BosonSampling with lost photons

Scott Aaronson and Daniel J. Brod

BosonSampling with lost photons

Scott Aaronson[*]

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Daniel J. Brod[†]

Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, ON, N2L 2Y5, Canada

(Dated: November 30, 2015)

BosonSampling is an intermediate model of quantum computation where linear-optical networks are used to solve sampling problems expected to be hard for classical computers. Since these devices are not expected to be universal for quantum computation, it remains an open question of whether any error-correction techniques can be applied to them, and thus it is important to investigate how robust the model is under natural experimental imperfections, such as losses and imperfect control of parameters. Here we investigate the complexity of BosonSampling under photon losses—more specifically, the case where an unknown subset of the photons is randomly lost at the sources. We show that, if k out of n photons are lost, then we cannot sample classically from a distribution that is $1/n^{\Theta(k)}$-close (in total variation distance) to the ideal distribution, unless a BPP^{NP} machine can estimate the permanents of Gaussian matrices in $n^{O(k)}$ time. In particular, if k is constant, this implies that simulating lossy BosonSampling is hard for a classical computer, under exactly the same complexity assumption used for the original lossless case.

I. INTRODUCTION

BOSONSAMPLING is a computational problem that involves sampling from a certain kind of probability distribution, defined in terms of the permanents of matrices. While it does not have any obvious applications, the BOSONSAMPLING problem has two features of interest: (i) it is "easy" to solve by a restricted, single-purpose quantum computer composed only of linear-optical elements, and (ii) it is hard to solve on a classical computer [1], under plausible complexity-theoretic assumptions. Its natural implementation as a linear-optical experiment led, since the initial proposal, to a flurry of experiments of increasing complexity [2–9].

The fact that these single-purpose linear-optical devices are not expected to be universal for quantum computing also means that they are not likely to have access to the usual tools of fault-tolerance and error correction, and thus it is important to investigate the resilience of the underlying models to real-world experimental imperfections. The original proposal of [1] already focused on this issue, giving evidence that BOSONSAMPLING is hard even if the classical computer is allowed to sample from a distribution that only approximates the ideal one (in variation distance). However, this does not suffice to address all realistic sources of error and imperfections that one may encounter. Thus, since then, further work has been done to investigate the robustness of BOSONSAM-PLING devices to errors such as losses and mode mismatch [10, 11], imperfections on the linear-optical network [12, 13], and other sources of noise [14].

There are a few reasons why photon losses are a central scalability issue in BOSONSAMPLING devices. The first is that they are pervasive in linear-optical experiments—to give an example, most recent BOSONSAMPLING experiments were performed using integrated photonic devices, where rather than setting up a linear-optical circuit using beam splitters in free space, the photons propagate along waveguides etched in a small chip. The most realistic models of loss in these devices give a per-photon probability of arrival that decreases exponentially with the depth of the circuit, and this is even before taking into account source and detector inefficiencies. The second reason is that often losses are dealt with using post-selection. By only accepting the experimental runs where the correct number of photons is observed, one can argue that the losses manifest only as an attenuation of the number of observed events per unit time. The problem with this approach, of course, is that even if each photon has only a constant probability of being lost, this already leads to an exponential overhead in experimental time as the desired number of photons increases.

In this paper, we give a formal treatment of the problem of BOSONSAMPLING with a few lost photons. More specifically, we show how the original argument of [1] can be modified to allow for the loss of a constant number (i.e. $O(1)$) of photons, thus showing that BOSONSAMPLING in this regime has the same complexity-theoretic status as in the ideal case. Of course, this falls short of the result one would want, since in practice one expects that at least a constant fraction (i.e. $\Theta(n)$) of the photons will be lost. Nevertheless, the result here constitutes a nontrivial step toward understanding experimentally-realistic BOSONSAMPLING regimes.

Let us point out in particular that, if more than $k = O(1)$ photons are lost, then we still get *some* hard-

[*] aaronson@csail.mit.edu
[†] dbrod@perimeterinstitute.ca

ness result: the issue is "just" that the strength of the result degrades rapidly as a function of k. More concretely, if k out of n photons are lost, then our argument will imply that BOSONSAMPLING cannot be done in classical polynomial time, to within a $1/n^{\Theta(k)}$ error in total variation distance, assuming that permanents of Gaussian matrices cannot be estimated in $n^{O(k)}$ time. So for example, even when a (sufficiently small) constant fraction of the photons is lost, $k = \epsilon n$, we can conclude that *exact* BOSONSAMPLING cannot be done in classical polynomial time, assuming that Gaussian permanent estimation is exponentially hard.

We also point out, in the appendix, how the proof of our main result can be repurposed to other physically-interesting error models, such as BOSONSAMPLING with dark counts, or a combination of both losses and dark counts.

Notation: Throughout this paper, we denote by $\mathcal{N}(0,1)_{\mathbb{C}}$ the complex Gaussian distribution with mean 0 and variance 1, and by $\mathcal{N}(0,1)_{\mathbb{C}}^{m \times n}$ the distribution over $m \times n$ matrices of i.i.d. complex Gaussian entries. For two probability distributions $\mathcal{D}_1 = \{p_x\}_x$ and $\mathcal{D}_2 = \{q_x\}_x$, the total variation distance $||\mathcal{D}_1 - \mathcal{D}_2||$ and the Kullback-Leibler divergence $D_{KL}(\mathcal{D}_1||\mathcal{D}_2)$ are given, respectively, by:

$$||\mathcal{D}_1 - \mathcal{D}_2|| := \frac{1}{2} \sum_x |p_x - q_x|, \qquad (1)$$

$$D_{KL}(\mathcal{D}_1||\mathcal{D}_2) := \sum_x p_x \ln \frac{p_x}{q_x}. \qquad (2)$$

The O and Θ notation have the usual meaning: for two functions $f(n)$ and $g(n)$, we say that $f(n)$ is $\mathrm{O}(g(n))$ if $f(n)$ is upper-bounded by some positive multiple of $g(n)$ in the asymptotic limit $n \to \infty$, and we say $f(n)$ is $\Theta(g(n))$ if both $f(n)$ is $\mathrm{O}(g(n))$ and $g(n)$ is $\mathrm{O}(f(n))$.

II. DEFINITION OF THE PROBLEM

Let U be an $m \times m$ Haar-random unitary matrix (i.e. the "interferometer"), and let $\Phi_{m,n}$ be the set of all lists of non-negative integers $S = (s_1, \ldots s_m)$ such that $\sum s_i = n$, (i.e. the "input/output states"). The transition probability from a given input state $S = (s_1, \ldots s_m) \in \Phi_{m,n}$ to an output state $T = (t_1, \ldots t_m) \in \Phi_{m,n}$ is given by

$$\Pr[S \to T] = \frac{|\mathrm{Per}(U_{S,T})|^2}{s_1! \ldots s_m! t_1! \ldots t_m!}, \qquad (3)$$

where $U_{S,T}$ is a submatrix of U constructed as follows: (i) first, construct a $m \times n$ matrix U_T by taking t_1 copies of the first row of U, t_2 copies of the second row, etc; (ii) then, form $U_{S,T}$ by taking s_1 times the first column of U_T, s_2 times its second column, etc. If we now fix some input state S (which, for simplicity, we will take to be a

string of n ones followed by $m-n$ zeroes), Eq. (3) defines a distribution $\mathcal{D}_U$ given by

$$\Pr_{\mathcal{D}_U}[T] = \frac{|\mathrm{Per}(U_{S,T})|^2}{t_1! \ldots t_m!}. \qquad (4)$$

The original BOSONSAMPLING problem then can simply be defined as producing a sample from $\mathcal{D}_U$, or at least from some other distribution that is close to it in total variation distance. It can also be shown that, if $m \gg n^2$ and U is Haar-random, then one only needs to consider "no-collision outputs," where each t_i is only 0 or 1, since these dominate the distribution. For simplicity we will restrict our attention to these states from now on, replacing $\Phi_{m,n}$ by $\Lambda_{m,n}$ (defined as the subset of $\Phi_{m,n}$ with only no-collision states) and dropping the denominator of Eq. (4), but for a full discussion see [1] or Sec. 4 of [15].

In this paper, we will consider a modified version of the BOSONSAMPLING problem, where a subset of the photons is lost along the way. Before giving the formal definition, consider the following illustrative example. Suppose we input one photon in each of the first $n + 1$ modes of U (i.e. $S \in \Lambda_{m,n+1}$ is a string of $n + 1$ ones followed by all zeroes), but we observe an outcome with only n photons (i.e. $T \in \Lambda_{m,n}$). With no knowledge of the mechanism behind the losses, what conditional probability should we ascribe to this event, given that we know that exactly one photon was randomly lost at the input ports? Clearly, the probability of this event is just the average of the probabilities of $n + 1$ different BosonSampling experiments that use only n out of the $n+1$ initial photons. In other words, we can write this as

$$\Pr[T] = \frac{1}{n+1} \sum_i |\mathrm{Per}(U_{S_i,T})|^2, \qquad (5)$$

where $U_{S_i,T}$ is obtained by deleting the ith column of $U_{S,T}$. Note that, by the way we defined $U_{S,T}$ it is, in this setting, an $n \times (n+1)$ matrix, so after deleting one of its columns we obtain a square matrix and the permanent function is well-defined.

We can now easily generalize the above to the case where exactly k out of $n + k$ photons were lost. In this case, the probability of outcome $T \in \Lambda_{m,n}$ is given by

$$\Pr[T] = \frac{1}{|\Lambda_{n+k,n}|} \sum_{S \in \bar{\Lambda}_{m,n+k,n}} |\mathrm{Per}(U_{S,T})|^2, \qquad (6)$$

where now $\bar{\Lambda}_{m,k,n}$ denotes the set of all possible no-collision states of n photons in the first $n + k$ out of m modes. It is easy to see that $|\bar{\Lambda}_{m,n+k,n}|$, which is the total number of states that contribute to this probability, is equal to $|\Lambda_{n+k,n}| = \binom{n+k}{n}$.

Two remarks are in order about this particular choice of loss model. The first is that we are supposing that exactly k photons have been lost in the experiment, but the more physically realistic model is one where each photon

has an independent probability p of being lost. These models may seem very different at first glance, especially since the latter is described by a state which does not even have a fixed photon number. Nevertheless, we can connect them as follows: if we start with N total photons, and each has probability p of being lost, then the average number of lost photons is $k = pN$. Furthermore, the probability of observing this exact value of k is non-negligible (i.e. inverse-polynomial), and in fact we are likely to observe it at least once after repeating the experiment $O(\sqrt{N})$ times. This shows that the model considered here can be simulated (with only polynomial overhead) by the realistic one if we set $p = k/N$—and hence, that a hardness result for our model carries over to the realistic one, which is what we need.

The second remark is that, to simplify the analysis, we are assuming that the photons are lost only at the input to the interferometer (e.g. in the sources). In reality, however, we should expect them to be lost inside the circuit or at the detectors as well, in which case the above equation for the probability would be different. We will return to this issue in Section IV.

Given our choice of loss model, we can view the modified problem more abstractly as BOSONSAMPLING with the $|\mathrm{Per}(X)|^2$ function replaced by the following function:

$$\Phi(A) := \frac{1}{|\Lambda_{n+k,n}|} \sum_S |\mathrm{Per}(A_S)|^2, \qquad (7)$$

where $\Phi(A)$ is a function defined on $n \times (n + k)$ matrices, and the sum is taken over all square $n \times n$ proper submatrices of A.[1]

The main questions we address in this paper, then, are: is this generalized version of BOSONSAMPLING, where probabilities are given by objects like $\Phi(A)$, as hard to simulate approximately as the non-lossy case where probabilities are given by $|\mathrm{Per}(X)|^2$? For what values of k?

We will show that the first question has an affirmative answer, but alas, we are only able to show a strong hardness result in the case where k is a constant (i.e. does not grow with n). We leave, as our main open problem, to give a fuller understanding of what happens in the realistic case that k grows with n.

Let us first give a (very) brief outline of the reasoning behind the result of Aaronson and Arkhipov [1]. We begin by defining the following problem:

Problem 1 ($|\mathrm{GPE}|_\pm^2$) [1]. Given as input a $n \times n$ matrix $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{n \times n}$ of i.i.d. Gaussians, together with error bounds $\epsilon, \delta > 0$, estimate $|\mathrm{Per}(X)|^2$ to within additive error $\pm \epsilon \cdot n!$ with probability at least $1 - \delta$ over X in $\mathrm{poly}(n, 1/\epsilon, 1/\delta)$ time.

$|\mathrm{GPE}|_\pm^2$ is the main problem addressed in [1]. Aaronson and Arkhipov show that, if there is a classical algo-

rithm that efficiently produces a sample from a distribution close in total variation distance to that sampled by a BOSONSAMPLING device, then $|\mathrm{GPE}|_\pm^2 \in \mathrm{BPP}^{\mathrm{NP}}$. They then conjecture that $|\mathrm{GPE}|_\pm^2$ is #P-hard (by means of two natural conjectures, the Permanent-of-Gaussians Conjecture and the Permanent Anti-Concentration Conjecture; see [1] for more details). Under that conjecture, they conclude that, if there was an efficient classical algorithm C as described above, the polynomial hierarchy would collapse. This argument provides evidence that, despite being a very restricted quantum computer, a BOSONSAMPLING device is exponentially hard to simulate on a classical computer.

Here, we are going to "intercept" the result of Aaronson and Arkhipov halfway through: we will show that $|\mathrm{GPE}|_\pm^2$ can be reduced to another problem, which is its natural generalization when we replace $|\mathrm{Per}(X)|^2$ by $\Phi(A)$, as long as k is constant. More concretely, consider the following problem:

Problem 2 ($\Sigma|\mathrm{GPE}|_\pm^2$). Given as input a $n \times (n + k)$ matrix $A \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{n \times (n+k)}$, together with error bounds $\epsilon', \delta > 0$, estimate $\Phi(A) := \Sigma_S |\mathrm{Per}(A_S)|^2$ to within additive error $\pm \epsilon' n!$ with probability at least $1 - \delta$ over A in $\mathrm{poly}(n, 1/\epsilon', 1/\delta)$ time.

Then our main result is:

Theorem 1. *If $\mathcal{O}$ is an oracle that solves $\Sigma|\mathrm{GPE}|_\pm^2$ with $\epsilon' = O\left(\frac{\epsilon \delta^{k+1/2} k^{k/2}}{n^{k/2}(n+k)^k}\right)$, then $|\mathrm{GPE}|_\pm^2$ can be solved in $\mathrm{BPP}^{\mathcal{O}}$.*

Notice that the precision demanded of the oracle $\mathcal{O}$ is exponential in k, which is why we can only use Theorem 1 to make claims about BOSONSAMPLING if k is constant. However, in this case Theorem 1 immediately allows us to replace all further claims made by Aaronson and Arkhipov regarding $|\mathrm{GPE}|_\pm^2$ by their equivalent versions with $\Sigma|\mathrm{GPE}|_\pm^2$, and thus show that BOSONSAMPLING with a few lost photons has the same complexity as the original BOSONSAMPLING problem.

III. REDUCTION FROM $|\mathbf{GPE}|_\pm^2$ TO $\Sigma|\mathbf{GPE}|_\pm^2$

In this section, we prove Theorem 1.

The idea behind the proof is as follows. Given $X \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{n \times n}$, we can trivially embed X as the leftmost submatrix of another matrix $A \sim \mathcal{N}(0, 1)_{\mathbb{C}}^{n \times (n+k)}$. Now define $A[c]$ to be the matrix obtained by multiplying the k rightmost columns of A by the real number c, and let $\mathcal{N}_A[c]$ be the resulting distribution over $A[c]$. Then it is easy to see that

$$\Phi(A[c]) := \frac{1}{|\Lambda_{n+k,n}|} |\mathrm{Per}(X)|^2 + |c|^2 Q_1 + |c|^4 Q_2 \\ + \ldots + |c|^{2k} Q_k, \qquad (8)$$

---

[1] Isaac Chuang has suggested to us that $\Phi(A)$ be called the "temperament," perhaps because it generalizes the mod-squared permanent to more temperamental experimental apparatus.

where each Q_i is the sum of the absolute squares of the permanents of all submatrices of A that include i out of the k rightmost columns of A, scaled by a factor of $\frac{1}{|\Lambda_{n+k,n}|}$. If c is sufficiently close to 1, then $\mathcal{N}_A[c]$ is close to $\mathcal{N}_A[1]$ $\left(\text{which coincides with } \mathcal{N}(0,1)_{\mathbb{C}}^{n \times (n+k)}\right)$, and we can use the oracle $\mathcal{O}$ to estimate $\Phi(A[c])$. By calling $\mathcal{O}$ for $k+1$ different values of c, we can then estimate $|\mathrm{Per}(X)|^2$ using standard polynomial interpolation.

Let us start with the following simple lemma, which relates the total variation distance between $\mathcal{N}_A[c]$ and $\mathcal{N}_A[1]$ to the distance between c and 1.

Lemma 1. *If* $|c - 1| \leq \frac{\delta}{\sqrt{nk}}$, *then* $||\mathcal{N}_A[c] - \mathcal{N}_A[1]|| = O(\delta)$.

Proof. First notice that, since $\mathcal{N}_A[1]$ is just the joint distribution of $n(n+k)$ i.i.d. complex Gaussians $x_i \sim \mathcal{N}(0,1)_{\mathbb{C}}$, then $\mathcal{N}_A[c]$ is the joint distribution over $n(n+k)$ i.i.d. complex Gaussians $y_i \sim \mathcal{N}(0, \sigma_i)_{\mathbb{C}}$, where σ_i is 1 for n^2 of the variables (those corresponding to the leftmost $n \times n$ submatrix of A) and c for the remaining nk variables. By Pinsker's inequality, we have

$$||\mathcal{N}_A[c] - \mathcal{N}_A[1]|| \leq \sqrt{\frac{1}{2} D_{KL}(\mathcal{N}_A[1]||\mathcal{N}_A[c])},$$

where $D_{KL}(P||Q)$ is the Kullback-Leibler divergence. When P and Q are multivariate Gaussian distributions, there is a closed form for $D_{KL}(P||Q)$—in particular, if P (respectively Q) corresponds to the distribution over K i.i.d. complex variables $\{x_i\}$ with means 0 and corresponding variances $\{\sigma_{P,i}\}$ (respectively $\{\sigma_{Q,i}\}$), we can write [16]:

$$D_{KL}(P||Q) = \sum_i \left(\frac{\sigma_{P,i}}{\sigma_{Q,i}}\right)^2 - K + 2\sum_i \ln \frac{\sigma_{Q,i}}{\sigma_{P,i}}.$$

By setting K equal to $n(n+k)$, $\sigma_{Q,i}$ equal to 1 for all i, and $\sigma_{P,i}$ equal to 1 for n^2 of the i's and c for the other kn we get

$$D_{KL}(\mathcal{N}_A[1]||\mathcal{N}_A[c]) = nk\left(\frac{1}{c^2} - 1 + 2\ln c\right).$$

If $|1 - c| = a$, we get

$$D_{KL}(\mathcal{N}_A[1]||\mathcal{N}_A[c]) = 2nka^2 + O(a^3)$$

and hence

$$||\mathcal{N}_A[c] - \mathcal{N}_A[1]|| \leq a\sqrt{nk} + O(a^2).$$

Setting $a = \frac{\delta}{\sqrt{nk}}$ completes the proof. $\qquad \square$

By Lemma 1, to satisfy the definition of $\mathcal{O}$ we must have $|1 - c| = O(\delta/\sqrt{kn})$. One might worry about how stable the estimate of $|\mathrm{Per}(X)|^2$ produced by the polynomial interpolation will be, if we are only allowed to probe a very small region of values of c. We now show how to relate the precision in the output of $\mathcal{O}$ to the resulting estimate of $|\mathrm{Per}(X)|^2$.

To begin, consider the equivalent problem of estimating the parameters $\{\beta_i\}$ of the polynomial

$$w = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k \qquad (9)$$

by choosing $k+1$ distinct values x_i in the interval $[1 - a, 1 + a]$ and estimating the corresponding values w_i each with error $e_i < \epsilon' n!$. This results in the following linear system (written in vector notation):

$$\mathbf{w} = X\beta + \mathbf{e}, \qquad (10)$$

where X is the Vandermonde matrix

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^k \\ 1 & x_2 & x_2^2 & \ldots & x_2^k \\ 1 & x_3 & x_3^2 & \ldots & x_3^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{k+1} & x_{k+1}^2 & \ldots & x_{k+1}^k \end{pmatrix}. \qquad (11)$$

Note that X is invertible as long as we choose all x_i's distinct.

We can now use, as an estimator for β, the one given by the ordinary least squares method, namely

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T \mathbf{w}. \qquad (12)$$

Let us now bound the variance on the estimator $\hat{\beta}_0$, which corresponds to the parameter that we are trying to estimate (i.e. $|\mathrm{Per}(X)|^2$ in Eq. (8)). Since the errors e_i are all in the interval $[-\epsilon' n!, \epsilon' n!]$ by assumption we can, without loss of generality, assume that $\mathrm{E}(e_i) = 0$ and $\mathrm{Var}(e_i) \leq (\epsilon' n!)^2$. Let us also denote by Ω the covariance matrix of the error vector $\mathbf{e}$. Then we can write:

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}\left[\left(X^T X\right)^{-1} X^T \mathbf{w}\right]$$
$$= X^{-1} \Omega \left(X^{-1}\right)^T, \qquad (13)$$

and thus

$$\mathrm{Var}(\hat{\beta}_0) = \left[X^{-1} \Omega \left(X^{-1}\right)^T\right]_{1,1}. \qquad (14)$$

But now we can bound this as:

$$\left[X^{-1} \Omega \left(X^{-1}\right)^T\right]_{1,1} \leq \max_{i,j} \left[X^{-1} \Omega \left(X^{-1}\right)^T\right]_{i,j}$$
$$\leq ||X^{-1} \Omega (X^{-1})^T||_\infty$$
$$\leq ||X^{-1}||_\infty ||\Omega||_\infty ||(X^{-1})^T||_\infty$$
$$= ||X^{-1}||_\infty ||\Omega||_\infty ||(X^{-1})||_1$$
$$\leq (k+1)||X^{-1}||_\infty^2 ||\Omega||_\infty, \qquad (15)$$

where $||A||_\infty$ and $||A||_1$ are the maximum row 1-norm and maximum column 1-norm, respectively, and we used the inequalities $||AB||_\infty \leq ||A||_\infty ||B||_\infty$ and $||A||_1 \leq (k+$

$1)||A||_\infty$, which hold for all $\{A, B\} \in M_{k+1}$ [17]. We can now use a result due to Gautschi [18], which bounds the norm of the inverse of the square Vandermonde matrix as

$$||X^{-1}||_\infty \leq \max_{1 \leq j \leq k+1} \prod_{i=1, i \neq j}^{k+1} \frac{1 + |x_j|}{|x_j - x_i|}.$$

Since, in our case, all $x_i's$ are bounded in the interval $[1 - a, 1 + a]$, we can write

$$||X^{-1}||_\infty \leq (2 + a)^k \max_{1 \leq j \leq k+1} \prod_{i=1, i \neq j}^{k+1} \frac{1}{|x_j - x_i|}. \quad (16)$$

In order to obtain the sufficiently tight bound from the expression above, it is helpful to choose the x_i's evenly spaced in the interval $[1 - a, 1 + a]$. In that case, the maximum in Eq. (16) is obtained by choosing x_j to be one of the central points, i.e. $j = k/2 + 1$ if k is even or $j = (k + 1)/2$ is k is odd. In this case, it is not hard to show that, if k is even, we get

$$||X^{-1}||_\infty \leq \left(\frac{2e}{a}\right)^k \frac{1}{\pi k}, \quad (17)$$

and if k is odd,

$$||X^{-1}||_\infty \leq \left(\frac{2e}{a}\right)^k \frac{1}{\pi \sqrt{k^2 - 1}} \frac{k^k}{(k-1)^{\frac{k-1}{2}}(k+1)^{\frac{k+1}{2}}}. \quad (18)$$

Clearly if k is sufficiently large, these bounds coincide. Finally note that, since every e_i has variance at most $(\epsilon'n!)^2$, we can write

$$||\Omega||_\infty \leq (k + 1)(\epsilon'n!)^2, \quad (19)$$

By plugging Eqs. (15), (17) or (18), and (19) back into Eq. (14) we obtain

$$\mathrm{Var}(\hat{\beta}_0) = O\left(\frac{(\epsilon'n!)^2}{a^{2k}}\right). \quad (20)$$

We are now in position to prove Theorem 1.

Proof. We begin by writing

$$\Phi(A[c]) := \frac{1}{|\Lambda_{n+k,n}|}|\mathrm{Per}(X)|^2 + |c|^2 Q_1 + |c|^4 Q_2$$
$$+ \ldots + |c|^{2k} Q_k.$$

We then choose $k + 1$ values of c equally spaced in the interval $\left[1 - \delta/\sqrt{nk}, 1 + \delta/\sqrt{nk}\right]$. By Lemma 1, each $A[c]$ obtained in this way is within δ of total variation distance to $\mathcal{N}(0, 1)_{\mathbb{C}}^{n \times (n+k)}$, so we can use the oracle $\mathcal{O}$ to obtain an estimate of $\Phi(A[c])$ to within additive error $\pm \epsilon'n!$. Using the ordinary least squares method, we can then give an estimate $\hat{P}$ for $\frac{1}{|\Lambda_{n+k,n}|}|\mathrm{Per}(X)|^2$ with

variance given by

$$\mathrm{Var}(\hat{P}) = O\left(\frac{(\epsilon'n!)^2}{(\delta^2/nk)^k}\right),$$

which is just Eq. (20) where we set $a = O(\delta/\sqrt{nk})$. Finally, by Chebyshev's inequality, we have

$$\Pr\left(|\hat{P} - \frac{1}{|\Lambda_{n+k,n}|}|\mathrm{Per}(X)|^2| \geq K\sqrt{\mathrm{Var}(\hat{P})}\right) \leq \frac{1}{K^2}.$$

For any $K > 0$. Thus, by setting $K = 1/\sqrt{\delta}$ and $\epsilon' = O\left(\frac{\epsilon \delta^{k+1/2} k^{k/2}}{n^{k/2}(n+k)^k}\right)$, we obtain an estimate for $|\mathrm{Per}(X)|^2$ to within additive error $\pm \epsilon \cdot n!$ with probability at least $1 - \delta$ (over X and over the errors in the outputs of $\mathcal{O}$) in poly$(n, 1/\epsilon, 1/\delta)$ time, thus solving $|\mathrm{GPE}|_\pm^2$. $\square$

Theorem 1 then guarantees, for example, that if k is constant, then as long as the error in each estimate of $\mathcal{O}$ is $\ll 1/n^{3k/2}$, solving $\Sigma|\mathrm{GPE}|_\pm^2$ is as least as hard as solving $|\mathrm{GPE}|_\pm^2$.

IV. SUMMARY AND OPEN PROBLEMS

We investigated the loss tolerance of BOSONSAMPLING in the regime where only a few photons are lost. In this case, the output distribution can become very far (in total variation distance) from the ideal one, and so the original argument of Aaronson and Arkhipov [1] does not work. Nevertheless, we showed how the problem of estimating the permanent, which usually describes the outcome probabilities in the ideal BOSONSAMPLING model, can be reduced to the problem of estimating the quantity Φ, defined in Eq. (7), which describes the probabilities in the lossy BOSONSAMPLING model. For the regime where the number of lost photons is constant (or, alternatively, where each photon is lost with probability $\sim c/n$), this allows us to replace $|\mathrm{Per}(X)|^2$ by $\Phi(A)$ in Aaronson and Arkhipov's argument, and thus to obtain similar complexity results as those in [1]. Note that it was not necessary to modify either the Permanent-of-Gaussians Conjecture or the Permanent Anti-Concentration Conjecture that Aaronson and Arkhipov [1] used.

Note that, if we restrict our attention to the loss model where each photon has an independent probability p of being lost [cf. the discussion after Eq. (6)], then there is a hardness argument that works for $k = O(1)$, based on postselection, that is much simpler than our proof. Namely, in this regime the probability that *zero* photons will be lost is only polynomially small (in n), so one could simply repeat the experiment a polynomial number of times, and postselect on observing all of the photons. As a related observation, if the scattering matrix U can be chosen arbitrarily (rather than from the Haar measure), then even if exactly k photons will be lost in each run, we could simply pick U to act as the identity on k of the n modes that initially have 1 photon each. If no

photons are observed in those k modes—which occurs with probability $\Theta(1/n^k)$—then we know that lossless BOSONSAMPLING has been performed on the remaining $n - k$ photons.

In our view, these observations underscore that our hardness result is only a first step, and that ultimately one wants a strong hardness result that holds even when k is non-constant. On the other hand, it is important to understand that our result is not itself obtainable by any simple postselection trick like those described above. Rather, our result requires analyzing output probabilities that are sums of absolute squares of many permanents, and showing that these sums inherit some of the hardness of the permanent itself. By demonstrating that such reductions are possible, we hope our result will inspire further work on the complexity of lossy BOSONSAMPLING. We also point out that the reasoning behind our proof of Theorem 1 can be used to deal with other issues besides just photon losses, such as e.g. the case of photon "shuffling" described in the appendix, where photon losses compounded with dark counts acts as noise in the probability distribution that cannot be dealt with just by postselection.

Our work leaves a few open questions. The obvious one is to investigate what happens in more realistic loss regimes, such as e.g. if a constant fraction of the photons is lost (i.e. in our notation, $k = \epsilon n$). In that case, Theorem 1 requires the oracle $\mathcal{O}$ to produce an estimate of Φ to within error $1/n^{\Theta(\epsilon n)}$, so it does not allow us to make any strong complexity claims. It would be reasonable to expect that, if k is too large, BOSONSAMPLING becomes classically simulable—in fact, this is easily seen to be true if all but $O(\log n)$ photons get lost, and it would be interesting if we could determine where the transition happens.

In that direction, an even more basic question that one can ask is whether the "temperament," $\Phi(A)$, of an i.i.d. Gaussian matrix $A \sim \mathcal{N}(0,1)_{\mathbb{C}}^{n \times (n+k)}$, becomes concentrated around some simple classical estimator when k is large compared to n. For example, one could look at $R(A)$, the product of the squared 2-norms of the rows of A. Aaronson and Arkhipov [15] previously studied this estimator in the case of *no* lost photons (i.e., $k = 0$), and showed that even there, $R(A)$ has a constant amount of correlation with $|\mathrm{Per}(A)|^2$. If (say) $n = 2$, then it is not hard to see that $R(A)$ has an excellent correlation with $\Phi(A)$, as k gets arbitrarily large. What happens when $n \sim \log^2 k$ or $n \sim \sqrt{k}$?

A second open problem concerns our choice to model photon losses assuming all of them were lost at the input, as described at the beginning of Section II. We leave as a question for future work to adapt our proof to the case where losses happen at the detectors. In that case, rather than summing over all possibilities that k out of $k + n$ input photons were lost, we would have to sum over all possibilities of a subset of k out of the $m - n$ detectors malfunctioning. When we do that, we obtain a probability expression that is very similar to that in Eq. (7),

but with three complications: (i) the permanents are of $(n+k) \times (n+k)$ matrices, since now $n+k$ photons actually traveled through the network, (ii) the sum is over $\binom{m-n}{k}$ possibilities, rather than $\binom{n+k}{n}$, and (iii) rather than considering A to be a $n \times (n + k)$ matrix of i.i.d. Gaussian entries, we need to consider it an $m \times (n + k)$ random matrix of orthonormal columns. Nevertheless, in the regime $m = \Theta(n^2)$ that is often assumed in BOSONSAMPLING, it should be possible (although somewhat cumbersome) to adapt our polynomial interpolation argument to this case. Then, as a next step, one could consider the case where the photons can be randomly lost both at the sources *and* at the detectors, or indeed anywhere in the network.

Appendix: Other applications of main result

In this appendix, we point out how our main result in Section III can be repurposed to include other scenarios besides photon losses, such as dark counts. In the same way that our main result (Theorem 1) only extends the loss-tolerance of BOSONSAMPLING to allow for a constant number k of photons to be lost, the other scenarios described in this Appendix also have some parameter k (e.g. the number of dark counts) that can be at most a constant.

Recall that our main result concerns the setting where $n + k$ photons were prepared, but only n of them were measured, and the corresponding probability is given by replacing the permanent with $\Phi(A)$ defined in Eq. (7), which is a function of $n \times (n + k)$ matrices. We then embed the $n \times n$ matrix X whose permanent we wish to estimate in a matrix $A[c]$ defined as

$$A[c] = (X \ cY), \qquad (A.1)$$

for some $k \times k$ Gaussian Y. We then write, as in Eq. (8),

$$\Phi(A[c]) := \frac{1}{|\Lambda|}(|\mathrm{Per}(X)|^2 + |c|^2 Q_1 + |c|^4 Q_2$$
$$+ \ldots + |c|^{2k} Q_k), \qquad (A.2)$$

where each Q_i is given in terms of permanents of different

submatrices of $A[c]$, other than X, and their values are irrelevant for our purposes. Finally, we can use a polynomial regression to produce an estimate for $|\text{Per}(X)|^2$ from estimates of $\Phi(A[c])$ for several values of c sufficiently close to 1.

Now, let us show how other scenarios can be cast in the same form.

1. Dark counts

Suppose that, rather than preparing $n+k$ photons and measuring n of them, we prepare n and measure $n+k$ photons—that is, we have events known as dark counts, when an optical detector "clicks" even in the absence of a photon. This means that we must now ascribe, to our detected event, a probability that is the average of the probabilities obtained by considering that each possible subset k of the $k+n$ detectors were the ones that misfired. In other words, we can write the probability as

$$\Phi_{\text{dark}}(A) := \frac{1}{|\Lambda|} \sum_{T \in \Lambda} |\text{Per}(A_T)|^2 \qquad \text{(A.3)}$$

where again Λ denotes the set of all possible no-collision states of n photon in $n+k$ modes, but A is now a $(n+k) \times n$ matrix, and A_T is the submatrix of A obtained by taking n of its rows corresponding to the 1's in T. We once more embed the $n \times n$ matrix X in a larger matrix, but which now is defined as

$$A[c] = \begin{pmatrix} X \\ cY \end{pmatrix}. \qquad \text{(A.4)}$$

Note that this leads to an identical expression as Eq. (8),

$$\Phi_{\text{dark}}(A[c]) = \frac{1}{|\Lambda|}(|\text{Per}(X)|^2 + |c|^2 Q_1 + |c|^4 Q_2$$
$$+ \ldots + |c|^{2k} Q_k). \qquad \text{(A.5)}$$

As before, the Q_i's correspond to permanents of different submatrices of $A[c]$, which are now distributed vertically rather than horizontally. It is clear from the form of $\Phi_{\text{dark}}(A[c])$ that the arguments of Section III follow through unchanged.

2. Combination of losses and dark counts

Consider now the following scenario: we prepared $n+k$ photons and measured $n+k$ photons, but during the process k of our photons were lost, and there were k dark counts (let us call this event of losing a photon and getting a dark count as "shuffling" a photon). One difference compared to the previous scenarios is that a shuffling event is not *heralded* in any way—in principle, any experiment which performs n-photon BosonSampling

should take all probabilities of up to n-fold shuffling into account as noise in the final distribution. As we now argue, in a situation where up to k out of $n+k$ photons can be shuffled, for constant k, the reasoning behind our hardness result goes through.

First, suppose that, for whatever reason, we know that exactly k photons were shuffled during the process. In this case, the probability we ascribe to the observed event is given by

$$\Phi_{\text{shuf}}(A,k) := \frac{1}{|\Lambda|^2} \Sigma_{S,T \in \Lambda} |\text{Per}(A_{ST})|^2, \qquad \text{(A.6)}$$

where now S and T both run over all possible collision-free basis states corresponding to n photons in $k+n$ modes and index rows and columns, respectively, of A. In other words, again the sum is over all $n \times n$ proper submatrices of A, which now is $(n+k) \times (n+k)$. As before, we can build a matrix of the form

$$A[c] = \begin{pmatrix} X & cY \\ cV & c^2W \end{pmatrix}, \qquad \text{(A.7)}$$

and write

$$\Phi_{\text{shuf}}(A[c],k) = \frac{1}{|\Lambda|^2}(|\text{Per}(X)|^2 + |c|^2 Q_1 + |c|^4 Q_2$$
$$+ \ldots + |c|^{4k} Q_2 k). \qquad \text{(A.8)}$$

In this case, each Q_j is a sum that collects all submatrices of $A[c]$ which use j rows and/or columns of A that contain the factor c, and the total polynomial is of degree $4k$, rather than $2k$ as in previous cases. Nonetheless, it is straightforward to adapt the least squares method to this setting, and our main argument goes through with no significant changes.

However, since shufflings are not heralded, restricting ourselves to the case where exactly k photons are shuffled is not very natural. To make it (slightly) more natural, we can consider the case where all j-fold shufflings can happen, from j ranging from 1 to k, with corresponding known probabilities p_j. In this case, the probability we assign to the event is

$$\Phi_{\text{shuf}}(A[c]) = \sum_{j=0}^{k} p_j \Phi_{\text{shuf}}(A[c],j). \qquad \text{(A.9)}$$

But note that, in each $\Phi_{\text{shuf}}(A[c],j)$ for j between 0 and $k-1$, the permanents in Eq. (A.8) are over submatrices of dimensions $(n+k-j) \times (n+k-j)$, and thus each of them must include at least one column or row that contains a factor c. This means that, in the sum in Eq. (A.9), the only term that does not have a factor of c is the one containing $|\text{Per}(X)|^2$. Thus we can write once

again

$$\Phi_{\text{shuf}}(A[c],k) = \frac{1}{|\Lambda|^2}(p_k|\text{Per}(X)|^2 + |c|^2 Q_1' \\ + |c|^4 Q_2' + \ldots + |c|^{4k} Q_2' k). \quad \text{(A.10)}$$

where now all Q_i's may depend in complicated but unimportant ways on all j-fold shuffling events. We can again apply the least-squares method, but now there is an important caveat: rather than obtaining an estimate of $|\text{Per}(X)|^2$ to within additive precision, we only obtain an estimate of $p_k|\text{Per}(X)|^2$. This introduces the additional restriction that p_k is large enough that the additive approximation does not become useless.

[1] S. Aaronson and A. Arkhipov, The computational complexity of linear optics, *Theory of Computing* **9**, 143–252 (2013).

[2] M. A. Broome, A. Fedrizzi, S. Rahimi-Keshari, J. Dove, S. Aaronson, T. C. Ralph, and A G. White, Photonic boson sampling in a tunable circuit, *Science* **339**, 794–798 (2013).

[3] A. Crespi, R. Osellame, R. Ramponi, D. J. Brod, E. F. Galvão, N. Spagnolo, C. Vitelli, E. Maiorino, P. Mataloni, and F. Sciarrino, Integrated multimode interferometers with arbitrary designs for photonic boson sampling, *Nat. Phot.* **7**, 545–549 (2013).

[4] M. Tillmann, B. Dakić, R. Heilmann, S. Nolte, A. Szameit, and P. Walther, Experimental boson sampling, *Nat. Phot.* **7**, 540–544 (2013).

[5] J B. Spring, B. J. Metcalf, P. C. Humphreys, W. S. Kolthammer, X.-M. Jin, M. Barbieri, A. Datta, N. Thomas-Peter, N. K. Langford, D. Kundys, J. C. Gates, B. J. Smith, P. G. R. Smith, and I. A. Walmsley, Boson sampling on a photonic chip, *Science* **339**, 798–801 (2013).

[6] N. Spagnolo, C. Vitelli, M. Bentivegna, D. J. Brod, A. Crespi, F. Flamini, S. Giacomini, G. Milani, R. Ramponi, P. Mataloni, R. Osellame, E. F. Galvão, and F. Sciarrino, Experimental validation of photonic boson sampling, *Nat. Phot.* **8**, 615–620 (2014).

[7] J. Carolan, J. D. A. Meinecke, P. J. Shadbolt, N. J. Russell, N. Ismail, K. Wörhoff, T. Rudolph, M. G. Thompson, J. L. O'Brien, J. C. F. Matthews, and A. Laing, On the experimental verification of quantum complexity in linear optics, *Nat. Phot.* **8**, 621–626 (2014).

[8] M. Bentivegna, N. Spagnolo, C. Vitelli, F. Flamini, N. Viggianiello, L. Latmiral, P. Mataloni, D. J. Brod, E. F. Galvão, A. Crespi, R. Ramponi, R. Osellame, and F. Sciarrino, Experimental scattershot boson sampling, *Sci. Adv.* **1**, e1400255 (2015).

[9] J. Carolan, C. Harrold, C. Sparrow, E. Martín-López, N. J. Russell, J. W. Silverstone, P. Shadbolt, N. Matsuda, M. Oguma, M. Itoh, G. D. Marshall, M. G. Thompson, J. C. F. Matthews, T. Hashimoto, J. L. O'Brien, and A. Laing Universal linear optics, *Science* **349**, 711–716 (2015).

[10] P. P. Rohde and T. C. Ralph, Error tolerance of the boson-sampling model for linear optics quantum computing, *Phys. Rev. A* **85**, 022332 (2012).

[11] V. Shchesnovich, Sufficient condition for the mode mismatch of single photons for scalability of the boson-sampling computer, *Phys. Rev. A* **89**, 022333 (2014).

[12] A. Leverrier and R. García-Patrón, Analysis of circuit imperfections in BosonSampling, *Quant. Inf. Comp.* **15**, 0489–0512, (2015).

[13] A. Arkhipov, Boson sampling is robust to small errors in the network matrix, (2014), arXiv:1412.2516 [quant-ph].

[14] G. Kalai and G. Kindler, Gaussian noise sensitivity and BosonSampling, (2014), arXiv:1409.3093 [quant-ph].

[15] S. Aaronson and A. Arkhipov, BosonSampling is far from uniform, *Quant. Inf. Comp.*, **14**, 1383–1423 (2014).

[16] S. Kullback, *Information Theory and Statistics*, (John Wiley & Sons, New York, 1959).

[17] R. Horn and C. Johnson, *Matrix Analysis*, (Cambridge University Press, New York, 1985).

[18] W. Gautschi, On inverses of Vandermonde and confluent Vandermonde matrices, *Numer. Math.*, **4**, 117–123 (1962).