

Gaussian Boson Sampling with Pseudo-Photon-Number-Resolving Detectors and Quantum Computational Advantage

Yu-Hao Deng,^{1,2,3} Yi-Chao Gu,^{1,2,3} Hua-Liang Liu,^{1,2,3} Si-Qiu Gong,^{1,2,3} Hao Su,^{1,2,3} Zhi-Jiong Zhang,^{1,2,3}
 Hao-Yang Tang,^{1,2,3} Meng-Hao Jia,^{1,2,3} Jia-Min Xu,^{1,2,3} Ming-Cheng Chen,^{1,2,3} Jian Qin,^{1,2,3} Li-Chao Peng,^{1,2,3}
 Jiarong Yan,^{1,2,3} Yi Hu,^{1,2,3} Jia Huang,⁴ Hao Li,⁴ Yuxuan Li,⁵ Yaojian Chen,⁵ Xiao Jiang,^{1,2,3} Lin Gan,⁵ Guangwen Yang,⁵
 Lixing You,⁴ Li Li,^{1,2,3} Han-Sen Zhong,^{1,2,3} Hui Wang,^{1,2,3} Nai-Le Liu,^{1,2,3} Jelmer J. Renema,⁶
 Chao-Yang Lu^{1,2,3,7} and Jian-Wei Pan^{1,2,3}

¹*Hefei National Laboratory for Physical Sciences at Microscale and School of Physical Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China*

²*CAS Centre for Excellence and Synergetic Innovation Centre in Quantum Information and Quantum Physics, University of Science and Technology of China, Shanghai 201315, China*

³*Hefei National Laboratory, University of Science and Technology of China, Hefei 230088, China*

⁴*State Key Laboratory of Functional Materials for Informatics, Shanghai Institute of Micro system and Information Technology (SIMIT), Chinese Academy of Sciences, 865 Changning Road, Shanghai 200050, China*

⁵*Department of Computer Science and Technology and Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China*

⁶*Adaptive Quantum Optics Group, Mesa+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, Netherlands*

⁷*New Cornerstone Science Laboratory, Shenzhen 518054, China*



(Received 10 April 2023; revised 23 July 2023; accepted 1 September 2023; published 10 October 2023)

We report new Gaussian boson sampling experiments with pseudo-photon-number-resolving detection, which register up to 255 photon-click events. We consider partial photon distinguishability and develop a more complete model for the characterization of the noisy Gaussian boson sampling. In the quantum computational advantage regime, we use Bayesian tests and correlation function analysis to validate the samples against all current classical spoofing mockups. Estimating with the best classical algorithms to date, generating a single ideal sample from the same distribution on the supercomputer Frontier would take ~ 600 yr using exact methods, whereas our quantum computer, Jiǔzhāng 3.0, takes only $1.27 \mu\text{s}$ to produce a sample. Generating the hardest sample from the experiment using an exact algorithm would take Frontier $\sim 3.1 \times 10^{10}$ yr.

DOI: [10.1103/PhysRevLett.131.150601](https://doi.org/10.1103/PhysRevLett.131.150601)

Quantum computational advantage (QCA) [1–5] marks an important milestone in the development of quantum computers. By solving certain quantum sampling problems [6,7] that are intractable for classical supercomputers, QCA experiments [8–12] have provided strong evidence for the long-anticipated quantum speed-up theoretically conceived ~ 40 yr ago. Similar to Bell tests [13] which were designed to refute the local hidden variable theories [14], these QCA experiments offered evidence of the violations of the extended Church-Turing thesis [3].

Interestingly, these QCA experiments have motivated growing study of faster classical simulations. For boson sampling, these efforts can be divided into three approaches. The first approach is reducing the classical simulation overhead of exactly simulating an ideal implementation of the Gaussian boson sampling protocol [15–17]. In this case, one expects that the exponential gap

between the classical and the quantum persists, but it is possible to bring down the overhead. The leading result is Ref. [17] which uses photon collisions to achieve better exact classical simulation that could reduce the computational overhead by orders of magnitudes lower than brute-force algorithms. The second approach is using imperfections in the experimental setup [18–20], such as photon loss and partial photon distinguishability, to come to a faster simulation of the experimental data produced by an imperfect experiment. In this case, one can hope for more efficient classical simulation if the level of imperfections or experimental noise becomes strong enough. The final approach is spoofed distributions [17,21–23], i.e., distributions which are not constructed to be close in variational distance to the ideal boson sampling distribution, but which are designed to reproduce some statistical aspect of the experiment. In particular, the recently proposed more

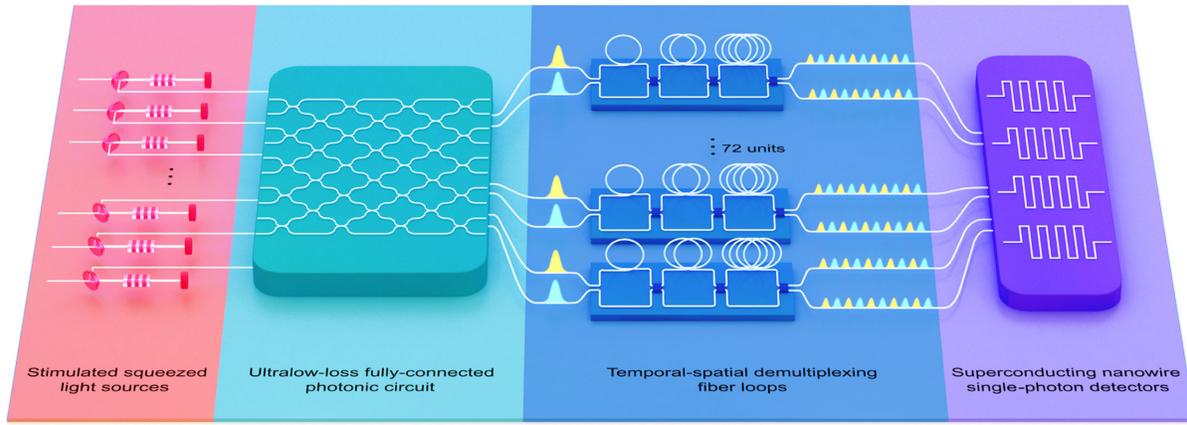


FIG. 1. The experimental setup. 25 stimulated two-mode squeezed state photon sources are all phase locked to each other and sent into a 144-mode ultralow-loss fully-connected optical interferometer. The photons go through 72 units of fiber loop setups for temporal-spatial demultiplexing and are detected by 144 superconducting nanowire single-photon detectors, which together constitute the pseudo-photon-number resolving detection scheme. Each fiber loop setup includes two input modes as represented by distinct colors. Photons from each mode are temporally demultiplexed by the fiber beam splitters and the delay lines into four time bins, and each time bin is furthermore split into two path bins at the last fiber beam splitter. The photons corresponding to each of the two input modes of the same fiber loop setup can be distinguished by their parity in time bin with a coincidence event analyzer (not shown).

competitive squashed state mockup [12] exhibits better agreement with the ground truth than the thermal state mockup. Moreover, the treewidth sampler [22], designed to spoof experiments with restricted circuit connectivity, and the independent pairs and singles (IPS) sampler [17], designed to emulate experiments of limited quantum interference, have not been reported to be ruled out in previous Gaussian boson sampling (GBS) QCA experiments.

In return, these new challenges from the classical counterparts motivate the development of higher-fidelity and larger-scale quantum computers, new methods for validation and characterization [24–27], as well as better modeling and understanding of the increasingly complex system [28–36], which is a fundamental endeavor in its own right [37–39]. Indeed, only by such a continuous quantum-classical competition can the QCA milestone be progressively better established.

In this direction, we report in this Letter a new, higher-efficiency GBS [40,41] machine with up to 255 photon clicks in the output using pseudo-photon-number-resolving detectors (PPNRDs), which overcomes both the most powerful exact classical simulation algorithms and all known approximate algorithms and spoofing algorithms. The PPNRD scheme overcomes the shortcoming of threshold detection that it is incapable of resolving photon-number information, and significantly increases the exact sampling task’s computational complexity by more than 6 orders of magnitude. We validate the samples against all the emerging classical mockups, particularly including the squashed state [12], the treewidth sampler [22], and the IPS sampler [17] mentioned above. A new model that includes the partial photon indistinguishability is used to

characterize the system, and exhibits better agreement with the experiment than previous modeling methods. The computational complexity of this new GBS device, Jiǔzhāng 3.0, is analyzed and a new QCA frontier is established.

The first GBS experiments [9,10] in the QCA regime used threshold detectors to register the samples. In those cases, there was a possibility of photon collision, that is, the photons can bunch at the outputs. New classical algorithms [17] could exploit photon collision to reduce the simulation overhead. A recent work [12] has reported time-bin-encoded GBS with photon-number resolved detection, in a fiber loop-based configuration similar to an earlier single-photon boson sampling experiment [42]. However, the relatively high photon loss in the fiber loops has allowed only three loops in the implementation, restricting the depth and universality of the interferometer. Additionally, the long recovery time of the transition-edge sensors could render it an unsuitable choice for high-repetition-rate experiments.

A schematic of our GBS experiment is shown in Fig. 1. Transform-limited laser pulses double pass periodically poled potassium titanyl phosphate (PPKTP) crystals to create 25 pairs of two-mode squeezed states (TMSSs) by a stimulated emission process [10]. The TMSSs have an average coupling efficiency of 88.4% and photon indistinguishability of 96.2%, simultaneously. In the experiment, the laser power is tuned to generate different squeezing parameters ranging from 1.2 to 1.6. The TMSSs are then fed into an ultralow-loss three-dimensional interferometer with full connectivity among all 144 modes. The transmission rate of the interferometer is 97% for each mode, and the average wave-packet overlap inside the

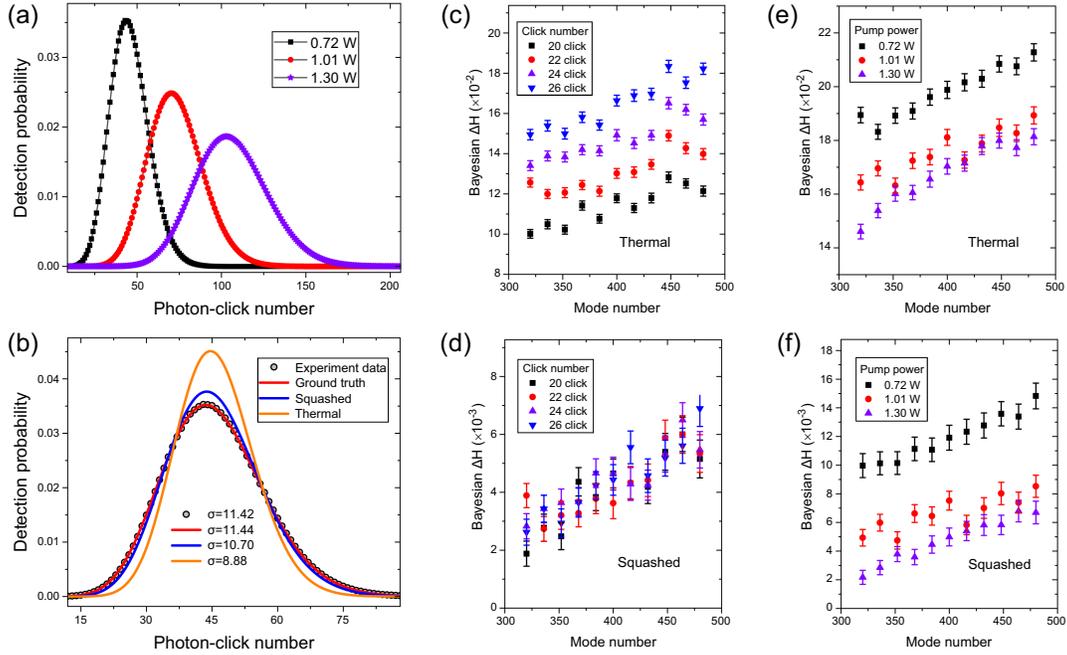


FIG. 2. The experimental photon-click number distribution and the Bayesian validation results. (a) Photon-click number distribution of this Letter. Data from experiments of three pump laser power ranging from 0.72 W and 1.30 W are displayed. A maximum photon-click number of 129, 203, and 255 are registered for each of the experiments. (b) Photon-click number distribution of the experimental results, the ground-truth theory, the squashed state and the thermal state mockups for the lowest laser intensity configuration. The plot includes labels indicating the standard deviation of the photon-click number for each distribution. It is evident that the experimental distribution agrees with the ground truth best. The error bars are too small to be displayed. (c),(d) Bayesian confidence of the ground-truth theory as a function of subsystem size, against the thermal state (c) and the squashed state (d) hypothesis in the high power (1.30 W) experiment. It can be observed that as the subsystem size grows, the Bayesian confidence is above zero and exhibits a clearly increasing trend, indicating an exclusion of the two mockups and a stronger Bayesian confidence for the full system. Bayesian results for the photon click number changing from 20 to 26 are displayed in data points of different colors and show similar results. (e),(f) Bayesian confidence of the ground-truth theory as a function of the subsystem size for experiments of the three distinct pump laser power. The results show the Bayesian results against the thermal state (e) and the squashed state (f) hypothesis, for a fixed photon-click number (25). For all pump laser powers, the Bayesian ΔH is above zero and exhibits a clearly increasing trend with subsystem size, implying stronger Bayesian confidence for the full system. In all plots, error bars indicate the standard error. The statistical fluctuation from photon-click number probability estimated with Monte-Carlo methods [28] has been included in the error bars. For results of each mode number, the Bayesian scores are obtained by averaging over an ensemble of randomly selected subsystems for unbiased benchmarking.

interferometer is above 99.5%. The whole setup is actively phase locked within a precision of 15 nm.

For detection, we implement pseudo-photon number resolution using one-to-eight demultiplexing of the optical modes (see Fig. 1). We conduct detector tomography [43] of each mode to validate the pseudo-photon-number-resolved detection (PPNRD) scheme [44]. As shown in Fig. S1(a), the probability of the nine photon-click number response of the PPNRD agrees well with the theoretical model. The positive-operator valued measures of the nine photon-click number measurement can be constructed from the data, which is shown in Fig. S1(c), together with its Wigner function and fidelity with the theoretical prediction. While we model the experiment with the PPNRD detection scheme in the following discussions, we address the question of how well our detection scheme emulates true photon number-resolving detection in [44].

The overall linear efficiency of the whole experimental setup is 43%, including the quantum light sources, transmission, and detection. This efficiency is much higher than that in Ref. [12], which was 33%, although involving only three loops. The photon-click number distribution under three laser powers is shown in Fig. 2(a). The maximum photon-click number reaches 255, which is higher than all the previous GBS experiments.

We analyze the obtained GBS samples and validate them against known classical hypotheses in the QCA regime. The most powerful method to spoof the GBS is to use classical states which can maximally approximate the experiment's quantum light sources which, under photon loss, can be gradually degraded into squeezed thermal states [18]. Thus, in this Letter, in addition to the thermal state hypothesis which has been tested in Jiǔzhāng 1.0 and 2.0, we consider the more competitive hypothesis, namely,

the squashed state [12,18]. The squashed state is a classical mixture of coherent states and has vacuum fluctuations in one quadrature and larger fluctuation in the other. It can be optimized to have the same mean photon number as the input lossy squeezed state [47], which is the convention we take in this Letter. Note that there are other plausible hypotheses such as using coherent light and distinguishable photons, which are much easier to be ruled out. As shown in [44], the validation strength against these other hypotheses is typically 2–4 orders of magnitude larger than the squashed states. Therefore, we will focus on the discussion of the latter in the main text.

We start by comparing the experimental photon number distribution with the ground-truth model and the possible mockups in Fig. 2(b). The GBS dataset taken at 0.72 W laser power (dot) well overlaps with the ground truth (red), while obviously deviates from the classical mockups based on the thermal state (yellow) and the squashed state (blue). These distributions can be quantitatively distinguished by their standard deviation as labeled on the plot, where the experiment exhibits the best agreement with the ground truth.

Then, we continue with the Bayesian test, [48], where two hypothetical theoretical models are compared against each other based on the likelihood to generate the experimental samples. We define the Bayesian test score ΔH as the difference between the ground-truth model H_0 and the classical adversary model hypothesis H_1 on a set of n -photon-click samples:

$$\Delta H = \frac{1}{N} \ln \prod_{i=1}^N \frac{P^{(0)}(\vec{s}_i)P^{(1)}(n)}{P^{(1)}(\vec{s}_i)P^{(0)}(n)} \quad (1)$$

where N is the number of samples, \vec{s}_i represents the i th sample, $P^{(0/1)}(\vec{s}_i)$ represents the event probability of sample \vec{s}_i under the $H_{0/1}$ hypothesis, and $P^{(0/1)}(n)$ is the coarse-grained probability for the n -photon click under hypothesis $H_{0/1}$. When $\Delta H > 0$, it is proven that the experimental samples are more likely from the ground-truth GBS rather than the mockup. A higher Bayesian test score indicates larger confidence.

For the validation, each of the 144 modes is treated as a one-to-eight fan-out, yielding a total mode number of 1152 for the ground truth. Further, in our model we consider the noise of the partial photon distinguishability for the Bayesian test, where the probability of the sample event is calculated by a modified version of the Torontonian [44]. We find that our model gives a closer description of the experiment than the previous ones. We use the Sunway TaihuLight supercomputer to calculate the probability of large photon number samples in the QCA regime.

The Bayesian test score of the ground truth against the mockups using the thermal state and the squashed state at 1.3W pump power are shown in Figs. 2(c) and 2(d).

Because of the classical computational overhead, we first start from a subsystem with fewer output bosonic modes, and gradually increase the subsystem size. We observe not only all the Bayesian scores in Figs. 2(c) and 2(d) are higher than zero, but also they show an evident increasing trend as the subsystem size ramps up. The positive scores demonstrate that the experimental GBS samples are more likely generated from the ground-truth distribution rather than these mockup distributions. More importantly, the rising trend of the Bayesian score indicates a higher score can be inferred for the full system. This increasing trend could be understood from that, as subsystem size increases, more complete information of the full system is incorporated into the Bayesian test, and therefore results in stronger validation strength. Therefore, we conclusively infer that for the full-mode system, though it is not directly computable, stronger Bayesian confidence for the ground-truth theory is expected.

We continue to investigate the power dependence of the Bayesian test. In Figs. 2(e) and 2(f), the Bayesian validation strength is plotted for three different pump laser powers from 0.72 W to 1.3 W. The results show that the lower pump power can generate higher Bayesian score. This is expected because the thermal noise and photon loss become increasingly sensitive with larger squeezing parameters [18].

Another important tool for characterizing the GBS is the correlation function. The k -order correlation function is recursively defined as

$$\kappa(X_1 X_2 \cdots X_k) = E(X_1 X_2 \cdots X_k) - \sum_{p \in P_k} \prod_{b \in p} \kappa[(X_i)_{i \in b}] \quad (2)$$

where X_k represents the experimental measurement operator at the k th output mode, and P_k represent all partitions of the set $\{1, 2, \dots, k\}$ excluding the universal set. Based on the correlation function, we validate our experimental samples against other recently proposed mockups based on approximating the experimental GBS through low-order marginal distributions [21].

Reference [21] proposed a greedy algorithm to sample from a distribution that approximates all first- and second-order correlations of the experiment, which we call a greedy sampler below. The same work also used the single-mode marginals and two-mode correlations to configure a Boltzmann machine for sampling under the Thouless, Anderson, and Palmer (TAP) mean field approximation. These methods can be generalized to higher orders in principle, but a full enumeration of all the marginal probabilities of the chosen order is required, which grows combinatorially, and thus limits it within the low-order regime. However, these mockups do not capture higher-order correlations, and can thus be ruled out from this aspect.

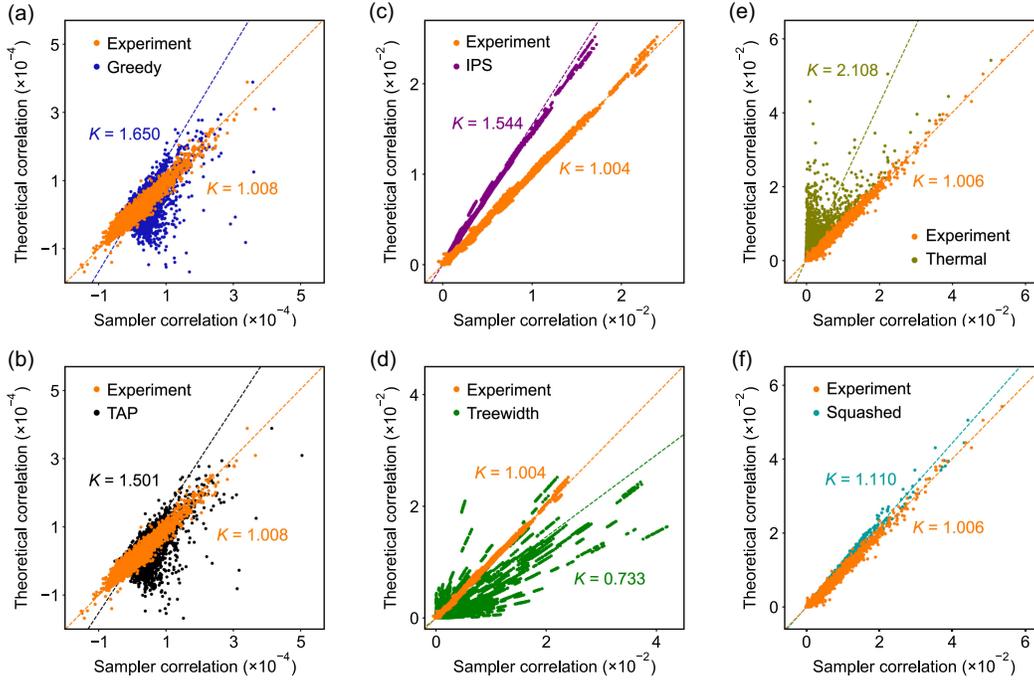


FIG. 3. Correlation function analysis. The experiment clearly exhibits a better agreement with the ground truth than the mockups, with two-norm distance D and slope K (marked in colored dashed lines on each plot). (a),(b) Three-order correlation function in the scatter plot showing the experiment ($D = 0.0054$, $K = 1.008$), the greedy sampler (a) ($D = 0.0062$, $K = 1.650$), and the TAP sampler (b) ($D = 0.0067$, $K = 1.501$) in comparison versus the ground-truth theory. For (a) and (b), 10 million samples are used for estimation of the experimental and mockup's correlation. All three-order correlation functions of 144 out of 1152 modes (one from each one-to-eight fan-out mode) are displayed. (c),(d) Scatter plot of the two-order correlation function of the experiment ($D = 0.045$, $K = 1.004$), the IPS sampler (c) ($D = 0.254$, $K = 1.544$), and the treewidth sampler (d) ($D = 0.739$, $K = 0.733$), versus the ground-truth theory. In (c) and (d), all pairs of 1152 modes are displayed, and 10 million samples are used for estimation of the experimental correlation whereas the exact theoretical value is used for the mockup samplers. (e),(f) The two-order correlation function of the experiment ($D = 0.040$, $K = 1.006$), the thermal state mockup (e) ($D = 0.275$, $K = 2.108$), and the squashed state mockup (f) ($D = 0.052$, $K = 1.110$) versus the ground truth theory, in 144 coarse-grained modes. In (e),(f), 10 million samples are used for estimation of the experimental correlation and the mockup samplers' correlation. All data are from the highest power (1.30 W) experiment.

In Figs. 3(a) and 3(b), we directly compare the third-order correlation functions between the experimental samples at the QCA regime and the mockup samplers from the order-2 greedy and the order-2 TAP samplers. The statistics of the experimental samples agree with the ground truth where they cluster around the identity line at 45° , whereas the mockup samplers' third-order correlations show significant divergence from the ground-truth predictions.

Additionally, Ref. [17] designed the IPS sampler, which was shown to yield a higher heavy output generation (HOG) score than the mockups using the thermal states and distinguishable photons. We compare the second-order cumulants of the sampler with the experimental results in Fig. 3(c), where the IPS sampler shows an evident deviation from the results of the ground truth and the experimental data.

The last mockup to be ruled out is the recently proposed treewidth sampler [22] which can use local connectivity of the circuits to reduce the overhead of classical approximate sampling. The sampler could generate a higher HOG score than some GBS experiments, which, however, is due to the

limitation of HOG test itself. In fact, one can show that the treewidth mockup sampler can even yield a higher HOG score than an ideal ground-truth sampler (see Supplemental Material [44]), which therefore makes the test unreliable [23] (related arguments on the limitation of the cross-entropy benchmarking in random circuit sampling have also been reported [49,50]). Thus, it is necessary to rule out the treewidth mockup in other ways. In Fig. 3(d), we show the second-order correlation of a treewidth mockup (with a propagation length of 65) and the experimental sampler. The former deviates significantly from the ground-truth theory, and is thus unambiguously ruled out.

As a complementary test, we also show in Fig. 3(e) the second order correlation of the experiment, the thermal state, and the squashed state mockup. The correlation of the experiment, with two-norm distance $D = 0.040$ and slope $K = 1.006$, clearly agrees with the ground truth better the thermal state mockup ($D = 0.275$, $K = 2.108$) and the squashed mockup ($D = 0.052$, $K = 1.110$).

Having excluded all currently proposed mockups, we now benchmark the classical computational cost to

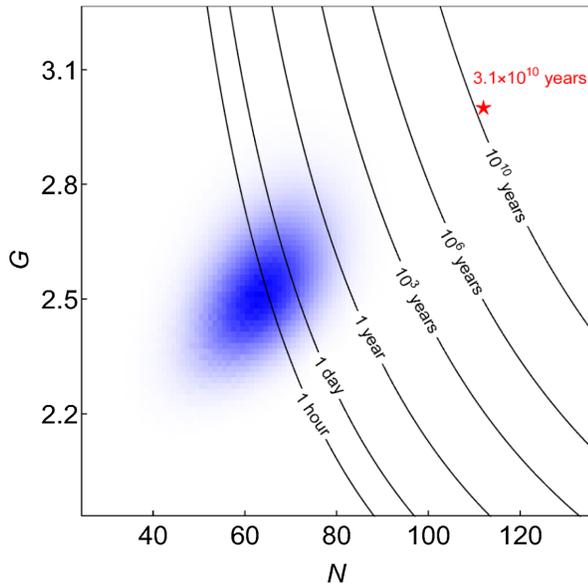


FIG. 4. Quantum computational advantage. The distribution of the estimated classical simulation time on Frontier with exact methods. We calculate the paired value (G, N) for each of the samples obtained in the highest power experiment, and construct the two-dimensional heat map. $G > 2$ indicates the feature of photon number resolving. The hardest sample is marked with a star. Equal simulation time given by Eq. (3) is shown as the lines. As a comparison, it takes $1.27 \mu\text{s}$ to produce a sample in our experiment.

simulate a noiseless version of our GBS experiment. Our pseudo-photon number resolving detection scheme can be modeled by treating all the 1152 fan-out bins as individual output modes, each with threshold detection. We can use results from [17] to exploit the fact that different output bins corresponding to the same optical mode result in repeated entries in the loop Hafnian during the simulation, further reducing the dominant computational cost to

$$T(\vec{N}) = \frac{1}{2} C_{\text{Frontier}} M N^3 G^{N/2} \quad (3)$$

where $\vec{N} = \{n_1, n_2, \dots, n_{144}\}$ represents the PPNRD sample with n_i being the photon-click number of the i th mode, N is the number of clicked modes, M is the mode number, $G = [\prod_i^M (n_i + 1)]^{1/N}$, and C_{Frontier} is estimated based on [12]. The addition of PPNRD therefore substantially increases the computational complexity of our experiment compared to threshold detection (which always has $G = 2$), due to the strong increase in the number of registered detection events.

We now estimate the time cost on Frontier, currently the most powerful supercomputer. For each sample of our high power experiment, we estimate it would on average take Frontier at least ~ 600 yr to generate using the exact methods, while it only takes our machine $1.27 \mu\text{s}$ to produce a sample, showing an overwhelming QCA of

1.5×10^{16} . Moreover, the hardest sample from our experiment would take Frontier more than $\sim 3.1 \times 10^{10}$ yr to generate using an exact algorithm. We show in Fig. 4 the distribution of the estimated classical overhead of all the experimental samples. We hope future work will further consider realistic experimental imperfections such as photon loss and partial photon distinguishability for a better benchmark [51].

Note added.—Recently, we became aware of a related theoretical proposal for balanced-fan-out PPNRD GBS [54].

The dataset of this experiment can be accessed via [53].

We thank Changhun Oh, Minzhao Liu, and Liang Jiang for interesting discussions. We thank H.-T. Zhu, Z.-M. He, R.-X. Wang, X.-Y. Wu, Z.-M. Zhang, and S.-T. Zheng for helpful assistance during the experiment. We thank Y. Liang and Q.-H. Shi for their help in preparation of the manuscript. This work was supported by the National Key R&D Program of China (2019YFA0308700), the National Natural Science Foundation of China, the Chinese Academy of Sciences, the Anhui Initiative in Quantum Information Technologies, the Science and Technology Commission of Shanghai Municipality (2019SHZDZX01), the Innovation Program for Quantum Science and Technology (No. ZD0202010000), the XPLORER PRIZE, the New Cornerstone Science Foundation, and NWO Veni.

Y.-H. D., Y.-C. G., H.-L. L., S.-Q. G., and H. S. contributed equally to this work.

-
- [1] E. Bernstein and U. Vazirani, in *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing—STOC '93* (ACM Press, San Diego, California, United States, 1993), pp. 11–20, [10.1145/167088.167097](https://doi.org/10.1145/167088.167097).
 - [2] J. Preskill, [arXiv:1203.5813](https://arxiv.org/abs/1203.5813).
 - [3] A. W. Harrow and A. Montanaro, *Nature (London)* **549**, 203 (2017).
 - [4] A. P. Lund, M. J. Bremner, and T. C. Ralph, *npj Quantum Inf.* **3**, 15 (2017).
 - [5] D. Hangleiter and J. Eisert, *Rev. Mod. Phys.* **95**, 035001 (2023).
 - [6] S. Aaronson and A. Arkhipov, in *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing—STOC '11* (ACM Press, New York, New York, USA, 2011), pp. 333–342, [10.1145/1993636.1993682](https://doi.org/10.1145/1993636.1993682).
 - [7] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, *Nat. Phys.* **14**, 595 (2018).
 - [8] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin *et al.*, *Nature (London)* **574**, 505 (2019).
 - [9] H. S. Zhong, H. Wang, Y. H. Deng, M. C. Chen, L. C. Peng *et al.*, *Science* **370**, 1460 (2020).

- [10] H.-S. Zhong, Y.-H. Deng, J. Qin, H. Wang, M.-C. Chen *et al.*, *Phys. Rev. Lett.* **127**, 180502 (2021).
- [11] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen *et al.*, *Phys. Rev. Lett.* **127**, 180501 (2021).
- [12] L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. F. Bulmer *et al.*, *Nature (London)* **606**, 75 (2022).
- [13] J. S. Bell, *Phys. Phys. Fiz.* **1**, 195 (1964).
- [14] A. Einstein, B. Podolsky, and N. Rosen, *Phys. Rev.* **47**, 777 (1935).
- [15] Ágoston Kaposi, Z. Kolarovszki, T. Kozsik, Z. Zimborás, and P. Rakyta, [arXiv:2109.04528](https://arxiv.org/abs/2109.04528).
- [16] N. Quesada, R. S. Chadwick, B. A. Bell, J. M. Arrazola, T. Vincent, H. Qi, and R. García-Patrón, *PRX Quantum* **3**, 010306 (2022).
- [17] J. F. F. Bulmer, B. A. Bell, R. S. Chadwick, A. E. Jones, D. Moise *et al.*, *Sci. Adv.* **8**, eab9236 (2022).
- [18] H. Qi, D. J. Brod, N. Quesada, and R. García-Patrón, *Phys. Rev. Lett.* **124**, 100502 (2020).
- [19] J. J. Renema, *Phys. Rev. A* **101**, 063840 (2020).
- [20] J. Shi and T. Byrnes, *npj Quantum Inf.* **8**, 54 (2022).
- [21] B. Villalonga, M. Y. Niu, L. Li, H. Neven, J. C. Platt, V. N. Smelyanskiy, and S. Boixo, [arXiv:2109.11525](https://arxiv.org/abs/2109.11525).
- [22] C. Oh, Y. Lim, B. Fefferman, and L. Jiang, *Phys. Rev. Lett.* **128**, 190501 (2022).
- [23] C. Oh, L. Jiang, and B. Fefferman, *Phys. Rev. Lett.* **131**, 010401 (2023).
- [24] V. Shchesnovich, *Quantum* **5**, 423 (2021).
- [25] A. S. Delliios, M. D. Reid, B. Opanchuk, and P. D. Drummond, [arXiv:2211.03480](https://arxiv.org/abs/2211.03480).
- [26] B. Seron, L. Novo, A. Arkhipov, and N. J. Cerf, [arXiv:2212.09643](https://arxiv.org/abs/2212.09643).
- [27] T. Giordani, V. Mannucci, N. Spagnolo, M. Fumero, A. Rampini, E. Rodolà, and F. Sciarrino, *Quantum Sci. Technol.* **8**, 015005 (2023).
- [28] P. D. Drummond, B. Opanchuk, A. Delliios, and M. D. Reid, *Phys. Rev. A* **105**, 012427 (2022).
- [29] A. S. Popova and A. N. Rubtsov, [arXiv:2106.01445](https://arxiv.org/abs/2106.01445).
- [30] D. Grier, D. J. Brod, J. M. Arrazola, M. B. d. A. Alonso, and N. Quesada, *Quantum* **6**, 863 (2022).
- [31] Y. Lim and C. Oh, [arXiv:2211.07184](https://arxiv.org/abs/2211.07184).
- [32] A. Delliios, P. D. Drummond, B. Opanchuk, R. Y. Teh, and M. D. Reid, *Phys. Lett. A* **429**, 127911 (2022).
- [33] J. T. Iosue, A. Ehrenberg, D. Hangleiter, A. Deshpande, and A. V. Gorshkov, *Quantum* **7**, 1017 (2023).
- [34] Y. Qiao, J. Huh, and F. Grossmann, *SciPost Phys.* **15**, 007 (2023).
- [35] M. Liu, C. Oh, J. Liu, L. Jiang, and Y. Alexeev, [arXiv:2301.12814](https://arxiv.org/abs/2301.12814).
- [36] C. Oh, L. Jiang, and B. Fefferman, [arXiv:2301.11532](https://arxiv.org/abs/2301.11532).
- [37] S. Rahimi-Keshari, T. C. Ralph, and C. M. Caves, *Phys. Rev. X* **6**, 021039 (2016).
- [38] A. Deshpande, B. Fefferman, M. C. Tran, M. Foss-Feig, and A. V. Gorshkov, *Phys. Rev. Lett.* **121**, 030501 (2018).
- [39] U. Chabaud and M. Walschaers, *Phys. Rev. Lett.* **130**, 090602 (2023).
- [40] C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, *Phys. Rev. Lett.* **119**, 170501 (2017).
- [41] N. Quesada, J. M. Arrazola, and N. Killoran, *Phys. Rev. A* **98**, 062322 (2018).
- [42] Y. He, X. Ding, Z.-E. Su, H.-L. Huang, J. Qin, C. Wang *et al.*, *Phys. Rev. Lett.* **118**, 190501 (2017).
- [43] J. S. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. L. Pregnell, Ch. Silberhorn, T. C. Ralph, J. Eisert, M. B. Plenio, and I. A. Walmsley, *Nat. Phys.* **5**, 27 (2009).
- [44] See Supplemental Material, which includes Ref. [45,46], at <http://link.aps.org/supplemental/10.1103/PhysRevLett.131.150601> for detailed description of (1) detector tomography of the pseudo-photon-number resolving detectors; (2) measurement for the matrix phase; (3) partial photon indistinguishability model; (4) Bayesian test results against weaker classical hypotheses; (5) on HOG test with the treewidth mockup.
- [45] M. K. Akhlaghi, A. H. Majedi, and J. S. Lundeen, *Opt. Express* **19**, 21305 (2011).
- [46] A. Christ, K. Laiho, A. Eckstein, K. N. Cassemiro, and C. Silberhorn, *New J. Phys.* **13**, 033027 (2011).
- [47] J. Martínez-Cifuentes, K. M. Fonseca-Romero, and N. Quesada, *Quantum* **7**, 1076 (2023).
- [48] M. Bentivegna, N. Spagnolo, C. Vitelli, D. J. Brod, A. Crespi *et al.*, *Int. J. Quantum. Inform.* **12**, 1560028 (2014).
- [49] B. Barak, C.-N. Chou, and X. Gao, [arXiv:2005.02421](https://arxiv.org/abs/2005.02421).
- [50] X. Gao, M. Kalinowski, C.-N. Chou, M. D. Lukin, B. Barak, and S. Choi, [arXiv:2112.01657](https://arxiv.org/abs/2112.01657).
- [51] After submission, we became aware of a newly developed faster classical approximate simulation algorithm which exhibits an error level of the HOG score and up-to five-order correlation function comparable to the experiment, by taking advantage of the photon loss [52]. Reference [52] takes ~ 10 min to generate the tensor network state corresponding to a specific experiment, while in our experiment it takes $1.27 \mu\text{s}$ to generate a sample and ~ 0.1 ms to reconfigure a new experiment (in Jiuzhang 2.0, only limited by the speed of the physical elements). In addition, Ref. [52] used a simplified model of ground truth for benchmarking of this Letter, without taking pseudo-photon detection and partial distinguishability into account. We expect future GBS experiments with higher efficiency, as well as better characterization of the experimental imperfections, in particular the calibration noise, could also help enhance the quantum advantage against the approximate simulation algorithms.
- [52] C. Oh, M. Liu, Y. Alexeev, B. Fefferman, and L. Jiang, [arXiv:2306.03709](https://arxiv.org/abs/2306.03709).
- [53] Raw data of Jiuzhang 3.0, <https://quantum.ustc.edu.cn/web/node/1121>.
- [54] G. Bressanini, H. Kwon, and M. Kim, [arXiv:2305.00853](https://arxiv.org/abs/2305.00853).