Chapter 1

Classically simulating near-term partially-distinguishable and lossy boson sampling

1.1 Introduction

Since the idea of quantum computers was first proposed, speedups have been theoretically demonstrated for a variety of problems [1]. More recently, this has turned to interest in near term quantum speedups on realistic hardware [2]. Boson Sampling is an example of such a problem [3], which considers the complexity of sampling from the same probability distribution as n collision-free indistinguishable photons output from an m-mode linear optical interferometer drawn uniformly at random. Aaronson and Arkhipov showed that the ability to exactly simulate Boson Sampling in polynomial time would imply that $P^{\#P} = BPP^{NP}$, and the Polynomial Hierarchy would collapse to the third level. This is because the output probabilities of a linear optical interferometer are proportional to the permanent of an $n \times n$ complex matrix [4], which is $\#P$-Hard to compute. The same result was also shown for approximately sampling from the same distribution, modulo two conjectures related to the permanents of Gaussian matrices [3]. This led to significant interest in Boson Sampling, with experimental demonstrations ranging from two to six photons [5, 6, 7, 8, 9, 10, 11, 12].

This has driven interest in improving classical simulations of Boson Sampling, in order to verify any speedup gained from an experimental implementation. Neville *et al.* [13] showed that Boson Sampling can be classically performed for 30 bosons across 900 modes at a rate of 250 samples in under five hours on a standard laptop. This was further improved by Clifford and Clifford [14], who showed that sampling from n photons across m modes can be classically performed in $O(n2^n + \text{poly}(n, m))$ time and $O(m)$ space.

Recently classical simulations have been expanded to consider practical issues such as photon distinguishability, based on a rich collection of theoretical work in this area [15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27]. Renema *et al.* [28] demonstrated that Boson Sampling with partially-distinguishable photons can be simulated in time which grows polynomially with n. This was later expanded to consider loss as well [29]. However, the runtime might still not be efficient in practice, as the polynomial can be large. There is also a further disadvantage in that the error bounds are the average case for a random linear optical interferometer, meaning that there could be interferometers for which the algorithm performs significantly worse. A

significant improvement could be achieved through adapting the method of Clifford & Clifford to this algorithm, but there are challenges with this approach.

Another motivation for adapting Clifford & Clifford to photonic imperfection is the task of classically simulating other photonic regimes, such as Gaussian Boson Sampling [30]. The probability distribution of Gaussian Boson Sampling is that of n indistinguishable squeezed states at the output of an m-mode linear optical interferometer, and depends on the Hafnian of a matrix. Unlike Boson Sampling, there is no known polynomial time classical algorithm for computing the outcome of n fully distinguishable squeezed states. On the other hand, it is classically efficient to sample distinguishable squeezed states, via a similar approach to that used for classically sampling distinguishable single photons [31]. As a result, adapting the Clifford & Clifford algorithm to non-ideal Boson Sampling models provides a first step towards being able to classically simulate imperfections in Gaussian Boson Sampling.

Here we consider the cost of classically simulating Boson Sampling when the photons are partially distinguishable or lossy. We look at the same model of distinguishability as considered in [28, 29], and use techniques for modelling photon distinguishability in first quantization [32, 33] to show that this is akin to choosing the indistinguishable photons of a Boson Sampling experiment via the binomial distribution. We combine this with the well-studied model of uniform loss, which also follows a binomial distribution. This gives rise to a method which is able to naturally apply the Clifford & Clifford algrithm and take advantage of its efficiency. This algorithm also offers a worst-case error bound for *any* linear optical interferometer, rather than a random one. Although this approach only offers a polynomial improvement compared to the runtime for ideal Boson Sampling (unlike the exponential improvement shown in [28, 29]) we use analytical bounds to show that for photon numbers of experimental interest our algorithm can make a significant improvement over alternative approaches.

This article is laid out as follows. In Sec. 1.2, we give an overview of distinguishability in first quantization, and summarise the work of classical simulation algorithms including [28, 29]. In Sec. 1.3, we show what the model of distinguishability looks like in first quantization, and provide an alternative classical simulation. In Sec. 1.4, we consider average error bounds for a Haar-random unitary interferometer, via the methods explained in Sec. 1.2.3. In Sec. 1.5, we improve this bound to a worst-case error bound, by computing an upper bound for the trace distance between our approximation and the model. In Sec. 1.6, we expand these results to consider uniform loss, and show how distinguishability and loss relate to each other. In Sec. 1.7, we explore these error bounds for experimentally interesting numbers of photons, and show that there are some cases where our algorithm offers an improvement. Finally, we briefly consider *non*-uniform loss, where loss is is a function of the number of optical components, and use the methods of [34, 35] to show that classical simulations with non-uniform loss also become easier when distinguishability is introduced. We conclude with some open research questions in Sec. 1.9.

1.2   Preliminaries

1.2.1   Partial distinguishability in first quantization

Studying photonics in first quantization allows an arbitrary linear optical experiment to be expressed in the quantum circuit model [32, 33, 35]. Bosonic symmetrisation can be implemented efficiently, e.g. through use of the quantum Schur transform [36], and the main sources of errors in linear optics can be modelled: loss as erasure of qudits and distinguishability as correlation to extra photonic degrees of freedom, both of which lead to decoherence of the quantum circuit.

We start by considering arbitrary distinguishability of photonic scattering, models of which

have been proposed in [20, 21, 22]. Let m, n be integers such that $m \in O(n^2)$. Let $U \in \mathrm{U}(m)$ be an m-mode linear optical interferometer, and let S, S' be n-photon m-mode Fock vectors, meaning that $S = (S_0, \ldots, S_m)$, $S' = (S'_0, \ldots, S'_m)$, and $\sum_{i=0}^{m} S_i = \sum_{i=0}^{m} S'_i = n$. The probability of seeing output occupation S' from input S with interferometer U is

$$\Pr[S'] = \frac{1}{\prod_{i=1}^{m} S_i! S'_i!} \sum_{\tau, \tau' \in \mathrm{S}_n} \prod_{k=1}^{n} U_{s'_k, s_{\tau(k)}} U^*_{s'_k, s_{\tau'(k)}} \mathcal{S}_{\tau'(k), \tau(k)}, \tag{1.1}$$

where $\mathcal{S}_{k,l}$ is an $n \times n$ Gram matrix describing the (in)distinguishability of pairs of photons, S is a Fock array describing the arrangement of input photons, s is the corresponding array of single particle states in first quantisation, likewise for S' and s', U is a $m \times m$ unitary matrix describing the interferometer, and S_n is the group of permutations of n particles. Of particular note is that when $\mathcal{S}$ is the all-1 matrix, we have

$$\Pr[S'] = \frac{1}{\prod_{i=1}^{m} S_i! S'_i!} \sum_{\tau, \tau' \in \mathrm{S}_n} \prod_{k=1}^{n} U_{s'_k, s_{\tau(k)}} U^*_{s'_k, s_{\tau'(k)}} \tag{1.2}$$

$$= \frac{1}{\prod_{i=1}^{m} S_i! S'_i!} \sum_{\tau \in \mathrm{S}_n} \prod_{k=1}^{n} U_{s'_k, s_{\tau(k)}} \sum_{\tau' \in \mathrm{S}_n} \prod_{k=1}^{n} U^*_{s'_k, s_{\tau'(k)}} \tag{1.3}$$

$$= \frac{1}{\prod_{i=1}^{m} S_i! S'_i!} \operatorname{per}(U_{S,S'}) \operatorname{per}(U^*_{S,S'}) \tag{1.4}$$

$$= \frac{1}{\prod_{i=1}^{m} S_i! S'_i!} |\operatorname{per}(U_{S,S'})|^2, \tag{1.5}$$

where $U_{S,S'}$ is a matrix defined by taking rows and columns of U according to S and S' [4]. This is the output probability distribution for Boson Sampling with indistinguishable photons. Likewise, we can see that when $\mathcal{S}$ is the identity matrix, we have

$$\Pr[S'] = \frac{1}{\prod_{i=1}^{m} S_i! S'_i!} \sum_{\tau \in \mathrm{S}_n} \prod_{k=1}^{n} U_{s'_k, s_{\tau(k)}} U^*_{s'_k, s_{\tau(k)}} \tag{1.6}$$

$$= \frac{1}{\prod_{i=1}^{m} S_i! S'_i!} \sum_{\tau \in \mathrm{S}_n} \prod_{k=1}^{n} |U_{s'_k, s_{\tau(k)}}|^2 \tag{1.7}$$

$$= \frac{1}{\prod_{i=1}^{m} S_i! S'_i!} \operatorname{per}(|U_{S,S'}|^2), \tag{1.8}$$

which is the distribution for fully distinguishable photons. In first quantisation, the input state can be written as [32]

$$\rho = \frac{1}{n! \prod_{i=1}^{m} S_i!} \sum_{\sigma, \sigma' \in \mathrm{S}_n} \sigma |s\rangle \langle s| \sigma'^\dagger \prod_{k=1}^{n} \mathcal{S}_{\sigma'^{-1}(k), \sigma^{-1}(k)}, \tag{1.9}$$

For convenience, we will assume here and throughout the rest of this paper that our photons start in the Fock state $|S\rangle = |1^n 0^{m-n}\rangle$. We will use the corresponding first quantized states of this form in the following.

1.2.2 Classical simulation of fully (in)-distinguishable lossless Boson Sampling

Boson Sampling under ideal conditions (lossless indistinguishable single photons) is intractable for sufficiently large n. Until recently the only classical simulation method explicitly known was to

compute the entire probability distribution before taking a sample, though it was widely believed that more efficient, albeit still exponential time, approaches existed. A brute force method cannot scale, due to both the number of possible outcomes and the complexity of computing even one $n \times n$ complex matrix permanent.

Two major results gave the first explicit classical simulation strategies which were faster than brute-force sampling. The first, by Neville *et al.* [13], demonstrated that Boson Sampling experiments with up to 30 photons could be simulated on a single laptop, and suggests that a supercomputer could handle up to 50 photons. This was achieved by starting with the classical distribution of n distinguishable photons, and then using Metropolised Independence Sampling to adapt the distribution to that of ideal Boson Sampling. The second result, by Clifford & Clifford [14], gave a classical algorithm for exact Boson Sampling and runs in time equivalent to computing two $n \times n$ matrix permanents per sample with a polynomial overhead. This is through a combination of optimizations, particularly computing marginal probabilities and sampling via the chain rule. Our approach here is to make these more efficient techniques applicable to realistic situations with distinguishability and loss.

It is also worth discussing at this point how to classically simulate boson sampling with completely distinguishable photons, for which a polynomial time exists [31]. In this case, there is no photon interference, so photons can be sampled individually. This is done by taking a photon which starts in mode i, and sampling output mode j with probability $|U_{j,i}|^2$. Repeating for all photons gives us the complete sample in $O(mn)$ time.

1.2.3 Expanding in terms of fixed points

In [28, 29], Renema *et al.* consider a model where inter-photon distinguishability is measured by an inner product of pure states [16, 15, 19, 18, 20, 22]. The probability distribution of arbitrarily distinguishable bosons is modelled as

$$\Pr[S'] = \sum_{\sigma \in \mathrm{S}_n} \prod_{i=1}^{n} \mathcal{S}_{i,\sigma(i)} \operatorname{per}(M * M^*_{1,\sigma}), \tag{1.10}$$

where $\mathcal{S}$ is the same matrix describing the distinguishability as in the previous section, M is a matrix defined by the rows and columns of our interferometer U selected based on our photon output S' and input S, $M^*_{1,\sigma}$ is the conjugate matrix with the identity permutation applied to rows and permutation σ applied to columns, and $*$ denotes element-wise multiplication. They further restrict to a model where the indistinguishability overlap is defined by a single parameter $\mathcal{S}_{i,j} = x + (1-x)\delta_{i,j}, x \in [0,1]$. The sum over permutations can be ordered based on how many *fixed points* a permutation has, giving

$$\Pr[S'] = \sum_{j=0}^{n} \sum_{\sigma^j} x^j \operatorname{per}(M * M^*_{1,\sigma}), \tag{1.11}$$

where σ^j denotes permutations which have $n - j$ elements as fixed points. Each permanent can be broken down via the Laplace expansion into a sum of a complex matrix permanent multiplied by a positive matrix permanent:

$$\Pr[S'] = \sum_{j=0}^{n} \sum_{\sigma^j} x^j \sum_{\substack{J' \leq S' \\ |J'| = j}} \operatorname{per}(M_{J',1} * M^*_{J',\sigma_p}) \operatorname{per}(|M_{\bar{J}',\sigma_u}|^2). \tag{1.12}$$

Here we are now choosing submatrices of M, with J' representing the $\binom{n}{j}$ possible combinations of rows from M, $\bar{J}'$ representing the remaining rows, and σ_p and σ_u representing permuted and unpermuted elements of σ respectively. The $J' \leq S'$ notation is used to indicate that J' is a Fock state such that $J_i \leq S'_i \, \forall i \in [m]$.

The classical simulation method used truncates the number of fixed points in a permutation as being at most k, with the remainder of the probability treated as an error margin. It is important to note while these approximations are real, they are not necessarily positive. This is due to the truncation, where positive higher order terms which would have corrected the probability to be positive are now missing from the approximation. To correct this, any negative approximations are rounded up to 0. These probabilities are then used to train a Metropolised Independence Sampler, akin to the technique of [13]. Training this sampler requires approximating a number of probabilities dependent on the underlying distribution, each of which involves computing $O(n^{2k})$ permanents of $k \times k$ complex matrices, and the same number of permanents of $(n-k) \times (n-k)$ matrices with non-negative entries. The permanents of $k \times k$ complex matrices can be computed classically in $O(k2^k)$ time via Ryser's algorithm, and the permanents of matrices with non-negative entries can be approximated up to multiplicative error in polynomial time [37, 38]. As long as k is independent of n, this means that there is a polynomial runtime.

To work out a suitable value k, define coefficients c_j as

$$c_j = \sum_{\substack{J' \leq S' \\ |J'| = j}} \operatorname{per}(M_{J',1} * M^*_{J',\sigma_p}) \operatorname{per}(|M_{\bar{J}',\sigma_u}|^2). \tag{1.13}$$

Assuming the matrices are Gaussian, the variance of each permanent can be bounded as

$$\operatorname{Var}[\operatorname{per}(M_{J',1} * M^*_{J',\sigma_p})] = \frac{j!}{m^{2j}}, \tag{1.14}$$

and

$$\operatorname{Var}[\operatorname{per}(|M_{\bar{J}',\sigma_u}|^2)] < \frac{(n-j)!}{m^{2(n-j)}} \sum_{l=0}^{n-j} \frac{1}{l!}. \tag{1.15}$$

This leads to two key results. The first is that the variance of c_j tends towards a constant value:

$$\operatorname{Var}[c_j] < \left(\frac{n!}{m^n}\right)^2 \frac{1}{e} \sum_{l=0}^{n-j} \frac{1}{l!} \tag{1.16}$$

$$\to \left(\frac{n!}{m^n}\right)^2 \text{ as } n \to \infty, \tag{1.17}$$

and the second is that the covariance for different values of j is zero:

$$\operatorname{Cov}[c_j, c'_j] = 0 \, \forall \, j \neq j'. \tag{1.18}$$

From this one can approximate the variance of the error as a geometric series, which as $n \to \infty$ tends towards the inequality

$$\operatorname{Var}[\Delta \operatorname{Pr}[S']] = \operatorname{Var}[\operatorname{Pr}[S'] - \operatorname{Pr}_k[S']] \tag{1.19}$$

$$= \operatorname{Var}\left[\sum_{j=k+1}^{n} x^j \operatorname{Var}[c_j]\right] \tag{1.20}$$

$$< \left(\frac{n!}{m^n}\right)^2 \left(\frac{x^{2(k+1)}}{1 - x^2}\right), \tag{1.21}$$

where P_k is the probability distribution when truncated at $j \leq k$.

Finally one can use a Markov inequality to show that if the variance of the error is of the form $(n!/m^n)^2 \epsilon^2$, the average error of the simulation is at most ϵ [29]. Crucially, this value of ϵ is only dependent on x and k and no longer dependent on n. This means that for any value of x, one can choose a suitable value of k to achieve a required error ϵ, and run a classical simulation in time polynomial in n.

Although this runtime is polynomial in terms of n and can therefore be considered asymptotically efficient, it might not be classically simulable in practice. There are three main contributions to this: First, the algorithm is reliant on Metropolised Independence Sampling, which potentially requires many probabilities to be approximated per sample. Second, approximating each probability requires $O(n^{2k})$ permanents of $k \times k$ matrices, which even for small k could be a large number of permanents. And third, approximating each probability requires $O(n^{2k})$ permanents of $(n-k) \times (n-k)$ matrices with non-negative elements. Although approximating permanents of matrices with non-negative elements can be achieved in polynomial time, classical algorithms still have a runtime ranging from $O((n-k)^4 \log(n-k))$ to $O((n-k)^7 \log^4(n-k))$, depending on the sparsity of the matrix [38]. These issues are the main points to address in order to achieve a practical classical algorithm for Boson Sampling. Clifford & Clifford could help to alleviate these issues, but there is a challenge due to the fact that the approximation used in Renema et al. does not correspond to a bosonic state. This in turn leads to negative probabilities, which are not clear how to correct for the Clifford & Clifford algorithm.

1.2.4 Classical simulation of lossy Boson Sampling

Another common source of imperfections in linear optics is that of photon loss, which arises through a number of different means. Indeed, any large-scale demonstration of Boson Sampling is bound to face photon loss, and therefore needs to take such issues into account. Some results have already shown instances where hardness is still retained, such as when only a constant number of photons are lost [39, 40].

Neville *et al.* compared the classical simulation of their approach to a Boson Sampling experiment where any photon loss was considered a rejected experiment [13]. In [29], the method described in Sec. 1.2.3 was adapted to consider uniform loss, showing that the same result can be found, with the only difference being that x is now replaced by $\alpha = \sqrt{\eta}x$, where η is the probability of each individual photon surviving. Crucially, this result demonstrated that Boson Sampling where a constant fraction of photons were lost can be simulated in $O(\ell^{2k}k2^k)$, where ℓ is the number of photons which survive and k is only dependent on the constant ℓ/n, distinguishability x, and the desired accuracy of the simulation. This can be expanded to classically simulating Boson Sampling under uniform loss by sampling ℓ from the binomial distribution before sampling output photons, which offers a runtime of $O(n^{2k}k2^k)$. Novel classical simulations for Boson Sampling under loss have also been considered by use of classically simulable states such as thermal [34] or separable [35] states.

There has also been some consideration of how classical simulations can be generalised to non-uniform loss. This usually means photon loss that is dependent on the number of optical components, with each component having transmission probability τ. Classical simulation methods can be generalised to this model by identifying a layer of uniform losses from the circuit, followed a non-uniform lossy circuit which can be simulated classically through the use of additional modes for lost photons [34, 35]. These results showed that Boson Sampling under non-uniform loss can be classically simulated as long as the smallest number of components a photon encounters is logarithmic in n.

1.3 Expanding in terms of states

We will now introduce a new expression for partially distinguishable particles. We start by writing the input state of n photons, with pairwise distinguishability parameter x as in the previous section, in first quantization

$$\rho_{n,x} = \frac{1}{n!} \left(\sum_{\sigma,\sigma' \in S_n} \sigma \left| s \right\rangle \left\langle s \right| \sigma' x^{\sigma \cdot \sigma'} \right), \tag{1.22}$$

where we have used $\sigma \cdot \sigma'$ to denote the number of places where permutations σ and σ' match. For reference, the expansion of [28, 29] is carried out by identifying σ and σ' that match for a fixed set of i points:

$$\rho_{n,x} = \frac{1}{n!} \left(\sum_{i=0}^{n} x^i \sum_{\substack{\sigma,\sigma' \in S_n \\ \exists I \subseteq [n] \text{ s.t. } |I|=i \\ \sigma^{-1}(j) \neq \sigma'^{-1}(j) \forall j \in I \\ \sigma^{-1}(j) = \sigma'^{-1}(j) \forall j \notin I}} \sigma \left| s \right\rangle \left\langle s \right| \sigma'^{\dagger} \right). \tag{1.23}$$

Note here that the sums over permutations do not correspond to physical states. This can be seen by the fact that for $i \neq 0$ this summation has no elements along the diagonal of the density matrix, as σ and σ' need to differ in *exactly* i places.

We instead look at an alternative expansion, in order to decompose the model into a linear combination of physical states:

$$\rho_{n,x} = \frac{1}{n!} \left(\sum_{i=0}^{n} p_i \sum_{\substack{\sigma,\sigma' \in S_n \\ \exists I \subseteq [n] \text{ s.t. } |I|=i \\ \sigma^{-1}(j) = \sigma'^{-1}(j) \forall j \notin I}} \sigma \left| s \right\rangle \left\langle s \right| \sigma'^{\dagger} \right) \tag{1.24}$$

$$= \sum_{i=0}^{n} p_i \sum_{\substack{I \subseteq [n] \\ |I|=i}} \left(\frac{1}{n!} \sum_{\substack{\sigma,\sigma' \in S_n \\ \sigma^{-1}(j) = \sigma'^{-1}(j) \forall j \notin I}} \sigma \left| s \right\rangle \left\langle s \right| \sigma'^{\dagger} \right) \tag{1.25}$$

$$= \sum_{i=0}^{n} p_i \sum_{\substack{I \subseteq [n] \\ |I|=i}} \rho_I, \tag{1.26}$$

where ρ_I is the state where photons in modes $j \in I$ are fully indistinguishable from each other, all other photons are fully distinguishable, and $0 \leq p_i \leq 1$ is a coefficient dependent on x and n determining the probability of a state with i indistinguishable single photons.

Note that unlike Eq. (1.23), where permutations must differ in exactly i points, in Eq. (1.24) we allow permutations to differ in *at most* i points. This means that elements closer to and along the diagonal of the density matrix are also part of this summation, and this means that each sum over $\sigma, \sigma' \in S_n$ forms a valid density matrix.

Already we can see how a classical simulation might work – if we are able to sample p_i efficiently, then we can choose ρ_I by selecting i photons uniformly at random to be indistinguishable. These i photons can be classically simulated using Clifford & Clifford [14], while the remaining $n - i$ photons are treated as fully distinguishable photons, each of which can be simulated individually in polynomial time [31, 13].

1.3.1 The p_i are binomially distributed

Here, we will show that the coefficients p_i follow the binomial distribution

$$p_i = x^i(1-x)^{n-i}. \tag{1.27}$$

To see that the matrix elements of Eq. (1.26) with p_i binomially distributed equal those of Eq. (1.23), consider σ, σ' which differ at points in the set I, where $|I| = i$; the coefficient here should be x^i. Contributing to this element of the density matrix will be the state ρ_I, as well as other states $\rho_{I'}$, where $I \subseteq I'$. The number of such sets I' is $\binom{n-i}{i'-i}$, as it is equivalent to choosing i' from n elements when i elements have already been chosen. The corresponding matrix element is

$$\frac{1}{n!}\left(\sum_{i'=i}^{n} x^{i'}(1-x)^{n-i'}\binom{n-i}{i'-i}\right) \tag{1.28}$$

$$= \frac{1}{n!}\left(x^i\sum_{i'=i}^{n} x^{i'-i}(1-x)^{n-i'}\binom{n-i}{i'-i}\right) \tag{1.29}$$

$$= \frac{1}{n!}\left(x^i\sum_{i'=0}^{n-i} x^{i'}(1-x)^{n-i-i'}\binom{n-i}{i'}\right) \tag{1.30}$$

$$= \frac{x^i(x+1-x)^{n-i}}{n!} \tag{1.31}$$

$$= \frac{x^i}{n!}. \tag{1.32}$$

It is not hard to see that the state is normalised

$$\mathrm{Tr}[\rho_{n,x}] = \sum_{i=0}^{n} x^i(1-x)^{n-i}\sum_{\substack{I\subseteq[n]\\|I|=i}} \mathrm{Tr}[\rho_I] \tag{1.33}$$

$$= \sum_{i=0}^{n} x^i(1-x)^{n-i}\sum_{\substack{I\subseteq[n]\\|I|=i}} 1 \tag{1.34}$$

$$= \sum_{i=0}^{n} x^i(1-x)^{n-i}\binom{n}{i} \tag{1.35}$$

$$= (x+1-x)^n \tag{1.36}$$

$$= 1. \tag{1.37}$$

Thus this model of fixed pairwise distinguishability can be written as an expansion in terms of valid states, where indistinguishable photons are drawn from a binomial distribution.

1.3.2   Classical simulation

We can now see explicitly how a simulation for Boson Sampling with distinguishable photons would work. First, we sample an integer $i \in [n]$ according to the Binomial distribution with coefficients n and x. Next, we sample a subset I of the photons uniformly at random from the $\binom{n}{i}$ possible subsets of size i. These are the indistinguishable photons of our simulation, which we simulate using Clifford and Clifford in $O(i2^i + \text{poly}(i, m))$ time. The remaining $n - i$ photons are considered to be distinguishable. Rather than needing to compute the output probabilities of these photons colletively, which could take between $O(n-i)^4 \log(n-i)$ and $O(n-i)^7 \log^4(n-i)$ time via permanents of matrices with non-negative entries [38], we can instead sample each distinguishable photon individually. To do so, we take a distinguishable photon in mode a, and compute the probability of this photon being measured in mode b as $|U_{b,a}|^2$. Thus we can compute all output probabilities and obtain a sample for a single distinguishable photon in $O(m)$ time, meaning that we can obtain a sample for all $n - i$ distinguishable photons in $O(m(n-i))$ time [31, 13].

The run time is dominated by the time taken to sample our indistinguishable photons, which can be as large as $O(n2^n + \text{poly}(n, m))$ if we are unlucky. By truncating our Binomial sampling up to some level k, we can simulate Boson Sampling up to some level of error. The extent of this error will be the focus of Secs. 1.4 & 1.5.

1.4   Average case error

We can use the same strategies used in [28, 29] to derive an error bound for Boson Sampling via state truncation for a Haar-random interferometer. We shall do this by considering the total variation distance between our approximation and the model for partial distinguishability for a Gaussian matrix. This is given by

$$\mathbb{E}[\Delta P] = \sum_{S'} \mathbb{E} \, | \Pr[S'] - \Pr_k[S']|, \tag{1.38}$$

where $\Pr_k$ is the probability distribution truncated at k indistinguishable photons via our approximation. For a specific outcome S', we can expand the right hand side to

$$\mathbb{E} \left[| \Pr[S'] - \Pr_k[S']| \right] = \mathbb{E} \left| \sum_{i=0}^{k} (p_i - p_i') \sum_{\substack{I \subseteq [n] \\ |I| = i}} P_I[S'] \right. $$
$$\left. + \sum_{i=k+1}^{n} p_i \sum_{\substack{I \subseteq [n] \\ |I| = i}} P_I[S'] \right|, \tag{1.39}$$

where P_I is now the probability distribution with indistinguishable photons defined by set I, and $p_i' = p_i / (\sum_{i=0}^{k} x^i (1-x)^{n-i} \binom{n}{i})$ is the normalised version of the p_i coefficients defined in Sec. 1.3. Note that P_I is akin to the distribution arising from state ρ_I in Sec. 1.3. We can use the triangle

inequality to bound this value to

$$\mathbb{E}\left[|\Pr[S'] - \Pr_k[S']|\right] \leq \mathbb{E}\left|\sum_{i=0}^{k}(p_i - p_i')\sum_{\substack{I\subseteq[n]\\|I|=i}}P_I[S']\right|$$

$$+ \mathbb{E}\left|\sum_{i=k+1}^{n}p_i\sum_{\substack{I\subseteq[n]\\|I|=i}}P_I[S']\right| \tag{1.40}$$

$$= \mathbb{E}[\Delta P_{\leq k}] + \mathbb{E}[\Delta P_{>k}], \tag{1.41}$$

where we have introduced variables $\Delta P_{\leq k}$ and $\Delta P_{>k}$ for convenience. We shall consider the expected values of these terms for a Gaussian matrix separately, starting with the latter. Using the Laplace expansion, we find that

$$\mathbb{E}[\Delta P_{>k}] = \sum_{i=k+1}^{n}p_i\sum_{\substack{I\subseteq[n]\\|I|=i}}\mathbb{E}\left[P_I[S']\right], \tag{1.42}$$

$$= \sum_{i=k+1}^{n}p_i\sum_{\substack{I\subseteq[n]\\|I|=i}}\sum_{\substack{J\leq S'\\|J|=i}}\mathbb{E}\left[|\operatorname{per}(U_{I,J})|^2\operatorname{per}(|U_{\bar{I},\bar{J}}|^2)\right], \tag{1.43}$$

where $U_{I,J}$ is a matrix defined from our interferometer U by selecting columns according to I and rows according to J, and $|U_{\bar{I}\bar{J}}|^2$ is a matrix whose elements are the absolute values squared of $U_{\bar{I},\bar{J}}$.

We next need to consider the expected values of the matrix permanents in Eq. (1.43) for a Haar random unitary. To do this, we shall assume the matrix describing our interferometer is Gaussian. This allows us to assume that each entry of $U_{I,J}$ and $|U_{\bar{I},\bar{J}}|^2$ is independent, and that the two matrices are independent of each other. Starting with $\operatorname{per}(|U_{\bar{I},\bar{J}}|^2)$, we note that this is a Gaussian matrix of size $(n-i)\times(n-i)$, and each entry is the square of two independent Gaussians, meaning that each entry of $|U_{\bar{I},\bar{J}}|^2$ has expected value $1/m$. From this, we can calculate the expected value as

$$\mathbb{E}[\operatorname{per}(|U_{\bar{I},\bar{J}}|^2)] = \frac{(n-i)!}{m^{n-i}}. \tag{1.44}$$

For $U_{I,J}$, we note that each element of $U_{I,J}$ is an independent Gaussian entry, with mean value 0 due to symmetry, and second order moment $\mathbb{E}[|U_{i,j}|^2] = 1/m$. The second order moment

for the permanent can then be calculated using the same methods as in [3]:

$$\mathbb{E}[|\operatorname{per}(U_{I,J})|^2] = \mathbb{E}\left[\sum_{\sigma,\sigma'\in S_i}\prod_{l=1}^{n}(U_{I,J})_{l,\sigma(l)}(U^*_{I,J})_{l,\sigma'(l)}\right] \tag{1.45}$$

$$= \mathbb{E}\left[\sum_{\sigma\in S_i}\prod_{l=1}^{n}|(U_{I,J})_{l,\sigma(l)}|^2\right] \tag{1.46}$$

$$= \sum_{\sigma\in S_i}\prod_{l=1}^{n}\mathbb{E}\left[|U_{l,\sigma(l)}|^2\right] \tag{1.47}$$

$$= \frac{i!}{m^i}. \tag{1.48}$$

Because the two matrices are independent [28], we can express the expected value of their product as

$$\mathbb{E}\left[|\operatorname{per}(U_{I,J})|^2\operatorname{per}(|U_{\bar{I},\bar{J}}|^2)\right] = \frac{(n-i)!}{m^{n-i}}\times\frac{i!}{m^i} \tag{1.49}$$

$$= \frac{(n-i)!i!}{m^n}. \tag{1.50}$$

Plugging this into Eq. (1.43), we find that

$$\mathbb{E}[\Delta P_{>k}] = \sum_{i=k+1}^{n}p_i\sum_{\substack{I\subseteq[n]\\|I|=i}}\sum_{\substack{J\leq S'\\|\bar{J}|=j}}\frac{(n-i)!i!}{m^n} \tag{1.51}$$

$$= \sum_{i=k+1}^{n}p_i\binom{n}{i}^2\frac{(n-i)!i!}{m^n} \tag{1.52}$$

$$= \frac{n!}{m^n}\sum_{i=k+1}^{n}p_i\binom{n}{i} \tag{1.53}$$

$$= \frac{n!}{m^n}\sum_{i=k+1}^{n}x^i(1-x)^{n-i}\binom{n}{i}, \tag{1.54}$$

where first we use the fact that since our input and output are both collision-free, there are $\binom{n}{i}$ ways of choosing I and J, then apply cancellation, and finally substitute the values of p_i. This gives our error bound for terms not in our approximation.

Next we shall consider the term $\Delta P_{\leq k}$. First we can note that $p'_i\geq p_i\forall i\leq k$, due to the normalisation of p'_i. Thus we can rewrite this term as

$$\mathbb{E}[\Delta P_{\leq k}] = \sum_{i=0}^{k}(p'_i-p_i)\sum_{\substack{I\subseteq[n]\\|I|=i}}\mathbb{E}\left[P_I[S']\right]. \tag{1.55}$$

Using the same techniques for computing the Laplace expansion and calculating the expected

value of permanents of Gaussian matrices, we can show that

$$\mathbb{E}[\Delta P_{\leq k}] = \sum_{i=0}^{k} (p_i' - p_i) \binom{n}{i}^2 \frac{(n-i)!i!}{m^n} \tag{1.56}$$

$$= \frac{n!}{m^n} \sum_{i=0}^{k} (p_i' - p_i) \binom{n}{i}. \tag{1.57}$$

Next we expand p_i':

$$\sum_{i=0}^{k} p_i' \binom{n}{i} = \frac{\sum_{i=0}^{k} x^i (1-x)^{n-i} \binom{n}{i}}{\sum_{i=0}^{k} x^i (1-x)^{n-i} \binom{n}{i}} \tag{1.58}$$

$$= 1, \tag{1.59}$$

and use this expansion as well as the value of p_i to calculate

$$\mathbb{E}[\Delta P_{\leq k}] = \frac{n!}{m^n} \left(1 - \sum_{i=0}^{k} x^i (1-x)^{n-i} \binom{n}{i} \right) \tag{1.60}$$

$$= \frac{n!}{m^n} \sum_{i=k+1}^{n} x^i (1-x)^{n-i} \binom{n}{i}. \tag{1.61}$$

Finally, we use these values and sum over all collision-free S', of which there are $\binom{m}{n} \approx m^n/n!$ to bound our total variation distance for a Haar-random unitary as

$$\mathbb{E}[\Delta P] \leq 2 \sum_{i=k+1}^{n} x^i (1-x)^{n-i} \binom{n}{i}. \tag{1.62}$$

It seems like there should be some room for improvement in this bound. In particular, the use of the triangle inequality suggests that there might be more precise approximations of the expected distance.

1.5 Worst case error

We shall now give an improved error bound, using a different technique. This bound will be an improvement in two ways: first by improving the error bound by a factor of two, and second by being a bound for the worst-case error of any linear-optical interferometer, rather than a bound for a Haar-random unitary interferometer. We do so by finding an upper bound for the trace distance between our ideal partially distinguishable state and the approximation that results from truncating at some k, the size of the largest indistinguishable set of particles. As the trace distance is an upper bound for any POVM measurement, we know that this will provide an upper bound for the difference in distribution produced by any interferometer.

Denoting our truncated state

$$\rho_{\leq k, x} = \frac{\sum_{i=0}^{k} x^i (1-x)^{n-i} \sum_{\substack{I \subseteq [n] \\ |I|=i}} \rho_I}{\sum_{i=0}^{k} x^i (1-x)^{n-i} \binom{n}{i}}, \tag{1.63}$$

where the denominator is a normalising factor, and similarly

$$\rho_{>k,x} = \frac{\sum_{i=k+1}^{n} x^i (1-x)^{n-i} \sum_{\substack{I \subseteq [n] \\ |I|=i}} \rho_I}{\sum_{i=k+1}^{n} x^i (1-x)^{n-i} \binom{n}{i}}. \tag{1.64}$$

We can now estimate the trace distance by rewriting $\rho_{n,x}$ as

$$\rho_{n,x} = \sum_{i=0}^{k} x^i (1-x)^{n-i} \binom{n}{i} \rho_{\leq k,x} \tag{1.65}$$

$$+ \sum_{i=k+1}^{n} x^i (1-x)^{n-i} \binom{n}{i} \rho_{>k,x}, \tag{1.66}$$

which, by convexity, gives

$$\delta_{\mathrm{tr}}(\rho_{n,x}, \rho_{\leq k,x}) \leq \sum_{i=k+1}^{n} x^i (1-x)^{n-i} \binom{n}{i}. \tag{1.67}$$

If we want this bound to be small, we can use techniques like those used for working out the tails of the binomial distribution. For example, it is known that for $nx < k < n$ that

$$\delta_{\mathrm{tr}}(\rho_{n,x}, \rho_{\leq k,x}) \leq \exp\left(-nD\left(\frac{k}{n}||x\right)\right), \tag{1.68}$$

where $D(k/n||x)$ is the relative entropy between coins with bias k/n and x, respectively [41]. Choosing a value of $k = n\alpha$ for $\alpha > x$ will give an error bound of $\exp(-nD(\alpha||x))$. Thus, choosing such a value for k would give an error that decreases as n increases, albeit at the cost of needing to increase k linearly with n. While this is still an exponential time algorithm, it would offer a polynomial speedup over the Clifford & Clifford method for Boson Sampling with indistinguishable photons.

We can also note that the trace distance is only dependent on the initial states and not the measurement outcomes. As a result, this error bound also applies in the case where the output is not collision free.

1.6 Incorporating loss

We now consider how to adapt this simulation to Boson Sampling under uniform loss. We shall assume that each photon survives with probability η.

In [35, 32], it was shown that the initial state for Boson Sampling with a fixed number of lost photons can be represented in the first quantisation as the initial state

$$\frac{1}{\binom{n}{\ell}} \sum_{\substack{L \subseteq [n] \\ |L|=\ell}} \frac{1}{\ell!} \sum_{\sigma,\sigma' \in S_n} \sigma |s_L\rangle \langle s_L| \sigma'^\dagger, \tag{1.69}$$

where s_L is the state where photons in the subset L of the original input photons have survived. In order to generalise this to uniform loss, we append $n-\ell$ "lost" photons in an additional spatial mode (single particle state 0) which isn't affected by the interferometer:

$$\left(\frac{1}{\binom{n}{\ell}} \sum_{\substack{L \subseteq [n] \\ |L|=\ell}} \frac{1}{\ell!} \sum_{\sigma,\sigma' \in S_n} \sigma |s\rangle \langle s| \sigma'^\dagger \right) \otimes (|0\rangle \langle 0|)^{\otimes n-\ell}. \tag{1.70}$$

Note that in the same way that it doesn't matter which particles are traced out when initially applying the loss, it similarly doesn't matter which particles are replaced with the $|0\rangle\langle 0|$ state. Uniform loss matches that of choosing which subset of photons survive according to the binomial distribution [35, 29]. We can combine this model with the distinguishability model of Sec. 1.3, giving

$$\rho_{n,\eta,x} = \sum_{\ell=0}^{n} \eta^{\ell}(1-\eta)^{n-\ell} \sum_{\substack{L\subseteq[n]\\|L|=\ell}} \sum_{i=0}^{\ell} x^{i}(1-x)^{\ell-i} \sum_{\substack{I\subseteq L\\|I|=i}} \left(\frac{1}{\ell!} \sum_{\substack{\sigma,\sigma'\in S_n\\\sigma^{-1}(j)=\sigma'^{-1}(j)\forall j\notin I}} \sigma|s\rangle\langle s|\sigma'^{\dagger} \right) \otimes (|0\rangle\langle 0|)^{\otimes n-\ell}$$ (1.71)

$$= \sum_{\ell=0}^{n} \eta^{\ell}(1-\eta)^{n-\ell} \sum_{\substack{L\subseteq[n]\\|L|=\ell}} \sum_{i=0}^{\ell} x^{i}(1-x)^{\ell-i} \sum_{\substack{I\subseteq L\\|I|=i}} \rho_{L,I}.$$ (1.72)

We can now see how our classical simulation for Boson Sampling under distinguishability can be adapted to accommodate loss as well. First, we choose a subset of photons L to indicate the photons that were not lost. From this subset, we choose another subset of photons $I \subseteq L$ to indicate the indistinguishable photons, which are simulated via the Clifford & Clifford algorithm. The photons in $L \setminus I$ are all distinguishable photons, and can be simulated classically as before. The classical complexity of this algorithm depends on the number of indistinguishable photons we choose. As in Section 1.3, by truncating this to be some maximum size k we can get an algorithm that runs in $O(k2^k + \text{poly}(k,m))$ time.

To understand the precision of this algorithm, we first note that if $|L| \leq k$, then we can classically simulate any number of indistinguishable photons within our desired runtime. As a result, we only need to truncate when $|L| > k$, and only need to do so up to $|I| \leq k$.

As with Sec. 1.4, we start by considering the error bound for a random interferometer. We note that in cases where at most k photons survive our approximation is exact, so these outcomes do not contribute to our total variation distance. For the remainder, we see that

$$\mathbb{E}[\Delta P] = \sum_{\ell=k+1}^{n} \eta^{\ell}(1-\eta)^{n-\ell} \sum_{\substack{L\subseteq[n]\\|L|=\ell}} \mathbb{E}[\Delta P_L],$$ (1.73)

where ΔP_L denotes the error of our simulation with photons in input modes denoted by L. Using our bound in Sec. 1.4 as well as the rule of conditional binomial distributions (see below), we can bound this as

$$\mathbb{E}[\Delta P] \leq 2 \sum_{i=k+1}^{n} (\eta x)^{i}(1-\eta x)^{n-i} \binom{n}{i}.$$ (1.74)

Again, an improvement over the use of the triangle inequality can lead to an improvement in this bound.

For the worst-case error, we construct analogous states to those in Sec. 1.5:

$$\rho_{\leq k,\eta,x} = \frac{\sum_{\ell=0}^{n} \eta^{\ell}(1-\eta)^{n-\ell} \sum_{\substack{L\subseteq[n]\\|L|=\ell}} \sum_{i=0}^{\min(\ell,k)} x^{i}(1-x)^{\ell-i} \sum_{\substack{I\subseteq L\\|I|=i}} \rho_{L,I}}{\sum_{\ell=0}^{n} \eta^{\ell}(1-\eta)^{n-\ell} \binom{n}{\ell} \sum_{i=0}^{\min(\ell,k)} x^{i}(1-x)^{\ell-i} \binom{\ell}{i}},$$ (1.75)

$$\rho_{>k,\eta,x} = \frac{\sum_{\ell=k+1}^{n} \eta^{\ell}(1-\eta)^{n-\ell} \sum_{\substack{L\subseteq[n]\\|L|=\ell}} \sum_{i=k+1}^{\ell} x^{i}(1-x)^{\ell-i} \sum_{\substack{I\subseteq L\\|I|=i}} \rho_{L,I}}{\sum_{\ell=k+1}^{n} \eta^{\ell}(1-\eta)^{n-\ell} \binom{n}{\ell} \sum_{i=k+1}^{\ell} x^{i}(1-x)^{\ell-i} \binom{\ell}{i}},$$ (1.76)

and note that $\rho_{n,\eta,x}$ is a linear combination of these states. As a result, the worst-case error of this simulation, using the convex properties of the trace distance, can be bounded as

$$\delta_{\mathrm{tr}}(\rho_{n,\eta,x}, \rho_{\leq k,\eta,x}) \leq \sum_{\ell=k+1}^{n} \eta^{\ell}(1-\eta)^{n-\ell}\binom{n}{\ell}$$

$$\times \sum_{i=k+1}^{\ell} x^{i}(1-x)^{\ell-i}\binom{\ell}{i} \tag{1.77}$$

$$= \sum_{i=k+1}^{n} (\eta x)^{i}(1-\eta x)^{n-i}\binom{n}{i}, \tag{1.78}$$

where in the second line we have used the rule of conditional binomial distributions. Using the same result as used for Eq. (1.68), we can bound the error to a constant value if $k = n\alpha$ for $\alpha > \eta x$ [41]. This shows a relationship between distinguishability and loss similarly to, but not exactly the same, as the one found in [29]: the more distinguishable photons are, the more we can classically simulate photon loss, and vice versa. It is remarkable to see that these two algorithms have a different dependence on x and η: while state truncation depends on ηx, point truncation depends on ηx^2. It is not immediately clear where this difference comes from, and we leave it as an open question.

1.7 Empirical Errors

A natural question at this point is how to assess the performance of this new approach over that of [28, 29]. It is not immediately clear how to find a fair comparison, as each approach has its own strengths and weaknesses. Truncating based on fixed points has the benefit of the error asymptotically tending towards a constant as n increases, which means that k can be chosen independently of n and does not need to increase. But this comes at the cost of a potentially large, albeit polynomial, runtime of at least $O(n^{2k}k2^k(n-k)^4 \log(n-k))$ [28, 38]. Truncating based on states, on the other hand, provides a significant improvement in runtime based on k, and is able to run in $O(2^k + \mathrm{poly}(k,m,n))$ time, but at the cost of k increasing linearly with n for constant error.

We therefore consider a variety of comparisons. In Sec. 1.7.1, we start by considering the highest value of x and η simulable by each approach when given the same values of n and k. We then introduce the runtime for each algorithm in Sec. 1.7.2, by comparing how fast they can simulate particular values of x and η for increasing n. Finally in Sec. 1.7.3, we compare the highest value of x and η simulable by both algorithms for a 90-photon Boson Sampling experiment, where k is varying but under the condition that the algorithms have similar run times. The motivation for this is that 90 photons has been suggested as strict upper bound for what is achievable using classical computation [42].

Before we go further, we make a few observations on the calculations of error bounds and runtimes used in this section. Rather than using the asymptotic error bounds for fixed point truncation, which assume $n \to \infty$, we have used bounds for finite n. This provides an improvement in the error of up to $1/\sqrt{e}$. For the runtime of fixed point truncation, we explicitly calculate $\sum_{i=0}^{k}\binom{n}{i}R(n,n-i)i2^i(n-i)^4 \log(n-i)$, where $R(n,n-i) = \binom{n}{i}\lceil i!/e \rceil$ is the number of permutations with $n-i$ fixed points, and we have assumed that computing every permanent of the $(n-i) \times (n-i)$ matrices with non-negative entries requires $O((n-i)^4 \log(n-i))$
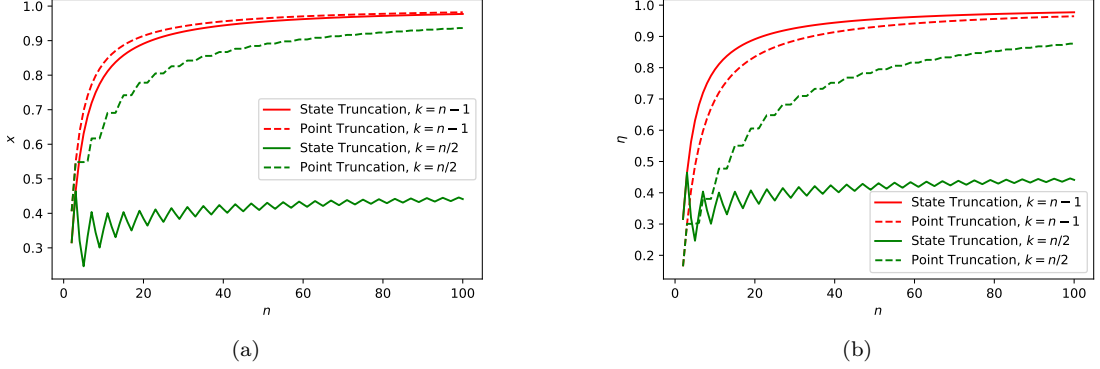
Figure 1.1: Highest value of (1.1a) x when $\eta = 1$ and (1.1b) of η when $x = 1$ simulable via state (solid) or point (dashed) truncation up to 10% error ($\epsilon = 0.1$). The number of photons, n, is varying, with k chosen as either $k = n - 1$ (red) or $k = n/2$ (green). The oscillatory behaviour is due to rounding $k = n/2$ when n is odd.

time [1]. For Metropolised Independence Sampling, we choose the number of probabilities to approximate via state truncation as 100, which matches currently used "burn in" and "thinning" times[13], though it is worth noting that this number could be different depending on the distribution of partially-distinguishable and lossy bosons. For the runtime of state truncation, we use $2k2^k + mk(k-1)/2 + m(n-k)$, where we have used the fact that the first term is approximately equivalent to computing two matrix permanents, the second term is the polynomial overhead of the Clifford & Clifford approach, and the final term is the polynomial overhead of sampling the fully distinguishable photons [14]. For these calculations, we have also assumed that $m = n^2$ [2].

Finally, we note that in Secs. 1.7.1 and 1.7.3, we only consider distinguishability and loss separately, by comparing the highest value of x simulable in a lossless system, and the highest value of η simulated in a full indistinguishable system. Ideally one would compare highest combinations of η and x which are classically simulable. However, doing so is complicated by the fact that both methods handle combinations of noise differently: point truncation handles them as the parameter ηx^2, whereas state truncation handles them as ηx. With this in mind, we plot values for distinguishability and loss separately, and note that, for the same performance, a reduction in one implies an increase in the other.

1.7.1 Comparison at same level of truncation

We start by comparing the performance of the two algorithms when truncated at the same level k. This is of interest as in both approaches k is considered to be a parameter defining the interference between photons. To do so, we consider the error bounds of classically simulating n-photon Boson Sampling for n ranging between 2 and 100. The values chosen for k depend on n: we consider $k = n - 1$ as the upper limit of what the two algorithms can achieve without

---

[1]Note that this computation could in the worst case take $O(n-i)^7 \log^4(n-i)$ time, depending on the matrix sparsity [38].

[2]Note that $m \in O(n^2)$ is only sufficient to ensure that the probability of seeing collisions from a Haar-random interferometer are small [43]. The classical hardness of Boson Sampling is also dependent on entries of U being drawn independently with high probability. To ensure this, m could be required to be as large as n^6 [3]. However, it is widely believed, and often referenced in Boson Sampling experiments, that $m = n^2$ should be sufficient [3, 5, 6, 7, 8, 9, 10, 11].

simulating the full distribution, and also $k = n/2$ as a more feasible, though still exponential time, value.

The result is plotted in FIG. 1.1, where in (1.1a) we show the highest value of x simulable assuming no loss ($\eta = 1$) and in (1.1b) we show the highest value of η simulable assuming the photons are fully indistinguishable ($x = 1$). For all cases, we are considering simulations up to 10% error.

There are a number of things we can note from FIG. 1.1. First is that when $k = n - 1$, we can see that both algorithms tend to the same maximum values of distinguishability and loss. In the case of distinguishability, we can easily see why by considering the error bounds of both algorithms. One can see from Eq. (1.67) that state truncation will have a simple error bound in this case of $\epsilon \leq x^n$, meaning that for constant error the largest value of x simulable is $x = \epsilon^{1/n}$. For point truncation, Eq. (1.20) shows that the error is similarly bounded as $\epsilon \leq x^n/\sqrt{e}$, leading to a largest value of $x = (\epsilon\sqrt{e})^{1/n}$. Thus, although the highest value of x simulable via point truncation is higher than that via state truncation, the difference will decrease in the limit of large n. Curiously we see the same effect as well in the case of loss, but now the highest value of η simulable via state truncation is higher than that of point truncation. Again, this can be shown to hold theoretically: For state truncation the error scales as $\epsilon \leq \eta^n$ according to Eq. (1.78), corresponding to $\eta = \epsilon^{1/n}$; whereas for point truncation we see from Eq. (1.20) and substituting $x = \sqrt{\eta}$ that the error scales as $\epsilon \leq \eta^{n/2}/\sqrt{e}$, meaning a maximum value of η is $\eta = (e\epsilon^2)^{1/n}$. In the limit of large n these differences will also tail off.

For $k = n/2$, we see that for both distinguishability and loss point truncation is more powerful than state truncation. Although this is harder to formally prove, there is intuition to see why this is the case. For state truncation, we know that for a small error to be achievable we need $k \geq n\eta x$, as this is the mean of the binomial distribution. Thus for $k = n/2$, we have that $\eta x \leq 1/2$, and in both cases we see the highest value of x and η tending to a value below $1/2$. For point truncation on the other hand, we know that the error tends to a constant value only dependent on k and ηx^2 in the limit of large n. As a result, it is unsurprising that for k increasing linearly with n the highest values of x and η will increase.

1.7.2 Comparison of runtimes

We next consider the runtime required to simulate n-photon Boson Sampling up to 10% error via either method. The motivation for this comparison is that the runtime of the two algorithms at the same value of k are significantly different. In particular, the runtime of state truncation is only dependent on k and not n, whereas the runtime for point truncation depends on a scaling of approximately $O(n^{2k})$.

To understand how the runtimes scale, in FIG. 1.2 we plot the runtimes of classically simulating n-photon Boson Sampling experiments via the two approaches for fixed values of η and x. The values of k chosen for each algorithm are the smallest values for an error of at most 10%. For choosing η and x, we give two example cases. The first (FIG. 1.2, red), where $\eta = 0.755$ and $x = 0.975$, is an example of a hypothetical best experiment we could build with current technology, with the most lossless sources (82%) [44], interferometers (99%) [40] and detectors (93%) [45], and the highest level of photon indistinguishability (97.6%) [46]. The second case (FIG. 1.2, green), where $\eta = x = 0.5$, is an example of how the two algorithms perform in what would be considered a poor experiment for both distinguishability and loss. Actual Boson Sampling experiments are likely to fall between these two extremes.

In both cases, state truncation appears to outperform point truncation for near-term photon experiments, with point truncation eventually being able to perform faster for larger values of n. When $\eta = x = 0.5$, point truncation performs better when n is approximately larger than 230. In
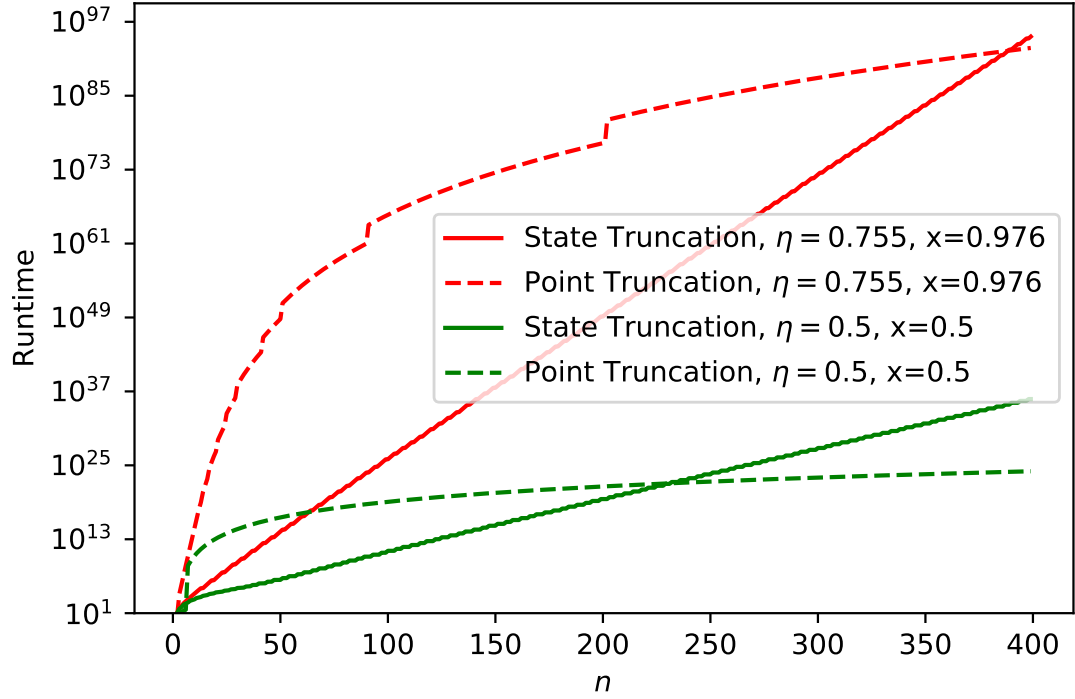
Figure 1.2: Approximated runtime in terms of number of operations to simulate n-photon Boson Sampling with chosen values of η and x up to 10% error ($\epsilon = 0.1$) via state (solid) or point (dashed) truncation.
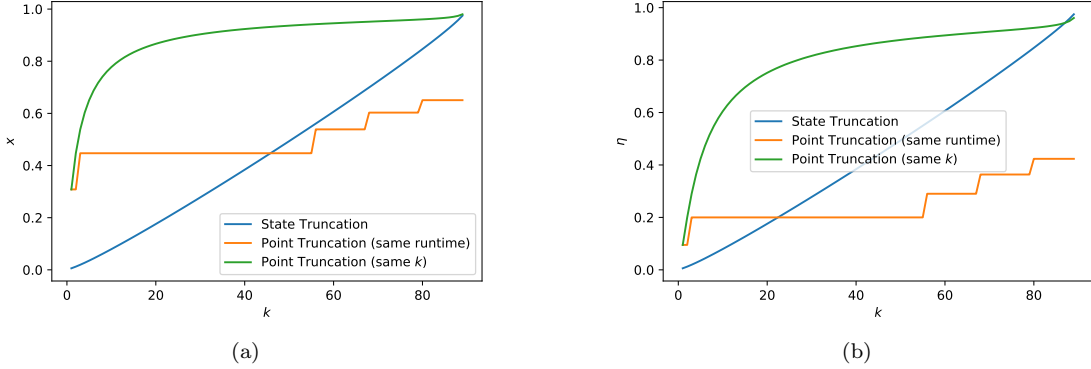
(a) (b)

Figure 1.3: Highest value of (1.3a) x when $\eta = 1$ and (1.3b) η when $x = 1$ simulable for 90-photon boson sampling at truncation level k up to 10% error ($\epsilon = 0.1$). Blue line indicates highest values simulable via state truncation at level k, green lines indicate highest values simulable via point truncation at level k, orange lines indicate highest values simulable at point truncation at level k' such that k' is the smallest level of truncation such that the approximate runtime of point truncation at level k' is longer than the runtime of state truncation at level k.

the case of $\eta = 0.755, x = 0.976$, point truncation only performs better when n is approximately larger than 390 photons. This gives an idea of the regions in which the polynomial runtime of point truncation can be better or worse than the exponential runtime of state truncation.

It is also worth noting that just because point truncation is faster than state truncation for large enough n does not necessarily mean that either algorithm is efficient in these cases. When $\eta = x = 0.5$, point truncation only becomes more efficient at instances where both algorithms already require the order of 10^{22} operations. And in the case where $\eta = 0.755, x = 0.976$, both algorithms have runtimes on the order of 10^{92} operations before point truncation outperforms state truncation.

1.7.3 Comparison at same runtime

Now we consider both truncation level and runtime, and compare the algorithms when restricted to comparable runtimes. To do this, we shall consider the challenge of simulating a 90-photon Boson Sampling experiment, and the largest values of distinguishability and loss that can be simulated at level k with 10% error. The motivation for this is that 90 photons has been suggested as strict upper bound for what is achievable using classical computation [42].

The results are shown in Fig. 1.3, detailing for each algorithm the highest value of x simulable when $\eta = 1$ (1.3a) and the highest value of η simulable when $x = 1$ (1.3b). In both figures, the blue line indicates state truncation at level k, and the green line indicates point truncation at level k. However, the runtime of state truncation at level k and point truncation at level k are likely to be drastically different. To take runtime into consideration as well, we consider the orange line which indicates point truncation at level k', where k' is the smallest integer such that the approximated runtime of point truncation at level k' is longer than that of state truncation at level k. This allows us to compare the performance of the two algorithms when restricted to similar runtimes.

Considering distinguishability in FIG. 1.3a, we can note that point truncation with comparable runtime performs better up to $k \leq 45$, after which the methods are roughly comparable with

state truncation performing marginally better, before becoming more dominant for $k \geq 60$. It has been suggested that boson sampling with 50 indistinguishable photons is roughly the limit of what can be classically simulated on a supercomputer [13, 14, 47], so it appears that when considering distinguishability, the algorithms are roughly comparable in this case.

When considering loss in FIG. 1.3b, we see a noticeable improvement for state truncation. Now point truncation under the same runtime only performs better up to $k \leq 22$, with state truncation performing considerably better for higher values of k. Boson Sampling with up to 30 indistinguishable photons is already known to be classically simulable on a standard laptop [13], so this appears to offer a noticeable improvement even for fast classical simulations.

1.8 Non-uniform loss

We finish by briefly considering non-uniform loss, where each photon survives a lossy optical component with probability τ. This model of loss has been considered before [34, 35], but without the incorporation of distinguishability. We can do this using the same methods as other non-uniform loss results, by extracting non-uniform losses into a layer of uniform losses followed by a lossy interferometer. The uniform loss layer means that each photon has probability $\eta = \tau^s$ of surviving, where τ is the loss of each optical component and s is the smallest number of lossy optical components a photon interacts with. If we take the total number of lossy components to be d, the remaining lossy circuit can be modelled as an $(m + d)$-mode interferometer, with lost photons ending up in the additional d modes. Thus we can achieve the same error as Eq. (1.78) in $O(k2^k + \text{poly}(k, m, d))$ time. In typical schemes for linear interferometers, d is at most polynomial in m [48, 49], so the overhead from these additional modes is small. We can bound the error to a constant if $k > nx\tau^s$. Taking the logarithm on both sides and rearranging for s, we find that this holds if

$$s > \frac{\log n - \log 1/x - \log k}{\log 1/\tau}. \tag{1.79}$$

This matches results in [34, 35], showing that boson sampling can be classically simulated if each photon encounters at least a logarithmic number of lossy components. It also shows how distinguishability can affect the simulability of lossy components in Boson Sampling: if our photons are more distinguishable, corresponding to a smaller value of x, then we can simulate shallower (i.e. less total loss) optical circuits.

1.9 Conclusion

In recent years significant improvements have been made in the ability of classical computers to simulate Boson Sampling under various imperfections. However, while it is of theoretical interest to demonstrate asymptotic improvements in classical simulation, the whole reason for proposals such as Boson Sampling is to offer speedups for near-term devices. Although our algorithm will not scale polynomially as the number of photons increases, we find that a substantial improvement over current classical algorithms can be achieved for the numbers of photons that experimentalists are currently aiming for. In doing so, we have effectively set a benchmark for what is required of a 50-90 photon Boson Sampling device.

There are a number of ways one could improve this classical simulation. In particular, the approach of Ref. [28] for truncation when looking at near-term devices is dependent on Metropolised Independence Sampling. A direct adaptation of the Clifford & Clifford algorithm to this approach would almost certainly offer an improvement over our algorithm. However, such an adaptation

is non-trivial, due to the fact that the term in the expansion are not states, something that motivated our work here.

There are other open questions we would like to consider as well. The first would be to improve the average-case error bounds to match, if not improve over, our worst-case error bound. This would most likely involve an alternative to using the triangle inequality. The second would be to find a way of explaining the difference in dependence on η and x between point and state truncation, and ideally improving either algorithm in the process.

1.9.1 Note Added

During this work we were made aware of independent work by V. Shchesnovich, which also shows that the model of distinguishability considered by Renema *et al.* corresponds to that of selecting indistinguishable photons via the binomial distribution [50]. This is derived using significantly different methods from those used in this manuscript, and does not consider classical simulation of distinguishability via the above method (though this has been anticipated [51]).

No underlying data was produced during this study.

Bibliography

[1] Ashley Montanaro. Quantum algorithms: an overview. *Npj Quantum Information*, 2:15023, 2016.

[2] Aram W. Harrow and Ashley Montanaro. Quantum computational supremacy. *Nature*, 549:203–209, 2017.

[3] Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. In *Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing*, STOC '11, pages 333–342, New York, NY, USA, 2011. ACM.

[4] Stefan Scheel and Stefan Yoshi Buhmann. Macroscopic quantum electrodynamics - concepts and applications. *Acta Physica Slovaca*, 58(5):675–809, 2008.

[5] Matthew A. Broome, Alessandro Fedrizzi, Saleh Rahimi-Keshari, Justin Dove, Scott Aaronson, Timothy C. Ralph, and Andrew G. White. Photonic boson sampling in a tunable circuit. *Science*, 339(6121):794–798, 2013.

[6] Justin B. Spring, Benjamin J. Metcalf, Peter C. Humphreys, W. Steven Kolthammer, Xian-Min Jin, Marco Barbieri, Animesh Datta, Nicholas Thomas-Peter, Nathan K. Langford, Dmytro Kundys, James C. Gates, Brian J. Smith, Peter G. R. Smith, and Ian A. Walmsley. Boson sampling on a photonic chip. *Science*, 339(6121):798–801, 2013.

[7] Max Tillmann, Borivoje Dakić, René Heilmann, Stefan Nolte, Alexander Szameit, and Philip Walther. Experimental boson sampling. *Nature Photonics*, 7:540, 2013.

[8] Andrea Crespi, Roberto Osellame, Roberta Ramponi, Daniel J. Brod, Ernesto F. Galvão, Nicolò Spagnolo, Chiara Vitelli, Enrico Maiorino, Paolo Mataloni, and Fabio Sciarrino. Integrated multimode interferometers with arbitrary designs for photonic boson sampling. *Nature Photonics*, 7:545, 2013.

[9] Jacques Carolan, Christopher Harrold, Chris Sparrow, Enrique Martín-López, Nicholas J. Russell, Joshua W. Silverstone, Peter J. Shadbolt, Nobuyuki Matsuda, Manabu Oguma, Mikitaka Itoh, Graham D. Marshall, Mark G. Thompson, Jonathan C. F. Matthews, Toshikazu Hashimoto, Jeremy L. O'Brien, and Anthony Laing. Universal linear optics. *Science*, 349(6249):711–716, 2015.

[10] Hui Wang, Yu He, Yu-Huai Li, Zu-En Su, Bo Li, He-Liang Huang, Xing Ding, Ming-Cheng Chen, Chang Liu, Jian Qin, Jin-Peng Li, Yu-Ming He, Christian Schneider, Martin Kamp, Cheng-Zhi Peng, Sven Höfling, Chao-Yang Lu, and Jian-Wei Pan. High-efficiency multiphoton boson sampling. *Nature Photonics*, 11:361—365, 2017.

[11] Han-Sen Zhong, Yuan Li, Wei Li, Li-Chao Peng, Zu-En Su, Yi Hu, Yu-Ming He, Xing Ding, Weijun Zhang, Hao Li, Lu Zhang, Zhen Wang, Lixing You, Xi-Lin Wang, Xiao Jiang, Li Li, Yu-Ao Chen, Nai-Le Liu, Chao-Yang Lu, and Jian-Wei Pan. 12-photon entanglement and scalable scattershot boson sampling with optimal entangled-photon pairs from parametric down-conversion. *Phys. Rev. Lett.*, 121:250505, Dec 2018.

[12] Stefano Paesani, Yunhong Ding, Raffaele Santagati, Levon Chakhmakhchyan, Caterina Vigliar, Karsten Rottwitt, Leif K. Oxenløwe, Jianwei Wang, Mark G. Thompson, and Anthony Laing. Generation and sampling of quantum states of light in a silicon chip, 2018. arXiv:1812.03158v1.

[13] Alex Neville, Chris Sparrow, Raphael Clifford, Eric Johnston, Patrick M. Birchall, Ashley Montanaro, and Anthony Laing. Classical boson sampling algorithms with superior performance to near-term experiments. *Nature Physics*, 13:1153—1157, 2017.

[14] Peter Clifford and Raphaël Clifford. The classical complexity of boson sampling. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 146–155, 2018.

[15] Hubert de Guise, Si-Hui Tan, Isaac P. Poulin, and Barry C. Sanders. Coincidence landscapes for three-channel linear optical networks. *Phys. Rev. A*, 89:063819, Jun 2014.

[16] Vincenzo Tamma. Sampling of bosonic qubits. *International Journal of Quantum Information*, 12(07n08):1560017, 2014.

[17] V. S. Shchesnovich. Sufficient condition for the mode mismatch of single photons for scalability of the boson-sampling computer. *Physical Review A*, 89:022333, Feb 2014.

[18] Peter P. Rohde. Boson sampling with photons of arbitrary spectral structure. *Physical Review A*, 91:012307, Jan 2015.

[19] V. S. Shchesnovich. Partial indistinguishability theory for multiphoton experiments in multiport devices. *Phys. Rev. A*, 91:013844, Jan 2015.

[20] Vincenzo Tamma and Simon Laibacher. Multiboson correlation interferometry with arbitrary single-photon pure states. *Phys. Rev. Lett.*, 114:243601, Jun 2015.

[21] Simon Laibacher and Vincenzo Tamma. From the physics to the computational complexity of multiboson correlation interference. *Physical Review Letters*, 115:243605, Dec 2015.

[22] Malte C. Tichy. Sampling of partially distinguishable bosons and the relation to the multidimensional permanent. *Phys. Rev. A*, 91:022316, Feb 2015.

[23] Max Tillmann, Si-Hui Tan, Sarah E. Stoeckl, Barry C. Sanders, Hubert de Guise, René Heilmann, Stefan Nolte, Alexander Szameit, and Philip Walther. Generalized multiphoton quantum interference. *Physical Review X*, 5:041015, Oct 2015.

[24] Adrian J. Menssen, Alex E. Jones, Benjamin J. Metcalf, Malte C. Tichy, Stefanie Barz, W. Steven Kolthammer, and Ian A. Walmsley. Distinguishability and many-particle interference. *Physical Review Letters*, 118:153603, Apr 2017.

[25] Simon Laibacher and Vincenzo Tamma. Spectral correlations among interfering nonidentical photons in universal linear optics, 2017. arXiv:1706.05578v2.

[26] Nicolò Spagnolo, Chiara Vitelli, Marco Bentivegna, Daniel J. Brod, Andrea Crespi, Fulvio Flamini, Sandro Giacomini, Giorgio Milani, Roberta Ramponi, Paolo Mataloni, Roberto Osellame, Ernesto F. Galvão, and Fabio Sciarrino. Experimental validation of photonic boson sampling. *Nature Photonics*, 8:615, 2014.

[27] Taira Giordani, Fulvio Flamini, Matteo Pompili, Niko Viggianiello, Nicolò Spagnolo, Andrea Crespi, Roberto Osellame, Nathan Wiebe, Mattia Walschaers, Andreas Buchleitner, and Fabio Sciarrino. Experimental statistical signature of many-body quantum interference. *Nature Photonics*, 12:173, 2018.

[28] 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, May 2018.

[29] Jelmer Renema, Valery Shchesnovich, and Raul Garcia-Patron. Classical simulability of noisy boson sampling, 2018. arXiv:1809.01953.

[30] Craig S. Hamilton, Regina Kruse, Linda Sansoni, Sonja Barkhofen, Christine Silberhorn, and Igor Jex. Gaussian boson sampling. *Phys. Rev. Lett.*, 119:170501, Oct 2017.

[31] Scott Aaronson and Alex Arkhipov. Bosonsampling is far from uniform. *Quantum Information & Computation*, 14(15-16):1383–1423, November 2014.

[32] Alexandra E. Moylett and Peter S. Turner. Quantum simulation of partially distinguishable boson sampling. *Phys. Rev. A*, 97:062329, Jun 2018.

[33] Stasja Stanisic and Peter S. Turner. Discriminating distinguishability. *Phys. Rev. A*, 98:043839, Oct 2018.

[34] Raúl García-Patrón, Jelmer J. Renema, and Valery Shchesnovich. Simulating boson sampling in lossy architectures, 2018. arXiv:1712.10037v2.

[35] Michał Oszmaniec and Daniel J Brod. Classical simulation of photonic linear optics with lost particles. *New Journal of Physics*, 20(9):092002, sep 2018.

[36] Dave Bacon, Isaac L. Chuang, and Aram W. Harrow. The quantum schur and clebsch-gordan transforms: I. efficient qudit circuits. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, pages 1235–1244, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.

[37] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, July 2004.

[38] Mark Huber and Jenny Law. Fast approximation of the permanent for very dense problems. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '08, pages 681–689, Philadelphia, PA, USA, 2008. Society for Industrial and Applied Mathematics.

[39] Scott Aaronson and Daniel J. Brod. Bosonsampling with lost photons. *Phys. Rev. A*, 93:012335, Jan 2016.

[40] Hui Wang, Wei Li, Xiao Jiang, Y.-M. He, Y.-H. Li, X. Ding, M.-C. Chen, J. Qin, C.-Z. Peng, C. Schneider, M. Kamp, W.-J. Zhang, H. Li, L.-X. You, Z. Wang, J. P. Dowling, S. Höfling, Chao-Yang Lu, and Jian-Wei Pan. Toward scalable boson sampling with photon loss. *Phys. Rev. Lett.*, 120:230502, Jun 2018.

[41] R. Arratia and L. Gordon. Tutorial on large deviations for the binomial distribution. *Bulletin of Mathematical Biology*, 51(1):125–131, Jan 1989.

[42] Alexander M. Dalzell, Aram W. Harrow, Dax Enshan Koh, and Rolando L. La Placa. How many qubits are needed for quantum computational supremacy? Technical Report MIT-CTP/5019, Massachusetts Institute of Technology, 2018.

[43] Alex Arkhipov and Greg Kuperberg. The bosonic birthday paradox, 2011. arXiv:1106.0849v1.

[44] Sergei Slussarenko, Morgan M. Weston, Helen M. Chrzanowski, Lynden K. Shalm, Varun B. Verma, Sae Woo Nam, and Geoff J. Pryde. Unconditional violation of the shot-noise limit in photonic quantum metrology. *Nature Photonics*, pages 1749–4893, 2017.

[45] F. Marsili, V. B. Verma, J. A. Stern, S. Harrington, A. E. Lita, T. Gerrits, I. Vayshenker, B. Baek, M. D. Shaw, R. P. Mirin, and S. W. Nam. Detecting single infrared photons with 93 *Nature Photonics*, pages 210—214, 2013.

[46] Yu-Ming He, Hui Wang, Stefan Gerhardt, Karol Winkler, Jonathan Jurkat, Ying Yu, Ming-Cheng Chen, Xing Ding, Si Chen, Jin Qian, Zhao-Chen Duan, Jin-Peng Li, Lin-Jun Wang, Yong-Heng Huo, Siyuan Yu, Sven Höfling, Chao-Yang Lu, and Jian-Wei Pan. Polarized indistinguishable single photons from a quantum dot in an elliptical micropillar, 2018. arXiv:1809.10992v1.

[47] Baida Zhang, Huiquan Wang, Yang Wang, Yong Liu, Junjie Wu, Xuejun Yang, and Xianmin Jin. A benchmark test of boson sampling on Tianhe-2 supercomputer. *National Science Review*, 5(5):715–720, 07 2018.

[48] Michael Reck, Anton Zeilinger, Herbert J. Bernstein, and Philip Bertani. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.*, 73:58–61, Jul 1994.

[49] William R. Clements, Peter C. Humphreys, Benjamin J. Metcalf, W. Steven Kolthammer, and Ian A. Walmsley. Optimal design for universal multiport interferometers. *Optica*, 3(12):1460–1465, Dec 2016.

[50] Valery Shchesnovich. Noise in BosonSampling and the threshold of efficient classical simulability, 2019. arXiv:1902.02258v4.

[51] V. S. Shchesnovich. personal communication, April 2019.