Classical simulation of photonic linear optics with lost particles

Michał Oszmaniec\(^1\) and Daniel J Brod\(^2\)

\(^1\) Institute of Theoretical Physics and Astrophysics, National Quantum Information Centre, Faculty of Mathematics, Physics and Informatics, University of Gdansk, Wita Stwosza 57, 80-952 Gdansk, Poland
\(^2\) Instituto de Física, Universidade Federal Fluminense, Av. Gal. Milton Tavares de Souza s/n, Niterói, RJ, 24210-340, Brazil

E-mail: michal.oszmaniec@gmail.com and danieljostbrod@id.uff.br

Keywords: boson sampling, quantum computing, quantum optics, linear optics, classical simulation

Abstract

We explore the possibility of efficient classical simulation of linear optics experiments under the effect of particle losses. Specifically, we investigate the canonical boson sampling scenario in which an \(n\)-particle Fock input state propagates through a linear-optical network and is subsequently measured by particle-number detectors in the \(m\) output modes. We examine two models of losses. In the first model a fixed number of particles is lost. We prove that in this scenario the output statistics can be well approximated by an efficient classical simulation, provided that the number of photons that is left grows slower than \(\sqrt{n}\). In the second loss model, every time a photon passes through a beamsplitter in the network, it has some probability of being lost. For this model the relevant parameter is \(s\), the smallest number of beamsplitters that any photon traverses as it propagates through the network. We prove that it is possible to approximately simulate the output statistics already if \(s\) grows logarithmically with \(m\), regardless of the geometry of the network. The latter result is obtained by proving that it is always possible to commute \(s\) layers of uniform losses to the input of the network regardless of its geometry, which could be a result of independent interest. We believe that our findings put strong limitations on future experimental realizations of quantum computational supremacy proposals based on boson sampling.

1. Introduction

Quantum computers are expected to offer an advantage in a wide variety of computational problems relative to their classical counterparts. However, experimental limitations mean quantum computers are notoriously hard to build, and to date it remains unclear which technological architecture is the most promising for this purpose. Thus, a full-purpose, large-scale universal quantum computer still seems like a long-term goal for the field.

With this in mind, a recent intermediate milestone was proposed in the form of the quantum computational supremacy paradigm \([1, 2]\). This consists of a series of proposed restricted models of quantum computing which are not expected to be universal, but for which there are strong arguments that they outperform classical computers in some computational task. One example of such proposals is boson sampling \([3]\), where the quantum device is restricted to single-photon states, linear optics and photon-number detectors, which we focus on in this work. Other examples include circuits of commuting gates \([4]\), a variant of the one-clean-qubit model \([5, 6]\), random quantum circuits \([7, 8]\), among others. These results follow a similar reasoning: one identifies some computational problem that is expected to be hard (such as computing the permanent of Gaussian matrices, in the case of boson sampling), posits a hardness conjecture regarding that problem, and shows formally how that conjecture implies that an efficient classical algorithm for simulating the physical system in question would have unexpected or surprising complexity-theoretic consequences. And so, the computational problem that is ‘solved’ by these restricted models is just to simulate their own behavior (according to the predictions of Quantum Mechanics), which a classical computer should not be able to do efficiently.

© 2018 The Author(s). Published by IOP Publishing Ltd on behalf of Deutsche Physikalische Gesellschaft
Shortly after the seminal boson sampling [3] paper was published, several works started investigating the effects of realistic experimental imperfections on the idealized theoretical model. The robustness of boson sampling was analyzed under the effect of partial photon distinguishability [9, 10], fabrication imperfections on the linear-optical transformation [11–13], losses [14, 15], probabilistic sources [16] and so on. On the experimental side, several small-scale implementations of boson sampling have been reported so far [17–28], with state-of-the-art implementations with up to four and five photons using near-deterministic quantum dot sources [25–27]. Through the development of the field there has been much interaction between the experimental and theoretical communities in order to assess the most important obstacles towards a large-scale implementation of boson sampling. One successful example of such an interaction was the conception of scattershot boson sampling [24, 29, 30], an alternative approach that allows circumvention of the exponential slowdown incurred due to the use of probabilistic sources. From the complementary perspective, recent works have pushed the development of classical algorithms for boson sampling [31, 32], and current supercomputers are expected to be able to simulate 50-photon boson sampling experiments without much difficulty [32, 33]. A more refined analysis of the underlying complexity-theoretic arguments has recently suggested 90 photons as a concrete milestone for the demonstration of quantum computational supremacy based on boson sampling [34].

One of the main obstacles that still hinders the scaling of boson sampling, both on the theoretical and experimental sides, is photon loss. In [14], the authors investigated the complexity of lossy boson sampling, and concluded that it remains as hard as standard boson sampling when a constant number of photons is lost. Unfortunately, this falls short of the realistic regime where we expect some fraction of the photons to be lost. Even worse, most current implementations [17–27] are expected to suffer from exponential losses. The intuition behind this claim is the following. Every time a photon traverses a beamsplitter, it has some probability of being lost, say \( (1 - \eta) \). And so, if it must traverse \( s \) beamsplitters in a network, the probability that it arrives at the output should decay as \( \eta^s \). The best known lower bound on the depth of a linear-optical circuit for it to satisfy the complexity requirements laid out in [3] is that it grows as \( n \log n \). Thus, the probability that each photon survives the experiment decays exponentially and quickly becomes negligible.

It is then a timely question to ask how much loss a boson sampling device can support before it admits an efficient classical simulation. To our knowledge, the best known bound of this type states that an efficient classical simulation becomes possible when all but \( \log n \) photons are lost (this was stated without proof in [14], but is a straightforward consequence of the more recent algorithm of [32]). In [15], the authors also investigate a similar question, and show that linear optics becomes classically simulable (via a positive quasiprobability representation) if a fraction of the photons is lost, as long as this is accompanied by a constant dark-count probability per detector.

In this work we propose a procedure to classically simulate certain linear-optical processes in a regime where losses are large. We do this by considering different physically-motivated models of loss. First we assume that exactly \( n/2 \) out of \( n \) photons have been lost, and show that efficient classical simulation is possible whenever \( l \) scales slower than \( \sqrt{n} \). We then consider a more usual loss model, when each photon has some (mode-independent) probability of being lost in the circuit, and so the output state does not have a well-defined number of particles. We show that, also in that case, if the average number of remaining photons is less than \( \sqrt{n} \) an efficient classical simulation is possible. Finally, we consider the most general case where the photons can traverse some linear-optical network of beamsplitters and where the loss probability can depend on the path followed by each photon. In this case we prove that, if the smallest number of lossy elements any input photon has to traverse in its path through the network is \( s \), we can effectively model that network as \( s \) rounds of uniform losses followed by a linear-optical channel with an efficient classical description. This leads to an efficient classical algorithm for lossy linear-optical networks of depth greater than \( C \log n \), for some suitable constant \( C \).

Our results also have implications outside boson sampling. First, in order to obtain our main result in theorem 1 we need to find the best approximation, in trace distance, to lossy bosonic states via particle-separable states (i.e. convex combinations of symmetric product pure states). To this end we use techniques that can be useful in the study of entanglement properties of symmetric states [35–37]. Our results might also be relevant for the design and characterization of complex linear-optical networks [23, 38, 39]. In theorem 2 we prove that a network where the smallest number of lossy elements in any path between inputs and outputs is \( s \) (but that has an otherwise arbitrary geometry) is equivalent to another network with \( s \) rounds of mode-independent losses at the input. This proves the intuition expressed previously regarding the exponential losses in current implementations of linear-optical networks, but also formally justifies the approximation of mode-independent losses which is often assumed in the literature [14, 15, 31, 40].

Our paper is organized as follows. In section 2 we give some theoretical background that underlies our work. In particular we review how linear-optical processes can be described both in first and second quantization formalisms, how different models of loss can be described mathematically, and some foundations on the boson sampling problem. In section 3 we describe the general idea of our simulation algorithm. In section 4 we describe our main results and discuss their consequences, although we defer the technical proofs to section 6. In section 5...
we present the relation of our results to other works, with particular emphasis on how they fit in the complexity-theoretic formalism of boson sampling. In section 6 we detail the proofs of the results of section 4. Finally, in section 7 we conclude with some general remarks and open questions.

**Notation:** Throughout the paper we use the following notation. Given two positive-valued functions \( f \) and \( g \), we write \( f = o(g) \) if \( \lim_{x \to \infty} f(x)/g(x) = 0 \) and \( f = O(g) \) if \( \lim_{x \to \infty} f(x)/g(x) < \infty \). Likewise, we say \( g = \omega(f) \) if \( f = o(g) \). Finally, \( f \approx g \) will refer to the situation when \( \lim_{x \to \infty} f(x)/g(x) = 1 \).

2. Theoretical background

2.1. Description of bosonic states

In this work we consider systems of bosons that can occupy \( m \) modes. One natural description of these systems is in the language of second quantization, commonly used in the context of quantum optics [41, 42] to describe the boson sampling problem. Here we focus on the complementary description based on first quantization, more natural for the description of the so-called particle entanglement, a concept that will prove useful in what follows.

In the language of first quantization, we begin with the Hilbert space for a single particle which can occupy \( m \) modes, \( \mathbb{C}^m \), with a basis denoted by \( \{ |i\rangle \}^m_{i=1} \). Then, an \( n \)-boson system is described by the Hilbert space \( \mathcal{S}_{n,m} := \text{Sym}(\mathbb{C}^m) \), i.e. the symmetric subspace of \( \mathcal{H}_{n,m} := (\mathbb{C}^m)^\otimes n \). This reflects the fact that bosonic wave-functions are symmetric upon exchange of particles. A basis for \( \mathcal{S}_{n,m} \) is the so-called Dicke basis, which can be defined in terms of the single-particle states. Elements of this basis are labeled by \( m \)-element tuples \( n = (n_1, \ldots, n_m) \) of non-negative integers \( n_i \) satisfying \( \sum_{i=1}^m n_i = n \). The Dicke state \( |n\rangle \) can be defined by \( |n\rangle = N(n) |\psi_{\text{sym}}(n)\rangle \), where \( |\psi_{\text{sym}}(n)\rangle \) is a projector onto \( \mathcal{S}_{n,m} \) (acting on \( \mathcal{H}_{n,m} \)) and \( |i\rangle \equiv |i_1, \ldots, i_n\rangle \). The relation between \( n \) and \( i \) is as follows: \( n_k \) equals the number of times \( |k\rangle \) appears in \( |i\rangle \), and so the basis \( |n\rangle \) is also called occupation-number basis. Finally, \( N(n) \) is a normalization factor (see equation (A10)). We also define the shorthand for the standard state

\[
|\Psi_0\rangle := |1, \ldots, 1, 0, \ldots, 0\rangle,
\]

which is often assumed to be at the input of a boson sampling device (see section 2.4). Additionally, by \( \Psi_0 \) we denote the projector onto \( |\Psi_0\rangle \).

In our analysis, a predominant role will be played by (bosonic) particle-separable or symmetric separable states. For a fixed number of particles \( l \), they are defined [43, 44] as states \( \sigma \) that can be written as convex combinations of pure product bosonic states, i.e.

\[
\sigma = \sum_{n} p_n |\phi_n\rangle \langle \phi_n|^\otimes n,
\]

where \( \{ p_n \} \) is some probability distribution and \( |\phi_n\rangle \langle \phi_n| \) are states on a single-particle space \( \mathbb{C}^m \). In what follows we denote the set of symmetric separable \( n \)-particle states by \( \text{Sep}(\mathcal{S}_{n,m}) \). Bosonic \( n \)-particle states that cannot be decomposed in this form are called particle-entangled. When we restrict the standard notion of entanglement of distinguishable particles from \( \mathcal{H}_{n,m} \) to \( \mathcal{S}_{n,m} \), it coincides with the above definition of particle entanglement. It is also a resource for quantum sensing [40, 45] and necessary for violations of Bell inequalities in the presence of superselection rules [46].

If we remove the constraint on the total number of particles, the corresponding Hilbert space (known as the bosonic Fock space) is the direct sum of the corresponding Hilbert spaces \( \mathcal{S}_{n,m} := \bigoplus_{n=0}^\infty \mathcal{S}_{n,m} \), where \( \mathcal{S}_{0,m} \) is the one-dimensional space spanned by the Fock vacuum \( |0\rangle \). The occupation number states \( |n\rangle \) (also called Fock states in this context) constitute a basis of \( \mathcal{S}_{n,m} \). The notion of particle separability extends naturally to \( \mathcal{S}_{n,m} \): a state \( \sigma \) supported on \( \mathcal{S}_{n,m} \) is particle-separable (i.e. \( \sigma \in \text{Sep}(\mathcal{S}_{n,m}) \)) if it can be expressed as

\[
\sigma = \sum_{n=0}^\infty p_n \sigma^{(n)},
\]

where \( \{ p_n \} \) is a probability distribution and \( \sigma^{(n)} \in \text{Sep}(\mathcal{S}_{n,m}) \). From our perspective, particle-separable states will be important as they yield classically-simulable instances of boson sampling (see section 2.4).

Second quantization constitutes a complementary language for the description of bosonic systems. In this formalism, a central role is played by creation \( (a_i^\dagger) \) and annihilation \( (a_i) \) operators that act on \( \mathcal{S}_{n,m} \) and satisfy canonical commutation relations \( [a_i, a_j^\dagger] = \delta_{ij} \). These operators define the number operator on mode \( |i\rangle \), \( \hat{n}_i := a_i^\dagger a_i \), which satisfies the relation \( \hat{n}_i |n\rangle = n_i |n\rangle \). Fock states can then be expressed as

\[
|n\rangle = |n_1, \ldots, n_m\rangle = \prod_{i=1}^m \frac{(a_i^\dagger)^{n_i}}{\sqrt{n_i!}} |0\rangle.
\]
Every operator acting on $S_m$ can be expressed as a polynomial in creation and annihilation operators. This, together with the commutation relations, leads to the mode-dependent tensor product decomposition

$$S_m \approx \bigotimes_{i=1}^{m} S_1^{(i)},$$

where $S_1^{(i)}$ is the Fock space associated to the $i$th mode. This decomposition leads to the notion of mode separability (and, by negation, mode entanglement), which differs from that of particle separability. In particular, Fock states admit the decomposition

$$n_1 \ldots n_m \mapsto |n_1 \ldots n_m\rangle,$$

while being particle-entangled whenever more than two modes are occupied. Conversely, particle-separable states $|\phi\rangle^\otimes$ are generically mode-entangled.

2.2. Linear optics in first and second quantization

A lossless linear-optical transformation is encoded by an $m \times m$ unitary matrix $U$. In the first quantization description, this matrix induces the independent evolution of particles as

$$\rho \mapsto U^\otimes \rho (U^\otimes)^\dagger,$$ for $\rho$ a state on $S_{n,m}$. (5)

In second quantization, the action of $U$ can be conveniently described as a transformation on creation operators

$$a_{i,\text{in}}^\dagger \mapsto a_{i,\text{out}}^\dagger = \sum_{j=1}^{m} U_{ij} a_{j,\text{in}}^\dagger,$$ (6)

where $a_{i,\text{in}}^\dagger$ and $a_{i,\text{out}}^\dagger$ are the operators for the $i$th mode at the input and output, respectively, of the linear-optical transformation. Transformations described by the above equations give rise to the quantum channel $\Lambda_U$ acting on states defined on $S_m$.

Consider now the paradigmatic boson sampling instance, where an $n$-photon input state $\Psi_n$ (see equation (1)) evolves according to an $m$-mode linear-optical transformation $U$ and is measured at the end in the photon-number basis, as represented in figure 1. One way to obtain a classical simulation of this system is to map it to a regular quantum circuit, and then leverage known results for classical simulation of quantum circuits. The representations defined in equations (5) and (6) lead to two methods to perform this mapping.

In the first-quantized description arising from equation (5), each of the $n$ photons gets mapped to a single $m$-level system, which now labels in which mode that photon is. We represent the corresponding quantum circuit in figure 2(a). In order to simulate a bosonic state, the system must be initialized in a fully symmetric state. The linear-optical transformation $U$ is then implemented as the local unitary transformation $U^\otimes$. This description treats all linear-optical transformations in the same footing, and does not bring much insight e.g. in to geometrical properties of a circuit. On the other hand, several properties of the input state are transparent. For example, if one photon is perfectly distinguishable from the others (maybe by some internal degree of freedom such as polarization), the simulation of figure 2(a) just leaves that particle out of the symmetrization step. Thus, particle indistinguishability in the physical system translates in to entanglement in the first-quantized simulation. Another effect with a natural manifestation in the first-quantized simulation is photon loss, as we discuss in more detail in the next section.

In the second-quantized description that arises from equation (6), we have a decomposition according to equation (4) where each of the $m$ modes is mapped to an $n$-level system labeling its occupation number. In this case, each 2-mode transformation in the original linear-optical circuit is mapped to a gate acting only on the two corresponding subsystems in the simulation, as in figure 2(b). This representation has the benefit of preserving geometrical properties of the circuit, and we might try to adapt results that limit the computational power of quantum circuits based e.g. on their depth [47–49].

Figure 1. A linear-optical circuit with ‘boson sampling resources’, i.e. Fock input states and number-resolving detectors.
2.3. Losses in linear optics

Let us now discuss how losses affect linear-optical systems. We survey three models of losses: the fixed-loss model and the beamsplitter loss model, the latter which can be divided in the uniform case (i.e. where all modes are subject to the same loss) and the non-uniform case describing more general (passive) linear-optical networks.

Fixed-loss model—In the fixed-loss model, we have initially $n$ particles and assume that exactly $n - l$ of them are lost. If we further assume that losses are mode-independent, i.e. any photon is equally likely to have been lost, this model has a very simple interpretation in the first-quantization picture. The loss corresponds to tracing out $n - l$ of the $n$ particles, as shown in figure 3, since $\rho$ is symmetric it does not matter which $n - l$ particles we trace over. Moreover, a partial trace does not change when a unitary transformation is applied to a subsystem which is disregarded and therefore

$$
\text{tr}_{n-l}(U^{\otimes n}\rho(U^{\dagger})^{\otimes n}) = \text{tr}_{n-l}[(U^{\otimes n} \otimes 1^{n-l})\rho(U^{\dagger})^{\otimes n}(U^{\otimes n}\otimes 1^{n-l})].
$$

Finally, by using the fact that partial trace commutes with channels applied to a subsystem which is not traced over, we get

$$
\text{tr}_{n-l}(U^{\otimes n}\rho(U^{\dagger})^{\otimes n}) = U^{\otimes n}\text{tr}_{n-l}(\rho)U^{\dagger}(U^{\dagger})^{\otimes n}.
$$

This is important as it implies that it does not matter whether losses occurred before or after the linear-optical transformation.

The fixed-loss model was used in [14] and in the context of quantum metrology in [50]. Physically, this model is suitable for cases where number of lost particles can be effectively controlled and accounted for, e.g. in atomic interferometry [51–54]. It is also more convenient for defining lossy boson sampling as a computational

---

3 In fact, the authors of [14] only investigate boson sampling when losses occur at the input to the circuit, and leave as an open question whether their results also apply when losses happen at the output. The observation below equation (8) answers that question affirmatively.
problem, since we can restrict our attention to the number of losses observed in a single run of the device with fewer assumptions about the physical process that caused them [14]. In the context of optics, however, a more faithful physical model of losses is given by the beamsplitter model [55], which we describe now.

Uniform beamsplitter loss model—In the beamsplitter loss model, whenever a photon reaches a lossy element, it is lost with some probability \(1 - \eta\). For example, a state \(|1\rangle\) with a single photon, when passing through a lossy element, leaves in state \((1 - \eta)|0\rangle|0\rangle + \eta|1\rangle|1\rangle\). We call this the beamsplitter model as it is equivalent to replacing the loss by a beamsplitter with transmissivity \(\sqrt{\eta}\), as in figure 4. This is the most common way of modeling losses in the quantum optics literature [40, 55]. If losses are mode-independent they can be conveniently described in the first quantization picture: \(n\)-particle states are transformed by the channel \(\Lambda_\eta\),

\[
\rho \rightarrow \Lambda_\eta(\rho) \equiv \sum_{l=0}^{\infty} \binom{n}{l} \eta^l (1 - \eta)^{n-l} \text{tr}_{n-l}(\rho),
\]

where \(\eta^l (1 - \eta)^{n-l} \binom{n}{l}\) is the probability that exactly \(l\) particles remain. Comparing equations (7) and (9) we see that the fixed-loss and uniform beamsplitter loss models are closely connected, which we use often in this work. In particular, in section 3.1 we show that in certain regimes of parameters \(l\) and \(\eta\) they are also computationally equivalent from the perspective of complexity theory.

General lossy optical networks—To analyze the case when losses affect each mode differently we turn our attention to general passive linear-optical networks. This provides a convenient mathematical framework to formulate this question, but also corresponds to what is commonly done in experiments. Since arbitrary multimode transformations are not readily available, the simplest way to implement a complex \(m\)-mode linear-optical operation is to decompose it as a network (or circuit) of elements acting on a few modes at a time, such as in [38, 39]. If each of the smaller elements has some loss associated with it, this might lead to a distribution of losses that affects a photon differently depending on which path it takes within the network, and which is determined by the overall geometry of the circuit.

Throughout most of this work we assume that a linear-optical network is composed entirely of 2-mode transformations, i.e. beamsplitters and phase shifters. We further assume, for simplicity, that losses are caused only by beamsplitters, disregarding those associated with phase shifters, detectors, sources and transmission paths between beamsplitters, and furthermore that each beamsplitter induces the same loss probability \((1 - \eta)\) in each of its input arms. Thus, the building block of a network is the element depicted in figure 5, where red elements represent pure losses (as in figure 4) and blue collectively represents beamsplitters and any neighboring phase shifters (although we refer to them only as beamsplitters from hereon). From equations (8) and (9) it follows that a round of uniform losses on \(m\) modes commutes with an \(m\)-mode passive linear-optical transformation, and in particular this holds for \(m = 2\) as shown in figure 5. For a network of arbitrary geometry, however, the analysis is more complicated since it is not immediately obvious to what extent losses can be commuted within the network. We return to this question in section 4.

\[\text{Figure 4. The beamsplitter model of losses. On the left we have the photonic mode passing through a lossy element with survival probability } \eta. \text{ On the right we have an equivalent circuit where the lossy element is replaced by a beamsplitter with transmission probability } \sqrt{\eta}, \text{ in which case each photon in the input is routed in to the unmeasured mode with probability } 1 - \eta.\]

\[\text{Figure 5. The building block of a linear-optical network used throughout the paper. Red elements are loss elements defined as in figure 4, while blue are lossless two-mode linear-optical transformations, including beamsplitters and phase shifters. It is not hard to show that a round of uniform losses on } m \text{ commutes with any linear optics on the same } m \text{ modes, and here we depict this for } m = 2.\]

\[\text{\textsuperscript{4} For the formal proof of equation (9) see the appendix of [56].}\]
We will describe the action of a general passive linear-optical network $\mathcal{N}$ by a channel $\Lambda_{\mathcal{N}}$, which can be obtained by composition of the corresponding channels for loss elements and beamsplitters. The explicit description of $\Lambda_{\mathcal{N}}$ in terms of equations (9) and (5) is possible in principle, but can be very cumbersome. For a more practical description, we direct the interested reader to a formalism that describes it in terms of the action of linear-optical transformations on coherent states [56].

### 2.4. Boson sampling

Here we consider the simulation of general lossy linear optics, but the main setup we consider is the boson sampling model of quantum computing [3]. The ideal boson sampling instance is defined by three ingredients: (i) preparation of an $n$-photon, $m$-mode input state, (ii) application of an $m$-mode unitary linear-optical transformation $U$, and (iii) measurement of occupation number in every output mode. This is exactly embodied in the physical setup described in figure 1. From the action of $U$ on bosonic operators (see equation (6)), it is easy to show that the transition probability from some input state $|n\rangle$ to some output state $|p\rangle$ is given by

$$\Pr(n \rightarrow p) = \frac{|\text{Per}(U_{n,p})|^2}{n! \cdots m! p_1! \cdots p_m!},$$

(10)

where $U_{n,p}$ is an $n \times n$ submatrix of $U$ constructed by repeating $n_i$ times the $i$th column of $U$, and $p_j$ times its $j$th row [3, 57], and $\text{Per}(A)$ denotes the permanent of matrix $A$ [38]. Throughout this paper we assume that the input is the state $|\Psi_0\rangle$ from equation (1), i.e. that the photons initially occupy the first $n$ modes, without loss of generality.

In [3] the authors showed that, modulo certain complexity-theoretic conjectures, it would be impossible for a classical computer to efficiently simulate the output of this idealized process, even in an approximate sense. We defer a more technical discussion of the results of [3] to section 5, when we discuss the complexity-theoretic consequences of our own results. For now it suffices to say that the existence of an efficient classical algorithm capable of simulating a distribution sufficiently close to the ideal boson sampling one would imply a collapse of the polynomial hierarchy to the third level, which is considered a very unlikely outcome in the complexity theory literature.

The results of [3] concern the possibility of classical simulation of boson sampling, but there are several notions of simulation one can use. A strong simulation of a quantum process is a classical algorithm that outputs the probability of any outcome, which may also include marginal probabilities. A weak simulation is a classical algorithm that produces a sample from the same distribution as the quantum device. It is generally regarded that, from the point of view of assessing whether some quantum device outperforms classical computers, the notion of weak simulation is more adequate, as it requires the classical algorithm to perform the same task as the quantum device.

Within the notion of weak simulation we can consider an exact simulator (which produces samples from the exact ideal distribution) or an approximate simulator. Given the ideal distribution $p_x$, an approximate simulator samples from some distribution $q_x$ such that

$$\frac{1}{2} \sum_x |p_x - q_x| \leq \epsilon$$

(11)

for some $\epsilon > 0$. In the original boson sampling paper [3], the authors require a classical algorithm to simulate the quantum device only in the approximate weak sense described above. Furthermore, their result also requires that the runtime of the classical algorithm is poly ($n, 1/\epsilon$). In other words, the classical algorithm should be able to give better approximations at the cost of only polynomial more time.

In our main results of section 4, by contrast, the classical algorithm is only able to perform weak simulation up to some threshold precision $\epsilon (n)$. We cannot improve the simulation simply by spending more computational time, but under some conditions the precision of our algorithm improves with the size of the problem. This means it is a different notion of approximation from the one used e.g. in [3], although other recent papers use similar definitions to ours [59]. We discuss these distinctions and complexity-theoretic implications of our results in section 5. For a more thorough discussion on this subject, see [60, 61].

One last important thing to point out is that there are cases where boson sampling is known to be classically simulable, such as when inputs and measurements are in the Gaussian basis [62], or when there is a combined high rate of losses in the system and dark counts in the detectors [15]. One such situation which will be important later on is when photons are perfectly distinguishable (e.g. due to some undetected degree of freedom such as polarization). In this case the transition probabilities are given by

$$\Pr(n \rightarrow p) = \frac{\text{Per}(U_{n,p})^2}{p_1! \cdots p_m!},$$

(12)
where $|U_{n,p}|^2$ is obtained from $U_{n,p}$ by taking the absolute value squared of each of its elements. This equation is surprisingly similar to equation (10), except that now the matrix has only positive elements. This, however, is crucial in enabling an efficient classical simulation. One way to see this is to note that, although the permanent is a hard function in general, for matrices with only positive elements it can be approximated efficiently by a classical algorithm due to Jerrum et al [63]. Alternatively, if one is satisfied with a weak simulation, one can do it simply by sampling the outcomes of the photons one at a time, as they do not interfere. The first-quantization counterpart of this statement is that distinguishable particles correspond, in the simulation of figure 2(a), to particles that do not undergo the symmetrization procedure. So, if all photons are distinguishable, the system simply evolves as a number of parallel $m$-level systems.

Although we do not actually consider partial photon distinguishability, note that the bosons in a particle-separable state such as in equation (2) can be considered as effectively distinguishable. To see this, first let $T$ be linear-optical transformation that takes a single photon from state $|1\rangle$ in to state $|\phi\rangle$. Then write
\[
(|\phi\rangle\langle\phi|)^{\otimes n} = T^{\otimes n}(|1\rangle\langle1|^{T})^{\otimes n}.
\] (13)
That is, any symmetric product state is equivalent to initializing all photons in the same mode and applying some linear-optical transformation. But now note that if all photons start in the same mode, the submatrix constructed in equation (10) has only $n$ repeated columns, and in this case the permanent collapses simply to a combinatorial factor multiplied by the product of all elements in that column. This, in turn, becomes identical to same probability defined for distinguishable particles in equation (12). Thus, we conclude that any bosonic particle-separable state effectively behaves as classical mixture of states of distinguishable particles, which is crucial for our simulation scheme in what follows. In the first-quantization picture, this statement corresponds to the fact that the only $n$-particle state that is symmetric and where all particles are in the same state is a product state.

3. General outline of approximate classical simulation

Our classical simulation results all follow from the same general reasoning, which we present in this section. Let us first introduce two definition of distance to measure the quality of our simulations. The first is the total variation distance (TVD), which measures the distance between two probability distributions $P = \{p_i\}$ and $Q = \{q_i\}$ and can be written as
\[
\|P - Q\| = \frac{1}{2} \sum_i |p_i - q_i|.
\] (14)
The TVD is the typical figure of merit in the context of boson sampling, as in equation (11). The second measure is the trace distance between two quantum states $\rho$ and $\sigma$
\[
d_T(\sigma, \rho) = \frac{1}{2} \text{tr} [\sqrt{(\rho - \sigma)(\rho - \sigma)^\dagger}].
\] (15)
The trace distance can be also defined [64] as the maximum, over all POVMs $\{M_\alpha\}$, of the TVD between the probability distributions $D_\rho$ and $D_\sigma$, that arise when one measures $M_\alpha$ on states $\rho$ and $\sigma$, respectively. Therefore, all such probability distributions $D_\rho$ and $D_\sigma$ satisfy
\[
\| D_\rho - D_\sigma \| \leq d_T(\sigma, \rho).
\] (16)
We are now ready to outline our classical simulation scheme. In all cases, we are interested in simulating a linear-optical protocol (as in figure 1) that uses as input some state $\rho$, corresponding to the action of some lossy channel on the initial state $|\psi_0\rangle$. The simulation is achieved by constructing an alternative state, $\sigma$, with the following properties (see figure 6)

(i) $\sigma$ is as a convex combination of particle-separable states (possibly with different numbers of particles):
\[
\sigma = \sum_{l=0}^\infty p_l \sigma_l,
\] (17)
with $\sigma_l \in \text{Sep} (S_{1,n})$, and so $\sigma \in \text{Sep} (S_{mn})$. Every $\sigma_l$ can be further decomposed as $\sigma_l = \sum_{\alpha} q_{l,\alpha}^1 |\phi_{l,\alpha}\rangle\langle\phi_{l,\alpha}|^{\otimes l}$. We also require that the joint probability distribution $p_l q_{l,\alpha}^1$ can be efficiently sampled on a classical computer.

(ii) The trace distance $d_T(\sigma, \rho)$ between $\sigma$ and $\rho$ is small, in a sense to be made precise later.

Let $D(\rho; \Lambda)$ and $D(\sigma; \Lambda)$ be the probability distributions obtained by applying some linear-optical channel $\Lambda$ to $\rho$ and $\sigma$, respectively, and measuring the resulting state in the particle-number basis. By equation (16) we have $\| D(\rho; \Lambda) - D(\sigma; \Lambda) \| \leq d_T(\sigma, \rho)$. In other words, simulating any linear-optical channel $\Lambda$ acting on $\sigma$
will give a simulation of the corresponding channel acting on $\rho$, up to some error determined by $d_{\text{TVD}}(\sigma, \rho)$. In the following sections $\Lambda$ might be $\Lambda_U$, $\Lambda_{\alpha}$, or $\Lambda_X$, as appropriate from context (see definitions in section 2.1).

Our claim now is that sampling from the distribution $\mathcal{D}(\sigma; \Lambda)$ can be done efficiently on a classical computer. We restrict our attention to weak simulation, i.e. the task of classically producing a sample from the correct distribution. This can be done in three steps. First, we draw a pair $(\alpha, l)$ according to the probability distribution $(p_{fi} q_{fi}^l)$, which can be done efficiently by assumption. Second, it is easy to simulate the effect of the linear-optical channel $\Lambda$ on $|\phi_{\Lambda,\alpha}f\rangle|\phi_{\Lambda,\alpha}l\rangle|^\otimes l$. Indeed, in the first-quantization formalism (see equation (5)) any linear-optical channel that we care about will act as $U^\otimes l$, i.e. as a tensor product of single-particle unitaries.\footnote{At worst, the linear-optical channel will act as some channel $\Lambda$, but that can easily be described as a linear-optical transformation on at twice the number of modes.}

Finally, since the particle-number measurements can also be implemented by a product of single-particle measurements, the last step corresponds simply to simulating $O(n)$ independent parallel $m$-level quantum systems, which can be trivially be done on a classical computer.

Combining these observations, it is clear this procedure efficiently samples from a distribution closer than $d_{\text{TVD}}(\sigma, \rho)$ in TVD to the desired one $\mathcal{D}(\rho; \Lambda)$. In section 4 we obtain explicit expressions for the separable states and their trace distances to the desired state under different physically-motivated loss models, and in section 5 we return to the complexity-theoretic consequences of our simulations and further interpretations.

### 3.1. Computational equivalence of loss models

Our main results include claims about the loss models described in section 2.3. There we described how the models in which losses are mode-independent are mathematically related, namely via the binomial distribution $\eta^{n(1-\eta)}$. From that expression it is clear that, in the beamsplitter model with uniform transmission probability $\eta$ per mode, the average number of lost photons is $(1-\eta)n$. Thus, if we set $\eta = 1 - k/n$, we obtain any instance of the fixed-loss model with exactly $k$ losses as the expected outcome in the beamsplitter model. However, that does not suffice to prove that the models are computationally equivalent. To achieve that, we need to show under which conditions one model can simulate the other efficiently, i.e. with only polynomial (in $n$) overhead. This will allow us to re-purpose results obtained from one model in to the other.

Consider two black boxes that simulate lossy $\eta$-photon boson sampling experiments. The first, BS, outputs samples according to the beamsplitter model for some uniform loss probability $(1-\eta)$ per photon, per mode. The second, FL, outputs samples according to the fixed-loss model with $k$ losses, for some range of values of $k$. The formal question is whether we can use poly $(n)$ samples produced by one black box to output a single sample from the distribution generated by the other (with high probability), and what are the corresponding asymptotic behaviors of $k$ and $\eta$.

To prove that BS can simulate FL note that, for large $n$, we can use Stirling’s approximation to write

$$
\Pr(m\eta) = \binom{n}{m\eta} \eta^{m\eta}(1-\eta)^{n(1-\eta)} \\
\approx \frac{1}{\sqrt{2\pi n(1-\eta)}} \left(\frac{n}{m\eta}\right)^n.
$$

That is, since $\eta \in [0, 1]$, the probability of observing exactly the expected number of losses is $O(1/\sqrt{n})$. Thus, if we set $\eta = 1 - k/n$, we only need $O(\sqrt{n})$ samples from BS in order to obtain one sample from FL with high probability. This was used in [14] to extend their hardness claim from the fixed-loss model (i.e. that it is computationally as hard as ideal boson sampling if $k = O(1)$) carried over to the beamsplitter model in the regime $1 - \eta = O(1/n)$.

Figure 6. A schematic view of our simulation strategy. The set Sep is the subset of separable states within some state space of interest, and $\sigma$ is the closest state in Sep to the lossy state $\rho$.\[\text{Figure 6. A schematic view of our simulation strategy. The set Sep is the subset of separable states within some state space of interest, and $\sigma$ is the closest state}$

[\text{in Sep to the lossy state $\rho$.}]]
Consider now the reverse question, i.e. when can FL simulate BS? We already saw that, by setting $k_n h = -()$, FL samples from a distribution with the same number of photons as the average of the BS distribution, but that is not enough. We also need to show that the distribution over photon numbers produced by BS, i.e. the distribution of equation (9), is sufficiently concentrated around its mean. If that holds, then a FL black box that samples from events with $n O h ()$ transmitted photons can approximately simulate a BS black box by first sampling $k$ according to equation (9) truncated to a small range of values around the mean, and then outputting a distribution for that value of $k$.

The number of transmitted photons can be described by a random variable $X = \sum_{i=1}^{n} X_i$, where each $X_i$ is an independent random variable which takes value 0 if the $i$th photon was lost and 1 otherwise. Clearly, $X$ has the expectation $E(X) = \eta n$. Then, by a standard multiplicative Chernoff bound \[65\] we can write:

$$\Pr(X < (1 - \delta) \eta n) \leq e^{-\frac{\delta^2 \eta n}{2}},$$

$$\Pr(X > (1 + \delta) \eta n) \leq e^{-\frac{\delta \eta n}{3}}.$$ (19)

We can repeat this argument for $Y = \sum_{i=1}^{n} (1 - X_i)$, which is the number of lost photons and has expectation value $E(Y) = n(1 - \eta)$. The Chernoff bounds for $X$ and $Y$ show that, in the extremal regimes where the photons are either very likely to survive or to be lost (specifically if either $n(1 - \eta)$ or $n\eta$ go to 0 sufficiently fast with $n$), then the number of photons at the end of the circuit concentrates around its average as $n$ increases. This happens, for instance, if the average number of photons lost is constant (as in \[14\]), or if there are only an average of $o(\sqrt{n})$ photons left after the losses (as is the regime of our main results in following sections).

### 4. Simulation of lossy linear optics

In this section we describe our main results, namely under which loss regimes linear optics can be simulated classically by approximating the lossy states by with particle-separable states. We defer most technical proofs to section 6, rather focusing here on the interpretation and consequences of the main theorems.

We begin in section 4.1 by considering lossy linear optics in the fixed-loss model described in section 2.3. Although this model is less realistic from a physical standpoint, it is mathematically elegant and leads to more straightforward results.

In subsequent sections we focus on lossy linear-optical networks, defined in section 2.3. Two common examples of these networks are depicted in figure 7 \[38, 39\]. These networks model good approximations for experiments based on integrated photonics such as \[17–20, 23\].

In section 4.2 we initially address losses for ‘uniform’ (or ‘balanced’) networks, namely those where every mode is affected by losses in the exact same amount. This is a simpler case where loss elements commute with
unitary transformation (see section 2.3), and so all losses can be assumed to happen at the input. The results in this section rely heavily on the Chernoff bounds from section 3.1 and the results from the fixed-loss model.

Finally, in section 4.3 we justify the use of the uniform loss model for networks with arbitrary geometries, such as figure 7(b). To this end we prove that it is possible to `extract' a certain amount of uniform losses from an arbitrary network. The relevant parameter for how much loss can be pulled out of the network is the minimal number of beamsplitters that a photon must traverse between any input and any output.

4.1. Simulation in the fixed-loss model

Suppose the input of the bosonic state is \(|\Psi_0\rangle\) (i.e. the first \(n\) out of \(m\) modes are occupied by a single particle each). Suppose then that we trace \(n - l\) out of the \(n\) particles, corresponding to losing exactly \(n - l\) photons at the input (see section 2.3). We can write the resulting state as

\[
\rho_{l,n} = \frac{1}{\binom{n}{l}} \sum_{|n| = l} |n\rangle \langle n|
\]

(20)

where the sum is over all occupation-number vectors of length \(|n| = l\) satisfying \(n_i \leq 1\). We are now interested in finding the separable symmetric state \(\sigma_b \in \text{Sep}(S_{l,n})\) that is closest in trace distance to \(\rho_{l,n}\). In other words we seek for \(\sigma_b\) satisfying \(d_f(\sigma_b, \rho_{l,n}) = \Delta_{l,n}\), where

\[
\Delta_{l,n} = \min_{\sigma \in \text{Sep}(S_{l,n})} d_f(\sigma, \rho_{l,n})
\]

(21)

As discussed in section 3, if we find \(\sigma_b\) such that \(\Delta_{l,n}\) is suitably small, we can use \(\sigma_b\) to simulate any boson to the set of symmetric separable \(l\)-particle states \(\text{Sep}(S_{l,n})\) is given by

\[
\Delta_{l,n} = 1 - \frac{n!}{l!(n-l)!}
\]

(22)

Moreover, a separable state \(\sigma_b \in \text{Sep}(S_{l,n})\) that attains \(\Delta_{l,n}\) is

\[
\sigma_b = \frac{1}{(2\pi)^n} \int_0^{2\pi} d\phi_1 \ldots d\phi_n \left[ \prod_{i=1}^n |\phi_i\rangle \langle \phi_i| \right] V_{\phi_1,\ldots,\phi_n}\dagger
\]

(23)

where \(|\phi_i\rangle = (1/\sqrt{n}) \sum_{j=1}^n |i\rangle\) and \(V_{\phi_1,\ldots,\phi_n} = \exp(-i\sum_{i=1}^n \phi_i a_i^\dagger a_i)\).

Let us look at the asymptotic behavior of \(\Delta_{l,n}\) in the limit of large and small losses. Using the bounds \(ck^{1/2} \exp(-k) \leq k! \leq Ck^{1/2} \exp(-k)\), where \(c = \sqrt{2\pi}\) and \(C = e\), we can easily prove that \(\Delta_{l,n} \geq 1 - \frac{e}{l} \exp\left(-\frac{e^2}{2l}\right)\) and hence if \(l = \omega(\sqrt{n})\) we have \(\Delta_{l,n} \to 1\). On the other hand, using Stirling approximation for \(l = o(\sqrt{n})^6\) we can write \(\log\left(\frac{n!}{l!(n-l)!}\right) = -\frac{l^2}{2n} + o\left(\frac{l^2}{n}\right)\). This leads to the following result.

**Corollary 1.** If all but \(l = o(\sqrt{n})\) photons are lost we have

\[
\Delta_{l,n} \approx \frac{e}{2n}
\]

(24)

and therefore \(\Delta_{l,n}\) asymptotically tends to 0. Conversely, if \(l = \omega(\sqrt{n})\), then \(\Delta_{l,n}\) tends to 1.

All that remains is to show that it is indeed possible to use \(\sigma_b\) efficiently in a classical simulation. But this is quite straightforward. The state \(\sigma_b\) is obtained by starting with the state

\[
|\phi\rangle^\otimes l = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n e^{i\phi_i} |i\rangle\right)^\otimes l
\]

and averaging over all phases \(\phi_i\) uniformly in the \([0, 2\pi]\) interval. A weak simulation then consists of choosing a set of phases \(\phi_i\) uniformly at random, then simulating the action of \(U^\otimes l\) on the \(|\phi\rangle^\otimes l\), together with the final

---

6 The explicit computation reads as follows. For any integer \(k\) we have \(\log(k!) = k \log(k) - k + (1/2) \log(k) + \gamma + O(1/k)\). Therefore, for \(l = \omega(\sqrt{n})\) we have

\[
\log\left(\frac{n!}{l!(n-l)!}\right) = -(n-l)/2 + O(1/l) + (1/2)\log(n-l) + O(1/n).
\]

By using the expansion

\[
\log(1-x) = -x - (1/2)x^2 + O(x^3)
\]

and keeping the leading terms we finally obtain

\[
-\frac{l^2}{2n} + O\left(\frac{l^2}{n}\right).
\]
single-qudit computational basis measurements. This is simply the simulation of $l$ parallel $m$-level systems, and can be done efficiently on a classical computer. Combining this with the discussion in section 3 we obtain the following result.

**Corollary 2.** Let $\mathcal{D}(\rho_{h,n}; \Lambda_U)$ be the probability distribution obtained by applying the linear-optical transformation $U$ to state $\rho_{h,n}$ of equation (20) and measuring in the particle-number basis. There exists a probability distribution $\mathcal{D}(\sigma_{h}; U)$ that can be classically sampled from efficiently and such that $\| \mathcal{D}(\rho_{h,n}; \Lambda_U) - \mathcal{D}(\sigma_{h}; \Lambda_U) \| \leq \Delta_{l,n}$, where $\Delta_{l,n}$ is given in equation (22). Moreover, in the limit $l = o(n^{1/2})$ we have $\Delta_{l,n} \approx \frac{c}{n}$.

To our knowledge, the previous best known bound of this type [14] gave an efficient classical simulation when all but $l = O(\log n)$ of the photons were lost, since in that regime, the transition probabilities are associated with permanents of log-sized matrices, which can be computed classically in polynomial time. Although this is stated in [14] without proof, it follows straightforwardly from a recent boson sampling simulation algorithm [32].

As discussed previously, the model where a fixed number of particles is lost is of limited practical relevance, since in an experiment this number is typically a random variable. However, as we show next, insights gained from the analysis of this simplified model are useful to derive classical simulations for the more realistic ones.

### 4.2. Uniform losses

Consider now a lossy linear-optical network under the assumption of uniform losses, i.e. every particle in every input can independently be lost with probability $1 - \eta$. As discussed in section 2.3, the action of such a loss channel onto the initial $n$-particle Fock state $|\Psi_n\rangle$ leads to the state

$$\rho_{\eta,n} = \sum_{l=0}^{n} \eta^l (1 - \eta)^{n-l} \binom{n}{l} \rho_{l,n}$$

To obtain an approximate classical simulation of the resulting linear-optical process, we choose to approximate $\rho_{\eta,n}$ by states from $\text{Sep}(\mathcal{S}_n)$ i.e. convex combinations of symmetric separable states with different numbers of particles,

$$\tilde{\sigma} = \sum_{l=0}^{\infty} p_l \sigma^{(l)}_s,$$

where $\{p_l\}$ is a probability distribution and $\sigma^{(l)}_s$ is a symmetric separable state supported on $l$ particles and $n$ modes. As an ansatz for a good approximation to $\rho_{\eta,n}$ we take the mixture of optimal $l$-particle separable states with appropriate binomial weights,

$$\sigma_{\eta} = \sum_{l=0}^{n} \eta^l (1 - \eta)^{n-l} \binom{n}{l} \sigma^{(l)}_s,$$

where $\sigma^{(l)}_s$ are given in equation (23). The resulting trace distance between $\sigma_{\eta}$ and $\rho_{\eta,n}$ is

$$\Delta(\eta, n) = d_{\text{tr}}(\sigma_{\eta}, \rho_{\eta,n}) = \sum_{l=0}^{n} \eta^l (1 - \eta)^{n-l} \binom{n}{l} \Delta_{l,n},$$

where $\Delta_{l,n}$ is given in equation (22). The following auxiliary result gives bounds on $\Delta(\eta, n)$.

**Lemma 1.** Let $\Delta(\eta, n)$ be the trace distance between states $\sigma_{\eta}$ and $\rho_{\eta,n}$ defined in equations (25) and (26), respectively. We have the following inequalities

$$\Delta(\eta, n) \leq \frac{\eta^2 n^2}{2} + \frac{\eta(1 - \eta)}{2},$$

$$\Delta(\eta, n) \geq \Delta_{[\eta n - \delta \eta n]}(1 - \exp(-\delta^2 \eta^2/2)).$$

where $\delta \in [0,1]$ and $[x]$ denotes the smallest integer greater than or equal to $x$.

The proof of the above lemma is given in appendix A. Combining this result with the results of section 4.1 and the arguments of section 3 we conclude that it is possible to approximately simulate any boson sampling protocol that uses the lossy state $\rho_{\eta,n}$ as input, provided that $\langle l \rangle = \eta n = o(n^{1/2})$, where $\langle l \rangle$ is the average number of photons that are left in the network. On the other hand, if $\langle l \rangle = \omega(\sqrt{n})$, then $\Delta(\eta, n)$ goes asymptotically to 1.

7 Recall that in general the total number of modes $m$ is greater than the number of modes $n$ on which the state $\rho_{l,n}$ is supported.
Figure 8. A lossy linear-optical network that is not balanced, in the sense that different paths within the network suffer different amounts of losses. In these cases, it is not always obvious how many layers of losses can be ‘extracted’ from the circuit.

**Corollary 3.** Let $U$ be a linear-optical unitary matrix. Let $\mathcal{D}(\rho_{\eta,n}; \Lambda_U)$ be the probability distribution obtained by applying $\Lambda_U$ to $\rho_{\eta,n}$ (see equation (25)) and measuring the resulting state in the particle-number basis. There exists a probability distribution $\mathcal{D}(\sigma_{\eta}; \Lambda_U)$ that can be sampled efficiently on a classical computer such that

$$\| \mathcal{D}(\rho_{\eta,n}; \Lambda_U) - \mathcal{D}(\sigma_{\eta}; \Lambda_U) \| = \Delta(\eta, n),$$

where $\Delta(\eta, n)$ is given in equation (28) and satisfies $\Delta(\eta, n) \leq \frac{\eta^2 n}{2} + \frac{n(1 - \eta)}{2}$.

Our arguments give an approximate classical weak simulation to $\mathcal{D}(\rho_{\eta,n}; \Lambda_U)$ only when $\eta = o(n^{-1/2})$, which at first sight does not seem physically relevant. Usually, lossy devices (such as sources, beamsplitters or detectors) have loss probabilities that are constant, rather than decreasing with $n$. However, as we discuss in the next section, standard practical realizations of linear-optical networks often lead to dependence of the effective overall transmissivity of the network, $\eta_{\text{eff}}$, on the size of the system. Since the usual boson sampling paradigm requires networks that increase with the number of photons, this can easily ensure that $\eta_{\text{eff}}$ decreases at least as $\eta_{\text{eff}} = o(n^{-1/2})$.

### 4.3. Non-uniform losses

The effect of losses on linear-optical protocols has been extensively studied in the literature [14, 15, 31, 50, 55]. A common assumption in these works is that losses are uniform, as we considered in the previous section. This model can be justified, for example, by assuming that every path that a photon can take through a network is equally lossy. For some geometries of the network this can indeed be the case, such as that shown in figure 7(a) [39]. Moreover, this is a natural assumption for losses that model e.g. $m$ identical imperfect particle detectors [15, 55].

However, if the network has an unbalanced geometry, different paths through it accumulate different amounts of losses, as in the case of figure 7(b) [38]. The analysis of such networks is more complex, as losses affecting only certain modes in general do not commute with linear-optical elements connecting them to different modes (see figure 8). Indeed, the main point of [39] is to propose a balanced decomposition of a universal linear-optical network with the goal of making it more loss-resistant. The authors argue that, in a balanced network such as that of figure 7(a), losses ‘only’ act as overall losses, i.e. uniformly, whereas in the more commonly used decomposition of figure 7(b) losses also behave as non-unitary noise that is harder to mitigate and characterize.

To the best of our knowledge, until now there was no rigorous justification for the phenomenological assumption that losses are uniform, unless the network has a very special geometry such as in figure 7(a). The following result remedies this by showing that it is always possible to rewrite a lossy network of arbitrary geometry as a round of uniform losses followed by another (possibly still lossy) linear-optical network.

**Theorem 2.** Let $\mathcal{N}$ be an $m$-mode linear-optical network constructed out of the two-mode elements of figure 5, each inducing the same loss probability $(1 - \eta)$ in both of its input modes. Let $s$ be the smallest number of beamsplitters traversed by any path connecting an input mode to an output mode in the network. Let $\Lambda_{\mathcal{N}}$ be the quantum channel associated with the entire network $\mathcal{N}$. Then we have the following decomposition of $\Lambda_{\mathcal{N}}$:

$$\Lambda_{\mathcal{N}} = \Lambda_{\tilde{\mathcal{N}}} \circ \Lambda_{\text{loss}},$$

where $\Lambda_{\text{loss}}$ is a channel describing uniform losses with effective transmissivity $\eta_{\text{eff}} = \eta'$, and $\tilde{\mathcal{N}}$ is a linear-optical network obtained by removing all losses from $\mathcal{N}$. The description of $\tilde{\mathcal{N}}$ from $\mathcal{N}$ can be obtained efficiently.

A schematic representation of this theorem is shown in figure 9. We present the complete proof of theorem 2 (including a prescription to obtain $\tilde{\mathcal{N}}$ from $\mathcal{N}$) in section 6 and focus now on its consequences.

A version of theorem 2 is well-known if one can assume that the network $\mathcal{N}$ can be decomposed in to layers of commuting linear-optical elements, each layer covering all modes. Since uniform $m$-mode losses commute...
with arbitrary $m$-mode linear-optical networks (see section 2.3), whenever a network has a layer of uniform losses it can be commuted out of the network as in figure 10. Thus, if a network has $s$ layers of this type they can be extracted layer by layer, leading to a decomposition as in theorem 2 in a straightforward manner. Theorem 2 strengthens this statement by showing that it holds in more general geometries, as long as every path between an input and an output passes through $s$ loss elements.

Theorem 2 already formalizes the claim, stated in section 1, that losses in current implementations of boson sampling increase exponentially with the depth of the circuit (if one identifies the depth of the circuit with the
definition of $s$ in theorem 2), which is a serious scalability issue. Let us now consider the consequences of this result for classical simulability.

First note that, from the efficient description of $\tilde{N}$, we can efficiently sample from the probability distribution $D(|\phi \rangle \langle \phi |^T, \Lambda_N)$ (see section 2.4 and section 3). Furthermore, from equation (31), we have that $D(\Psi_l; \Lambda_N) = D(\Lambda_{\eta l}(\Psi_l; \Lambda_{\eta l}))$. From the strategy outlined in section 3, if $\Lambda_{\eta l}(\Psi_l)$ is close in trace distance to Sep(S), it is possible to classically sample from a distribution approximating the desired one $D(\Psi_l; \Lambda_N)$. On the other hand, from lemma 1 we know that there exist a state $\sigma_{\eta l} \in$ Sep(S) such that $d_\text{tr}(\Lambda_{\eta l}(\Psi_l), \sigma_{\eta l}) \leq \Delta(\eta_{\text{eff}}, n)$, where $\Delta(\eta_{\text{eff}}, n)$ is defined in equation (28). The simplest way to simulate the remaining network $\tilde{\Lambda}_\chi$ acting on $\sigma_{\eta l}$ is to replace each loss element by a corresponding beamsplitter that redirects that mode in to a new mode, as described in figure 4. Supposing there are at most $m^2$ beamsplitters in the network (which is sufficient to construct an arbitrary interferometer), this would require adding at most $m^2$ new modes. Thus, this non-unitary optical $m$-mode transformation can be described as a linear-optical transformation on $m + m^2$ modes and the simulation is still efficient. Using standard methods [42], it is also possible to decompose $\tilde{\Lambda}_\chi$ in to a pure $m$-mode linear-optical transformation followed by an array of non-uniform losses. Although we do not describe the details of this construction here it follows that, in general, the number of additional modes required is $m$ regardless of the network geometry. Putting this all together leads to the following result.

Corollary 4. Let $N$ be the $m$-mode linear-optical network of theorem 2 and $\Lambda_N$ the corresponding linear-optical channel. Let $s$ be the smallest number of beamsplitters traversed by any path between an input and an output mode. Let $D(\Psi_l; \Lambda_N)$ be the probability distribution obtained by applying $\Lambda_N$ to the standard boson sampling input state $\Psi_l$ and measuring the resulting state in the particle-number basis. There exists a probability distribution $D(\sigma_{\eta l}; \tilde{\Lambda}_N)$ that can be sampled efficiently on a classical computer and such that $\| D(\Psi_l; \Lambda_N) - D(\sigma_{\eta l}; \tilde{\Lambda}_N) \| = \Delta(\eta_{\text{eff}}, n)$, where $\Delta(\eta, n)$ is given in equation (28) and satisfies $\Delta(\eta, n) \leq \frac{\eta n}{2} + \frac{m^2 - n}{2}$, and $\eta_{\text{eff}} = \eta^*$. It is clear from corollary 4 that our simulation gives a good approximation whenever there are less than $\text{O}(\sqrt{n})$ photons left on average. In terms of $s$, the regime of $\alpha \sqrt{n}$ photons left corresponds to $s = -\log n / (2 \log \eta) + \log \alpha / \log \eta$. Thus, our results show that boson sampling becomes classically simulable for networks of depth greater $C \log n$, where $C$ is some suitably large constant that depends on $\eta$. We should note that a similar scaling could also be drawn from the results described at the end of section 4.1, since the regime where only $\alpha \log n$ photons are left on average corresponds to depth $s = -\log n / \log \eta + \log \log n / \log \eta + \log \alpha / \log \eta$.

It is curious to note that our corollary 4 generates a tension with the conclusions drawn by the authors of [39]. There, they argue that balanced networks are better, since unbalanced losses act as non-unitary noise. But here we conclude the opposite: if a network is too balanced, more losses can be uniformly extracted from the network, improving the quality of our simulation of corollary 4.

We conclude this section by pointing out a few ways that the above results can be strengthened. For the first note that, if photons are only inserted in particular input modes, we can replace the network by another effective network where unused inputs are ignored. For example, consider the standard decomposition of figure 7(b). One could think that theorem 2 would not be relevant there since, in principle, $s = 1$. However, if it is known that photons will only enter the first $n$ modes we can restrict definition of $s$ in theorem 2 to only take in to account paths starting from those inputs, effectively increasing $s$. This can also be relevant if one does not have control over the input, such as in scattershot boson sampling [24, 30].

Another way theorem 2 can be strengthened is that it can be applied even if losses are not all equal over all beamsplitters. As long as every beamsplitter has some nonzero loss probability in each input mode (which may even differ between the two inputs of the same beamsplitter), we can repeat the procedure of theorem 2 by pulling out $s$ layers of uniform loss, each with loss probability $1 - \eta^*$, where $\eta^*$ is the highest among all value of $\eta$ for all beamsplitters. Finally, we conjecture that corollary 4 holds even if a network is composed not only of two-mode elements, but linear-optical elements of various sizes, and we leave this as a question for future work.

5. Relation to other works and relevance for boson sampling

5.1. Complexity-theoretic considerations

All main simulation theorems contained in sections 4.1–4.3 have a common feature, in that they allow us to approximate boson sampling with the desired lossy state $\rho$, by replacing it by another state $\sigma$, up to some TVD $\epsilon$. The precision of the simulation, $\epsilon$, is not a controllable parameter, but rather a function of $n$ and possibly other physical parameters such as the transmission probability $\eta$. In suitable regimes (for instance, if $l = o(n^{1/2})$ in
corollary 2), $\epsilon$ decreases as the problem size $n$ increases. Let us investigate the complexity-theoretic consequences of these results.

In the original paper on boson sampling, [3], the authors define the following computational problem, which they call Gaussian permanent estimation:

**Problem 1** ([GPE]$_n^e$ [3]). Given as input a matrix $X$ of i.i.d. elements taken from the standard (complex) normal distribution, together with bounds $\epsilon', \delta > 0$, estimate $|\text{Per}(X)|^2$ to within additive error $\pm \epsilon' \cdot n!$ with probability at least $1 - \delta$ over choices of $X$ in poly $(n, 1/\epsilon', 1/\delta)$ time.

Their main result (theorem 1.3 of [3]) then can be stated roughly as follows. Let $U$ be an $m \times m$ Haar-random matrix, and denote by $\mathcal{D}$ the probability distribution over outcomes generated by a boson sampling computer, i.e. with probabilities given by equation (10). Suppose there is a classical algorithm $C$ that takes as input the unitary matrix $U$ and an error bound $\epsilon > 0$ and outputs a sample from some distribution $\mathcal{D}'$ such that $\|\mathcal{D} - \mathcal{D}'\| < \epsilon$, in time poly $(n, 1/\epsilon)$. Then it would be possible to solve $|\text{GPE}|_n^e$ in the complexity class $\text{BPP}^{\text{NP}}$ (for definitions and discussions on these complexity classes we direct the reader to [3] or to the Complexity Zoo [66]).

The authors argue that such a consequence would be very surprising, and so an efficient classical algorithm $C$ is unlikely, but they also take a step further. By positing two additional complexity conjectures, the permanent-Gaussian conjecture and the permanent anti-concentration conjecture, which together state that estimating $|\text{Per}(X)|^2$ for Gaussian $X$ to high precision and with high probability is a $\not\in \text{P}$-hard problem, they show that the existence of an efficient classical algorithm $C$ would imply a collapse of the polynomial hierarchy. The polynomial hierarchy is a hierarchy of complexity classes which, in the complexity theory community, is widely believed to be infinite, and so this results further strengthens the initial claim that boson sampling cannot be efficiently simulated on a classical computer.

By inspecting our results, we see that they do not follow the same definition of efficient classical simulation as used in [3]. There, the classical algorithm should be able to produce samples from a distribution that is $\epsilon$-close to the ideal one, for arbitrarily small $\epsilon$, just by spending poly $(1/\epsilon)$ more computational time. This definition of classical simulation has also recently been discussed in detail in [61], where it is called an $\epsilon$-simulation.

One might ask whether the authors of [3] could have used a less stringent definition of simulation, where $\epsilon$ is some fixed threshold value rather than a controllable parameter that can be made arbitrarily small. This is used, for example, in [59], where the authors apply a similar reasoning used for boson sampling to the model of commuting quantum circuits, concluding that it would be unlikely for a classical computer to be able to efficiently sample from a distribution that is closer than 1/192 in TVD to the ideal one. By adapting the argument of [3] one can see that, if there existed a classical algorithm that sampled from a distribution $\mathcal{D}'$ such that $\|\mathcal{D} - \mathcal{D}'\| < \epsilon$ for fixed $\epsilon$, it would be possible to solve, in $\text{BPP}^{\text{NP}}$, an easier version of $|\text{GPE}|_n^e$ where $\delta$ and $\epsilon'$ are not free parameters, but rather satisfy $\epsilon' \cdot \delta = \Omega(\epsilon)$. That is, there would be a trade-off between the probability that the $\text{BPP}^{\text{NP}}$ algorithm outputs an estimate for $|\text{Per}(X)|^2$ for Gaussian $X$ and the quality of that estimate. That would be still an unlikely complexity-theoretic outcome, albeit a less surprising one than an efficient classical algorithm $C$ described above.

This discussion does not explicitly influence our conclusions, however. Ultimately, our main results are not addressing any complexity conjecture, given that to our knowledge no one has formally conjectured that boson sampling is hard in the limit of large losses or that the probabilities generated by a lossy device are associated with any natural hard problem. At the same time, for any given problem size $n$ our simulations work only up to fixed precision $\epsilon(n)$ as described in the corresponding theorems. Thus our results do not rule out, technically, that lossy boson sampling has interesting computational capabilities, if one defines the corresponding computational task in terms of the ‘arbitrarily small $\epsilon$’ prescription advocated in [3, 61]. On the other hand our results do place stringent requirements on proposed experimental demonstrations of quantum computational supremacy based on lossy boson sampling. The fact that the precision of our simulations improves with the size of the problems means that, in order for a realistic boson sampling machine to outperform (asymptotically) our simulation, the experimental imperfections (most crucially the loss probability per-mode per-component) must decrease sufficiently fast as experiments move to larger problem size.

One final point we wish to make regards two possible choices of loss models we consider, the fixed-loss model and the uniform beamsplitter model. In the uniform beamsplitter model, there is always some nonzero probability that zero photons are lost. Thus it is always possible, in principle, to postselect on the outcome where all photons make it to the output, and conditioned on that fact we just have regular boson sampling. This postselection is not efficient, but nonetheless it allows us to use the postselection-based argument described in [3, 4, 49] to show that there is no efficient exact classical simulation of lossy boson sampling in the beamsplitter model for any degree of loss, unless the polynomial hierarchy collapses. Conversely, any proposal for such a classical algorithm must be only approximate, up to (possibly arbitrarily small) $\epsilon$. 
The above argument seems to contradict our claim, given just below corollary 2, that if only $O(\log n)$ photons remain efficient classical simulation of boson sampling is possible (which follows from the result of [32]). The apparent contradiction arises because the simulation that would follow from [32] is in fact exact. The resolution comes from the fact that this claim holds only for the fixed-loss model with $n = O(\log n)$ losses, and the mapping between the two models we described in section 3.1 is not exact. More specifically, if one wanted to use this classical simulation to simulate lossy boson sampling in the beamsplitter model, one would have to choose the regime where the average number of remaining photons is $O(\log n)$ and truncate the binomial distribution of equation (9) to some number of standard deviations around its average. The error incurred in this truncation is the $\epsilon$ error in the final simulation, and so this algorithm is not ruled out by the postselection argument.

5.2. Relation to other work

Consequences for current experiments—The results discussed up to now concern only an asymptotic regime. We can also analyze the relevance of our results for finite-size experiments aimed at giving concrete demonstrations of quantum computational supremacy, for example in the range of 50 [31, 32] or 90 [34] photons.

Let us initially consider some range of behaviors for $\eta_{\text{eff}}$, i.e., the overall transmissivity of the network, such that the expected number of photons arriving at detectors is $\langle l \rangle = m \eta_{\text{eff}}$. Suppose that $\eta_{\text{eff}} = C/n^s$ for some constant $C$ and some $x \in [1/2, 1]$. The case $x = 1$ is the threshold where $\langle l \rangle$ is constant regardless of how many photons are prepared, and so scaling up experiments in this limit is clearly pointless (for $x > 1$, we get less photons out the more photons we put in). If $x = 1/2$ we have $\langle l \rangle = C\sqrt{n}$, which is informally the threshold for our simulation to improve asymptotically in $n$, as discussed after our corollary 3. So $x \in [1/2, 1]$ is the range where our simulation is (asymptotically) good while the number of detected photons is still expected to grow with $n$, i.e. where the main obstacle to demonstrations of quantum supremacy would be our simulation rather than intrinsic non-scalability due to losses.

For concreteness, now, suppose $\eta_{\text{eff}} = n^{-3/5}$, and thus $\langle l \rangle = n^{2/5}$. In that case, it is not hard to show that the error in our simulation is roughly 7.5% if $n$ is greater than $10^4$, in which case $\langle l \rangle \sim 40$. These numbers suggest that our simulation might not be very relevant for future experiments. It seems experimentally inconceivable to perform a boson sampling experiment in a loss regime where one needs to prepare $10^4$ photons in order to get 40 out.

On the other hand, it is important to stress that part of the reason why our simulation will not be too relevant for near-future experiments is that losses that increase exponentially with circuit depth are already a much more stringent obstacle to scalability. Recall that if $\eta_{\text{eff}} = C/n^s$ for $x > 1$, then $\langle l \rangle$ actually decreases with $n$, and trying to scale an experiment in this regime would be pointless. Suppose now that we have a circuit where we applied our theorem 2 and extracted $s$ layers of uniform losses. Suppose also that $\eta$ is the transmissivity for each beamsplitter. Then we have $\eta_{\text{eff}} = \eta^s$. Thus, one can show that $\eta_{\text{eff}} = 1/\eta^s$ corresponds to $s = x \log(n) / \log(1/\eta)$. This means the limit $x = 1$ is already achieved for circuit depth $\log(n) / \log(1/\eta)$, which corresponds to 96 layers if we assume a loss probability of 4% per beamsplitter [23] and $n = 50$. Additionally, our classical simulation can formally also be avoided (i.e. $x < 1/2$) if one is restricted to depths smaller than 46.

Connection to the mean-field state—Consider the closest separable state to $\rho_{\text{bs}}$ from equation (20), namely the state $\sigma_n$ in theorem 1. This state has appeared previously in the boson sampling literature, where it was called a mean-field state [67]. A device that performs a boson sampling experiment using this state as input has been proposed in [67] as a semi-classical approximation to a boson sampling device that would be classically simulable, but that would be able to fool methods capable of distinguishing boson sampling distributions from alternative ones [22, 68, 69]. In particular, the mean-field state has the ability to effectively display more ‘bunching’ effects than distinguishable classical particles, which is usually considered a genuinely bosonic signature. It is interesting then that this state has also appeared in our main results as the closest symmetric separable state to the lossy bosonic state, as answer to a question that is seemingly unrelated to the matter of verification of boson sampling devices.

Quantum de Finetti theorem—It is natural to investigate the connection of our results to quantum de Finetti theorem [70–72]. In particular, the quantum de Finetti theorem for bosonic systems is concerned with the best approximation (in trace distance) of the state $\rho_{n-\langle \Psi \rangle}$ by separable product states $\sigma \in \text{Sep}(S_{\text{in}})$, where $\Psi$ is an $n$-particle $m$-mode pure bosonic state. This is exactly the setting that we consider in the case of the fixed-loss model considered in theorem 1. Applying directly the bosonic de Finetti [71] theorem yields

$$\min_{\sigma \in \text{Sep}(S_{\text{in}})} d_{\text{tr}}(\rho_{n-\langle \Psi \rangle}, \sigma) \leq \frac{2bm}{n}. \quad (32)$$
In the context of boson sampling we always have \( m \geq n \) and thus the above bound becomes useless. The much stronger result given in theorem 1 was possible only because it concerns separable approximation to the particular lossy state \( \rho_{1,n} / (\vert \psi \rangle \langle \psi \vert) \), where \( \psi \) is defined in equation (1).

**Entanglement of symmetric states**—Our results can be also linked to studies of entanglement of symmetric states, which recently received a lot attention [35–37, 73, 74]. In particular, due to the difficulty of the general separability problem, many works consider the problem of deciding whether states diagonal in the Dicke basis are entangled [36, 37, 74]. However, little attention has been given to computing entanglement measures or entanglement indicators for such states. The trace distance to the set of symmetric separable states equation (21) can be used as an indicator of entanglement and as a lower bound the geometric measure of entanglement [75].

Thus, our results can be also understood as quantitative description for a class of lossy bosonic states. We believe that tools introduced in this work will be useful in quantitative studies of entanglement of symmetric states diagonal in the Dicke basis.

**Boson sampling certification**—One issue that was recognized in the original boson sampling proposal [3] and has been discussed at length in the literature since [21, 22, 67–69] is the issue of certification of a boson sampling device. The main question is whether it is possible to apply some statistical test to a set of samples to develop confidence that the data indeed comes from the ideal boson sampling distribution. The certification complexity of the test depends on the number of samples needed for a given confidence level, as well as the computational complexity of applying the test itself. This issue was raised more noticeably in [69], where it was argued that the ideal boson sampling distribution might be hard to distinguish even from the uniform distribution. Subsequent work [21, 22, 67, 68] has developed tests to validate boson sampling not only against the uniform distribution, but also against other physically-motivated classically-simulable distributions.

Consider the question of certification in the fixed-loss model (although our arguments can be easily extended to the other loss models as well). From corollary 2 we know that, when \( l = o(n^{1/2}) \) photons are left, the distribution generated in our simulation is closer than \( \Delta_{1,n} \approx \frac{l^2}{2n} \) in TVD to the one generated in a lossy boson sampling device. The TVD is closely connected to hypothesis testing, and can give us estimates on the number of samples required to distinguish two distributions [76]. More concretely, consider two hypotheses \( \mathcal{H}_L \) and \( \mathcal{H}_S \) associated with probability distributions \( \mathcal{D}_L \) and \( \mathcal{D}_S \), and let \( \| \mathcal{D}_L - \mathcal{D}_S \| = \delta \). The probability of making an error in identifying the hypothesis (including both false positives and false negatives) is \( P_{err} = (1 - \delta) / 2 \). If we now draw \( k \) samples from either distribution, the relevant TVD is that between \( k \) copies of the distributions. This can be bounded as

\[
1 - 2e^{-\frac{k\delta^2}{2}} \leq \| \mathcal{D}_L^k - \mathcal{D}_S^k \| \leq k\delta.
\]

From these we conclude that, if we choose enough samples such that \( k = O(\Delta_{1,n}^2) = O(n^{1/2}) \), we can make the error probability arbitrarily low. On the other hand, if only few samples are taken such that \( k = o(\Delta_{1,n}) = o(n^{1/2}) \), then the error probability approaches 1/2 (i.e. no test can perform better than just ignoring the samples and randomly tossing a fair coin to decide from which hypothesis that data came from). These arguments give estimates of the sample complexity of some test that would distinguish our simulation from real-world lossy boson sampling data, but we leave as an open question whether there also exists a test that is computationally efficient.

**6. Proofs of the main results**

In this section we prove the main results of the paper: theorems 1 and 2. The proofs given below are essentially self-contained, however along the way we use some auxiliary technical statements whose justification we defer to the Appendix.

**6.1. Best separable approximation with respect to trace distance**

In order to find the best separable approximation to \( \rho_{1,n} \) (in trace distance) we prove a few intermediate results. First, in lemma 2 we show that the symmetries of the state \( \rho_{1,n} \) allow us to limit our attention to separable approximations having a particularly simple ‘twirled’ structure, which we describe explicitly in corollary 5 and lemma 3. Then, in lemma 4 we show that the closest symmetric separable state can be chosen as the twirling of a pure product state. Finally, in lemma 5 we carry out the optimization over pure product seed states which directly yields theorem 1. Proofs of some of the intermediate results are technical and can be left out during the first reading.

---

8. These bounds are standard knowledge in the field of statistics. Slightly tighter bounds can be proven in the context of quantum hypothesis testing using [69] (mainly the proof of theorem 3 therein) and [77], although the final conclusions are essentially the same.
Lemma 2 (Stabilizing subgroups and the structure of the best separable approximation). Let \( \sigma_n \in \text{Sep}(S_{l,n}) \) be a symmetric \( l \)-particle separable state supported on \( n \) modes satisfying

\[
d_{U}(\rho_{l,n}, \sigma_n) = \Delta_{l,n} = \min_{\sigma} d_{U}(\rho_{l,n}, \sigma).
\]

Let \( K \) be some subgroup of \( U(n) \) consisting of unitaries stabilizing the state \( \rho_{l,n} \), i.e.

\[
V \in K \Rightarrow V^\otimes l \rho_{l,n} (V^\otimes l)^\dagger = \rho_{l,n}.
\]

Then the state

\[
\tilde{\sigma}_n = \Lambda_K(\sigma_n) = \int_{K} dV \ V^\otimes l \sigma_n (V^\otimes l)^\dagger
\]

satisfies \( d_{U}(\rho_{l,n}, \tilde{\sigma}_n) = d_{U}(\rho_{l,n}, \sigma_n) = \Delta_{l,n} \). The integration in equation (35) is done with respect to the Haar measure on \( K \).

**Proof.** First note that \( \Lambda_K \) satisfies \( \Lambda_K(\rho_{l,n}) = \rho_{l,n} \). Moreover, \( \Lambda_K \) is a CPTP map and so, by the data-processing inequality \([64]\) we obtain

\[
d_{U}(\rho_{l,n}, \sigma_n) \geq d_{U}(\Lambda_K(\rho_{l,n}), \Lambda_K(\sigma_n)) = d_{U}(\rho_{l,n}, \tilde{\sigma}_n).
\]

Now note that transformations of the form \( V^\otimes l \) preserve separable symmetric states. This means that \( \Lambda_K(\text{Sep}(S_{l,n})) \subseteq \text{Sep}(S_{l,n}) \), and hence \( \tilde{\sigma}_n \in \text{Sep}(S_{l,n}) \). From the definition of \( \tilde{\sigma}_n \) we thus get

\[
d_{U}(\rho_{l,n}, \tilde{\sigma}_n) \geq d_{U}(\rho_{l,n}, \sigma_n) \quad \text{which together with equation (36) concludes the proof.} \]

**Remark 1.** Inspection of the above proof shows that an analogous result holds if \( \rho_{l,n} \) is replaced by arbitrary \( l \)-particle symmetric \( n \)-mode state \( \rho \in \mathcal{D}(S_{l,n}) \) and \( K \) is replaced by any subgroup of the stabilizer of \( \rho \) in \( U(n) \).

The following subgroups of the stabilizer of \( \rho_{l,n} \) will prove especially useful in the computation of \( \Delta_{l,n} \) and finding \( \tilde{\sigma}_n \).

\[
K_c := \{ V_{\phi_1, \ldots, \phi_n} | \phi_k \in [0, 2\pi) \},
\]

\[
K_d := \{ V_\pi | V_{\pi|i} := \pi(i), \pi \in \mathcal{P}(n) \},
\]

where \( V_{\phi_1, \ldots, \phi_n} = \exp\left(i \sum_{k=1}^{n} \phi_k |i\rangle\langle i| \right) \) and \( \mathcal{P}(n) \) is the permutation group of the set \( \{n\} \). Intuitively speaking, \( K_c \) represents linear-optical transformations that are diagonal in the Fock basis, while \( K_d \) represents permutations of the \( n \) modes. Using group-theoretic techniques similar to the ones used in \([78, 79]\) we can show that, by composing elements from \( K_c \) and \( K_d \), one can obtain the entire stabilizer of \( \rho_{l,n} \) in \( U(n) \) (however, in what follows we do not need this result). The twirling operations for \( K_c \) and \( K_d \) read

\[
\Lambda_{K_c}(\rho) = \frac{1}{(2\pi)^n} \int_{0}^{2\pi} d\phi_1 \cdots d\phi_n (V_{\phi_1, \ldots, \phi_n})^\otimes l \rho (V_{\phi_1, \ldots, \phi_n})^\otimes l,
\]

\[
\Lambda_{K_d}(\rho) = \frac{1}{n!} \sum_{\pi \in \mathcal{P}(n)} V_\pi^\otimes l \rho (V_\pi^\otimes l)^\dagger.
\]

**Corollary 5.** By application of lemma 2 first to \( K_c \) and then to \( K_d \), we obtain that the closest separable symmetric state to \( \rho_{l,n} \) (in trace distance) can be chosen as \( \tilde{\sigma_n} = \Lambda_{K_c} \circ \Lambda_{K_d}(\sigma) \), for \( \sigma \in \text{Sep}(S_{l,n}) \).

We now describe the action of the operations of equation (38) on product symmetric states \( |\phi\rangle^\otimes l \). To this end we need to introduce some auxiliary notation. We define the type of the \( n \)-mode occupation number string \( n = (n_1, \ldots, n_n) \) as the string \( \tau(n) = (\tau_1, \ldots, \tau_n) \) consisting of non-increasingly ordered components of \( n \). Two Fock states \( |n\rangle \) and \( |n'\rangle \) can be mapped to each other by a linear-optical transformations from \( V_{n} \in K_d \) if and only if \( \tau(n) = \tau(n') \). Just like for occupation-number strings, we define the length of the type \( \tau \) as

\[
|\tau| = \sum_{i=1}^{n} \tau_i.
\]

We use the following notation for multinomial coefficients:

\[
\binom{|\tau|}{\tau} = \frac{|\tau|!}{\prod_{i} \tau_i!}.
\]

For a given type \( \tau \) we can define the corresponding mixed state

\[
\rho_\tau = \frac{1}{N(\tau)} \sum_{\mathbf{n} : \tau(n) = \tau} |\mathbf{n}\rangle\langle \mathbf{n}|,
\]

where \( N(\tau) \) is a normalization constant equal to the cardinality of the set \( \{\mathbf{n} | \tau(n) = \tau\} \). In particular, for \( \tau_0 = (1, 1, \ldots, 1, 0, \ldots, 0) \) we have \( N(\tau_0) = \binom{n}{l} \) and \( \rho_{\tau_0} = \rho_{l,n} \). For a vector \( \alpha = (\alpha_1, \ldots, \alpha_n) \) and an

\[9\]

In other words, \( K \) is a subgroup of the stabilizer group \( \text{Stab}(\rho_{l,n}) \) of \( \rho_{l,n} \) in \( U(n) \).
occupation number string \( n \), define \(|\alpha|^n := \prod_{i=1}^n |\alpha_i|^n \). Finally, to state our results we also need the following polynomial in \( \alpha \) depending on the type \( \tau \):

\[
m_{\tau}(\alpha) := \sum_{\mathbf{n}: \mathbf{n}(\mathbf{m}) = \tau} |\alpha|^n.
\]  

**Lemma 3 (Structure of the twirled product states).** Let \(|\phi\rangle = \sum_{i=1}^n \alpha_i |i\rangle\rangle\) be a pure state. Let \( K_{n}, K_{q} \) be as in equation (37). Then \( \Lambda_{K_{n}} \) be a twirling map defined by equation (35). Then we have the following result

\[
\Lambda_{K_{n}} \circ \Lambda_{K_{q}}(|\phi\rangle\langle\phi|) = \sum_{\tau|\tau| = l} m_{\tau}(\alpha) \left( \frac{1}{\tau} \right)^{l} \rho_{\tau}
\]

with the notation defined below corollary 5.

The proof of the above lemma is cumbersome, so we present it in appendix A. The above characterization of the twirled product states allows us to obtain the following important result.

**Lemma 4 (A seed state can be chosen to be a pure product state).** Let \( \sigma = \sum \rho_{\alpha} |\phi_{\alpha}\rangle\langle\phi_{\alpha}| \) be an arbitrary symmetric product states on \( \mathbb{C}^n \) be as in equation (37) and let \( \Lambda_{K_{n}} \) be the map defined by equation (35). Then the trace distance between \( \Lambda_{K_{n}} \circ \Lambda_{K_{q}}(\sigma) \) and \( \rho_{\alpha} \) is given by

\[
d_{\alpha}(\rho_{\alpha}, \Lambda_{K_{q}} \circ \Lambda_{K_{q}}(\sigma)) = 1 - \sum_{k} \rho_{k} q(|\phi_{k}\rangle\langle\phi_{k}|),
\]

where for \(|\phi\rangle = \sum_{i=1}^n \alpha_i |i\rangle\rangle\) we have defined \( q(|\phi\rangle\langle\phi|) := \prod_{i=1}^n m_{\alpha_i}(\alpha) \). Therefore, the closest separable symmetric state to \( \rho_{\alpha} \) (in trace distance) can be chosen as \( \sigma_{\alpha} = \Lambda_{K_{n}} \circ \Lambda_{K_{q}}(|\phi_{\alpha}\rangle\langle\phi_{\alpha}|) \), for a suitable product state \(|\phi_{\alpha}\rangle\langle\phi_{\alpha}|\).

**Proof.** Let \(|\phi\rangle = \sum_{i=1}^n \alpha_i |i\rangle\rangle\). We note that equation (41) implies the following convex decomposition

\[
\Lambda_{K_{n}} \circ \Lambda_{K_{q}}(|\phi\rangle\langle\phi|) = q \rho_{\alpha} + (1 - q) \rho_{\alpha}',
\]

where \( \rho_{\alpha} \) is supported on the subspace orthogonal to the support of \( \rho_{\alpha} \) and

\[
q = q(|\phi\rangle\langle\phi|) = \prod_{i=1}^n m_{\alpha_i}(\alpha).
\]

Now, for an arbitrary separable symmetric state \( \sigma = \sum \rho_{\alpha} |\phi_{\alpha}\rangle\langle\phi_{\alpha}| \) we obtain

\[
\Lambda_{K_{n}} \circ \Lambda_{K_{q}}(\sigma) = q \rho_{\alpha} + (1 - q) \rho_{\alpha}',
\]

where \( q' = \sum \rho_{\alpha} q_{\alpha} q_{\alpha}' = q(|\phi_{\alpha}\rangle\langle\phi_{\alpha}|) \) and again \( \rho_{\alpha}' \) is supported on the subspace orthogonal to the support of \( \rho_{\alpha} \). Now, using the explicit formula for the trace distance given in equation (15), we obtain directly equation (42).

We now state the result that settles the optimization problem for seed symmetric product states. This together with lemma 4 and corollary 5 concludes the proof theorem 1.

**Lemma 5 (Optimization over seed symmetric product states).** Let \(|\phi\rangle = \sum_{i=1}^n \alpha_i |i\rangle\rangle\) be a pure state on \( \mathbb{C}^n \). The maximal value of the function

\[
q(|\phi\rangle\langle\phi|) = \prod_{i=1}^n m_{\alpha_i}(\alpha)
\]

is attained for the ‘maximally coherent’ state \(|\phi_{\alpha}\rangle = \frac{1}{n!} \sum_{i=1}^n |i\rangle\rangle\). Consequently,

\[
\min_{|\phi\rangle \in \mathbb{C}^n, |\langle\phi|\rangle = 1} d_{\alpha}(\rho_{\alpha}, \Lambda_{K_{n}} \circ \Lambda_{K_{q}}(|\phi_{\alpha}\rangle\langle\phi_{\alpha}|)) = 1 - \frac{n!}{n^n (n - 1)!}.
\]

**Proof.** Note that \( q(|\phi\rangle\langle\phi|) \) is proportional to the elementary symmetric polynomial of degree \( l \) in variables \( p = (p_1, \ldots, p_n) \), where \( p_i = |\alpha_i|^2 \). Elementary symmetric polynomials are Schur-concave [80] i.e. satisfy \( q'(p) \leq q(S p) \) for any doubly-stochastic \( n \times n \) matrix \( S \). Setting \( S = \frac{1}{n!} \) we obtain that for any pure state \(|\phi\rangle\) we have the inequality \( q(|\phi_{\alpha}\rangle\langle\phi_{\alpha}|) > q(|\phi\rangle\langle\phi|) \). Finally, equation (47) follows by inserting \( |\alpha|^2 = 1/n \) in to equation (46) and using equation (42).

**Remark 2.** For the optimal pure state \(|\phi_{\alpha}\rangle\) from (47) we have

\[
\Lambda_{K_{n}} \circ \Lambda_{K_{q}}(|\phi_{\alpha}\rangle\langle\phi_{\alpha}|) = \Lambda_{K_{q}}(|\phi_{\alpha}\rangle\langle\phi_{\alpha}|),
\]

which gives exactly the form of the optimal separable state given in theorem 1.
6.2. Extraction of uniform losses from a lossy network

Let us now prove theorem 2. Our proof is constructive, in the sense that it follows directly from an algorithm to obtain $\tilde{\mathcal{N}}$ from $\mathcal{N}$. A central ingredient of our proof is the equivalence shown in figure 5, namely that two equal losses at the output modes of a beamsplitter can be commuted to left to act on its input modes. Throughout this section we will say that losses can be ‘pulled out’ from a network if, by repeated application of that identity to individual beamsplitters, they can moved all the way to the input of the circuit. Whenever we pull out one loss element for each input in the network, we say we pulled out one uniform layer of losses. The network $\tilde{\mathcal{N}}$ will be obtained by pulling out $s$ layers of losses from $\mathcal{N}$, one at a time.

Recall that, by assumption, $\mathcal{N}$ is such that every path from the input to the output has to traverse at least $s$ beamsplitters. Let us temporarily call the length of a path within the network the number of loss elements that a photon would traverse following that path, irrespective of the actual number of beamsplitters. This distinction is initially irrelevant since every beamsplitter has exactly one loss element in each of its input paths, but it becomes important in intermediate stages after we start to commute loss elements around. Let us call a path that connects an input to an output an IO-path. By assumption, the shortest IO-path in $\mathcal{N}$ has length $s$.

We now describe the procedure step by step.

STEP 0: Label every beamsplitter in $\mathcal{N}$ by the shortest path between it and any input. There might be exponentially many paths within the network, but this procedure can be done efficiently if it is performed in a breadth-first manner. First, label all beamsplitters immediately connected to the inputs as 1. Then, move to all beamsplitters that are connected to those and label them. Repeat this procedure until all beamsplitters have been labeled. For any beamsplitter, its label is at most 1 greater than the smaller of the labels of the two that immediately precede it. This procedure can be done in time linear in the number of beamsplitters, and is exemplified in figure 11.

STEP 1: Consider the set of loss elements that lie between a beamsplitter labeled 1 and an input (e.g. the highlighted ones in figure 11). Pull these losses out to the input as an uniform layer, and remove them from the network. This reduces the shortest IO-path by 1. Recompute the labels of all beamsplitters accordingly.

We now apply mathematical induction. Suppose that, on STEP $j - 1$, it was possible to pull out losses from every beamsplitter labeled 1 at that stage, and that these losses were pulled out as a uniform layer (which we then removed from the network, relabeling all beamsplitters in the process). Furthermore, suppose that, by this procedure, the length of the shortest IO-path was reduced by 1. Let $G_{j-1}$ be the set of beamsplitters that were relabeled from 1 to 0 in that step. Then apply the following.

STEP $j$: If the shortest IO-path has length 0, terminate the procedure, as $s$ layers of losses have already been pulled out. Otherwise consider the beamsplitters in $G_{j-1}$. Each of them must be in one of the configurations of figures 12(a)–(c), since figure 12(d) would imply that the shortest IO-path has length 0. For each, if there are two loss elements at their output, commute them through to the two inputs. We now claim that this must be possible for every beamsplitter in $G_{j-1}$ in every configuration of figures 12(a)–(c). To see this, note that we only ever commute losses through beamsplitters labeled 0. But at the beginning of STEP $j - 1$, every beamsplitter in $G_{j-1}$ was
labeled 1, and they were only relabeled from 1 to 0 because losses that preceded them were pulled out. This means that the procedure has not yet touched the losses that follow them along both output modes.

After losses have been commuted through all beamsplitters in $G_{j - 1}$, note that this would reverse their label from 0 back to 1. But this places them in the configuration they were at the beginning of step $j - 1$, relative to the part of the network that precedes them. This means that we can pull out a uniform layer of losses, by the inductive hypothesis. After this uniform layer is pulled out and removed from the circuit, relabel all beamsplitter accordingly. Note that a new set of beamsplitters have been relabeled from 1 to 0, all which lie in the immediate future of the beamsplitters in $G_{j - 1}$. This set is $G_j$, to be used in the next step. By analyzing the configurations of figures 12(a–c), it is also easy to see that this step has reduced the length of the shortest IO-path by 1.

Thus we have shown that the inductive hypothesis also holds for $j$, and so it holds all the way up to step $s$. This procedure is clearly efficient and pulls out $s$ uniform layers of losses. The network that remains afterwards is $\tilde{N}$, as in theorem 2, and is a valid linear-optical network with an efficient description, which completes the proof.

For the particular example of figure 11, the network $\tilde{N}$ is shown in figure 13. In appendix B we provide a detailed walkthrough of the steps that lead to $\tilde{N}$. In figure 13 we notice that some segments in the network seem to accumulate losses in the process of pulling out the uniform layers. This is a curious feature, but does not affect the proof that the procedure works in any way.

7. Discussion and open problems

In this paper we proposed a method for efficient classical simulation of linear-optical experiments, most notably boson sampling, in the high-loss regime. We showed that, in an experiment containing initially $n$ photons, if all but $l = o(\sqrt{n})$ of them are lost we can approximate the resulting state by a symmetric separable state of $l$.
particles. This state has trace distance \( o(l^2/n) \) to the state of interest, and so a simulation based on it becomes better and better as the size of the system increases. The fact that this state is separable (in the first quantization representation of linear optics) and has a simple description implies a straightforward efficient classical simulation of any linear-optical protocol that uses it as input. We also showed that in the more physically-realistic model of losses, where each photon has some probability \( (1 - \eta) \) of being lost (rather than there being a definite number of photons lost) our classical simulation is efficient whenever \( \eta = o(n^{-1/2}) \). Finally, we showed how to perform this simulation in a non-uniform model of loss, i.e. when we have different paths in the linear-optical circuit incurring different losses. We showed that, by assuming that every beamsplitter in a circuit induces some loss \( (1 - \eta) \) and that the shortest path that a photon can take through the circuit passes through \( s \) beamsplitters, our simulation is efficient whenever \( \eta_{\text{eff}} = \sqrt{s} \eta^s = o(n^{-1/2}) \), or alternatively for networks of minimal depth \( s = O(\log n) \) for some suitably large constant (depending on \( \eta \)).

Our results imply that, in order to demonstrate some computational advantage within the boson sampling paradigm, experiments will likely have to shift to different architectures. One possibility is to improve the original results of \([3]\) to show that boson sampling on linear-optical circuits of logarithmic depth is already hard, for example, because such a circuit can already exhibit sufficient Haar-randomness for the purposes of the proofs of \([3]\). Our results place limitations on the capabilities of log-depth circuits, but do not rule out the possibility that sufficiently shallow circuits would lead to hard boson sampling instances—although it is possible to show, using tensor-network methods, that these circuits would require beamsplitters that act between arbitrarily distant modes \([48, 81]\). This at least rules out boson sampling with log-depth networks if one is restricted to the two-dimensional integrated architectures of \([17–20, 23]\).

We leave several open questions for future work, including:

(i) In \([14]\) the authors proved that boson sampling remains hard when \( O(1) \) photons are lost, and here we proved that it becomes classically simulable when only \( O(\sqrt{n}) \) photons remain. The main problem we leave open is at which point between \( O(\sqrt{n}) \) and \( n - O(1) \) losses the transition between these two complexity regimes occurs.

(ii) Can we get better approximations if we allow for particle-separable states that are not symmetric? Alternatively, can we prove that the closest state to our desired state \( \rho_{l,n} \) must be symmetric?

(iii) Our results bound the trace distance between the mean-field state and the lossy input state, as in theorem 1. This also acts as a bound on the TVD between the corresponding distributions when both states are input in to arbitrary linear-optical transformations. What happens with the TVD after a Haar-random unitary \( U \)? Would the distance between the output states after a random \( U \) be considerably smaller than the worst-case bound we obtain? Would this change the scaling of our main result and our conclusions?

(iv) We showed that, for a network where the shortest path connecting any input to any output is \( s \), one can remove \( s \) layers of uniform losses. But it is possible to decompose an arbitrary linear-optical unitary such that \( s = 1 \), e.g. the standard decomposition of Reck et al in figure 7(b) \([38]\). In \([39]\), the authors propose a balanced decomposition, as in figure 7(a), since there losses behave uniformly and induce less non-unitary noise. What if one instead tried to engineer the unbalanced noise in a helpful way? That is, can we argue that boson sampling must be hard even under the non-unitary noise coming from the unbalanced losses in figure 7(b)? If so, that might provide a way to engineer long networks with small \( s \) tailored to circumvent our argument.

(v) As discussed in section 5, our simulations get better as the size of the problem increases, but seemingly cannot be made arbitrarily precise at the cost of more computational time. This means we do not have an \( \epsilon \)-simulation as per the definition of \([3, 61]\), although our notion of simulation agrees with the one used e.g. in \([59]\). Can our simulation be improved in to the more stringent form of an \( \epsilon \)-simulation? Alternatively, do these different notions of simulation actually lead to fundamentally different complexity regimes?

(vi) Can our simulation be adapted for other variants of boson sampling \([24, 29, 30]\), or for other candidates of quantum computational supremacy experiments \([4, 6–8, 59]\)?

(vii) Is it possible to derive a version of the de Finetti theorem that, for symmetric states diagonal in Dicke basis, yields a bound stronger than the one in equation (32)?

**Note:** Upon completion of this work, we became aware of a simultaneous work \([81]\) that seems to arrive at similar results for classical simulation of boson sampling in the case of uniform losses. In addition, after the completion of our work, another paper was published \([82]\). It investigates complexity-theoretic aspects of boson sampling also using a first-quantization formalism, in particular addressing photon distinguishability and losses.


Acknowledgments

The authors would like to thank Scott Aaronson, Alexander Strelcov, Michal Horodecki, Hakop Pashayan, Rafał Demkowicz-Dobrzański, Zbigniew Puchała, Martin Kleisch, Ashley Montanaro, and others, for helpful discussions.

MO acknowledges the support of the Homing programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund. DB acknowledges financial support by CAPESBrazilian Federal Agency for Support and Evaluation of Graduate Education within the Ministry of Education of Brazil.

Appendix A. Proofs of technical results

In this part of the appendix we collect proofs of some of the auxiliary results that were needed to obtain main findings of the paper.

Lemma A1. Let $\Delta(\eta, n)$ be the trace distance between states $\sigma_\eta$ and $\rho_{\eta,n}$ defined in equations (25) and (26), respectively. We have the following inequalities

\[
\Delta(\eta, n) \leq \frac{\eta^2 n}{2} + \frac{\eta(1 - \eta)}{2},
\]

\[
\Delta(\eta, n) \geq \Delta_{\{\eta(1 - \delta)n\}, n}(1 - \exp(-\delta^2 n/2)),
\]

where $\delta \in [0, 1]$ and $[x]$ denotes the smallest integer greater than $x$.

Proof. We first observe that $\Delta(\eta, n)$ given in equation (28) can be interpreted as the average of $\Delta_{l,n}$, where $l$ is distributed according to the binomial distribution $\mathcal{B}(n\eta, \eta)$. A 3

Now, in order to prove equation (A1) we use the upper bound $\ln l \leq l^2/2$, which can be immediately proven using mathematical induction in $l$. Applying this bound to equation (A3) yields

\[
\Delta(\eta, n) \leq \sum_{l=0}^{n} \Pr(l) \Delta_{l,n} = \frac{\eta^2 n}{2} + \frac{\eta(1 - \eta)}{2},
\]

where in the last equality we have used the standard properties of the binomial distribution.

The proof of equation (A2) also uses equation (A3). Since $\Delta_{l,n} \geq \sum_{l=0}^{n} \Pr(l) \Delta_{l,n}$, for $l_n = \lceil \eta n (1 - \delta) \rceil$, we obtain

\[
\sum_{l=0}^{n} \Pr(l) \Delta_{l,n} \geq \Delta_{\{\eta(1 - \delta)n\}, n} \Pr(l \geq \eta n (1 - \delta)).
\]

Finally, equation (A2) follows from the above inequality by applying the Chernoff bound given in equation (19).

Lemma A2 (Structure of the twirled product states). Let $\ket{\phi} = \sum_{i=1}^{n} \alpha_i \ket{i}$ be a pure state on $\mathbb{C}^n$. Let $K_c, K_d$ be subgroups of $U(n)$ defined in equation (37). Let $\Lambda_K$ be a twirling map defined by equation (35). Then, we have the following result

\[
\Lambda_{K_c} \circ \Lambda_K (\ket{\phi} \bra{\phi}) = \sum_{\tau \mid |\tau| = l} m_{\tau} \left( \frac{l}{\tau} \right) \rho_{\tau}.
\]

where we have used the notation introduced below corollary 5.

Proof. In many places of this proof we will be using the notation introduced below corollary 5 in the main text. We will deal first with the averaging over the group $K_c$. Consider first the pure state $|\phi\rangle \langle \phi| \in \mathcal{D}(S_{\eta,n})$, for $|\phi\rangle = \sum_{i=1}^{n} \alpha_i |i\rangle$ and $\sum_{i=1}^{n} |\alpha_i|^2$. To simplify the calculations, let us write $|\phi\rangle \langle \phi| = \sum_{i} \alpha_i |i\rangle \langle i|$, where we defined...
The set $G_1$ is identified by the red-labeled beamsplitter. In panel (III), losses have been commuted through the marked beamsplitters, and now occupy the same configuration as before STEP 1, which means that by assumption they can now be pulled out as a uniform layer (this is trivial to see in this particular example, because the losses that will be pulled out are already at the input, but the argument holds in general). Panels (IV) and (V) show the configuration and labeling after STEP 2 and STEP 3, respectively. After STEP 3, three layers of losses have been pulled out and the shortest IO-path has length 0, so the procedure terminates.
Figure B1. An example of our procedure to pull out losses from a linear-optical network. Beamsplitters labeled in red zeroes are the ones that, in step \( j \), are in \( G_{j-1} \), as defined in the main text. Empty boxes correspond to loss elements that have been commuted elsewhere. The blue loss elements in panel (I) are the same as in figure 11, and are the first to be pulled out. In panel (III) they exemplify the argument, given in section 6.2, that in every step, after losses are commuted through the red-labeled beamsplitters, they recover the configuration of the previous step, and by the inductive hypothesis they can always be pulled out.

ORCID iDs

Daniel J Brod @ https://orcid.org/0000-0002-5925-4158

References

[1] Lund A, Bremner M and Ralph T 2017 npj Quantum Inf. 3 115
[2] Harrow A and Montanaro A 2017 Nature 549 203
[3] Aaronson S and Arkhipov A 2013 Theory Comput. 4 143
[4] Bremner M J, Jozsa R and Shepherd D J 2011 Proc. R. Soc. A 467 459
[5] Knill E and Laflamme R 1998 Phys. Rev. Lett. 81 5672
[6] Morimae T, Fujii K and Fitzsimons J F 2014 Phys. Rev. Lett. 112 130502
[7] Aaronson S and Chen L 2017 32nd Computational Complexity Conference (CCC 2017) (Dagstuhl: Schloss Dagstuhl) (https://doi.org/10.4230/LIPIcs.CCC.2017.22)
[8] Boixo S, Isakov S V, Smelyanskiy V N, Babbush R, Ding N, Jiang Z, Bremner M J, Martinis J M and Neven H 2016 Nat. Phys. 14 S95
[9] Shchesnovich V S 2015 Phys. Rev. A 91 063842
[10] Renema J J, Menssen A, Clements W R, Triginer G, Kolthammer W S and Walmsley I A 2018 Phys. Rev. Lett. 120 220502
[11] Kalai G and Kindler G 2014 arXiv:1409.3093
[12] Arkhipov A 2015 Phys. Rev. A 92 062326
[13] Leverrier A and Garcia-Patron R 2015 Quantum Inf. Comput. 15 489
[14] Aaronson S and Brod D J 2016 Phys. Rev. A 93 012335
[15] Rahimi-Keshari S, Ralph T C and Caves C M 2016 Phys. Rev. X 6 021039
[16] Motes K R, Dowling J P and Rohde P P 2013 Phys. Rev. A 88 063822
[17] Broome M A, Fedrizzi A, Rahimi-Keshari S, Dove J, Aaronson S, Ralph T C and White A G 2013 Science 339 794
[18] Crespi A, Osellame R, Ramponi R, Brod D J, Galvão E F, Spagnolo N, Vitelli C, Maiorino E, Mataloni P andSciarrino F 2013 Nat. Photon. 7 545
[19] Spring J B et al 2013 Science 339 798
[20] Tillmann M, Đukić B, Heilmann R, Nolte S, Szameit A and Walther P 2013 Nat. Photon. 7 540
[21] Spagnolo N et al 2014 Nat. Photon. 8 615
[22] Carolan J et al 2014 Nat. Photon. 8 621
[23] Carolan J et al 2015 Science 349 711
[24] Bentivegna M et al 2015 Sci. Adv. 1 e1400235
[25] Loredo J C, Broome M A, Hilaire P, Gazzano O, Sagnes I, Lemaître A, Almeida M P, Senellart P and White A G 2017 Phys. Rev. Lett. 118 130503
[26] He Y et al 2017 Phys. Rev. Lett. 118 199501
[27] Wang H et al 2017 Nat. Photon. 11 361
[28] Wang H et al 2018 Phys. Rev. Lett. 120 230502
[29] Lund A P, Laing A, Rahimi-Keshari S, Rudolph T, O’Brien J L and Ralph T C 2014 Phys. Rev. Lett. 113 100502
[30] S. Aaronson’s blog, acknowledged to S. Kothlhammer, http://scottaronson.com/blog/?p=1579
[31] Neville A, Sparrow C, Clifford R, Johnston E, Birchall P M, Montanaro A and Laing A 2017 Nat. Phys. 13 1153
[32] Clifford P and Clifford R 2018 The classical complexity of boson sampling Proc. 29th Annual ACM-SIAM Symp. on Discrete Algorithms pp 146–55
[33] Wu J, Liu Y, Zhang B, Jin X, Wang Y, Wang H and Yang X 2018 Nat. Sci. Rev. nwx079
[34] Dzellel A M, Harrow A W, Koh D E and La Placa R L 2018 arXiv:1805.05224
[35] Hübenre R, Kleinmann M, Wei T, González-Guillén C and Gühne O 2009 Phys. Rev. A 80 032324
[36] Yu N 2016 Phys. Rev. A 94 060101
[37] Quesada R, Rana S and Sanpera A 2017 Phys. Rev. A 95 042128
[38] Reck M, Zeilinger A, Bernstein H and Bertani P 1994 Phys. Rev. Lett. 73 58
[39] Clements W R, Humphreys P C, Metcalf B J, Kohlhammer W S and Walsmea I A 2016 Optica 3 1460
[40] Demkowicz-Dobrzański R, Jarzyna M and Kolo J 2015 Progress in Optics ed J E Wolf vol 60 (Amsterdam: Elsevier) pp 345–435
[41] Bachor H and Ralph T C 2004 A Guide to Experiments in Quantum Optics (New York: Wiley)
[42] Weedbrook C, Pirandola S, García-Patrón R, Cerf N J, Ralph T C, Shapiro J H and Lloyd S 2012 Rev. Mod. Phys. 84 621
[43] Eckert K, Schlemmern J, Bruß D and Lewenstein M 2002 Ann. Phys., NY 299 88
[44] Killoran N, Cramer M and Plenio M B 2014 Phys. Rev. Lett. 112 150501
[45] Tóth G and Apellaniz I 2014 J. Phys. A: Math. Theor. 47 424006
[46] Wasak T, Smerzi A and Chwvedenczuk J 2016 arXiv:1609.01576
[47] Terhal B and DiVincenzo D 2004 Quantum Inf. Comput. 4 134
[48] Jozsa R 2006 arXiv:quant-ph/0603163
[49] Brod D 2015 Phys. Rev. A 91 042316
[50] Oszmaniec M, Augusiak R, Gogolin C, Koldyrev J, Acín A and Lewenstein M 2016 Phys. Rev. X 6 041044
[51] Schumun T, Hofferberth S, Andersson L M, Wildermuth S, Groth S, Bar-Joseph I, Schmiedmayer J and Krüger P 2005 Nat. Phys. 1 57
[52] Sebby-Strabley J, Brown B L, Anderlini M, Lee P J, Phillips W D, Porto J V and Johnson P R 2007 Phys. Rev. Lett. 98 200405
[53] Zhang H, McConnell R, Ćuk S, Lin Q, Schleier-Smith M H, Leroux I D and Vuletić V 2012 Phys. Rev. Lett. 109 133603
[54] Stroescu I, Hume D B and Oberthaler M K 2015 Phys. Rev. A 91 013412
[55] Barnett S M, Jeffers J, Gatti A and Loudon R 1998 Phys. Rev. A 57 2134
[56] Rahimi-Keshari S, Broome M A, Pickler R, Fedrizzi A, Ralph T C and White A G 2013 Opt. Express 21 13430
[57] Schell S 2004 arXiv:quant-ph/0406127
[58] Valiante I 1979 Theor. Comput. Sci. 8 189
[59] Bremner M, Montanaro A and Shepherd D 2016 Phys. Rev. Lett. 117 080501
[60] van den Nest M 2011 Nat. Commun. 2 1
[61] Pashayan H, Bartlett S and Gross D 2017 arXiv:1712.02806
[62] Bartlett S D and Sanders B C 2003 J. Mod. Opt. 50 2331
[63] Jerrum M, Sinclair A and Vigoda E 2004 J. ACM 51 671
[64] Nielsen M A and Chuang I L 2010 Quantum Computation and Quantum Information (Cambridge: Cambridge University Press)
[65] Chernoff H 1952 Ann. Math. Stat. 23 493
[66] Aaronson S and Kuperberg G Complexity Zoo https://complexityzoo.waterturtle.ca/Complexity_Zoo, (Accessed: 10 January 2018)
[67] Tichy M C, Mayer K, Buchleitner A and Molmer K 2014 Phys. Rev. Lett. 113 020502
[68] Aaronson S and Arkhipov A 2014 Quantum Inf. Comput. 14 1383
[69] Gogolin C, Kliesch M, Aolita L and Eisert J 2013 arXiv:1306.3995
[70] Doherty A C, Parrilo P A and Spedalieri F M 2002 Phys. Rev. Lett. 88 187904
[71] Christandl M, König R, Mitchison G and Renner R 2007 Commun. Math. Phys. 273 473
[72] A W Harrow 2013 arXiv:1308.6595
[73] Oszmaniec M and Vuškovic 2014 Phys. Rev. A 90 010302
[74] Tura J, Aloy A, Quesada R, Lewenstein M and Sanpera A 2017 Quantum 2 45
[75] Streltsov A, Kampermann H and Bruss D 2010 New J. Phys. 12 123004
[76] Lehmann E L and Romero P 2005 Testing Statistical Hypotheses (Springer Texts in Statistics) (New York: Springer)
[77] Audenaert K M R, Mosonyi M and Verstraete F 2012 J. Math. Phys. 53 122205
[78] Bouland A and Aaronson S 2014 Phys. Rev. A 89 062316
[79] Sawicki A 2016 Quantum Inf. Comput. 16 0291
[80] Cleve R and Watkin W 1991 Math. Mag. 64 183
[81] García-Patrón R, Renema J and Schenovich V 2017 arXiv:1712.10037
[82] Moylett A E and Turner P S 2018 Phys. Rev. A 97 062329