

Wave-Function Engineering for Spectrally Uncorrelated Biphotons in the Telecommunication Band Based on a Machine-Learning Framework

Chaohan Cui^{1,*}, Reeshad Arian,^{2,3} Saikat Guha,^{1,2} N. Peyghambarian,^{1,4} Quntao Zhuang,^{2,1} and Zheshen Zhang^{4,1}

¹*James C. Wyant College of Optical Sciences, The University of Arizona, Tucson, Arizona 85721, USA*

²*Department of Electrical and Computer Engineering, The University of Arizona, Tucson, Arizona 85721, USA*

³*Department of Mathematics and Computational Sciences, The University of Arizona, Tucson, Arizona 85721, USA*

⁴*Department of Materials Science and Engineering, The University of Arizona, Tucson, Arizona 85721, USA*

 (Received 22 May 2019; revised manuscript received 30 July 2019; published 30 September 2019)

Indistinguishable single photons are key ingredients for a plethora of quantum-information-processing applications, ranging from quantum communications to photonic quantum computing. A mainstream platform to produce indistinguishable single photons over a wide spectral range is based on biphoton generation through spontaneous parametric down-conversion in nonlinear crystals. The purity of the biphotons produced is, however, limited by their spectral correlations. Here we present a design recipe, based on a machine-learning framework, for the engineering of biphoton joint spectral amplitudes over a wide spectral range. By customizing the poling profile of the KTiOPO_4 crystal, we show, numerically, that spectral purities of 99.22%, 99.99%, and 99.82%, respectively, can be achieved in the 1310-, 1550-, and 1600-nm bands after applying a moderate 8-nm filter. The machine-learning framework thus enables the generation of near-indistinguishable single photons over the entire telecommunication band without resorting to the KTiOPO_4 crystal's group-velocity-matching wavelength window near 1582 nm.

DOI: [10.1103/PhysRevApplied.12.034059](https://doi.org/10.1103/PhysRevApplied.12.034059)

I. INTRODUCTION

Quantum-information science is an emerging area of study that creates new opportunities for the next-generation communication, computing, and sensing applications. Photons are unique quantum-information carriers as they can be transmitted over long distances for entanglement distribution [1,2], secure communication [3,4], and sensing [5–7]. In addition, single photons would be critical resources in near-term quantum-computing devices for, for example, boson sampling [8–12], to demonstrate a performance advantage over any classical computing platforms, also known as the quantum supremacy.

Specifically, the quantum Internet [13,14] will be empowered by single photons that herald the creation of entanglement between network nodes at a distance [15]. Such a capability underpins distributed quantum computing [16–18] and distributed quantum sensing [19–22]. The quality of the heralded entanglement, produced by interference of two single photons on a beam splitter to erase the which-way information, is critically dependent on the indistinguishability and the purity of the two photons. To ensure high performance for quantum-information

applications, the interfering photons need to share the same spectral, temporal, spatial, and polarization states [23,24]. In addition, it is desirable that the single photons are situated in the telecommunication band to leverage the abundant modulation, transmission, and detection devices for long-distance quantum communications.

Nonlinear crystals are widely used to produce entangled and heralded single photons [25–28]. Compared with solid-state single-photon emitters such as quantum dots and nitrogen-vacancy centers, nonlinear crystals enjoy room-temperature operations, the capability of generating photons in the telecommunication band, and the absence of spectral diffusion that degrades the purity of the photons produced. KTiOPO_4 (KTP), in this regard, is a widely used nonlinear-crystal material by virtue of its high nonlinearity and broad transparency window. In particular, KTP possesses a group-velocity-matching (GVM) wavelength of around 1582 nm [26], vouchsafed by its material dispersion in type-II spontaneous parametric down-conversion (SPDC) processes. Such a unique property has been harnessed to generate spectrally uncorrelated biphotons near the telecommunication C band at 1550 nm. To achieve phase matching, two crystal poling strategies have been pursued. In the conventional periodic poling strategy, shown in Fig. 1 (top), the positive and negative polarities

*chaohancui@email.arizona.edu

both constitute half of the duty cycle in each poling period, resulting in a sinc phase-matching profile the side lobes of which limit the spectral purity. As a result, a narrowband filter is typically used to cut off the side lobes, at the cost of reducing the flux and the heralding efficiency. To mitigate the limitation of periodic poling, Brańczyk *et al.* [28] gave a solution by adding another type of poling. Then Dixon *et al.* [29] introduced a customized aperiodic poling profile, illustrated in Fig. 1 (bottom), to achieve a Gaussian phase-matching profile at the GVM wavelength more practically. In conjunction with a Gaussian-spectrum pump, a spectral purity of 99.5% was measured after application of a Gaussian spectral filter with a full width at half maximum of 8.5 nm. The spectral purity of the biphotons produced in the scheme of Dixon *et al.*, however, degrades to 97.12% after application of a 40-nm filter in the 1550-nm band due to the higher-order terms (e.g., group-velocity dispersion) [30]. In follow-up studies [31–36], several poling-design optimization approaches were introduced to improve on the spectral purity in Ref. [29]. Like the scheme of Dixon *et al.*, these approaches require operation under the validity of a first-order approximation. Since KTP crystal has a smooth group-velocity-mismatch around the GVM wavelength, the former designs can still be valid for spectrally uncorrelated biphoton generation over the entire telecommunication window from approximately 1300 nm to 1600 nm [30]. However, the purity is confined as a result of the first-order approximation, which is no longer efficient. In addition, these approaches rely on binary optimization that limits the achievable purity, due to a lack of access to the full parameter space. To generate high-purity biphotons at wavelengths away from 1582 nm, the period-poled KTP’s output spectral purity is limited to merely approximately 81% [37], which is insufficient for many applications. Apart from KTP, Refs. [38–40] compiled a list of other nonlinear materials, each operating at

a specific wavelength dictated by its GVM property. However, a scheme for generating high-spectrally-uncorrelated biphotons at any target wavelength remains elusive.

Here we present a general machine-learning framework that seeks the optimum poling design for the generation of spectrally uncorrelated biphotons. Unlike in prior work, our approach exploits an optimization fully empowered by machine learning to obviate the need for either the first-order approximation or the GVM property of nonlinear crystals. This yields a purity in excess of 99.8% with an 8-nm filter over the range from 1310 to 1600 nm. In particular, a spectral purity of 99% in the 1550-nm band is achieved after application of a 40-nm wideband filter that nicely maintains the flux and the heralding efficiency. Our result demonstrates the power of machine learning in tackling hard quantum-information problems.

In this paper, we focus on optimizing KTP’s type-II phase matching to generate spectrally pure biphotons, but our method can be applied to other phase-matching types.

II. BIPHOTON JOINT SPECTRAL AMPLITUDE AND PHASE MATCHING

In SPDC, the generated biphotons can be represented in the frequency domain as a superposition of different frequency modes:

$$|\psi\rangle_{SI} = \iint d\omega_S d\omega_I f(\omega_S, \omega_I) \hat{a}_{\omega_S}^\dagger \hat{b}_{\omega_I}^\dagger |0\rangle_S |0\rangle_I, \quad (1)$$

where $\hat{a}_{\omega_S}^\dagger$ is the creation operator for the signal photon, $\hat{b}_{\omega_I}^\dagger$ is the creation operator for the idler photon, and $f(\omega_S, \omega_I)$ is the biphoton joint spectral amplitude (JSA), which contains complete information about the spectral-temporal properties of the photon pair. The biphoton JSA is determined by the pump spectrum and the properties of the nonlinear crystal by $f(\omega_S, \omega_I) \propto \alpha(\omega_P) G(\Delta k)$, where $\omega_P = \omega_S + \omega_I$ relates the pump, signal, and idler frequencies by energy conservation, $\alpha(\omega_P)$ describes the pump spectral profile, and $G(\Delta k)$ encompasses information about the phase-matching properties of the nonlinear crystal [26]. Specifically, the phase-matching function

$$G(\Delta k) = \frac{1}{L} \int_0^L dz g(z) \exp(-i\Delta k z), \quad (2)$$

where $g(z) = \{1, -1\}$, describes the poling profile along the propagation z axis. The phase mismatch

$$\begin{aligned} \Delta k(\omega_S, \omega_I) &= k_P(\omega_P) - k_S(\omega_S) - k_I(\omega_I) \\ &= 2\pi n(\omega_P)/\lambda_P - 2\pi n(\omega_S)/\lambda_S - 2\pi n(\omega_I)/\lambda_I, \end{aligned} \quad (3)$$

where $n(\omega)$ is the frequency-dependent refractive index, which determines the material dispersion, and λ_P , λ_S , and

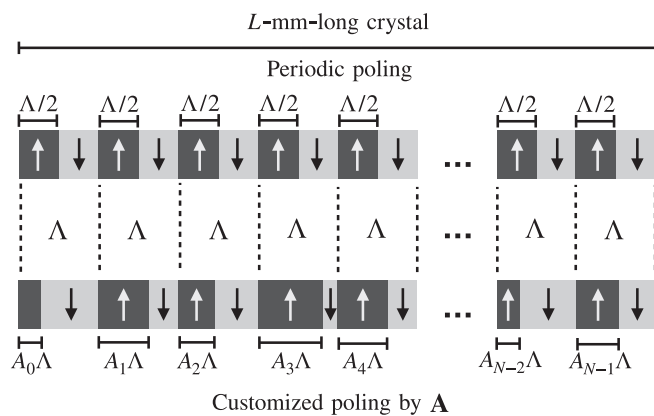


FIG. 1. Poling profiles for a L -mm-long nonlinear crystal with N poling periods: periodic poling with a period of Λ (top) and customized poling with a profile embedded in the array \mathbf{A} (bottom).

λ_I are the pump, signal, and idler wavelengths [26,29,35]. To quantitatively describe the spectral correlation between the signal and idler photons, their JSA is decomposed [29,36] as $f(\omega_S, \omega_I) = \sum_n \xi_n \beta_{S,n}(\omega_S) \beta_{I,n}(\omega_I)$, where $\{\beta_{S,n}(\omega_S)\}$ and $\{\beta_{I,n}(\omega_I)\}$ are two sets of orthogonal functions, viz.,

$$\int d\omega \beta_{j,n}^*(\omega) \beta_{j,m}(\omega) = \delta_{n,m}, \quad (4)$$

where $j \in \{S, I\}$. The purity is then defined as

$$\mathcal{P} = \sum_n |\xi_n|^4 / \left(\sum_n |\xi_n|^2 \right)^2. \quad (5)$$

Equivalently, $\mathcal{P} = \text{Tr}[\text{Tr}_I(|\psi\rangle\langle\psi|_{SI})^2]$, which quantifies how pure the signal-photon state is after tracing out the idler photon. If $f(\omega_S, \omega_I) = \beta_S(\omega_S) \beta_I(\omega_I)$ (i.e., $\mathcal{P} = 1$), the JSA then describes a product state of two spectrally uncorrelated photons, which is particularly useful for boson sampling, entanglement distribution, and photonic quantum-information processing [11,27,41].

Another quantitative property of a biphoton JSA is the spectral indistinguishability defined as

$$\mathcal{I} = \text{Re} \left[\iint d\omega_s d\omega_f^* (\omega_s, \omega_i) f(\omega_i, \omega_s) \right], \quad (6)$$

where \mathcal{I} quantifies the difference in the spectra of two photons [42–47] and equals the visibility of Hong-Ou-Mandel interference. \mathcal{I} is a useful metric for single-photon sources such as quantum dots for which spectral diffusion is a practical limitation. In this paper we consider broadband heralded single photons from independent SPDC sources, so the purity is a more-appropriate performance metric (see the Appendix for more information). We do, however, calculate the indistinguishability for the signal and idler photons.

To engineer a desired biphoton JSA $f(\omega_S, \omega_I)$, one has two tunable knobs: the pump spectral profile $\alpha(\omega_P)$ and the crystal poling profile $g(z)$. In this paper, we introduce a general recipe that harnesses a machine-learning framework to automate the design for $\alpha(\omega_P)$ and $g(z)$. We first formulate the JSA engineering problem and provide some insights into its connection with machine learning.

The poling profile embedded in $g(z)$ involves the poling period Λ and an array $\mathbf{A} = \{A_i\}$ that specifies the duty cycle in each of the N periods. Λ is given by the phase-matching condition at the pump's central frequency ω_{P_0} when signal and idler photons are wavelength degenerate at $\omega_{P_0}/2$: $\Lambda = 2\pi/\Delta k_0$, where $\Delta k_0 = \Delta k(\omega_S = \omega_I =$

$\omega_{P_0}/2)$. $g(z)$ thus relates to Λ and \mathbf{A} by

$$g(z, \mathbf{A}) = -1 + 2 \sum_{j=0}^{N-1} \left\{ \Theta(z - j\Lambda) - \Theta[z - (j + A_j)\Lambda] \right\}, \quad (7)$$

where Θ is the unit step function (see Fig. 1). Plugging $g(z)$, given by Eq. (7), into the phase-matching function yields

$$G(\Delta k, \mathbf{A}) = \frac{1}{iL\Delta k} \sum_{j=0}^{N-1} \left[e^{-i\Lambda_j \Delta k} + e^{-i\Lambda(j+1)\Delta k} - 2e^{-i\Lambda(j+A_j)\Delta k} \right]. \quad (8)$$

In practice, the design of $g(z)$ is limited by the minimum poling length Λ_{\min} . To accommodate the practical limitation, we modify the duty cycles as $A_i \in [\Lambda_{\min}/\Lambda, 1 - \Lambda_{\min}/\Lambda]$ (the ratio of the positive-polarity portion within each poling period) in tuning the JSA.

A. The GVM condition and Gaussian phase-matching profile

To generate spectrally uncorrelated biphotons, we desire $f(\omega_S, \omega_I) = \alpha(\omega_S + \omega_I) G[\Delta k(\omega_S, \omega_I)] = \beta_S(\omega_S) \beta_I(\omega_I)$. A conventional approach to engineer such a product-state wave function is picking the signal and idler wavelengths, ω_{S_0} and ω_{I_0} , that satisfy $\partial \Delta k / \partial \omega_S|_{\omega_{S_0}} = -\partial \Delta k / \partial \omega_I|_{\omega_{I_0}} = \gamma_{\text{GVM}}$, known as the GVM condition. Around ω_{S_0} and ω_{I_0} , the phase mismatch is fully determined by the frequency difference between the signal and idler photons; that is, $\Delta k \approx \Delta k_0 + \gamma_{\text{GVM}}(\omega_S - \omega_I)$. This leads to a phase-matching function $G(\Delta k)$ that is a function solely of $\omega_S - \omega_I$, viz., $G(\omega_S - \omega_I)$. Under the GVM condition, it is possible to engineer a Gaussian phase-matching function [30]

$$G(\omega_S - \omega_I) \approx G_0 \exp \left[-\frac{(\omega_S - \omega_I)^2}{\sigma_C^2} \right], \quad (9)$$

where σ_C is determined by the poling profile $g(z|\mathbf{A}_{\text{GVM}})$. In conjunction with a Gaussian pump spectral profile

$$\alpha(\omega_P) = a_0 \exp \left[-\frac{(\omega_S + \omega_I - \omega_{P_0})^2}{\sigma_P^2} \right] \quad (10)$$

and by choosing $\sigma_P = \sigma_C$, one obtains

$$f(\omega_S, \omega_I) = \alpha(\omega_S + \omega_I) G(\omega_S - \omega_I) \propto \exp \left[-\frac{2(\omega_S - \omega_{S_0})^2}{\sigma_P^2} \right] \exp \left[-\frac{2(\omega_I - \omega_{I_0})^2}{\sigma_P^2} \right]; \quad (11)$$

that is, a spectrally uncorrelated product state of biphotons.

The JSA engineering approach based on GVM and Gaussian phase matching, albeit ingenious, is limited by the dispersion properties of the nonlinear materials, resulting in only a handful of GVM wavelengths, each associated with a specific nonlinear optical material. For example, the GVM wavelengths are approximately 1582 nm for KTP, approximately 830 nm for KH_2PO_4 , and approximately 922 nm for $\text{NH}_4\text{H}_2\text{AsO}_4$ [40]. Such a restriction impedes the generation of spectrally uncorrelated biphotons that covers the entire telecommunication band from approximately 1300 nm to approximately 1600 nm and precludes interfacing SPDC photons with solid-state quantum emitters in the visible to near-infrared wavelength range as a means to entangle qubits at a distance. To engineer spectrally uncorrelated biphoton JSAs over a wide spectral range, we need to further understand the limitations of the GVM approach.

B. General phase matching without the GVM condition

We now consider working at non-GVM wavelengths. Let the phase mismatch at the degenerate wavelength be Δk_0 . The phase mismatch at any wavelengths in the vicinity of the degenerate wavelength can be expressed, in a first-order approximation, as $\Delta k - \Delta k_0 \approx \gamma_S \omega_S - \gamma_I \omega_I$, where $\gamma_S \neq \gamma_I$ for a non-GVM case. Without loss of generality, both γ_S and γ_I are chosen positive. The validity of the first-order approximation is verified in Fig. 2, which shows the linearity of Δk around the degenerate wavelengths at the intersections between the solid and dashed lines. If we apply the poling profile \mathbf{A}_{GVM} developed at the GVM wavelengths,

$$\begin{aligned}
 G(\Delta k, \mathbf{A}_{\text{GVM}}) &\approx G'_0 \exp \left[-\frac{(\Delta k - \Delta k_0)^2}{\gamma_{\text{GVM}}^2 \sigma_C^2} \right] \\
 &\times \left[1 + \frac{\Delta k - \Delta k_0}{\Delta k_0} + \mathcal{O}((\Delta k - \Delta k_0)^2) \right] \\
 &\approx G'_0 \exp \left[-\frac{(\gamma_S \omega_S - \gamma_I \omega_I)^2}{\gamma_{\text{GVM}}^2 \sigma_C^2} \right] \\
 &\times \left[1 + \frac{\gamma_S \omega_S - \gamma_I \omega_I}{\Delta k_0} \right. \\
 &\left. + \mathcal{O}((\gamma_S \omega_S - \gamma_I \omega_I)^2) \right], \quad (12)
 \end{aligned}$$

where the second term includes high-order contributions. Since $(\gamma_S \omega_S - \gamma_I \omega_I)/\Delta k_0$ is small but not negligible, it causes a reduced spectral purity. The spectral purity will, nonetheless, still be high after the application of narrowband filters as the higher-order terms primarily contribute to areas of large detuning from the central wavelength. It appears that by choosing the pump bandwidth $\sigma_P = \gamma_{\text{GVM}} \sigma_C / \sqrt{\gamma_S \gamma_I}$, one obtains,

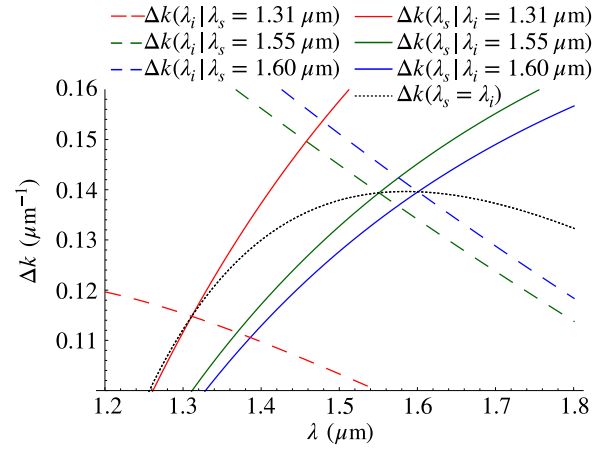


FIG. 2. KTP's type-II SPDC phase mismatch $\Delta k(\omega_S, \omega_I)$ with fixed idler (solid lines) or signal (dashed lines) wavelengths in the 1200–1800-nm range. The intersection points between the solid and dashed lines of the same color are where a first-order approximation for the phase mismatch is performed. The dotted black line is formed by all intersections at different wavelengths, giving the phase mismatch, Δk_0 , at the degenerate wavelengths.

in a non-GVM regime, a product-state JSA $f(\omega_S, \omega_I) \propto \exp[-(\omega_S - \omega_{S_0})^2/\sigma_S^2] \exp[-(\omega_I - \omega_{I_0})^2/\sigma_I^2]$, where

$$\begin{aligned}
 \sigma_S &= \frac{\sigma_C \gamma_{\text{GVM}}}{\sqrt{\gamma_S(\gamma_S + \gamma_I)}}, \\
 \sigma_I &= \frac{\sigma_C \gamma_{\text{GVM}}}{\sqrt{\gamma_I(\gamma_S + \gamma_I)}}. \quad (13)
 \end{aligned}$$

However, a number of caveats about this result need be noted. First, a practical phase-matching function $G(\Delta k, \mathbf{A}_{\text{GVM}})$ is only approximately Gaussian due to the finite number of periods and the minimum duty cycle Λ_{min} , which induces discretization errors. Second, the higher-order Hermite-Gaussian terms in Eq. (12) generate side lobes, which should be accounted for in an effective design. As a result, a 99.5% filtered spectral purity [29] at the GVM wavelength and Gaussian phase matching reduces to 97.13% in the 1550-nm band after application of a 40-nm filter.

III. THE MACHINE-LEARNING FRAMEWORK

We introduce a machine-learning framework to cope with the limitations associated with the GVM and Gaussian-phase-matching approach. The machine-learning framework enables the suppression of higher-order terms and compensations for discretization errors, leading to high-spectral-purity biphotons over a wide range of wavelengths. In addition, the machine-learning framework is capable of designing a poling profile that corrects deviations from a perfect Gaussian-spectrum pump.

A general machine-learning framework comprises multiple building blocks, as illustrated in Fig. 3. A nonlinear

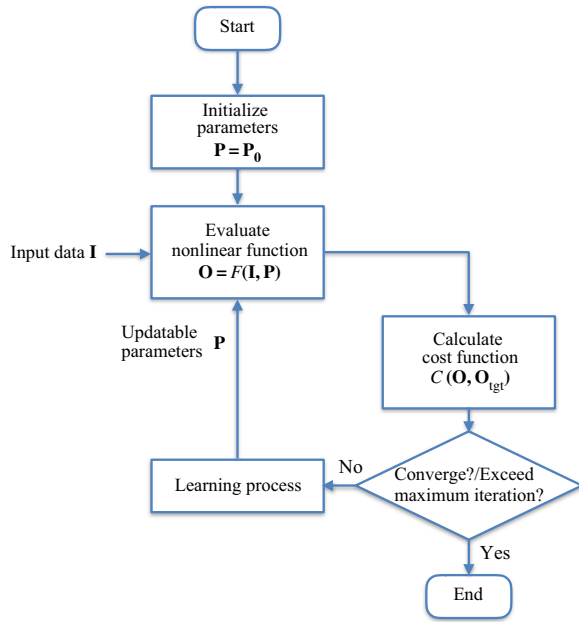


FIG. 3. A general machine-learning model. The learning process is an algorithm that updates the parameters on the basis of the cost. \mathbf{I} , input data; \mathbf{O} , output data; \mathbf{P} , updatable parameters; \mathbf{P}_0 , initial parameters; $F(\mathbf{I}, \mathbf{P})$, nonlinear function; \mathbf{O}_{tgt} , target output data; $C(\mathbf{O}, \mathbf{O}_{\text{tgt}})$, cost function that determines the distance between the present output data and the target output data.

function F is characterized by a set of updatable parameters \mathbf{P} . F takes input data \mathbf{I} and produces output data $\mathbf{O} = F(\mathbf{I}, \mathbf{P})$. The objective of learning is to seek the optimum \mathbf{P} so that \mathbf{O} converges to the target output data \mathbf{O}_{tgt} for different \mathbf{I} . To this end, we define a cost function, $C(\mathbf{O}, \mathbf{O}_{\text{tgt}})$, that quantifies the distance between \mathbf{O} and \mathbf{O}_{tgt} . \mathbf{P} is initialized by a preset of parameters \mathbf{P}_0 and updated by a learning process supplied with the calculated cost in each iteration.

The most-common class of learning processes are based on gradient descent, in which \mathbf{P} is updated on the basis of a linear scaling of the negativity of the gradient of the cost function [48,49]. The linear scaling is defined as the learning rate. Adam is an upgraded version of the gradient-descent learning process. Adam introduces the adaptive momentum method, which adjusts the learning rate on the basis of the learning history [50,51]. Such a feature differentiates Adam from a conventional gradient-descent learning process with a fixed learning rate.

To date, several optimization approaches based on forward propagation have been used to design the poling profile for JSA engineering [31–33]. These approaches rest on equal poling periods and static strategies to optimize the poling profile. In contrast, gradient descent and Adam fall into back-propagation learning processes because the gradient of the cost function is calculated as a partial derivative over elements in \mathbf{P} . Critically, the back-propagation

approach in our machine-learning framework automates the optimization procedure. Moreover, the search time in the parameter space is significantly reduced by virtue of Adam’s adaptive parameter update strategy.

To use the machine-learning framework to engineer the JSAs for spectrally uncorrelated biphotons, we assume that a pump with an ideal Gaussian spectrum and bandwidth is available at the desired wavelength, with the understanding that deviations from the ideal conditions can be accommodated by the machine-learning framework. Consider the function $H(\omega_S, \mathbf{A}) \equiv |G[\Delta k(\omega_S, \omega_{P_0} - \omega_I), \mathbf{A}]|$, that is, a slice of the phase-matching function $|G[\Delta k(\omega_S, \omega_I), \mathbf{A}]|$ along the $\omega_S + \omega_I = \omega_{P_0}$ axis, where ω_{P_0} is the central frequency for the pump. If $H(\omega_S, \mathbf{A})$ is Gaussian, $|G[\Delta k(\omega_S, \omega_I), \mathbf{A}]|$ will have the same form as the first term in Eq. (12). Therefore, our objective is to seek the optimum poling profile \mathbf{A} such that $H(\omega_S, \mathbf{A}) \rightarrow H_0 \exp[-(\omega_S - \omega_{T_0})^2/2\sigma_T^2]$, where ω_{T_0} and σ_T are the central frequency and the standard deviation for this Gaussian target. The purity of the biphotons is tied to the distance between $H(\omega_S, \mathbf{A})$ and an ideal Gaussian form. In the machine-learning framework, the input data \mathbf{I} consist of an array of sampled signal frequencies: $\omega_S = \{\omega_{S_1}, \omega_{S_2}, \dots, \omega_{S_k}\}$. The updatable parameters \mathbf{P} include the duty cycle array \mathbf{A} , the target central frequency ω_{T_0} , and the target bandwidth σ_T . With the updatable parameters, the nonlinear function returns the output data $\mathbf{O} = F(\mathbf{I}, \mathbf{P}) = H(\omega_S, \mathbf{A})$, which aims to approach the target output data:

$$\mathbf{O}_{\text{tgt}} = H_0 \exp\left[-\frac{(\omega_S - \omega_{T_0})^2}{2\sigma_T^2}\right]. \quad (14)$$

The cost is then defined as the distance between the output data and an ideal Gaussian function as

$$C(\mathbf{O}, \mathbf{O}_{\text{tgt}}) = \sum_{l=1}^k \left\{ H(\omega_{S_l}, \mathbf{A}) - H_0 \exp\left[-\frac{(\omega_{S_l} - \omega_{T_0})^2}{2\sigma_T^2}\right] \right\}^2. \quad (15)$$

IV. POLING-DESIGN RECIPE

We customize the poling for an $L = 10$ mm KTP crystal comprises approximately 200 poling periods. Such a size of the parameter space would be a challenge for any analytic attempt to devise the optimum poling profile. The machine-learning framework, by contrast, optimizes the duty cycle array \mathbf{A} by gradually minimizing the cost function over each iteration. In doing so, the optimization does not rely on any approximation, nor does it place any requirement on the pump profile. It can thus cope with any

target output data under any pump profile. In what follows, we describe the four main steps of our poling-design recipe.

Step 1: initialization. Set the target wavelength and the bandwidth of interest. Set the poling period $\Lambda = 2\pi/\Delta k_0$ at the target wavelength. Set the minimum poling length A_{\min} allowed by the fabrication processes. As a reference, $A_{\min} = 0$ is an upper bound for the performance. Initialize $\mathbf{A} = \mathbf{A}_{\text{GVM}}$ and choose the maximum number of machine-learning iterations.

Step 2: cost function calculation. Obtain and normalize $\mathbf{O} = H(\omega_S, \mathbf{A})$. Calculate the cost function $C(\mathbf{O}, \mathbf{O}_{\text{tgt}})$.

Step 3a: pump optimization. Use Adam with learning rate r_a to update ω_{T_0} and σ_T and reduce the cost function.

Step 3b: poling optimization. Use Adam with learning rate r_b to update \mathbf{A} and reduce the cost function. Repeat from step 2 until the maximum number of iterations is reached or the cost function converges to the minimum.

Step 4: poling period adjustment. Fine-tune Λ without modifying \mathbf{A} to eliminate the nondegeneracy between the signal and idler photons.

V. SIMULATION RESULTS

The machine-learning framework for the poling design is realized in PYTHON with the TensorFlow library [52]. The material dispersion profile of KTP is derived by the Sellmeier equation reported in Ref. [53]. The learning rates used are $r_a = 0.005$ and $r_b = 0.015$.

Working in the 1550-nm band, we find the cost converges after 100 machine-learning iterations, as shown in Fig. 4. Figure 5 displays the initial poling profile and the optimized poling profile obtained by the machine-learning framework. The simulation derives a minimum poling length of approximately $0.9 \mu\text{m}$ (2.04% of a poling period), even though we set $A_{\min} = 0$ to obtain the optimum poling profile. The periodic peaks in the machine-learned poling profile may be responsible for the compensation of discretization errors.

Since our main focus is on broadband heralded single photons from independent SPDC sources, the visibility of the Hong-Ou-Mandel interference is primarily determined by the purity [24,42,45–47] (see also the Appendix). As such, we use purity as a main performance metric to compare our design with others [54].

After applying ideal rectangular filters with various bandwidths, we calculate the spectral purity of the biphotons produced by the coefficients of Schmidt decomposition [29,36] and Eq. (5). We apply our recipe to generating spectrally uncorrelated biphotons in the 1310-, 1400-, 1500-, and 1600-nm bands. The spectral purities with filters of various bandwidths are derived for both optimized-poling KTP and periodically poled KTP, as depicted in Fig. 6 and summarized in the inset. Even with a 40-nm filter, our machine-learning-based poling design

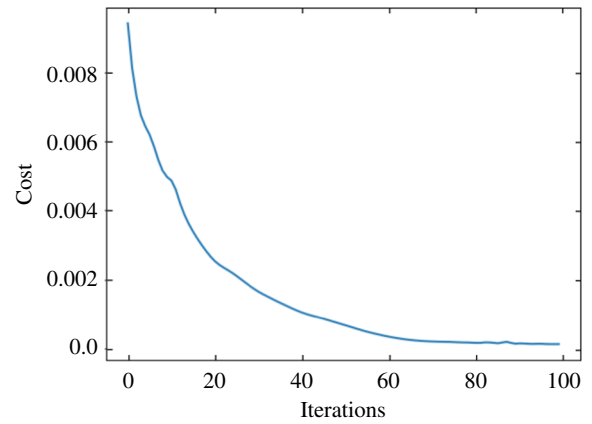


FIG. 4. The evolution of the cost during poling optimization for the generation of spectrally uncorrelated biphotons in the 1550-nm band. The cost approaches 0 after 100 iterations, indicating near-spectrally-uncorrelated biphotons can be produced with the optimized crystal poling.

achieves purities in excess of 98%, marked as the dashed line in Fig. 6) [10,55].

The filters cut the JSA at finite wavelengths and thus simplify the numerical estimation for the purity. The filter bandwidths are critical parameters: Fig. 6 shows that a narrower filter bandwidth leads to a higher purity. A narrowband filter, however, decreases the photon flux and the heralding efficiency due to its increased insertion loss compared with that of a wideband filter. In our simulation at 1550 nm, the output signal's full width at half maximum is about 4.5 nm, so we choose 4, 8, and 40 nm to keep approximately 68%, approximately 95%, and approximately 100% of the photon flux. Filters of the same linewidths are used at other wavelengths. To achieve a high heralding efficiency, our customized poling crystals

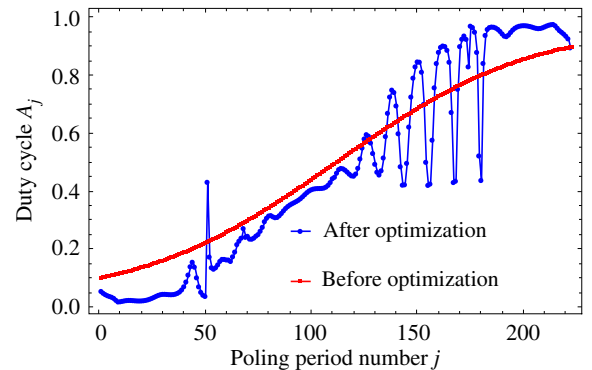


FIG. 5. Poling profiles for the generation of spectrally uncorrelated biphotons in the 1550-nm band. The red curve shows the initial poling profile before application of machine learning, obtained from the GVM condition and Gaussian phase matching [29]. Blue dots show the optimized poling profile obtained by the machine-learning framework.

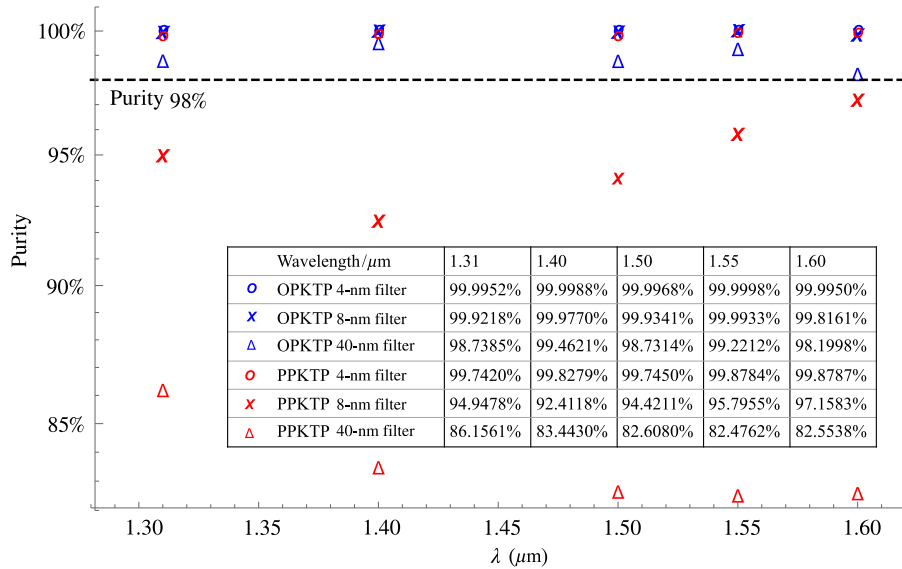


FIG. 6. Simulated purities at different wavelengths after application of filters of different bandwidths. Circles represent purities obtained with our machine-learning-based poling design and crosses represent purities obtained with periodic poling. The inset shows a table that summarizes the purities for both optimized-poling KTP (OPKTP) and periodically poled KTP (PPKTP). For PPKTP, a longer wavelength yields a larger bandwidth for the biphotons. An 8-nm filter then eliminates the first side lobe to increase the purity. The dashed line corresponds to a purity of 98%. The purities obtained by our poling-design recipe in conjunction with a 40-nm filter all sit above the dashed line. The relation between the purity and the filter bandwidth is discussed in detail in the main text.

need be pumped by loosely focused beams [56,57] in conjunction with high-efficiency detectors [58,59]. Since the signal photon and the idler photon are nearly spectrally separable, the heralding efficiency is close to the system efficiency of either arm. Assuming a 5% insertion loss for the optical filter and a 90% quantum efficiency for the detector, the heralding efficiency will be close to 85%.

To further illustrate the poling optimization procedure, Fig. 7 shows the joint spectral intensity, defined as $|f(\omega_S, \omega_I)|^2$, in the 1550-nm band after each step. One observes that side lobes are suppressed after steps 2 and 3, and step 4 eliminates the nondegeneracy between the biphotons.

Key metrics reported in this work and prior work on poling design for spectrally-uncorrelated-photon generation

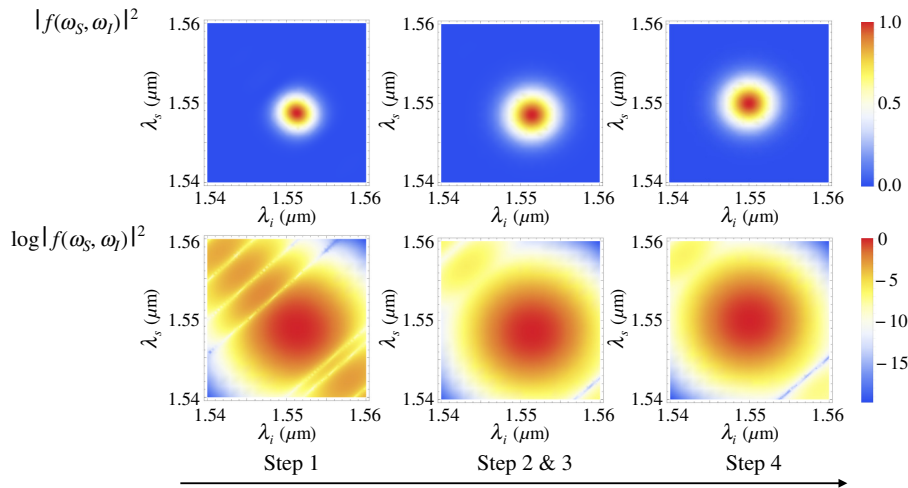


FIG. 7. $|f(\omega_S, \omega_I)|^2$ and $\log |f(\omega_S, \omega_I)|^2$ after each design step. Step 1 takes a prior poling design, which, without any optimization, suffers from side lobes that degrade the purity. After steps 2 and 3, the side lobes are suppressed by the machine-learning framework but the biphotons are left nondegenerate. After step 4, the biphotons become degenerate at 1550 nm. The logarithmic scale amplifies the visibility of the side lobes.

are collected in Table I. Since the different studies use filters of different bandwidths and set different resolutions for JSAs in calculating the purity, the highest purity numbers do not accurately reflect the performance for different schemes. Notably, our machine-learning framework is able to compensate higher-order terms by directly calculating the cost function from the sampled biphoton JSA, whereas prior studies were all based on first-order approximations, as discussed in Sec. III B. Because our machine-learning framework relies on neither a first-order approximation nor the GVM condition, high spectral purity can be achieved over the entire telecommunication band from 1300 to 1600 nm. We note that the biphoton JSA centered at non-GVM wavelengths (e.g., a central wavelength at 1310 nm) is elliptical since $\sigma_S \neq \sigma_I$. The unequal bandwidth of the signal and idler photons, however, does not prevent signal or idler photons from independent sources from interfering, as a building block for entanglement swapping and quantum teleportation.

VI. DISCUSSION

The machine-learning framework is a general optimization strategy particularly suitable for complex problems that have clear objectives but cannot be tackled by approximation methods due to the sensitivity to small variations in their solutions. Problems of this kind typically involve a large parameter space that is hard to solve by conventional optimization methods. Intriguingly, the solutions sought by the machine-learning framework may in turn offer new insights for the complex problems at hand.

The machine-learning framework uses Adam as the learning process. Since Adam converges at a zero-gradient minimum point, the poling profile is robust against small fabrication errors. Adam has shown great performance in many nonconvex optimization problems. To fully unleash the potential of Adam, careful choice of the learning rate is sometimes required. Since one cannot ensure convergence of the machine-learning algorithm to the global optimum, it is recommended to set the initial parameters on the basis of a good existing design and subsequently leverage machine learning to achieve substantial improvement. In this work, we set a known Gaussian poling profile obtained at the GVM wavelength as the initial parameters for the machine-learning framework. In the learning process, a finer sampling resolution will slightly improve the performance at the cost of requiring more computational resources. Apart from KTP, our machine-learning-based design recipe is applicable to nonlinear crystals, as long as a first-order approximation dominates higher-order terms in the phase-mismatch function at the working wavelength.

The machine-learning framework can also be used to seek poling profiles for other forms of biphoton JSAs by simply modifying the target JSA in the cost function. For example, nondegenerate biphoton states and non-Gaussian

TABLE I. Comparison of the performance metrics between prior work and the present work. Numbers with the symbol \sim are inferred on the basis of the context of the references. The number of domains equals the total number of free variables being optimized. The numbers in parentheses denote the final numbers of domains after the combining of domains with the same polarity after optimization. The phase-matching bandwidth is determined by the crystal length. As such, the comparison of the filter bandwidths should be done among crystals with similar lengths.

Reference	Poling strategy	Optimization method	Central wavelength (nm)	Number of domains	Crystal length (mm)	Filter type and bandwidth	Achieved highest purity (%)
Brańczyk <i>et al.</i> [28]	Customized poling order	Analytic design	1576	~ 900	24.2	NA	99.0
Dixon <i>et al.</i> [29]	Customized duty cycle	Analytic design	1582	~ 520	12	Gaussian 8.5 nm	99.5
Tambasco <i>et al.</i> [32]	Customized orientation	Binary coordinate descent	1550	~ 532	12	Rectangular ~ 16 nm [60]	99.6
Dosseva <i>et al.</i> [31]	Customized orientation	Binary simulated annealing	1582	1300	~ 14.1	Rectangular ~ 10 nm [60]	99.9
Graffitti <i>et al.</i> [33]	Customized orientation and tuning periods	Binary coordinate descent & simulated annealing	1582	~ 870 (60)	2	Rectangular 107 nm	99.0
Graffitti <i>et al.</i> [33]	Subcoherence engineering	Binary coordinate descent	1582	~ 870 (502)	2	Rectangular 107 nm	99.4
This work	Customized duty cycle and tuning periods	Adam	1300–1600	~ 400 to ~ 500	10	Rectangular 8 nm	> 99.8

NA, not available.

states can be engineered by our approach as a means for entangling solid-state qubits at a distance [61–63]. To engineer an arbitrary biphoton JSA, a learning process based solely on sampling the signal frequencies becomes insufficient. In such a general situation, one should sample both the signal’s frequencies and the idler’s frequencies, in ω_S and ω_I , and feed them to the nonlinear function to obtain $H(\omega_S, \omega_I|A)$ as the output. The cost function should also be modified accordingly. There are, however, no fundamental constraints that prevent the machine-learning framework from engineering an arbitrary biphoton JSA, but given a finite number of poling periods, the machine-learning framework may approach a target biphoton JSA down to only a certain precision [64].

VII. CONCLUSIONS

We develop a machine-learning framework to solve the problem of the generation of indistinguishable biphotons in the telecommunication band from 1300 to 1600-nm. Our approach leads to a spectral purity in excess of 99.99% for biphotons in the 1550-nm band after application of a 8-nm filter and in excess of 98% over the entire telecommunication band while maintaining a high photon flux and heralding efficiency by using a wide 40-nm filter. This work demonstrates the potential of machine learning to advance quantum-information science. We hope that this work will spur the pursuit of other machine-learning-enhanced quantum-communication, sensing, and information-processing applications.

ACKNOWLEDGMENTS

This work was funded by National Science Foundation Major Research Instrumentation Program Award No. 1828132 with matching funds provided by the University of Arizona. C.C. acknowledges the Nicolaas Bloembergen Graduate Student Scholarship. Z.Z. is grateful for support from the University of Arizona. S.G. acknowledges the Office of Naval Research program Communications and Networking with Quantum Operationally-Secure Technology for Maritime Deployment (CONQUEST), awarded under Raytheon BBN Technologies Prime Contract No. N00014-16-C2069, and a subcontract to the University of Arizona. N.P. acknowledges support from Regents Innovation Funds. Q.Z. is supported by the University of Arizona. We thank F. Graffitti and A.M. Brańczyk for helpful discussions.

APPENDIX: DEFINITION OF THE PURITY AND INDISTINGUISHABILITY OF THE BIPHOTON STATE

1. Purity

In general, a biphoton state can be described by a density matrix ρ_{12} . A common definition for the biphoton spectral

purity is the purity of one photon after tracing out the other:

$$\mathcal{P} \equiv \text{Tr} \{ [\text{Tr}_2(\rho_{12})]^2 \}. \quad (\text{A1})$$

In our cases, a biphoton pure state with a JSA $f(\omega_S, \omega_I)$ is described as

$$|\psi\rangle_{SI} \equiv \iint d\omega_S d\omega_I f(\omega_S, \omega_I) \hat{a}_{\omega_S}^\dagger \hat{b}_{\omega_I}^\dagger |0\rangle_S |0\rangle_I. \quad (\text{A2})$$

Since $f(\omega_S, \omega_I)$ can be Schmidt decomposed to $f(\omega_S, \omega_I) = \sum_n \xi_n \beta_{S,n}(\omega_S) \beta_{I,n}(\omega_I)$, where $\{\beta_{S,n}\}$ and $\{\beta_{I,n}\}$ are two sets of orthogonal functions,

$$\int d\omega \beta_{j,n}^*(\omega) \beta_{j,m}(\omega) = \delta_{n,m}, \quad (\text{A3})$$

where $j \in \{S, I\}$. We can then rewrite the biphoton pure state as

$$|\psi\rangle_{SI} = \frac{\sum_n \xi_n \hat{a}_{S,n}^\dagger \hat{b}_{I,n}^\dagger |0\rangle_S |0\rangle_I}{\sqrt{\sum_n |\xi_n|^2}}, \quad (\text{A4})$$

in which the operators obey

$$\begin{aligned} \hat{a}_n^\dagger &= \int d\omega_S \beta_{S,n}(\omega_S) \hat{a}_{\omega_S}^\dagger, \\ \hat{b}_n^\dagger &= \int d\omega_I \beta_{I,n}(\omega_I) \hat{b}_{\omega_I}^\dagger, \\ [\hat{a}_m, \hat{a}_n^\dagger] &= \delta_{m,n}, \\ [\hat{b}_m, \hat{b}_n^\dagger] &= \delta_{m,n}. \end{aligned} \quad (\text{A5})$$

Finally, the purity of this biphoton pure state is given by

$$\begin{aligned} \mathcal{P} &\equiv \text{Tr} \{ [\text{Tr}_I(|\psi\rangle\langle\psi|_{SI})]^2 \} \\ &= \text{Tr} \left[\left(\frac{\sum_n |\xi_n|^2 \hat{a}_{S,n}^\dagger |0\rangle\langle 0| \hat{a}_{S,n}}{\sum_n |\xi_n|^2} \right)^2 \right] \\ &= \sum_n |\xi_n|^4 / \left(\sum_n |\xi_n|^2 \right)^2, \end{aligned} \quad (\text{A6})$$

which agrees with the definition in Eq. (5).

2. Indistinguishability

Indistinguishability, on the other hand, is usually defined as the visibility of Hong-Ou-Mandel interference: $\mathcal{I} \equiv (R_{\max} - R_{\min})/R_{\max}$, where R_{\max} (R_{\min}) is the maximum (minimum) coincidence rate at different time delays between the signal and idler photons [24,42–47]. R_{\max} occurs when the two photons arrive at the beam splitter at completely different times, while R_{\min} arises when they arrive at the same time. Therefore, given photon-pair

flux R_{in} , $R_{\text{max}} = R_{\text{in}}/2$ because no quantum interference occurs and the two photons randomly take the two output ports of the beam splitter. Assuming ideal detectors and constant biphoton flux, the coincidence rate at zero time delay depends only on the JSA of the biphoton. The two detection events can be written as two projection measurements:

$$\begin{aligned} \langle M_1 | &= \langle 0 | \frac{1}{\sqrt{2}} \int d\omega (a_\omega + b_\omega), \\ \langle M_2 | &= \langle 0 | \frac{1}{\sqrt{2}} \int d\omega (a_\omega - b_\omega). \end{aligned} \quad (\text{A7})$$

The minimal coincidence rate is then derived as

$$R_{\text{min}} \equiv R_{\text{in}} \text{Tr}(|M_1\rangle\langle M_1| \otimes |M_2\rangle\langle M_2| \rho_{12}). \quad (\text{A8})$$

We estimate indistinguishability for two states. The first one is the postselected biphoton pure state [i.e., Eq. (A4)] from the SPDC process. The indistinguishability in this case equals the visibility of Hong-Ou-Mandel interference between the signal and idler photons from the same source. The minimum coincidence rate is derived as

$$\begin{aligned} R_{\text{min}} &\equiv R_{\text{in}} \text{Tr}(|M_1\rangle\langle M_1| \otimes |M_2\rangle\langle M_2| |\psi\rangle\langle\psi|_{SI}) \\ &= \frac{1}{4} R_{\text{in}} \iint d\omega d\omega' |f(\omega, \omega') - f(\omega', \omega)|^2 \\ &= \frac{1}{2} R_{\text{in}} \left\{ 1 - \text{Re} \left[\iint d\omega d\omega' f^*(\omega, \omega') f(\omega', \omega) \right] \right\}. \end{aligned} \quad (\text{A9})$$

The indistinguishability between such a pair of signal and idler photons is in turn

$$\begin{aligned} \mathcal{I} &\equiv (R_{\text{max}} - R_{\text{min}})/R_{\text{max}} \\ &= \text{Re} \left[\iint d\omega_s d\omega_i f^*(\omega_s, \omega_i) f(\omega_i, \omega_s) \right]. \end{aligned} \quad (\text{A10})$$

A more useful scenario (e.g., in a quantum network) is the interference between two photons from two independent sources (e.g., two heralded-single-photon sources or two quantum dots). In this case, the biphoton state is in the form of a product state; that is, $\rho_{12} = \rho_1 \otimes \rho_2$, where ρ_1 (ρ_2) is the density matrix of the first (second) photon. The minimum coincidence rate in the Hong-Ou-Mandel interference is given by

$$\begin{aligned} R_{\text{min}} &\equiv R_{\text{in}} \text{Tr}(|M_1\rangle\langle M_1| \otimes |M_2\rangle\langle M_2| \rho_1 \otimes \rho_2) \\ &= \frac{1}{2} R_{\text{in}} [1 - \text{Tr}(\rho_1 \rho_2)] \\ &= \frac{1}{4} R_{\text{in}} [2 + \|\rho_1 - \rho_2\|^2 - \text{Tr}(\rho_1^2) - \text{Tr}(\rho_2^2)], \end{aligned} \quad (\text{A11})$$

where $\|\rho\|^2 \equiv \text{Tr}(\rho^\dagger \rho)$ [27,46]. The indistinguishability turns out to be

$$\begin{aligned} \mathcal{I} &\equiv (R_{\text{max}} - R_{\text{min}})/R_{\text{max}} \\ &= \frac{1}{2} [\text{Tr}(\rho_1^2) + \text{Tr}(\rho_2^2) - \|\rho_1 - \rho_2\|^2]. \end{aligned} \quad (\text{A12})$$

Because of the broad spectrum of single photons generated from heralded SPDC sources, $\|\rho_1 - \rho_2\|^2$ is much smaller than each photon's purity $\text{Tr}(\rho_1^2)$ and $\text{Tr}(\rho_2^2)$. So the Hong-Ou-Mandel visibility of heralded single photons from SPDC sources is primarily limited by the purity of each photon. As such, we take purity as the principal performance metric to compare our customized poling source with others. In contrast, the indistinguishability of single photons generated from quantum dots is primarily constrained by $\|\rho_1 - \rho_2\|^2$ due to spectral wandering and spectral diffusion.

-
- [1] Juan Yin *et al.*, Satellite-based entanglement distribution over 1200 kilometers, *Science* **356**, 1140 (2017).
 - [2] Chin-Wen Chou, Julien Laurat, Hui Deng, Kyung Soo Choi, Hugues De Riedmatten, Daniel Felinto, and H. Jeff Kimble, Functional quantum nodes for entanglement distribution over scalable quantum networks, *Science* **316**, 1316 (2007).
 - [3] Sheng-Kai Liao *et al.*, Satellite-to-ground quantum key distribution, *Nature* **549**, 43 (2017).
 - [4] Alberto Boaron, Gianluca Boso, Davide Rusca, Cédric Vullicz, Claire Autebert, Misael Caloz, Matthieu Perrenoud, Gaëtan Gras, Félix Bussières, Ming-Jun Li, Daniel Nolan, Anthony Martin, and Hugo Zbinden, Secure Quantum Key Distribution Over 421 km of Optical Fiber, *Phys. Rev. Lett.* **121**, 190502 (2018).
 - [5] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone, Advances in quantum metrology, *Nat. Photonics* **5**, 222 (2011).
 - [6] Zheshen Zhang, Sara Mouradian, Franco N. C. Wong, and Jeffrey H. Shapiro, Entanglement-Enhanced Sensing in a Lossy and Noisy Environment, *Phys. Rev. Lett.* **114**, 110506 (2015).
 - [7] Stefano Pirandola, Bhaskar Roy Bardhan, Tobias Gehring, Christian Weedbrook, and Seth Lloyd, Advances in photonic quantum sensing, *Nat. Photonics* **12**, 724 (2018).
 - [8] Malte C. Tichy, Sampling of partially distinguishable bosons and the relation to the multidimensional permanent, *Phys. Rev. A* **91**, 022316 (2015).
 - [9] Alex Neville, Chris Sparrow, Raphaël Clifford, Eric Johnston, Patrick M. Birchall, Ashley Montanaro, and Anthony Laing, Classical boson sampling algorithms with superior performance to near-term experiments, *Nat. Phys.* **13**, 1153 (2017).
 - [10] 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).

- [11] Murphy Yuezhen Niu, Isaac L. Chuang, and Jeffrey H. Shapiro, Qudit-Basis Universal Quantum Computation Using $\chi^{(2)}$ Interactions, *Phys. Rev. Lett.* **120**, 160502 (2018).
- [12] Bryn A. Bell and Benjamin J. Eggleton, Multiphoton Interference in the Spectral Domain by Direct Heralding of Frequency Superposition States, *Phys. Rev. Lett.* **121**, 033601 (2018).
- [13] H. Jeff Kimble, The quantum internet, *Nature* **453**, 1023 (2008).
- [14] Stephanie Wehner, David Elkouss, and Ronald Hanson, Quantum internet: A vision for the road ahead, *Science* **362**, eaam9288 (2018).
- [15] L.-M. Duan, M. D. Lukin, J. Ignacio Cirac, and Peter Zoller, Long-distance quantum communication with atomic ensembles and linear optics, *Nature* **414**, 413 (2001).
- [16] Yuan Liang Lim, Almut Beige, and Leong Chuan Kwek, Repeat-Until-Success Linear Optics Distributed Quantum Computing, *Phys. Rev. Lett.* **95**, 030505 (2005).
- [17] Alessio Serafini, Stefano Mancini, and Sougato Bose, Distributed Quantum Computation Via Optical Fibers, *Phys. Rev. Lett.* **96**, 010503 (2006).
- [18] Pieter Kok, William J. Munro, Kae Nemoto, Timothy C. Ralph, Jonathan P. Dowling, and Gerard J. Milburn, Linear optical quantum computing with photonic qubits, *Rev. Mod. Phys.* **79**, 135 (2007).
- [19] Timothy J. Proctor, Paul A. Knott, and Jacob A. Dunningham, Multiparameter Estimation in Networked Quantum Sensors, *Phys. Rev. Lett.* **120**, 080501 (2018).
- [20] Wenchao Ge, Kurt Jacobs, Zachary Eldredge, Alexey V. Gorshkov, and Michael Foss-Feig, Distributed Quantum Metrology With Linear Networks and Separable Inputs, *Phys. Rev. Lett.* **121**, 043604 (2018).
- [21] Quntao Zhuang, Zheshen Zhang, and Jeffrey H. Shapiro, Distributed quantum sensing using continuous-variable multipartite entanglement, *Phys. Rev. A* **97**, 032329 (2018).
- [22] Yi Xia, Quntao Zhuang, William Clark, and Zheshen Zhang, Repeater-enhanced distributed quantum sensing based on continuous-variable multipartite entanglement, *Phys. Rev. A* **99**, 012328 (2019).
- [23] Changchen Chen, Cao Bo, Murphy Yuezhen Niu, Feihu Xu, Zheshen Zhang, Jeffrey H. Shapiro, and Franco N. C. Wong, Efficient generation and characterization of spectrally factorable biphotons, *Opt. Express* **25**, 7300 (2017).
- [24] Changchen Chen, Jane E. Heyes, Kyung-Han Hong, Murphy Yuezhen Niu, Adriana E. Lita, Thomas Gerrits, Sae Woo Nam, Jeffrey H. Shapiro, and Franco N. C. Wong, Indistinguishable single-mode photons from spectrally engineered biphotons, *Opt. Express* **27**, 11626 (2019).
- [25] K. Fradkin, A. Arie, A. Skliar, and G. Rosenman, Tunable midinfrared source by difference frequency generation in bulk periodically poled KTiOPO₄, *Appl. Phys. Lett.* **74**, 914 (1999).
- [26] Warren P. Grice, Alfred B. U'Ren, and Ian A. Walmsley, Eliminating frequency and space-time correlations in multiphoton states, *Phys. Rev. A* **64**, 063815 (2001).
- [27] Peter J. Mosley, Jeff S. Lundeen, Brian J. Smith, Piotr Wasylczyk, Alfred B. U'Ren, Christine Silberhorn, and Ian A. Walmsley, Heralded Generation of Ultrafast Single Photons in Pure Quantum States, *Phys. Rev. Lett.* **100**, 133601 (2008).
- [28] Agata M. Brańczyk, Alessandro Fedrizzi, Thomas M. Stace, Tim C. Ralph, and Andrew G. White, Engineered optical nonlinearity for quantum light sources, *Opt. Express* **19**, 55 (2011).
- [29] P. Ben Dixon, Jeffrey H. Shapiro, and Franco N. C. Wong, Spectral engineering by Gaussian phase-matching for quantum photonics, *Opt. Express* **21**, 5879 (2013).
- [30] Nicolás Quesada and Agata M. Brańczyk, Gaussian functions are optimal for waveguided nonlinear-quantum-optical processes, *Phys. Rev. A* **98**, 043813 (2018).
- [31] Annamaria Dosseva, Łukasz Cincio, and Agata M. Brańczyk, Shaping the joint spectrum of down-converted photons through optimized custom poling, *Phys. Rev. A* **93**, 013801 (2016).
- [32] J. L. Tambasco, A. Boes, L. G. Helt, M. J. Steel, and A. Mitchell, Domain engineering algorithm for practical and effective photon sources, *Opt. Express* **24**, 19616 (2016).
- [33] Francesco Graffitti, Dmytro Kundys, Derryck T. Reid, Agata M. Brańczyk, and Alessandro Fedrizzi, Pure down-conversion photons through sub-coherence-length domain engineering, *Quant. Sci. Technol.* **2**, 035001 (2017).
- [34] Francesco Graffitti, Peter Barrow, Massimiliano Proietti, Dmytro Kundys, and Alessandro Fedrizzi, Independent high-purity photons created in domain-engineered crystals, *Optica* **5**, 514 (2018).
- [35] Francesco Graffitti, Jérémy Kelly-Massicotte, Alessandro Fedrizzi, and Agata M. Brańczyk, Design considerations for high-purity heralded single-photon sources, *Phys. Rev. A* **98**, 053811 (2018).
- [36] Vahid Ansari, John M. Donohue, Benjamin Brecht, and Christine Silberhorn, Tailoring nonlinear processes for quantum optics with pulsed temporal-mode encodings, *Optica* **5**, 534 (2018).
- [37] Rui-Bo Jin, Ryosuke Shimizu, Kentaro Wakui, Hugo Benichi, and Masahide Sasaki, Widely tunable single photon source with high purity at telecom wavelength, *Opt. Express* **21**, 10659 (2013).
- [38] Rui-Bo Jin, Pei Zhao, Peigang Deng, and Qing-Lin Wu, Spectrally Pure States at Telecommunications Wavelengths from Periodically Poled MTiOXO₄ (M = K, Rb, Cs; X = P, As), *Phys. Rev. Appl.* **6**, 064017 (2016).
- [39] Fabian Laudenbach, Rui-Bo Jin, Chiara Greganti, Michael Hentschel, Philip Walther, and Hannes Hübel, Numerical Investigation of Photon-Pair Generation in Periodically Poled MTiOXO₄ (M = K, Rb, Cs; X = P, As), Phred and Telecom Wavelengths, *Phys. Rev. Appl.* **8**, 024035 (2017).
- [40] Rui-Bo Jin, Neng Cai, Ying Huang, Xiang-Ying Hao, Shun Wang, Fang Li, Hai-Zhi Song, Qiang Zhou, and Ryosuke Shimizu, Theoretical Investigation of a Spectrally Pure-State Generation from Isomorphs of KDP Crystal at Near-Infrared and Telecom Wavelengths, *Phys. Rev. Appl.* **11**, 034067 (2019).
- [41] Qi-Chao Sun, Yang-Fan Jiang, Ya-Li Mao, Li-Xing You, Wei Zhang, Wei-Jun Zhang, Xiao Jiang, Teng-Yun Chen, Hao Li, Yi-Dong Huang, Xian-Feng Chen, Zhen Wang, Jingyun Fan, Qiang Zhang, and Jian-Wei Pan, Entanglement swapping over 100 km optical fiber with independent entangled photon-pair sources, *Optica* **4**, 1214 (2017).

- [42] Chong-Ki Hong, Zhe-Yu Ou, and Leonard Mandel, Measurement of Subpicosecond Time Intervals Between Two Photons by Interference, *Phys. Rev. Lett.* **59**, 2044 (1987).
- [43] H. Fearn and R. Loudon, Theory of two-photon interference, *J. Opt. Soc. Am. B* **6**, 917 (1989).
- [44] Charles Santori, David Fattal, Jelena Vučković, Glenn S. Solomon, and Yoshihisa Yamamoto, Indistinguishable photons from a single-photon device, *Nature* **419**, 594 (2002).
- [45] Atature Kiraz, M. Atatüre, and A. Imamoğlu, Quantum-dot single-photon sources: Prospects for applications in linear optics quantum-information processing, *Phys. Rev. A* **69**, 032305 (2004).
- [46] C. I. Osorio, N. Sangouard, and Robert Thomas Thew, On the purity and indistinguishability of down-converted photons, *J. Phys. B* **46**, 055501 (2013).
- [47] Agata M. Brańczyk, Hong-Ou-Mandel interference, arXiv:1711.00080 (2017).
- [48] Sebastian Ruder, An overview of gradient descent optimization algorithms, arXiv:1609.04747 (2016).
- [49] Quoc V. Le, Jiquan Ngiam, Adam Coates, Abhik Lahiri, Bobby Prochnow, and Andrew Y. Ng, in *Proceedings of the 28th International Conference on International Conference on Machine Learning* (Omnipress, Bellevue, Washington, USA, 2011), p. 265–272. SEP
- [50] Diederik P. Kingma and Jimmy Ba, Adam: A method for stochastic optimization, arXiv:1412.6980 (2014).
- [51] Soham De, Anirbit Mukherjee, and Enayat Ullah, Convergence guarantees for RMSProp and ADAM in non-convex optimization and an empirical comparison to Nesterov acceleration, arXiv:1807.06766 (2018).
- [52] The source code is available on reasonable request.
- [53] Kiyoshi Kato and Eiko Takaoka, Sellmeier and thermo-optic dispersion formulas for KTP, *Appl. Opt.* **41**, 5040 (2002).
- [54] We calculate the indistinguishability defined in Ref. [45] in the 1550-nm band. After application of a 4-nm filter, the indistinguishability between the signal and the idler is 99.4%, decreasing to 98.7% (98.1%) with an 8-nm (40-nm) filter. The decrease in the indistinguishability is caused by the asymmetry between the signal's spectrum and the idler's spectrum at non-GVM wavelengths.
- [55] In addition, Fig. 4 in Ref. [10] shows that with assumptions the minimum indistinguishability required for quantum advantages in boson sampling asymptotically approaches around 98% by the dotted black line. Indistinguishability of photons from independent heralded SPDC sources is dominated by the purity of each photon. As such, we set 98% as a baseline to compare the performance of our design recipe with that of others.
- [56] Ryan S. Bennink, Optimal collinear gaussian beams for spontaneous parametric down-conversion, *Phys. Rev. A* **81**, 053805 (2010).
- [57] P. Ben Dixon, Danna Rosenberg, Veronika Stelmakh, Matthew E. Grein, Ryan S. Bennink, Eric A. Dauler, Andrew J. Kerman, Richard J. Molnar, and Franco N. C. Wong, Heralding efficiency and correlated-mode coupling of near-IR fiber-coupled photon pairs, *Phys. Rev. A* **90**, 043804 (2014).
- [58] Lutfi Arif Ngah, Olivier Alibert, Laurent Labonté, Virginia d'Auria, and Sébastien Tanzilli, Ultra-fast heralded single photon source based on telecom technology, *Laser Photon. Rev.* **9**, L1 (2015).
- [59] Nicola Montaut, Linda Sansoni, Evan Meyer-Scott, Raimund Ricken, Viktor Quiring, Harald Herrmann, and Christine Silberhorn, High-Efficiency Plug-and-Play Source of Heralded Single Photons, *Phys. Rev. Appl.* **8**, 024021 (2017).
- [60] Inferred from figures in related references.
- [61] Nicolas Sangouard, Christoph Simon, Hugues De Riedmatten, and Nicolas Gisin, Quantum repeaters based on atomic ensembles and linear optics, *Rev. Mod. Phys.* **83**, 33 (2011).
- [62] Matthäus Halder, Alexios Beveratos, Nicolas Gisin, Valerio Scarani, Christoph Simon, and Hugo Zbinden, Entangling independent photons by time measurement, *Nat. Phys.* **3**, 692 (2007).
- [63] Xiyuan Lu, Qing Li, Daron A. Westly, Gregory Moille, Anshuman Singh, Vikas Anant, and Kartik Srinivasan, Chip-integrated visible–telecom entangled photon pair source for quantum communication, *Nat. Phys.* **15**, 373 (2019).
- [64] Kurt Hornik, Maxwell Stinchcombe, and Halbert White, Multilayer feedforward networks are universal approximators, *Neural Netw.* **2**, 359 (1989).