APL Quantum ARTICLE pubs.aip.org/aip/apq

Automated Bell inequality violation searches for estimating entanglement quality in fiber

Cite as: APL Quantum 2, 016119 (2025); doi: 10.1063/5.0239623
Submitted: 20 September 2024 • Accepted: 2 February 2025 •
Published Online: 18 February 2025

Evan Dowling,[1,2,a)] **Gerald Baumgartner,**[3] **Richard Brewster,**[1,4] **Mark Morris,**[3] **Anne Marie Richards,**[3] **Rajarshi Roy,**[1,2,4] **and Thomas E. Murphy**[1,4]

AFFILIATIONS

[1] Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, Maryland 20740, USA
[2] Department of Physics, University of Maryland, College Park, Maryland 20740, USA
[3] Laboratory for Telecommunication Sciences, College Park, Maryland 20740, USA
[4] Department of Electrical and Computer Engineering, University of Maryland, University of Maryland, College Park, Maryland 20740, USA

[a)] **Author to whom correspondence should be addressed:** edowling@terpmail.umd.edu

ABSTRACT

In a future quantum network, two spatially separated individuals sharing a polarization entanglement source may need to assess the entanglement quality of the source without the presence of classical auxiliary signals. When the two are separated by a single-mode fiber, automated methods are necessary for the validation and estimation of polarization entanglement. We experimentally examine real-time iterative methods to search for maximal Bell violations, specifically the Clauser–Horne–Shimony–Holt (CHSH) inequality, between two observers sharing polarization entangled photons to evaluate the entanglement quality of the source. Our source allows us to tune the degree of entanglement of their shared photon pair by changing the temporal overlap of the two photons at an entangling 50/50 beam splitter so that we can compare the CHSH parameter these methods find to the theoretical values our source produces. The iterative methods used within our experiment are the Nelder–Mead optimization method, stochastic gradient descent, and Bayesian optimization. This is the first feedback experiment to study automated Bell violations in fiber and the first to compare all three of these iterative methods to one another in a quantum polarization control experiment. In our experiment, all methods are able to find Bell violations, but the Nelder–Mead method performed the best in terms of the speed and accuracy in finding the maximal violation.

I. INTRODUCTION

Quantum entanglement is a fundamental resource in quantum communication,[1,2] quantum key distribution (QKD),[3,4] quantum computing,[2,5,6] and quantum teleportation.[7,8] While entanglement can occur in many physical systems, photons are the only practical choice for transmitting entangled states over long distances, whether through free space[9–11] or over low-loss optical fibers.[12,13]

While photons can be entangled in many possible degrees of freedom, polarization is among the most commonly studied.[14–21] However, distributing polarization-entangled photons through optical fibers is challenging since most deployed fiber channels do not preserve the state of polarization. Even with a known input polarization, unpredictable temperature, bending, and stress variations within the fiber make the emerging state's polarization unpredictable and time varying.[22–26] In addition, widely used low loss piezoelectric actuator fiber squeezer polarization controllers (PCs) are difficult to calibrate and experience hysteresis effects that further complicate precise control of the orientation of coordinated measurement bases between the two individuals receiving the entangled photons.[27,28]

In addition to the uncontrolled polarization transformations from controllers and optical fibers, nonideal entanglement sources and experimental components introduce impairments such as multi-photon-pair production, polarization mode dispersion, and polarization-dependent loss (PDL), reducing the quality of the received entanglement.[29–33] A method to evaluate the quality of entanglement in a system in an automated way is therefore highly desirable. Such a characterization would guarantee, in practical real-world scenarios, that the quality of the distributed polarization entangled photon pairs is sufficient for the relevant quantum information protocols that will use them.

Many approaches for ensuring entanglement quality require knowledge of the density matrix acquired from quantum state tomography. For a bipartite entangled state, quantum state tomography requires the two individuals to make nine coordinated measurements from all combinations of three aligned measurement directions ($3^2 = 9$).[34] In the presence of unknown fiber transformations and with uncalibrated polarization controllers, this becomes a demanding task. One way to reduce the number of aligned and coordinated measurement bases is to use an entanglement witness as a proxy for full quantum state tomography.[1,35,36] A common entanglement witness for two qubits is the measurement of nonlocality through the violation of a Bell inequality.[1,37–39] This approach does not reconstruct the density matrix but creates an observable that distinguishes entangled vs separable states. The Bell inequality we use in this paper is the version first studied by Clauser–Horne–Shimony–Holt (CHSH),[40] where the measurement of a parameter S that exceeds 2 indicates the presence of nonlocality and, importantly, entanglement. Although entanglement could exist in mixed states when $S < 2$, the Bell parameter S is nonetheless a common, useful measure of entanglement quality. Efforts to self-test quantum sources for entanglement by violating CHSH inequalities by exceeding certain thresholds is well understood in the literature.[41–43] The value of S is determined by two individuals using four measurements consisting of all combinations of two measurement directions ($2^2 = 4$). The reduction from three to two measurement projection directions per observer reduces the experimental complexity from quantum state tomography and is also the commonly used experimental configuration in QKD experiments. This simplification is still true with iterative alignment techniques, such as those in this paper, which take many measurements of S because knowledge of the intended transmitted state and the transformation of the polarization controller are not assumed. In addition, our investigation does not attempt to close all locality loopholes to rigorously prove nonlocality, like Giustina *et al.*,[44] rather it aims to use the calculation of S as an entanglement measure when its value is above 2 and assumes fair sampling.

To find optimal measurement projection vectors in the presence of uncontrolled polarization transformations due to fiber, various experimental and computational methods have been explored, including multiplexing classical pilot tone calibration signals with the quantum signals[45–51] or using only quantum signal-based polarization control.[52–57] To specifically find maximal Bell violations using only feedback from the measurements of the entangled photons, computational studies using various optimization techniques and machine learning have been investigated,[58–64] but experimentally only a free space implementation with slow motorized waveplates and higher insertion loss at 800 nm has been studied.[65] Metroscale quantum networks will utilize near-IR photons transmitted over an optical fiber—a medium that introduces inherent polarization unpredictability that is absent in free-space. Moreover, the commercially available piezoelectric polarization actuators used here have ultra-low insertion loss and are electrically actuated, without the need for rotating waveplates. The experiment reported here thus represents a more practical and realistic setting for quantum networks.

We report here on an experiment to find the optimal orientation for the four needed projection vectors using a time and polarization multiplexed detection system modeled after the QKD experiment by Grünenfelder *et al.*[66] This setup has a few key advantages: it reduces the number of detectors needed by half, which are often the most expensive optical component, and it allows the value of S to be determined in one measurement integration period without needing to toggle between different polarization settings, which could otherwise introduce hysteresis errors in the polarization actuators. Our entanglement source also allows us to tune the amount of entanglement shared between two observers by varying the relative arrival time of two photons at an entangling 50/50 beam splitter. Using this setup, we examined several optimization algorithms' ability to maximize the CHSH inequality in the presence of different levels of entanglement. The methods we investigated were Nelder–Mead optimization,[67] stochastic gradient descent, and Bayesian optimization.[68]

This paper is organized as follows: we begin with a brief overview of the CHSH parameter given random fiber birefringence and polarization controller transformations. Following this is a description of our entanglement generation and our time-multiplexed measurement setups. Finally, we present experimental data on the three optimization algorithms when different levels of entanglement are shared by two parties, Alice and Bob.

II. THEORY ON CHSH PARAMETER ESTIMATION

The value of the CHSH parameter, S, is calculated using four sets of local measurements from two separated parties, Alice and Bob, on a two particle quantum state $\hat{\rho}$. To describe S, we assume that Alice performs polarization-resolved measurements along two different projection directions denoted by the unit vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in Stokes space. Bob likewise performs his polarization-projected measurements along directions $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$. For two qubits, the CHSH parameter[40] is

$$\begin{aligned} S(\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{b}_1, \boldsymbol{b}_2) = \operatorname{Tr}\left(\hat{\rho}\hat{A}_1 \otimes \hat{B}_1\right) + \operatorname{Tr}\left(\hat{\rho}\hat{A}_1 \otimes \hat{B}_2\right) \\ + \operatorname{Tr}\left(\hat{\rho}\hat{A}_2 \otimes \hat{B}_1\right) - \operatorname{Tr}\left(\hat{\rho}\hat{A}_2 \otimes \hat{B}_2\right), \end{aligned} \tag{1}$$

where, for example, $\hat{A}_i = \boldsymbol{a}_i \cdot \boldsymbol{\sigma}$ and $\boldsymbol{a}_i$ is the Stokes measurement projection direction for Alice's ith measurement and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ are Pauli matrices.[2] This is the same for Bob's jth measurement $\hat{B}_j = \boldsymbol{b}_j \cdot \boldsymbol{\sigma}$. When the shared state is the maximally entangled singlet state $\hat{\rho} = |\Psi^-\rangle\langle\Psi^-|$, where $|\Psi^-\rangle = 1/\sqrt{2}(|H_a V_b\rangle - |V_a H_b\rangle)$, Eq. (1) reduces to

$$S(\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{b}_1, \boldsymbol{b}_2) = -(\boldsymbol{a}_1 \cdot \boldsymbol{b}_1 + \boldsymbol{a}_1 \cdot \boldsymbol{b}_2 + \boldsymbol{a}_2 \cdot \boldsymbol{b}_1 - \boldsymbol{a}_2 \cdot \boldsymbol{b}_2). \tag{2}$$

Using the well-known Bell linear polarization projection orientations $\boldsymbol{a}_1 = (1, 0, 0)$, $\boldsymbol{a}_2 = (0, 1, 0)$, $\boldsymbol{b}_1 = 1/\sqrt{2}(1, 1, 0)$, and $\boldsymbol{b}_2 = 1/\sqrt{2}(1, -1, 0)$, we get the maximal Bell violation $|S| = 2\sqrt{2}$.

The CHSH parameter S can be experimentally measured using the coincidence counts from the four combinations of Alice's and Bob's two measurement bases,

$$\begin{aligned} S &= C(\boldsymbol{a}_1, \boldsymbol{b}_1) + C(\boldsymbol{a}_1, \boldsymbol{b}_2) + C(\boldsymbol{a}_2, \boldsymbol{b}_1) - C(\boldsymbol{a}_2, \boldsymbol{b}_2), \\ C(\boldsymbol{a}, \boldsymbol{b}) &= \frac{N_{++}(\boldsymbol{a}, \boldsymbol{b}) + N_{--}(\boldsymbol{a}, \boldsymbol{b}) - N_{+-}(\boldsymbol{a}, \boldsymbol{b}) - N_{-+}(\boldsymbol{a}, \boldsymbol{b})}{N_{++}(\boldsymbol{a}, \boldsymbol{b}) + N_{--}(\boldsymbol{a}, \boldsymbol{b}) + N_{+-}(\boldsymbol{a}, \boldsymbol{b}) + N_{-+}(\boldsymbol{a}, \boldsymbol{b})}, \end{aligned} \tag{3}$$

where $C(\boldsymbol{a}, \boldsymbol{b})$ is typically referred to as the correlation measure and $N_{+-}(\boldsymbol{a}, \boldsymbol{b})$, for instance, is the coincidence count between Alice's plus detector and Bob's minus detector for an integration time T, where the polarization measurement projections on $\hat{\rho}$ given the fiber and Alice's and Bob's polarization controllers' rotational transformations are $\boldsymbol{a}$ and $\boldsymbol{b}$. The correlation measure is the sum of the probability that like detectors fire together ($++$ and $--$) minus the probability that opposite detectors fire together ($+-$ and $-+$). For example, the probability that $+-$ detectors fire together is

$$P_{+-} = \frac{N_{+-}(\boldsymbol{a}, \boldsymbol{b})}{N_{++}(\boldsymbol{a}, \boldsymbol{b}) + N_{--}(\boldsymbol{a}, \boldsymbol{b}) + N_{+-}(\boldsymbol{a}, \boldsymbol{b}) + N_{-+}(\boldsymbol{a}, \boldsymbol{b})}. \tag{4}$$

In maximizing the CHSH parameter, the optimal orientation for the measurement projection directions is for $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ to lie within the plane spanned by $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$. In particular, $\boldsymbol{a}_1 = \frac{1}{\sqrt{2}}(\boldsymbol{b}_1 + \boldsymbol{b}_2)$ and $\boldsymbol{a}_2 = \frac{1}{\sqrt{2}}(\boldsymbol{b}_1 - \boldsymbol{b}_2)$.

Alice can adjust the relative orientation of her measurement bases by preceding her receiver by a polarization controller. Figure 1(a) shows the configuration of the polarization controllers employed here, which use piezoelectric transducers to produce a voltage-controlled retardance. Three independent stages, nominally oriented at 0°, 45°, and 0°, provide sufficient degrees of freedom to adjust and orient the projection direction $\boldsymbol{a}$ arbitrarily on the Poincaré sphere. The cumulative transformation of the controller can be described as a product of three rotation matrices,

$$\boldsymbol{R}_{\mathrm{PC}}(V_1, V_2, V_3) = \boldsymbol{R}_3(\boldsymbol{r}_3, \theta_3(V_3))\boldsymbol{R}_2(\boldsymbol{r}_2, \theta_2(V_2))\boldsymbol{R}_1(\boldsymbol{r}_1, \theta_1(V_1)), \tag{5}$$

where V_i is the voltage applied to the ith piezoelectric actuator and $\boldsymbol{R}(\boldsymbol{r}, \theta)$ is a 3×3 matrix describing a rotation by an angle θ about a unit Stokes vector $\boldsymbol{r}$,[69]

$$\boldsymbol{R}(\boldsymbol{r}, \theta) = (\boldsymbol{r}\boldsymbol{r}\cdot) + \sin\,\theta(\boldsymbol{r}\times) - \cos\,\theta(\boldsymbol{r} \times \boldsymbol{r}\times). \tag{6}$$

This analysis would apply for any polarization controller that uses fixed-axis, variable-retardance actuators, including liquid crystals or electro-optic controllers. In principle, the three input voltages could be computed based upon observations in order to reverse the polarization misalignment.[28,70] However, for the piezoelectric controllers used here, the dependence of θ_i on the applied voltage V_i and the orientation $\boldsymbol{r}_i$ are unknown *a priori* and can exhibit hysteresis under large steps. The optimization methods reported here therefore treat $\boldsymbol{R}_{\mathrm{PC}}$ as a black box with three voltage inputs that are adjusted iteratively.

As shown in Fig. 1(b), if the photon emerging from the controller is detected using a polarizing beam splitter and photon counters, the preceding polarization controller can arbitrarily steer the polarization projection $\boldsymbol{a}$. To resolve the photon along two independent directions, a non-polarizing 50/50 coupler can be used to randomly pass the photon to one of two polarizing beam splitters, which define $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ as shown in Fig. 1(c). In this configuration, the relative angle between $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ is preserved by the up-stream polarization controller. Figure 1(d) shows the arrangement used here, in which the four output channels are polarization multiplexed with a relative time delay ΔT.[66] This configuration allows two detectors to distinguish between four different outcomes by their relative arrival times.

III. EXPERIMENT

Figure 2 shows the experimental apparatus used to create and measure polarization-entangled photons. Polarization-entangled photons are generated from the arrival of two cross-polarized photons produced from a type-II spontaneous parametric downconversion (SPDC) process at an entangling 50/50 beam splitter. This process is described more extensively in prior work.[56]

This setup affords us specific control over the amount of entanglement generated by controlling the temporal overlap of the two cross-polarized (horizontal and vertical) photons at the entangling beam splitter from a controllable time delay stage (τ) in one arm of the entangling interferometer. When Alice and Bob post-select on events that are measured by both parties, the output state is cast into a superposition of a pure maximally entangled

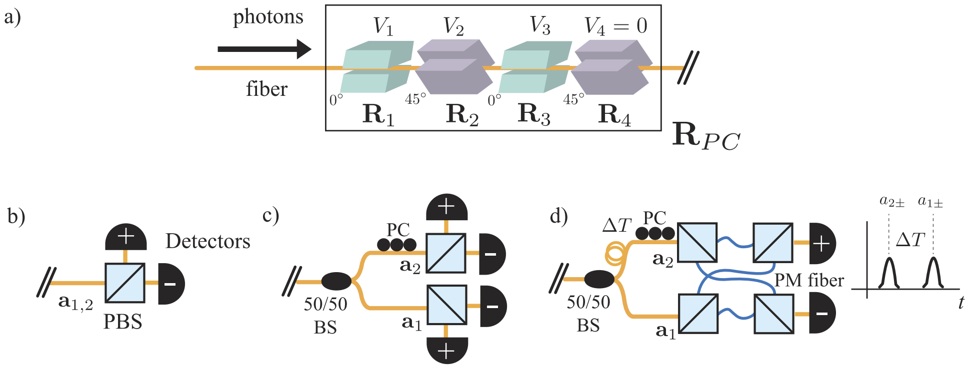

FIG. 1. (a) Schematic of our piezoelectric actuator polarization controller. Four in-series actuators expand due to an applied voltage, V_i, and compress the fiber along specific orientations around the circumference of the fiber. The photons get rotated by a total rotational transformation of $\boldsymbol{R}_{PC}$. The photons can be projected onto two projection bases, $\boldsymbol{a}_{1,2}$, from toggling known polarization transforms controlled by the polarization controller (b), from probabilistic splitting of a 50/50 beam splitter and four photon detectors (c), or from probabilistic splitting of a 50/50 beam splitter and two photon detectors from a polarization and time-multiplexed setup (d).

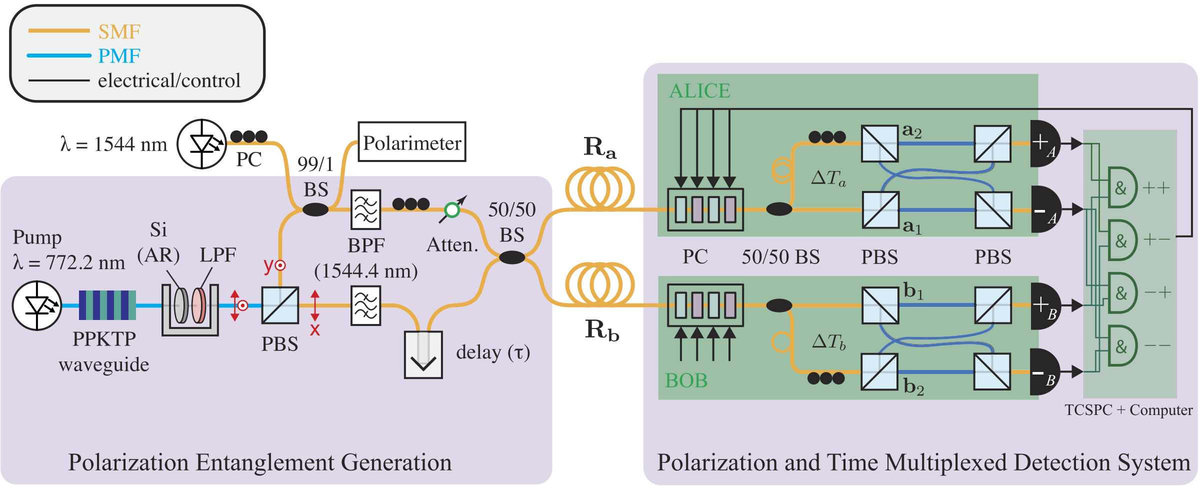

FIG. 2. Experimental diagram of our polarization entanglement generation setup and our time-multiplexed polarization detection system, both outlined in purple. Entanglement is controlled through the temporal overlap of each photon (τ), at a 50/50 beam splitter. SMF: single-mode fiber, PMF: polarization-maintaining fiber, AR: anti-reflection coating, LPF: longpass filter, PC: polarization controller, BPF: bandpass filter, $\boldsymbol{R}_a$ $(\boldsymbol{R}_b)$: rotational transformation of Alice's (Bob's) fiber span, PBS: polarization beam splitter, BS: beam splitter, and TCSPC: time-correlating single-photon counting module.

state, $|\Psi^-\rangle\langle\Psi^-|$, and a classically correlated mixed state, $\hat{\rho}_{\text{class}} = \frac{1}{2}(|HV\rangle\langle HV| + |VH\rangle\langle VH|)$,[15,71]

$$\hat{\rho} = C|\Psi^-\rangle\langle\Psi^-| + (1-C)\rho_{\text{class}} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -C & 0 \\ 0 & -C & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{7}$$

where C is the concurrence of $\hat{\rho}$ and defines the amount of entangled vs classical correlation between the two exiting photons. With central frequency ω_0 and bandwidth $\Delta\omega$ for each photon, the concurrence can be written as

$$C = e^{-\Delta\omega^2\tau^2}, \tag{8}$$

where τ is the controllable time delay between the photons incident on the entangling beam splitter. The bandwidth, $\Delta\omega$, is related to the full-width-half-maximum (FWHM) of the filter measured to be 1.4 nm,[56]

$$\omega = \frac{\pi c \Delta\lambda_{\text{FWHM}}}{\sqrt{2\ \ln 2}\lambda_0^2} = 0.46\ \text{rad/ps}. \tag{9}$$

Imperfections in the splitting ratios of the entangling beam splitter, the relative location and spectral width of the filters, white noise sources, and the polarization orientation of the classically mixed state can have an impact on the density matrix, CHSH parameter, and entanglement quality. The impact of these can be seen in Appendixes A and B and in Brewster *et al.*[30] When the polarization projection directions are optimally aligned, substituting Eqs. (7) and (8) into (1) gives

$$|S(\tau)| = \sqrt{2}\left(1 + e^{-\Delta\omega^2\tau^2}\right). \tag{10}$$

The maximal Bell violation of $2\sqrt{2}$ occurs for $\tau = 0$. This theoretical line, data from our experiment, and the linear fit of $|S(\tau)| = A + Be^{-\Delta\omega^2\tau^2}$ are shown in Fig. 3. The fit parameters, A and B, are discussed in relation to experimental imperfections in Appendixes A and B.

The alignment for the linear polarization basis, shown within Fig. 3(a), was done using a classical 10 mW CW 1544 nm pilot tone with a polarimeter and powermeters to measure all eight output ports of $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{b}_1, \boldsymbol{b}_2$. The alignment of the proper projection directions, say $\boldsymbol{a}_1$ to (1, 0, 0), was done by first adjusting the PC after the pilot tone until the polarization state recorded by the polarimeter was (1, 0, 0), and then, Alice's piezoelectric actuator PC was adjusted until all power was transmitted through the positive port of $\boldsymbol{a}_1$'s polarization beam splitter and near zero through the negative port. This was done for $\boldsymbol{a}_2$ to (0, 1, 0) through the same method; only a manual polarization controller before $\boldsymbol{a}_2$ was adjusted, leaving $\boldsymbol{a}_1$ unchanged. The same technique was used to align $\boldsymbol{b}_1$ to $1/\sqrt{2}(1,1,0)$ and $\boldsymbol{b}_2$ to $1/\sqrt{2}(1,-1,0)$. The alignment pilot tone and polarimeter were inserted into one of the entangling interferometer arms with a 99/01 beam splitter to minimize photon loss from the SPDC source, but to still permit classical power measurements above the powermeters' noise floor. After aligning to the known maximal projection directions, turning off the pilot tone, disconnecting the powermeters, and reconnecting the superconducting single photon counting nanowire detectors (SNSPDs), care was taken to attempt to align the two photon's polarization within the entanglement interferometer arms to the horizontal and vertical polarization orientations set by $\boldsymbol{a}_1$. In theory, near-perfect alignment to the horizontal and vertical polarization direction is possible, but was not achieved here due to uncontrolled polarization rotations from the need to disconnect fibers. More details are provided in Appendixes A and B.

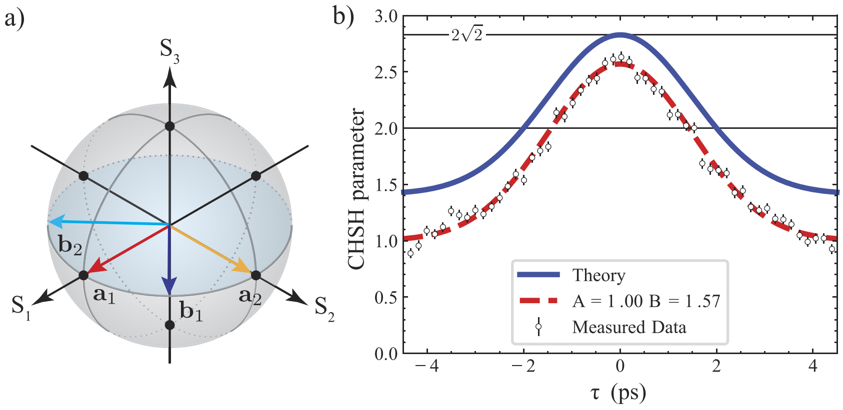

FIG. 3. (a) Measurement of the projectors $\boldsymbol{a}_{1,2}$ and $\boldsymbol{b}_{1,2}$. (b) CHSH parameter as a function of the temporal mismatch of the two photons at our entangling 50:50 beam splitter. A fit to the data $|S(\tau)| = A + Be^{-\Delta\omega^2\tau^2}$ is shown in red alongside the theoretical line in blue from Eq. (10). Error bars on the measured data are determined from Poisson shot noise.

The two spatially separated receivers (here denoted Alice and Bob) each include a fiber-based polarization controller with four piezoelectric actuators [General Photonics PolaRITE III], 50/50 beam splitters, polarization beam splitters, a pair of SNSPDs with quantum efficiencies up to 80% [Opus One, Quantum Opus LLC.], and a time-correlating single-photon counting (TCSPC) instrument [ID900, ID Quantique], where real time coincidence measurements can be processed and the results can be fed back to the upstream polarization controllers.

The polarization and time-multiplexed detection setup that both Alice and Bob have has been constructed so that the path difference between a photon heading toward $\boldsymbol{a}_1$ is 49 cm (2.35 ns) shorter

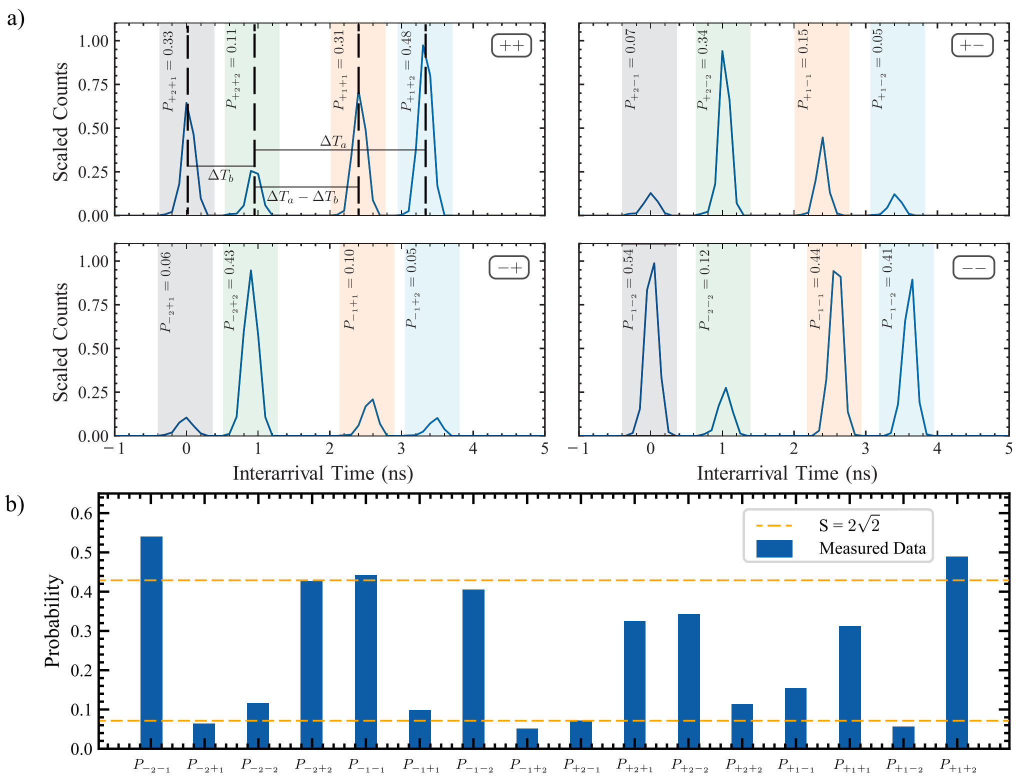

FIG. 4. (a) Example of a 16 peak signal from the four histograms across Alice's and Bob's two detectors $(\pm\pm)$ due to the time-multiplexed setup of Fig. 2. The detector pair used for each histogram is given in the top right of each plot. These measurements were made after a Nelder–Mead optimization routine maximized the CHSH parameter with a final value of $|S| = 2.56$. The total counts within each colored portion can be used to determine the coincidental probabilities from Eq. (4). Within the first plot, the temporal separation of the four signals is shown with respect to the temporal delays, ΔT_a and ΔT_b, from the different paths within the polarization and time-multiplexed setup. (b) For ideal $S = 2\sqrt{2}$ and state $\hat{\rho} = |\Psi^-\rangle\langle\Psi^-|$, the probability should be either $\frac{1}{4}(1 \pm \frac{1}{\sqrt{2}}) \approx 0.07$ or 0.43, shown as the orange dashed line with respect to measured probabilities.

than the photon heading toward $\boldsymbol{a}_2$. Similarly, the photon heading toward $\boldsymbol{b}_1$ experiences a 20 cm (0.96 ns) shorter path than a photon heading toward $\boldsymbol{b}_2$, leaving $\Delta T_a \approx 2\Delta T_b$. All other fiber paths within the detection setup were kept as close to equal as possible. The coincidence peaks in our experiment exhibit a statistical width of ~300 ps due to timing jitter in detection, and the path delays of roughly 1 and 2 ns were chosen to ensure that the correlation peaks are clearly distinguishable. In this configuration, the interarrival time histogram between one of Alice's detectors and one of Bob's detectors will exhibit four distinct correlation peaks, representing early–early, late–early, early–late, and late–late photons. This temporal correlation pattern can be measured even though the photons are continuous-wave (not pulsed) and spontaneously generated. This allows us to make the 16 different measurement combinations needed within Eq. (3) in one time integration period without toggling between polarization projection directions for Alice and Bob. An example of a given 16 signal coincidence measurement and the related joint detection probability, leading to a Bell violation ($S = 2.56$), is shown in Fig. 4.

One assumption of this method is that the photon-counters can detect both polarization states with comparable efficiency. Because the superconducting nanowire detectors employed here have a polarization-dependent efficiency, a manual polarization controller (not shown) was inserted prior to each detector in order to equalize the quantum efficiency between the early and late arriving photon channels. If the efficiencies are not equalized, the resulting polarization dependent loss (PDL) can impact the count rates and alter the calculated CHSH value (3). We note that newer nanowire constructions such as winding different nanowire geometries together in stacks, using spiral winding geometries, or stacking different materials with near-field optic effects can reduce this polarization-dependent efficiency.[72–75]

An example of Alice's and Bob's measurement bases before and after Alice tries to align to a random configuration of Bob's measurement bases from the maximization of S is shown in Fig. 5. The orientation of the projection bases is determined with the pilot tone laser and powermeters; like before, only Alice's and Bob's downstream PCs are left untouched and only the PC after the pilot tone is adjusted until it finds the orientation that maximizes the power through the positive port of that projection base. The insets of Fig. 5 show an initially arbitrary orientation of Alice's and Bob's projection bases, but after the maximization of S, Alice's and Bob's projection bases are nearly interleaved on the same plane, shown as red and blue dashed lines, as is expected.

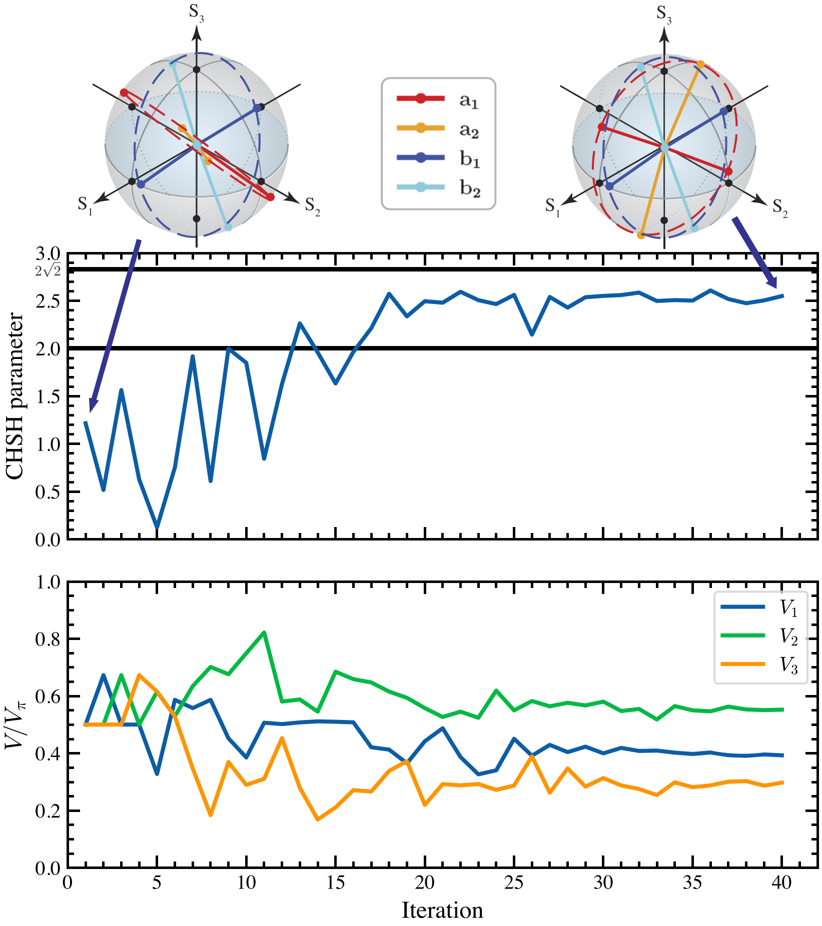

FIG. 5. Representative example of maximizing the CHSH parameter by changing the voltages of Alice's polarization controller. Here, the optimization algorithm was the Nelder–Mead method. The orientation of Alice's and Bob's projection direction is measured before and after maximizing the CHSH parameter using the alignment method mentioned in the body and shown above the CHSH plot.

IV. FINDING MAXIMAL CHSH VIOLATIONS

Being able to change the relative degree of entanglement between Alice and Bob allowed us to examine how various optimization algorithms compare to each other in a real-world setting. Nelder–Mead methods have been used in quantum experiments on entangled photons to orient measurement frames and determine CHSH values.[56,65,76] Stochastic gradient descent methods have been used for aligning measurement bases for BB84 QKD protocols.[26,54,58] Bayesian optimization algorithms have been used for control of filters and polarization controllers to maximize the indistinguishability of two photons.[55] This prior work motivated us to compare these algorithms in our experiment to maximize the CHSH parameter for photon pairs with varying entanglement qualities.

The Nelder–Mead method, also known as the downhill simplex method, creates a four-vertex tetrahedral simplex in the 3D parameter space defined by the voltages on the first three piezoelectric actuators within Alice's polarization controller. It iteratively modifies the simplex vertex voltages until they converge on a setting that maximizes the CHSH parameter.[67] Since the CHSH parameter is maximized through the relative location of Alice's and Bob's projector when they share the singlet state, only one person's polarization controller needs to be varied. This algorithm has the advantage that it does not require knowledge of the function's derivative, which would require additional measurements and whose measured accuracy can be low when the system has significant noise. The Nelder–Mead algorithm used in our experiment and simulations is implemented using Python's SciPy library.[77]

The stochastic gradient descent algorithm is an iterative method that moves through parameter space along randomized directions using gradient estimates along those directions. For our 3D parameter space having only one direction along which to evaluate the gradient saves four measurements compared to if the total derivative were to be estimated. The method works by first selecting a random direction in 3D parameter space, $\boldsymbol{V}_{\text{rand}}$, and evaluating the symmetric derivative,

$$\Delta S_n(\boldsymbol{V}_n) = \frac{S(\boldsymbol{V}_n + dV\boldsymbol{V}_{\text{rand}}) - S(\boldsymbol{V}_n - dV\boldsymbol{V}_{\text{rand}})}{2dV}, \tag{11}$$

and the next parameter point is then updated to

$$\boldsymbol{V}_{n+1} = \boldsymbol{V}_n + \lambda \cdot \Delta S_n(\boldsymbol{V}_n) \cdot \boldsymbol{V}_{\text{rand}}. \tag{12}$$

The hyperparameters λ and dV control how much to step in the direction of the gradient and how much to step in parameter space. For large values of λ, convergence can become a problem due to oscillations about the extrema, and for small values of λ, the algorithm converges slowly. Noise can impact the accuracy of the derivative estimate for too small of a value for dV, but too large of a value of dV can cause an inaccurate symmetric derivative. In addition, if the polarization controllers display hysteresis, all voltage jumps should be as small as possible. The optimum values were determined through simulations and found to be $\lambda = 0.8$ and $dV = V_\pi/12$. For details on how these were determined, see Appendixes A and B.

Bayesian optimization uses prior knowledge from measurements to build a probability distribution over possible functions to fit the measured data. Here, measurements are values of the CHSH parameter with the variables being the voltages on Alice's polarization controller. The probability distribution of possible functions is generated through the use of a Gaussian process. Because there is a probability distribution over possible fitting functions, it is possible to calculate a mean function and variances to indicate confidence regions. More details about how Bayesian optimization works is given in Appendixes A and B. Using the mean function and the variances, an acquisition function then picks the next point in parameter space, weighing areas of parameter space that are not sampled (exploration) and areas where the mean function is estimated to be maximal (exploitation) with some built in exploration vs exploitation trade-off setting established. When exploitation is preferred, the acquisition function picks points where the mean function is maximal with smaller regard to the variance caused from unsampled parameter space locations, but if exploration is preferred, the acquisition function's next point is skewed to unsampled parameter space. Both the Gaussian process covariance kernels and acquisition functions have different functional forms and hyperparameters that come with them. For a full mathematical description of Gaussian process Bayesian optimization and its application to optimizing photon indistinguishability, see the recent work of Cortes *et al.*[55] We implement Bayesian optimization through the use of the Python package bayes_opt[78] for its ease of implementation. Due to hysteresis in our polarization controllers, we chose hyperparameters that biased toward nearby parameter space exploitation rather than exploring non-sampled parameter space because they performed better. The acquisition function that performed the best in our simulations was the upper confidence bound with hyperparameter $\kappa = 0.4$ and Gaussian process kernel hyperparameter $\alpha = 10^{-8}$.

We used these methods to analyze experiments with various amounts of entanglement to determine how well they perform relative to each other. For ten different delay values, τ, we determined the maximum CHSH parameter each method attained relative to Eq. (10). At each τ value, we performed ten alignment procedures to maximize the CHSH parameter for each method: Nelder–Mead, Bayesian optimization, and stochastic gradient descent. For each given alignment, we randomly assign each piezoelectric actuator in Bob's polarization controller a value in $[0, V_\pi]$ to randomize the location of his projectors, $\boldsymbol{b}_{1,2}$, relative to Alice's projectors $\boldsymbol{a}_{1,2}$. Alice's initial polarization controller is set to the middle of her parameter space $\boldsymbol{V}_0 = [V_\pi/2, V_\pi/2, V_\pi/2]$ so that going outside the bounds ($V < 0$ and $V > V_\pi$) occurs rarely within the algorithm. When this does happen, the control code converts the voltage values used by the algorithm to be in the range $[0, V_\pi]$ by modulo V_π, which may introduce hysteresis effects if frequent jumps across this boundary occur. Note that reset-free actuation is possible but was not implemented here.[79]

Several results using different values of τ can be seen in Fig. 6(a), where each faint orange, blue, and green line represents one of the ten runs for the Nelder–Mead, Bayesian optimization, and stochastic gradient descent algorithms, respectively. For all three algorithms, the best vertex found up to that point is plotted since not every sampled parameter space point yields a better CHSH value, as can be seen from the Nelder–Mead example shown in Fig. 5. The bold lines are the averaged convergence behavior across the 10 runs on their historical staircasing ascent.

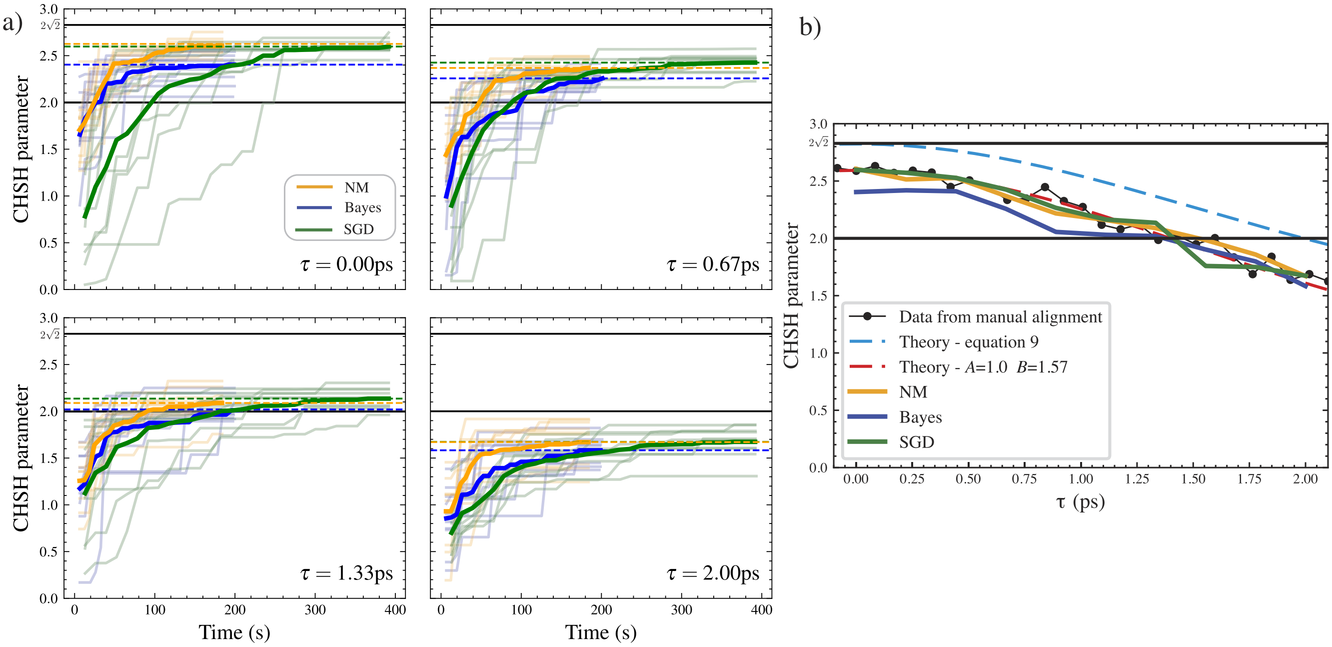

FIG. 6. (a) Experimental data comparing the Nelder–Mead (orange), Bayesian optimization (blue), and stochastic gradient descent (green) methods for maximizing the CHSH parameter. In the top left, $\tau = 0$, Alice and Bob share the approximately maximally entangled singlet state. For increasing values of τ, the entanglement shared by Alice's and Bob's photons decreases, thereby reducing their ability to violate Bell's inequality. Light lines represent the individual alignment runs, and bold lines are the statistical average of the 10 runs. Each measurement of S was based on an integration time of 5 s. The dashed lines show the final averaged S values the optimization algorithms converged to. (b) The average final convergence S value from each optimization method as a function of the relative delay within the entanglement interferometer.

The first observation one can make is that the Nelder–Mead and Bayesian optimization in general start with a higher CHSH value than the stochastic gradient descent. This is because they initially sample four points around, including $\boldsymbol{V}_0$, and each point may generate up to four CHSH values depending on where the minus sign is located within (1). This amounts to changing one of the four projectors to its orthogonal counterpart, as an example $\boldsymbol{a}_2 \rightarrow -\boldsymbol{a}_2$. After this initial sampling, both the Nelder–Mead and Bayesian optimization algorithms then maximize the CHSH parameter that was highest. When $\tau = 0$, both the Nelder–Mead and Bayesian optimization algorithms essentially found states where $S > 2$ (i.e., greater than the classical limit) in the first four measurements. This shows the effectiveness of random measurements that has been demonstrated for pairs of qubits, N qubits, and also higher dimensional quantum states.[80,81] This is in contrast to the stochastic gradient descent algorithm that must pick one of the four CHSH values to maximize at the beginning and therefore does not get the potential initial boost from reducing parameter space by picking one of the four CHSH parameters that is maximal near the initial starting condition.

Another observation is that the stochastic gradient descent takes about twice as long to reach a steady state value. This is largely because each of these algorithms takes roughly the same number of iterations to converge to a maximum (≈30 iterations), but the gradient calculation requires two evaluations of the CHSH parameter at each iteration. The stochastic gradient descent and Nelder–Mead algorithms both perform nearly identically when looking at the final values they converge to. The Bayesian optimization appears to perform the worst in finding the true maximum. This is largely attributed to the hysteresis effects of the polarization controllers and the large jumps the Bayesian optimization algorithm performs to sample unknown regions of parameter space. This happens even when the acquisition function is still weighted for exploitation rather than exploration. After returning from sampling a distant region of parameter space, the polarization settings that were giving near-optimal CHSH parameter values are now slightly altered, but all past priors do not take this into account. Since gradient based and also simplex based algorithms, to some extent, do not have this changing parameter space memory problem, they are not as affected by hysteresis.

The final CHSH value the averaged methods converged to is plotted vs the temporal mismatch τ in Fig. 6(b). As was stated earlier, the Nelder–Mead and stochastic gradient descent perform nearly identically and reproduce the results from the optimal configuration shown in Fig. 3 with Bayesian optimization performing slightly worse. All the optimization methods can detect a Bell violation up for $\tau < 1.3$ ps. The density matrix with this τ has a concurrence of ≈0.42, which means that the photons are still entangled, but the entanglement can no longer be detected by Bell's inequalities.

V. CONCLUSION

Rather than performing measurement intensive quantum state tomography to estimate the entanglement quality, proving the violation of a Bell's inequality is a much easier experimental task. An evaluation of the CHSH parameter can give a quantitative measure of the entangled systems' concurrence, or entanglement quality. This still requires the careful alignment of two parties' measurement bases, which is difficult in a fiber optic environment. In this paper, we have experimentally demonstrated the ability to find the optimal configuration of Alice's polarization controller, when Alice and Bob share entangled photon pairs, which maximizes the violation of Bell's inequality in the presence of uncontrolled birefringent fiber spans. We investigated several iterative optimization methods that have been used within the quantum communication literature—namely, Nelder–Mead optimization, stochastic gradient descent, and Bayesian optimization—for their ability to maximize the CHSH parameter. These methods were tested using various levels of entanglement shared between the photons sent to Alice and Bob. All three methods were able to violate Bell's inequality for a temporal mismatch in our entangling interferometer of $\tau < 1.3$ ps, corresponding to an entanglement concurrence value of 0.42. The Nelder–Mead and Bayesian optimization found the optimum condition the quickest, while the Nelder–Mead and stochastic gradient descent were the most accurate in finding the optimum value.

Our experiment used a polarization and time-multiplexed detection system that reduces the experimental demands on photon detection instruments. Using half the number of detectors allows off-the-shelf coincidence counting electronic boards to process the 16 signals in real time so that feedback measurements can be sent upstream for polarization control. In addition, this time-multiplexed detection system is more robust to polarization controller hysteresis problems, where jumping between two nonorthogonal polarization projection bases introduces measurement errors. This setup and evaluation methods may prove useful for the entanglement quantification of distributed entangled particles in future quantum networks where no alignment pilot tones are used and only feedback signals from the entangled particles are leveraged.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Evan Dowling: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Gerald Baumgartner**: Funding acquisition (supporting); Methodology (supporting); Validation (equal); Writing – review & editing (equal). **Richard Brewster**: Methodology (equal); Validation (equal); Writing – review & editing (supporting). **Mark Morris**: Data curation (supporting); Resources (equal). **Anne Marie Richards**: Funding acquisition (supporting); Resources (equal); Writing – review & editing (supporting). **Rajarshi Roy**: Funding acquisition (equal); Visualization (supporting); Writing – review & editing (supporting). **Thomas E. Murphy**: Conceptualization (equal); Formal analysis (equal); Methodology (equal); Supervision (equal); Visualization (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: IMPAIRMENTS ON CHSH PARAMETER

Experimental imperfections can impact the density matrix, Eq. (7), and also the CHSH parameter. Appendix A 1 deals with spectral imperfections in the entanglement interferometer, and Appendix A 2 deals with deviations from general photon indistinguishability and polarization misalignment between the classically mixed state and the projection bases of Alice and Bob.

1. Filter impairments

Instead of having identical filters in the entanglement interferometer, let us assume that the two filters can differ in both spectral width and location. The emitted photon pair coming out of the PPKTP waveguide is best described by state

$$|\psi\rangle = \int d\omega_s d\omega_i f_0(\omega_s, \omega_i)\hat{a}_x^\dagger(\omega_s)\hat{a}_y^\dagger(\omega_i)|0\rangle, \tag{A1}$$

where x (y) is the horizontal (vertical) polarization and f_0 is the joint spectral distribution for the signal and idler photons $\omega_{s,i}$. For continuous wave pump sources with pump frequency ω_p, pumping a second-order nonlinear waveguide with length L, the joint spectrum is seen to produce a sinc profile,

$$f_0(\omega_s, \omega_i) \propto \operatorname{sinc}\left(\frac{\Delta k(\omega_s, \omega_i)L}{2}\right)\sqrt{\delta(\omega_s + \omega_i - \omega_p)}, \tag{A2}$$

where $\Delta k = k_p(\omega_p) - k_s(\omega_s) + k_i(\omega_i) - \frac{2\pi}{\Lambda}$ is the quasi-phase matching bandwidth of the PPKTP waveguide. The delta function, $\delta(\omega_s + \omega_i - \omega_p)$, is there for energy conservation in the downconversion process. The polarization beam splitter sends the horizontal photon to the x path and the vertical photon to the y path. Both photons undergo filtering in their paths due to tunable bandpass filters. Each filter can be assumed to be of Gaussian form with a center frequency $\Omega_{x,y}$ and bandwidth $\Delta\omega_{x,y}$,

$$f_{\text{filter}}(\omega) \propto \exp\frac{-(\omega - \Omega)^2}{2\Delta\omega^2}. \tag{A3}$$

If the filters are narrower than the central sinc width, then the joint spectrum can be approximated as

$$f_0(\omega_s, \omega_i) \propto \exp\frac{-(\omega_s - \Omega_x)^2}{2\Delta\omega_x^2}\exp\frac{-(\omega_i - \Omega_y)^2}{2\Delta\omega_y^2} \times \sqrt{\delta(\omega_s + \omega_i - \omega_p)}. \tag{A4}$$

In Ref. 82, the CHSH parameter is determined through the formula

$$S(\tau) = \sqrt{2}\left(1 + \frac{K(\tau)}{I}\right) \tag{A5}$$

for integral quantities

$$I = \frac{1}{2}\int d\omega_s d\omega_i |f_0(\omega_s,\omega_i)|^2, \tag{A6}$$

$$K(\tau) = \frac{1}{2}\int d\omega_s d\omega_i e^{i(\omega_s-\omega_i)\tau} f_0^*(\omega_i,\omega_s) f_0(\omega_s,\omega_i). \tag{A7}$$

Assuming small central frequency and frequency width errors $\Omega_x = \Omega_y + \varepsilon = \Omega$ and $\Delta\omega_x = \Delta\omega_y + \delta = \Delta\omega$ gives a value for $S(\tau)$ as shown as follows:

$$S(\tau) = \sqrt{2}\left(1 + \exp\left[-\frac{\left(2\Delta\omega^2\varepsilon + (\delta^2 + 2\Delta\omega\delta)(\omega_p - 2\Omega)\right)^2}{4\Delta\omega^2(\Delta\omega+\delta)^2[\Delta\omega^2 + (\Delta\omega+\delta)^2]}\right] \times \exp\left[-\frac{\Delta\omega^2(\Delta\omega+\delta)^2\tau^2}{\Delta\omega^2 + (\Delta\omega+\delta)^2}\right]\right). \tag{A8}$$

The first exponential term represents the reduction factor from the ideal case of $S(\tau = 0) = 2\sqrt{2}$. For $\Omega = \omega_p/2$, it is interesting to note that any $\varepsilon \neq 0$ reduces S, but increasing the relative width of one filter with respect to the other enhances S.

2. Photon indistinguishability and the orientation of the classically mixed state

In general, even when $\tau = 0$, the temporal overlap of the two cross-polarized photons at our entangling 50/50 beam splitter will not produce the $|\Psi^-\rangle$ state 100% of the time. This could be due to the filter imperfections, shown above, transmission and reflection coefficients that differ from the ideal 50/50 ratio,[30] and other frequency and temporal mode imperfections. In fact, this is identical to the reason why, when polarization aligned photons meet at a 50/50 beam splitter, the extinction of coincidences from the exit ports of the beam splitter never perfectly reaches zero.[82] This experiment, often named the HOM experiment, gives a measure of the two photons indistinguishably.[83] The visibility, V_{HOM}, is the ratio,

$$V_{\mathrm{HOM}} = \frac{N(\tau\to\infty) - N(\tau=0)}{N(\tau\to\infty)}, \tag{A9}$$

where τ is the relative temporal delay between the two photon wave packet at the beam splitter and N is the coincidence rate at the exit ports. Accounting for this, we can modify our density matrix for experiment, Eq. (7), by letting $C = V_{\mathrm{HOM}}e^{-\Delta\omega^2\tau^2}$.

In addition to deviations from having identical photons, having Alice's and Bob's projection measurements be at orientations different from the polarization of the classically mixed state will impact the value of S. Assuming Alice and Bob measure at the normal linear polarization orientations, $\boldsymbol{a}_1 = (1,0,0)$, $\boldsymbol{a}_2 = (0,1,0)$, $\boldsymbol{b}_1 = 1/\sqrt{2}(1,1,0)$, and $\boldsymbol{b}_2 = 1/\sqrt{2}(1,-1,0)$, but that the classically mixed polarization state is created on the basis

$$\hat{\rho}_{\mathrm{class}} = \frac{1}{2}(|nn_\perp\rangle\langle nn_\perp| + |n_\perp n\rangle\langle n_\perp n|), \tag{A10}$$

where $|n\rangle = \cos(\theta/2)|H\rangle + e^{i\phi}\sin\theta/2|V\rangle$ and $|n_\perp\rangle = -e^{-i\phi}\sin(\theta/2)|H\rangle + \cos\theta/2|V\rangle$. Written out in detail, the classically mixed density matrix is

$$\hat{\rho}_{\mathrm{class}} = \frac{1}{2}\begin{pmatrix} \frac{1}{2}s^2 & -cse^{-i\phi} & -cse^{-i\phi} & -\frac{1}{2}s^2e^{-i2\phi} \\ -cse^{i\phi} & 1-\frac{1}{2}s^2 & -\frac{1}{2}s^2 & cse^{-i\phi} \\ -cse^{i\phi} & -\frac{1}{2}s^2 & 1-\frac{1}{2}s^2 & cse^{-i\phi} \\ -\frac{1}{2}s^2e^{i2\phi} & cse^{i\phi} & cse^{i\phi} & \frac{1}{2}s^2 \end{pmatrix}, \tag{A11}$$

where $s = \sin\theta$ and $c = \cos\theta$. The density matrix of our experiment is then written as

$$\hat{\rho} = V_{\mathrm{HOM}}e^{-\Delta\omega^2\tau^2}|\Psi^-\rangle\langle\Psi^-| + \left(1 - V_{\mathrm{HOM}}e^{-\Delta\omega^2\tau^2}\right)\hat{\rho}_{\mathrm{class}}. \tag{A12}$$

Plugging Eq. (A12) into (1) gives a CHSH parameter,

$$\begin{aligned} |S(\tau)| &= \sqrt{2}\left[1 + V_{\mathrm{HOM}}e^{-\Delta\omega^2\tau^2} - (\sin\,\theta\,\sin\,\phi)^2\left(1 - V_{\mathrm{HOM}}e^{-\Delta\omega^2\tau^2}\right)\right] \\ &= \sqrt{2}\left[1 + V_{\mathrm{HOM}}e^{-\Delta\omega^2\tau^2} - |\boldsymbol{n}\cdot\hat{s}_3|^2\left(1 - V_{\mathrm{HOM}}e^{-\Delta\omega^2\tau^2}\right)\right] \\ &= \sqrt{2}\left[1 + V_{\mathrm{HOM}}e^{-\Delta\omega^2\tau^2} - (\cos\,\varphi)^2\left(1 - V_{\mathrm{HOM}}e^{-\Delta\omega^2\tau^2}\right)\right], \end{aligned} \tag{A13}$$

where $\boldsymbol{n}$ is the Stokes vector from the Jones projection vector $|n\rangle$, $\hat{s}_3$ is the Stokes unit vector in the right circular polarization direction, and φ is the angle between $\boldsymbol{n}$ and $\hat{s}_3$. With $\varphi = 0$, we recover the CHSH parameter expression given within Brewster *et al.*[82] The maximum CHSH parameter value attainable at $\tau = 0$ is between a maximum value of $|S| = \sqrt{2}(1 + V_{\mathrm{HOM}})$ and a minimum value $|S| = \sqrt{2}(2V_{\mathrm{HOM}})$ depending upon φ. The HOM visibility for our filtered SPDC source was measured with a maximum value of 93%.[56] The theoretical best CHSH parameter given this HOM visibility is between $S = 2.63$ and $S = 2.73$ dependent on the orientation of the classically mixed state.

Brewster *et al.* have shown that any additional experimental white noise added to the source by the transformation $\hat{\rho} \to \eta\hat{\rho} + \frac{1-\eta}{4}\hat{I}$ affects the CHSH parameter by a scaling term $|S(\tau)| \to \eta|S(\tau)|$.[30] Measuring $|S(\tau)|$ along the linear polarization directions is shown in Fig. 3, along with a fit to $|S(\tau)| = \sqrt{2}A + Be^{-\Delta\omega^2\tau^2}$. Using Eq. (A13), the fit parameters are determined to be $A = \eta\left(1 - \cos^2\varphi\right) = 1.00$ and $B = \sqrt{2}\eta V_{\mathrm{HOM}}\left(1 + \cos^2\varphi\right) = 1.57$. Assuming the best visibility $V_{\mathrm{HOM}} = 0.93$, the white noise term is found to be equal to $\eta = 0.95$ and the misalignment term $\cos^2\varphi = 0.25$. This equates to a misalignment from the linear polarization equator of 30°.

APPENDIX B: HYPERPARAMETERS FOR BAYESIAN OPTIMIZATION AND STOCHASTIC GRADIENT DESCENT

To simulate our experiment, we use Eq. (1), where $\boldsymbol{a}_i$ and $\boldsymbol{b}_j$ for $i,j \in \{1,2\}$ are initially randomly distributed to simulate the fiber's birefringence with the condition $\boldsymbol{a}_1\cdot\boldsymbol{a}_2 = 0$ and likewise $\boldsymbol{b}_1\cdot\boldsymbol{b}_2 = 0$. Alice's polarization controller then rotates her projectors to new states,

$$\boldsymbol{a}_i \to \boldsymbol{R}_{PC}(\theta_1,\theta_2,\theta_3)\boldsymbol{a}_i = \boldsymbol{R}(\hat{s}_1,\theta_3)\boldsymbol{R}(\hat{s}_2,\theta_2)\boldsymbol{R}(\hat{s}_1,\theta_1)\boldsymbol{a}_i \tag{B1}$$

with $\boldsymbol{R}(\boldsymbol{r},\theta) = (\boldsymbol{r}\boldsymbol{r}\cdot) + \sin\theta(\boldsymbol{r}\times) - \cos\theta(\boldsymbol{r}\times\boldsymbol{r}\times)$, where $\boldsymbol{r}$ is the rotation vector around which the polarization projection directions rotate by angle θ. The exact values for θ_i are determined by the algorithm and the prior values for S with the goal of iteratively converging to the maximal condition of $S \to 2\sqrt{2}$.

1. Bayesian optimization

For an introduction to Bayesian optimization and application of Bayesian optimization to the task of engineering photon indistinguishability, see the textbook *Bayesian Optimization* by Garnett[68] and the paper of Cortes *et al.*[55] Bayesian optimization attempts to find the maximum of an objective function $\phi(\boldsymbol{x})$, where the variable $\boldsymbol{x} \in \mathbb{R}^D$ and D is the dimension size of the input variables. Here, our objective function is the CHSH parameter and the variable is $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \in \mathbb{R}^3$. In pursuit of this goal, the Bayesian optimization incrementally builds a probabilistic model of our objective function, which reflects the current knowledge of it from prior measurements. The probabilistic model is generated via a Gaussian process (GP). A GP generates a probabilistic distribution of functions over the domain,

$$p(f,\mathscr{D}) = GP(f;\mu,K), \tag{B2}$$

where μ is a mean function, K is a positive semidefinite covariance function (or kernel), and $\mathscr{D}$ is the total sampled data defined as $\mathscr{D} = (y, X)$. The quantities $y = \{y_1, y_2, \ldots, y_n\}$ are samples of the objective function using input parameters $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$. The mean function $\mu(\boldsymbol{x})$ is the average value over all functions in the distribution for all points $\boldsymbol{x}$ within the parameter space. The covariance function determines how deviations from the mean are structured, encoding expected properties of the function's behavior. Intuitively, one can think of a GP as analogous to a function, but instead of returning a scalar $f(\boldsymbol{x})$ for an arbitrary $\boldsymbol{x}$, it returns the mean and variance of a normal distribution over the possible values of f at $\boldsymbol{x}$. Importantly, the mean function and knowledge of the variance can be used within the acquisition function, defined below, to determine where the best point in parameter space to sample next.

Common kernels are the squared exponential covariance kernel,

$$K(\boldsymbol{x},\boldsymbol{x}') = \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|^2}{2\alpha^2}\right) \tag{B3}$$

with α being a hyperparameter length scale, which ensures that nearby points have similar function values, and the Matern kernel,

$$K(\boldsymbol{x},\boldsymbol{x}') = \frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}}{\alpha}\|\boldsymbol{x}-\boldsymbol{x}'\|^2\right)^{\nu} \times \mathscr{K}_\nu\left(\frac{\sqrt{2\nu}}{\alpha}\|\boldsymbol{x}-\boldsymbol{x}'\|^2\right) \tag{B4}$$

with Γ being the Gamma function, $\mathscr{K}_\nu$ being the modified Bessel function, and ν being a hyperparameter that controls the smoothness of the function. The GP is updated using Bayesian inference via Bayes' theorem where new measurements of the objective function change the probabilistic model to better fit the data. Bayes' theorem, at a simple level, can be stated as the posterior probability of a model M given data $\mathscr{D}$ are proportional to the likelihood of $\mathscr{D}$ given M multiplied by the prior probability of M,

$$p(M|\mathscr{D}) \propto p(\mathscr{D}|M)p(M). \tag{B5}$$

To sample efficiently, Bayesian optimization uses an acquisition function on the GP (after a new measurement) and the acquisition function generates the best point to sample the objective function next. Like GP kernels, many acquisition functions exist, such as the probability of improvement, expected improvement, and upper confidence bound (UCB). An optimization of the acquisition function over all possible points within the parameter space determines the next sample point. As an example, the UCB acquisition function is

$$a_{\mathrm{UCB}} = \mu(\boldsymbol{x}) + \kappa\sigma(\boldsymbol{x}). \tag{B6}$$

When the hyperparameter κ is large, the acquisition function will be maximal in locations where the uncertainty, σ, is large giving exploratory behaviors to the optimization process. When κ is small, the acquisition function will be maximal in locations closer to the maximal of the mean function.

An extensive examination of different Bayesian kernels and acquisition functions, as done in Cortes work, was not done here and remains a project for future investigation. Instead, we used Cortes' findings that the Matern kernel is the optimal kernel with $\nu = 5/2$ and the upper confidence bound acquisition function being the optimal acquisition function. Since Cortes was also concerned with photon polarization, along with temporal alignment and spectral filtering for photon indistinguishability, this assumption seems to be a good starting point. We numerically simulated various size hyperparameters, α within the Matern kernel and κ values within the upper confidence bound acquisition function, to determine what parameters would be best to implement within our experiment. We used the bayes_opt Python package, which builds upon the sklearn library. Since we are concerned with the speed of convergence to the maximal CHSH parameter value, we average over 100 simulated alignments to determine how many measurements of S are needed before the Bayesian optimization routine finds a condition where the sampled S is within 1% of $2\sqrt{2}$. The plot of this is shown in Fig. 7(a). In general, a smaller length scale with $\alpha = 10^{-8}$ and $\kappa = 0.4$, shown as a dashed line, performed best and was determined as the best points to implement within our experiment.

In addition, we want to reduce the distance between consecutive points chosen by the acquisition function due to hysteresis within our polarization controllers. The distance between parameter points $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_{i+1}$ is the Euclidean distance $\|\theta_{i+1} - \theta_i\|$. After a given alignment run with N samples, we can get an overall average distance between all consecutive samples,

$$D(\boldsymbol{\theta}_{i-1},\boldsymbol{\theta}_i) = \frac{1}{N-1}\sum_{i}^{N-1}\|\boldsymbol{\theta}_{i+1}-\boldsymbol{\theta}_i\|. \tag{B7}$$

A plot of this distance averaged over the 100 simulated alignments for each value of κ and α is shown in Fig. 7(b). The $\alpha = 10^{-8}$ and $\kappa = 0.4$ parameters chosen also perform well in minimizing this distance.

2. Stochastic gradient descent

For the stochastic gradient descent, the size of the steps in the direction of the gradient, λ, and the parameter space step size to determine the gradient, dV, are hyperparameters. To find the optimal value of λ within Eq. (12), sometimes referred to as the learning parameter, we ran 200 simulated alignments maximizing Eq. (1)

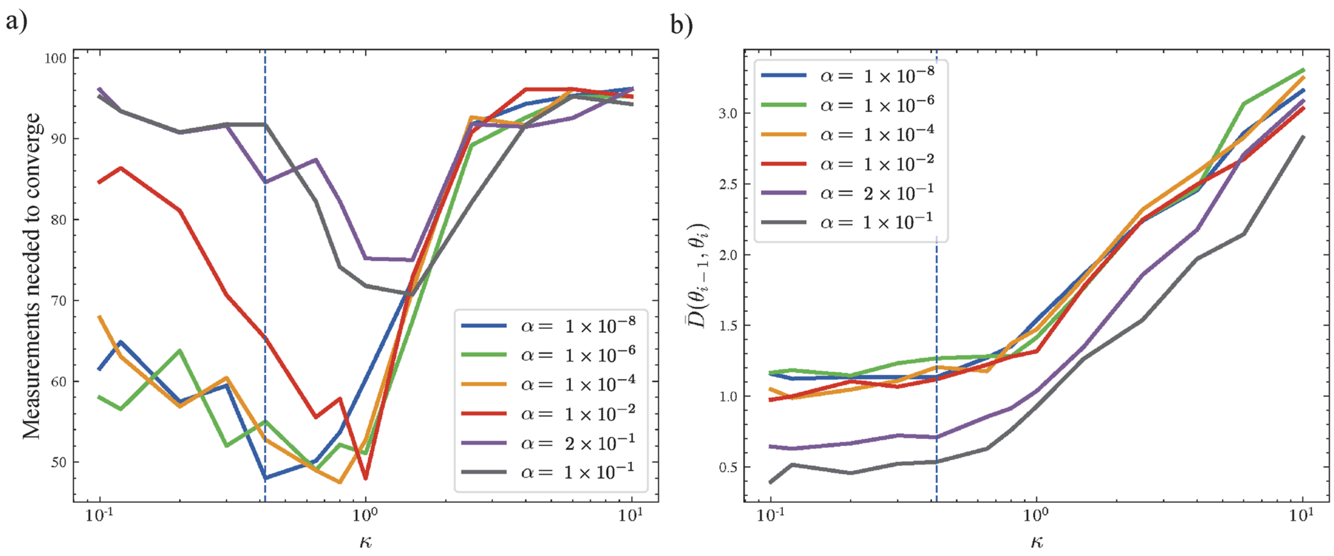

FIG. 7. Simulated experiments performing Bayesian optimization with various hyperparameters to maximize the CHSH parameter. (a) Number of measurements needed for the Bayes optimization to converge S to 99% of $2\sqrt{2}$ vs the hyperparameter κ used within the upper confidence bound acquisition function. Plotted are various values of the hyperparameter α used within the Matern covariance kernel of the Gaussian process. (b) Average distance over the course of CHSH parameter maximization in $\boldsymbol{\theta}$ parameter space between successive points picked by the acquisition function.

for 30 values of λ ranging from 0.3 to 1.1. Because we are interested in choosing a setting that maximizes the value of S, we set the stopping condition at 99% of the optimal value $2\sqrt{2}$. The number of measurements needed for Alice's polarization controller to reach this threshold average over the 200 runs is shown in Fig. 8(a). The dashed line corresponds to $\lambda = 0.8$, which performed the best at finding the true maximum with the fewest number of measurements at 50. With gradient estimates taking two measurements,

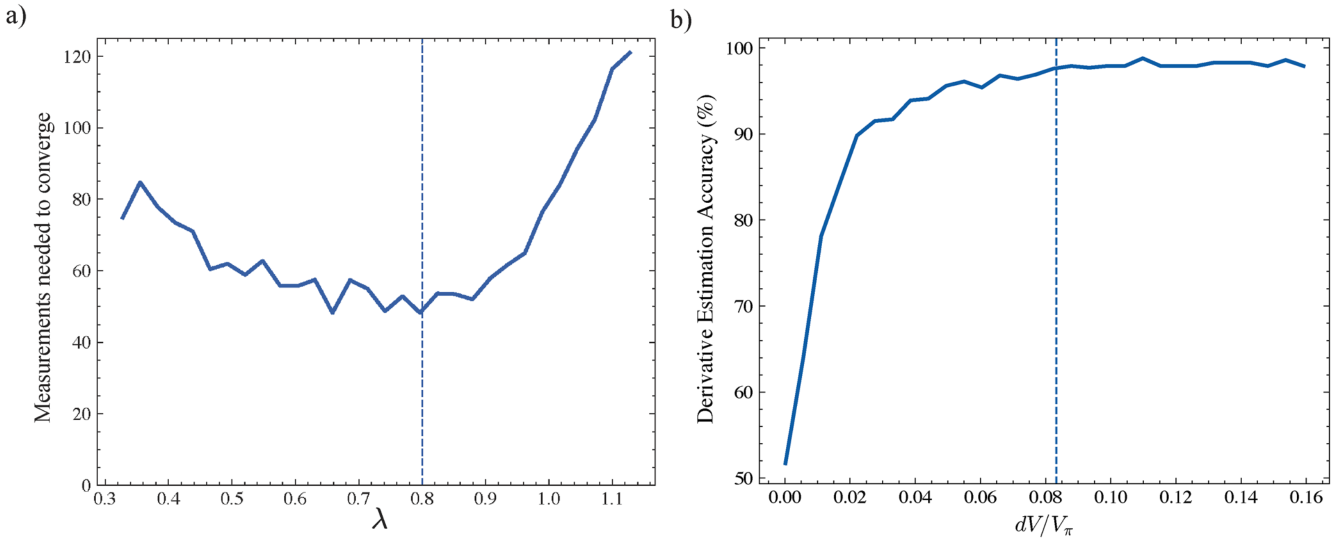

FIG. 8. (a) Number of measurements needed for the stochastic gradient descent method to converge S to 99% of $2\sqrt{2}$ vs the hyperparameter λ, which determines how much to step in the direction of the gradient. (b) In the presence of background counts and Poissonian statistics, the estimation of the derivative of S from Eq. (11) needs a large enough step in parameter space to overcome the noise from accidental coincidence counts, but cannot be so large as to introduce errors and bring hysteresis effects into play. With $r_s = 400$ pairs per second, $r_b = 80$ pairs per second, and an integration time $T = 5$ s, we get an optimal $dV = V_\pi/12$.

this amounts to roughly 25 iterations of the stochastic gradient descent method, which is what was observed experimentally in Fig. 8(a).

To determine the parameter step needed for an accurate estimation of the symmetric derivative, Eq. (11), we model the coincidence count rate between any two projectors, i for Alice and j for Bob, based on Poisson statistics with the signal rate, r_s, the probability of detection $\mathrm{Tr}\,(\hat{\rho}\hat{A}_i \otimes \hat{B}_j)$, a constant background coincidence count rate, r_b, and integration time T.[56] Taken together, the number of coincidences, N, measured in an integration interval, T, can be modeled by a discrete Poisson distribution with a mean that depends on Alice's rotational angles $\theta_1, \theta_2, \theta_3$ from her polarization controller,

$$\begin{aligned} \mathrm{N}(\boldsymbol{a}_i, \boldsymbol{b}_j)_{N=k} &= \frac{(r_{i,j}(\theta_1,\theta_2,\theta_3)T)^k}{k!} e^{-r_{i,j}(\theta_1,\theta_2,\theta_3)T}, \\ r_{i,j}(\theta_1,\theta_2,\theta_3) &= r_s\, \mathrm{Tr}\,(\hat{\rho}\hat{A}_i(\theta_1,\theta_2,\theta_3) \otimes \hat{B}_j) + r_b, \\ \hat{A}_i(\theta_1,\theta_2,\theta_3) &= \boldsymbol{R}_{PC}(\theta_1,\theta_2,\theta_3)\boldsymbol{a}_i \cdot \boldsymbol{\sigma}. \end{aligned} \tag{B8}$$

In our experiment, we are concerned with taking a large enough step so that the noise from the background counts does not impair the accuracy of the symmetric derivative (11). In the time-multiplexed setup, we observe a maximal count with coaligned measurement bases of $r_s = 400$ pairs/sec and a background accidental coincidence count of $r_b = 80$ pairs/sec. We use an integration time of $T = 5$ s. The simulated coincidence probability of a photon from Alice's ith projector and Bob's jth projector is then

$$P_{i,j} = \frac{\mathrm{N}(\boldsymbol{a}_i, \boldsymbol{b}_j)}{\mathrm{N}(\boldsymbol{a}_i, \boldsymbol{b}_j) + \mathrm{N}(\boldsymbol{a}_i, -\boldsymbol{b}_j) + \mathrm{N}(-\boldsymbol{a}_i, \boldsymbol{b}_j) + \mathrm{N}(-\boldsymbol{a}_i, -\boldsymbol{b}_j)}. \tag{B9}$$

The accuracy of the symmetric derivative is determined from the percent overlap of the derivative with and without background accidentals ($r_b = 0$) shown in Fig. 8(b). With the observed rates, we determined $dV = V_\pi/12$, shown as the dashed line, performed optimally while being small enough that hysteresis effects from the voltage jumps were not as impactful.

REFERENCES

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).

[2] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*, 10th Anniversary Edition (Cambridge University Press, 2012).

[3] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).

[4] C. H. Bennett, G. Brassard, and N. D. Mermin, Phys. Rev. Lett. **68**, 557 (1992).

[5] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, Phys. Rev. A **52**, 3457 (1995).

[6] D. Deutsch, Proc. R. Soc. London, Ser. A **425**, 73 (1989).

[7] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).

[8] R.-B. Jin, M. Takeoka, U. Takagi, R. Shimizu, and M. Sasaki, Sci. Rep. **5**, 9333 (2015).

[9] M. Aspelmeyer, H. R. Böhm, T. Gyatso, T. Jennewein, R. Kaltenbaek, M. Lindenthal, G. Molina-Terriza, A. Poppe, K. Resch, M. Taraba, R. Ursin, P. Walther, and A. Zeilinger, Science **301**, 621 (2003).

[10] R. Ursin, F. Tiefenbacher, T. Schmitt-Manderbach, H. Weier, T. Scheidl, M. Lindenthal, B. Blauensteiner, T. Jennewein, J. Perdigues, P. Trojek, B. Ömer, M. Fürst, M. Meyenburg, J. Rarity, Z. Sodnik, C. Barbieri, H. Weinfurter, and A. Zeilinger, Nat. Phys. **3**, 481 (2007).

[11] J. Yin, Y. Cao, Y.-H. Li, S.-K. Liao, L. Zhang, J.-G. Ren, W.-Q. Cai, W.-Y. Liu, B. Li, H. Dai, G.-B. Li, Q.-M. Lu, Y.-H. Gong, Y. Xu, S.-L. Li, F.-Z. Li, Y.-Y. Yin, Z.-Q. Jiang, M. Li, J.-J. Jia, G. Ren, D. He, Y.-L. Zhou, X.-X. Zhang, N. Wang, X. Chang, Z.-C. Zhu, N.-L. Liu, Y.-A. Chen, C.-Y. Lu, R. Shu, C.-Z. Peng, J.-Y. Wang, and J.-W. Pan, Science **356**, 1140 (2017).

[12] J. F. Dynes, H. Takesue, Z. L. Yuan, A. W. Sharpe, K. Harada, T. Honjo, H. Kamada, O. Tadanaga, Y. Nishida, M. Asobe, and A. J. Shields, Opt. Express **17**, 11440 (2009).

[13] T. Inagaki, N. Matsuda, O. Tadanaga, M. Asobe, and H. Takesue, Opt. Express **21**, 23241 (2013).

[14] P. G. Kwiat, K. Mattle, H. Weinfurter, A. Zeilinger, A. V. Sergienko, and Y. Shih, Phys. Rev. Lett. **75**, 4337 (1995).

[15] S. Bose and D. Home, Phys. Rev. Lett. **88**, 050401 (2002).

[16] C. E. Kuklewicz, M. Fiorentino, G. Messin, F. N. C. Wong, and J. H. Shapiro, Phys. Rev. A **69**, 013807 (2004).

[17] A. Aspect, P. Grangier, and G. Roger, Phys. Rev. Lett. **49**, 91 (1982).

[18] O. Kuzucu and F. N. C. Wong, Phys. Rev. A **77**, 032314 (2008).

[19] F. Kaiser, A. Issautier, L. A. Ngah, O. Alibart, A. Martin, and S. Tanzilli, Laser Phys. Lett. **10**, 045202 (2013).

[20] E. Ji, Q. Liu, X. Fu, P. Du, M. Nie, F. Zhang, and M. Gong, IEEE J. Quantum Electron. **53**, 9300206 (2017).

[21] P. J. Thomas, C. J. Chunnilall, D. J. M. Stothard, D. A. Walsh, and M. H. Dunn, Opt. Express **18**, 26600 (2010).

[22] T. Imai and T. Matsumoto, J. Lightwave Technol. **6**, 1366 (1988).

[23] Y. Namihira and H. Wakabayashi, J. Lightwave Technol. **7**, 1201 (1989).

[24] M. Karlsson, J. Brentel, and P. A. Andrekson, J. Lightwave Technol. **18**, 941 (2000).

[25] D. S. Waddy, L. Chen, and X. Bao, Opt. Fiber Technol. **11**, 1 (2005).

[26] Y.-Y. Ding, H. Chen, S. Wang, D.-Y. He, Z.-Q. Yin, W. Chen, Z. Zhou, G.-C. Guo, and Z.-F. Han, Opt. Express **25**, 27923 (2017).

[27] H. Shimizu, S. Yamazaki, T. Ono, and K. Emura, J. Lightwave Technol. **9**, 1217 (1991).

[28] M. Jofre, G. Anzolin, F. Steinlechner, N. Oliverio, J. P. Torres, V. Pruneri, and M. W. Mitchell, Opt. Express **20**, 12247 (2012).

[29] R. A. Brewster, G. Baumgartner, and Y. K. Chembo, Phys. Rev. A **104**, 022411 (2021).

[30] R. A. Brewster, G. Baumgartner, and Y. K. Chembo, Phys. Rev. A **107**, 022225 (2023).

[31] B. T. Kirby, D. E. Jones, and M. Brodsky, J. Lightwave Technol. **37**, 95 (2019).

[32] Y. Liu, Opt. Quantum Electron. **55**, 720 (2023).

[33] L. Ruan, B. T. Kirby, M. Brodsky, and M. Z. Win, Phys. Rev. A **103**, 032425 (2021).

[34] J. B. Altepeter, E. R. Jeffrey, and P. G. Kwiat, Adv. At., Mol., Opt. Phys. **52**, 105 (2005).

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

[36] M. Barbieri, F. De Martini, G. Di Nepi, P. Mataloni, G. M. D'Ariano, and C. Macchiavello, Phys. Rev. Lett. **91**, 227901 (2003).

[37] R. Horodecki, P. Horodecki, and M. Horodecki, Phys. Lett. A **200**, 340 (1995).

[38] G. Chen, W. Zhang, P. Yin, C. Li, and G. Guo, Fundam. Res. **1**, 27 (2021).

[39] W. J. Munro, K. Nemoto, and A. G. White, J. Mod. Opt. **48**, 1239 (2001).

[40] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).

[41] D. Mayers and A. Yao, Quantum Inf. Comput. **4**, 273 (2004).

[42] M. McKague, T. H. Yang, and V. Scarani, J. Phys. A: Math. Theor. **45**, 455304 (2012).

[43] I. Šupić and J. Bowles, Quantum **4**, 337 (2020).

[44] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-A. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger, Phys. Rev. Lett. **115**, 250401 (2015).

[45] F. Heismann, J. Lightwave Technol. **12**, 690 (1994).

[46] M. Martinelli and R. A. Chipman, J. Lightwave Technol. **21**, 2089 (2003).

[47] M. Martinelli, P. Martelli, and S. M. Pietralunga, J. Lightwave Technol. **24**, 4172 (2006).

[48] G. B. Xavier, G. V. de Faria, G. P. Temporão, and J. P. von der Weid, Opt. Express **16**, 1867 (2008).

[49] G. B. Xavier, N. Walenta, G. V. de Faria, G. P. Temporão, N. Gisin, H. Zbinden, and J. P. von der Weid, New J. Phys. **11**, 045015 (2009).

[50] J. Chen, G. Wu, L. Xu, X. Gu, E. Wu, and H. Zeng, New J. Phys. **11**, 065004 (2009).

[51] D. Li, S. Gao, G. Li, L. Xue, L. Wang, C. Lu, Y. Xiang, Z. Zhao, L. Yan, Z. Chen, G. Yu, and J. Liu, Opt. Express **26**, 22793 (2018).

[52] Y. Ding, W. Chen, H. Chen, C. Wang, Y. Li, S. Wang, Z. Yin, G. Guo, and Z. Han, Opt. Lett. **42**, 1023 (2017).

[53] C. Agnesi, M. Avesani, L. Calderaro, A. Stanco, G. Foletto, M. Zahidy, A. Scriminich, F. Vedovato, G. Vallone, and P. Villoresi, Optica **7**, 284 (2020).

[54] Y. Shi, H. S. Poh, A. Ling, and C. Kurtsiefer, Opt. Express **29**, 37075 (2021).

[55] C. L. Cortes, P. Lefebvre, N. Lauk, M. J. Davis, N. Sinclair, S. K. Gray, and D. Oblak, Phys. Rev. Appl. **17**, 034067 (2022).

[56] E. Dowling, M. Morris, G. Baumgartner, R. Roy, and T. E. Murphy, Opt. Express **31**, 2316 (2023).

[57] M. Peranić *et al.*, EPJ Quantum Technol. **10**, 30 (2023).

[58] L. Yang, G. Chen, W. Zhang, X. Peng, S. Yu, X. Ye, C. Li, and G. Guo, Phys. Rev. A **96**, 052310 (2017).

[59] J. Cortés-Vega, J. F. Barra, L. Pereira, and A. Delgado, Quantum Inf. Process. **22**, 203 (2023).

[60] Y. C. Ma and M. H. Yung, npj Quantum Inf. **4**, 34 (2018).

[61] D. L. Deng, Phys. Rev. Lett. **120**, 240402 (2018).

[62] A. Canabarro, S. Brito, and R. Chaves, Phys. Rev. Lett. **122**, 200401 (2019).

[63] K. Bharti, T. Haug, V. Vedral, and L. Kwek, "How to teach AI to play bell non-local games: Reinforcement learning," arXiv:1912.10783 [quant-ph] (2019). eprint

[64] K. Bharti, T. Haug, V. Vedral, and L.-C. Kwek, AVS Quantum Sci. **2**, 034101 (2020).

[65] D. Poderini, E. Polino, G. Rodari, A. Suprano, R. Chaves, and F. Sciarrino, Phys. Rev. Res. **4**, 013159 (2022).

[66] F. Grünenfelder, A. Boaron, D. Rusca, A. Martin, and H. Zbinden, Appl. Phys. Lett. **112**, 051108 (2018).

[67] J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).

[68] R. Garnett, *Bayesian Optimization* (Cambridge University Press, 2023).

[69] J. N. Damask, *Polarization Optics in Telecommunications* (Springer Science & Business Media, 2004), Vol. 101.

[70] M. F. Ramos, N. A. Silva, N. J. Muga, and A. N. Pinto, Opt. Express **28**, 5035 (2020).

[71] M. C. Tichy, J. Phys. B: At., Mol. Opt. Phys. **47**, 103001 (2014).

[72] F. Zheng, R. Xu, G. Zhu, B. Jin, L. Kang, W. Xu, J. Chen, and P. Wu, Sci. Rep. **6**, 22710 (2016).

[73] X. Chi, K. Zou, C. Gu, J. Zichi, Y. Cheng, N. Hu, X. Lan, S. Chen, Z. Lin, V. Zwiller, and X. Hu, Opt. Lett. **43**, 5017 (2018).

[74] S. N. Dorenbos, E. M. Reiger, N. Akopian, U. Perinetti, V. Zwiller, T. Zijlstra, and T. M. Klapwijk, Appl. Phys. Lett. **93**, 161102 (2008).

[75] V. B. Verma, F. Marsili, S. Harrington, A. E. Lita, R. P. Mirin, and S. W. Nam, Appl. Phys. Lett. **101**, 251114 (2012).

[76] E. Dowling, M. Morris, G. Baumgartner, R. Roy, and T. E. Murphy, in *CLEO 2023* (Optica Publishing Group, 2023), p. FM1A.2.

[77] P. Virtanen *et al.*, Nat. Methods **17**, 261 (2020).

[78] F. Nogueira (2014). "Bayesian optimization: Open source constrained global optimization tool for Python," GitHub. https://github.com/fmfn/BayesianOptimization/

[79] R. Noé, H. Heidrich, and D. Hoffmann, J. Lightwave Technol. **6**, 1199 (1988).

[80] P. Shadbolt, T. Vértesi, Y. Liang, C. Branciard, N. Brunner, and J. L. O'Brien, Sci. Rep. **2**, 470 (2012).

[81] M. C. Tran, B. Dakić, W. Laskowski, and T. Paterek, Phys. Rev. A **94**, 042302 (2016).

[82] R. A. Brewster, J. Goldhar, M. Morris, G. Baumgartner, and Y. K. Chembo, IEEE Trans. Quantum Eng. **3**, 4100310 (2022).

[83] C. K. Hong and L. Mandel, Phys. Rev. A **31**, 2409 (1985).