

# Quantum-enhanced interferometry with large heralded photon-number states

G.S. Thekkadath,<sup>1</sup> M.E. Mycroft,<sup>2</sup> B.A. Bell,<sup>1</sup> C.G. Wade,<sup>1</sup> A. Eckstein,<sup>1</sup> D.S. Phillips,<sup>1</sup> R.B. Patel,<sup>1</sup> A. Buraczewski,<sup>2</sup> A.E. Lita,<sup>3</sup> T. Gerrits,<sup>3,4</sup> S.W. Nam,<sup>3</sup> M. Stobińska,<sup>2</sup> A.I. Lvovsky,<sup>1</sup> and I.A. Walmsley<sup>1,5</sup>

<sup>1</sup>*Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK\**

<sup>2</sup>*Faculty of Physics, University of Warsaw, ul. Pasteura 5, 02-093 Warsaw, Poland*

<sup>3</sup>*National Institute of Standards and Technology, 325 Broadway, Boulder, Colorado 80305, USA*

<sup>4</sup>*National Institute of Standards and Technology,*

*100 Bureau Drive, Gaithersburg, Maryland 20899, USA*

<sup>5</sup>*Department of Physics, Imperial College London, Prince Consort Rd, London SW7 2AZ, UK*

Quantum phenomena such as entanglement can improve fundamental limits on the sensitivity of a measurement probe. In optical interferometry, a probe consisting of  $N$  entangled photons provides up to a  $\sqrt{N}$  enhancement in phase sensitivity compared to a classical probe of the same energy. Here, we employ high-gain parametric down-conversion sources and photon-number-resolving detectors to perform interferometry with heralded quantum probes of sizes up to  $N = 8$  (i.e. measuring up to 16-photon coincidences). Our probes are created by injecting heralded photon-number states into an interferometer, and in principle provide quantum-enhanced phase sensitivity even in the presence of significant optical loss. Our work paves the way towards quantum-enhanced interferometry using large entangled photonic states.

## INTRODUCTION

Optical interferometry provides a means to sense very small changes in the path of a light beam. These changes may be induced by a wide range of phenomena, from pressure and temperature variations that impact refractive index, to modifications of the space-time metric that characterize gravitational waves. In its simplest form, interferometry measures distortions via the phase difference  $\phi$  between the two paths of the interferometer. The uncertainty  $\Delta\phi$  in a measurement of this phase difference is limited fundamentally by the quantum noise of the illuminating light beams. This noise can be reduced by employing light exhibiting nonclassical properties such as entanglement and squeezing in order to improve the sensitivity of an interferometer beyond classical limits [1]. Quantum states of light are most effective when it is desirable to maximize the phase sensitivity per photon inside an interferometer, such as in gravitational wave detectors [2, 3] or when characterizing delicate photosensitive samples [4–8].

In principle,  $N$ -photon quantum states of light such as the highly entangled N00N state can provide up to a  $\sqrt{N}$  precision enhancement over a classical state of equal energy [9–17]. Unfortunately these highly entangled states are vulnerable to decoherence, especially at large photon numbers. In practice, their enhanced sensitivity disappears in the presence of loss which may originate from interactions inside the interferometer (e.g. absorption in a sample) as well as external losses in the state preparation and detection [18].

Although a  $\sqrt{N}$  enhancement is not achievable in the presence of loss, one can engineer states that trade-away sensitivity for loss-tolerance in order to still achieve some advantage over classical limits [19, 20]. For example, squeezed light [21–24] and non-maximally entangled

states such as Holland-Burnett states [25–31] can surpass classical limits despite some losses. Importantly, the precision enhancement achievable with such states can grow with  $N$ , even in the presence of loss [19]. Experimental demonstrations have prepared unheralded  $N = 6$  [29, 30] (or heralded  $N = 2$  [28]) Holland-Burnett states, but further increase of  $N$  is constrained by source brightness as well as detector efficiency and number-resolution. This motivates developing experimental protocols that can produce and detect loss-tolerant states with larger photon numbers.

In this work, we address a number of key challenges in order to scale-up quantum-enhanced interferometry using definite photon-number states of light. Firstly, we introduce probe states that are prepared by combining two photon-number states on a beam splitter similarly to Holland-Burnett states. However, unlike the latter, we allow the initial photon-number states to be unequal. We show that these generalized Holland-Burnett states are more sensitive than both Holland-Burnett and N00N states in the presence of loss and approximate the performance of the optimal probe [19]. Secondly, we experimentally implement our scheme using high-gain parametric down-conversion sources [32, 33] and state-of-the-art photon-number-resolving detectors [34] in order to access a large photon-number regime. We herald entangled probes of sizes up to  $N = 8$  and measure up to 16-photon coincidences, thereby further increasing the scale of experimental multiphoton quantum technologies [35–37].

The idea is illustrated in Fig. 1(a). Two type-II parametric down-conversion (PDC) sources each produce pairs of light beams that are quantum-correlated in photon number, i.e. a two-mode squeezed vacuum state

$$|\chi\rangle = \sqrt{1 - \lambda^2} \sum_{n=0}^{\infty} \lambda^n |n, n\rangle. \quad (1)$$

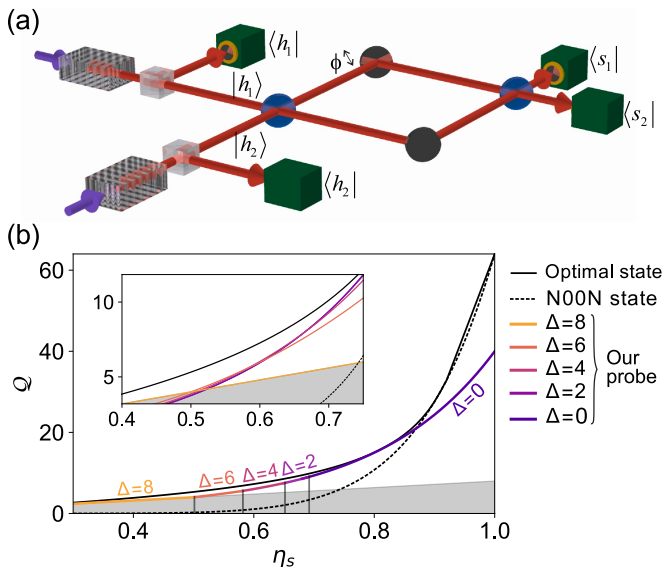


FIG. 1. **Interferometric scheme.** (a) Two type-II parametric down-conversion sources each produce orthogonally-polarized pairs of beams that are separated using polarizing beam splitters. By measuring one of the beams from each source with a photon-number-resolving detector, we herald a pair of photon-number states  $|h_1, h_2\rangle$ . We inject this probe into an interferometer and perform photon counting at the output to estimate the unknown phase difference  $\phi$ . (b) Quantum Fisher information  $\mathcal{Q}$  calculated for 8-photon ( $N = h_1 + h_2 = 8$ ) probes inside the interferometer as a function of the signal transmissivity  $\eta_s$  which is assumed to be equal in both interferometer modes. Coloured curve in the main figure plots  $\mathcal{Q}$  of the probe with the optimal  $\Delta = |h_1 - h_2|$  for a given  $\eta_s$ , while the inset shows the full curves of each probe for  $\eta_s \in [0.4, 0.75]$ . Our probe approximates the performance of the optimal state [black line] and surpasses that of the N00N state [dashed line] for efficiencies below  $\sim 90\%$ . The grey filled region indicates performance below the shot-noise limit.

Here,  $\lambda$  is a parameter that determines the average number of photons in each beam,  $\langle n \rangle = \lambda^2 / (1 - \lambda^2)$ . Measuring one of the beams with an ideal lossless photon-number-resolving detector projects the second beam onto a known photon-number state  $|h_1\rangle$ . Duplicating this procedure with a second independent source and detector, we herald pairs of photon-number states that are not necessarily identical, i.e. the probe  $|h_1, h_2\rangle$ . When these states are combined on the first beam splitter, multiphoton interference generates a path-entangled probe inside the interferometer [38].

We quantify the phase-sensitivity of the probe inside the interferometer by calculating the quantum Fisher information  $\mathcal{Q}$ . The quantity  $\mathcal{Q}$  provides a lower limit on the best achievable phase uncertainty via the quantum Cramer-Rao bound,  $\Delta\phi \geq 1/\sqrt{\mathcal{Q}}$ . The bound can be saturated using the optimal measurement strategy, which in the absence of loss is photon counting for the probes considered here [39, 40].

In Fig. 1(b), we plot  $\mathcal{Q}$  for several probes with the same total photon number  $N = h_1 + h_2 = 8$ , but different  $\Delta = |h_1 - h_2|$ , as a function of the signal transmissivity  $\eta_s$  which we assume to be equal in both interferometer modes. Probes with a small  $\Delta$  provide a greater advantage over the classical shot-noise limit but are more sensitive to losses. Since the probe is heralded in our scheme, one can choose the optimal  $\Delta$  for a given  $\eta_s$ .

Also shown in Fig. 1(b) is  $\mathcal{Q}$  for the optimal state that maximizes this parameter for a given  $N$  and  $\eta_s$ . This state has been found in Ref. [19]; the derivation is reproduced in the Supplementary Method 1. For the loss-free case ( $\eta_s = 1$ ), the optimal state is the N00N state. However, for efficiencies below  $\sim 90\%$ , our probes significantly surpass the N00N state in terms of  $\mathcal{Q}$ , exhibiting performance close to optimal. Moreover, in contrast to the N00N and Holland-Burnett states, our probe performs at least as well as the shot-noise limit for any amount of loss.

We now turn to the experiment. Both PDC sources are periodically poled potassium titanyl phosphate (pp-KTP) waveguides pumped with  $\sim 0.5$  ps long pulses from a mode-locked laser at a repetition rate of 100 kHz. The four detectors are superconducting transition edge sensors which we use to count up to 10 photons with a detection efficiency exceeding 95% [34]. The interferometer is a fiber-based device in which we can control the distance between two evanescently-coupled fibers using a micrometer to vary  $\phi$ , much like changing the path length difference between two arms of an interferometer. Further details on the experimental setup can be found in the Methods.

We measure interference fringes given by  $\text{pr}_{s_1, s_2, h_1, h_2}(\phi)$ , the joint photon-number probability per pump pulse to obtain the herald outcome  $(h_1, h_2)$  and measure  $(s_1, s_2)$  at the output of the interferometer when the phase difference is  $\phi$ . We will refer to this as the  $(s_1, s_2, h_1, h_2)$  rate. To quantify the phase sensitivity of the rates measured with a particular herald outcome  $(h_1, h_2)$ , we calculate the Fisher information:

$$\mathcal{F}_{h_1, h_2}(\phi) = \sum_{s_1, s_2} \frac{[\partial_\phi \tilde{\text{pr}}_{s_1, s_2, h_1, h_2}(\phi)]^2}{\tilde{\text{pr}}_{s_1, s_2, h_1, h_2}(\phi)}, \quad (2)$$

where  $\partial_\phi$  denotes the partial derivative with respect to  $\phi$ , and  $\tilde{\text{pr}}_{s_1, s_2, h_1, h_2}(\phi)$  is a model fitted to the measured rates (see Supplementary Method 2). Note that  $\mathcal{F}_{h_1, h_2}(\phi)$  quantifies the amount of information about  $\phi$  in our measurement results, i.e. for a specific measurement strategy, and so  $\mathcal{F}_{h_1, h_2}(\phi) \leq \mathcal{Q}$ . We compare the performance of our photon counting strategy to the optimal measurement strategy in the Supplementary Discussion 1.

Our primary figure of merit is the Fisher information per detected signal photon conditioned on measuring

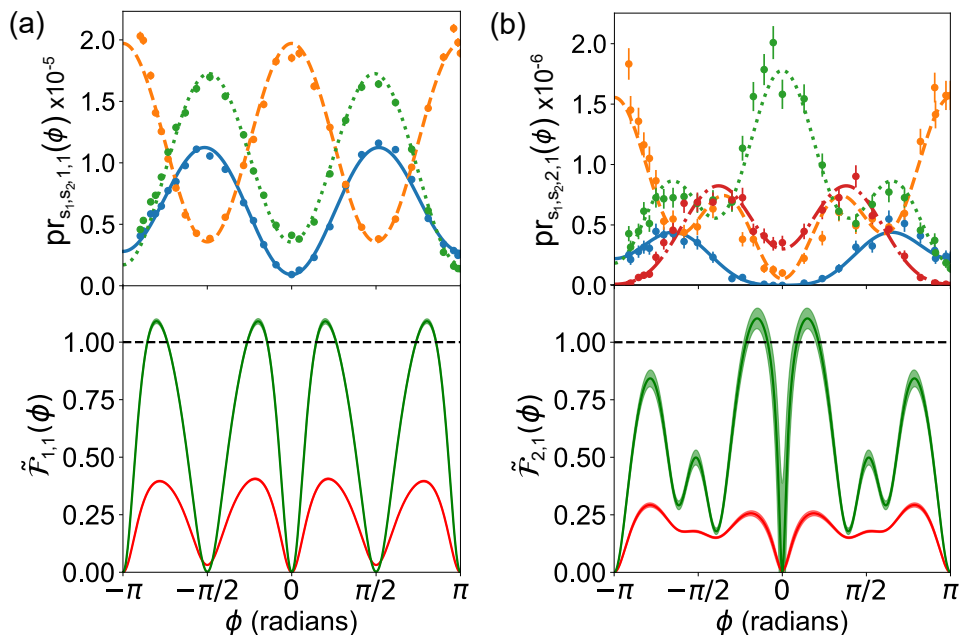


FIG. 2. **The weak gain regime.** (a) Rates measured with the probe  $|1, 1\rangle$ :  $(s_1, s_2, h_1, h_2) = (2, 0, 1, 1)$  [blue],  $(1, 1, 1, 1)$  [orange],  $(0, 2, 1, 1)$  [green]. (b) Rates measured with the probe  $|2, 1\rangle$ :  $(3, 0, 2, 1)$  [blue],  $(2, 1, 2, 1)$  [orange],  $(1, 2, 2, 1)$  [green],  $(0, 3, 2, 1)$  [red]. Error bars are one standard deviation assuming Poissonian counting statistics. Lines are a model fitted to  $\text{pr}_{s_1, s_2, h_1, h_2}(\phi)$ . Bottom panels show the normalized Fisher information  $\tilde{\mathcal{F}}_{h_1, h_2}(\phi)$  calculated using two methods: (i) post-selecting on events where  $s_1 + s_2 = h_1 + h_2$  [green] and (ii) using all events [red]. Line thicknesses show  $1\sigma$  confidence intervals obtained by fitting 50 simulated data sets that are calculated with a Monte Carlo method. The dashed black line indicates the shot-noise limit.

$(h_1, h_2)$  at the heralding detectors,

$$\tilde{\mathcal{F}}_{h_1, h_2}(\phi) = \mathcal{F}_{h_1, h_2}(\phi) / \langle \tilde{n} \rangle,$$

where

$$\langle \tilde{n} \rangle = \sum_{s_1, s_2} (s_1 + s_2) \tilde{\text{pr}}_{s_1, s_2, h_1, h_2}(\phi) \quad (3)$$

is the total number of detected signal photons. Injecting a coherent state into our interferometer would in principle yield the Fisher information  $\mathcal{F} = \langle \tilde{n} \rangle$  when the detected mean photon number is  $\langle \tilde{n} \rangle$  [18]. Thus, our figure of merit can be easily compared to the shot-noise limit which corresponds to  $\tilde{\mathcal{F}}_{h_1, h_2}(\phi) = 1$ .

We measured the total efficiency of both the heralding and signal modes to be between 47 – 55% (see Supplementary Method 3). This includes  $\sim 90\%$  waveguide transmission,  $\sim 70\%$  mode coupling efficiency into fibers, 90% interferometer transmission, and  $\gtrsim 95\%$  detector efficiency. Due to the latter two losses, the detected  $\langle \tilde{n} \rangle$  is 10-15% smaller than the mean photon number inside the interferometer. As such, the Fisher information per photon inside the interferometer (which is the relevant resource when e.g. probing a delicate sample) is about 10-15% smaller than  $\tilde{\mathcal{F}}_{h_1, h_2}(\phi)$ .

## RESULTS

We begin with low pump power to test our setup in the weak gain regime ( $\lambda \sim 0.25$ , 10  $\mu\text{W}$  per source). In Fig. 2, we show results for two different probes, (a)  $|1, 1\rangle$ , the well-studied  $N = 2$  N00N or Holland-Burnett state, and (b)  $|2, 1\rangle$ , a probe studied here for the first time. We calculate  $\tilde{\mathcal{F}}_{h_1, h_2}(\phi)$  using two methods. In the first, we discard events in which we know photons were lost by only including rates where  $s_1 + s_2 = h_1 + h_2$  in the sums of Eqs. (2) and (3). These rates are shown in the top panels of Fig. 2. Using this first method,  $\tilde{\mathcal{F}}_{h_1, h_2}(\phi)$  [green curves] surpasses the shot-noise limit by  $0.09 \pm 0.01$  for  $|1, 1\rangle$  and  $0.10 \pm 0.04$  for  $|2, 1\rangle$  at its highest point. In the second method, we include all measured events. Note that this may include events where  $s_1 + s_2 < h_1 + h_2$  due to loss in the signal modes, but also  $s_1 + s_2 > h_1 + h_2$  due to loss in the herald modes. Conditioned on obtaining the herald outcome  $(h_1, h_2)$ , the probability of the latter occurring can be minimized by reducing the pump power and hence  $\lambda$ . This increases the purity of the probe at the cost of reducing its heralding rate. Without post-selection,  $\tilde{\mathcal{F}}(\phi)$  [red curves] drops below the shot-noise limit mainly due to losses.

In addition to loss, the spectral purity and distinguishability of our photons are also sources of imperfection that reduce the contrast of the fringes and hence

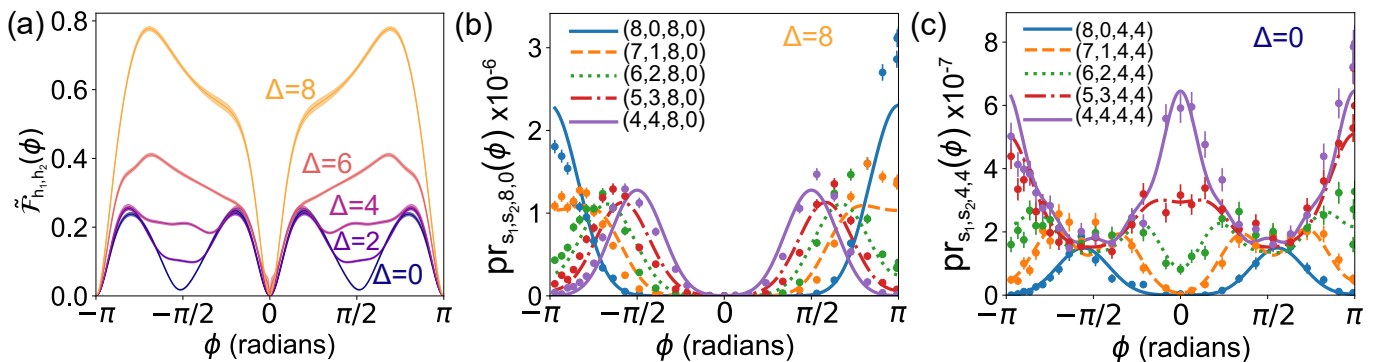


FIG. 3. **The high gain regime.** (a)  $\tilde{\mathcal{F}}_{h_1, h_2}(\phi)$  of 8-photon probes ( $N = 8$ ) parameterized by  $\Delta = |h_1 - h_2|$ . Curves are calculated using the data and Eqs. (2) and (3) without post-selection. Probes with a larger  $\Delta$  have a larger  $\tilde{\mathcal{F}}_{h_1, h_2}(\phi)$  and hence greater phase sensitivity due to their increased robustness to loss. Line thicknesses show  $1\sigma$  confidence intervals obtained by fitting 50 simulated data sets that are calculated with a Monte Carlo method. (b) and (c) show a subset of rates for the probe with  $\Delta = 8$  and  $\Delta = 0$ , respectively. Error bars are one standard deviation assuming Poissonian counting statistics. The lines are a model fitted to  $\text{pr}_{s_1, s_2, h_1, h_2}(\phi)$ .

diminish  $\tilde{\mathcal{F}}_{h_1, h_2}(\phi)$  [41]. Consider the probe  $|1, 1\rangle$ , for example. For  $\phi = \pm\pi/2$ , the whole interferometer acts as a balanced beam splitter, in which case Hong-Ou-Mandel interference should lead to a complete suppression in coincidences at its output. However, as can be seen in the orange  $(1, 1, 1, 1)$  fit in Fig. 2(a), the visibility of this interference effect is  $\sim 75\%$ . This visibility exceeds  $\sqrt{0.5}$ , which is the minimum required for demonstrating post-selected quantum-enhanced sensitivity with the probe  $|1, 1\rangle$  [12, 28, 42]. In addition to spectral mismatch between the signal modes, the visibility is degraded by uncorrelated background photons ( $\sim 5\%$  of detected photons) and the slight multi-mode nature of our sources, both of which reduce the purity of our heralded photons. We discuss source imperfections in more detail in the Supplementary Discussion 2. The finite detector energy resolution also plays a small role as the detectors have a  $\sim 1\%$  chance to mislabel an event by  $\pm 1$  photon [43].

Next, we increase the pump power to reach a high-gain regime ( $\lambda \sim 0.75$ , 135  $\mu\text{W}$  per source) in which we can herald large photon numbers. We detect 16-photon events at a rate of roughly 7 per second, which is much higher than the state-of-the-art achievable with bulk crystal PDC sources [36] or quantum dots [37]. In Fig. 3(a), we plot  $\tilde{\mathcal{F}}_{h_1, h_2}(\phi)$  calculated without post-selection for all probes with  $N = 8$ . As expected given the amount of loss in our experiment, probes with larger  $\Delta$  are more phase sensitive due to their increased robustness to loss [Fig. 1(b)]. In particular, the sensitivity of the  $\Delta = N$  probe should be shot-noise limited regardless of losses [44]. However in practice, the heralded detection of 0 photons could occur due to photon loss in the corresponding herald mode, resulting in the contamination of the signal with states for which  $\Delta \neq N$ . This degrades the performance of the  $\Delta = 8$  probe [orange curve]. In the Supplementary Discussion 3, we show that shot-noise

limited performance with the  $\Delta = N$  probe is recovered by blocking one of the sources.

## DISCUSSION

The fringes produced by our probes exhibit a number of different features compared to those measured with N00N or Holland-Burnett states. For example, with these two states, the expected signature of  $N$ -photon interference are fringe oscillations that vary as  $\cos(N\phi)$ . While our measured fringes do not exhibit such oscillations in the high gain regime, they do exhibit sharper features than classical fringes. We show this explicitly by comparing our rates to those measured with distinguishable photons. This is achieved by temporally delaying photons coming from the top source with respect to photons coming from the bottom source by more than their coherence time. As an example, we consider the probe  $|3, 2\rangle$  in Fig. 4. When the photons are injected inside the interferometer at the same time, the fringe contrast is significantly higher than when they are temporally delayed [Fig. 4(a)]. Likewise, when we calculate  $\tilde{\mathcal{F}}_{3, 2}(\phi)$  without post-selection, we find an improvement in the probe's sensitivity in the former case [Fig. 4(b)]. This demonstrates that the probe sensitivity derives from multiphoton interference even at high photon numbers.

With any finite amount of loss,  $\tilde{\mathcal{F}}_{h_1, h_2}(\phi)$  vanishes when all fringes share a common turning point such as at  $\phi = 0$ . In the case of Holland-Burnett ( $\Delta = 0$ ) and N00N states, there are also common turning points at  $\phi = \pm\pi/2$  which causes the reduction in  $\tilde{\mathcal{F}}_{h_1, h_2}(\phi)$  around these phase values [Fig. 3(c)]. In contrast, the probes with  $\Delta = 4, 6, 8$  do not have a dip in  $\tilde{\mathcal{F}}_{h_1, h_2}(\pm\pi/2)$ . The origin of this effect for  $\Delta = 8$  can be seen directly in the rates shown in Fig. 3(b). The region of the fringe with

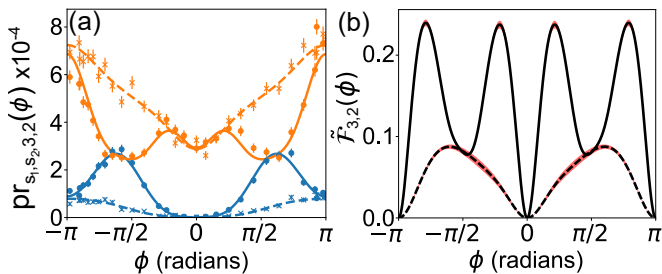


FIG. 4. **Testing multiphoton interference.** Benefits of multiphoton interference using the probe  $|3, 2\rangle$ . (a) Two sets of rates [blue:  $(5, 0, 3, 2)$ , orange:  $(3, 2, 3, 2)$ ] measured when the photons are injected inside the interferometer at the same time (data: circles, theory: bold lines) or at different times (data: crosses, theory: dashed line). In the latter case, the photons are well modelled by classical distinguishable particles. Error bars are one standard deviation assuming Poissonian counting statistics. (b)  $\tilde{F}_{3,2}(\phi)$  shows a significant improvement in sensitivity in the former case (bold line) compared to the latter case (dashed line), demonstrating that multiphoton interference improved the sensitivity of our probe. Red shaded regions shows  $1\sigma$  confidence intervals obtained by fitting 50 simulated data sets that are calculated with a Monte Carlo method.

high sensitivity to  $\phi$  (i.e. large gradient) is different for different values of  $|s_1 - s_2|$ . This feature of  $\tilde{F}_{h_1, h_2}(\phi)$  allows estimating  $\phi$  without prior knowledge of the range in which it lies, as is required for N00N or Holland-Burnett states, and thus provides a means for global phase estimation without using an adaptive protocol [27, 45].

Finally, we briefly compare our results to other works reporting Fisher information per detected photon. The highest achieved here is  $\sim 1.1$  using the herald outcome  $(2, 1)$ , i.e. a  $N = 3$  probe. Ref. [31] and Ref. [17] respectively report  $\sim 1.25$  and  $\sim 1.2$  using a  $N = 2$  probe. The latter work also achieves a Fisher information per photon inside the interferometer (i.e. accounting for undetected photons) of  $\sim 1.15$  which thus far is the only experiment demonstrating an unconditional improvement to the shot-noise limit. In the Supplementary Discussion 4, we estimate that an efficiency of 80% (in all four modes) and quantum interference visibility of 85% would be sufficient to demonstrate an improvement to the shot-noise limit with  $N = 8$  photons without post-selection. Although we do not attain these parameters in our experiment, our results do demonstrate the robustness of our probes to losses despite their large size. For example, the Fisher information per photon calculated without post-selection for the  $N = 8$  probe with  $\Delta = 6$  [Fig. 3(a)] is slightly higher than that of the  $N = 2$  N00N state [Fig. 2(a)]. This contradicts the usual expectation that large entangled probes will necessarily be more fragile to noise and loss.

In summary, we proposed and experimentally demonstrated a scheme for quantum-enhanced interferometry that exploits bright two-mode squeezed vacuum sources

and photon-number-resolving detectors. We measured interference fringes involving up to 16 photons which is significantly higher than the previous state-of-the-art [35, 36]. Crucially, our scheme prepares probes that are nearly optimally robust to losses and hence addresses one of the principal challenges when scaling-up to large entangled photonic states. With further improvements in the quality (e.g. coupling efficiency into optical fiber and purity) of bright two-mode squeezed vacuum sources compatible with transition edge sensors [33, 46], we believe our loss-tolerant scheme provides a promising route towards achieving quantum-enhanced resolution using large entangled photonic states.

## METHODS

### Sources

We pick 150-fs pulses from a mode-locked Ti:Sapphire laser (Coherent Mira-HP) at a rate of 100 kHz using a Pockels-cell-based pulse picker having a 50 dB extinction ratio. This repetition rate is chosen to accommodate the recovery time of the transition edge sensor detectors. The pump pulses are filtered to  $783 \pm 2$  nm [full-width at half maximum] using a pair of angle-tuned bandpass filters. We split the pulses into two paths that are matched in length using a translation stage. In each path, we pump a 8 mm long ppKTP waveguide that is phase-matched for type-II parametric down-conversion. At the exit of the waveguide, the pump light is rejected with a longpass filter, and the orthogonally-polarized down-converted modes are separated using a polarizing beam splitter. Each down-converted mode is filtered with a bandpass filter whose bandwidth is chosen to transmit the main feature of the down-converted spectrum but reject its side-lobes. The herald modes ( $1566 \pm 7$  nm) are coupled into single-mode fibers and sent directly to the detectors. The signal modes ( $1567 \pm 7$  nm) are coupled into polarization-maintaining single-mode fibers and sent into the interferometer. Details on the coupling efficiency and the spectral indistinguishability of the signal modes are provided in the Supplementary Discussion 2.

### Interferometer

The interferometer is a fiber-based variable beam splitter (Newport F-CPL-1550-P-FP). The splitting ratio is adjusted by controlling the distance between two evanescently-coupled fibers using a micrometer, which is analogous to changing the path length difference between two arms of an interferometer. In fact, any variable beam splitter that coherently splits light into two modes can be described by the same transformation as a Mach-Zender-type interferometer [47].



During data acquisition, we scan the distance  $x$  between the two evanescently-coupled fibers. To display our data as a function of the interferometer phase, we first calculate the transmission coefficient  $T(x)$  of the variable beam splitter using the measured  $(1, 0, 1, 0)$  and  $(0, 1, 1, 0)$  rates:

$$T(x) = \frac{\text{pr}_{1,0,1,0}(x)}{\text{pr}_{1,0,1,0}(x) + \text{pr}_{0,1,1,0}(x)}. \quad (4)$$

At low powers, we find that the quantity  $T(x)$  typically varies within  $[0.02, 0.98]$ . To obtain the corresponding phase, we correct for the imperfect visibility:

$$T_{\text{corr}}(x) = \frac{T(x) - \min[T(x)]}{\max[T(x)] - \min[T(x)]} \quad (5)$$

such that  $T_{\text{corr}}(x)$  varies between  $[0,1]$ . For a single photon injected into a Mach-Zender type interferometer with phase difference  $\phi$  between its two arms, one expects  $T_{\text{corr}}(x) = [1 - \cos(\phi)]/2$ . Solving for  $\phi$ , we find:

$$\phi(x) = \arccos(2T_{\text{corr}}(x) - 1). \quad (6)$$

### Detectors

Our detectors are superconducting transition edge sensor detectors that operate at a temperature of 85 mK inside a dilution refrigerator. Details on their physical operation can be found in Ref. [34]. An electrical trigger signal from the pump laser begins a 6  $\mu\text{s}$  time window of data acquisition during which the detector outputs are amplified and recorded with an analogue-to-digital converter. We use a matched-filter technique in real-time to convert each detector's output trace into a scalar value [48]. The scalar value is then converted into a photon number using bins that are set during an initial calibration run prior to data acquisition.

### DATA AVAILABILITY

The data sets generated and/or analyzed during this study are available from the corresponding author on reasonable request. Correspondence and requests for materials should be addressed to G.S.T.\*

### ACKNOWLEDGEMENTS

We thank B. Vlastakis for his assistance with the operation of the dilution refrigerator. This work was supported by the following: the Natural Sciences and Engineering Research Council of Canada (NSERC); the Networked Quantum Information Technologies Hub (NQIT) as part of the UK National Quantum Technologies Programme GrantEP/N509711/1) and within "First Team"

project No. POIR.04.04.00-00-220E/16-00 (originally: FIRST TEAM/2016-2/17) of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund.

### COMPETING INTERESTS

The authors declare no competing interests.

### AUTHOR CONTRIBUTIONS

Both G.S.T. and M.E.M. contributed equally. G.S.T. performed the experiment with assistance from B.A.B., C.G.W., A.E., D.S.P.; M.E.M. and G.S.T. performed numerical calculations with assistance from A.B.; A.E.L., T.G., and S.W.N. developed the detectors; R.B.P., M.S., A.I.L., and I.A.W. initiated and/or supervised the project; G.S.T. and M.E.M. wrote the manuscript with input from all authors.

---

\* guillaume.thekkadath@physics.ox.ac.uk

- [1] R. Demkowicz-Dobrzański, M. Jarzyna, and J. Kolodyński, in *Progress in Optics*, Vol. 60 (Elsevier, 2015) pp. 345–435.
- [2] C. M. Caves, *Phys. Rev. D* **23**, 1693 (1981).
- [3] M. Tse *et al.*, *Phys. Rev. Lett.* **123**, 231107 (2019).
- [4] A. Crespi, M. Lobino, J. C. Matthews, A. Politi, C. R. Neal, R. Ramponi, R. Osellame, and J. L. O'Brien, *App. Phys. Lett.* **100**, 233704 (2012).
- [5] F. Wolfgramm, C. Vitelli, F. A. Beduini, N. Godbout, and M. W. Mitchell, *Nat. Photonics* **7**, 28 (2013).
- [6] M. A. Taylor, J. Janousek, V. Daria, J. Knittel, B. Hage, H.-A. Bachor, and W. P. Bowen, *Nat. Photon.* **7**, 229 (2013).
- [7] M. A. Taylor and W. P. Bowen, *Phys. Rep.* **615**, 1 (2016).
- [8] V. Cimini, M. Mellini, G. Rampioni, M. Sbroscia, L. Leoni, M. Barbieri, and I. Gianani, *Opt. Express* **27**, 35245 (2019).
- [9] J. J. Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, *Phys. Rev. A* **54**, R4649 (1996).
- [10] M. W. Mitchell, J. S. Lundeen, and A. M. Steinberg, *Nature* **429**, 161 (2004).
- [11] P. Walther, J.-W. Pan, M. Aspelmeyer, R. Ursin, S. Gasparoni, and A. Zeilinger, *Nature* **429**, 158 (2004).
- [12] T. Nagata, R. Okamoto, J. L. O'Brien, K. Sasaki, and S. Takeuchi, *Science* **316**, 726 (2007).
- [13] H. Kim, H. S. Park, and S.-K. Choi, *Opt. Express* **17**, 19720 (2009).
- [14] I. Afek, O. Ambar, and Y. Silberberg, *Science* **328**, 879 (2010).
- [15] J. C. F. Matthews, A. Politi, D. Bonneau, and J. L. O'Brien, *Phys. Rev. Lett.* **107**, 163602 (2011).
- [16] A. E. Ulanov, I. A. Fedorov, D. Sychev, P. Grangier, and A. Lvovsky, *Nat. Commun.* **7**, 11925 (2016).

- [17] S. Slussarenko, M. M. Weston, H. M. Chrzanowski, L. K. Shalm, V. B. Verma, S. W. Nam, and G. J. Pryde, *Nat. Photonics* **11**, 700 (2017).
- [18] A. Datta, L. Zhang, N. Thomas-Peter, U. Dorner, B. J. Smith, and I. A. Walmsley, *Phys. Rev. A* **83**, 063836 (2011).
- [19] U. Dorner, R. Demkowicz-Dobrzański, B. J. Smith, J. S. Lundeen, W. Wasilewski, K. Banaszek, and I. A. Walmsley, *Phys. Rev. Lett.* **102**, 040403 (2009).
- [20] M. Kacprowicz, R. Demkowicz-Dobrzański, W. Wasilewski, K. Banaszek, and I. Walmsley, *Nat. Photonics* **4**, 357 (2010).
- [21] R. Demkowicz-Dobrzański, K. Banaszek, and R. Schnabel, *Phys. Rev. A* **88**, 041802 (2013).
- [22] M. D. Lang and C. M. Caves, *Phys. Rev. Lett.* **111**, 173601 (2013).
- [23] H. Yonezawa, D. Nakane, T. A. Wheatley, K. Iwasawa, S. Takeda, H. Arao, K. Ohki, K. Tsumura, D. W. Berry, T. C. Ralph, H. M. Wiseman, E. H. Huntington, and A. Furusawa, *Science* **337**, 1514 (2012).
- [24] A. A. Berni, T. Gehring, B. M. Nielsen, V. Händchen, M. G. Paris, and U. L. Andersen, *Nat. Photon.* **9**, 577 (2015).
- [25] M. J. Holland and K. Burnett, *Phys. Rev. Lett.* **71**, 1355 (1993).
- [26] F. Sun, B. Liu, Y. Gong, Y. Huang, Z. Ou, and G. Guo, *EPL* **82**, 24001 (2008).
- [27] G.-Y. Xiang, B. L. Higgins, D. Berry, H. M. Wiseman, and G. Pryde, *Nat. Photonics* **5**, 43 (2011).
- [28] N. Thomas-Peter, B. J. Smith, A. Datta, L. Zhang, U. Dorner, and I. A. Walmsley, *Phys. Rev. Lett.* **107**, 113603 (2011).
- [29] G. Xiang, H. Hofmann, and G. Pryde, *Sci. Rep.* **3**, 2684 (2013).
- [30] R.-B. Jin, M. Fujiwara, R. Shimizu, R. J. Collins, G. S. Buller, T. Yamashita, S. Miki, H. Terai, M. Takeoka, and M. Sasaki, *Sci. Rep.* **6**, 36914 (2016).
- [31] J. C. Matthews, X.-Q. Zhou, H. Cable, P. J. Shadbolt, D. J. Saunders, G. A. Durkin, G. J. Pryde, and J. L. O'Brien, *npj Quantum Inf.* **2**, 1 (2016).
- [32] A. Eckstein, A. Christ, P. J. Mosley, and C. Silberhorn, *Phys. Rev. Lett.* **106**, 013603 (2011).
- [33] G. Harder, T. J. Bartley, A. E. Lita, S. W. Nam, T. Gerrits, and C. Silberhorn, *Phys. Rev. Lett.* **116**, 143601 (2016).
- [34] A. E. Lita, A. J. Miller, and S. W. Nam, *Opt. Express* **16**, 3032 (2008).
- [35] W.-B. Gao, C.-Y. Lu, X.-C. Yao, P. Xu, O. Gühne, A. Goebel, Y.-A. Chen, C.-Z. Peng, Z.-B. Chen, and J.-W. Pan, *Nat. Phys.* **6**, 331 (2010).
- [36] X.-L. Wang, L.-K. Chen, W. Li, H.-L. Huang, C. Liu, C. Chen, Y.-H. Luo, Z.-E. Su, D. Wu, Z.-D. Li, H. Lu, Y. Hu, X. Jiang, C.-Z. Peng, L. Li, N.-L. Liu, Y.-A. Chen, C.-Y. Lu, and J.-W. Pan, *Phys. Rev. Lett.* **117**, 210502 (2016).
- [37] H. Wang, J. Qin, X. Ding, M.-C. Chen, S. Chen, X. You, Y.-M. He, X. Jiang, L. You, Z. Wang, C. Schneider, J. J. Renema, S. Höfling, C.-Y. Lu, and J.-W. Pan, *Phys. Rev. Lett.* **123**, 250503 (2019).
- [38] M. Stobińska, A. Buraczewski, M. Moore, W. Clements, J. J. Renema, S. Nam, T. Gerrits, A. Lita, W. Kolthammer, A. Eckstein, *et al.*, *Sci. Adv.* **5**, eaau9674 (2019).
- [39] H. F. Hofmann, *Phys. Rev. A* **79**, 033822 (2009).
- [40] W. Zhong, Y. Huang, X. Wang, and S.-L. Zhu, *Phys. Rev. A* **95**, 052304 (2017).
- [41] P. M. Birchall, J. Sabines-Chesterking, J. L. O'Brien, H. Cable, and J. C. Matthews, arXiv preprint arXiv:1603.00686 (2016).
- [42] K. J. Resch, K. L. Pregnell, R. Prevedel, A. Gilchrist, G. J. Pryde, J. L. O'Brien, and A. G. White, *Phys. Rev. Lett.* **98**, 223601 (2007).
- [43] P. C. Humphreys, B. J. Metcalf, T. Gerrits, T. Hiemstra, A. E. Lita, J. Nunn, S. W. Nam, A. Datta, W. S. Kolthammer, and I. A. Walmsley, *New J. Phys.* **17**, 103044 (2015).
- [44] L. Pezzé, A. Smerzi, G. Khoury, J. F. Hodelin, and D. Bouwmeester, *Phys. Rev. Lett.* **99**, 223602 (2007).
- [45] S. Daryanoosh, S. Slussarenko, D. Berry, H. Wiseman, and G. Pryde, *Nat. Commun.* **9**, 2041 (2018).
- [46] V. Vaidya, B. Morrison, L. Helt, R. Shahrokhshahi, D. Mahler, M. Collins, K. Tan, J. Lavoie, A. Repington, M. Menotti, *et al.*, *Sci. Adv.* **6**, eaba9186 (2020).
- [47] J. Flórez, N. J. Carlson, C. H. Nacke, L. Giner, and J. S. Lundeen, *Rev. Sci. Instrum.* **89**, 023108 (2018).
- [48] E. Figueroa-Feliciano, B. Cabrera, A. Miller, S. Powell, T. Saab, and A. Walker, *Nucl. Instrum. Methods Phys. Res.* **444**, 453 (2000).
- [49] S. L. Braunstein and C. M. Caves, *Phys. Rev. Lett.* **72**, 3439 (1994).
- [50] R. Demkowicz-Dobrzański, U. Dorner, B. J. Smith, J. S. Lundeen, W. Wasilewski, K. Banaszek, and I. A. Walmsley, *Phys. Rev. A* **80**, 013825 (2009).
- [51] U. Leonhardt, *Essential quantum optics: from quantum measurements to black holes* (Cambridge University Press, 2010).
- [52] A. Worsley, H. Coldenstrodt-Ronge, J. Lundeen, P. Mosley, B. Smith, G. Puentes, N. Thomas-Peter, and I. Walmsley, *Opt. Express* **17**, 4397 (2009).
- [53] T. Ono and H. F. Hofmann, *Phys. Rev. A* **81**, 033819 (2010).
- [54] C. Oh, S.-Y. Lee, H. Nha, and H. Jeong, *Phys. Rev. A* **96**, 062304 (2017).
- [55] L. Padberg, M. Santandrea, M. Rüsing, J. Brockmeier, P. Mackwitz, G. Berth, A. Zrenner, C. Eigner, and C. Silberhorn, *Opt. Express* **28**, 24353 (2020).
- [56] M. F. Volk, C. E. Rüter, M. Santandrea, C. Eigner, L. Padberg, H. Herrmann, C. Silberhorn, and D. Kip, *Opt. Mater. Express* **8**, 82 (2018).
- [57] A. Eckstein, *Mastering quantum light pulses with nonlinear waveguide interactions*, Ph.D. thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (2012).

SUPPLEMENTARY INFORMATION

**Supplementary Note 1: Quantum Fisher information of generalized Holland-Burnett states**

Here we derive the quantum Fisher information of our generalized Holland-Burnett states. We first consider the ideal lossless case. In general, the quantum Fisher information of a pure state  $|\Psi(\phi)\rangle$  that depends on some parameter  $\phi$  is given by [49]:

$$\mathcal{Q} = 4 \left( \langle \partial_\phi \Psi(\phi) | \partial_\phi \Psi(\phi) \rangle - |\langle \partial_\phi \Psi(\phi) | \Psi(\phi) \rangle|^2 \right) \quad (\text{S7})$$

where  $|\partial_\phi \Psi(\phi)\rangle \equiv \partial_\phi |\Psi(\phi)\rangle$ . In our case,  $|\Psi\rangle$  is the two-mode state inside the interferometer (before the phase shift) and  $|\Psi(\phi)\rangle = e^{i\hat{c}^\dagger \hat{c} \phi} |\Psi\rangle$  is the state after the phase shift  $\phi$  is applied in the upper interferometer mode  $c$ . After some simple algebra, one finds that  $\mathcal{Q}$  is independent of  $\phi$  and is determined by:

$$\mathcal{Q} = 4 \left[ \langle \Psi | (\hat{c}^\dagger \hat{c})^2 | \Psi \rangle - \langle \Psi | \hat{c}^\dagger \hat{c} | \Psi \rangle^2 \right]. \quad (\text{S8})$$

We wish to calculate  $\mathcal{Q}$  for the particular probe  $|\Psi\rangle = \hat{U}_{\text{BS}} |h_1, h_2\rangle$  where  $\hat{U}_{\text{BS}}$  is the balanced beam splitter unitary transformation. The second term in Eq. (S8) is given by:

$$\langle \Psi | \hat{c}^\dagger \hat{c} | \Psi \rangle = \langle h_1, h_2 | \hat{U}_{\text{BS}}^\dagger \hat{c}^\dagger \hat{c} \hat{U}_{\text{BS}} | h_1, h_2 \rangle \quad (\text{S9})$$

$$= \langle h_1, h_2 | \hat{U}_{\text{BS}}^\dagger \hat{c}^\dagger \hat{U}_{\text{BS}} \hat{U}_{\text{BS}}^\dagger \hat{c} \hat{U}_{\text{BS}} | h_1, h_2 \rangle \quad (\text{S10})$$

$$= \langle h_1, h_2 | \left( \frac{a^\dagger + b^\dagger}{\sqrt{2}} \right) \left( \frac{a + b}{\sqrt{2}} \right) | h_1, h_2 \rangle \quad (\text{S11})$$

$$= \frac{1}{2} \langle h_1, h_2 | a^\dagger a + b^\dagger b | h_1, h_2 \rangle \quad (\text{S12})$$

$$= \frac{1}{2} (h_1 + h_2), \quad (\text{S13})$$

where in line (S10) we used the fact that  $\hat{U}_{\text{BS}}$  is unitary and in line (S11) we transformed mode  $c$  to the input modes  $a$  and  $b$ . The first term in equation (S8) is calculated in a similar manner:

$$\langle \Psi | (\hat{c}^\dagger \hat{c})^2 | \Psi \rangle = \langle h_1, h_2 | \hat{U}_{\text{BS}}^\dagger \hat{c}^\dagger \hat{c} \hat{c}^\dagger \hat{c} \hat{U}_{\text{BS}} | h_1, h_2 \rangle \quad (\text{S14})$$

$$= \langle h_1, h_2 | \left( \hat{U}_{\text{BS}}^\dagger \hat{c}^\dagger \hat{U}_{\text{BS}} \hat{U}_{\text{BS}}^\dagger \hat{c} \hat{U}_{\text{BS}} \hat{U}_{\text{BS}}^\dagger \hat{c} \hat{U}_{\text{BS}} \right)^2 | h_1, h_2 \rangle \quad (\text{S15})$$

$$= \langle h_1, h_2 | \left( \frac{a^\dagger + b^\dagger}{\sqrt{2}} \right)^2 \left( \frac{a + b}{\sqrt{2}} \right)^2 | h_1, h_2 \rangle \quad (\text{S16})$$

$$= \frac{1}{4} \langle h_1, h_2 | (a^\dagger + b^\dagger)^2 (a + b)^2 | h_1, h_2 \rangle \quad (\text{S17})$$

$$= \frac{1}{4} \langle h_1, h_2 | a^\dagger a a^\dagger a + b^\dagger b b^\dagger b + 4a^\dagger a b^\dagger b + a^\dagger a + b^\dagger b | h_1, h_2 \rangle \quad (\text{S18})$$

$$= \frac{1}{4} (h_1^2 + h_2^2 + 4h_1 h_2 + h_1 + h_2). \quad (\text{S19})$$

Therefore  $\mathcal{Q}$  is given by

$$\mathcal{Q} = \frac{1}{4} (h_1^2 + h_2^2 + 4h_1 h_2 + h_1 + h_2) - \frac{1}{4} (h_1 + h_2)^2 = 2h_1 h_2 + h_1 + h_2. \quad (\text{S20})$$

Eq. (S20) only applies when there are no losses in the system. In the presence of losses, the probe  $|\Psi\rangle$  is transformed to a mixed state  $\hat{\rho}$ . Then,  $\mathcal{Q}$  is calculated using

$$\mathcal{Q} = \text{Tr} \left\{ \hat{\rho}(\phi) \hat{\Lambda}^2 [\hat{\rho}(\phi)] \right\} \quad (\text{S21})$$

where  $\hat{\rho}(\phi) = e^{-i\hat{c}^\dagger \hat{c} \phi} \hat{\rho} e^{i\hat{c}^\dagger \hat{c} \phi}$  is the probe state after the phase shift  $\phi$  and  $\hat{\Lambda}[\hat{\rho}(\phi)]$  is a Hermitian operator called the ‘‘symmetric logarithmic derivative’’ defined implicitly via

$$\partial_\phi \hat{\rho}(\phi) = \frac{1}{2} \left\{ \hat{\Lambda}[\hat{\rho}(\phi)] \hat{\rho}(\phi) + \hat{\rho}(\phi) \hat{\Lambda}[\hat{\rho}(\phi)] \right\}. \quad (\text{S22})$$



We notice that by combining Eq. (S22) with Eq. (S21) we obtain an alternative equation for the QFI

$$\mathcal{Q} = \text{Tr} \left\{ \partial_\phi \hat{\rho}(\phi) \Lambda[\hat{\rho}(\phi)] \right\}. \quad (\text{S23})$$

By writing  $\hat{\rho}$  in its eigenbasis,  $\hat{\rho} = \sum_i p_i |e_i\rangle \langle e_i|$  and writing out the derivative  $\partial_\phi \hat{\rho}(\phi) = ie^{-i\hat{c}^\dagger \hat{c} \phi} [\hat{\rho}, \hat{c}^\dagger \hat{c}] e^{i\hat{c}^\dagger \hat{c} \phi}$ , it can be shown that  $\mathcal{Q}$  is given by [1]

$$\mathcal{Q} = \sum_{i,j} \frac{2 |\langle e_i | \hat{c}^\dagger \hat{c} | e_j \rangle|^2 (p_i - p_j)^2}{p_i + p_j} \quad (\text{S24})$$

which is independent of  $\phi$ . The sum is taken over all terms with a non-vanishing denominator.

### Supplementary Method 1: Optimal states

In the main text we compare the performance of our probes to the ‘‘optimal states’’ which provide the largest possible quantum Fisher information given some amount of loss [19, 50]. A general  $N$ -photon pure state inside the interferometer can be written in the Fock basis as

$$|\Psi\rangle = \sum_{n=0}^N \alpha_n |n, N-n\rangle. \quad (\text{S25})$$

In the absence of loss, the optimal state is found by optimizing the coefficients  $\{\alpha_n\}$  to maximize the quantum Fisher information  $\mathcal{Q} = 4 \left[ \langle \Psi | (\hat{c}^\dagger \hat{c})^2 | \Psi \rangle - \langle \Psi | \hat{c}^\dagger \hat{c} | \Psi \rangle^2 \right]$ .

In the presence of loss,  $|\Psi\rangle$  turns into a mixture  $\hat{\rho}$  which can be written in the following form

$$\hat{\rho} = \sum_j p_j |\Psi_j\rangle \langle \Psi_j|, \quad (\text{S26})$$

where  $|\Psi_j\rangle$  do not have to be orthogonal. Due to the convexity of quantum Fisher information,  $\mathcal{Q}'$  of  $\hat{\rho}$  is upper bounded by

$$\mathcal{Q}' \leq \mathcal{Q} = 4 \sum_j p_j \left( \langle \Psi_j | (\hat{c}^\dagger \hat{c})^2 | \Psi_j \rangle - \langle \Psi_j | \hat{c}^\dagger \hat{c} | \Psi_j \rangle^2 \right). \quad (\text{S27})$$

The bound is attained if the kets  $|\Psi_j\rangle$  are orthogonal, which is the case for e.g. N00N states or if photon losses are present in only one interferometer mode.

Applying Eq. (S27) to Eq. (S25), we obtain

$$\mathcal{Q} = 4 \left( \sum_{n=0}^N n^2 x_n - \sum_{l=0}^N \sum_{m=0}^{N-l} \frac{(\sum_{k=l}^{N-m} x_n n B_{lm}^n)^2}{\sum_{n=l}^{N-m} x_n B_{lm}^n} \right), \quad (\text{S28})$$

where  $x_n = |\alpha_n|^2$ ,  $B_{lm}^n \equiv \binom{n}{l} \binom{N-n}{m} \eta_{s_1}^n (\eta_{s_1}^{-1} - 1)^l \eta_{s_2}^{N-n} (\eta_{s_2}^{-1} - 1)^m$  and  $\eta_{s_1}, \eta_{s_2}$  denote the transmittances in the signal modes.

The optimal states are found by numerically maximizing  $\mathcal{Q}$  over the probabilities  $\{x_n\}$ . Since  $\mathcal{Q}$  is a concave function of  $\{x_n\}$  [50], any maximum is global. Although  $\mathcal{Q}' < \mathcal{Q}$  [Eq. (S27)] when losses are present in both modes, the difference between the two quantities is small relative to the difference between the shot-noise limit and the Heisenberg limit [50]. Due to this approximation, the optimized  $\mathcal{Q}$  is a slight over-estimate of true quantum Fisher information  $\mathcal{Q}'$  of the optimal states.

Fig. 1(b) in the main text shows  $\mathcal{Q}$  of our probes and the optimal state as a function of equal transmissivity in the signal modes  $\eta_{s_1} = \eta_{s_2} = \eta_s$  which varied from 0 to 1 in steps of 0.01. The optimal state was calculated in MATHEMATICA by maximizing over coefficients  $\{x_n\}$  in Eq. (S28), assuming they all sum up to 1 and are real and positive. We computed  $\mathcal{Q}$  of our probes in Python using the following method. We started with two copies of the state in Eq. (1) in the main text, inserted a beam splitter in each signal mode, and traced over the reflected port to model signal transmissivities  $\eta_s$ . The two matrices were then combined on the first interferometer beam splitter

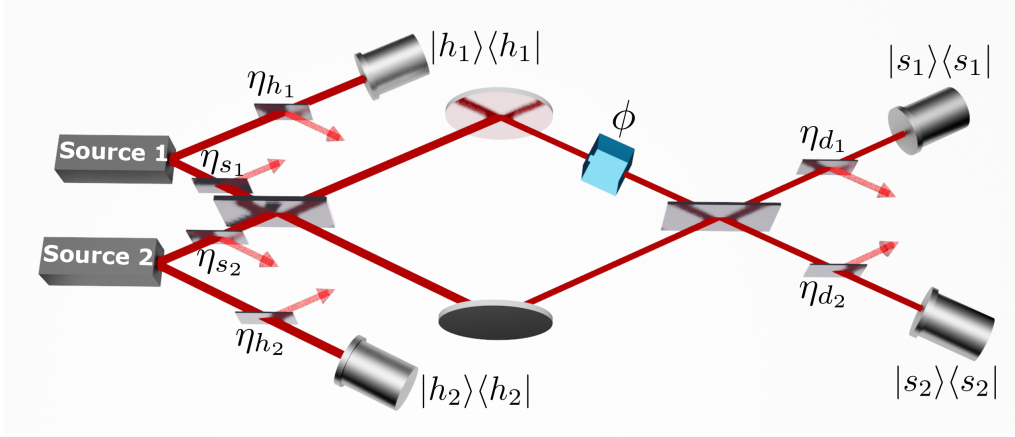


FIG. S5. Losses are modelled by placing fictitious beam splitters in all four modes before the interferometer and just before the signal detectors. Coefficients show the transmission of the beam splitters.

forming a four mode density matrix, which was then reduced to two modes by projectively measuring the two herald modes. Eigenvalues and eigenvectors were found for the two-mode density matrix inside the interferometer which were then used to calculate  $\mathcal{Q}$  via Eq. (S24). To compare the ideal performance of our probes with the optimal state, we exclude the effect of imperfect heralding on the former by using  $\eta_{h_1} = \eta_{h_2} = 1$  in the calculation of  $\mathcal{Q}$ .

From the calculations described above we show that for  $\eta_s \in (0, 0.5)$  the best approximation to the optimal state is given by the probe with  $\Delta = 8$ ; for  $\eta_s \in (0.5, 0.58)$  by the probe with  $\Delta = 6$ ; for  $\eta_s \in (0.58, 0.66)$  by the probe with  $\Delta = 4$ ; for  $\eta_s \in (0.66, 0.69)$  by the probe with  $\Delta = 2$ ; and for  $\eta_s > 0.69$  by the probe with  $\Delta = 0$ , as shown by the colored line in Fig. 1(b).

### Supplementary Method 2: Modelling the measured rates

Here we describe the model  $\tilde{\text{pr}}(s_1, s_2, h_1, h_2, \phi)$  used to fit the experimentally measured rates. We model optical loss by placing fictitious beam splitters (see Fig. S5) and tracing over the reflected modes. For now, we assume  $\eta_{d_1} = \eta_{d_2} = 1$ . We will treat the effect of these detection losses at the end.

The sources produce two-mode squeezed vacuum states:

$$|\chi_i\rangle = \sqrt{1 - \lambda_i^2} \sum_{n=0}^{\infty} \lambda_i^n |n, n\rangle, \quad (\text{S29})$$

where  $i = 1, 2$  represent sources 1 and 2, respectively. The joint photon-number distribution of this two-mode squeezed vacuum state after the losses is given by:

$$\tilde{\text{pr}}_i(x, y) = (1 - \lambda_i^2) \sum_{n=\max(x, y)}^{\infty} \binom{n}{x} \binom{n}{y} \lambda_i^{2n} \eta_{h_i}^x \eta_{s_i}^y (1 - \eta_{h_i})^{n-x} (1 - \eta_{s_i})^{n-y}. \quad (\text{S30})$$

The intuition for the expression above is as follows. Imagine that there are two detectors after the fictitious beam splitters that give the detection outcome  $(x, y)$ . The source must have produced at least  $\max(x, y)$  photon pairs and perhaps some photons were lost (i.e. reflected at the beam splitters). The probability to produce  $n$  pairs of photons is  $(1 - \lambda_i)^2 \lambda_i^{2n}$ . Having produced  $n$  pairs, the probability to reflect  $n - x$  [ $n - y$ ] photons and transmit  $x$  [ $y$ ] photons in the herald [signal] mode is  $\binom{n}{x} \eta_{h_i}^x (1 - \eta_{h_i})^{n-x}$  [ $\binom{n}{y} \eta_{s_i}^y (1 - \eta_{s_i})^{n-y}$ ]. In principle,  $n$  can range up to  $\infty$ , but in practice it suffices to truncate this sum at some value where  $(1 - \lambda_i)^2 \lambda_i^{2n}$  becomes small. In our numerics, we truncate the sum at  $n = 50$ .

If we obtain the herald outcome  $(h_1, h_2)$ , then the (unnormalized) state that is injected into the interferometer is given by:

$$\hat{\rho} = \sum_{m, n=0}^{\infty} \tilde{\text{pr}}_1(h_1, m) \tilde{\text{pr}}_2(h_2, n) |m, n\rangle \langle m, n|. \quad (\text{S31})$$

Losses occurring inside the interferometer can be absorbed into  $\eta_s$  or  $\eta_d$  if they are equal in both interferometer modes, which was approximately the case in our experiment. Thus, the interferometer transformation can be described by a unitary operator  $\hat{U}(\phi)$  which depends on the phase difference  $\phi$  between both arms. The probability that we wish to calculate is given by:

$$\tilde{\text{pr}}(s_1, s_2, h_1, h_2, \phi) = \langle s_1, s_2 | \hat{U}(\phi) \hat{\rho} \hat{U}^\dagger(\phi) | s_1, s_2 \rangle. \quad (\text{S32})$$

Knowing that there are a total of  $s_1 + s_2$  photons before the interferometer, we can constrain  $n = s_1 + s_2 - m$  and truncate the sum at  $s_1 + s_2$  in Eq. (S31). Thus, we obtain:

$$\tilde{\text{pr}}(s_1, s_2, h_1, h_2, \phi) = \sum_{m=0}^{s_1+s_2} \tilde{\text{pr}}_1(h_1, m) \tilde{\text{pr}}_2(h_2, s_1 + s_2 - m) \left| \langle s_1, s_2 | \hat{U}(\phi) | m, s_1 + s_2 - m \rangle \right|^2. \quad (\text{S33})$$

The matrix element  $\left| \langle s_1, s_2 | \hat{U}(\phi) | m, s_1 + s_2 - m \rangle \right|^2$  is derived in Ref. [51] and is given by:

$$\begin{aligned} \left| \langle s_1, s_2 | \hat{U}(\phi) | m, s_1 + s_2 - m \rangle \right|^2 &= \frac{m!(s_1 + s_2 - m)!}{s_1!s_2!} (\sin[\phi/2])^{2(s_1+m)} (\cos[\phi/2])^{2(s_2-m)} \\ &\times \left( \sum_{k=0}^{s_1} \binom{s_1}{k} \binom{s_2}{s_2+k-m} (-1)^k \tan[\phi/2]^{-2k} \right)^2. \end{aligned} \quad (\text{S34})$$

Alternatively, the matrix element can also be evaluated using Kravchuk polynomials [38].

The model for temporally distinguishable photons follows the same approach as above. While the derivation below focuses on temporal distinguishability, the same equations are valid to describe distinguishability in any other degree of freedom. We adopt a heuristic approach (e.g. as in Ref [41]) in which the temporal mode of the photons produced in the top source is decomposed into a component completely indistinguishable ( $\parallel$ ) to the temporal mode of the bottom source photons as well as a component completely distinguishable ( $\perp$ ). With this decomposition, Eq. (S31) becomes:

$$\hat{\rho}^{dist} = \sum_{m,n=0}^{\infty} \sum_{l=0}^m \binom{m}{l} \mathcal{M}^l (1 - \mathcal{M})^{m-l} \tilde{\text{pr}}_1(h_1, m) \tilde{\text{pr}}_2(h_2, n) |l, n\rangle_{\parallel} \langle l, n|_{\parallel} \otimes |m-l, 0\rangle_{\perp} \langle m-l, 0|_{\perp}. \quad (\text{S35})$$

where  $\mathcal{M} \in [0, 1]$  is a mode overlap parameter characterizing the distinguishability of the photons. For  $\mathcal{M} = 0$  ( $\mathcal{M} = 1$ ), the photons from top and bottom sources are completely distinguishable (indistinguishable). Since our detectors cannot resolve the time difference between  $\perp$  and  $\parallel$ , they convolve the probabilities for the photons to have originated from either temporal mode. This measurement is described by the following incoherent sum of projectors:

$$\hat{\Pi} = \sum_{x=0}^{s_1} \sum_{y=0}^{s_2} |s_1 - x, s_2 - y\rangle_{\parallel} \langle s_1 - x, s_2 - y|_{\parallel} \otimes |x, y\rangle_{\perp} \langle x, y|_{\perp}. \quad (\text{S36})$$

Many of the terms in the sum of Eq. (S36) can be eliminated due to constraints on the photon numbers. For example, a total of  $m - l$  photons are produced in mode  $\perp$  and so  $x + y = m - l$ . Moreover,  $s_1 + s_2 = m + n$ . After applying these constraints, the final joint probability is given by:

$$\begin{aligned} \tilde{\text{pr}}^{dist}(s_1, s_2, h_1, h_2, \phi) &= \text{Tr} \left( \hat{\Pi} \hat{U}(\phi) \hat{\rho}^{dist} \hat{U}^\dagger(\phi) \right) \\ &= \sum_{m=0}^{s_1+s_2} \sum_{l=0}^m \sum_{x=\max(0, m-s_2)}^{\min(s_1, m-l)} \binom{m}{l} \mathcal{M}^l (1 - \mathcal{M})^{m-l} \tilde{\text{pr}}_1(h_1, m) \tilde{\text{pr}}_2(h_2, s_1 + s_2 - m) \\ &\times \left| \langle s_1 - x, l + s_2 - m + x | \hat{U}(\phi) | l, s_1 + s_2 - m \rangle \right|^2 \\ &\times \left| \langle x, m - l - x | \hat{U}(\phi) | m - l, 0 \rangle \right|^2. \end{aligned} \quad (\text{S37})$$

Finally, we can now consider the effect of the losses just before the detectors. These losses can be modelled with a transformation analogous to Eq. (S30). Applying this transformation on Eq. (S33), we obtain:

$$\tilde{\text{pr}}(s_1, s_2, h_1, h_2, \phi; \eta_{d_1}, \eta_{d_2}) = \sum_{j=s_1}^{\infty} \sum_{k=s_2}^{\infty} \binom{j}{s_1} \binom{k}{s_2} \eta_{d_1}^{s_1} \eta_{d_2}^{s_2} (1 - \eta_{d_1})^{j-s_1} (1 - \eta_{d_2})^{k-s_2} \tilde{\text{pr}}(j, k, h_1, h_2, \phi) \quad (\text{S38})$$

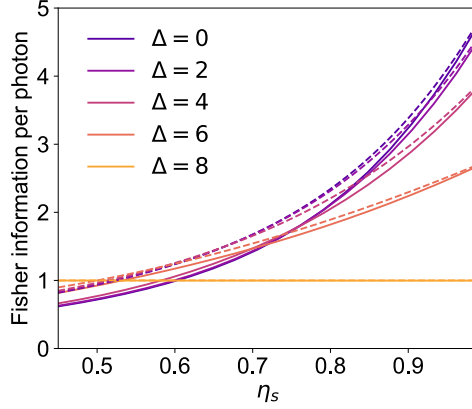


FIG. S6. Quantum Fisher information per photon,  $Q/\eta_s N$  [dashed lines], and the maximum classical Fisher information per photon,  $\max[\tilde{F}(\phi)]$  [continuous lines], for  $N = 8$  probes with different  $\Delta = |h_1 - h_2|$ . We assumed ideal heralding ( $\eta_{h_1} = \eta_{h_2} = 1$ ) and balanced loss inside the interferometer ( $\eta_{s_1} = \eta_{s_2} = \eta_s$ ). This plot compares the phase sensitivity obtainable with photon counting [continuous lines] to the optimal measurement strategy [dashed lines].

The same method is used for the distinguishable photons model, i.e. replace  $\tilde{\text{pr}}(j, k, h_1, h_2, \phi)$  with  $\tilde{\text{pr}}^{\text{dist}}(j, k, h_1, h_2, \phi)$  in Eq. (S38). In our numerics, we truncate the sums in Eq. (S38) to only include the effect of losing a few photons, which is a good approximation given the high efficiency of our number-resolving detectors.

The equations above are evaluated numerically and fitted to the experimentally measured  $\text{pr}(s_1, s_2, h_1, h_2, \phi)$  by varying the fit parameters  $\eta_{h_1}, \eta_{h_2}, \eta_{s_1}, \eta_{s_2}, \eta_{d_1}, \eta_{d_2}, \lambda_1, \lambda_2$ . Fitting is performed using the Python package LMFIT with a least squares method. Note that, for the sake of increasing the speed of the fitting, we used  $\mathcal{M} = 1$  for all data except for the dashed lines in Fig. 4 of the main text where we used  $\mathcal{M} = 0$ . Thus, the fit parameters generally did not correspond to the measured efficiencies and squeezing parameters (see below). Instead, the fitting procedure converged on larger  $\lambda$  values and smaller  $\eta$  values to emulate the effect of imperfect interference (i.e. reduced fringe visibility). We tested the full model (i.e. including  $\mathcal{M}$ ) by fitting a subset of rates measured in the high gain regime and found the fit parameters:  $\eta_{h_1} = 0.50$ ,  $\eta_{h_2} = 0.50$ ,  $\eta_{s_1} = 0.61$ ,  $\eta_{s_2} = 0.50$ ,  $\eta_{d_1} = 0.9$ ,  $\eta_{d_2} = 0.99$ ,  $\lambda_1 = 0.68$ ,  $\lambda_2 = 0.68$ , and  $\mathcal{M} = 0.73$ . These efficiency values are within error to the measured values (see below), and  $\mathcal{M} = 0.73$  is roughly consistent with the measured  $\sim 75\%$  quantum interference visibility of the  $(1, 1, 1, 1)$  rate.

### Supplementary Method 3: Estimating efficiencies

We characterize the efficiency of our setup using a Klyshko-like method that is generalized to photon-number-resolving detection [52]. We set the variable beam splitter to maximize reflection and measure the joint photon-number distribution  $\text{pr}_i(x, y)$  pumping one source at a time. We fit the measured  $\text{pr}_i(x, y)$  to  $\tilde{\text{pr}}_i(x, y)$  [see Eq. (S30)] using three parameters: the PDC gain  $\lambda_i$  and the total efficiency of the herald mode ( $\eta_{h_i}$ ) and the signal mode ( $\eta_{s_i}$ ). Note that the latter will also include the detection efficiency  $\eta_{d_i}$ . By repeating the procedure with five different pump powers, we find that  $\eta_{s_1}\eta_{d_1} = 56 \pm 3\%$  and  $\eta_{h_1} = 47 \pm 1\%$  for the first source, and  $\eta_{s_2}\eta_{d_2} = 52 \pm 4\%$  and  $\eta_{h_2} = 51 \pm 1\%$  for the second source.

### Supplementary Discussion 1: Comparing the performance of photon counting and the optimal measurement

Here we compare the performance of our measurement strategy, photon counting, to the optimal measurement strategy. In Fig. S6, we plot the classical Fisher information per photon obtained with photon counting for various  $N = 8$  probes, assuming ideal heralding ( $\eta_{h_1} = \eta_{h_2} = 1$ ) and balanced loss inside the interferometer ( $\eta_{s_1} = \eta_{s_2} = \eta_s$ ). Since this quantity generally depends on  $\phi$  when  $\eta_s < 1$ , we focus on the region with the largest phase sensitivity, i.e.  $\max[\tilde{\mathcal{F}}(\phi)]$  [continuous lines]. In the same plot, we reproduce the quantum Fisher information curves from Fig. 1(b) normalized by the number of detected photons, i.e.  $Q/\eta_s N$  [dashed lines]. For  $\eta_s = 1$ , the quantum and classical Fisher information are equal which means that photon counting is the optimal measurement strategy, as expected [39].

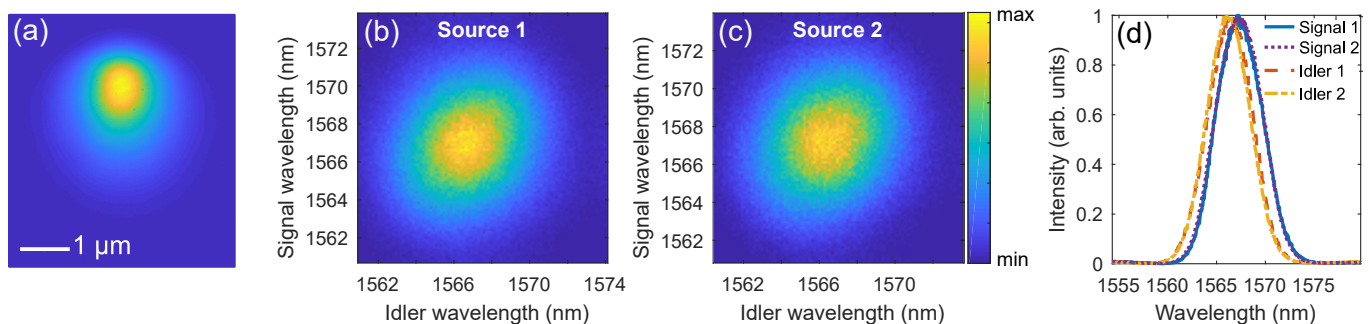


FIG. S7. Characterization of the sources. (a) Typical intensity distribution of the waveguide spatial mode measured at 1550 nm. Its non-Gaussian features limit the fiber coupling efficiency to 70%. Scale bar shows dimension in the object plane, i.e. at facet of waveguide. (b) and (c) Joint spectral intensities of sources 1 and 2, respectively. (d) Spectra of the four modes in the experiment.

For  $\eta_s < 1$ , photon counting is no longer optimal (except for  $\Delta = N$ ) but still provides quantum-enhanced phase sensitivity for  $\eta_s \gtrsim 0.55$ . To the best of our knowledge, the optimal measurement strategy for lossy Holland-Burnett interferometry is not known. We note that homodyne and weak-field homodyne (i.e. combining a signal with local oscillator then performing photon counting) have been shown to be more loss-tolerant than photon counting using Gaussian probes [53, 54]. However, it remains an open question whether these measurement strategies would be advantageous using our non-Gaussian probes in the presence of loss.

### Supplementary Discussion 2: Source imperfections

The main imperfections limiting the estimation precision are (i) photon loss and (ii) incomplete interference of the input photons. The main contribution to photon loss ( $\sim 50\%$  end-to-end, see Supplementary Method 3) are the inefficiencies caused by the spatial overlap between the waveguide mode and the fiber mode. The diffusion process used to produce KTP waveguides leads to non-Gaussian and asymmetric features in the waveguide spatial mode [Fig. S7(a)] which limit the fiber coupling efficiency to  $\sim 70\%$ . It may be possible to improve the spatial mode of the waveguide by optimizing the diffusion process [55] or employing ridge waveguides [56]. The quantum interference visibility ( $\sim 75\%$ ) is mainly limited by spectral mode mismatch between the signal modes as well as the spectral purity of the sources. We show the joint spectral intensities [Fig. S7(b-c)] and marginal spectra [Fig. S7(d)] of the sources measured using a time-of-flight spectrometer (resolution  $\pm 0.1\text{nm}$ ). Although the joint spectral intensities appear decorrelated and the signal spectra are well overlapped, we suspect that non-uniform spectral phase (perhaps due to pump chirp or dispersion through optical elements) may have reduced the quantum interference visibility. Moreover, the spectral purity is degraded by uncorrelated background photons ( $\sim 5\%$  of detected photons). These background photons are generated in a continuum of spectral modes and likely originate from processes where one photon from a down-converted pair is generated in an unguided waveguide mode [57], in which case their contribution could be reduced by minimizing propagation losses inside the waveguide.

### Supplementary Discussion 3: Recovering shot-noise limited performance

Using the  $\Delta = N$  probe, all photons injected into the interferometer originate from one source. As such, imperfections such as spectral purity and mode matching should not affect the performance of the probe. The  $\Delta = N$  probe is prepared by considering trials where e.g.  $(h_1, h_2) = (N, 0)$ . However, even when  $h_2 = 0$ , this second source can still inject unwanted light into the interferometer due to losses in the herald modes. This generally degrades the performance of the  $\Delta = N$  probe. Here we show that shot-noise limited performance is recovered by blocking one of the sources. In Fig. S8, we plot  $\tilde{\mathcal{F}}_{5,0}(\phi)$  for the  $\Delta = N = 5$  probe calculated without post-selection. We performed the measurement with a single source blocked and with both sources unblocked. In the latter case, we find that  $\tilde{\mathcal{F}}_{5,0}(\phi)$  reaches  $0.991 \pm 0.001$  at its highest point, demonstrating shot-noise limited performance. Ideally,  $\tilde{\mathcal{F}}_{5,0}(\phi)$  should be flat with  $\phi$ . However, experimental imperfections such as imbalanced detector efficiency, detector dark counts ( $\sim 1\%$ ), and imperfect interferometer visibility cause the dips in  $\tilde{\mathcal{F}}_{5,0}(0)$  and  $\tilde{\mathcal{F}}_{5,0}(\pm\pi)$  where the photons should ideally always

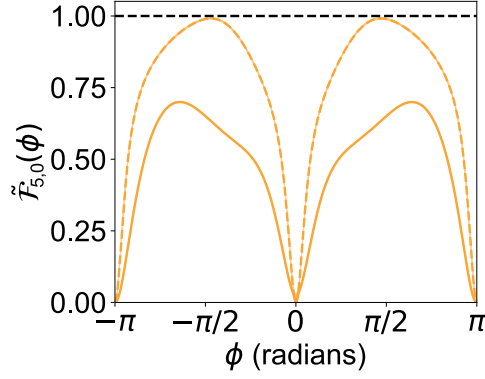


FIG. S8.  $\tilde{\mathcal{F}}_{5,0}(\phi)$  calculated without post-selection. Bold yellow line shows result with both sources unblocked, whereas dashed yellow line is result with one of the sources blocked. The thicknesses of the yellow lines show  $1\sigma$  confidence intervals obtained by fitting 50 simulated experiments. Dashed black line is the shot-noise limit.

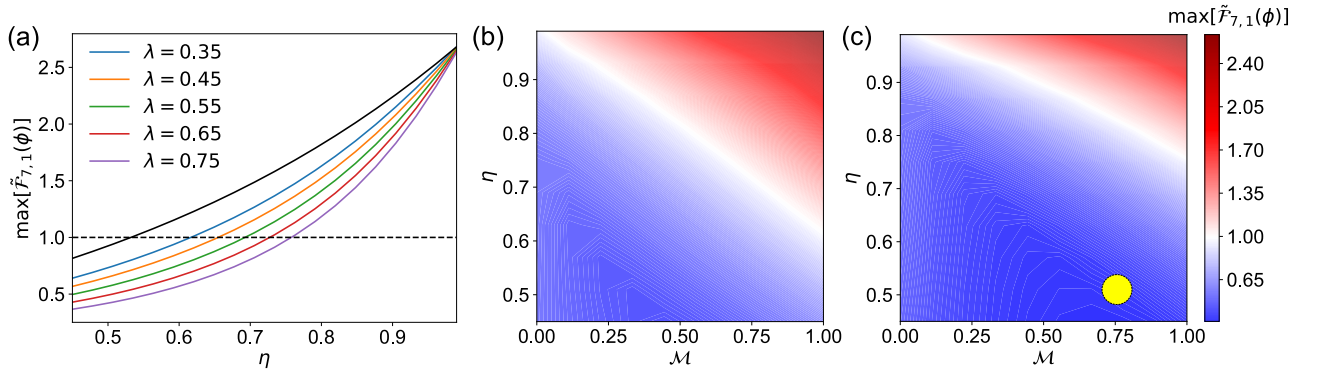


FIG. S9. (a) Effect of squeezing strength on imperfect heralding. We plot  $\max[\tilde{\mathcal{F}}_{7,1}(\phi)]$  as a function of the efficiency  $\eta$  (equal in all four modes) for various PDC gain parameters  $\lambda$  and assuming  $\mathcal{M} = 1$ . Black line shows the perfect heralding case ( $\eta_{h_1} = \eta_{h_2} = 1$ ). Dashed line shows the shot-noise limit. (b) and (c) plots  $\max[\tilde{\mathcal{F}}_{7,1}(\phi)]$  as a function of  $\eta$  and  $\mathcal{M}$  for  $\lambda = 0.35$  and  $\lambda = 0.75$ , respectively. Blue (red) indicates a parameter regime where expected performance is below (above) the shot-noise limit. The yellow circle in (c) shows roughly the parameter regime achieved in the experiment.

exit the interferometer from one port.

#### Supplementary Discussion 4: Parameters required to surpass shot-noise limit without post-selection

Here we provide an analysis on estimating the efficiency and quality of the two-mode vacuum sources required to surpass the shot-noise limit without post-selection. We focus on the  $\Delta = 6$  [i.e.  $(h_1, h_2) = (7, 1)$ ] probe as this is the most loss-tolerant  $N = 8$  probe that can surpass the shot-noise limit in our scheme. For simplicity, we assume equal efficiency  $\eta$  in all four modes of the experiment ( $\eta = \eta_{h_1} = \eta_{h_2} = \eta_{s_1} = \eta_{s_2}$ ) and equal PDC gain parameters  $\lambda$ . There are three main experimental parameters to consider: (i) the efficiency  $\eta$ , (ii) the distinguishability  $\mathcal{M}$  of photons between the top and bottom sources, (iii) the PDC gain  $\lambda$ . Given these parameters, we estimate the sensitivity of the probe by calculating the classical Fisher information  $\tilde{\mathcal{F}}_{7,1}(\phi)$  (per detected signal photon). Since this quantity generally depends on  $\phi$ , we focus on region in phase with the largest possible sensitivity, i.e.  $\max[\tilde{\mathcal{F}}_{7,1}(\phi)]$ .

We begin by focusing on the effect of the PDC gain  $\lambda$  and assume  $\mathcal{M} = 1$  for now. As shown in Fig. S9(a), a smaller  $\lambda$  provides a larger  $\max[\tilde{\mathcal{F}}_{7,1}(\phi)]$ . This is because lowering  $\lambda$  increases the photon-number purity of the heralded probe in the presence of loss in the heralding arms, i.e. it reduces the probability that the herald detectors under-counted the true number of photon pairs produced by the sources. As a reference, we include the perfect heralding case which is shown by the black line. While reducing  $\lambda$  minimizes the detrimental effects of imperfect heralding, it also drastically decreases the heralding rate. For example, assuming  $\eta = 0.5$  and a 100 kHz laser repetition rate,  $\lambda = 0.75$  would produce a  $N = 8$  probe roughly once per second whereas  $\lambda = 0.35$  would produce such a probe only about once per



day.

Next we consider the combined effect of imperfect distinguishability and efficiency. In Fig. S9(b) [(c)], we plot  $\max[\tilde{\mathcal{F}}_{7,1}(\phi)]$  as a function of  $\eta$  and  $\mathcal{M}$  for  $\lambda = 0.75$  [ $\lambda = 0.35$ ]. The approximate region achieved in our experiment is shown in yellow. Improvements in both  $\eta$  and  $\mathcal{M}$  are necessary to unconditionally surpass the shot-noise limit. As a reference point, a distinguishability of  $\mathcal{M} \sim 0.85$  was achieved in Ref. [38] using the same type of high-gain PDC sources as used in our experiment. With such a distinguishability, the efficiency would need to be improved to  $\sim 80\%$  [ $\sim 70\%$ ] when using  $\lambda = 0.75$  [ $\lambda = 0.35$ ].