Sample-efficient benchmarking of multiphoton interference on a boson sampler in the sparse regime

Jelmer J. Renema,[1,*] Hui Wang,[2,3] Jian Qin,[2,3] Xiang You,[2,3] Chaoyang Lu,[2,3] and Jianwei Pan[2,3]

[1]*Mesa+ Institute for Nanotechnology, University of Twente, PO Box 217, 7500 AE Enschede, Netherlands*
[2]*Hefei National Laboratory for Physical Sciences at Microscale and Department of Modern Physics,*
University of Science and Technology of China, Hefei, Anhui 230026, China
[3]*CAS Centre for Excellence and Synergetic Innovation Centre in Quantum Information and Quantum Physics,*
University of Science and Technology of China, Hefei, Anhui 230026, China

Verification of a quantum advantage in the presence of noise is a key open problem in the study of near-term quantum devices. In this work, we show how to assess the quality of photonic interference in a linear optical quantum device (boson sampler) by using a maximum likelihood method to measure the strength at which various noise sources are present in the experiment. This allows us to use a sparse set of samples to test whether a given boson-sampling experiment meets known upper bounds on the level of noise permissible to demonstrate a quantum advantage. Furthermore, this method allows us to monitor the evolution of noise in real time, creating a valuable diagnostic tool. Finally, we observe that sources of noise in the experiment compound, meaning that the observed value of the mutual photon indistinguishability, which is the main imperfection in our study, is an effective value taking into account all sources of error in the experiment.

I. INTRODUCTION

With the increasing computational power of quantum information processing devices [1–3], a pressing problem is how to demonstrate the computational advantage of a quantum device over a classical one (also known as "quantum supremacy"). The key issue is *verification*, the problem of checking that the quantum device is truly performing its task. Ideally, a quantum advantage demonstration would be arranged so that verification can be done efficiently in the size of the quantum system. For example, an efficiently verifiable quantum advantage can be demonstrated on a universal, fault-tolerant computer by a fast solution to a problem such as prime factoring, i.e., one whose output can be efficiently checked. If the quantum machine can solve the posed problem much faster than a classical computer, this would constitute very strong evidence of the quantum nature of the device.

However, for the foreseeable future, the main devices of experimental interest are noisy intermediate-scale quantum (NISQ) devices, which are small-to-medium sized devices without quantum error correction. These devices usually work by solving a sampling problem over some probability distribution more efficiently than is believed to be possible classically [4,5]. In photonics, this takes the form of sampling over the output distribution of interfering photons in a linear optical network, a problem known as boson sampling (see Fig. 1) [6]. For such sampling problems, a no-go result says that no device-independent sample-efficient verification protocol exists [7].

A further problem in the demonstration of a quantum advantage using NISQ devices is the influence of noise, i.e., any unwanted physical effect which degrades the performance of the device. For NISQ devices, the physical picture is that noise pushes the distribution over which the device samples closer to the distribution associated with classical operation of the device [8]. For quantum advantage demonstrations based on random circuit sampling [1], it was recently suggested that if the noise pushes the system toward the uniform distribution, any level of noise which does not completely revert the device to sampling over a classical distribution is sufficient to maintain a quantum advantage [9].

For boson sampling, in contrast, pseudoprobability distributions are known which approximate the sampling distribution in the presence of noise [10–15], and from which it is believed to be possible to sample efficiently. The distance between these distributions and the output distribution of the device depends on the level of noise, and for sufficiently high levels of noise the two distributions are close, ruling out a quantum advantage in that regime.

However, the application of such classicality thresholds requires knowledge of which noise sources are present in a boson sampler, and at what strength. The main sources of noise in a boson sampler are photon loss [16,17] and partial indistinguishability [18]. So far, the level of noise in boson sampling experiments was mostly inferred by measuring the complete output distribution and comparing it to the ideal (i.e., noiseless) output distribution, where the experiment is declared a success if the distance between these is sufficiently low. However, this approach is inherently nonscalable in the number of measurements required [6] since the probability of observing any single outcome decreases exponentially with the number of photons. The other approach was to compare the hypothesis of fully distinguishable photons with that of fully indistinguishable ones [19], which does not capture the

full range of possible effects in the device [18]. This issue recently became pressing with the demonstration of the first boson sampler operating in the sparse regime [3], i.e., where only a sparse set of samples can be collected due to the size of the Hilbert space. With "sparse" we mean a set of samples which is of insufficient size to reconstruct the complete probability distribution to fixed accuracy. Since the number of samples required to reconstruct the distribution grows exponentially with the number of photons, the ability to infer photon indistinguishabilities from a sparse set of samples is a strongly desired feature.

In this work, we show that all information required to assess the quality of quantum interference in a boson sampler is contained in just a sparse set of samples. We use a maximum likelihood method to infer the relevant properties of the photons, in particular their mutual indistinguishability. We apply this method to samples produced on the experimental apparatus reported in [3]. There is a discrepancy between indistinguishability as inferred from the list of samples and that inferred from a direct, independent measurement of the indistinguishability. We explain this by demonstrating that all other imperfections result in a decrease of the indistinguishability. The value inferred from the samples must therefore be interpreted as an effective value taking into account many sources of noise. Using this fact, we can construct an error budget for a boson sampler. We also show how to use this method to monitor the quality of photonic interference in real time. Our method is an extension of Hong-Ou-Mandel interference [20] to measure the mutual overlap of photons in a highly complex multiphoton, multimode case. We show that our method is efficient in the number of samples required to produce an estimate of the indistinguishability. It is, however, not efficient in the computational resources required to obtain this estimate.

Together with the results on the permissible level of noise in a boson sampler [10–15], these results enable testing whether a given boson sampling experiment passes the known requirements to exhibit a quantum advantage.

Our work is structured as follows: We begin with an overview of the experimental apparatus. We then detail our maximum likelihood analysis method. Finally, we conduct numerical simulations to show that the effect of other imperfections is to decrease the degree of indistinguishability.

II. EXPERIMENTAL IMPERFECTIONS

The experimental apparatus on which our data set was generated consists of an InAs/GaAs quantum-dot source operating at 893 nm pumped with a pulsed laser with 1.2 nW intensity and a repetition rate of 76 MHz, corresponding to a pulse energy of 1.6×10^{-17} J. This single-photon source is then demultiplexed using a tree of 19 Pockels cells, and fed into a fixed free-space interferometer with 60 optical modes, where quantum interference occurs. Finally, detection is done by a bank of superconducting single-photon detectors which are fiber-coupled to these modes. Samples consist of the set of detectors which are triggered in each run of the experiment, and are recorded using standard correlation electronics. More details of the setup are provided in [3].

The data set under consideration here consists of lists of samples from this device. For simplicity of the analysis, postselection was used to focus on those samples where the number of photons incident is equal to the number of detection events reported, thereby removing the effect of photon loss from consideration. The number of detected photons varied between $n = 3$ and $n = 7$, with the length of the list of samples varying from 161 at $n = 6$ to 3×10^4 at $n = 3$. Besides the samples, the data set also contains the independently characterized transmission matrix M of the interferometer. A further restriction on the list of samples arises from the threshold nature of the detectors, namely that all samples must be *collision-free* in order to be registered; i.e., all n photons must emerge from distinct output modes.

This experimental setup contains four imperfections of note, which result in samples which do not correspond to ideal n-photon quantum transmission in the interferometer. First, the generated photons are not perfectly indistinguishable. Second, the input states occasionally contain additional noise photons. This results in samples where one single photon is lost, but where this is compensated by one of the modes containing two photons instead, preserving the total photon number. Third, the detectors in the system produce dark counts, resulting in events where one photon is lost and replaced by a dark count. Finally, there is uncertainty in measuring the transmission matrix M, resulting in differences between the actual and expected interference patterns.

III. EXPERIMENTAL RESULTS

In this section, we will measure the decree of imperfections present in the experiment. We will model all imperfections as decreasing the level of indistinguishability. In the next section, we will show a numerical model that demonstrates that this approximation is appropriate.

To measure the level of partial indistinguishability, we use a standard maximum likelihood approach. The intuition behind the maximum likelihood method is that if we have some probability distribution p, which is a function of a set of parameters θ, we can estimate θ from a list of samples by the following expression: $\theta_{ml} = \mathrm{argmax}_\theta[\Lambda(\theta)]$, with $\Lambda(\theta) = \prod_i p_i(\theta)$; i.e., we must maximize with respect to θ the total likelihood to find the observed series of samples.

In a boson sampler operating at some level of partial indistinguishability, the probability of a given outcome can be written as [21–23]

$$P(M, S) = \sum_{\sigma \in S_n} \left(\prod_j S_{j,\sigma_j} \right) \mathrm{Perm}\big(M \circ M_{\sigma,1}^\dagger\big), \qquad (1)$$

where S is a matrix of indistinguishabilities defined elementwise as $S_{ij} = \langle x_i | x_j \rangle$, $|x\rangle_i$ is the internal wave function of the ith photon, Perm is the permanent function, σ is a permutation of size n, indices on matrices denote permutation according to those indices, M is the submatrix of U connecting the modes containing an input photon to the outputs of interest [24], and $\circ$ is the elementwise product.

It is this matrix S which we are interested in estimating. Throughout this work, we will parametrize S elementwise as $S_{ij} = x + (1 - x)\delta_{ij}$; i.e., we assume that all pairs of photons
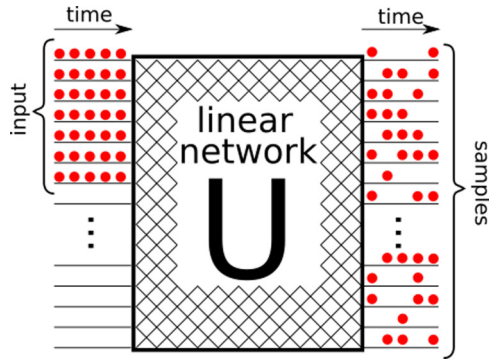
FIG. 1. A sketch of a boson sampler. A series of n modes out of a large linear interferometer are fed with single photons, and samples from the resulting distribution are recorded at the output.

have equal overlap with each other, which we can then set to be real without loss of generality. We will show in the Appendix that it is an appropriate parametrization of S for our experimental data.

A complication arises because we are sampling over the subset of collision-free events due to the threshold nature of the detector. This means that we must assign probability 0 to events containing a collision, and increase the probability of noncollision events by a factor C, where $C(x) = 1/\sum_i P_i(x)$ is the normalization factor obtained by summing all noncollision events. The total probability of obtaining a noncollision event depends on the level of indistinguishability, meaning that $C(x)$ must form part of the estimation. We estimate $C(x)$ using a Monte Carlo procedure, sampling uniformly over 10^4 possible output modes.

Figure 2 illustrates the intuition behind the maximum likelihood approach for boson sampling. The main figure shows ten arbitrarily chosen samples from our experiment, for $n = 7$, as a function of the level of mutual indistinguishability x. By noting which samples we have observed, we can obtain information about the level of indistinguishability at which
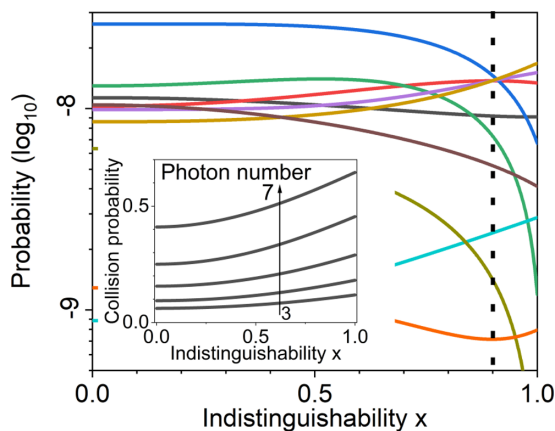

FIG. 2. The probability of observing 10 arbitrarily chosen output configurations (i.e., sets of photon detection events) as a function of partial photon indistinguishability x. Note the nonmonotonic behavior in some of these curves. Inset: Monte Carlo estimate of the correction factor arising from the restriction of our data set to collisionless samples.
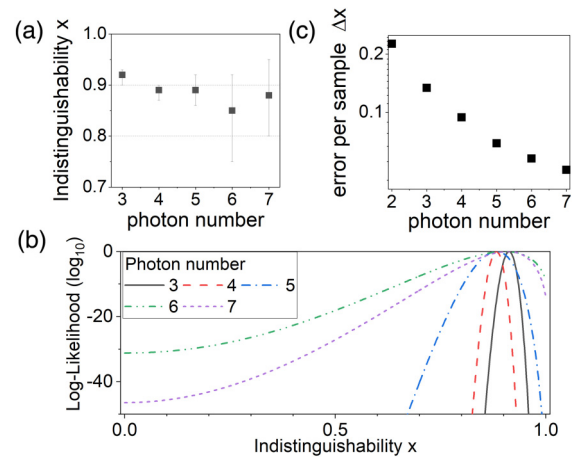

FIG. 3. (a) Maximum likelihood estimation of the partial indistinguishability $x = \langle \psi_i | \psi_j \rangle$, from a sparse series of samples. (b) Log-likelihood functions plotted between $0 \leqslant x \leqslant 1$. The colors indicate the number of photons. (c) Numerical simulation of the accuracy of our experiment, normalized to a constant number of samples.

our sampler is operating. The inset shows our Monte Carlo estimate of the fraction of collision events [i.e., $1 - 1/C(x)$]. Note that the probability of a collision event goes up strongly as a function of photon number.

Figure 3(a) shows the results of our estimation. We compute that across all our experiments, our photons have a mean wave function overlap of $x = 0.89 \pm 0.02$, as shown by the black points in Fig. 3(a), corresponding to a HOM dip depth of $x^2 = 0.79$. The error bars in Fig. 3(a) are given by the point where the relative likelihood is smaller than 0.05. There is no statistically significant trend in our measured indistinguishabilities as a function of photon number. Figure 3(b) shows the likelihood functions evaluated from $0 < x < 1$. Note that for all n, $\Lambda(x = 1) \gg \Lambda(x = 0)$, which explains why the effect of partial indistinguishability was not detected by previous tests, which compared only fully distinguishable and fully indistinguishable photons [3], or combinations of such [19], as was noted previously by [25].

Figure 3(c) shows an estimate of the accuracy of our method, as a function of photon number. The variation in the size of the error bars in Fig. 3(a) is mainly attributable to the vastly differing numbers of samples. To get an estimate of the efficiency per sample, we numerically simulate processing 10 000 samples of each photon number at $x = 1$, and compute the resulting uncertainty in estimating x by looking at the relative likelihood. We find that the error of our method decreases with the number of photons. The intuition behind this is that since the indistinguishability is a polynomial of degree n in x [10], increasing the number of photons increases the changes in probability around the maximum of the likelihood function.

IV. MODELING THE EXPERIMENT

Figure 3 raises a question: If the wave function overlap between our photons is approximately $x = 0.89 \pm 0.02$, why does an independent measurement of the overlap between our photons via the Hong-Ou-Mandel (HOM) effect measure

TABLE I. Error budget for a boson sampler. The first column shows the value of the indistinguishability when that particular imperfection and the ones above it in the table are applied. The second column shows the relative likelihood of the corresponding value of x as derived from the simulation, normalized to the maximum value of the likelihood function derived from the experimental data.

Imperfection	Value of x inferred	$\log_{10}(\Lambda_s/\Lambda_e)$
x from HOM experiment	0.981 ± 0.002	-130
Misspecification of M	0.970 ± 0.008	-83
Multiphoton states	0.924 ± 0.009	-2.5
Dark counts	0.913 ± 0.010	0

$x = 0.981$ [26]? The answer to this question is that imperfections in boson sampling compound one another, resulting in an effective reduction of the photon indistinguishability. This was shown theoretically for indistinguishability and loss [11], as well as for indistinguishability and noise on M [12]. We must look to the other noise sources in our experiment to explain the discrepancy between our inferred x and that observed from HOM measurements, which are themselves not affected by the other imperfections mentioned in Sec. II.

To demonstrate this physical mechanism, we have performed a full numerical simulation of our experiment, for $n = 3$. We generate a set of samples which contains all imperfections, at the strengths at which we have independently measured them to be present in the experiment [27]: a 3% probability of a dark count, a 1.2% probability of double photon emission per mode, and approximately 1% error on measuring the elements of M. We then analyze these samples using the same procedure which we used to analyze the actual experimental data, and report on the observed value of x. We account for dark counts by adding some samples with $n = 2$ among all three possible combinations of pairs of input modes, and generating the third detection event according to the measured dark count rate of our detectors. We account for multiphoton emission by generating samples from all nine combinations of states containing two of the wanted photons and one noise photon. We account for misspecification on the interferometer by perturbing our observed interferometer M with elementwise Gaussian noise.

Table I shows the impact of these noise sources. By performing a series of simulations where we switch these imperfections on one at a time, and reconstructing x, we can get an idea of the cumulative impact which each of these imperfections has on our boson sampler. Table I therefore constitutes an error budget for our boson sampler. In particular, it shows that the effect of multiphoton states is almost a factor 4 larger than that of misspecification of M and of dark counts, but of the same order as the reduction of x due to true indistinguishability. This shows that the main challenges in improving the quantumness of our boson sampler lie in improving the quantum dot source.

To compare our simulation to our experimental data, we also report the likelihood of the value of x from our simulations (Λ_s) relative that of the experimental data (Λ_e). This shows that when we "switch on" all imperfections, our simulation predicts a value of x within the error bar of the one which we measure in our experiment, and hence that our simulation
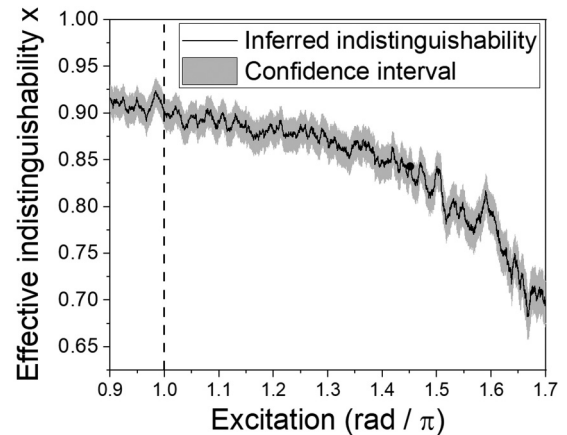

FIG. 4. Real-time monitoring of the effective degree of photon indistinguishability in a boson sampler as a function of excitation power. To decrease the effective degree of indistinguishability, the pump power on the quantum dot single photon source was increased to push the excitation away from the ideal π pulse (indicated with a dashed line). The black curve shows a rolling average estimate of x, averaging over 10 000 consecutive samples, with the gray band showing the 95% confidence interval.

and our measurement are consistent. This validates the picture of our inferred maximum likelihood indistinguishability value being essentially an effective value that takes into account all other known sources of error in the experiment.

V. REAL-TIME MONITORING

A practical application of our result is that we can monitor the strength of the noise sources in our experiment in real time. In this way, we open up a diagnostic tool for stabilizing an experimental boson sampling setup.

To illustrate this procedure, Fig. 4 shows a case in which we have artificially increased the $g^{(2)}$ of our photon source by increasing the excitation power of the pump pulse from 1.19×10^{-17} mJ to 4.58×10^{-17} mJ. To obtain the ith point in this graph, we compute the maximum likelihood estimate of the ith to $i + 10\,000$th consecutive samples from the experiment. The resulting estimate shows a clear continuous decrease in the effective photon indistinguishability. It takes approximately 0.1 ms on a standard laptop to analyze each sample, demonstrating the feasibility of using this monitoring method in real time.

VI. DISCUSSION AND CONCLUSION

An important restriction in our approach is that Eq. (1) is expensive to evaluate, since it requires the computation of $n!$ many permanents, each of which comes at a cost of $n2^n$ computational steps. This can be reduced to $2n2^{2n}$ by applying a multidimensional extension to Ryser's formula [18], but this still restricts application of this approach to approximately 25 photons [28]. Two approaches are possible: either to use the approximate probabilities of [10–15], or to use simpler circuits for the measurement of x, as was done in the benchmarking of Google's random circuit sampling experiment [1]. We leave the issue of finding the appropriate circuits for the

photonic case as an open problem for future work. Another further extension of this work would be to strengthen our belief in the thresholds for classicality by connecting them to a complexity conjecture, as was done for random circuit sampling [9].

In conclusion, we have demonstrated how to infer the quality of a boson sampler from a sparse series of samples. We have demonstrated that the measured indistinguishability is an effective value, which accounts for a series of imperfections. These results show how to demonstrate that a candidate quantum advantage demonstration using photonics outperforms the best known simulation algorithms.

ACKNOWLEDGMENTS

APPENDIX

1. Model selection

In the main text, we assert that the optimal model for assessing the effect of distinguishability is that of a single wave function overlap x, which is equal for all pairs of photons. In this section, we provide the reasoning for this claim.

To maximize our likelihood function, we must find an appropriate parametrization of S. Since S is a Gram matrix (i.e., a matrix of inner products), it can be of rank up to n, but it cannot have arbitrary form. For example, it must be positive semidefinite. Parametrizing S amounts to parametrizing the internal states of the photons $|\psi\rangle$, which in principle can live in an infinite-dimensional Hilbert space. However, since we are only interested in the overlap $\langle\psi_i|\psi_j\rangle$, for full generality it suffices to specify the coefficients of $|\psi_i\rangle$ on n basis vectors $|\beta_i\rangle$, as well as on a single "remainder" state $|\phi_i\rangle$ per photon, which is defined as $\langle\phi_i|\phi_j\rangle = \delta_{ij}$, which covers the part of the wave function not interfering with any other photon.

However, we find that parametrizing S in its full generality is impractical, since this introduces order n^2 parameters which must be estimated simultaneously. Instead, we use insight into the workings of our photon source to build a series of models of escalating complexity. We then use a standard model selection method to test which level of model complexity is necessary to explain our data.

In the simplest model, all pairs of photons have identical mutual distinguishability x, which then can be set to be real without loss of generality [10]; i.e., we set $|\psi_i\rangle = \sqrt{x}|\beta\rangle + \sqrt{1-x}|\phi_i\rangle$. This results in a matrix S of the form explored in Fig. 2. This model has a single free parameter. This model is motivated by the idea that our quantum dot source and demultiplexer are operating in a regime where all photons incident into the interferometer should have equal wave function overlap [3].

In the next simplest model, we allow the overlap between photons to vary, but we still constrain the overlaps to be real, i.e., $|\psi_i\rangle = \sqrt{x_i}|\beta\rangle + \sqrt{1-x_i}|\phi_i\rangle$. This model has n free parameters, namely the x_i, and allows for fluctuations in the states produced by our quantum dot source. Finally,

it is known that when operating at partial distinguishability, the phase of the wave function overlap can play a role as well, provided there are two orthogonal internal states in which the photons overlap [29]. To account for this possibility, we also attempt the state $|\psi_i\rangle = \sqrt{x_1}|\beta_1\rangle + \sqrt{x_2}e^{i\phi_2}|\beta_2\rangle + \sqrt{1-x_1-x_2}|\phi_i\rangle$. This model has $3n$ free parameters, namely two wave function overlaps per photon and one phase.

To distinguish which of these parametrizations is appropriate, we use the Akaike information criterion (AIC), a standard method of selecting models given some data set. The AIC asserts that when we have competing models, the model that minimizes the parameter $\text{AIC} = 2k - \log(\Lambda)$ is the one to be preferred, where Λ is the maximum value of the likelihood function associated with that model, and k is the number of free parameters. The AIC balances in a meaningful way between model complexity and goodness of fit, and protects against overfitting.

For each observed model, we compute the likelihood function belonging to that parametrization, and optimize numerically using standard Matlab routines. We then use the value of the likelihood function at the maximum and the number of parameters to evaluate which model is preferred. We ignore the effect of only observing collision-free samples, for the moment, due to the difficulty in estimating $C(S)$ for arbitrary S with sufficient accuracy to perform numerical optimization.

For $n = 3$ and $n = 4$, we find that the optimal model is one with arbitrary wave function overlaps, but no phases. For $n \geqslant 4$, we find that the optimal model is one with a single wave function overlap. The model with phases is never strongly preferred. This last fact is physically understandable, as arranging for such phases to enter the problem requires the wave function of the photons to overlap substantially in two orthogonal internal states, while in a quantum interference experiment such as this much care has been put into making sure the photons all overlap along a single internal state. Furthermore, as all photons come from the same quantum dot, it is perhaps no surprise that their degree of overlap is approximately equal for all pairs.

We proceed with the parametrization of X by a single wave function overlap x for all n, for the following two reasons. First, the differences in wave function overlap predicted by the second model are small (approximately 3%), and while the difference is statistically significant, the effect of these variations on the complexity is slight. Second, for reasons of computation time and numerical stability we can only estimate $C(S)$ in the simplest parametrization, and for $n \geqslant 4$ the effect of $C(S)$ is highly significant in estimating S. Therefore, we proceed by computing $C(x)$ in the single-overlap parametrization by Monte Carlo simulation and maximizing the likelihood function as a function of x.

2. Error budget of the experiment

In this subsection, we describe in more detail the simulation behind the error budget (Table I) in the main text.

We identify the following noise sources in the experiment: finite accuracy to which the transmission matrix M can be measured (i.e., a static version of Kalai-Kindler noise [30]), dark counts, and multiphoton emissions [i.e., nonzero $g^{(2)}(0)$].

To assess the effect of these imperfections, we performed a series of numerical simulations of the entire experiment and data processing. These simulations consist of two steps: first, we generate data simulated sets of samples, to mimic the experiment and its imperfections. For this, we use a Markov chain Monte Carlo sampler, in the style of Neville *et al.* [31], with a burn-in of 100 samples and a thinning factor of 100 samples. The sampler generates 3×10^4 samples (after thinning) per numerical simulation, and each simulation was performed 100 times to ensure statistical accuracy. The sampler takes into account the fact that the experiment cannot generate collision samples, and we include a numerical estimate of $C(x)$ in our data processing. In this sampler, we incorporate the imperfections as detailed below. Then, we process the data using the same code we used to process the experimental data, and compare the values of x which we obtain.

Below, we will detail the main sources of imperfections, how we include them in the sample generation. Since previously the effect of g^2 was not taken into account separately, the value of x inferred from the HOM experiment has to be corrected for finite g^2 as well. Note that we simulate all imperfections only to first order; i.e., we assume no double imperfections, or mixtures of imperfections (e.g., two simultaneous dark counts). Since the probability of all our imperfections is at the 10^{-2} level, the second-order correction is at least an order of magnitude lower than that, and hence this is justified.

In what follows, we will focus on the $n = 3$ experiment. This has a few reasons. First, this is computationally the least intensive data set to work with. Second, it is the data set that suffers least from the problem that the experiment cannot register collision-carrying samples, which might otherwise distort the data. Third, this is the largest data set, which means that the statistical error on estimating x is the lowest. This data set has $x = 0.913^{+0.008}_{-0.009}$ using a 95% confidence level.

a. Simulation of mischaracterization of M

To estimate the accuracy to which M was measured, the following procedure was carried out: from the experimentally measured matrix M, we compute the product $M^\dagger M$, which should be the identity. In practice, measurement errors mean that it is equal to $M^\dagger M = I + \delta$, where δ is a matrix with nonzero off-diagonal elements, where the mean value of the mod-square of the elements of δ is $\langle |\delta_{ij}|^2 \rangle \approx 0.02$. We model the measurement noise on M as multiplicative i.i.d. Gaussian, i.e., $M_{\text{exp}} = M_{\text{true}}(I + \epsilon)$, where ϵ is an i.i.d. Gaussian matrix. Solving the equations gives that $\langle |\epsilon|^2 \rangle = 0.01$; i.e., 1% intensity noise on M produces the observed 2% intensity deviations in $M^\dagger M$.

We then simulate the effect of mischaracterization of M by generating our samples with M_{true}, and analyzing our samples with M_{exp} We also perform the estimate of the correction factor $C(x)$, which accounts for the lack of collision-carrying samples, using M_{exp}.

b. Estimate of dark count rate

To estimate the dark count rate, the following procedure was carried out: a set of dark count measurements with the same integration time as the 3-photon experiment was used. Summing over all detectors gives a total number of counts of $\sim 9 \times 10^7$. The total number of dark counts is approximately 9×10^5, leading to an estimate of the dark count probability of approximately 1%. This means that in a 3-photon experiment, approximately 3% of all samples contain a dark count (since the probability of 2 simultaneous dark counts is order 10^{-4}, this is negligible).

We model the effect of dark counts by replacing 900 of our 3×10^4 samples with samples where we generate 2-photon samples (with the same value of x as for the 3-photon samples) and add a uniformly randomly chosen third mode to represent the dark count. The 900 samples are made up of 300 samples from each of the 3 possible ways a photon can be lost (i.e., numbering the input modes 1 to 3, $s_{\text{in}} = \{[1, 2], [1, 3], [2, 3]\}$). This simulation makes the approximations that both the dark count rate of the detector and the loss in each channel are uniform.

c. Estimate of g^2

The g^2 rate was measured to be $g^2(0) = 0.023$. We assume that this is entirely due to 2-photon contributions in the state. Inverting the definition of g^2 shows that in 1.2% of the cases, when prompted for a single photon, instead, two will be present. This is due to pump light of the laser leaking through into the experiment; we therefore model this photon as being completely distinguishable from the other photons in the experiment. Elementary combinatorics shows that there are 9 ways in which the additional photon can replace one of the photons in the experiment. We simulate the effect of this error by generating samples from all these 9 possible configurations. These samples are then used to replace the same number of samples from the "good" data set, to simulate this imperfection.

d. Correction of distinguishability for finite g^2

To correct the raw value of $x = 0.961$ for finite g^2, we calculated the HOM dip depth as a function of the real photon indistinguishability, taking into account the effect of multiphoton states. We compute the output probability of a coincidence count for a 3-photon $s_{\text{in}} = [1, 1, 2]$ state, where the second photon is distinguishable from the others, which together with the usual expressions for the 2-photon case gives us the probability of a coincidence detection event on and outside of the HOM dip (i.e., when $x = x_{\text{corr}}$ and when $x = 0$, respectively). Then, we invert the equation for the HOM dip ($D_{\text{HOM}} = 1 - P_{x_{\text{true}}}/P_{x=0}$). Solving this equation gives $x = 0.981$.

[1] F. Arute *et al.*, Nature (London) **574**, 505 (2019).

[2] J. M. Pino *et al.*, arXiv:2003.01293.

[3] H. Wang *et al.*, Phys. Rev. Lett. **123**, 250503 (2019).

[4] J. Preskill, Quantum **2**, 79 (2018).

[5] A. Harrow and A. Montanaro, Nature (London) **549**, 203 (2017).

[6] S. Aaronson and A. Arkhipov, Theory Comput. **9**, 143 (2013).

[7] D. Hangleiter, M. Kliesch, J. Eisert, and C. Gogolin, Phys. Rev. Lett. **122**, 210502 (2019).

[8] S. Boixo *et al.*, Nat. Phys. **14**, 595 (2018).

[9] S. Aaronson and S. Gunn, Theory Comput. **16**, 1 (2020).

[10] J. J. Renema *et al.*, Phys. Rev. Lett. **120**, 220502 (2018).

[11] J. Renema, V. Shchesnovich, and R. Garcia-Patron, arXiv:1809.01953.

[12] V. S. Shchesnovich, Phys. Rev. A **100**, 012340 (2019).

[13] V. Shchesnovich, Int. J. Quantum Inf. **18**, 2050044 (2020).

[14] A. E. Moylett, R. García-Patrón, J. J. Renema, and P. S. Turner, Quantum Sci. Technol. **5**, 015001 (2019).

[15] J. J. Renema, Phys. Rev. A **101**, 063840 (2020).

[16] S. Aaronson, Introducing some British people to P vs. NP (Comment 84), https://www.scottaaronson.com/blog/?p=2408.

[17] While photon loss can be removed by postselecting on those events where all photons make it through the optical system, this comes at the cost of an exponential overhead in the time required to produce a sample.

[18] M. C. Tichy, Phys. Rev. A **91**, 022316 (2015).

[19] M. Bentivegna *et al.*, Int. J. Quantum Inf. **12**, 1560028 (2014).

[20] C. K. Hong, Z. Y. Ou, and L. Mandel, Phys. Rev. Lett. **59**, 2044 (1987).

[21] V. S. Shchesnovich, Phys. Rev. A **89**, 022333 (2014).

[22] V. S. Shchesnovich, Phys. Rev. A **91**, 063842 (2015).

[23] V. S. Shchesnovich, Phys. Rev. A **91**, 013844 (2015).

[24] S. Scheel and S. Buhmann, Acta Physica Slovaca: Reviews and Tutorials **58**, 675 (2008).

[25] Z. Dai *et al.*, Sci. China: Phys., Mech. Astron. **63**, 250311 (2020).

[26] H. Wang *et al.*, Nat. Photonics **11**, 361 (2017).

[27] The details of these experiments are given in the Appendix.

[28] J. Wu *et al.*, Nat. Sci. Rev. **5**, 715 (2018).

[29] A. Menssen *et al.*, Phys. Rev. Lett. **118**, 153603 (2017).

[30] G. Kalai and G. Kindler, arXiv:1409.3093.

[31] A. Neville *et al.*, Nat. Phys. **13**, 1153 (2017).