_{(n_1, \dots, n_\ell)}$ as $F_{(p_1^{r_1}, \dots, p_\ell^{r_\ell})} Q$, where Q is a permutation matrix, corresponding to potentially choosing different roots of unity as above.

To be concrete, we consider the correspondence between $F_{n_1} \otimes F_{n_2}$ and F_n , where n_1, n_2 are relatively prime and $n = n_1 n_2$. (This gives the general case by induction.) By Bezout's theorem, we find c_1, c_2 such that $c_1 n_1 + c_2 n_2 = 1$. We note that c_1 is uniquely determined modulo n_2 and c_2 is uniquely determined modulo n_1 . We have a canonical ring isomorphism $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \rightarrow \mathbb{Z}_n$ by

$$(k_1, k_2) \mapsto k_1 c_2 n_2 + k_2 c_1 n_1. \quad (\text{C10})$$

The inverse is simply $k \mapsto (k \bmod n_1, k \bmod n_2)$. We then have $F_{n_1}^{(c_2)} \otimes F_{n_2}^{(c_1)}$ similar to F_n , with the change of basis simply being a relabeling of the modes according to Eq. (C10), specifically $|k_1\rangle \otimes |k_2\rangle \leftrightarrow |k_1 c_2 n_2 + k_2 c_1 n_1\rangle$. [Note this is a different labeling from Eq. (C3) in general!] In particular, this gives a *canonical* way of identifying $F_{n_1}^{(c_2)} \otimes F_{n_2}^{(c_1)}$ and F_n . Further, as discussed above, the suppression laws are the same as for $F_{n_1} \otimes F_{n_2}$ when expressed in terms of the natural basis for $\mathbb{C}^{n_1} \otimes \mathbb{C}^{n_2}$. To express them in terms of the natural basis for \mathbb{C}^n using Eq. (C10), let g_0, \dots, g_{n-1} be mode labels in $0, \dots, n-1$, with each $g_j = k_1^{(j)} c_2 n_2 + k_2^{(j)} c_1 n_1$. By Eq. (C9), we have $\sum_{j=0}^{n-1} k_i^{(j)} \equiv 0 \pmod{n_i}$. Then we see that

$$\sum_{j=0}^{n-1} g_j = \left(\sum_{j=0}^{n-1} k_1^{(j)} \right) c_2 n_2 + \left(\sum_{j=0}^{n-1} k_2^{(j)} \right) c_1 n_1 \equiv 0 \pmod{n}. \quad (\text{C11})$$

Then if we label the modes according to Eq. (C10), we recover the standard ZTL for F_n .

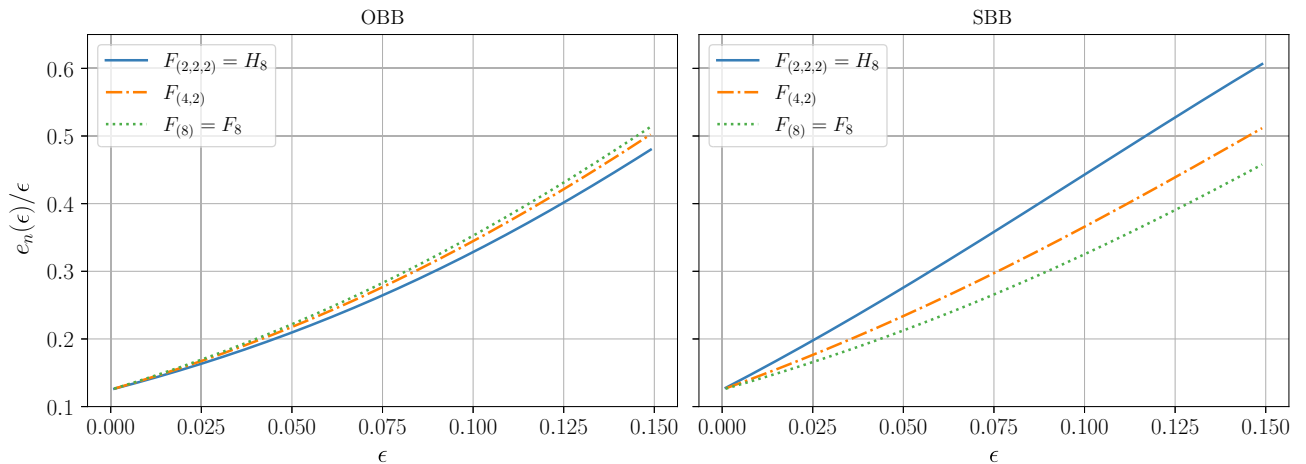


FIG. 13. Performance comparison of all $n = 8$ distillation protocols under the OBB and SBB noise models. The heralding rates $h_n(\epsilon)$ do not vary much between the models (see Fig. 8 for F_8, H_8).

For example, we consider the case of F_6 . Writing $n_1 = 2, n_2 = 3$, we have corresponding $c_1 = -1, c_2 = 1$, so that F_6 and $F_2 \otimes F_3^{(-1)}$ are similar, with both having the same suppression laws as $F_2 \otimes F_3$. We may explicitly check that for Q fixing column 0 and permuting columns 1 and 2, we have $F_3^{(-1)} Q = F_3$. The similarity between F_6 and $F_2 \otimes F_3^{(-1)}$ is seen by choosing $|j\rangle$ to correspond to the j th element of the sequence

$$(0, 0), (1, 2), (0, 1), (1, 0), (0, 2), (1, 1). \quad (\text{C12})$$

In summary, it suffices to consider the case of Eq. (C1) where the n_i are weakly decreasing prime powers. From this perspective, for each of $n = 8, 9, 12$, we obtain one protocol beyond the standard F_n and H_n (the latter only applicable when n is a power of 2), namely $F_{(4,2)}$, $F_{(3,3)}$, and $F_{(6,2)}$. For $n = 16$, there are five total options: F_{16} , $F_{(8,2)}$, $F_{(4,4)}$, $F_{(4,2,2)}$, and $F_{(2,2,2,2)} = H_{16}$.

In Fig. 13 we show the performance for all three $n = 8$ protocols (i.e., including the $F_{(4,2)}$ protocol as well as $F_8 = F_{(8)}$ and $H_8 = F_{(2,2,2,2)}$ studied in the main text). Note that the $F_{(4,2)}$ protocol falls in between F_8 and H_8 in terms of performance, under both error models. Therefore, if one is agnostic to the noise model, it may in fact be advantageous to use the $F_{(4,2)}$ protocol (if one knows the noise model, then either F_8 or H_8 is optimal).

APPENDIX D: HAAR RANDOM CASE

For comparison with the distillation schemes outlined in the main text and how they utilize interference to assist in distilling errors, we provide analysis in the case where random unitaries are considered for distillation. Interestingly, although these unitaries have no special structure, they can still be used for distillation purposes. However, as we will see, the protocols based around random matrices are significantly less resource efficient than the Fourier protocol of the main text. This highlights the significance of the specific constructive and destructive interference in the latter scheme.

We consider generating Haar random unitaries from $U(n)$ (where the number of photons, n , is the same as the number of modes) and, as a warm up, running the same distillation protocols as outlined in this work. In particular, we evolve the all 1 state $|\bar{1}\rangle = |1, \dots, 1\rangle$ under a sampled random matrix, compute the ideal heralding patterns that have exactly one photon out in mode 0 (which as the matrix is random, it will typically include all possible patterns), and study how the heralding probabilities scale. By concentration of measure [43], for large enough n we expect a typical sample to be well described by the average, and so we can use Haar averaging to compute quantities of interest.

First we can ask what the heralding rate $h_n(0)$ is. This is the probability that a single photon exits in mode 0, from

the initial all 1 state. We compute this as follows:

$$h_n(0) = \int d\mu_{\hat{U}} \langle \bar{1} | \hat{U}^\dagger (|1\rangle\langle 1| \otimes \mathbf{I}_{n-1,n-1}) \hat{U} | \bar{1} \rangle. \quad (\text{D1})$$

That is, the state $|\bar{1}\rangle$ is evolved under unitary \hat{U} , we compute the probability of PNRD giving an output pattern with a single photon in mode 0, then we average over all unitaries. Note that here, the \hat{U} are $d_{n,n} \times d_{n,n}$ matrices, where $d_{n,m}$ is the dimension of the space of n photons in m modes [$d_{n,m} = \binom{n+m-1}{n}$]. The identity matrix $\mathbf{I}_{n-1,n-1}$ on $n-1$ modes and $n-1$ photons has dimension $d_{n-1,n-1}$.

Whilst the underlying linear optical unitary is drawn from a Haar random distribution on $U(n)$, it does not imply the $d_{n,n} \times d_{n,n}$ matrices \hat{U} are Haar random in $U(d_{n,n})$. However, as shown in Ref. [25], linear optical unitaries are a continuous 1-design. As such, we can still treat the integral in the above equation as a Haar random integral ($\int d\mu_U U X U^\dagger = \text{Tr}[X] \mathbb{I}/d$), which we can analytically evaluate as

$$h_n(0) = \frac{\text{Tr}[\mathbf{I}_{n-1,n-1}]}{d_{n,n}} = \frac{d_{n-1,n-1}}{d_{n,n}} = \frac{1}{4} \frac{1}{1 - (2n)^{-1}} > 1/4. \quad (\text{D2})$$

Of course this makes intuitive sense; for a random unitary, one expects any particular state is equally likely with probability $1/d_{n,n}$ (when averaged over unitaries), and the total number of states with one photon in mode 0 is $d_{n-1,n-1}$. Notice that as with Theorem III.5 and Conjecture III.7 (pertaining to the Fourier and Hadamard matrices), this is decreasing monotonically in n and tends to $1/4$. We additionally conjecture that Eq. (D2) is an upper bound for the same sized protocol with the Fourier matrices. See Fig. 14 for some numerics.

We can further calculate what happens at first-order errors (where there are $n-1$ identical photons, and 1 in

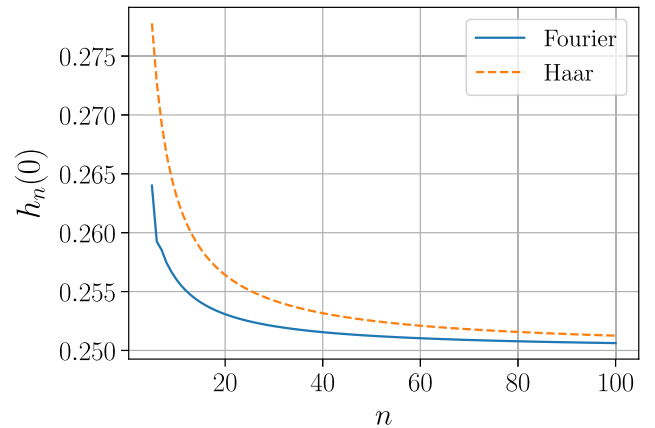


FIG. 14. Haar [Eq. (D2)] and Fourier (Theorem III.5) distillation schemes heralding rate at 0 error, as a function of n (i.e., the probability a single photon exits from mode 0, in the absence of error).

an orthogonal internal state). Now there are four integrals similar to Eq. (D1); we care about getting 1 or 0 photons out starting with $n - 1$ identical photons in n modes, and conversely 0 or 1 photons out starting with 1 (error) photon. Multiplying these as written gives, respectively, the probability of observing an ideal photon out of the scheme, or an error photon, in the case of exactly one error. These can be computed in the same manner as Eq. (D1), and combining the results gives two probabilities:

$$\begin{aligned} P_{\text{ideal}} &= \frac{d_{n-2,n-1}}{d_{n-1,n}} \frac{d_{1,n-1}}{d_{1,n}} = \frac{1}{4} \frac{(1 - 1/n)^2}{1 - \frac{3}{2n}}, \\ P_{\text{err}} &= \frac{d_{n-1,n-1}}{d_{n-1,n}} \frac{d_{0,n-1}}{d_{1,n}} = \frac{1}{2n}. \end{aligned} \quad (\text{D3})$$

The first term corresponds to the probability an ideal photon exits mode 0, and the second term the probability the error (distinguishable) photon exits mode 0, given exactly 1 distinguishability error at the input.

We can compute the output fidelity of the scheme at first order using $h_n(0)$, P_{ideal} , P_{err} :

$$\begin{aligned} f_n(\epsilon) &:= 1 - e_n(\epsilon) \\ &\approx \frac{h_n(0)(1 - \epsilon)^n + n\epsilon(1 - \epsilon)^{n-1}P_{\text{ideal}}}{h_n(0)(1 - \epsilon)^n + n\epsilon(1 - \epsilon)^{n-1}(P_{\text{ideal}} + P_{\text{err}})} \\ &\approx 1 - \frac{\epsilon}{2h_n(0)}. \end{aligned} \quad (\text{D4})$$

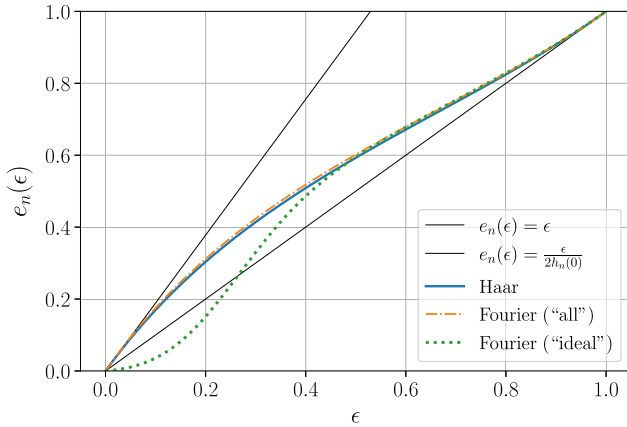


FIG. 15. “Distillation” from a single instance drawn from the Haar random distribution for $n = 9$, with the OBB error model, where we postselect on all possible patterns (solid blue line). Distillation is useful when $e_n(\epsilon) < \epsilon$ (lower solid black line), which is never the case here. For small $\epsilon \lesssim 0.1$ we see it follows the upper solid line, Eq. (D4). For comparison, we also include the case where we postselect on the ideal patterns in the Fourier protocol at $n = 9$ (dotted line), which we see for $\epsilon \lesssim 0.25$ results in $e_n(\epsilon) < \epsilon$. If we run the Fourier protocol but allow one to postselect on all valid patterns (dash-dot), the scaling is similar to the Haar case.

Since $h_n(0) < 0.5$ for $n \geq 3$, we see the output error $e_n(\epsilon) \approx \epsilon/2h_n(0)$ has actually *increased*, and for large enough n , $e_n(\epsilon) \approx 2\epsilon$. We verify this scaling numerically in Fig. 15. Note that we numerically obtain similar scaling $e_n(\epsilon) \approx 2\epsilon$ in the Fourier case if we postselect on all possible outcomes with one output photon, as opposed to only ideal patterns (see Fig. 15). Therefore, this result is more about the fact that we are postselecting on all possible patterns, not necessarily related to interference or lack thereof in the Haar vs. Fourier case.

A more interesting comparison is where we take a particular (random) unitary, and use only the highest weight heralding patterns resulting from the evolution of the initial all 1 state (such that there is one photon out in mode 0). The intuition here is that from a random linear optical unitary, one expects the probabilities to be roughly Porter-Thomas distributed, as demonstrated in Fig. 16 (also see Refs. [38,44,45]). As such, there will be a set of patterns that will have a relatively high weight (i.e., above the mean probability, $1/d$), distributed according to de^{-dp} , where p is the probability and $d = d_{n,n}$ (the expected number of states with probability greater than p^* is $\approx de^{-dp^*}$). The high weight patterns undergo a degree of constructive interference, akin to the patterns used for Fourier and

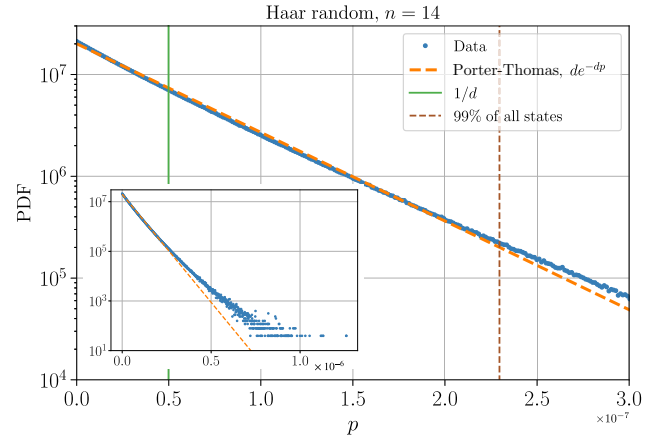


FIG. 16. Probability density function (PDF) of the output probabilities p from linear optical evolution of a sampled Haar random unitary in $U(n)$. Here the number of photons and modes is $n = 14$, with initial state $|1\rangle^{\otimes n}$. We see the data points very closely follow the conjectured Porter-Thomas distribution. We also plot for reference the vertical solid line $p = 1/d$ (where $d = d_{n,n}$ is the dimension), as well as the vertical dash line, left of which contains 99% of all states, approximately where the data begins to diverge from Porter-Thomas (likely due to random sampling and small size effects). In the main figure we exclude the long sparse tail to focus on the majority of the data. The inset contains all data (the points that visibly diverge from the Porter-Thomas line account for less than 1% of all states). The histogram data is constructed using 1000 equally sized bins over the range of probabilities p .

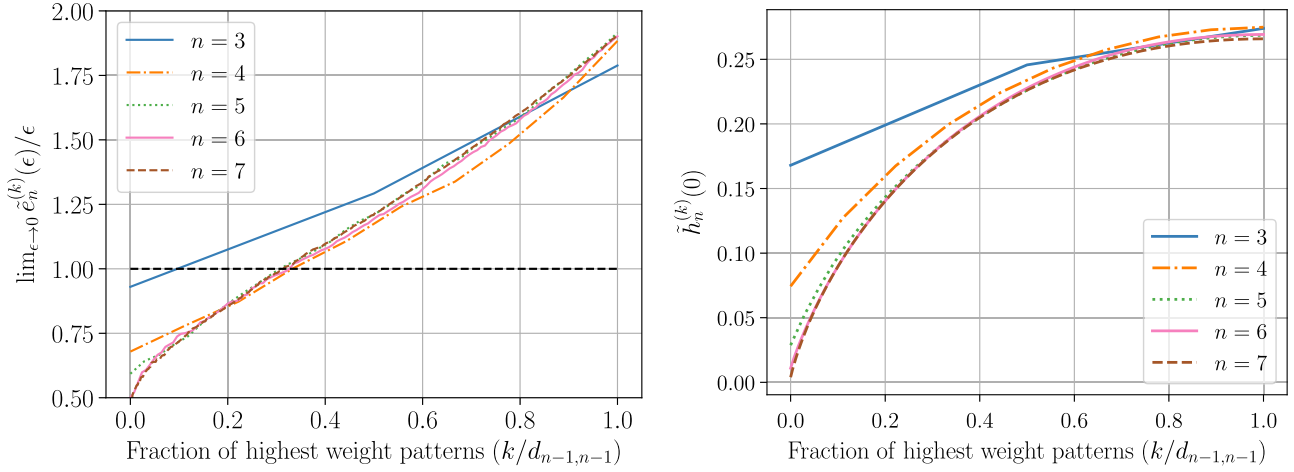


FIG. 17. Distillation using Haar random unitaries, where postselection is on a fraction of the highest weight (probability) output patterns (i.e., if one enumerated all states, those with the most constructive interference would be positioned towards the left on the x axis). To show the data for different n on a consistent axis, $x = 0$ corresponds to including the single highest weight pattern, and $x = 1$ to including all patterns. Each curve is an average of at least 100 Haar random unitaries. Left: error reduction relative to the initial error ϵ , at “small” ϵ . We see for $n = 4, 5, 6, 7$, there is a distillation effect when the top $\lesssim 30\%$ of the highest weight patterns are used. As expected from Eq. (D4), this tends to $\tilde{e}_n(\epsilon)/\epsilon \approx 2$ when all patterns are used. Right: heralding rate at 0 error for the protocol that selects a certain percentage of the highest weight patterns. As expected from Eq. (D2) it tends to a little over $1/4$, when using all possible patterns.

Hadamard distillation. As such, one may expect these patterns could be used for photon distillation, despite the underlying unitary being random.

We provide numerics for this protocol in Fig. 17, which shows that if one postselects on only the several highest-weighted patterns for that particular unitary, up to around 30% of the total weight, distillation does work (i.e., it reduces the error, when the initial error ϵ is small). However, as can also be seen in the figure (right), the heralding probability decreases as we reduce the number of postselection patterns and thus the error $\tilde{e}_n(\epsilon)$ (we put “tilde” above to differentiate this from the protocols of the main text). This is in contrast to the Fourier protocol in the main text, where at low error, the heralding probability is close to $1/4$ for all n , while the error reduces by a factor of n . Moreover, here the error reduction factor at small error seems to be at most by a factor of approximately 2 (i.e., at small ϵ , $\tilde{e}_n(\epsilon) \approx \epsilon/2$). Again, we can contrast this to the Fourier protocol, which has $e_n(\epsilon) \approx \epsilon/n$.

We can provide some analytics on this protocol too. First, consider the case where one uses only the highest weight pattern for postselection (this corresponds to the x axis at 0 in Fig. 17). From the Porter-Thomas distribution, we can estimate the largest expected probability to be around $p^* = 1/d \log d$ (i.e., the expected number of states with probability greater than p^* is 1). However, since the fraction of states that have one photon out in mode 0 is $d_{n-1,n-1}/d_{n,n}$ [see Eq. (D2)] we need to solve instead the

following equation:

$$N_{p>p^*} = d_{n,n} e^{-d_{n,n} p^*} = \frac{d_{n,n}}{d_{n-1,n-1}} \implies p^* = \frac{1}{d_{n,n}} \log d_{n-1,n-1}. \quad (\text{D5})$$

This is the expected probability of the highest weight state with one photon out in mode 0. Whilst the distillation scheme that selects only the highest weighted pattern has the greatest error reduction, the heralding probability decreases with the dimension.

If instead we select the highest weighted k patterns (with one photon out in mode 0), by similar reasoning as above, the cumulative probability is

$$\begin{aligned} \tilde{h}_n^{(k)}(0) &= \frac{1}{d_{n,n}} \sum_{i=1}^k \log \frac{d_{n-1,n-1}}{i} \\ &= \frac{1}{d_{n,n}} (k \log d_{n-1,n-1} - \log k!) \\ &\approx r \frac{d_{n-1,n-1}}{d_{n,n}} (1 - \log r). \end{aligned} \quad (\text{D6})$$

For the fraction of all possible patterns $r = k/d_{n-1,n-1}$ “large enough,” the equation can be rewritten via Stirling’s approximation as on the right-hand side. For $r \rightarrow 1$ this reproduces $\tilde{h}_n^{(k)} \approx (d_{n-1,n-1}/d_{n,n})$ as expected from

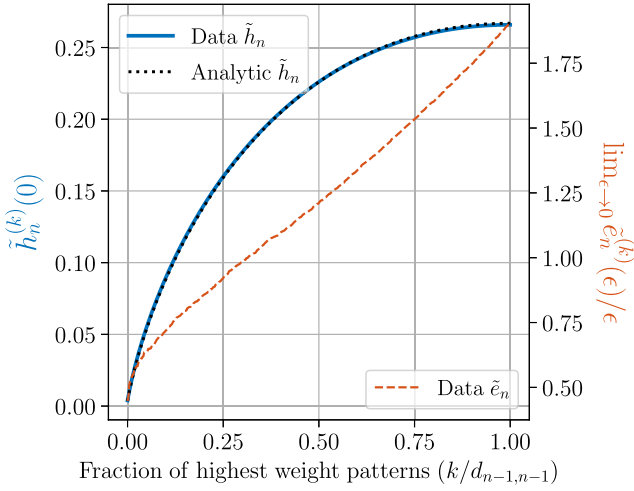


FIG. 18. The same data as in Fig. 17 for the $n = 7$ case, with the analytically computed heralding rate, Eq. (D6) (dotted line). The dashed line (red) pertains to the right y axis.

Eq. (D2) ($r = 1$ is equivalent to postselecting on all possible patterns). We show how the analytic function compares to data for $n = 7$ in Fig. 18.

This analysis shows that distillation is not a unique phenomena, in fact, it is typically possible for an arbitrary unitary, so long as there are a subset of patterns that have relatively high constructive interference. However, distillation from random unitaries is generally much more expensive (compared to say the Fourier protocol) as shown in Fig. 17 and the subsequent analysis above; the heralding rates are lower and the error is reduced by a lesser amount. The nature of the constructive interference in the Fourier (and Hadamard) unitaries is what allows for an efficient protocol, where the weight concentrates on a relatively small subset of all possible patterns, as per the zero-transmission laws, ultimately as a result of symmetry in the matrices (Sec. A 1).

APPENDIX E: SIMULATIONS

Here we discuss how we performed our numerics for the data in the main text and the Appendix (e.g., Sec. F). For simplicity, we restrict our attention to the OBB and SBB error models. First we set up some basics. We call $d_{n,m} = \binom{n+m-1}{n}$ the dimension of the Hilbert space of n photons in m modes. To compute an evolved Fock state under linear optics, in dimension $d_{n,m}$, requires time scaling like $O(nd_{n,m})$ and space (i.e., memory) like $O(d_{n,m})$ [46]. For an n photon distillation protocol, the full state space is of dimension $d_{n,n}$; however, as discussed below, we only need to consider outcomes with 0 and 1 photon in mode 0, and hence the relevant dimension is $d_{n,n-1} + d_{n-1,n-1}$ (for $n = 16$, this is around 220 million). As the simulations are carried out by adding one photon at a time, as soon as a configuration has more than two photons in mode 0, we

can discard it, thus constraining the dynamics in the relevant subspace. We will now describe how we simulate distinguishability errors in this work. Though not necessary, some additional details on this topic can be found in Ref. [25].

Naively to compute the distillation properties (i.e., $h_n(\epsilon), e_n(\epsilon)$) requires one to evaluate all 2^n initial states with $n - k$ “ideal” photons and k “error” photons. (See the discussions of Appendix B.) For example, for $n = 3$ the relevant input states can be represented $(1, 1, 1), (1, 1, 1'), (1, 1', 1), (1', 1, 1), (1, 1', 1'), (1', 1, 1'), (1', 1', 1), (1', 1', 1')$, where 1 indicates an “ideal” photon, and 1' an error. Since the errors do not interfere with the ideal photons, we can split the simulation into different subspaces, and combine results in postprocessing.

In the SBB case each initial state splits into two independent simulations, one of $n - k$ identical photons, and another with k identical photons, in the complementary positions (by symmetry of the states, this reduces the number of unique initial states to 2^{n-1}). In each case we only care about getting 0 or 1 photon out in mode 0, since anything else would be heralded as a non-ideal pattern. In particular, this is equivalent to projection of the full evolved state onto the subspace with 0 or 1 photons in mode 0. The two evolved (unnormalized) “states,” $|\psi_{n-k}\rangle, |\phi_k\rangle$, can be combined to compute the relevant quantities. Let us write each as a sum over configurations, where we separate out those that have 0 or 1 photon in mode 0:

$$\begin{aligned} |\psi_{n-k}\rangle &= \sum_{r^{(0)}} b_r^{(0)} |r^{(0)}\rangle + \sum_{r^{(1)}} b_r^{(1)} |r^{(1)}\rangle \\ |\phi_k\rangle &= \sum_{s^{(0)}} c_s^{(0)} |s^{(0)}\rangle + \sum_{s^{(1)}} c_s^{(1)} |s^{(1)}\rangle, \end{aligned} \quad (\text{E1})$$

where $|r^{(0)}\rangle$ ($|r^{(1)}\rangle$) is an $n - k$ photon configuration with 0 (1) photons in mode 0. Likewise for $|s^{(0)}\rangle, |s^{(1)}\rangle$, but for k photons.

As these two states do not interfere, we can classically combine their outcomes. Note that only pairs of the form $(r^{(0)}, s^{(1)})$ and $(r^{(1)}, s^{(0)})$ can combine to give a heralded pattern with 1 photon out. In particular, we can compute two relevant probabilities

$$\begin{aligned} P_{01} &:= \sum_{r^{(0)}, s^{(1)}: r^{(0)} + s^{(1)} \in \mathcal{I}} |b_r^{(0)}|^2 |c_s^{(1)}|^2, \\ P_{10} &= \sum_{r^{(1)}, s^{(0)}: r^{(1)} + s^{(0)} \in \mathcal{I}} |b_r^{(1)}|^2 |c_s^{(0)}|^2, \end{aligned} \quad (\text{E2})$$

where \mathcal{I} denotes the set of ideal heralding patterns, and $r + s$ is notation to add the length n tuples elementwise (since PNRD does not distinguish between ideal and error photons). The first term corresponds to the probability that an error photon exits mode 0, and the second term an

ideal photon. Therefore, the heralding probability given this particular initial state is $P_{01} + P_{10}$.

The numerical time cost to compute these terms by checking membership of \mathcal{I} for all pairs is, respectively, (order of)

$$d_{n-k,n-1}d_{k-1,n-1}, \quad d_{n-k-1,n-1}d_{k,n-1}, \quad (\text{E3})$$

which in fact dominates the total simulation cost, i.e., this is generally more expensive than computing the evolved states themselves $|\psi_{n-k}\rangle, |\phi_k\rangle$ (discussed above). Note, for the largest size we considered, $n = 16$, we have $|\mathcal{I}| \approx 2^{22}$; however, membership can be checked in time $O(1)$ by hashing. For $n = 16$, computing these probabilities in the worst case requires around 2^{36} numerical evaluations, which is achievable on a single processor given a few hours.

If we further denote $P_{01}^{x_k}, P_{10}^{x_k}$ to be the two above probabilities given the initial state x_k of k errors [with $\binom{n}{k}$ total states for each $k = 0, \dots, n$], we can compute the relevant rates as [see Eq. (40)]:

$$\begin{aligned} h_n(\epsilon) &= \sum_{k=0}^n \epsilon^k (1-\epsilon)^{n-k} \sum_{x_k} (P_{01}^{x_k} + P_{10}^{x_k}) \\ \bar{e}_n(\epsilon) &= \sum_{k=0}^n \epsilon^k (1-\epsilon)^{n-k} \sum_{x_k} P_{01}^{x_k}. \end{aligned} \quad (\text{E4})$$

In the OBB case, since the k errors are all independent, the statistics are classical. In particular, the probability to observe an output “error” pattern $s = (s_0, \dots, s_{n-1})$ from an initial state with $k = \sum_i s_i$ independent errors is given by the multinomial expression:

$$p_s = \frac{k!}{n^k} \prod_i \frac{1}{s_i!}. \quad (\text{E5})$$

Since the final calculation only requires probabilities as per Eq. (E2) (i.e., as opposed to the quantum amplitudes), we do not per se require any simulation for the k error state; given the evolved $n-k$ identical photon state $|\psi_{n-k}\rangle$ as above, along with Eq. (E5), we can compute Eqs. (E2) and (E4).

Thus far we have considered a brute-force approach, where we evaluate all 2^n error configurations. However as remarked several times in this work, the systems of interest for distillation have a great deal of symmetry, which we can use to reduce the total number of unique error configurations. Consider the Hadamard case first. Recall the discussion Appendix A3, where it is noted the 2×2 Hadamard matrix $H = F_2$ has the symmetry $HX = ZH$ where X, Z are Pauli matrices (i.e., swapping the columns is equivalent to multiplication of row 1 by -1). This symmetry extends amongst the tensor factors

of $H^{\otimes r}$ in the obvious way. Let us write these symmetries as before: $UP = DU$ where P is a permutation matrix (swapping columns), and D a diagonal matrix (of phases). Notice that this implies initial states that are equivalent up to permutation, i.e., $|y_k\rangle = P|x_k\rangle$, will have the same output amplitudes, up to phase: $\langle s|U|y_k\rangle = \langle s|UP|x_k\rangle = \langle s|DU|x_k\rangle = e^{i\phi} \langle s|U|x_k\rangle$. Since eventually we only care about the probabilities [Eq. (E2)], the phase is unimportant for computing heralding rates for distillation.

We can find another kind of symmetry relevant for distillation; if there exists a column swap that is equivalent to swapping rows (apart from the 0th), the permuted input state will have the same output amplitudes on configurations with modes permuted. In other words, we look for cases in which swapping certain input modes is equivalent to swapping certain corresponding output modes. If we denote the column and row swap by matrices C, R , respectively, so that $UC = RU$, we note that for any pattern s with $\langle s|U|\bar{1}\rangle \neq 0$, its permuted version $C|s\rangle$ leads to the same amplitude: $0 \neq \langle \bar{1}|U|s\rangle = \langle \bar{1}|RU|s\rangle = \langle \bar{1}|UC|s\rangle$. If such a transformation exists, this implies the relevant statistics for an input configuration x_k will be the same as the version permuted by C . Intuitively this is saying that the amplitudes do not care precisely which mode a photon ends up in; they only care if the full configuration is an ideal pattern. Note that our definition of ideal pattern $s \in \mathcal{I}$ does care that exactly one photon is in the 0th mode; therefore, to ensure this process maps ideal patterns to other ideal patterns, we must consider only C and R fixing the 0th mode.

We provide the symmetries for $H_4 = H \otimes H$ as an example (we ignore normalization for convenience). Recall,

$$H_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}. \quad (\text{E6})$$

There are three nontrivial symmetries of the usual $UP = DU$ form, arising from applying $HX = ZH$ on either or both tensor factors (i.e., $X \otimes I, I \otimes X, X \otimes X$ with $Z \otimes I, I \otimes Z, Z \otimes Z$). However, there are three further symmetries of the form $RU = UC$ that can be seen, involving swapping columns (1,2), (1,3), and (2,3); these are, respectively, equivalent to row swaps (1,2), (2,3), (1,3). For $H_{16} = H_4 \otimes H_4$, this results in only 32 unique error configurations (compared to 2^{16} naively).

In the Fourier case, F_n , as described in Appendix A2 there is a cyclic symmetry of the form $F_n P = D F_n$ where D is a diagonal matrix (of phases) and P a cyclic permutation. This implies any initial states that are cyclically equivalent will have the same statistics as discussed above. This reduces the number of unique states by around a factor of n . We can find another symmetry to further reduce

the terms, however, swapping both columns and rows similarly to the Hadamard case discussed above. Recall that the matrix elements of F_n (again ignoring normalization) are of the form ω_n^{ij} . For r relatively prime to n , a unique multiplicative inverse r^{-1} exists (modulo n). Therefore, if we map $i \mapsto r^{-1}i, j \mapsto rj$, the matrix elements are unchanged: $ij = r^{-1}irj \pmod n$. This corresponds to sending column j to column rj (permuting modes of the input state), and then sending the output row i to $r^{-1}i$. (Note that we index the modes modulo n , so the 0th row and column are untouched.) For $n = 16$, these symmetries reduce the total number of unique errors from 2^{16} to 693. The sequence of

integers counting the number of unique errors for different F_n can be found in the OEIS, sequence number A002729 [47].

APPENDIX F: NUMERICS

Here we list the functions for the heralding rate $h_n(\epsilon)$ and $\bar{e}_n(\epsilon) := h_n(\epsilon)e_n(\epsilon)$ [see Eq. (40) for reference], where the coefficients are given up to six decimal places (for some of the smaller protocols we write it as a fraction). The superscript denotes Hadamard (H) or Fourier (F) protocol, along with the error model (SBB or OBB). We go up to protocols of size $n = 16$.

$$\begin{aligned}
h_3^{(F,OBB)}(\epsilon) &= \frac{1}{3}(1-\epsilon)^3 + \frac{1}{9}\binom{3}{1}\epsilon^1(1-\epsilon)^2 + \frac{2}{9}\binom{3}{2}\epsilon^2(1-\epsilon)^1 + \frac{2}{9}\epsilon^3 \\
\bar{e}_3^{(F,OBB)}(\epsilon) &= \frac{1}{27}\binom{3}{1}\epsilon^1(1-\epsilon)^2 + \frac{4}{27}\binom{3}{2}\epsilon^2(1-\epsilon)^1 + \frac{2}{9}\epsilon^3 \\
h_4^{(F,OBB)}(\epsilon) &= \frac{1}{4}(1-\epsilon)^4 + \frac{1}{16}\binom{4}{1}\epsilon^1(1-\epsilon)^3 + \frac{1}{12}\binom{4}{2}\epsilon^2(1-\epsilon)^2 + \frac{3}{32}\binom{4}{3}\epsilon^3(1-\epsilon)^1 + \frac{3}{32}\epsilon^4 \\
\bar{e}_4^{(F,OBB)}(\epsilon) &= \frac{1}{64}\binom{4}{1}\epsilon^1(1-\epsilon)^3 + \frac{5}{96}\binom{4}{2}\epsilon^2(1-\epsilon)^2 + \frac{9}{128}\binom{4}{3}\epsilon^3(1-\epsilon)^1 + \frac{3}{32}\epsilon^4 \\
h_5^{(F,OBB)}(\epsilon) &= 0.264000(1-\epsilon)^5 + 0.052800\binom{5}{1}\epsilon^1(1-\epsilon)^4 + 0.073600\binom{5}{2}\epsilon^2(1-\epsilon)^3 \\
&\quad + 0.076800\binom{5}{3}\epsilon^3(1-\epsilon)^2 + 0.083200\binom{5}{4}\epsilon^4(1-\epsilon)^1 + 0.083200\epsilon^5 \\
\bar{e}_5^{(F,OBB)}(\epsilon) &= 0.010560\binom{5}{1}\epsilon^1(1-\epsilon)^4 + 0.039680\binom{5}{2}\epsilon^2(1-\epsilon)^3 + 0.051840\binom{5}{3}\epsilon^3(1-\epsilon)^2 \\
&\quad + 0.066560\binom{5}{4}\epsilon^4(1-\epsilon)^1 + 0.083200\epsilon^5 \\
h_6^{(F,OBB)}(\epsilon) &= 0.259259(1-\epsilon)^6 + 0.043210\binom{6}{1}\epsilon^1(1-\epsilon)^5 + 0.034156\binom{6}{2}\epsilon^2(1-\epsilon)^4 \\
&\quad + 0.032562\binom{6}{3}\epsilon^3(1-\epsilon)^3 + 0.029630\binom{6}{4}\epsilon^4(1-\epsilon)^2 + 0.028292\binom{6}{5}\epsilon^5(1-\epsilon)^1 + 0.028292\epsilon^6 \\
\bar{e}_6^{(F,OBB)}(\epsilon) &= 0.007202\binom{6}{1}\epsilon^1(1-\epsilon)^5 + 0.016872\binom{6}{2}\epsilon^2(1-\epsilon)^4 + 0.021451\binom{6}{3}\epsilon^3(1-\epsilon)^3 \\
&\quad + 0.022085\binom{6}{4}\epsilon^4(1-\epsilon)^2 + 0.023577\binom{6}{5}\epsilon^5(1-\epsilon)^1 + 0.028292\epsilon^6 \\
h_7^{(F,OBB)}(\epsilon) &= 0.258523(1-\epsilon)^7 + 0.036932\binom{7}{1}\epsilon^1(1-\epsilon)^6 + 0.047072\binom{7}{2}\epsilon^2(1-\epsilon)^5 \\
&\quad + 0.050204\binom{7}{3}\epsilon^3(1-\epsilon)^4 + 0.052988\binom{7}{4}\epsilon^4(1-\epsilon)^3 + 0.055079\binom{7}{5}\epsilon^5(1-\epsilon)^2 \\
&\quad + 0.056660\binom{7}{6}\epsilon^6(1-\epsilon)^1 + 0.056660\epsilon^7
\end{aligned}$$

$$\begin{aligned}\bar{e}_7^{(F, OBB)}(\epsilon) = & 0.005276 \binom{7}{1} \epsilon^1 (1 - \epsilon)^6 + 0.019997 \binom{7}{2} \epsilon^2 (1 - \epsilon)^5 + 0.027561 \binom{7}{3} \epsilon^3 (1 - \epsilon)^4 \\ & + 0.034779 \binom{7}{4} \epsilon^4 (1 - \epsilon)^3 + 0.041589 \binom{7}{5} \epsilon^5 (1 - \epsilon)^2 + 0.048566 \binom{7}{6} \epsilon^6 (1 - \epsilon)^1 + 0.056660 \epsilon^7\end{aligned}$$

$$\begin{aligned}h_8^{(F, OBB)}(\epsilon) = & 0.257446 (1 - \epsilon)^8 + 0.032181 \binom{8}{1} \epsilon^1 (1 - \epsilon)^7 + 0.039616 \binom{8}{2} \epsilon^2 (1 - \epsilon)^6 \\ & + 0.042520 \binom{8}{3} \epsilon^3 (1 - \epsilon)^5 + 0.044800 \binom{8}{4} \epsilon^4 (1 - \epsilon)^4 + 0.046653 \binom{8}{5} \epsilon^5 (1 - \epsilon)^3 \\ & + 0.048085 \binom{8}{6} \epsilon^6 (1 - \epsilon)^2 + 0.049087 \binom{8}{7} \epsilon^7 (1 - \epsilon)^1 + 0.049087 \epsilon^8\end{aligned}$$

$$\begin{aligned}\bar{e}_8^{(F, OBB)}(\epsilon) = & 0.004023 \binom{8}{1} \epsilon^1 (1 - \epsilon)^7 + 0.014948 \binom{8}{2} \epsilon^2 (1 - \epsilon)^6 + 0.021359 \binom{8}{3} \epsilon^3 (1 - \epsilon)^5 \\ & + 0.027055 \binom{8}{4} \epsilon^4 (1 - \epsilon)^4 + 0.032378 \binom{8}{5} \epsilon^5 (1 - \epsilon)^3 + 0.037566 \binom{8}{6} \epsilon^6 (1 - \epsilon)^2 \\ & + 0.042951 \binom{8}{7} \epsilon^7 (1 - \epsilon)^1 + 0.049087 \epsilon^8\end{aligned}$$

$$\begin{aligned}h_9^{(F, OBB)}(\epsilon) = & 0.256678 (1 - \epsilon)^9 + 0.028520 \binom{9}{1} \epsilon^1 (1 - \epsilon)^8 + 0.034616 \binom{9}{2} \epsilon^2 (1 - \epsilon)^7 \\ & + 0.036834 \binom{9}{3} \epsilon^3 (1 - \epsilon)^6 + 0.038728 \binom{9}{4} \epsilon^4 (1 - \epsilon)^5 + 0.040313 \binom{9}{5} \epsilon^5 (1 - \epsilon)^4 \\ & + 0.041613 \binom{9}{6} \epsilon^6 (1 - \epsilon)^3 + 0.042628 \binom{9}{7} \epsilon^7 (1 - \epsilon)^2 + 0.043305 \binom{9}{8} \epsilon^8 (1 - \epsilon)^1 + 0.043305 \epsilon^9\end{aligned}$$

$$\begin{aligned}\bar{e}_9^{(F, OBB)}(\epsilon) = & 0.003169 \binom{9}{1} \epsilon^1 (1 - \epsilon)^8 + 0.012104 \binom{9}{2} \epsilon^2 (1 - \epsilon)^7 + 0.017083 \binom{9}{3} \epsilon^3 (1 - \epsilon)^6 \\ & + 0.021689 \binom{9}{4} \epsilon^4 (1 - \epsilon)^5 + 0.025991 \binom{9}{5} \epsilon^5 (1 - \epsilon)^4 + 0.030110 \binom{9}{6} \epsilon^6 (1 - \epsilon)^3 \\ & + 0.034208 \binom{9}{7} \epsilon^7 (1 - \epsilon)^2 + 0.038493 \binom{9}{8} \epsilon^8 (1 - \epsilon)^1 + 0.043305 \epsilon^9\end{aligned}$$

$$\begin{aligned}h_{10}^{(F, OBB)}(\epsilon) = & 0.256038 (1 - \epsilon)^{10} + 0.025604 \binom{10}{1} \epsilon^1 (1 - \epsilon)^9 + 0.025161 \binom{10}{2} \epsilon^2 (1 - \epsilon)^8 \\ & + 0.026194 \binom{10}{3} \epsilon^3 (1 - \epsilon)^7 + 0.027467 \binom{10}{4} \epsilon^4 (1 - \epsilon)^6 + 0.028630 \binom{10}{5} \epsilon^5 (1 - \epsilon)^5 \\ & + 0.029633 \binom{10}{6} \epsilon^6 (1 - \epsilon)^4 + 0.030501 \binom{10}{7} \epsilon^7 (1 - \epsilon)^3 + 0.031206 \binom{10}{8} \epsilon^8 (1 - \epsilon)^2 \\ & + 0.031657 \binom{10}{9} \epsilon^9 (1 - \epsilon)^1 + 0.031657 \epsilon^{10}\end{aligned}$$

$$\begin{aligned}\bar{e}_{10}^{(F, OBB)}(\epsilon) = & 0.002560 \binom{10}{1} \epsilon^1 (1 - \epsilon)^9 + 0.007890 \binom{10}{2} \epsilon^2 (1 - \epsilon)^8 + 0.011238 \binom{10}{3} \epsilon^3 (1 - \epsilon)^7 \\ & + 0.014343 \binom{10}{4} \epsilon^4 (1 - \epsilon)^6 + 0.017258 \binom{10}{5} \epsilon^5 (1 - \epsilon)^5 + 0.020034 \binom{10}{6} \epsilon^6 (1 - \epsilon)^4 \\ & + 0.022780 \binom{10}{7} \epsilon^7 (1 - \epsilon)^3 + 0.025578 \binom{10}{8} \epsilon^8 (1 - \epsilon)^2 + 0.028491 \binom{10}{9} \epsilon^9 (1 - \epsilon)^1 + 0.031657 \epsilon^{10}\end{aligned}$$

$$\begin{aligned}
h_{11}^{(F, OBB)}(\epsilon) &= 0.255512(1-\epsilon)^{11} + 0.023228 \binom{11}{1} \epsilon^1 (1-\epsilon)^{10} + 0.027306 \binom{11}{2} \epsilon^2 (1-\epsilon)^9 \\
&\quad + 0.028891 \binom{11}{3} \epsilon^3 (1-\epsilon)^8 + 0.030293 \binom{11}{4} \epsilon^4 (1-\epsilon)^7 + 0.031500 \binom{11}{5} \epsilon^5 (1-\epsilon)^6 \\
&\quad + 0.032535 \binom{11}{6} \epsilon^6 (1-\epsilon)^5 + 0.033413 \binom{11}{7} \epsilon^7 (1-\epsilon)^4 + 0.034138 \binom{11}{8} \epsilon^8 (1-\epsilon)^3 \\
&\quad + 0.034699 \binom{11}{9} \epsilon^9 (1-\epsilon)^2 + 0.035049 \binom{11}{10} \epsilon^{10} (1-\epsilon)^1 + 0.035049 \epsilon^{11} \\
\bar{e}_{11}^{(F, OBB)}(\epsilon) &= 0.002112 \binom{11}{1} \epsilon^1 (1-\epsilon)^{10} + 0.008118 \binom{11}{2} \epsilon^2 (1-\epsilon)^9 + 0.011588 \binom{11}{3} \epsilon^3 (1-\epsilon)^8 \\
&\quad + 0.014835 \binom{11}{4} \epsilon^4 (1-\epsilon)^7 + 0.017861 \binom{11}{5} \epsilon^5 (1-\epsilon)^6 + 0.020725 \binom{11}{6} \epsilon^6 (1-\epsilon)^5 \\
&\quad + 0.023484 \binom{11}{7} \epsilon^7 (1-\epsilon)^4 + 0.026204 \binom{11}{8} \epsilon^8 (1-\epsilon)^3 + 0.028964 \binom{11}{9} \epsilon^9 (1-\epsilon)^2 \\
&\quad + 0.031863 \binom{11}{10} \epsilon^{10} (1-\epsilon)^1 + 0.035049 \epsilon^{11} \\
h_{12}^{(F, OBB)}(\epsilon) &= 0.255068(1-\epsilon)^{12} + 0.021256 \binom{12}{1} \epsilon^1 (1-\epsilon)^{11} + 0.016566 \binom{12}{2} \epsilon^2 (1-\epsilon)^{10} \\
&\quad + 0.015836 \binom{12}{3} \epsilon^3 (1-\epsilon)^9 + 0.015671 \binom{12}{4} \epsilon^4 (1-\epsilon)^8 + 0.015559 \binom{12}{5} \epsilon^5 (1-\epsilon)^7 \\
&\quad + 0.015427 \binom{12}{6} \epsilon^6 (1-\epsilon)^6 + 0.015266 \binom{12}{7} \epsilon^7 (1-\epsilon)^5 + 0.015097 \binom{12}{8} \epsilon^8 (1-\epsilon)^4 \\
&\quad + 0.014939 \binom{12}{9} \epsilon^9 (1-\epsilon)^3 + 0.014821 \binom{12}{10} \epsilon^{10} (1-\epsilon)^2 + 0.014762 \binom{12}{11} \epsilon^{11} (1-\epsilon)^1 + 0.014762 \epsilon^{12} \\
\bar{e}_{12}^{(F, OBB)}(\epsilon) &= 0.001771 \binom{12}{1} \epsilon^1 (1-\epsilon)^{11} + 0.004568 \binom{12}{2} \epsilon^2 (1-\epsilon)^{10} + 0.006112 \binom{12}{3} \epsilon^3 (1-\epsilon)^9 \\
&\quad + 0.007409 \binom{12}{4} \epsilon^4 (1-\epsilon)^8 + 0.008507 \binom{12}{5} \epsilon^5 (1-\epsilon)^7 + 0.009451 \binom{12}{6} \epsilon^6 (1-\epsilon)^6 \\
&\quad + 0.010279 \binom{12}{7} \epsilon^7 (1-\epsilon)^5 + 0.011040 \binom{12}{8} \epsilon^8 (1-\epsilon)^4 + 0.011785 \binom{12}{9} \epsilon^9 (1-\epsilon)^3 \\
&\quad + 0.012584 \binom{12}{10} \epsilon^{10} (1-\epsilon)^2 + 0.013532 \binom{12}{11} \epsilon^{11} (1-\epsilon)^1 + 0.014762 \epsilon^{12} \\
h_{13}^{(F, OBB)}(\epsilon) &= 0.254691(1-\epsilon)^{13} + 0.019592 \binom{13}{1} \epsilon^1 (1-\epsilon)^{12} + 0.022512 \binom{13}{2} \epsilon^2 (1-\epsilon)^{11} \\
&\quad + 0.023702 \binom{13}{3} \epsilon^3 (1-\epsilon)^{10} + 0.024770 \binom{13}{4} \epsilon^4 (1-\epsilon)^9 + 0.025713 \binom{13}{5} \epsilon^5 (1-\epsilon)^8 \\
&\quad + 0.026544 \binom{13}{6} \epsilon^6 (1-\epsilon)^7 + 0.027272 \binom{13}{7} \epsilon^7 (1-\epsilon)^6 + 0.027904 \binom{13}{8} \epsilon^8 (1-\epsilon)^5 \\
&\quad + 0.028446 \binom{13}{9} \epsilon^9 (1-\epsilon)^4 + 0.028893 \binom{13}{10} \epsilon^{10} (1-\epsilon)^3 + 0.029234 \binom{13}{11} \epsilon^{11} (1-\epsilon)^2 \\
&\quad + 0.029438 \binom{13}{12} \epsilon^{12} (1-\epsilon)^1 + 0.029438 \epsilon^{13}
\end{aligned}$$

$$\begin{aligned}\bar{e}_{13}^{(F, OBB)}(\epsilon) = & 0.001507 \binom{13}{1} \epsilon^1 (1 - \epsilon)^{12} + 0.005824 \binom{13}{2} \epsilon^2 (1 - \epsilon)^{11} + 0.008382 \binom{13}{3} \epsilon^3 (1 - \epsilon)^{10} \\ & + 0.010789 \binom{13}{4} \epsilon^4 (1 - \epsilon)^9 + 0.013046 \binom{13}{5} \epsilon^5 (1 - \epsilon)^8 + 0.015182 \binom{13}{6} \epsilon^6 (1 - \epsilon)^7 \\ & + 0.017223 \binom{13}{7} \epsilon^7 (1 - \epsilon)^6 + 0.019200 \binom{13}{8} \epsilon^8 (1 - \epsilon)^5 + 0.021144 \binom{13}{9} \epsilon^9 (1 - \epsilon)^4 \\ & + 0.023090 \binom{13}{10} \epsilon^{10} (1 - \epsilon)^3 + 0.025082 \binom{13}{11} \epsilon^{11} (1 - \epsilon)^2 + 0.027174 \binom{13}{12} \epsilon^{12} (1 - \epsilon)^1 + 0.029438 \epsilon^{13}\end{aligned}$$

$$\begin{aligned}h_{14}^{(F, OBB)}(\epsilon) = & 0.254365 (1 - \epsilon)^{14} + 0.018169 \binom{14}{1} \epsilon^1 (1 - \epsilon)^{13} + 0.019987 \binom{14}{2} \epsilon^2 (1 - \epsilon)^{12} \\ & + 0.020964 \binom{14}{3} \epsilon^3 (1 - \epsilon)^{11} + 0.021899 \binom{14}{4} \epsilon^4 (1 - \epsilon)^{10} + 0.022751 \binom{14}{5} \epsilon^5 (1 - \epsilon)^9 \\ & + 0.023515 \binom{14}{6} \epsilon^6 (1 - \epsilon)^8 + 0.024196 \binom{14}{7} \epsilon^7 (1 - \epsilon)^7 + 0.024798 \binom{14}{8} \epsilon^8 (1 - \epsilon)^6 \\ & + 0.025322 \binom{14}{9} \epsilon^9 (1 - \epsilon)^5 + 0.025770 \binom{14}{10} \epsilon^{10} (1 - \epsilon)^4 + 0.026137 \binom{14}{11} \epsilon^{11} (1 - \epsilon)^3 \\ & + 0.026414 \binom{14}{12} \epsilon^{12} (1 - \epsilon)^2 + 0.026576 \binom{14}{13} \epsilon^{13} (1 - \epsilon)^1 + 0.026576 \epsilon^{14}\end{aligned}$$

$$\begin{aligned}\bar{e}_{14}^{(F, OBB)}(\epsilon) = & 0.001298 \binom{14}{1} \epsilon^1 (1 - \epsilon)^{13} + 0.004810 \binom{14}{2} \epsilon^2 (1 - \epsilon)^{12} + 0.006986 \binom{14}{3} \epsilon^3 (1 - \epsilon)^{11} \\ & + 0.009027 \binom{14}{4} \epsilon^4 (1 - \epsilon)^{10} + 0.010959 \binom{14}{5} \epsilon^5 (1 - \epsilon)^9 + 0.012798 \binom{14}{6} \epsilon^6 (1 - \epsilon)^8 \\ & + 0.014559 \binom{14}{7} \epsilon^7 (1 - \epsilon)^7 + 0.016260 \binom{14}{8} \epsilon^8 (1 - \epsilon)^6 + 0.017923 \binom{14}{9} \epsilon^9 (1 - \epsilon)^5 \\ & + 0.019567 \binom{14}{10} \epsilon^{10} (1 - \epsilon)^4 + 0.021218 \binom{14}{11} \epsilon^{11} (1 - \epsilon)^3 + 0.022909 \binom{14}{12} \epsilon^{12} (1 - \epsilon)^2 \\ & + 0.024678 \binom{14}{13} \epsilon^{13} (1 - \epsilon)^1 + 0.026576 \epsilon^{14}\end{aligned}$$

$$\begin{aligned}h_{15}^{(F, OBB)}(\epsilon) = & 0.254081 (1 - \epsilon)^{15} + 0.016939 \binom{15}{1} \epsilon^1 (1 - \epsilon)^{14} + 0.017300 \binom{15}{2} \epsilon^2 (1 - \epsilon)^{13} \\ & + 0.017856 \binom{15}{3} \epsilon^3 (1 - \epsilon)^{12} + 0.018498 \binom{15}{4} \epsilon^4 (1 - \epsilon)^{11} + 0.019093 \binom{15}{5} \epsilon^5 (1 - \epsilon)^{10} \\ & + 0.019624 \binom{15}{6} \epsilon^6 (1 - \epsilon)^9 + 0.020091 \binom{15}{7} \epsilon^7 (1 - \epsilon)^8 + 0.020500 \binom{15}{8} \epsilon^8 (1 - \epsilon)^7 \\ & + 0.020856 \binom{15}{9} \epsilon^9 (1 - \epsilon)^6 + 0.021164 \binom{15}{10} \epsilon^{10} (1 - \epsilon)^5 + 0.021426 \binom{15}{11} \epsilon^{11} (1 - \epsilon)^4 \\ & + 0.021643 \binom{15}{12} \epsilon^{12} (1 - \epsilon)^3 + 0.021808 \binom{15}{13} \epsilon^{13} (1 - \epsilon)^2 + 0.021907 \binom{15}{14} \epsilon^{14} (1 - \epsilon)^1 + 0.021907 \epsilon^{15}\end{aligned}$$

$$\begin{aligned}\bar{e}_{15}^{(F, OBB)}(\epsilon) = & 0.001129 \binom{15}{1} \epsilon^1 (1 - \epsilon)^{14} + 0.003946 \binom{15}{2} \epsilon^2 (1 - \epsilon)^{13} + 0.005654 \binom{15}{3} \epsilon^3 (1 - \epsilon)^{12} \\ & + 0.007271 \binom{15}{4} \epsilon^4 (1 - \epsilon)^{11} + 0.008791 \binom{15}{5} \epsilon^5 (1 - \epsilon)^{10} + 0.010224 \binom{15}{6} \epsilon^6 (1 - \epsilon)^9\end{aligned}$$

$$\begin{aligned}
& + 0.011582 \binom{15}{7} \epsilon^7 (1 - \epsilon)^8 + 0.012880 \binom{15}{8} \epsilon^8 (1 - \epsilon)^7 + 0.014133 \binom{15}{9} \epsilon^9 (1 - \epsilon)^6 \\
& + 0.015361 \binom{15}{10} \epsilon^{10} (1 - \epsilon)^5 + 0.016581 \binom{15}{11} \epsilon^{11} (1 - \epsilon)^4 + 0.017818 \binom{15}{12} \epsilon^{12} (1 - \epsilon)^3 \\
& + 0.019095 \binom{15}{13} \epsilon^{13} (1 - \epsilon)^2 + 0.020446 \binom{15}{14} \epsilon^{14} (1 - \epsilon)^1 + 0.021907 \epsilon^{15} \\
h_{16}^{(F, OBB)}(\epsilon) &= 0.253832 (1 - \epsilon)^{16} + 0.015865 \binom{16}{1} \epsilon^1 (1 - \epsilon)^{15} + 0.017781 \binom{16}{2} \epsilon^2 (1 - \epsilon)^{14} \\
& + 0.018617 \binom{16}{3} \epsilon^3 (1 - \epsilon)^{13} + 0.019369 \binom{16}{4} \epsilon^4 (1 - \epsilon)^{12} + 0.020048 \binom{16}{5} \epsilon^5 (1 - \epsilon)^{11} \\
& + 0.020663 \binom{16}{6} \epsilon^6 (1 - \epsilon)^{10} + 0.021217 \binom{16}{7} \epsilon^7 (1 - \epsilon)^9 + 0.021716 \binom{16}{8} \epsilon^8 (1 - \epsilon)^8 \\
& + 0.022162 \binom{16}{9} \epsilon^9 (1 - \epsilon)^7 + 0.022559 \binom{16}{10} \epsilon^{10} (1 - \epsilon)^6 + 0.022907 \binom{16}{11} \epsilon^{11} (1 - \epsilon)^5 \\
& + 0.023205 \binom{16}{12} \epsilon^{12} (1 - \epsilon)^4 + 0.023450 \binom{16}{13} \epsilon^{13} (1 - \epsilon)^3 + 0.023633 \binom{16}{14} \epsilon^{14} (1 - \epsilon)^2 \\
& + 0.023738 \binom{16}{15} \epsilon^{15} (1 - \epsilon)^1 + 0.023738 \epsilon^{16} \\
\bar{e}_{16}^{(F, OBB)}(\epsilon) &= 0.000992 \binom{16}{1} \epsilon^1 (1 - \epsilon)^{15} + 0.003835 \binom{16}{2} \epsilon^2 (1 - \epsilon)^{14} + 0.005592 \binom{16}{3} \epsilon^3 (1 - \epsilon)^{13} \\
& + 0.007241 \binom{16}{4} \epsilon^4 (1 - \epsilon)^{12} + 0.008802 \binom{16}{5} \epsilon^5 (1 - \epsilon)^{11} + 0.010285 \binom{16}{6} \epsilon^6 (1 - \epsilon)^{10} \\
& + 0.011704 \binom{16}{7} \epsilon^7 (1 - \epsilon)^9 + 0.013068 \binom{16}{8} \epsilon^8 (1 - \epsilon)^8 + 0.014391 \binom{16}{9} \epsilon^9 (1 - \epsilon)^7 \\
& + 0.015684 \binom{16}{10} \epsilon^{10} (1 - \epsilon)^6 + 0.016962 \binom{16}{11} \epsilon^{11} (1 - \epsilon)^5 + 0.018239 \binom{16}{12} \epsilon^{12} (1 - \epsilon)^4 \\
& + 0.019533 \binom{16}{13} \epsilon^{13} (1 - \epsilon)^3 + 0.020863 \binom{16}{14} \epsilon^{14} (1 - \epsilon)^2 + 0.022255 \binom{16}{15} \epsilon^{15} (1 - \epsilon)^1 + 0.023738 \epsilon^{16} \\
h_3^{(F, SBB)}(\epsilon) &= \frac{1}{3} (1 - \epsilon)^3 + \frac{1}{9} \binom{3}{1} \epsilon^1 (1 - \epsilon)^2 + \frac{1}{9} \binom{3}{2} \epsilon^2 (1 - \epsilon)^1 + \frac{1}{3} \epsilon^3 \\
\bar{e}_3^{(F, SBB)}(\epsilon) &= \frac{1}{27} \binom{3}{1} \epsilon^1 (1 - \epsilon)^2 + \frac{2}{27} \binom{3}{2} \epsilon^2 (1 - \epsilon)^1 + \frac{1}{3} \epsilon^3 \\
h_4^{(F, SBB)}(\epsilon) &= \frac{1}{4} (1 - \epsilon)^4 + \frac{1}{16} \binom{4}{1} \epsilon^1 (1 - \epsilon)^3 + \frac{1}{12} \binom{4}{2} \epsilon^2 (1 - \epsilon)^2 + \frac{1}{16} \binom{4}{3} \epsilon^3 (1 - \epsilon)^1 + \frac{1}{4} \epsilon^4 \\
\bar{e}_4^{(F, SBB)}(\epsilon) &= \frac{1}{64} \binom{4}{1} \epsilon^1 (1 - \epsilon)^3 + \frac{1}{24} \binom{4}{2} \epsilon^2 (1 - \epsilon)^2 + \frac{3}{64} \binom{4}{3} \epsilon^3 (1 - \epsilon)^1 + \frac{1}{4} \epsilon^4 \\
h_5^{(F, SBB)}(\epsilon) &= 0.264000 (1 - \epsilon)^5 + 0.052800 \binom{5}{1} \epsilon^1 (1 - \epsilon)^4 + 0.059200 \binom{5}{2} \epsilon^2 (1 - \epsilon)^3 \\
& + 0.059200 \binom{5}{3} \epsilon^3 (1 - \epsilon)^2 + 0.052800 \binom{5}{4} \epsilon^4 (1 - \epsilon)^1 + 0.264000 \epsilon^5 \\
\bar{e}_5^{(F, SBB)}(\epsilon) &= 0.010560 \binom{5}{1} \epsilon^1 (1 - \epsilon)^4 + 0.024960 \binom{5}{2} \epsilon^2 (1 - \epsilon)^3 + 0.034240 \binom{5}{3} \epsilon^3 (1 - \epsilon)^2 \\
& + 0.042240 \binom{5}{4} \epsilon^4 (1 - \epsilon)^1 + 0.264000 \epsilon^5
\end{aligned}$$

$$\begin{aligned}
h_6^{(F,SBB)}(\epsilon) &= 0.259259(1-\epsilon)^6 + 0.043210\binom{6}{1}\epsilon^1(1-\epsilon)^5 + 0.041152\binom{6}{2}\epsilon^2(1-\epsilon)^4 \\
&\quad + 0.038272\binom{6}{3}\epsilon^3(1-\epsilon)^3 + 0.041152\binom{6}{4}\epsilon^4(1-\epsilon)^2 + 0.043210\binom{6}{5}\epsilon^5(1-\epsilon)^1 + 0.259259\epsilon^6 \\
\bar{e}_6^{(F,SBB)}(\epsilon) &= 0.007202\binom{6}{1}\epsilon^1(1-\epsilon)^5 + 0.017284\binom{6}{2}\epsilon^2(1-\epsilon)^4 + 0.019136\binom{6}{3}\epsilon^3(1-\epsilon)^3 \\
&\quad + 0.023868\binom{6}{4}\epsilon^4(1-\epsilon)^2 + 0.036008\binom{6}{5}\epsilon^5(1-\epsilon)^1 + 0.259259\epsilon^6 \\
h_7^{(F,SBB)}(\epsilon) &= 0.258523(1-\epsilon)^7 + 0.036932\binom{7}{1}\epsilon^1(1-\epsilon)^6 + 0.042440\binom{7}{2}\epsilon^2(1-\epsilon)^5 \\
&\quad + 0.043909\binom{7}{3}\epsilon^3(1-\epsilon)^4 + 0.043909\binom{7}{4}\epsilon^4(1-\epsilon)^3 + 0.042440\binom{7}{5}\epsilon^5(1-\epsilon)^2 \\
&\quad + 0.036932\binom{7}{6}\epsilon^6(1-\epsilon)^1 + 0.258523\epsilon^7 \\
\bar{e}_7^{(F,SBB)}(\epsilon) &= 0.005276\binom{7}{1}\epsilon^1(1-\epsilon)^6 + 0.014952\binom{7}{2}\epsilon^2(1-\epsilon)^5 + 0.020046\binom{7}{3}\epsilon^3(1-\epsilon)^4 \\
&\quad + 0.023862\binom{7}{4}\epsilon^4(1-\epsilon)^3 + 0.027487\binom{7}{5}\epsilon^5(1-\epsilon)^2 + 0.031656\binom{7}{6}\epsilon^6(1-\epsilon)^1 + 0.258523\epsilon^7 \\
h_8^{(F,SBB)}(\epsilon) &= 0.257446(1-\epsilon)^8 + 0.032181\binom{8}{1}\epsilon^1(1-\epsilon)^7 + 0.042454\binom{8}{2}\epsilon^2(1-\epsilon)^6 \\
&\quad + 0.038256\binom{8}{3}\epsilon^3(1-\epsilon)^5 + 0.044332\binom{8}{4}\epsilon^4(1-\epsilon)^4 + 0.038256\binom{8}{5}\epsilon^5(1-\epsilon)^3 \\
&\quad + 0.042454\binom{8}{6}\epsilon^6(1-\epsilon)^2 + 0.032181\binom{8}{7}\epsilon^7(1-\epsilon)^1 + 0.257446\epsilon^8 \\
\bar{e}_8^{(F,SBB)}(\epsilon) &= 0.004023\binom{8}{1}\epsilon^1(1-\epsilon)^7 + 0.013932\binom{8}{2}\epsilon^2(1-\epsilon)^6 + 0.016147\binom{8}{3}\epsilon^3(1-\epsilon)^5 \\
&\quad + 0.022166\binom{8}{4}\epsilon^4(1-\epsilon)^4 + 0.022109\binom{8}{5}\epsilon^5(1-\epsilon)^3 + 0.028522\binom{8}{6}\epsilon^6(1-\epsilon)^2 \\
&\quad + 0.028158\binom{8}{7}\epsilon^7(1-\epsilon)^1 + 0.257446\epsilon^8 \\
h_9^{(F,SBB)}(\epsilon) &= 0.256678(1-\epsilon)^9 + 0.028520\binom{9}{1}\epsilon^1(1-\epsilon)^8 + 0.032563\binom{9}{2}\epsilon^2(1-\epsilon)^7 \\
&\quad + 0.036335\binom{9}{3}\epsilon^3(1-\epsilon)^6 + 0.034293\binom{9}{4}\epsilon^4(1-\epsilon)^5 + 0.034293\binom{9}{5}\epsilon^5(1-\epsilon)^4 \\
&\quad + 0.036335\binom{9}{6}\epsilon^6(1-\epsilon)^3 + 0.032563\binom{9}{7}\epsilon^7(1-\epsilon)^2 + 0.028520\binom{9}{8}\epsilon^8(1-\epsilon)^1 + 0.256678\epsilon^9 \\
\bar{e}_9^{(F,SBB)}(\epsilon) &= 0.003169\binom{9}{1}\epsilon^1(1-\epsilon)^8 + 0.009810\binom{9}{2}\epsilon^2(1-\epsilon)^7 + 0.014357\binom{9}{3}\epsilon^3(1-\epsilon)^6 \\
&\quad + 0.015994\binom{9}{4}\epsilon^4(1-\epsilon)^5 + 0.018299\binom{9}{5}\epsilon^5(1-\epsilon)^4 + 0.021978\binom{9}{6}\epsilon^6(1-\epsilon)^3 \\
&\quad + 0.022753\binom{9}{7}\epsilon^7(1-\epsilon)^2 + 0.025351\binom{9}{8}\epsilon^8(1-\epsilon)^1 + 0.256678\epsilon^9
\end{aligned}$$

$$\begin{aligned}
h_{10}^{(F,SBB)}(\epsilon) &= 0.256038(1-\epsilon)^{10} + 0.025604\binom{10}{1}\epsilon^1(1-\epsilon)^9 + 0.027144\binom{10}{2}\epsilon^2(1-\epsilon)^8 \\
&\quad + 0.024597\binom{10}{3}\epsilon^3(1-\epsilon)^7 + 0.025946\binom{10}{4}\epsilon^4(1-\epsilon)^6 + 0.025583\binom{10}{5}\epsilon^5(1-\epsilon)^5 \\
&\quad + 0.025946\binom{10}{6}\epsilon^6(1-\epsilon)^4 + 0.024597\binom{10}{7}\epsilon^7(1-\epsilon)^3 + 0.027144\binom{10}{8}\epsilon^8(1-\epsilon)^2 \\
&\quad + 0.025604\binom{10}{9}\epsilon^9(1-\epsilon)^1 + 0.256038\epsilon^{10} \\
\bar{e}_{10}^{(F,SBB)}(\epsilon) &= 0.002560\binom{10}{1}\epsilon^1(1-\epsilon)^9 + 0.007627\binom{10}{2}\epsilon^2(1-\epsilon)^8 + 0.009105\binom{10}{3}\epsilon^3(1-\epsilon)^7 \\
&\quad + 0.011405\binom{10}{4}\epsilon^4(1-\epsilon)^6 + 0.012791\binom{10}{5}\epsilon^5(1-\epsilon)^5 + 0.014541\binom{10}{6}\epsilon^6(1-\epsilon)^4 \\
&\quad + 0.015493\binom{10}{7}\epsilon^7(1-\epsilon)^3 + 0.019517\binom{10}{8}\epsilon^8(1-\epsilon)^2 + 0.023043\binom{10}{9}\epsilon^9(1-\epsilon)^1 + 0.256038\epsilon^{10} \\
h_{11}^{(F,SBB)}(\epsilon) &= 0.255512(1-\epsilon)^{11} + 0.023228\binom{11}{1}\epsilon^1(1-\epsilon)^{10} + 0.026221\binom{11}{2}\epsilon^2(1-\epsilon)^9 \\
&\quad + 0.027262\binom{11}{3}\epsilon^3(1-\epsilon)^8 + 0.027760\binom{11}{4}\epsilon^4(1-\epsilon)^7 + 0.027974\binom{11}{5}\epsilon^5(1-\epsilon)^6 \\
&\quad + 0.027974\binom{11}{6}\epsilon^6(1-\epsilon)^5 + 0.027760\binom{11}{7}\epsilon^7(1-\epsilon)^4 + 0.027262\binom{11}{8}\epsilon^8(1-\epsilon)^3 \\
&\quad + 0.026221\binom{11}{9}\epsilon^9(1-\epsilon)^2 + 0.023228\binom{11}{10}\epsilon^{10}(1-\epsilon)^1 + 0.255512\epsilon^{11} \\
\bar{e}_{11}^{(F,SBB)}(\epsilon) &= 0.002112\binom{11}{1}\epsilon^1(1-\epsilon)^{10} + 0.006888\binom{11}{2}\epsilon^2(1-\epsilon)^9 + 0.009509\binom{11}{3}\epsilon^3(1-\epsilon)^8 \\
&\quad + 0.011509\binom{11}{4}\epsilon^4(1-\epsilon)^7 + 0.013212\binom{11}{5}\epsilon^5(1-\epsilon)^6 + 0.014762\binom{11}{6}\epsilon^6(1-\epsilon)^5 \\
&\quad + 0.016252\binom{11}{7}\epsilon^7(1-\epsilon)^4 + 0.017753\binom{11}{8}\epsilon^8(1-\epsilon)^3 + 0.019333\binom{11}{9}\epsilon^9(1-\epsilon)^2 \\
&\quad + 0.021117\binom{11}{10}\epsilon^{10}(1-\epsilon)^1 + 0.255512\epsilon^{11} \\
h_{12}^{(F,SBB)}(\epsilon) &= 0.255068(1-\epsilon)^{12} + 0.021256\binom{12}{1}\epsilon^1(1-\epsilon)^{11} + 0.018466\binom{12}{2}\epsilon^2(1-\epsilon)^{10} \\
&\quad + 0.016542\binom{12}{3}\epsilon^3(1-\epsilon)^9 + 0.016350\binom{12}{4}\epsilon^4(1-\epsilon)^8 + 0.015541\binom{12}{5}\epsilon^5(1-\epsilon)^7 \\
&\quad + 0.016128\binom{12}{6}\epsilon^6(1-\epsilon)^6 + 0.015541\binom{12}{7}\epsilon^7(1-\epsilon)^5 + 0.016350\binom{12}{8}\epsilon^8(1-\epsilon)^4 \\
&\quad + 0.016542\binom{12}{9}\epsilon^9(1-\epsilon)^3 + 0.018466\binom{12}{10}\epsilon^{10}(1-\epsilon)^2 + 0.021256\binom{12}{11}\epsilon^{11}(1-\epsilon)^1 + 0.255068\epsilon^{12} \\
\bar{e}_{12}^{(F,SBB)}(\epsilon) &= 0.001771\binom{12}{1}\epsilon^1(1-\epsilon)^{11} + 0.004505\binom{12}{2}\epsilon^2(1-\epsilon)^{10} + 0.005473\binom{12}{3}\epsilon^3(1-\epsilon)^9 \\
&\quad + 0.006457\binom{12}{4}\epsilon^4(1-\epsilon)^8 + 0.006983\binom{12}{5}\epsilon^5(1-\epsilon)^7 + 0.008064\binom{12}{6}\epsilon^6(1-\epsilon)^6
\end{aligned}$$

$$\begin{aligned}
& + 0.008558 \binom{12}{7} \epsilon^7 (1 - \epsilon)^5 + 0.009893 \binom{12}{8} \epsilon^8 (1 - \epsilon)^4 + 0.011068 \binom{12}{9} \epsilon^9 (1 - \epsilon)^3 \\
& + 0.013961 \binom{12}{10} \epsilon^{10} (1 - \epsilon)^2 + 0.019484 \binom{12}{11} \epsilon^{11} (1 - \epsilon)^1 + 0.255068 \epsilon^{12} \\
h_{13}^{(F,SBB)}(\epsilon) & = 0.254691 (1 - \epsilon)^{13} + 0.019592 \binom{13}{1} \epsilon^1 (1 - \epsilon)^{12} + 0.021871 \binom{13}{2} \epsilon^2 (1 - \epsilon)^{11} \\
& + 0.022711 \binom{13}{3} \epsilon^3 (1 - \epsilon)^{10} + 0.023175 \binom{13}{4} \epsilon^4 (1 - \epsilon)^9 + 0.023439 \binom{13}{5} \epsilon^5 (1 - \epsilon)^8 \\
& + 0.023561 \binom{13}{6} \epsilon^6 (1 - \epsilon)^7 + 0.023561 \binom{13}{7} \epsilon^7 (1 - \epsilon)^6 + 0.023439 \binom{13}{8} \epsilon^8 (1 - \epsilon)^5 \\
& + 0.023175 \binom{13}{9} \epsilon^9 (1 - \epsilon)^4 + 0.022711 \binom{13}{10} \epsilon^{10} (1 - \epsilon)^3 + 0.021871 \binom{13}{11} \epsilon^{11} (1 - \epsilon)^2 \\
& + 0.019592 \binom{13}{12} \epsilon^{12} (1 - \epsilon)^1 + 0.254691 \epsilon^{13} \\
\bar{e}_{13}^{(F,SBB)}(\epsilon) & = 0.001507 \binom{13}{1} \epsilon^1 (1 - \epsilon)^{12} + 0.005089 \binom{13}{2} \epsilon^2 (1 - \epsilon)^{11} + 0.007104 \binom{13}{3} \epsilon^3 (1 - \epsilon)^{10} \\
& + 0.008677 \binom{13}{4} \epsilon^4 (1 - \epsilon)^9 + 0.010019 \binom{13}{5} \epsilon^5 (1 - \epsilon)^8 + 0.011221 \binom{13}{6} \epsilon^6 (1 - \epsilon)^7 \\
& + 0.012340 \binom{13}{7} \epsilon^7 (1 - \epsilon)^6 + 0.013420 \binom{13}{8} \epsilon^8 (1 - \epsilon)^5 + 0.014498 \binom{13}{9} \epsilon^9 (1 - \epsilon)^4 \\
& + 0.015608 \binom{13}{10} \epsilon^{10} (1 - \epsilon)^3 + 0.016782 \binom{13}{11} \epsilon^{11} (1 - \epsilon)^2 + 0.018085 \binom{13}{12} \epsilon^{12} (1 - \epsilon)^1 \\
& + 0.254691 \epsilon^{13} \\
h_{14}^{(F,SBB)}(\epsilon) & = 0.254365 (1 - \epsilon)^{14} + 0.018169 \binom{14}{1} \epsilon^1 (1 - \epsilon)^{13} + 0.021090 \binom{14}{2} \epsilon^2 (1 - \epsilon)^{12} \\
& + 0.020198 \binom{14}{3} \epsilon^3 (1 - \epsilon)^{11} + 0.021062 \binom{14}{4} \epsilon^4 (1 - \epsilon)^{10} + 0.020892 \binom{14}{5} \epsilon^5 (1 - \epsilon)^9 \\
& + 0.021293 \binom{14}{6} \epsilon^6 (1 - \epsilon)^8 + 0.021166 \binom{14}{7} \epsilon^7 (1 - \epsilon)^7 + 0.021293 \binom{14}{8} \epsilon^8 (1 - \epsilon)^6 \\
& + 0.020892 \binom{14}{9} \epsilon^9 (1 - \epsilon)^5 + 0.021062 \binom{14}{10} \epsilon^{10} (1 - \epsilon)^4 + 0.020198 \binom{14}{11} \epsilon^{11} (1 - \epsilon)^3 \\
& + 0.021090 \binom{14}{12} \epsilon^{12} (1 - \epsilon)^2 + 0.018169 \binom{14}{13} \epsilon^{13} (1 - \epsilon)^1 + 0.254365 \epsilon^{14} \\
\bar{e}_{14}^{(F,SBB)}(\epsilon) & = 0.001298 \binom{14}{1} \epsilon^1 (1 - \epsilon)^{13} + 0.004640 \binom{14}{2} \epsilon^2 (1 - \epsilon)^{12} + 0.005990 \binom{14}{3} \epsilon^3 (1 - \epsilon)^{11} \\
& + 0.007519 \binom{14}{4} \epsilon^4 (1 - \epsilon)^{10} + 0.008532 \binom{14}{5} \epsilon^5 (1 - \epsilon)^9 + 0.009695 \binom{14}{6} \epsilon^6 (1 - \epsilon)^8 \\
& + 0.010583 \binom{14}{7} \epsilon^7 (1 - \epsilon)^7 + 0.011598 \binom{14}{8} \epsilon^8 (1 - \epsilon)^6 + 0.012359 \binom{14}{9} \epsilon^9 (1 - \epsilon)^5 \\
& + 0.013543 \binom{14}{10} \epsilon^{10} (1 - \epsilon)^4 + 0.014208 \binom{14}{11} \epsilon^{11} (1 - \epsilon)^3 + 0.016449 \binom{14}{12} \epsilon^{12} (1 - \epsilon)^2 \\
& + 0.016871 \binom{14}{13} \epsilon^{13} (1 - \epsilon)^1 + 0.254365 \epsilon^{14}
\end{aligned}$$

$$\begin{aligned}
h_{15}^{(F,SBB)}(\epsilon) &= 0.254081(1-\epsilon)^{15} + 0.016939\binom{15}{1}\epsilon^1(1-\epsilon)^{14} + 0.017055\binom{15}{2}\epsilon^2(1-\epsilon)^{13} \\
&\quad + 0.017717\binom{15}{3}\epsilon^3(1-\epsilon)^{12} + 0.017592\binom{15}{4}\epsilon^4(1-\epsilon)^{11} + 0.017847\binom{15}{5}\epsilon^5(1-\epsilon)^{10} \\
&\quad + 0.017963\binom{15}{6}\epsilon^6(1-\epsilon)^9 + 0.017946\binom{15}{7}\epsilon^7(1-\epsilon)^8 + 0.017946\binom{15}{8}\epsilon^8(1-\epsilon)^7 \\
&\quad + 0.017963\binom{15}{9}\epsilon^9(1-\epsilon)^6 + 0.017847\binom{15}{10}\epsilon^{10}(1-\epsilon)^5 + 0.017592\binom{15}{11}\epsilon^{11}(1-\epsilon)^4 \\
&\quad + 0.017717\binom{15}{12}\epsilon^{12}(1-\epsilon)^3 + 0.017055\binom{15}{13}\epsilon^{13}(1-\epsilon)^2 + 0.016939\binom{15}{14}\epsilon^{14}(1-\epsilon)^1 + 0.254081\epsilon^{15} \\
\bar{e}_{15}^{(F,SBB)}(\epsilon) &= 0.001129\binom{15}{1}\epsilon^1(1-\epsilon)^{14} + 0.003541\binom{15}{2}\epsilon^2(1-\epsilon)^{13} + 0.005027\binom{15}{3}\epsilon^3(1-\epsilon)^{12} \\
&\quad + 0.006017\binom{15}{4}\epsilon^4(1-\epsilon)^{11} + 0.006997\binom{15}{5}\epsilon^5(1-\epsilon)^{10} + 0.007850\binom{15}{6}\epsilon^6(1-\epsilon)^9 \\
&\quad + 0.008601\binom{15}{7}\epsilon^7(1-\epsilon)^8 + 0.009345\binom{15}{8}\epsilon^8(1-\epsilon)^7 + 0.010113\binom{15}{9}\epsilon^9(1-\epsilon)^6 \\
&\quad + 0.010850\binom{15}{10}\epsilon^{10}(1-\epsilon)^5 + 0.011574\binom{15}{11}\epsilon^{11}(1-\epsilon)^4 + 0.012691\binom{15}{12}\epsilon^{12}(1-\epsilon)^3 \\
&\quad + 0.013513\binom{15}{13}\epsilon^{13}(1-\epsilon)^2 + 0.015810\binom{15}{14}\epsilon^{14}(1-\epsilon)^1 + 0.254081\epsilon^{15} \\
h_{16}^{(F,SBB)}(\epsilon) &= 0.253832(1-\epsilon)^{16} + 0.015865\binom{16}{1}\epsilon^1(1-\epsilon)^{15} + 0.018639\binom{16}{2}\epsilon^2(1-\epsilon)^{14} \\
&\quad + 0.018075\binom{16}{3}\epsilon^3(1-\epsilon)^{13} + 0.018841\binom{16}{4}\epsilon^4(1-\epsilon)^{12} + 0.018735\binom{16}{5}\epsilon^5(1-\epsilon)^{11} \\
&\quad + 0.019043\binom{16}{6}\epsilon^6(1-\epsilon)^{10} + 0.019007\binom{16}{7}\epsilon^7(1-\epsilon)^9 + 0.019174\binom{16}{8}\epsilon^8(1-\epsilon)^8 \\
&\quad + 0.019007\binom{16}{9}\epsilon^9(1-\epsilon)^7 + 0.019043\binom{16}{10}\epsilon^{10}(1-\epsilon)^6 + 0.018735\binom{16}{11}\epsilon^{11}(1-\epsilon)^5 \\
&\quad + 0.018841\binom{16}{12}\epsilon^{12}(1-\epsilon)^4 + 0.018075\binom{16}{13}\epsilon^{13}(1-\epsilon)^3 + 0.018639\binom{16}{14}\epsilon^{14}(1-\epsilon)^2 \\
&\quad + 0.015865\binom{16}{15}\epsilon^{15}(1-\epsilon)^1 + 0.253832\epsilon^{16} \\
\bar{e}_{16}^{(F,SBB)}(\epsilon) &= 0.000992\binom{16}{1}\epsilon^1(1-\epsilon)^{15} + 0.003711\binom{16}{2}\epsilon^2(1-\epsilon)^{14} + 0.004890\binom{16}{3}\epsilon^3(1-\epsilon)^{13} \\
&\quad + 0.006173\binom{16}{4}\epsilon^4(1-\epsilon)^{12} + 0.007048\binom{16}{5}\epsilon^5(1-\epsilon)^{11} + 0.007997\binom{16}{6}\epsilon^6(1-\epsilon)^{10} \\
&\quad + 0.008756\binom{16}{7}\epsilon^7(1-\epsilon)^9 + 0.009587\binom{16}{8}\epsilon^8(1-\epsilon)^8 + 0.010252\binom{16}{9}\epsilon^9(1-\epsilon)^7 \\
&\quad + 0.011046\binom{16}{10}\epsilon^{10}(1-\epsilon)^6 + 0.011687\binom{16}{11}\epsilon^{11}(1-\epsilon)^5 + 0.012668\binom{16}{12}\epsilon^{12}(1-\epsilon)^4 \\
&\quad + 0.013185\binom{16}{13}\epsilon^{13}(1-\epsilon)^3 + 0.014928\binom{16}{14}\epsilon^{14}(1-\epsilon)^2 + 0.014873\binom{16}{15}\epsilon^{15}(1-\epsilon)^1 + 0.253832\epsilon^{16}
\end{aligned}$$

$$h_4^{(H, OBB)}(\epsilon) = \frac{1}{4}(1-\epsilon)^4 + \frac{1}{16}\binom{4}{1}\epsilon^1(1-\epsilon)^3 + \frac{1}{16}\binom{4}{2}\epsilon^2(1-\epsilon)^2 + \frac{3}{32}\binom{4}{3}\epsilon^3(1-\epsilon)^1 + \frac{3}{32}\epsilon^4$$

$$\bar{e}_4^{(H, OBB)}(\epsilon) = \frac{1}{64}\binom{4}{1}\epsilon^1(1-\epsilon)^3 + \frac{1}{32}\binom{4}{2}\epsilon^2(1-\epsilon)^2 + \frac{9}{128}\binom{4}{3}\epsilon^3(1-\epsilon)^1 + \frac{3}{32}\epsilon^4$$

$$\begin{aligned} h_8^{(H, OBB)}(\epsilon) = & 0.257446(1-\epsilon)^8 + 0.032181\binom{8}{1}\epsilon^1(1-\epsilon)^7 + 0.037857\binom{8}{2}\epsilon^2(1-\epsilon)^6 \\ & + 0.042412\binom{8}{3}\epsilon^3(1-\epsilon)^5 + 0.044759\binom{8}{4}\epsilon^4(1-\epsilon)^4 + 0.046648\binom{8}{5}\epsilon^5(1-\epsilon)^3 \\ & + 0.048082\binom{8}{6}\epsilon^6(1-\epsilon)^2 + 0.049087\binom{8}{7}\epsilon^7(1-\epsilon)^1 + 0.049087\epsilon^8 \end{aligned}$$

$$\begin{aligned} \bar{e}_8^{(H, OBB)}(\epsilon) = & 0.004023\binom{8}{1}\epsilon^1(1-\epsilon)^7 + 0.012783\binom{8}{2}\epsilon^2(1-\epsilon)^6 + 0.021266\binom{8}{3}\epsilon^3(1-\epsilon)^5 \\ & + 0.027012\binom{8}{4}\epsilon^4(1-\epsilon)^4 + 0.032374\binom{8}{5}\epsilon^5(1-\epsilon)^3 + 0.037564\binom{8}{6}\epsilon^6(1-\epsilon)^2 \\ & + 0.042951\binom{8}{7}\epsilon^7(1-\epsilon)^1 + 0.049087\epsilon^8 \end{aligned}$$

$$\begin{aligned} h_{16}^{(H, OBB)}(\epsilon) = & 0.253832(1-\epsilon)^{16} + 0.015865\binom{16}{1}\epsilon^1(1-\epsilon)^{15} + 0.017582\binom{16}{2}\epsilon^2(1-\epsilon)^{14} \\ & + 0.018614\binom{16}{3}\epsilon^3(1-\epsilon)^{13} + 0.019368\binom{16}{4}\epsilon^4(1-\epsilon)^{12} + 0.020048\binom{16}{5}\epsilon^5(1-\epsilon)^{11} \\ & + 0.020663\binom{16}{6}\epsilon^6(1-\epsilon)^{10} + 0.021217\binom{16}{7}\epsilon^7(1-\epsilon)^9 + 0.021716\binom{16}{8}\epsilon^8(1-\epsilon)^8 \\ & + 0.022162\binom{16}{9}\epsilon^9(1-\epsilon)^7 + 0.022559\binom{16}{10}\epsilon^{10}(1-\epsilon)^6 + 0.022907\binom{16}{11}\epsilon^{11}(1-\epsilon)^5 \\ & + 0.023205\binom{16}{12}\epsilon^{12}(1-\epsilon)^4 + 0.023450\binom{16}{13}\epsilon^{13}(1-\epsilon)^3 + 0.023633\binom{16}{14}\epsilon^{14}(1-\epsilon)^2 \\ & + 0.023738\binom{16}{15}\epsilon^{15}(1-\epsilon)^1 + 0.023738\epsilon^{16} \end{aligned}$$

$$\begin{aligned} \bar{e}_{16}^{(H, OBB)}(\epsilon) = & 0.000992\binom{16}{1}\epsilon^1(1-\epsilon)^{15} + 0.003579\binom{16}{2}\epsilon^2(1-\epsilon)^{14} + 0.005590\binom{16}{3}\epsilon^3(1-\epsilon)^{13} \\ & + 0.007241\binom{16}{4}\epsilon^4(1-\epsilon)^{12} + 0.008802\binom{16}{5}\epsilon^5(1-\epsilon)^{11} + 0.010285\binom{16}{6}\epsilon^6(1-\epsilon)^{10} \\ & + 0.011704\binom{16}{7}\epsilon^7(1-\epsilon)^9 + 0.013068\binom{16}{8}\epsilon^8(1-\epsilon)^8 + 0.014391\binom{16}{9}\epsilon^9(1-\epsilon)^7 \\ & + 0.015684\binom{16}{10}\epsilon^{10}(1-\epsilon)^6 + 0.016962\binom{16}{11}\epsilon^{11}(1-\epsilon)^5 + 0.018239\binom{16}{12}\epsilon^{12}(1-\epsilon)^4 \\ & + 0.019533\binom{16}{13}\epsilon^{13}(1-\epsilon)^3 + 0.020863\binom{16}{14}\epsilon^{14}(1-\epsilon)^2 + 0.022255\binom{16}{15}\epsilon^{15}(1-\epsilon)^1 + 0.023738\epsilon^{16} \end{aligned}$$

$$h_4^{(H, SBB)}(\epsilon) = \frac{1}{4}(1-\epsilon)^4 + \frac{1}{16}\binom{4}{1}\epsilon^1(1-\epsilon)^3 + \frac{1}{8}\binom{4}{2}\epsilon^2(1-\epsilon)^2 + \frac{1}{16}\binom{4}{3}\epsilon^3(1-\epsilon)^1 + \frac{1}{4}\epsilon^4$$

$$\bar{e}_4^{(H, SBB)}(\epsilon) = \frac{1}{64}\binom{4}{1}\epsilon^1(1-\epsilon)^3 + \frac{1}{16}\binom{4}{2}\epsilon^2(1-\epsilon)^2 + \frac{3}{64}\binom{4}{3}\epsilon^3(1-\epsilon)^1 + \frac{1}{4}\epsilon^4$$

$$\begin{aligned}
h_8^{(H,SBB)}(\epsilon) &= 0.257446(1-\epsilon)^8 + 0.032181\binom{8}{1}\epsilon^1(1-\epsilon)^7 + 0.075714\binom{8}{2}\epsilon^2(1-\epsilon)^6 \\
&+ 0.037857\binom{8}{3}\epsilon^3(1-\epsilon)^5 + 0.062317\binom{8}{4}\epsilon^4(1-\epsilon)^4 + 0.037857\binom{8}{5}\epsilon^5(1-\epsilon)^3 \\
&+ 0.075714\binom{8}{6}\epsilon^6(1-\epsilon)^2 + 0.032181\binom{8}{7}\epsilon^7(1-\epsilon)^1 + 0.257446\epsilon^8
\end{aligned}$$

$$\begin{aligned}
\bar{e}_8^{(H,SBB)}(\epsilon) &= 0.004023\binom{8}{1}\epsilon^1(1-\epsilon)^7 + 0.025566\binom{8}{2}\epsilon^2(1-\epsilon)^6 + 0.015856\binom{8}{3}\epsilon^3(1-\epsilon)^5 \\
&+ 0.031158\binom{8}{4}\epsilon^4(1-\epsilon)^4 + 0.022001\binom{8}{5}\epsilon^5(1-\epsilon)^3 + 0.050148\binom{8}{6}\epsilon^6(1-\epsilon)^2 \\
&+ 0.028158\binom{8}{7}\epsilon^7(1-\epsilon)^1 + 0.257446\epsilon^8
\end{aligned}$$

$$\begin{aligned}
h_{16}^{(H,SBB)}(\epsilon) &= 0.253832(1-\epsilon)^{16} + 0.015865\binom{16}{1}\epsilon^1(1-\epsilon)^{15} + 0.035163\binom{16}{2}\epsilon^2(1-\epsilon)^{14} \\
&+ 0.017966\binom{16}{3}\epsilon^3(1-\epsilon)^{13} + 0.022741\binom{16}{4}\epsilon^4(1-\epsilon)^{12} + 0.018697\binom{16}{5}\epsilon^5(1-\epsilon)^{11} \\
&+ 0.020894\binom{16}{6}\epsilon^6(1-\epsilon)^{10} + 0.018984\binom{16}{7}\epsilon^7(1-\epsilon)^9 + 0.020593\binom{16}{8}\epsilon^8(1-\epsilon)^8 \\
&+ 0.018984\binom{16}{9}\epsilon^9(1-\epsilon)^7 + 0.020894\binom{16}{10}\epsilon^{10}(1-\epsilon)^6 + 0.018697\binom{16}{11}\epsilon^{11}(1-\epsilon)^5 \\
&+ 0.022741\binom{16}{12}\epsilon^{12}(1-\epsilon)^4 + 0.017966\binom{16}{13}\epsilon^{13}(1-\epsilon)^3 + 0.035163\binom{16}{14}\epsilon^{14}(1-\epsilon)^2 \\
&+ 0.015865\binom{16}{15}\epsilon^{15}(1-\epsilon)^1 + 0.253832\epsilon^{16}
\end{aligned}$$

$$\begin{aligned}
\bar{e}_{16}^{(H,SBB)}(\epsilon) &= 0.000992\binom{16}{1}\epsilon^1(1-\epsilon)^{15} + 0.007158\binom{16}{2}\epsilon^2(1-\epsilon)^{14} + 0.004791\binom{16}{3}\epsilon^3(1-\epsilon)^{13} \\
&+ 0.007453\binom{16}{4}\epsilon^4(1-\epsilon)^{12} + 0.007019\binom{16}{5}\epsilon^5(1-\epsilon)^{11} + 0.008774\binom{16}{6}\epsilon^6(1-\epsilon)^{10} \\
&+ 0.008742\binom{16}{7}\epsilon^7(1-\epsilon)^9 + 0.010296\binom{16}{8}\epsilon^8(1-\epsilon)^8 + 0.010242\binom{16}{9}\epsilon^9(1-\epsilon)^7 \\
&+ 0.012120\binom{16}{10}\epsilon^{10}(1-\epsilon)^6 + 0.011678\binom{16}{11}\epsilon^{11}(1-\epsilon)^5 + 0.015289\binom{16}{12}\epsilon^{12}(1-\epsilon)^4 \\
&+ 0.013174\binom{16}{13}\epsilon^{13}(1-\epsilon)^3 + 0.028005\binom{16}{14}\epsilon^{14}(1-\epsilon)^2 + 0.014873\binom{16}{15}\epsilon^{15}(1-\epsilon)^1 + 0.253832\epsilon^{16}
\end{aligned}$$

-
- [1] E. Knill, G. Ortiz, and R. D. Somma, Optimal quantum measurements of expectation values of observables, *Phys. Rev. A* **75**, 012328 (2007).
- [2] D. E. Browne and T. Rudolph, Resource-efficient linear optical quantum computation, *Phys. Rev. Lett.* **95**, 010501 (2005).
- [3] S. Bartolucci, P. Birchall, H. Bombin, H. Cable, C. Dawson, M. Gimeno-Segovia, E. Johnston, K. Kieling,

- N. Nickerson, M. Pant, *et al.*, Fusion-based quantum computation, *Nat. Commun.* **14**, 912 (2023).
- [4] C. Sparrow, *Quantum Interference in Universal Linear Optical Devices for Quantum Computation and Simulation* (Imperial College London, London, England, 2017).
- [5] R. D. Shaw, A. E. Jones, P. Yard, and A. Laing, Errors in heralded circuits for linear optical entanglement generation, [arXiv:2305.08452](https://arxiv.org/abs/2305.08452).
- [6] 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).
- [7] 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).
- [8] J. Marshall, Distillation of indistinguishable photons, *Phys. Rev. Lett.* **129**, 213601 (2022).
- [9] C. F. Faurby, L. Carosini, H. Cao, P. I. Sund, L. M. Hansen, F. Giorgino, A. B. Villadsen, S. N. Van den Hoven, P. Lodahl, S. Paesani, *et al.*, Purifying photon indistinguishability through quantum interference, *Phys. Rev. Lett.* **133**, 033604 (2024).
- [10] M. Halder, A. Beveratos, R. T. Thew, C. Jorel, H. Zbinden, and N. Gisin, High coherence photon pair source for quantum communication, *New J. Phys.* **10**, 023027 (2008).
- [11] Y. Tsujimoto, Y. Sugiura, M. Tanaka, R. Ikuta, S. Miki, T. Yamashita, H. Terai, M. Fujiwara, T. Yamamoto, M. Koashi, *et al.*, High visibility Hong-Ou-Mandel interference via a time-resolved coincidence measurement, *Opt. Express* **25**, 12069 (2017).
- [12] J. L. Tambasco, G. Corrielli, R. J. Chapman, A. Crespi, O. Zilberberg, R. Osellame, and A. Peruzzo, Quantum interference of topological states of light, *Sci. Adv.* **4**, eaat3187 (2018).
- [13] 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).
- [14] H. Ollivier, S. E. Thomas, S. C. Wein, I. M. de Buy Weninger, N. Coste, J. C. Lored, N. Somaschi, A. Harouri, A. Lemaitre, I. Sagnes, *et al.*, Hong-Ou-Mandel interference with imperfect single photon sources, *Phys. Rev. Lett.* **126**, 063602 (2021).
- [15] F. H. B. Somhorst, R. van der Meer, M. Correa Anguita, R. Schadow, H. J. Snijders, M. de Goede, B. Kassenberg, P. Venderbosch, C. Taballione, J. P. Epping, *et al.*, Quantum simulation of thermodynamics in an integrated quantum photonic processor, *Nat. Commun.* **14**, 3895 (2023).
- [16] K. Alexander, A. Bahgat, A. Benyamini, D. Black, D. Bonneau, S. Burgos, B. Burrage, G. Campbell, G. Catalano, A. Ceballos, *et al.*, A manufacturable platform for photonic quantum computing, *arXiv:2404.17570*.
- [17] M. C. Tichy, M. Tiersch, F. de Melo, F. Mintert, and A. Buchleitner, Zero-transmission law for multiport beam splitters, *Phys. Rev. Lett.* **104**, 220405 (2010).
- [18] A. Crespi, Suppression laws for multiparticle interference in Sylvester interferometers, *Phys. Rev. A* **91**, 013811 (2015).
- [19] C. Dittel, G. Dufour, M. Walschaers, G. Weihs, A. Buchleitner, and R. Keil, Totally destructive interference for permutation-symmetric many-particle states, *Phys. Rev. A* **97**, 062116 (2018).
- [20] M. C. Tichy, Sampling of partially distinguishable bosons and the relation to the multidimensional permanent, *Phys. Rev. A* **91**, 022316 (2015).
- [21] M. C. Tichy, K. Mayer, A. Buchleitner, and K. Mølmer, Stringent and efficient assessment of boson-sampling devices, *Phys. Rev. Lett.* **113**, 020502 (2014).
- [22] N. Viggianiello, F. Flamini, L. Innocenti, D. Cozzolino, M. Bentivegna, N. Spagnolo, A. Crespi, D. J. Brod, E. F. Galvão, R. Osellame, *et al.*, Experimental generalized quantum suppression law in Sylvester interferometers, *New J. Phys.* **20**, 033017 (2018).
- [23] F. H. B. Somhorst, B. K. Sauër, S. N. van den Hoven, and J. J. Renema, Photon distillation schemes with reduced resource costs based on multiphoton Fourier interference, *arXiv:2404.14262*.
- [24] M. Englbrecht, T. Kraft, C. Dittel, A. Buchleitner, G. Giedke, and B. Kraus, Indistinguishability of identical bosons from a quantum information theory perspective, *Phys. Rev. Lett.* **132**, 050201 (2024).
- [25] J. Saied, J. Marshall, N. Anand, S. Grabbe, and E. G. Rieffel, in *Quantum Computing, Communication, and Simulation IV* (SPIE, 2024), Vol. 12911, p. 37.
- [26] P. P. Rohde and T. C. Ralph, Error models for mode mismatch in linear optics quantum computing, *Phys. Rev. A* **73**, 062312 (2006).
- [27] H. Bombin, C. Dawson, R. V. Mishmash, N. Nickerson, F. Pastawski, and S. Roberts, Logical blocks for fault-tolerant topological quantum computation, *PRX Quantum* **4**, 020303 (2023).
- [28] H. Bombin, C. Dawson, N. Nickerson, M. Pant, and J. Sullivan, Increasing error tolerance in quantum computers with dynamic bias arrangement, *arXiv:2303.16122*.
- [29] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, *Phys. Rev. A—At. Mol. Opt. Phys.* **86**, 032324 (2012).
- [30] D. Litinski and N. Nickerson, Active volume: An architecture for efficient fault-tolerant quantum computers with limited non-local connections, *arXiv:2211.15465*.
- [31] OEIS Foundation Inc., Entry A277458 in the On-Line Encyclopedia of Integer Sequences (2024). Published electronically at <https://oeis.org/A277458>.
- [32] B. Salvy, Examples of automatic asymptotic expansions, *ACM Sigsum Bulletin* **25**, 4 (1991).
- [33] S. R. Valluri, D. J. Jeffrey, and R. M. Corless, Some applications of the Lambert W function to physics, *Can. J. Phys.* **78**, 823 (2000).
- [34] F. V. Mendes, C. Lima, and R. V. Ramos, Applications of the Lambert–Tsallis W_q function in quantum photonic Gaussian boson sampling, *Quantum Inf. Process.* **21**, 215 (2022).
- [35] M. E. O. Bezerra and V. Shchesnovich, Families of bosonic suppression laws beyond the permutation symmetry principle, *New J. Phys.* **25**, 093047 (2023).
- [36] M. Oszmaniec and D. J. Brod, Classical simulation of photonic linear optics with lost particles, *New J. Phys.* **20**, 092002 (2018).
- [37] R. Barak and Y. Ben-Aryeh, Quantum fast Fourier transform and quantum computation by linear optics, *JOSA B* **24**, 231 (2007).
- [38] S. Aaronson and A. Arkhipov, 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. 333.
- [39] H. Thomas, The number of terms in the permanent and the determinant of a generic circulant matrix, *J. Algebr. Comb.* **20**, 55 (2004).

- [40] L. Colarte, E. Mezzetti, R. Miró-Roig, and M. Salat, On the coefficients of the permanent and the determinant of a circulant matrix: Applications, *Proc. Am. Math. Soc.* **147**, 547 (2019).
- [41] J. Malenfant, On the matrix-element expansion of a circulant determinant, [arXiv:1502.06012](https://arxiv.org/abs/1502.06012).
- [42] V. V. Kocharovskiy and V. V. Kocharovskiy, Exact general solution to the three-dimensional Ising model and a self-consistency equation for the nearest-neighbors' correlations, [arXiv:1510.07327](https://arxiv.org/abs/1510.07327).
- [43] M. Ledoux, The Concentration of Measure Phenomenon (2005). <https://www.ams.org/surv/089>.
- [44] S. Nezami, Permanent of random matrices from representation theory: Moments, numerics, concentration, and comments on hardness of boson-sampling, [arXiv:2104.06423](https://arxiv.org/abs/2104.06423).
- [45] B. Go, C. Oh, L. Jiang, and H. Jeong, Exploring shallow-depth boson sampling: Towards scalable quantum supremacy, *Phys. Rev. A* **109**, 052613 (2024).
- [46] J. Marshall and N. Anand, Simulation of quantum optics by coherent state decomposition, *Optica Quantum* **1**, 78 (2023).
- [47] OEIS Foundation Inc., Entry A002729 in the On-Line Encyclopedia of Integer Sequences (2024). Published electronically at <https://oeis.org/A002729>.