

Quantum Computational Advantage of Noisy Boson Sampling with Partially Distinguishable Photons

Byeongseon Go¹, Changhun Oh^{2,*} and Hyunseok Jeong^{1,†}

¹*NextQuantum Innovation Research Center, Department of Physics and Astronomy, Seoul National University, Seoul 08826, Republic of Korea*

²*Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea*



(Received 22 February 2025; revised 8 July 2025; accepted 15 August 2025; published 25 September 2025)

Boson sampling stands out as a promising approach toward experimental demonstration of quantum computational advantage. However, the presence of physical noise in near-term experiments hinders the realization of quantum computational advantage with boson sampling. Since physical noise in near-term boson-sampling devices is inevitable, precise characterization of the boundary of noise rates where the classical intractability of boson sampling is maintained is crucial for quantum computational advantage using near-term devices. In this work, we identify the level of partial-distinguishability noise that upholds the classical intractability of boson sampling. We find that boson sampling with on average $O(\log N)$ distinguishable photons out of N input photons maintains the equivalent complexity to the ideal boson-sampling case. By providing strong complexity-theoretical evidence for the classical intractability of noisy boson sampling, we expect that our findings will ultimately facilitate the demonstration of quantum computational advantage with noisy boson-sampling experiments in the near future.

DOI: [10.1103/rflv-gc66](https://doi.org/10.1103/rflv-gc66)

I. INTRODUCTION

The development of noisy intermediate-scale quantum devices [1] in recent years has raised expectations for the experimental realization of quantum computational advantage. Various computational tasks have been proposed to achieve a quantum computational advantage, i.e., hard to solve with classical computers but efficiently solvable with quantum computers [2–6]. Among these tasks, boson sampling [6–9] has become one of the most promising candidates for experimental demonstration of quantum computational advantage due to its experimental feasibility and the strong complexity-theoretical evidence of its hardness. In fact, we have now seen several boson-sampling experiments [10–13] the system size of which has been sufficiently large to claim a quantum computational advantage.

However, despite the significant progress in boson-sampling experiments, the degrees of physical noise, such as photon loss and partial distinguishability of photons, in

current experimental setups are still large. This noise can hinder the experimental demonstration of quantum computational advantage with boson sampling, because such physical noise can considerably reduce its computational complexity. More specifically, numerous classical algorithms have been developed to simulate boson sampling subject to physical noise such as photon loss and partial distinguishability of photons [14–31]. These classical algorithms can efficiently simulate noisy boson sampling if the noise level is sufficiently large, possibly ruling out the quantum computational advantage of boson sampling when the noise rate increases quickly with the system size.

Since physical noise is unavoidable in near-term quantum devices and often significantly reduces the computational complexity of the systems, to achieve a quantum computational advantage using near-term quantum devices, it is crucial to identify the boundary of the regime that is classically intractable under the effect of noise, i.e., how large an amount of noise is tolerated for the hardness result. However, the current theoretical hardness results of boson sampling (as in Refs. [6–9,32–34]) are generally for the ideal case without any physically realistic noise; thus, these results may not provide classical-hardness evidence of boson sampling when the physical types of noise are applied.

To resolve this issue, it is essential to find the hardness evidence of noisy boson sampling itself. One promising approach proposed by Ref. [35] is to verify that classical

*Contact author: changhun0218@gmail.com

†Contact author: h.jeong37@gmail.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

simulation of noisy boson sampling is *as hard as* classical simulation of ideal boson sampling. This can be done by showing that the average-case hardness of ideal boson sampling, a key ingredient for the classical hardness of boson sampling [6–9,32–34], can carry over to the average-case hardness of noisy boson sampling with its noise rate below a certain threshold. Then, such noisy boson sampling is equivalently as hard to classically simulate as ideal boson sampling, because the average-case hardness of ideal boson sampling leads to the average-case hardness of noisy boson sampling. Reference [35] has conducted this analysis for photon loss, which is a dominant noise source in current optical setups. More specifically, in Ref. [35] it has been shown that for an N single-photon input, noisy boson sampling where at most a fixed $k = O(1)$ number of lost photons (i.e., $N - k$ photons survive) maintains the same complexity as the ideal boson-sampling case. Since the loss rate in current experiments is obviously much larger than this regime, increasing the noise threshold above the $k = O(1)$ number of lost photons in Ref. [35] remains an important open problem.

Meanwhile, current boson-sampling experiments are susceptible to another type of physical noise, partial distinguishability of photons. This is one of the major noise sources that hinder the quantum computational advantage of boson sampling [15–21]. More specifically, the indistinguishable nature of photons is key to the computational hardness of boson sampling, and the computational complexity of boson sampling reduces as photons become more distinguishable from each other. Also, partial distinguishability is a noise type that appears not only in optical systems [10–13] but also across broader bosonic systems such as atomic arrays and ion traps; those systems have recently served as platforms for boson-sampling experiments [36,37] and are expected to be promising candidates to experimentally demonstrate quantum computational advantage. Therefore, to achieve quantum computational advantage with near-term boson-sampling experiments, it is crucial to investigate the classical intractability of noisy boson sampling subject to partial-distinguishability noise. Here, as we have previously outlined, this can be accomplished by identifying the threshold of the partial-distinguishability noise that maintains the classical intractability of ideal boson sampling.

In this work, we show that noisy N -single photon boson sampling with $O(\log N)$ distinguishable photons, on average, still maintains the classical intractability of ideal boson sampling. Specifically, we show that the average-case estimation of output probabilities of noisy boson sampling with $O(\log N)$ distinguishable photons is complexity-theoretically equivalent to the ideal boson-sampling case. This indicates that boson sampling with a logarithmic number of distinguishable photons is classically hard under the conjecture that ideal boson sampling is classically hard. We also generalize our result to noisy

boson sampling when both photon loss and partial distinguishability are applied and find that at most an $O(\log N)$ number of lost photons and an $O(\log N)$ number of distinguishable photons still upholds the classical intractability of ideal boson sampling.

We expect that our findings can help us to understand the conditions of classical intractability for noisy boson sampling, which has been considerably underexplored compared to the numerous classical simulability arguments of noisy boson sampling [15–30]. Further, by providing strong complexity-theoretical evidence of the classical simulation hardness of noisy boson sampling, we expect that our results will enable one to fully demonstrate quantum computational advantage with noisy boson-sampling experiments in the near future.

Our paper is organized as follows. In Sec. II, we give a brief overview of results related to our work. In Sec. III, we establish the noise model for partial distinguishability of photons and introduce the output probability of noisy boson sampling with partial distinguishability. In Sec. IV, we formally define our main problem, the average-case estimation of the output probabilities of noisy boson sampling with partial distinguishability, and introduce our main result. In Sec. V, we prove that the average-case estimation problem of noisy boson sampling with partial distinguishability is equivalently hard to the ideal boson-sampling case. In Sec. VI, we generalize our main result to the classical intractability of noisy boson sampling when both photon-loss and partial-distinguishability noise are applied. Finally, in Sec. VII, we provide some concluding remarks and directions for future work.

II. RELATED WORKS

Before proceeding, we briefly review related results and clarify our contributions within the context of the existing literature.

We first go through the existing classical simulability results for partially distinguishable boson sampling. In Ref. [16], the authors present an efficient approximate classical algorithm for partially distinguishable boson sampling and find that it becomes more accurate as the distinguishability of photons increases. In Refs. [15,17], the authors establish the combined noise model for photon loss and partial distinguishability and similarly propose a classical algorithm to approximately simulate that noisy boson sampling. In Ref. [18], the authors modify the Gaussian noise proposed in Ref. [14] to realistic partial distinguishability noise and suggest a noise threshold for efficient classical simulability of that noisy boson sampling. Indeed, these results provide a noise threshold for the classical simulability of partially distinguishable boson sampling, providing a necessary condition on the noise rate to achieve quantum computational advantage. On the other hand, our work complements these analyses by providing

a noise threshold for the classical intractability of partially distinguishable boson sampling, establishing a sufficient condition on the noise rate required to achieve quantum computational advantage.

Meanwhile, for the classical-hardness result, in Ref. [38] it is shown that partially distinguishable boson sampling, the single-photon mode mismatch of which is given by $O(N^{-3/2})$ on N input photons, corresponding to an $O(N^{-1/2})$ number of distinguishable photons on average, is complexity-theoretically equivalent to ideal boson sampling. Further, in Ref. [35], complexity-theoretic arguments on lossy boson sampling are established, deriving that a fixed $O(1)$ number of lost photons on an input of N photons maintains the classical intractability of ideal boson sampling. One limitation of these earlier results is that the number of tolerable noisy photons for classical hardness does not *scale* with the system size. Unlike these previous results, we show that partially distinguishable boson sampling with an $O(\log N)$ number of distinguishable photons on average still maintains its classical hardness. Hence, our work significantly improves the noise robustness for classical hardness by allowing the number of tolerable noisy photons to scale with the system size, thereby paving the way for realizing quantum computational advantage with noisy boson-sampling experiments.

For further related works and details, see Appendix A, which provides a detailed overview of the existing literature on the complexity of noisy boson sampling.

III. BOSON SAMPLING WITH PARTIAL DISTINGUISHABILITY

In this section, we introduce the output probability of noisy boson sampling subject to partial-distinguishability noise. We first introduce the output probability of ideal boson sampling and later describe how it varies when partial-distinguishability noise is introduced.

We consider the standard Fock-state boson-sampling scheme in Ref. [6], with the following setup. We prepare N single-photon states and inject them into the first N modes over an M -mode linear optical network, characterized by an $M \times M$ Haar-random unitary matrix U . Here, M is polynomially related to N as $M = \Theta(N^\gamma)$ for a large constant γ , as depicted in Ref. [6]. After the unitary evolution, we measure the output photon number for each mode. Here, the measurement outcome can be represented as an N -dimensional integer vector $S = (s_1, \dots, s_N)$, with $1 \leq s_i \leq M$ and $s_1 \leq s_2 \leq \dots \leq s_N$, where each element represents the output mode in which a photon is detected. Then, the output probability $p_S(U)$ of obtaining the outcome S can be expressed as [6]

$$p_S(U) = \frac{1}{\mu(S)} |\text{Per}(U_{S,1})|^2, \quad (1)$$

where $\mu(S)$ denotes a product of the multiplicity of all possible values in S (e.g., $\mu(S) = 1$ for collision-free outcome S) and $U_{S,1}$ is an $N \times N$ matrix defined by taking rows of U according to S and taking the first N columns of U .

By the hiding property shown in Ref. [6], given $M \times M$ Haar-random unitary matrix U with a sufficiently large γ , $U_{S,1}$ is close to $X/\sqrt{M}$ for a complex independent identically distributed (IID) Gaussian random matrix $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$, for any collision-free outcome S . Therefore, for Haar-random unitary matrix U and for any collision-free outcome S , the output probability $p_S(U)$ can also be represented as $q(X)/M^N$ for $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$, where

$$q(X) = |\text{Per}(X)|^2 = \sum_{\sigma, \rho \in S_N} \prod_{i=1}^N X_{\sigma(i), i} X_{\rho(i), i}^* \quad (2)$$

is the rescaled version of the output probability $p_S(U)$ of ideal boson sampling multiplied by M^N .

We now describe how $q(X)$ in Eq. (2) changes when partial-distinguishability noise is applied. For the partial-distinguishability noise, we employ the noise model of *mutual indistinguishability* between photon pairs, depicted in Refs. [16, 17, 39]. More precisely, the mutual indistinguishability for N photons can be quantified by an $N \times N$ Gram matrix $S_{ij} = \langle \Psi_i | \Psi_j \rangle$ for Ψ_i being the i th single-photon wave function. We note that this mutual indistinguishability between photon pairs (i.e., the elements of the Gram matrix) can be experimentally inferred via Hong-Ou-Mandel interference [40], as demonstrated in the recent boson-sampling experiment [37]. Specifically, this can be done by experimentally obtaining the coincidence probability of a pair of photons and comparing it with that of a perfectly distinguishable pair of photons [37].

Throughout this work, similarly to Refs. [16, 17, 39], we consider the simplified Gram matrix $S_{ij} = x + (1 - x)\delta_{ij}$ for $x \in [0, 1]$, which represents the *uniform* mutual indistinguishability x among all photon pairs. Here and throughout, we refer to x as the indistinguishability rate of boson sampling. Using this noise model, photons become more distinguishable as x decreases; $x = 1$ corresponds to the fully indistinguishable case, $0 < x < 1$ corresponds to a partially distinguishable case, and $x = 0$ corresponds to the fully distinguishable case. After the indistinguishability x is introduced, the (rescaled) noisy output probability $q(x, X)$ transforms from $q(X)$ as [16, 39]

$$q(x, X) = \sum_{\sigma, \rho \in S_N} x^{N - (\sigma \cdot \rho)} \prod_{i=1}^N X_{\sigma(i), i} X_{\rho(i), i}^*, \quad (3)$$

where $(\sigma \cdot \rho)$ is the number of i 's such that $\sigma(i) = \rho(i)$ for all $i \in [N]$. One can readily check the two extremal cases: first, for $x = 1$, $q(x, X) = |\text{Per}(X)|^2$, which reproduces a

(rescaled) output probability of ideal boson sampling and is known to be #P-hard to calculate to within a multiplicative error [6]; and, second, for $x = 0$, $q(x, X) = \text{Per}(|X|^2)$, which can be calculated to within a multiplicative error in polynomial time due to the non-negativity of the matrix elements [41]. Clearly, this clarifies a tendency that the lower the indistinguishability rate x (i.e., photons become more distinguishable from each other), the easier it is to compute the output probability.

More importantly, as proposed in Ref. [17], the indistinguishability rate x in Eq. (3) can also be interpreted as the probability that an individual photon will remain indistinguishable (in other words, each photon becomes distinguishable with probability $1 - x$). Accordingly, the output probability $q(x, X)$ in Eq. (3) can also be represented by a linear sum of output probabilities for all possible indistinguishable photon numbers $j \in \{0, 1, \dots, N\}$, where the coefficients are given by the binomial distribution of N and x as

$$\begin{aligned} q(x, X) &= \sum_{j=0}^N x^j (1-x)^{N-j} \sum_{\substack{I \subseteq [N] \\ |I|=j}} \sum_{\substack{J \subseteq [N] \\ |J|=N-j}} |\text{Per}(X_{I,J})|^2 |\text{Per}(X_{\bar{I},\bar{J}})|^2 \\ &= \sum_{j=0}^N \binom{N}{j} x^j (1-x)^{N-j} q_j(X), \end{aligned} \quad (4)$$

where I and J represent all the possible subsets of $[N]$ with their size j , and $\bar{I}$ and $\bar{J}$ represent $[N] \setminus I$ and $[N] \setminus J$, each with their size $N - j$. Here, $X_{I,J}$ is a matrix defined by taking rows and columns of X according to I and J (also similarly for $X_{\bar{I},\bar{J}}$) and

$$q_j(X) = \binom{N}{j}^{-1} \sum_{\substack{I \subseteq [N] \\ |I|=j}} \sum_{\substack{J \subseteq [N] \\ |J|=N-j}} |\text{Per}(X_{I,J})|^2 |\text{Per}(X_{\bar{I},\bar{J}})|^2 \quad (5)$$

denotes the (rescaled) output probability of boson sampling with a fixed number $j \in \{0, 1, \dots, N\}$ of indistinguishable photons over N photons. For example, $j = N$ corresponds to the output probability with fully indistinguishable photons and, correspondingly, $q_N(X) = |\text{Per}(X)|^2$ is the ideal output probability of boson sampling. Hence, the noisy output probability $q(x, X)$ in Eq. (4) contains the ideal output probability $q_N(X)$ of boson sampling in the summation. This later becomes a crucial property for showing the classical intractability of partially distinguishable boson sampling.

IV. OUR MAIN PROBLEM AND RESULT

In this section, we introduce our main problem for the classical simulation hardness of partially distinguishable boson sampling, which is a modified version

of the main problem addressed in the original boson-sampling proposal [6] (which we have referred to as ideal boson sampling). More specifically, proving the classical intractability of ideal boson sampling relies on the #P-hardness of the average-case estimation of the (rescaled) output probability $q(X)$ in Eq. (2) over $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$, which is also referred to as Gaussian Permanent Estimation (GPE).

Problem 1 ($|\text{GPE}|_{\pm}^2$ [6]). Given as input a matrix $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$ of IID Gaussians, together with error bounds $\epsilon_0, \delta_0 > 0$, estimate $q(X)$ to within additive error $\pm \epsilon_0 \cdot N!$ with probability at least $1 - \delta_0$ over X in $\text{poly}(N, \epsilon_0^{-1}, \delta_0^{-1})$ time.

By Ref. [6], the above $|\text{GPE}|_{\pm}^2$ problem is #P-hard under plausible conjectures. Assuming that $|\text{GPE}|_{\pm}^2$ is #P-hard, one can show that simulating ideal boson sampling within an inverse-polynomial total variation distance error is classically intractable. More specifically, suppose that there exists a randomized classical sampler $\mathcal{S}_0$ that simulates ideal boson sampling: on input, a circuit unitary matrix U samples the outcome from the output distribution of boson sampling within total variation distance β . Given that the input U is an $M \times M$ Haar-random unitary matrix, the (rescaled) output probability of boson sampling can be represented as $q(X)$ for $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$. Then, by Stockmeyer's reduction [42], $|\text{GPE}|_{\pm}^2$ can be solved for $\beta = O(\epsilon_0 \delta_0)$ and, in this case, $|\text{GPE}|_{\pm}^2 \in \text{BPP}^{\text{NP}^{\mathcal{S}_0}}$ (for more details, see Ref. [6]). This implies that if any efficient algorithm for $\mathcal{S}_0$ with $\beta = O(\epsilon_0 \delta_0)$ exists, $|\text{GPE}|_{\pm}^2$ can be solved in BPP^{NP} , and then the infinite level of polynomial hierarchy collapses to the finite level BPP^{NP} , given that $|\text{GPE}|_{\pm}^2$ is #P-hard. In other words, given that the polynomial hierarchy does not collapse to the finite level, there cannot be any classical algorithm that can simulate ideal boson sampling within the total variation distance β in $\text{poly}(N, \beta^{-1})$ time.

Based on this understanding, we now extend this analysis to the noisy version as follows. First, we denote x^* as the *fixed* indistinguishability rate of a partially distinguishable boson sampling hereafter, such that $1 - x^*$ characterizes the noise rate of noisy boson sampling. To derive the classical-hardness result for this noisy boson sampling with indistinguishability rate x^* , which is the goal of this work, we now introduce our key strategy and assumption: we consider an indistinguishability rate x below x^* (i.e., $x \in [0, x^*)$) as an *input variable* to the classical sampler simulating the noisy boson sampling, and derive the hardness result for this classical sampler throughout our work. This constitutes the main distinction from the previous hardness results as in Ref. [35] and represents a crucial aspect of our approach, because, as it turns out, we utilize different values of x to derive our hardness results.

In fact, the consideration of $x \in [0, x^*]$ as an input variable is motivated by the crucial observation that noisy boson sampling becomes classically *easier* as x decreases [15–21]. That is, if noisy boson sampling with indistinguishability x^* is classically simulable, then so is noisy boson sampling with indistinguishability $x \leq x^*$. Let us assume that this observation holds generally for any x^* and $x \in [0, x^*]$, which is a plausible assumption supported by previous findings [15–21]. Then, one can deduce that the classical hardness of simulating noisy boson sampling that can choose any indistinguishability $x \in [0, x^*]$ implies the classical hardness of simulating noisy boson sampling with a fixed indistinguishability x^* .

To derive the classical hardness of simulating noisy boson sampling that takes an indistinguishability rate $x \in [0, x^*]$ as an input, we now set the problem for partially distinguishable boson sampling in correspondence to $|\text{GPE}|_\pm^2$, i.e., average-case estimation of the noisy output probability $q(x, X)$ in Eq. (4). We formally state our main problem, Partially-positive Gaussian Permanent Estimation, hereafter denoted as $|\text{PGPE}|_\pm^2$.

Problem 2 ($|\text{PGPE}|_\pm^2$). Given as input an indistinguishability rate $x \in [0, x^*]$ for a fixed x^* and a matrix $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$ of IID Gaussians, together with error bounds $\epsilon, \delta > 0$, estimate $q(x, X)$ to within additive error $\pm \epsilon \cdot N!$ with probability at least $1 - \delta$ over X in $\text{poly}(N, \epsilon^{-1}, \delta^{-1})$ time.

To highlight the changes in $|\text{PGPE}|_\pm^2$ from the ideal case $|\text{GPE}|_\pm^2$, we modify the output probability to account for noise (i.e., $q(X) \rightarrow q(x, X)$) and also introduce an additional input indistinguishability rate $x \in [0, x^*]$.

In correspondence to the $|\text{PGPE}|_\pm^2$ problem, we can formally define a randomized classical sampler $\mathcal{S}$ that simulates a partially distinguishable boson sampling within total variation distance error β , where now the inputs of $\mathcal{S}$ are (i) a circuit unitary matrix U and (ii) an indistinguishability rate $x \in [0, x^*]$. Then, similarly as to before, to show that the classical sampler $\mathcal{S}$ for partially distinguishable boson sampling is intractable, it suffices to prove that $|\text{PGPE}|_\pm^2$ is $\#P$ -hard. Specifically, given that the input U is a Haar-random unitary matrix, the (rescaled) output probability of noisy boson sampling can be represented as $q(x, X)$ for $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$. Then, given access to the classical sampler $\mathcal{S}$, on input a Haar-random unitary matrix U and an indistinguishability rate $x \in [0, x^*]$, $|\text{PGPE}|_\pm^2$ is in complexity class $\text{BPP}^{\text{NP}^{\mathcal{S}}}$ for $\beta = O(\epsilon\delta)$ by Stockmeyer’s reduction [6, 42]. Therefore, if $|\text{PGPE}|_\pm^2$ is $\#P$ -hard, there cannot be any efficient classical algorithm for $\mathcal{S}$ unless the polynomial hierarchy collapses. Moreover, under our key assumption that classical simulability does not degrade with decreasing x , hardness of $\mathcal{S}$ implies the classical hardness of simulating partially distinguishable boson sampling with a fixed indistinguishability rate x^* .

Next, to derive the $\#P$ -hardness of $|\text{PGPE}|_\pm^2$, we take a similar approach to Ref. [35], i.e., establishing reduction from $|\text{GPE}|_\pm^2$ to $|\text{PGPE}|_\pm^2$. Specifically, we aim to identify the minimum (threshold) x^* that allows *polynomial* reduction from $|\text{GPE}|_\pm^2$ to $|\text{PGPE}|_\pm^2$. Then, summarizing all the previous arguments, this implies that for x^* above such a threshold value, partially distinguishable boson sampling with its indistinguishability rate x^* is equivalently as hard to classically simulate as ideal boson sampling.

To quantify the threshold of x^* that allows the polynomial reduction from $|\text{GPE}|_\pm^2$ to $|\text{PGPE}|_\pm^2$, we employ a convention for the input indistinguishability rate as $x = 1 - k/N$. Here, because x can be interpreted as the probability of being indistinguishable for each photon [17], the average number of indistinguishable photons is $Nx = N - k$. Then, similarly to the photon-loss case as in Ref. [22], $k = (1 - x)N$ represents the average number of photons subject to noise and, in our case, it represents the average number of distinguishable photons. Also, the fixed indistinguishability rate can be represented as $x^* = 1 - k_{\min}/N$, where the subscript “min” comes from the fact that x^* is indeed a maximum value over input indistinguishability rate x .

Finally, we introduce our main result. We show that for $k_{\min} = O(\log N)$, $|\text{PGPE}|_\pm^2$ is polynomially reducible from $|\text{GPE}|_\pm^2$. Informally, our main result can be expressed as follows.

Theorem 1 (Informal). For x^* satisfying $(1 - x^*)N = O(\log N)$, $|\text{PGPE}|_\pm^2$ is polynomially reducible from $|\text{GPE}|_\pm^2$. In other words, if $\mathcal{O}$ is an oracle that solves $|\text{PGPE}|_\pm^2$ for $x^* = 1 - k_{\min}/N$ with $k_{\min} = O(\log N)$, together with $\epsilon, \delta = O(\text{poly}(N, \epsilon_0^{-1}, \delta_0^{-1})^{-1})$, then $|\text{GPE}|_\pm^2$ can be solved in $\text{BPP}^{\mathcal{O}}$.

We provide in Sec. V a formal statement of Theorem 1, followed by a detailed proof of the theorem. Here, let us emphasize that, by Theorem 1, partially distinguishable boson sampling that has on average an $(1 - x^*)N = O(\log N)$ number of distinguishable photons is complexity-theoretically equivalent to ideal boson sampling. More specifically, ϵ^{-1} and δ^{-1} in Theorem 1 are polynomially related to N , ϵ_0^{-1} , and δ_0^{-1} . Hence, for ϵ and δ given in Theorem 1, the allowed operation time to solve $|\text{PGPE}|_\pm^2$ problem is still in $\text{poly}(N, \epsilon_0^{-1}, \delta_0^{-1})$. Accordingly, by Theorem 1, if one can classically simulate noisy boson sampling with up to a $k_{\min} = c_{\min} \log N$ number of distinguishable photons on average, to within $O(\epsilon\delta)$ total variation distance, then one can also solve the $|\text{GPE}|_\pm^2$ problem in BPP^{NP} . Therefore, under the complexity-theoretic conjecture that $|\text{GPE}|_\pm^2$ is $\#P$ -hard, we can deduce that such a noisy boson sampling is hard to classically simulate.

In addition, by using a similar process to the proof of Theorem 2, we also find that if the average number of distinguishable photons k scales sublogarithmically with

N , then the requirement of the approximation error ϵ for the reduction from $|\text{GPE}\rangle_{\pm}^2$ to $|\text{PGPE}\rangle_{\pm}^2$ is comparably less stringent.

Corollary 1. If $\mathcal{O}$ is an oracle that solves $|\text{PGPE}\rangle_{\pm}^2$ for $x^* = 1 - k_{\min}/N$ with $k_{\min} = o(\log N)$, $\epsilon = O(\epsilon_0^{\alpha+1} \delta_0^{\alpha} N^{-\alpha})$ with any constant $\alpha > 0$, and $\delta = O(\delta_0/(\log N + \log \epsilon_0^{-1} \delta_0^{-1}))$, then $|\text{GPE}\rangle_{\pm}^2$ can be solved in $\text{BPP}^{\mathcal{O}}$.

Proof. See Appendix D. ■

V. REDUCING $|\text{GPE}\rangle_{\pm}^2$ TO $|\text{PGPE}\rangle_{\pm}^2$ FOR $x^* = 1 - k_{\min}/N$ WITH $k_{\min} = O(\log N)$

In this section, we establish a reduction from $|\text{GPE}\rangle_{\pm}^2$ to $|\text{PGPE}\rangle_{\pm}^2$, especially for the case in which the minimum average number of distinguishable photons scales logarithmically: $k_{\min} = O(\log N)$. Here, to simplify the analysis, we set $k_{\min}$ as $c_{\min} \log N$ for a constant $c_{\min}$, without loss of generality. Using this convention, we restate our main theorem for the reduction from $|\text{GPE}\rangle_{\pm}^2$ to $|\text{PGPE}\rangle_{\pm}^2$, which can be formally represented as follows.

Theorem 2. If $\mathcal{O}$ is an oracle that solves $|\text{PGPE}\rangle_{\pm}^2$ for $x^* = 1 - k_{\min}/N$ with $k_{\min} = c_{\min} \log N$, together with $\epsilon = O(\epsilon_0^{7.1} \delta_0^{6.1} N^{-40c_{\min}})$, and $\delta = O(\delta_0/(\log N + \log \epsilon_0^{-1} \delta_0^{-1}))$, then $|\text{GPE}\rangle_{\pm}^2$ can be solved in $\text{BPP}^{\mathcal{O}}$.

Because $c_{\min}$ is a constant, ϵ^{-1} and δ^{-1} in Theorem 2 are at most polynomially related to N , ϵ_0^{-1} , and δ_0^{-1} . As we have discussed in the previous section, this indicates the complexity-theoretical equivalence of classically simulating partially distinguishable boson sampling with an $O(\log N)$ number of distinguishable photons and classically simulating ideal boson sampling.

In the rest of the section, we provide a step-by-step proof of Theorem 2.

A. Proof sketch

Let us first sketch the proof of Theorem 2 by briefly describing the reduction process from $|\text{GPE}\rangle_{\pm}^2$ to $|\text{PGPE}\rangle_{\pm}^2$. As the noisy output probability $q(x, X)$ in Eq. (4) is a polynomial in x , given access to $q(x, X)$ values for different values of input $x \leq x^*$, we use the polynomial-interpolation technique to infer the value at $x = 1$, i.e., $q(x = 1, X) = |\text{Per}(X)|^2$. Indeed, this process is similar to one of the main techniques of error mitigation, i.e., zero-noise extrapolation [43–45]. However, since $q(x, X)$ is an N -degree polynomial in x , directly using the polynomial interpolation for $q(x, X)$ induces at least exponential additive imprecision blowup in N , as given in Ref. [46] (i.e., about $e^{O(N)}$), which requires the approximation error ϵ in $|\text{PGPE}\rangle_{\pm}^2$ to be exponentially small (about $e^{-O(N)}$) for the reduction.

To avoid this issue, our main strategy is to find a *low-degree polynomial* in x that (i) can be well estimated via the access to the estimated values of $q(x, X)$ for $x \leq x^*$ (i.e., access to an oracle that can solve the $|\text{PGPE}\rangle_{\pm}^2$ problem) and (ii) contains the desired value $|\text{Per}(X)|^2$ at $x = 1$. Here and throughout, we will use a notation $f_X^{(l)}(x)$ as the low-degree polynomial described above, with l corresponding to the degree of $f_X^{(l)}(x)$. Then, given the well-estimated values of $f_X^{(l)}(x)$ with each corresponding to $l + 1$ values of $x \leq x^*$, we can use the polynomial interpolation for $f_X^{(l)}(x)$ and infer the desired value $f_X^{(l)}(1) = |\text{Per}(X)|^2$. This overall process is illustrated in Fig. 1. Note that the degree l of the low-degree polynomial $f_X^{(l)}(x)$ should be at most $O(\log(\text{poly}(N, \epsilon_0^{-1}, \delta_0^{-1})))$, so that the polynomial-interpolation process induces an imprecision blowup that scales at most as $e^{O(\log(\text{poly}(N, \epsilon_0^{-1}, \delta_0^{-1})))} = \text{poly}(N, \epsilon_0^{-1}, \delta_0^{-1})$.

B. Conventions

Before proceeding to the main proof, we establish the parameters that will be frequently used throughout the proof. Given the fixed indistinguishability rate x^* which is indeed the upper bound of input indistinguishability values x , we similarly set the lower bound of x as $x_{\min}$, such that we further restrict the input $x \in [x_{\min}, x^*]$ to the oracle. As the indistinguishability rate x can be expressed as $1 - k/N$ by the average noisy photon number k , those boundary values of x can also be represented as

$$x_{\min} = 1 - \frac{k_{\max}}{N}, \quad x^* = 1 - \frac{k_{\min}}{N}, \quad (6)$$

where $k_{\min}$ ($k_{\max}$) denotes the lower (upper) bound of the average noisy photon number k , such that $k_{\min} \leq k \leq k_{\max}$.

For simplicity, as we are interested in logarithmically scaling k with N , we use a convention $k = c \log N$ for a constant c without loss of generality. Then, the boundary values of k can also be represented as

$$k_{\max} = c_{\max} \log N, \quad k_{\min} = c_{\min} \log N, \quad (7)$$

where $c_{\min}$ ($c_{\max}$) denotes the lower (upper) bound of the constant c , such that $c_{\min} \leq c \leq c_{\max}$.

Lastly, since we consider that the degree l of the low-degree polynomial $f_X^{(l)}(x)$ scales at most logarithmically with N , we similarly define l as

$$l = c_l \log N, \quad (8)$$

where we set l larger than any k such that $l > k_{\max}$ (in other words, $c_l > c_{\max}$). It is worth emphasizing that $c_{\min}$ is a given constant determined by the fixed indistinguishability rate x^* , whereas $c_{\max}$ and c_l are parameters we can arbitrarily set.

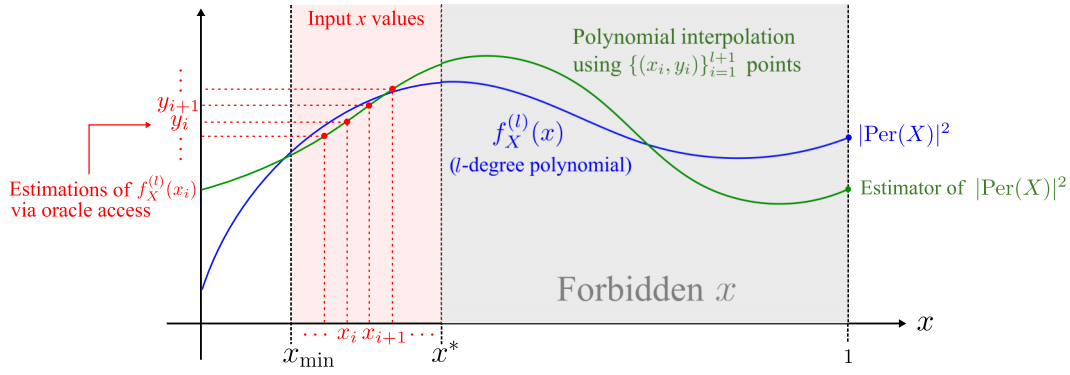


FIG. 1. The schematics for the proof of Theorem 2, i.e., how to solve the $|GPE|_{\pm}^2$ problem via oracle access to the $|PGPE|_{\pm}^2$ problem. We first find the l -degree polynomial $f_X^{(l)}(x)$ that can be well estimated via oracle access to the $|PGPE|_{\pm}^2$ problem for $x \leq x^*$, and satisfies $f_X^{(l)}(1) = |\text{Per}(X)|^2$ (blue line). Then, for each of the $l+1$ input indistinguishability rates $x_i \in [x_{\min}, x^*]$, we obtain the value of y_i that estimates $f_X^{(l)}(x_i)$ via the oracle access (red dots). Lastly, we conduct polynomial interpolation with the given $\{x_i, y_i\}_{i=1}^{l+1}$ points to infer the polynomial $f_X^{(l)}(x)$ (green line) and finally obtain the estimation value of $|\text{Per}(X)|^2$ at $x = 1$.

C. Finding a low-degree polynomial $f_X^{(l)}(x)$

Our key idea is to find a low-degree polynomial $f_X^{(l)}(x)$ that can be well estimated via the oracle access to the $|PGPE|_{\pm}^2$ problem and contains the desired value $|\text{Per}(X)|^2$. To obtain such an l -degree polynomial $f_X^{(l)}(x)$ from the noisy output probability $q(x, X)$ in Eq. (4), our approach is first to truncate the summation terms in $q(x, X)$ and only leave $l+1$ terms from $j = N-l$ to $j = N$.

Here, given $l > k_{\max}$, the summation terms in Eq. (4) from $j = 0$ to $j = N-l-1$ correspond to the tail of the binomial distribution $B(N, x)$, the expectation number of success of which is $Nx = N - k$. Hence, if l is sufficiently larger than $k_{\max}$, one can expect that those tail terms from $j = 0$ to $j = N-l-1$ would be small enough. In other words, we expect that

$$q(x, X) = \sum_{j=0}^N \binom{N}{j} x^j (1-x)^{N-j} q_j(X) \quad (9)$$

$$\approx \sum_{j=N-l}^N \binom{N}{j} x^j (1-x)^{N-j} q_j(X), \quad (10)$$

for l sufficiently larger than $k_{\max}$.

Hereafter, we newly define $q^{(l)}(x, X)$ as the output probability after truncating the tail terms from $j = 0$ to $j = N-l-1$ in $q(x, X)$, such that

$$q^{(l)}(x, X) := \sum_{j=N-l}^N \binom{N}{j} x^j (1-x)^{N-j} q_j(X). \quad (11)$$

We show that the truncated output probability $q^{(l)}(x, X)$ can be made arbitrarily close to $q(x, X)$, by increasing l

to be sufficiently larger than $k_{\max}$. Specifically, we prove the following lemma.

Lemma 1. For $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$ and $x \geq 1 - k_{\max}/N$, $q(x, X)$ and $q^{(l)}(x, X)$ are $\epsilon_1 N!$ -close for

$$\epsilon_1 = \delta^{-1} N^{-\frac{c_{\max}}{3}} \left(\frac{c_l}{c_{\max}} - 1 \right)^2, \quad (12)$$

with probability at least $1 - \delta$ over X .

We defer a detailed proof of Lemma 1 to Appendix B. One can easily check that ϵ_1 in Eq. (12) decreases with increasing c_l . Hence, by Lemma 1, $q^{(l)}(x, X)$ can be made close enough to $q(x, X)$ over a large portion of $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$, by setting c_l sufficiently large compared to $c_{\max}$.

Given the truncated output probability $q^{(l)}(x, X)$ that is close to the noisy output probability $q(x, X)$, now we can obtain the low-degree polynomial $f_X^{(l)}(x)$ by decreasing the polynomial degree of $q^{(l)}(x, X)$. Specifically, $q^{(l)}(x, X)$ can be expressed as $x^{N-l} f_X^{(l)}(x)$ for the l -degree polynomial $f_X^{(l)}(x)$ such that

$$f_X^{(l)}(x) = x^{-N+l} q^{(l)}(x, X) \quad (13)$$

$$= \sum_{j=N-l}^N \binom{N}{j} x^{j-N+l} (1-x)^{N-j} q_j(X) \quad (14)$$

$$= a_0 + a_1 x + a_2 x^2 + \cdots + a_l x^l, \quad (15)$$

where the coefficients $a_0, \dots, a_l$ of $f_X^{(l)}(x)$ are given by the combinations of $q_j(X)$ for $j \in [N-l, N]$.

Given that c_l is sufficiently larger than $c_{\max}$, the l -degree polynomial $f_X^{(l)}(x)$ can be well estimated for $x \leq x^*$ via

the oracle access to $|\text{PGPE}|_{\pm}^2$. Specifically, this can be done by first querying the $q(x, X)$ value (which is close to the $q^{(l)}(x, X)$ value as given in Lemma 1) and then multiplying x^{-N+l} . Also, one can easily check that $f_X^{(l)}(x)$ exhibits the desired value $|\text{Per}(X)|^2$ at $x = 1$, i.e., $f_X^{(l)}(1) = q_N(X) = |\text{Per}(X)|^2$ as given in Eq. (5). Hence, using the well-estimated values of $f_X^{(l)}(x)$ corresponding to $l+1$ values of $x \leq x^*$, we can infer the desired value $f_X^{(l)}(1)$ via polynomial interpolation. This overall process is described in Fig. 1.

D. Establishing reduction from $|\text{GPE}|_{\pm}^2$ to $|\text{PGPE}|_{\pm}^2$

Given the l -degree polynomial $f_X^{(l)}(x)$ as introduced in the previous section, we are now ready to prove Theorem 2. Specifically, we establish the BPP reduction from $|\text{GPE}|_{\pm}^2$ to $|\text{PGPE}|_{\pm}^2$ for $k_{\min} = c_{\min} \log N$, with the error parameters given by $\epsilon = O(\epsilon_0^{7.1} \delta_0^{6.1} N^{-40c_{\min}})$ and $\delta = O(\delta_0/(\log N + \log \epsilon_0^{-1} \delta_0^{-1}))$.

Proof of Theorem 2. Let $\mathcal{O}$ be the oracle introduced in Theorem 2, i.e., on input $x \leq 1 - k_{\min}/N$ and $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$, that estimates $q(x, X)$ to within $\epsilon N!$ over $1 - \delta$ of X such that

$$\Pr_X[|\mathcal{O}(x, X) - q(x, X)| > \epsilon N!] < \delta. \quad (16)$$

In the following, we show that given access to the oracle $\mathcal{O}$, one can obtain the desired value $p(1, X) = |\text{Per}(X)|^2$ to within $\epsilon_0 N!$ over $1 - \delta_0$ of X (i.e., solve the $|\text{GPE}|_{\pm}^2$ problem), for ϵ and δ given in Theorem 2.

Let us define $\epsilon' := \epsilon + \epsilon_1$ for ϵ_1 given in Eq. (12). By Lemma 1, the triangular inequality, and the union bound, we have

$$\begin{aligned} & \Pr_X[|\mathcal{O}(x, X) - q^{(l)}(x, X)| > \epsilon' N!] \\ & \leq \Pr_X[|\mathcal{O}(x, X) - q(x, X)| > \epsilon N!] \\ & \quad + \Pr_X[|q(x, X) - q^{(l)}(x, X)| > \epsilon_1 N!] \quad (17) \\ & < 2\delta, \quad (18) \end{aligned}$$

which implies that $\mathcal{O}$ can also well estimate $q^{(l)}(x, X)$ to within $\epsilon' N!$ with probability at least $1 - 2\delta$ over X .

As we have previously defined, $x_{\min}$ is a minimum value of x that we set, such that $x_{\min} = 1 - k_{\max}/N$ with $k_{\max} = c_{\max} \log N$. Let us define $\epsilon'' := \epsilon' N^{c_{\max}}$, such that $\epsilon' x^{-N+l} \leq \epsilon' x_{\min}^{-N+l} = \epsilon' e^{k_{\max} - \tilde{O}(N^{-1})} \leq \epsilon''$ for all possible

x . Then, the following inequality holds:

$$\begin{aligned} & \Pr_X[|\mathcal{O}(x, X)x^{-N+l} - f_X^{(l)}(x)| > \epsilon' N!] \\ & \leq \Pr_X[|\mathcal{O}(x, X)x^{-N+l} - f_X^{(l)}(x)| > \epsilon' x^{-N+l} N!] \quad (19) \\ & = \Pr_X[|\mathcal{O}(x, X) - q^{(l)}(x, X)| > \epsilon' N!] \quad (20) \\ & < 2\delta, \quad (21) \end{aligned}$$

by using Eq. (18).

Accordingly, for each access to the oracle $\mathcal{O}$ on input of $x \leq 1 - k_{\min}/N$ and $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$, one can estimate $f_X^{(l)}(x)$ to within $\epsilon'' N!$ with probability at least $1 - O(\delta)$ over X . The remaining problem is to infer the desired value $f_X^{(l)}(1)$ from the estimated values $f_X^{(l)}(x)$. Let $\{k_i\}_{i=1}^{l+1}$ be the set of $l+1$ equally spaced points in the interval $k_i \in [k_{\min}, k_{\max}]$. For each $x_i = 1 - k_i/N$, let $y_i = \mathcal{O}(x_i, X)x_i^{-N+l}$. By Eq. (21), each set of points (x_i, y_i) satisfies

$$\Pr[|y_i - f_X^{(l)}(x_i)| > \epsilon'' N!] < 2\delta. \quad (22)$$

For simplicity, we change the l -degree polynomial to $g(x) = f_X^{(l)}(z(x))$ for a linear function $z(x) = ((k_{\max} + k_{\min})/2N)(x - 1) + 1$, to rescale the input variable x . After rescaling the variable, we have the input $\{x_i\}_{i=1}^{l+1}$, which is the set of equally spaced points in the interval $x_i \in [-\Delta, \Delta]$ with a constant $\Delta := (k_{\max} - k_{\min})/(k_{\max} + k_{\min}) = (c_{\max} - c_{\min})/(c_{\max} + c_{\min})$. Likewise, for each input x_i , we have an estimator $y_i = \mathcal{O}(z(x_i), X)z(x_i)^{-N+l}$ for $g(x_i)$ such that

$$\Pr[|y_i - g(x_i)| > \epsilon'' N!] < 2\delta. \quad (23)$$

Using those estimation values, we infer the value $g(1) = f_X^{(l)}(1) = |\text{Per}(X)|^2$ via the polynomial-interpolation technique, as we have described previously.

To infer the value of $g(1)$, we use the Lagrange interpolation technique for $g(x)$; to do so, it should be promised that all the (x_i, y_i) points satisfy $|y_i - g(x_i)| \leq \epsilon'' N!$. Here, by simply using the union bound, the probability that all $l+1$ points y_i are $\epsilon'' N!$ -close to $g(x_i)$ is at least $(1 - 2\delta)^{l+1}$. Given that all the $l+1$ points are successful, we can use the Lagrange interpolation technique, the upper bound of the error of which has been shown in Ref. [47], as follows.

Lemma 2 (Kondo et al. [47]). Let $h(x)$ be a polynomial of degree at most l and let $\Delta \in (0, 1)$. Assume that $|h(x_j)| \leq \epsilon$ for all of the $l+1$ equally spaced points $x_j = -\Delta + (2j/l)\Delta$ for $j = 0, \dots, l$. Then,

$$|h(1)| < \epsilon \frac{\exp[l(1 + \log \Delta^{-1})]}{\sqrt{2\pi l}}. \quad (24)$$

By drawing l -degree polynomial $g_e(x)$ according to (x_i, y_i) points, we get a polynomial $h(x) := g_e(x) - g(x)$. Then, $|h(x_i)| \leq \epsilon'' N!$ for x_i satisfying $|y_i - g(x_i)| \leq \epsilon'' N!$. Therefore, given the error bound as in Eq. (24), we can obtain an estimator $g_e(1)$ for $g(1)$ satisfying $|g_e(1) - g(1)| < \epsilon'' (\exp[l(1 + \log \Delta^{-1})]) / \sqrt{2\pi l N!} < \epsilon'' e^{l(1 + \log \Delta^{-1})} N!$ via Lagrange interpolation whenever all the $l + 1$ points are successful, the probability of which is at least $(1 - 2\delta)^{l+1}$. In other words, we can estimate the desired value $g(1) = |\text{Per}(X)|^2$ as

$$\Pr \left[|g_e(1) - |\text{Per}(X)|^2| > \epsilon'' e^{l(1 + \log \Delta^{-1})} N! \right] < 1 - (1 - 2\delta)^{l+1}. \quad (25)$$

Therefore, to solve $|\text{GPE}|_\pm^2$ with Eq. (25), the error parameters ϵ and δ should satisfy the conditions $\epsilon_0 \geq \epsilon'' e^{l(1 + \log \Delta^{-1})} = \epsilon'' N^{c_l(1 + \log \Delta^{-1})}$ and $\delta_0 \geq 1 - (1 - 2\delta)^{l+1}$. Accordingly, we set δ satisfying $\delta_0 = 1 - (1 - 2\delta)^{l+1}$. Also, combining the results obtained so far, we get

$$\epsilon_0 \geq N^{c_{\max} + c_l(1 + \log \Delta^{-1})} (\epsilon + \epsilon_1), \quad (26)$$

which requires ϵ to satisfy the following condition to solve $|\text{GPE}|_\pm^2$ problem:

$$\epsilon \leq \epsilon_0 N^{-c_{\max} - c_l(1 + \log \Delta^{-1})} - \epsilon_1, \quad (27)$$

for $\epsilon_1 = \delta^{-1} N^{-\frac{c_{\max}}{3} \left(\frac{c_l}{c_{\max}} - 1 \right)^2}$ as given in Lemma 1.

Since the error parameter is positive such that $\epsilon > 0$, one should make sure that $\epsilon_0 N^{-c_{\max} - c_l(1 + \log \Delta^{-1})}$ is the leading term in the right-hand side of Eq. (27) as the system size scales. To do so, note that we have the freedom to choose the value c_l and $c_{\max}$ for given $c_{\min}$. Specifically, by setting c_l sufficiently larger than $c_{\max}$, we can make ϵ_1 arbitrarily smaller than $\epsilon_0 N^{-c_{\max} - c_l(1 + \log \Delta^{-1})}$. On the other hand, arbitrarily increasing c_l results in arbitrarily small approximation error ϵ in Eq. (27), because ϵ scales as approximately $N^{-c_l(1 + \log \Delta^{-1})}$ with c_l . Therefore, we need to find an *intermediate* size of c_l , to (i) make ϵ_1 small enough compared to $\epsilon_0 N^{-c_{\max} - c_l(1 + \log \Delta^{-1})}$ and (ii) prevent an excessive decrease in ϵ .

Considering all those arguments, we investigate the appropriate size of the parameters $c_{\max}$ and c_l in Appendix C. More specifically, we set $c_{\max}$ and c_l as

$$c_{\max} = \kappa c_{\min}, \quad (28)$$

$$c_l = \kappa \lambda \frac{5 + 3 \log \frac{\kappa+1}{\kappa-1}}{2} c_{\min} + \frac{3 \log \epsilon_0^{-1} \delta_0^{-1}}{\log N}, \quad (29)$$

with $\kappa = 2.108$ and $\lambda = 2.149$. Using these values, we find in Appendix C that the right-hand side of Eq. (27) becomes

$$\begin{aligned} & \epsilon_0 N^{-c_{\max} - c_l(1 + \log \Delta^{-1})} - \epsilon_1 \\ &= O \left(\epsilon_0 N^{-c_{\max} - c_l(1 + \log \Delta^{-1})} \right) \end{aligned} \quad (30)$$

$$= O \left(\epsilon_0^{7.094} \delta_0^{6.094} N^{-39.35 c_{\min}} \right), \quad (31)$$

where we have used $\Delta = (c_{\max} - c_{\min}) / (c_{\max} + c_{\min}) = (\kappa - 1) / (\kappa + 1)$. Hence, by setting $c_{\max}$ and c_l as the above, the condition for the error parameter ϵ in Eq. (27) can be expressed as

$$\epsilon \leq O \left(\epsilon_0^{7.094} \delta_0^{6.094} N^{-39.35 c_{\min}} \right). \quad (32)$$

To sum up, by setting the error parameters ϵ and δ as $\delta = 1 - (1 - 2\delta)^{l+1} = O(\delta_0 l^{-1}) = O(\delta_0 / (\log N + \log \epsilon_0^{-1} \delta_0^{-1}))$ and

$$\epsilon = O \left(\epsilon_0^{7.1} \delta_0^{6.1} N^{-40 c_{\min}} \right), \quad (33)$$

one can probabilistically estimate $|\text{Per}(X)|^2$ to within $\epsilon_0 N!$ with probability at least $1 - \delta_0$ over X given access to the oracle $\mathcal{O}$. This completes the proof. ■

VI. CLASSICAL INTRACTABILITY FOR PARTIALLY DISTINGUISHABLE AND LOSSY BOSON SAMPLING

Photon loss and partial distinguishability of photons are, indeed, typically regarded as the most crucial noise sources in boson-sampling experiments [10–13]. Accordingly, in this section we discuss whether the classical intractability of boson sampling can be maintained when *both* photon loss noise and partial-distinguishability noise are applied.

For the photon-loss noise, we consider the conventional beam-splitter loss channel described in Refs. [22–24], where the degree of noise is characterized by a uniform transmission rate $\eta \in [0, 1]$ of the entire circuit. Let us denote n as the output photon number after photon loss, such that $n \leq N$ for the input photon number N . When the uniform transmission rate η is applied for each of N input photons, the probability of observing n -photon outcomes follows a binomial distribution $\binom{N}{n} \eta^n (1 - \eta)^{N-n}$.

Now, let $q(x, \eta, X)$ be a (rescaled) output probability of obtaining the first n modes outcome over M modes, where the rescaling factor is given by M^n . Here, X is the first $n \times N$ submatrix of a rescaled $M \times M$ unitary matrix $\sqrt{M}U$, which can be approximated to the random Gaussian matrix $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{n \times N}$ for $M \times M$ Haar-random unitary matrix U given that $M = \Theta(N^\gamma)$ for sufficiently large γ . Then, as

proposed in Ref. [17], $q(x, \eta, X)$ can be represented as

$$\begin{aligned}
 q(x, \eta, X) &= \eta^n (1 - \eta)^{N-n} \sum_{\substack{K \subseteq [N] \\ |K|=n}} \sum_{j=1}^n x^j (1 - x)^{N-j} \\
 &\quad \times \sum_{\substack{I \subseteq K \\ |I|=j}} \sum_{\substack{J \subseteq K \\ |J|=j}} |\text{Per}((X_K)_{I,J})|^2 |\text{Per}((X_K)_{\bar{I},\bar{J}})|^2 \\
 &= \binom{N}{n} \eta^n (1 - \eta)^{N-n} \cdot \binom{N}{n}^{-1} \sum_{\substack{K \subseteq [N] \\ |K|=n}} q(x, X_K),
 \end{aligned} \tag{34}$$

where X_K is an $n \times n$ matrix defined by taking columns of X according to K . Also, $q(x, X_K)$ is the partially distinguishable output probability given in Eq. (4).

Here, let us remark that we consider the general output probability for an n -photon outcome, in the sense that the binomial factor $\binom{N}{n} \eta^n (1 - \eta)^{N-n}$ is included in the output probability on the right-hand side of Eq. (34). Note that when n , N , and η satisfy $\binom{N}{n} \eta^n (1 - \eta)^{N-n} = \Omega(\text{poly}(n)^{-1})$, one can instead consider a noisy boson sampler with a fixed output photon number n , which can be implemented by repeating the sampling polynomial a number of times and postselecting n -photon outcomes. In this case, the binomial factor $\binom{N}{n} \eta^n (1 - \eta)^{N-n}$ in the right-hand side of Eq. (34) vanishes, yielding a conditional output probability that is independent of η . However, since our analysis of the condition on η for the hardness result requires an explicit dependence of the output probability on η , we use the general output probability in Eq. (34) rather than the conditional output probability.

To examine the classical intractability of simulating noisy boson sampling with loss and partial distinguishability, we follow the approach in Sec. IV. That is, we similarly set the problem corresponding to the average-case estimation of the noisy output probability $q(x, \eta, X)$ over $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{n \times N}$, which can be formally stated as follows.

Problem 3 ($\sum |\text{PGPE}|_{\pm}^2$). Given as input a transmission rate $\eta \in [0, \eta^*]$, an indistinguishability rate $x \in [0, x^*]$, and a matrix $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{n \times N}$ of IID Gaussians with $N \geq n$, together with error bounds $\epsilon, \delta > 0$, estimate $q(x, \eta, X)$ to within additive error $\pm \epsilon \cdot n!$ with probability at least $1 - \delta$ over X in $\text{poly}(n, \epsilon^{-1}, \delta^{-1})$ time.

Similarly to x^* , η^* is a fixed transmission rate of noisy boson sampling, such that $1 - x^*$ and $1 - \eta^*$ characterize the noise rate for partial-distinguishability and photon-loss noises, respectively. Also, to avoid any confusion, we note that now the problem size is given by the output photon number n , while we set the input photon number N as a free parameter that we can arbitrarily choose, as long as $n \leq N \ll M$.

As we have previously argued in Sec. IV, the #P-hardness of the $\sum |\text{PGPE}|_{\pm}^2$ problem serves as a sufficient condition for the classical intractability of simulating noisy boson sampling subject to loss and distinguishability. Specifically, on input of a Haar-random unitary matrix U , $\eta \in [0, \eta^*]$ and $x \in [0, x^*]$, the output probability of noisy boson sampling for the first n -mode outcome can be represented as $q(x, \eta, X)$ in Eq. (34) (rescaled by M^n), for $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{n \times N}$. Then, considering a randomized classical sampler $\mathcal{S}$ that simulates noisy boson sampling to within total variation distance β , with η^* and x^* characterizing the noise rate for loss and partial distinguishability, $\sum |\text{PGPE}|_{\pm}^2$ is in complexity class $\text{BPP}^{\text{NP}^{\mathcal{S}}}$ for $\beta = O(\epsilon \delta)$ by Stockmeyer's reduction [42]. Therefore, showing the #P-hardness of $\sum |\text{PGPE}|_{\pm}^2$ implies that there cannot be any efficient classical sampler $\mathcal{S}$ for the approximate noisy boson sampling, unless the polynomial hierarchy collapses. We note that this hardness result requires that n , N , and η satisfy $\binom{N}{n} \eta^n (1 - \eta)^{N-n} = \Omega(\text{poly}(n)^{-1})$, ensuring that the n -photon outcomes that we are interested in occupy a sufficiently large portion of probability weight over all possible outcomes.

Following our main strategy, to show the #P-hardness of $\sum |\text{PGPE}|_{\pm}^2$, one way is to find whether $\sum |\text{PGPE}|_{\pm}^2$ can be reduced from $|\text{GPE}|_{\pm}^2$, showing that solving $\sum |\text{PGPE}|_{\pm}^2$ is at least as hard as solving $|\text{GPE}|_{\pm}^2$. More precisely, we investigate whether average-case estimation of $q(n, x, X)$ is polynomially reducible from the average-case estimation of ideal output probability $q(X_{K_0}) = |\text{Per}(X_{K_0})|^2$, for some fixed subset of modes $K_0 \subseteq [N]$ with $|K_0| = n$, corresponding to the $|\text{GPE}|_{\pm}^2$ problem with input $X_{K_0} \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{n \times n}$. If this reduction can be established, it implies that such a noisy boson sampling is equivalently as hard to classically simulate as the ideal boson sampling. Since we have already established the reduction from $|\text{GPE}|_{\pm}^2$ to $|\text{PGPE}|_{\pm}^2$ for an $O(\log n)$ number of distinguishable photons in Sec. V, the remaining problem is to establish the reduction from $|\text{PGPE}|_{\pm}^2$ to $\sum |\text{PGPE}|_{\pm}^2$, i.e., to estimate $q(x, X_{K_0})$ in Eq. (4) for a fixed K_0 from estimated values of $q(x, \eta, X)$ in Eq. (34).

To this end, a natural approach is to set n close to the mean output photon number (such that $N \simeq n/\eta$), which always satisfies the requirement $\binom{N}{n} \eta^n (1 - \eta)^{N-n} = \Omega(\text{poly}(n)^{-1})$ for classical hardness. In this case, one can employ the previous result in Ref. [35] for lossy boson sampling. Specifically, for $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{n \times N}$, Ref. [35] has established a reduction from the average-case estimation of $q(X_{K_0}) = |\text{Per}(X_{K_0})|^2$ to the average-case estimation of $\binom{N}{n}^{-1} \sum_K q(X_K)$ (where $K \subseteq [N]$ with $|K| = n$) for $N = n + k'$, with imprecision blowup by at most $n^{O(k')}$. Hence, it is polynomially reducible when $k' = O(1)$, denoting that a constant number of lost photons maintains the computational complexity of lossy boson sampling. By extending their techniques to our case and setting $N \simeq n/\eta$, one

can similarly establish a reduction from the average-case estimation of $q(x, X_{K_0})$ to the average-case estimation of $q(x, \eta, X)$, with imprecision blowup by at most $n^{O(k')}$, where now $k' \simeq (1 - \eta)n/\eta$ denotes the average number of lost photons for transmission rate η . Here, the imprecision blowup is minimized when $\eta = \eta^*$, because $(1 - \eta)n/\eta \geq (1 - \eta^*)n/\eta^*$. Therefore, given that $(1 - \eta^*)n/\eta^* = O(1)$, which implies a constant number of lost photons on average, $\sum |\text{PGPE}|_{\pm}^2$ is polynomially reducible from $|\text{PGPE}|_{\pm}^2$. And combining our result, $\sum |\text{PGPE}|_{\pm}^2$ is polynomially reducible from $|\text{GPE}|_{\pm}^2$, for an $O(1)$ number of lost photons and an $O(\log n)$ number of distinguishable photons on average.

We remark that there is a much simpler way, as also noted in Ref. [35]. That is, one can instead set $n = N$, corresponding to the *lossless* output probability. More specifically, from Eq. (34), the output probability for $n = N$ is given by

$$q(x, \eta, X) = \eta^N q(x, X), \quad (35)$$

for $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$, and $q(x, X)$ is the partially distinguishable output probability in Eq. (4). Hence, one can directly estimate $q(x, X)$ by estimating $q(x, \eta, X)$, but with a multiplicative imprecision blowup by η^{-N} . Observe that η^{-N} is minimized when $\eta = \eta^*$ and when η^* satisfies $(1 - \eta^*)n = O(\log n)$, we have $(\eta^*)^{-N} = e^{O(\log n)} = \text{poly}(n)$. Indeed, this indicates an $(1 - \eta^*)n/\eta^* = O(\log n/\eta^*) = O(\log n)$ number of lost photons on average. Therefore, under an $O(\log n)$ number of losses, $\sum |\text{PGPE}|_{\pm}^2$ is polynomially reducible from $|\text{PGPE}|_{\pm}^2$, with the required imprecision level decreases multiplicatively by $\text{poly}(n)$ during the reduction. Combining this argument with Theorem 2 for, on average, an $O(\log n)$ number of lost photons and an $O(\log n)$ number of distinguishable photons, $\sum |\text{PGPE}|_{\pm}^2$ is polynomially reducible from $|\text{GPE}|_{\pm}^2$, given that $\epsilon, \delta = O(\text{poly}(n, \epsilon_0^{-1}, \delta_0^{-1})^{-1})$. Although this simple approach can yield a stronger hardness result, it cannot be extended to higher loss rates, since $(\eta^*)^{-N}$ becomes quasipolynomial in n outside the log-loss regime. Therefore, this exhibits clear limitations beyond the log-loss regime, and establishing hardness results under higher loss rates necessitates fundamentally different techniques.

We now conclude this section by formally summarizing the arguments so far.

Corollary 2. If $\mathcal{O}$ is an oracle that solves $\sum |\text{PGPE}|_{\pm}^2$ for $(1 - \eta^*)n = O(\log n)$ and $(1 - x^*)n = O(\log n)$, together with $\epsilon, \delta = O(\text{poly}(n, \epsilon_0^{-1}, \delta_0^{-1})^{-1})$, then $|\text{GPE}|_{\pm}^2$ can be solved in $\text{BPP}^{\mathcal{O}}$.

VII. CONCLUDING REMARKS

In this work, we have shown that an $O(\log N)$ number of distinguishable photons, on average, for N -photon

input maintains an equivalent complexity to the ideal boson-sampling case. Specifically, we have established a complexity-theoretical reduction from the problem of estimating the output probabilities of ideal boson sampling to the problem of estimating the output probabilities of noisy boson sampling with a logarithmic number of distinguishable photons on average. This implies that noisy boson sampling with an $O(\log N)$ number of distinguishable photons preserves the classical intractability of ideal boson sampling. We have also extended our result to the partially distinguishable and lossy boson sampling and found that an $O(\log N)$ number of lost photons and an $O(\log N)$ number of distinguishable photons upholds the classical intractability of ideal boson sampling.

Let us first remark on the barrier to using our main technique in the photon-loss case, i.e., noisy boson sampling subject to photon loss. Our main strategy is that, given access to the noisy output probability for different noise rates, we first conduct polynomial interpolation to infer the noisy output probability in terms of the noise rate and then obtain the ideal output probability by setting the noise rate to zero. Therefore, the requirement of the noisy output probability to use this approach is that it becomes the ideal output probability when the noise rate goes to zero. For the partial-distinguishability noise, the noisy output probability $q(x, X)$ in Eq. (4) satisfies this condition, because it exhibits the ideal output probability at $x = 1$. However, for the photon-loss case with transmission rate η , the noisy output probability to obtain the first n modes outcome for $N \geq n$ input photons (without postselection) is [22]

$$q(\eta, X) = \eta^n (1 - \eta)^{N-n} \sum_{\substack{K \subseteq [N] \\ |K|=n}} |\text{Per}(X_K)|^2, \quad (36)$$

which does not converge to the ideal output probability by only adjusting the transmission rate η , except for the output probability without any loss (i.e., $N = n$, which is the ideal output probability itself). Moreover, when we postselect n -photon outcomes as in Ref. [35], the noisy output probability loses its dependence on η and thus we cannot take our approach for this case either. Therefore, our main strategy does not work for the output probability of lossy boson sampling.

We also remark that in contrast to the photon-loss case, we cannot employ the “postselection trick” for the partial-distinguishability noise, i.e., postselecting noise-free events. For the photon-loss case, we can postselect no-loss outcomes by repeating the sampling process until the output photon number of the outcome is equal to the input photon number. As described in Sec. VI, the probability of postselecting a no-loss outcome for a number k' of lost photons on average for an input of N photons is $(1 - k'/N)^N$. Hence, via postselection, one can directly deduce that lossy boson sampling with at most a

$k' = O(\log N)$ number of lost photons on average maintains the computational complexity of the ideal boson sampling. On the other hand, in the case of partial-distinguishability noise, we cannot postselect the noise-free outcome, i.e., the outcome without any distinguishable photons. Instead, the output probability of partially distinguishable boson sampling is represented as a summation of output probabilities corresponding to a fixed number of distinguishable photons from 0 to N , as depicted in Eq. (4). Therefore, in contrast to the photon-loss case, we cannot directly obtain the classical-hardness result for an $O(\log N)$ number of noisy photons via postselection in the partial-distinguishability case.

We now conclude with a few open problems.

- (1) A crucial open problem that needs to be addressed is the extension of our result for a larger rate of noise. It is reasonable to expect that the experimental implementation of boson sampling generates at least $\Theta(N)$ noisy photons over N input photons. That is, a constant fraction of input photons suffer from noise and, for our case, this corresponds to $x = \Theta(1)$. In this regime, the low-degree approximation of the noisy output probability in Eq. (4), which was our main strategy for the proof of Theorem 2 in Sec. V, no longer holds. However, as long as we use the polynomial interpolation to infer the noiseless output probability, it is crucial to decrease the degree of the polynomial corresponding to the noisy output probability to at most $O(\log(\text{poly}(N, \epsilon_0^{-1}, \delta_0^{-1})))$ degree as we have described previously. Hence, a more advanced technique beyond our low-degree approximation has to be developed to extend our result to larger noise rates; we leave this problem as an open question.
- (2) Another open problem is investigating the classical simulation hardness of partially distinguishable Gaussian boson sampling. Gaussian boson sampling is a variant of boson sampling with Gaussian state inputs instead of Fock states, which is more suitable for experimental setups while having similar computational complexity [7–9,48]. Accordingly, recent experiments have been conducted on Gaussian boson sampling for the experimental demonstration of quantum computational advantage [10–13]. Therefore, to achieve quantum computational advantage via near-term Gaussian boson-sampling experiments, it is crucial to verify that noisy Gaussian boson sampling is classically hard. One way to do this is to find that the noisy Gaussian boson sampling is as hard as the ideal Gaussian boson-sampling case, similar to Ref. [35] and our main result in this work. For the partial-distinguishability noise, Ref. [20] has identified the noisy output probability of Gaussian boson

sampling when partial-distinguishability noise is applied. However, since our proof technique cannot be directly applied to the noisy output probability in Ref. [20], other proof techniques should be developed to deal with it.

- (3) One can also consider more realistic noise models, i.e., beyond the uniform distinguishability noise parametrized by a single parameter x as in Ref. [39]. One example is nonuniform distinguishability noise parametrized by $x_1, \dots, x_N$ for each of the N input photons. In this case, since the output probability cannot be represented as a polynomial of a single parameter [as x in Eq. (4)], our technique cannot be applied directly. One suboptimal approach that we can take is to arbitrarily increase each x_i value for all $i \in [N]$, to the maximum value among $x_1, \dots, x_N$. Then, by defining the maximum value as a parameter x , we can use the same process as in the main result and obtain the same result. Another approach is to employ the results in Refs. [21,25], which dealt with the classical simulability of noisy boson sampling subject to nonuniform physical noise including loss and distinguishability. Although those results are for classical simulation algorithms, extending their techniques to the hardness arguments of noisy boson sampling with nonuniform types of noise would be an interesting open problem.

ACKNOWLEDGMENTS

B.G. was supported by the education and training program of the Quantum Information Research Support Center, funded through the National Research Foundation of Korea (NRF) by the Ministry of Science and ICT (MSIT) of the Korean government (Grant No. 2021M3H3A1036573). B.G. and H.J. were supported by the Korean government (MSIT), the NRF grants funded by the Korean government (MSIT) (Grants No. RS-2024-00413957, No. RS-2024-00438415, and No. RS-2023-NR076733), and the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (Grants No. IITP-2025-RS-2020-II201606 and No. IITP-2025-RS-2024-00437191). C.O. was supported by the NRF (Grants No. RS-2024-00431768 and No. RS-2025-00515456), funded by the Korean government (MSIT), and IITP grants funded by the Korean government (MSIT) (Grants No. IITP-2025-RS-2025-02283189 and No. IITP-2025-RS-2025-02263264).

DATA AVAILABILITY

No data were created or analyzed in this study.

APPENDIX A: COMPARISON TO PREVIOUS RESULTS

In this appendix, we provide a detailed overview of the existing literature on the complexity of noisy boson sampling and compare the presented work to these findings to further clarify our contribution.

1. Our main result

We first summarize our main results, emphasizing our key assumption, main technique, and noise model. In the present work, we consider partial-distinguishability noise, one of the major noise sources that hinders experimental demonstration of quantum computational advantage. We adopt the uniform indistinguishability rate x as a noise model, which indicates that each photon becomes distinguishable with probability $1 - x$ [16,17,39]. Our key ingredient for the hardness result is that, for a fixed x^* of the partially distinguishable boson sampling, we consider the indistinguishability rate $x \in [0, x^*]$ as an input for that noisy boson sampler and derive the hardness result for that sampling problem. Then, under the plausible assumption that classical simulability does not degrade with decreasing x , this implies the classical hardness of simulating partially distinguishable boson sampling with a fixed indistinguishability rate x^* .

To summarize our result, we show that noisy N -single photon boson sampling that has a $(1 - x^*)N = O(\log N)$ number of distinguishable photons on average maintains the classical intractability of ideal boson sampling. Our main strategy is the *low-degree polynomial approximation* of noisy output probability, which is also a key technique for classical simulation studies for noisy boson sampling. We also generalize our result to the photon-loss noise and show that noisy boson sampling that unavoidably has, on average, an $O(\log N)$ number of lost photons and an $O(\log N)$ number of distinguishable photons maintains the classical intractability of ideal boson sampling.

2. Classical-hardness results for noisy boson sampling

Hereafter, we compare our results to the existing literature on the complexity of noisy boson sampling. One of the works closely related to ours is Ref. [38], which shows that partially distinguishable boson sampling, the single-photon mode mismatch of which is given by $O(N^{-3/2})$, still preserves the classical intractability of ideal boson sampling. Despite a slight difference in definition, the single-photon mode mismatch in Ref. [38] represents a single-photon distinguishability, which is essentially the same quantity as $1 - x$ in our work; this implies the classical hardness of partially distinguishable boson sampling for an $O(N^{-1/2})$ number of distinguishable photons on average. Also, in Ref. [35] it is shown that lossy boson

sampling, where at most a fixed $k = O(1)$ number of photons are lost over N input photons, maintains the same complexity as the ideal boson-sampling case.

The main strategy for these previous works is to show that the output probability estimation problem for the ideal case (i.e., the $|\text{GPE}\rangle_{\pm}^2$ problem in the main text) can also be solved by a noisy boson sampler under Stockmeyer's reduction, when the noise rate is below a certain threshold. In our work, we similarly adopt this strategy while improving the tolerable noise threshold for the classical intractability of noisy boson sampling; specifically, we increase the number of tolerable noisy photons for classical hardness to $O(\log N)$, now allowing the number of noisy photons to *scale* with the system size. Hence, by providing a significantly improved noise threshold for the classical hardness of boson sampling, our work paves the way toward realizing quantum computational advantage with noisy boson samplers.

3. Classical simulability results for partially distinguishable boson sampling

Another closely related line of work is the classical simulability results of noisy boson sampling with partially distinguishable photons. First of all, in Ref. [14] it is found that when the rate of Gaussian-type noise is sufficiently large, the output probability of that noisy boson sampling can be efficiently approximated by using a *low-degree polynomial approximation*. Reference [16] generalizes this low-degree polynomial approximation technique in Ref. [14] to partially distinguishable boson sampling with a uniform indistinguishability rate x and proposes a classical algorithm that approximately simulates partially distinguishable boson sampling. References [15,17] further generalize to the combined noise model for loss and distinguishability, characterized by uniform transmission rate η and uniform indistinguishability x , and, similarly employing the low-degree polynomial approximation, the authors propose a classical algorithm to approximately simulate that noisy boson sampling. Reference [18] also generalizes Ref. [14] by converting the Gaussian noise to realistic partial-distinguishability noise and suggests a noise threshold for efficient classical simulability of that noisy boson sampling. References [19,20] extend the analysis for the Gaussian boson sampling and also propose an efficient classical simulation for sufficiently small indistinguishability rates. Reference [21] generalizes these previous classical algorithms from uniform distinguishability noise to the nonuniform distinguishability cases.

These numerous simulability results suggest a noise threshold for the classical simulability of partially distinguishable boson sampling, providing a necessary condition for achieving quantum computational advantage. Here, the major technique that the authors have employed is a low-degree polynomial approximation, which can be achieved

by truncation of high-order terms in the noisy output probability. In our work, we employ a similar low-degree polynomial approximation strategy for noisy output probabilities but use it to establish the classical-hardness result for partially distinguishable boson sampling. Our result provides a noise threshold for the classical intractability of partially distinguishable boson sampling, thereby complementing existing simulability results and ultimately offering a sufficient condition for the noise rate of boson sampling to achieve quantum computational advantage.

4. Classical simulability results for noisy boson sampling with photon loss

We list here the classical simulability results for noisy boson sampling under photon loss. Reference [49] presents a classical simulation method based on quasiprobability distribution of continuous-variable quantum states, showing that boson sampling becomes classically simulable when the loss and dark-count rate are sufficiently large. After that, Refs. [22,23] show that for uniform photon loss, the lossy boson sampling becomes efficiently simulable when $O(\sqrt{N})$ photons survive over N input photons. Reference [24] reproduces this scaling behavior for the Gaussian boson-sampling case, showing the classical simulability of lossy Gaussian boson sampling with uniform loss when $O(\sqrt{N})$ photons survive over N input photons on average. Reference [25] generalizes these previous analyses to the nonuniform loss case. For the nonasymptotic case, by using the effect of photon loss on Gaussian boson sampling, Refs. [26–28] develop classical algorithms that can simulate the recent Gaussian boson-sampling experiments [10–13].

Although not directly related to our work, these works illustrate how noise in experiments can significantly reduce the computational complexity of boson sampling, thereby hindering the realization of quantum computational advantage. Hence, these simulability results highlight the necessity for further classical-hardness results of noisy boson sampling, i.e., improving the tolerable noise rates that maintain the computational complexity of boson sampling.

5. Classical simulability and hardness results for shallow-depth boson sampling

We now provide an overview of shallow-depth boson sampling. For the classical simulability results, Refs. [23, 50] present classical algorithms for simulating one-dimensional (1D) log-depth boson sampling based on matrix product state (MPS) methods, for geometrically local circuit architectures. Reference [51] extends these arguments, showing that 1D local log-depth Gaussian boson sampling can be efficiently simulated. Classical algorithms for shallow-depth boson sampling with general-dimensional local circuit architectures have also been developed, including boson sampling in Refs. [52,53]

and Gaussian boson sampling in Refs. [53,54]. Reference [30] proposes a classical algorithm for noisy constant-depth boson sampling with general types of noise, including the partial-distinguishability noise that we mainly consider.

Meanwhile, for the classical-hardness results of shallow-depth boson sampling, Ref. [55] proposes a log-depth circuit architecture with geometrically nonlocal gates, that can potentially be used for hardness results. Subsequently, Ref. [56] claims the average-case hardness for the log-depth circuit architecture proposed in Ref. [55] up to a certain imprecision level, providing hardness evidence of shallow-depth boson sampling for that circuit architecture.

Indeed, analysis on shallow-depth boson sampling is crucial for achieving quantum computational advantage, because shallow-depth circuits can significantly reduce the effect of noise. However, since the hardness results on noisy boson sampling (including ours) consider Haar-random linear optical circuits that require a deep circuit depth [57,58], they cannot be directly applied to those analyses of shallow-depth boson sampling. Hence, extending the existing hardness arguments for noisy boson sampling to shallow-depth circuit architectures would pave the way toward the ultimate demonstration of quantum computational advantage.

APPENDIX B: PROOF OF LEMMA 1

In this appendix, we prove Lemma 1, which states that for $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$ and $x \geq 1 - k_{\max}/N$, the truncated output probability $q^{(l)}(x, X)$ in Eq. (11) is ϵ_1 -close to $q(x, X)$ in Eq. (4), for ϵ_1 given by

$$\epsilon_1 = \delta^{-1} N^{-\frac{c_{\max}}{3}} \left(\frac{c_l}{c_{\max}} - 1 \right)^2, \quad (\text{B1})$$

over at least $1 - \delta$ of $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$.

Proof of Lemma 1. The error induced by truncation from $q(x, X)$ to $q^{(l)}(x, X)$ can be expressed as

$$q(x, X) - q^{(l)}(x, X) = \sum_{j=0}^{N-l-1} \binom{N}{j} x^j (1-x)^{N-j} q_j(X). \quad (\text{B2})$$

Here, given $l > k$, the coefficients in Eq. (B2) correspond to the tail of the binomial distribution. We first investigate the maximum size of these binomial tails in terms of l and k .

To examine the bound of the binomial tails, we use the Chernoff-Hoeffding bound. Specifically, for N independent Bernoulli random variables $X_1, \dots, X_N$ with each $X_i \in \{0, 1\}$, the sum of these variables $S_N = \sum_{i=1}^N X_i$ satisfies

the following inequality [59]:

$$\Pr[S_N \geq (1 + \xi)\mathbb{E}[S_N]] \leq \exp\left(-\frac{\xi^2}{3}\mathbb{E}[S_N]\right). \quad (\text{B3})$$

Using this bound, we can find the upper bound of the following tail terms:

$$\sum_{j=0}^{N-l-1} \binom{N}{j} x^j (1-x)^{N-j} = \sum_{j=l+1}^N \binom{N}{j} (1-x)^j x^{N-j}. \quad (\text{B4})$$

Specifically, by considering the success probability of the Bernoulli trial (i.e., the probability that $X_i = 1$) as $1-x$, the right-hand side of Eq. (B4) corresponds to $\Pr[S_N \geq l+1]$. Hence, by using $\mathbb{E}[S_N] = (1-x)N = k$ and setting $\xi = (l+1)/\mathbb{E}[S_N] - 1 = (l+1)/k - 1$, we have

$$\sum_{j=0}^{N-l-1} \binom{N}{j} x^j (1-x)^{N-j} = \Pr[S_N \geq l+1] \quad (\text{B5})$$

$$\leq \exp\left(-\frac{k}{3} \left(\frac{l+1}{k} - 1\right)^2\right) \quad (\text{B6})$$

$$\leq \exp\left(-\frac{k}{3} \left(\frac{l}{k} - 1\right)^2\right) \quad (\text{B7})$$

$$\leq \exp\left(-\frac{k_{\max}}{3} \left(\frac{l}{k_{\max}} - 1\right)^2\right) \quad (\text{B8})$$

$$= N^{-\frac{c_{\max}}{3} \left(\frac{c_l}{c_{\max}} - 1\right)^2}. \quad (\text{B9})$$

We now examine the size of $q_j(X)$ in Eq. (B2). For any $j = 0, \dots, N$, the average value of $q_j(X)$ in Eq. (5) over $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$ is

$$\begin{aligned} \mathbb{E}_X [q_j(X)] &= \binom{N}{j}^{-1} \sum_{\substack{I \subseteq [N] \\ |I|=j}} \sum_{\substack{J \subseteq [N] \\ |J|=j}} \mathbb{E}_X [|\text{Per}(X_{I,J})|^2 \text{Per}(|X_{I,J}|^2)] \\ &= \binom{N}{j}^{-1} \sum_{\substack{I \subseteq [N] \\ |I|=j}} \sum_{\substack{J \subseteq [N] \\ |J|=j}} j!(N-j)! \end{aligned} \quad (\text{B10})$$

$$= \binom{N}{j}^{-1} \sum_{\substack{I \subseteq [N] \\ |I|=j}} \sum_{\substack{J \subseteq [N] \\ |J|=j}} j!(N-j)! \quad (\text{B11})$$

$$= N!, \quad (\text{B12})$$

where the equality in Eq. (B11) comes from the fact that for any $I, J \subseteq [N]$ with $|I| = |J| = j$, $X_{I,J} \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{j \times j}$ and

$X_{I,J} \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N-j \times N-j}$ are IID Gaussians that are independent of each other. Also, for $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{N \times N}$, one can obtain $\mathbb{E}_X [|\text{Per}(X)|^2] = \mathbb{E}_X [\text{Per}(|X|^2)] = N!$ as shown in Ref. [6]. By the above equation and using the fact that $q(x, X) \geq q^{(l)}(x, X)$, one can find that

$$\begin{aligned} \mathbb{E}_X [q(x, X) - q^{(l)}(x, X)] &= \sum_{j=0}^{N-l-1} \binom{N}{j} x^j (1-x)^{N-j} \mathbb{E}_X [q_j(X)] \\ &= N! \sum_{j=0}^{N-l-1} \binom{N}{j} x^j (1-x)^{N-j}. \end{aligned} \quad (\text{B13})$$

$$= N! \sum_{j=0}^{N-l-1} \binom{N}{j} x^j (1-x)^{N-j}. \quad (\text{B14})$$

Finally, by using Markov's inequality, given the size of ϵ_1 as in Eq. (B1), observe that

$$\begin{aligned} \Pr_X [q(x, X) - q^{(l)}(x, X) > \epsilon_1 N!] &\leq \frac{1}{\epsilon_1 N!} \mathbb{E}_X [q(x, X) - q^{(l)}(x, X)] \\ &= \frac{1}{\epsilon_1} \sum_{j=0}^{N-l-1} \binom{N}{j} x^j (1-x)^{N-j} \end{aligned} \quad (\text{B15})$$

$$\leq \delta, \quad (\text{B16})$$

$$\leq \delta, \quad (\text{B17})$$

concluding the proof. $\blacksquare$

APPENDIX C: SPECIFYING $c_{\max}$ AND c_l

In this appendix, we investigate the appropriate size of the parameters $c_{\max}$ and c_l , first to make $\epsilon_1 = \delta^{-1} N^{-\frac{c_{\max}}{3} \left(\frac{c_l}{c_{\max}} - 1\right)^2} \ll \epsilon_0 N^{-c_{\max} - c_l(1 + \log \Delta^{-1})}$, while avoiding $\epsilon_0 N^{-c_{\max} - c_l(1 + \log \Delta^{-1})}$ becoming exceedingly small. To do so, we aim to find a minimum c_l that satisfies the first condition.

More specifically, to satisfy the first condition, the following quantity should be small, to at most a small constant:

$$\begin{aligned} \frac{\epsilon_1}{\epsilon_0 N^{-c_{\max} - c_l(1 + \log \Delta^{-1})}} &= \delta^{-1} \epsilon_0^{-1} N^{-\frac{c_{\max}}{3} \left(\frac{c_l}{c_{\max}} - 1\right)^2 + c_{\max} + c_l(1 + \log \Delta^{-1})} \\ &= O\left(\delta_0^{-1} \epsilon_0^{-1} N^{-\frac{c_{\max}}{3} \left(\frac{c_l}{c_{\max}} - 1\right)^2 + c_{\max} + c_l(1 + \log \Delta^{-1})}\right) \end{aligned} \quad (\text{C1})$$

$$= O\left(\delta_0^{-1} \epsilon_0^{-1} N^{-\frac{c_{\max}}{3} \left(\frac{c_l}{c_{\max}} - 1\right)^2 + c_{\max} + c_l(1 + \log \Delta^{-1})}\right) \quad (\text{C2})$$

$$= O\left(\delta_0^{-1} \epsilon_0^{-1} N^{-\frac{c_{\max}}{3} \left(\frac{c_l}{c_{\max}} - 1\right)^2 + c_{\max} + c_l(1 + \log \Delta^{-1}) + \frac{\log \epsilon_0^{-1} \delta_0^{-1}}{\log N}}\right), \quad (\text{C3})$$

where we have used $\delta^{-1} = O(\delta_0^{-1} l)$ in Eq. (C2), given that $\delta_0 = 1 - (1 - 2\delta)^{l+1}$.

Our approach is to first parametrize c_l in terms of $c_{\max}$ as $c_l = \alpha c_{\max}$ for $\alpha > 1$. Then, the exponent in Eq. (C3) can be expressed as a quadratic function of α as

$$-\frac{c_{\max}}{3} \left(\alpha^2 - (5 + 3 \log \Delta^{-1})\alpha - 2 - \frac{3 \log \epsilon_0^{-1} \delta_0^{-1}}{c_{\max} \log N} \right). \quad (\text{C4})$$

We denote α^* as a (positive) root of the above quadratic function, which is

$$\alpha^* = \frac{5 + 3 \log \Delta^{-1}}{2} + \sqrt{\left(\frac{5 + 3 \log \Delta^{-1}}{2} \right)^2 + 2 + \frac{3 \log \epsilon_0^{-1} \delta_0^{-1}}{c_{\max} \log N}}. \quad (\text{C5})$$

To prevent the right-hand side of Eq. (C3) from increasing with system size N , the exponent in Eq. (C3) should be negative and thus α should be larger than α^* . Here, we set α as

$$\alpha = \lambda \frac{5 + 3 \log \Delta^{-1}}{2} + \frac{3 \log \epsilon_0^{-1} \delta_0^{-1}}{c_{\max} \log N} \quad (\text{C6})$$

for a constant λ . One can check that when the constant λ is larger than a certain value, $\alpha > \alpha^*$ always holds, by using the following inequalities:

$$\frac{5 + 3 \log \Delta^{-1}}{2} + \sqrt{\left(\frac{5 + 3 \log \Delta^{-1}}{2} \right)^2 + 2 + \frac{3 \log \epsilon_0^{-1} \delta_0^{-1}}{c_{\max} \log N}} \quad (\text{C7})$$

$$\leq \frac{5 + 3 \log \Delta^{-1}}{2} + \sqrt{\left(\frac{5 + 3 \log \Delta^{-1}}{2} \right)^2 + 2} + \frac{3 \log \epsilon_0^{-1} \delta_0^{-1}}{c_{\max} \log N} \quad (\text{C8})$$

$$< \lambda \frac{5 + 3 \log \Delta^{-1}}{2} + \frac{3 \log \epsilon_0^{-1} \delta_0^{-1}}{c_{\max} \log N}, \quad (\text{C9})$$

where the inequality in Eq. (C9) holds for $\lambda \geq 2.149$ given that $(5 + 3 \log \Delta^{-1})/2 > 2.5$ [as $\log \Delta^{-1} = \log((c_{\max} + c_{\min})/(c_{\max} - c_{\min})) > 0$]. Since we aim to minimize α (to minimize c_l), we set α as given in Eq. (C6) with $\lambda = 2.149$ hereafter.

We now examine the scaling behavior of the right-hand side term in Eq. (C3), for the α value that we have set.

More specifically, we find that

$$N^{-\frac{c_{\max}}{3} \left(\alpha^2 - (5 + 3 \log \Delta^{-1})\alpha - 2 - \frac{3 \log \epsilon_0^{-1} \delta_0^{-1}}{c_{\max} \log N} \right)} \quad (\text{C10})$$

$$= N^{-\frac{c_{\max}}{3} \left((2\alpha^* - 5 - 3 \log \Delta^{-1})(\alpha - \alpha^*) + (\alpha - \alpha^*)^2 \right)} \quad (\text{C11})$$

$$\leq N^{-\frac{c_{\max}}{3} \left((2\alpha^* - 5 - 3 \log \Delta^{-1})(\alpha - \alpha^*) \right)} \quad (\text{C12})$$

$$= N^{-\frac{2c_{\max}}{3} (\alpha - \alpha^*) \sqrt{\left(\frac{5 + 3 \log \Delta^{-1}}{2} \right)^2 + 2 + \frac{3 \log \epsilon_0^{-1} \delta_0^{-1}}{c_{\max} \log N}}}, \quad (\text{C13})$$

where we have used Taylor expansion of the quadratic function in Eq. (C11). Here, from the definition of α and α^* , one can check that $\alpha - \alpha^* = \Omega(1 + \log \epsilon_0^{-1} \delta_0^{-1} / \log N)$. Then the right-hand side of Eq. (C13) scales inverse polynomially with N and $\epsilon_0^{-1} \delta_0^{-1}$. Also, we have

$$l = c_l \log N \quad (\text{C14})$$

$$= \alpha c_{\max} \log N \quad (\text{C15})$$

$$= O(\log N + \log \epsilon_0^{-1} \delta_0^{-1}), \quad (\text{C16})$$

such that l scales logarithmically with N and $\epsilon_0^{-1} \delta_0^{-1}$. Combining these results, the right-hand side term in Eq. (C3) scales inverse polynomially with N and $\epsilon_0^{-1} \delta_0^{-1}$ up to a logarithmic factor; thus, from Eq. (C3), we have

$$\frac{\epsilon_1}{\epsilon_0 N^{-c_{\max} - c_l(1 + \log \Delta^{-1})}} = O(\text{poly}(N, \epsilon_0^{-1} \delta_0^{-1})^{-1}). \quad (\text{C17})$$

Next, we parametrize $c_{\max}$ in terms of $c_{\min}$ as $c_{\max} = \kappa c_{\min}$ for a constant κ . Then, $\Delta = (c_{\max} - c_{\min})/(c_{\max} + c_{\min}) = (\kappa - 1)/(\kappa + 1)$, and we have

$$\epsilon_0 N^{-c_{\max} - c_l(1 + \log \Delta^{-1})} = \epsilon_0 N^{-(1 + (1 + \log \Delta^{-1})\alpha)c_{\max}} \quad (\text{C18})$$

$$= \epsilon_0 N^{-\left(1 + (1 + \log \Delta^{-1}) \left(\lambda \frac{5 + 3 \log \Delta^{-1}}{2} + \frac{3 \log \epsilon_0^{-1} \delta_0^{-1}}{c_{\max} \log N} \right) \right) c_{\max}} \quad (\text{C19})$$

$$= \epsilon_0^{\frac{4 + 3 \log \frac{\kappa + 1}{\kappa - 1}}{\kappa - 1}} \delta_0^{\frac{3 + 3 \log \frac{\kappa + 1}{\kappa - 1}}{\kappa - 1}} \times N^{-\kappa \left(1 + \frac{\lambda}{2} \left(1 + \log \frac{\kappa + 1}{\kappa - 1} \right) \left(5 + 3 \log \frac{\kappa + 1}{\kappa - 1} \right) \right) c_{\min}}. \quad (\text{C20})$$

Since we have freedom to choose the constant κ , we choose $\kappa = 2.108$, which maximizes the exponent of N in Eq. (C20) to $-39.35 c_{\min}$.

To sum up, by setting $c_{\max}$ and c_l as

$$c_{\max} = \kappa c_{\min}, \quad (\text{C21})$$

$$c_l = \kappa \lambda \frac{5 + 3 \log \frac{\kappa+1}{\kappa-1}}{2} c_{\min} + \frac{3 \log \epsilon_0^{-1} \delta_0^{-1}}{\log N}, \quad (\text{C22})$$

with $\kappa = 2.108$ and $\lambda = 2.149$, we have

$$\epsilon_0 N^{-c_{\max}-c_l(1+\log \Delta^{-1})} - \epsilon_1 = O\left(\epsilon_0 N^{-c_{\max}-c_l(1+\log \Delta^{-1})}\right), \quad (\text{C23})$$

and the exponents in the right-hand side of Eq. (C23) can be specified as

$$\epsilon_0 N^{-c_{\max}-c_l(1+\log \Delta^{-1})} = \epsilon_0^{7.094} \delta_0^{6.094} N^{-39.35 c_{\min}}. \quad (\text{C24})$$

Note that one can further optimize Eq. (C24), by choosing α in Eq. (C6) more closely to α^* in Eq. (C5) while satisfying Eq. (C17). But, for our purpose, this result is sufficient.

APPENDIX D: PROOF OF COROLLARY 1: REDUCING $|\text{GPE}|_{\pm}^2$ TO $|\text{PGPE}|_{\pm}^2$ FOR $k_{\min} = o(\log N)$

To prove Corollary 1, we reproduce Lemma 1 for sublogarithmically scaling $k_{\min} = o(\log N)$. Similarly to our main result in Sec. V, let $x_{\min}$ be a minimum value of x that we set, such that $x_{\min} = 1 - k_{\max}/N$ with $k_{\max} = o(\log N)$. Specifically, $k_{\max}$ is larger than $k_{\min}$ and also proportional to $k_{\min}$ such that $k_{\max}/k_{\min} = \Theta(1)$.

We first examine the size of the binomial term in the noisy output probability $q(x, X)$ in Eq. (4), for $j = N - l$ with l larger than $k_{\max}$. Specifically, for any input indistinguishability rate $x = 1 - k/N$ with $k \in [k_{\min}, k_{\max}]$, the binomial term corresponding to $j = N - l$ is bounded by

$$\begin{aligned} & \binom{N}{N-l} x^{N-l} (1-x)^l \\ &= \binom{N}{N-l} \left(1 - \frac{k}{N}\right)^{N-l} \left(\frac{k}{N}\right)^l \end{aligned} \quad (\text{D1})$$

$$< \frac{1}{\sqrt{l}} \frac{N^N}{l^l (N-l)^{N-l}} \left(1 - \frac{k}{N}\right)^{N-l} \left(\frac{k}{N}\right)^l \quad (\text{D2})$$

$$= \frac{1}{\sqrt{l}} \left(\frac{N-k}{N-l}\right)^{N-l} \left(\frac{k}{l}\right)^l \quad (\text{D3})$$

$$< \frac{1}{\sqrt{l}} e^{-l \log(\frac{l}{k}) + l - k} \quad (\text{D4})$$

$$< e^{-l \log(\frac{l}{k}) + l - k} \quad (\text{D5})$$

$$\leq e^{-l \log(\frac{l}{k_{\max}}) + l - k_{\max}}, \quad (\text{D6})$$

where we have used Stirling's inequality in Eq. (D2) and we have used the following relation in Eq. (D4):

$$\begin{aligned} & (N-l) \log \left(\frac{N-k}{N-l}\right) \\ &= (N-l) \left(\log \left(1 - \frac{k}{N}\right) - \log \left(1 - \frac{l}{N}\right)\right) \end{aligned} \quad (\text{D7})$$

$$= (N-l) \sum_{m=1}^{\infty} \frac{1}{m!} \frac{l^m - k^m}{N^m} \quad (\text{D8})$$

$$= l - k + \sum_{m=1}^{\infty} \frac{l^{m+1} - k^{m+1} - (m+1)l(l^m - k^m)}{(m+1)!N^m} \quad (\text{D9})$$

$$\leq l - k + \sum_{m=1}^{\infty} \frac{-l^{m+1} + 2lk^m - k^{m+1}}{(m+1)!N^m} \quad (\text{D10})$$

$$= l - k - \sum_{m=1}^{\infty} \frac{k^{m+1}}{(m+1)!N^m} \left(\left(\frac{l}{k}\right)^{m+1} - 2\left(\frac{l}{k}\right) + 1 \right) \quad (\text{D11})$$

$$< l - k, \quad (\text{D12})$$

where we have used Taylor expansion in Eq. (D8) and we have used $m \geq 1$ and $l > k$ in Eqs. (D10) and (D12).

By the property of the binomial distribution, the binomial term $\binom{N}{j} x^j (1-x)^{N-j}$ monotonically decreases with decreasing j , as long as $j < Nx$. Hence, for any $x = 1 - k/N$, all the binomial terms corresponding to $j < N - l$ are smaller than the binomial term corresponding to $j = N - l$, given that l is larger than $k_{\max}$. Accordingly, the summation of the binomial terms from $j = 0$ to $j = N - l - 1$ is upper bounded as

$$\begin{aligned} & \sum_{j=0}^{N-l-1} \binom{N}{j} x^j (1-x)^{N-j} \\ & \leq (N-l) \cdot \binom{N}{N-l} x^{N-l} (1-x)^l \end{aligned} \quad (\text{D13})$$

$$< N e^{-l \log(\frac{l}{k_{\max}}) + l - k_{\max}}. \quad (\text{D14})$$

Then, similarly using the proof of Lemma 1, $q(x, X)$ and $q^{(l)}(x, X)$ are $\epsilon_1 N!$ -close over at least $1 - \delta$ of X , for $\epsilon_1 = \delta^{-1} N e^{-l \log(\frac{l}{k_{\max}}) + l - k_{\max}}$, i.e.,

$$\Pr_X \left[|q(x, X) - q^{(l)}(x, X)| > \epsilon_1 N! \right] < \delta. \quad (\text{D15})$$

By using this property and following the proof in Sec. VD, the proof of Corollary 1 is now straightforward.

Proof of Corollary 1. The proof is essentially the same as the proof of Theorem 2 except for some minor changes

in the size of the parameters. Let $\mathcal{O}$ be the oracle that, on input x and $X \sim \mathcal{N}(0, 1)^{N \times N}$, estimates $q(x, X)$ within $\epsilon N!$ over $1 - \delta$ of X , such that

$$\Pr_X[|\mathcal{O}(x, X) - q(x, X)| > \epsilon N!] < \delta. \quad (\text{D16})$$

For $\epsilon' = \epsilon + \epsilon_1$, using the triangular inequality,

$$\begin{aligned} \Pr_X[|\mathcal{O}(x, X) - q^{(l)}(x, X)| > \epsilon' N!] \\ \leq \Pr_X[|\mathcal{O}(x, X) - q(x, X)| > \epsilon N!] \\ + \Pr_X[|q(x, X) - q^{(l)}(x, X)| > \epsilon_1 N!] \end{aligned} \quad (\text{D17})$$

$$< 2\delta. \quad (\text{D18})$$

Let $\epsilon'' = \epsilon' e^{k_{\max}}$ such that $\epsilon' x^{-N+l} \leq \epsilon' x_{\min}^{-N+l} = \epsilon' e^{k_{\max} - \tilde{O}(N^{-1})} \leq \epsilon''$. Then, for $f_X^{(l)}(x) = x^{-N+l} q^{(l)}(x, X)$,

$$\begin{aligned} \Pr_X[|\mathcal{O}(x, X) x^{-N+l} - f_X^{(l)}(x)| > \epsilon'' N!] \\ \leq \Pr_X[|\mathcal{O}(x, X) x^{-N+l} - f_X^{(l)}(x)| > \epsilon' x^{-N+l} N!] \end{aligned} \quad (\text{D19})$$

$$= \Pr_X[|\mathcal{O}(x, X) - q^{(l)}(x, X)| > \epsilon' N!] \quad (\text{D20})$$

$$< 2\delta. \quad (\text{D21})$$

As Eq. (D21) is the same as Eq. (21) in the proof of Theorem 2, we can repeat all the steps identically to the proof of Theorem 2 and obtain the estimation value $|\text{Per}(X)|^2$ within $\epsilon'' e^{l(1+\log \Delta^{-1})} N!$ with probability at least $1 - (1 - 2\delta)^{l+1}$ [where $\Delta = (k_{\max} - k_{\min}) / (k_{\max} + k_{\min}) = \Theta(1)$]. Then, the conditions for the error parameters ϵ and δ are $\delta_0 \geq 1 - (1 - 2\delta)^{l+1}$ and $\epsilon_0 \geq \epsilon'' e^{l(1+\log \Delta^{-1})} = (\epsilon + \epsilon_1) e^{k_{\max}} e^{l(1+\log \Delta^{-1})}$ (in other words, $\epsilon \leq \epsilon_0 e^{-k_{\max} - l(1+\log \Delta^{-1})} - \epsilon_1$). Hence, similarly as to before, we set $\delta_0 = 1 - (1 - 2\delta)^{l+1}$, such that $\delta = O(l^{-1} \delta_0)$.

Here, note that the approximation error ϵ in the $|\text{PGPE}|^2_{\pm}$ problem is positive. Therefore, as we have previously discussed in Sec. VD, ϵ_1 should satisfy $\epsilon_1 \ll \epsilon_0 e^{-k_{\max} - l(1+\log \Delta^{-1})}$ such that $\epsilon_0 e^{-k_{\max} - l(1+\log \Delta^{-1})} - \epsilon_1 = O(\epsilon_0 e^{-k_{\max} - l(1+\log \Delta^{-1})})$. This means that the following quantity should be small enough:

$$\begin{aligned} \frac{\epsilon_1}{\epsilon_0 e^{-k_{\max} - l(1+\log \Delta^{-1})}} \\ = \epsilon_0^{-1} \delta^{-1} N e^{-l \log(\frac{l}{k_{\max}}) + l(2+\log \Delta^{-1})} \end{aligned} \quad (\text{D22})$$

$$= O\left(l e^{-l \log(\frac{l}{k_{\max}}) + l(2+\log \Delta^{-1}) + \log N + \log \epsilon_0^{-1} \delta_0^{-1}}\right), \quad (\text{D23})$$

where we have used $\delta^{-1} = O(l \delta^{-1})$ from $\delta_0 = 1 - (1 - 2\delta)^{l+1}$ in Eq. (D23). To make the right-hand side of

Eq. (D23) small, we choose l to scale logarithmically with N and $\epsilon_0^{-1} \delta_0^{-1}$, namely

$$l = a_0 \log N + a_1 \log \epsilon_0^{-1} \delta_0^{-1}, \quad (\text{D24})$$

for constants $a_0, a_1 > 0$. Note that the right-hand side of Eq. (D23) scales as $o(\text{poly}(N, \epsilon_0^{-1} \delta_0^{-1})^{-1})$ even for arbitrarily small constants a_0 and a_1 , because the leading term scales as $e^{-l \log l}$. Hence, for l given in Eq. (D24) with any $a_0, a_1 > 0$, $\epsilon_0 e^{-k_{\max} - l(1+\log \Delta^{-1})} - \epsilon_1 = O(\epsilon_0 e^{-k_{\max} - l(1+\log \Delta^{-1})})$.

Therefore, using l in Eq. (D24), $\delta = O(l^{-1} \delta_0) = O(\delta_0 (\log N + \log \epsilon_0^{-1} \delta_0^{-1})^{-1})$. Also, the condition for ϵ reduces to

$$\epsilon \leq O\left(\epsilon_0 e^{-k_{\max} - l(1+\log \Delta^{-1})}\right) \quad (\text{D25})$$

$$= O\left(\frac{\epsilon_0^{1+a_1(1+\log \Delta^{-1})} \delta_0^{a_1(1+\log \Delta^{-1})}}{N^{a_0(1+\log \Delta^{-1})}} e^{-k_{\max}}\right). \quad (\text{D26})$$

Because a_0 and a_1 can be made arbitrarily small constants, setting $\epsilon = O(\epsilon_0^{\alpha+1} \delta_0^{\alpha} N^{-\alpha})$ for any constant $\alpha > 0$ is sufficient to satisfy the condition in Eq. (D26). This completes the proof. $\blacksquare$

-
- [1] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
 - [2] R. P. Feynman, Simulating physics with computers, *Int. J. Theor. Phys.* **21**, 467 (1982).
 - [3] S. Lloyd, Universal quantum simulators, *Science* **273**, 1073 (1996).
 - [4] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM Rev.* **41**, 303 (1999).
 - [5] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani, On the complexity and verification of quantum random circuit sampling, *Nat. Phys.* **15**, 159 (2019).
 - [6] S. Aaronson and A. Arkhipov, in *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing* (ACM, New York, 2011), pp. 333–342.
 - [7] C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, Gaussian boson sampling, *Phys. Rev. Lett.* **119**, 170501 (2017).
 - [8] A. Deshpande, A. Mehta, T. Vincent, N. Quesada, M. Hinsche, M. Ioannou, L. Madsen, J. Lavoie, Ha. Qi, J. Eisert *et al.*, Quantum computational advantage via high-dimensional Gaussian boson sampling, *Sci. Adv.* **8**, eabi7894 (2022).
 - [9] D. Grier, D. J. Brod, J. Miguel Arrazola, M. Benicio de Andrade Alonso, and N. Quesada, The complexity of bipartite Gaussian boson sampling, *Quantum* **6**, 863 (2022).
 - [10] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu *et al.*, Quantum computational advantage using photons, *Science* **370**, 1460 (2020).

- [11] H.-S. Zhong, Y.-H. Deng, J. Qin, H. Wang, M.-C. Chen, L.-C. Peng, Y.-H. Luo, D. Wu, S.-Q. Gong, H. Su *et al.*, Phase-programmable Gaussian boson sampling using stimulated squeezed light, *Phys. Rev. Lett.* **127**, 180502 (2021).
- [12] L. S. Madsen, F. Laudenbach, M. Falamarzi Askarani, F. Rortais, T. Vincent, J. F. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins *et al.*, Quantum computational advantage with a programmable photonic processor, *Nature* **606**, 75 (2022).
- [13] Y.-H. Deng, Y.-C. Gu, H.-L. Liu, S.-Q. Gong, H. Su, Z.-J. Zhang, H.-Y. Tang, M.-H. Jia, J.-M. Xu, M.-C. Chen *et al.*, Gaussian boson sampling with pseudo-photon-number-resolving detectors and quantum computational advantage, *Phys. Rev. Lett.* **131**, 150601 (2023).
- [14] G. Kalai and G. Kindler, Gaussian noise sensitivity and bosonsampling, [arXiv:1409.3093](https://arxiv.org/abs/1409.3093).
- [15] J. Renema, V. Shchesnovich, and R. Garcia-Patron, Classical simulability of noisy boson sampling, [arXiv:1809.01953](https://arxiv.org/abs/1809.01953).
- [16] J. J. Renema, A. Menssen, W. R. Clements, G. Triginer, W. S. Kolthammer, and I. A. Walmsley, Efficient classical algorithm for boson sampling with partially distinguishable photons, *Phys. Rev. Lett.* **120**, 220502 (2018).
- [17] A. E. Moylett, R. García-Patrón, J. J. Renema, and P. S. Turner, Classically simulating near-term partially-distinguishable and lossy boson sampling, *Quantum Sci. Technol.* **5**, 015001 (2019).
- [18] V. S. Shchesnovich, Noise in boson sampling and the threshold of efficient classical simulatability, *Phys. Rev. A* **100**, 012340 (2019).
- [19] J. J. Renema, Simulability of partially distinguishable superposition and Gaussian boson sampling, *Phys. Rev. A* **101**, 063840 (2020).
- [20] J. Shi and T. Byrnes, Effect of partial distinguishability on quantum supremacy in Gaussian boson sampling, *npj Quantum Inf.* **8**, 54 (2022).
- [21] S. N. van den Hoven, E. Kanis, and J. J. Renema, Efficient classical algorithm for simulating boson sampling with heterogeneous partial distinguishability, *Phys. Rev. A* **111**, 052448 (2025).
- [22] M. Oszmaniec and D. J. Brod, Classical simulation of photonic linear optics with lost particles, *New J. Phys.* **20**, 092002 (2018).
- [23] R. García-Patrón, J. J. Renema, and V. Shchesnovich, Simulating boson sampling in lossy architectures, *Quantum* **3**, 169 (2019).
- [24] H. Qi, D. J. Brod, N. Quesada, and R. García-Patrón, Regimes of classical simulability for noisy Gaussian boson sampling, *Phys. Rev. Lett.* **124**, 100502 (2020).
- [25] D. J. Brod and M. Oszmaniec, Classical simulation of linear optics subject to nonuniform losses, *Quantum* **4**, 267 (2020).
- [26] B. Villalonga, M. Yuezhen Niu, L. Li, H. Neven, J. C. Platt, V. N. Smelyanskiy, and S. Boixo, Efficient approximation of experimental Gaussian boson sampling, [arXiv:2109.11525](https://arxiv.org/abs/2109.11525).
- [27] J. F. F. Bulmer, B. A. Bell, R. S. Chadwick, A. E. Jones, D. Moise, A. Rigazzi, J. Thorbecke, U.-U. Haus, T. Van Vaerenbergh, R. B. Patel *et al.*, The boundary for quantum advantage in Gaussian boson sampling, *Sci. Adv.* **8**, eabl9236 (2022).
- [28] C. Oh, M. Liu, Y. Alexeev, B. Fefferman, and L. Jiang, Classical algorithm for simulating experimental Gaussian boson sampling, *Nat. Phys.* **20**, 1461 (2024).
- [29] C. Oh, L. Jiang, and B. Fefferman, On classical simulation algorithms for noisy boson sampling, [arXiv:2301.11532](https://arxiv.org/abs/2301.11532).
- [30] C. Oh, Classical simulability of constant-depth linear-optical circuits with noise, *npj Quantum Inf.* **11**, 126 (2025).
- [31] C. Oh, Recent theoretical and experimental progress on boson sampling, *Curr. Opt. Photonics* **9**, 1 (2025).
- [32] A. Bouland, B. Fefferman, Z. Landau, and Y. Liu, in *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, New York, 2022), pp. 1308–1317.
- [33] A. Bouland, D. Brod, I. Datta, B. Fefferman, D. Grier, F. Hernandez, and M. Oszmaniec, Complexity-theoretic foundations of bosonsampling with a linear number of modes, [arXiv:2312.00286](https://arxiv.org/abs/2312.00286).
- [34] A. Bouland, I. Datta, B. Fefferman, and F. Hernandez, On the average-case hardness of bosonsampling, [arXiv:2411.04566](https://arxiv.org/abs/2411.04566).
- [35] S. Aaronson and D. J. Brod, Bosonsampling with lost photons, *Phys. Rev. A* **93**, 012335 (2016).
- [36] W. Chen, Y. Lu, S. Zhang, K. Zhang, G. Huang, M. Qiao, X. Su, J. Zhang, J.-N. Zhang, L. Bianchi *et al.*, Scalable and programmable phononic network with trapped ions, *Nat. Phys.* **19**, 877 (2023).
- [37] A. W. Young, S. Geller, W. J. Eckner, N. Schine, S. Glancy, E. Knill, and A. M. Kaufman, An atomic boson sampler, *Nature* **629**, 311 (2024).
- [38] V. S. Shchesnovich, Sufficient condition for the mode mismatch of single photons for scalability of the boson-sampling computer, *Phys. Rev. A* **89**, 022333 (2014).
- [39] M. C. Tichy, Sampling of partially distinguishable bosons and the relation to the multidimensional permanent, *Phys. Rev. A* **91**, 022316 (2015).
- [40] C.-K. Hong, Z.-Y. Ou, and L. Mandel, Measurement of subpicosecond time intervals between two photons by interference, *Phys. Rev. Lett.* **59**, 2044 (1987).
- [41] M. Jerrum, A. Sinclair, and E. Vigoda, A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries, *J. Assoc. Comput. Mach.* **51**, 671 (2004).
- [42] L. Stockmeyer, On approximation algorithms for # P, *SIAM J. Comput.* **14**, 849 (1985).
- [43] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Hugrins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, *Rev. Mod. Phys.* **95**, 045005 (2023).
- [44] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, *Phys. Rev. Lett.* **119**, 180509 (2017).
- [45] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, *Phys. Rev. X* **7**, 021050 (2017).
- [46] R. Paturi, in *Proceedings of the 24th Annual ACM Symposium on Theory of Computing (STOC '92)*, Victoria, British Columbia, Canada (Association for Computing Machinery, New York, 1992), pp. 468–474.

- [47] Y. Kondo, R. Mori, and R. Movassagh, in *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Denver, CO, USA, 2022), pp. 1296–1307.
- [48] R. Kruse, C. S. Hamilton, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, Detailed study of Gaussian boson sampling, *Phys. Rev. A* **100**, 032326 (2019).
- [49] S. Rahimi-Keshari, T. C. Ralph, and C. M. Caves, Sufficient conditions for efficient classical simulation of quantum optics, *Phys. Rev. X* **6**, 021039 (2016).
- [50] G. Vidal, Efficient classical simulation of slightly entangled quantum computations, *Phys. Rev. Lett.* **91**, 147902 (2003).
- [51] H. Qi, D. Cifuentes, K. Brádler, R. Israel, T. Kalajdzievski, and N. Quesada, Efficient sampling from shallow Gaussian quantum-optical circuits with local interactions, *Phys. Rev. A* **105**, 052412 (2022).
- [52] A. Deshpande, B. Fefferman, M. C. Tran, M. Foss-Feig, and A. V. Gorshkov, Dynamical phase transitions in sampling complexity, *Phys. Rev. Lett.* **121**, 030501 (2018).
- [53] C. Oh, Y. Lim, B. Fefferman, and L. Jiang, Classical simulation of boson sampling based on graph structure, *Phys. Rev. Lett.* **128**, 190501 (2022).
- [54] Z. Kolarovszki, Á. Kaposi, T. Kozsik, and Z. Zimborás, in *International Conference on Computational Science* (Springer, Prague, 2023), pp. 209–223.
- [55] B. Go, C. Oh, L. Jiang, and H. Jeong, Exploring shallow-depth boson sampling: Toward a scalable quantum advantage, *Phys. Rev. A* **109**, 052613 (2024).
- [56] B. Go, C. Oh, and H. Jeong, On computational complexity and average-case hardness of shallow-depth boson sampling. [arXiv:2405.01786](https://arxiv.org/abs/2405.01786).
- [57] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. Steven Kolthammer, and I. A. Walmsley, Optimal design for universal multiport interferometers, *Optica* **3**, 1460 (2016).
- [58] N. J. Russell, L. Chakhmakhchyan, J. L. O’Brien, and A. Laing, Direct dialling of Haar random unitary matrices, *New J. Phys.* **19**, 033007 (2017).
- [59] D. P. Dubhashi and A. Panconesi, *Concentration of Measure for the Analysis of Randomized Algorithms* (Cambridge University Press, New York, 2009).