Quantum Computation with Identical Bosons

by

Aleksandr Arkhipov

B.S., Massachusetts Institute of Technology (2010)
S.M., Massachusetts Institute of Technology (2012)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2017

Author ..... Signature redacted
Department of Electrical Engineering and Computer Science
August 31, 2017

Certified by ..... Signature redacted .....
Scott Aaronson
Visiting Associate Professor of Electrical Engineering
and Computer Science
Thesis Supervisor

Accepted by ..... Signature redacted ....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Quantum Computation with Identical Bosons

by

Aleksandr Arkhipov

Submitted to the Department of Electrical Engineering and Computer Science
on August 31, 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

We investigate the computational complexity of quantum computing with identical noninteracting bosons, such as that in a linear optical system. We explore the challenges in building devices that implement this model and in certifying their correctness.

In work done with Scott Aaronson, we introduce BOSONSAMPLING, a computational model of quantum linear optics [1]. We argue that the statistical distribution of outcomes cannot be reproduced by any classical device in a reasonable time span. This gives hands-on evidence of quantum advantage, that there are quantum phenomena are prohibitive to simulate in the classical world. Moreover, this quantum advantage is already present in limited optical systems, suggesting a lower bar to building devices that exhibit super-classical computation. We lay out the computational complexity argument for the classical difficulty of simulating BOSONSAMPLING. An efficient classical simulation would have unlikely complexity consequences for the polynomial hierarchy PH. We look into the difficulties in proving an analogous approximate result, including the conjectures that seem to be needed to push it through.

We then discuss experimental implementations of BOSONSAMPLING. The scalability of current implementations is limited by various sources of noise that accumulate as the problem size grows. We prove a result [5] that pertains to the inexactnesses of components that comprise the linear optical network, giving bounds on the tolerances that suffice to obtain an output distribution close to the ideal one.

Finally, we look at the challenge of certifying a BOSONSAMPLING device. We show the impossibility of one technique, to use a submatrix whose permanent is so large that its corresponding outcome appears very frequently. Joint work with Aaronson [2] argues that the outputs of a BOSONSAMPLING device can be verified not to come from a uniform distribution. Results on the statistical bunching of bosons obtained with Kuperberg [6] are another approach to certification. We further present a novel certification technique based on classically estimating the distribution of integer combinations of the boson counts.

Thesis Supervisor: Scott Aaronson
Title: Visiting Associate Professor of Electrical Engineering
and Computer Science

Acknowledgments

I'd like to thank my advisor Scott Aaronson, who has been with me through my whole graduate career and more. Even before I met Scott, his blog *Shtetl Optimized* interested me in what he was interested. My undergrad sophomore year, I took Scott's class *Great Ideas in Theoretical Computer Science*, a smorgasbord of Scott's favorite topics, and his contagious enthusiasm finally pushed me to seriously consider computer science as a major. He invited me to join him for an undergraduate research project, a seed that would grow into a sprawling body of research that this thesis is a part of, endlessly branching into new problems, extensions, and unexpected connections. I appreciate Scott's boundless supply of excitement and open problems, his drive to understand and explain, his patient guidance, and how he always pushed me to connect with his vast network of fellow researchers, to travel to conferences, and to give as many talks as possible.

I thank my family for their endless love and encouragement through this long process. Their enthusiastic kindling of my interest in math since I was young has put me where I am now. I thank my friends who have shared my bizarre enthusiasm for abstruse interests like board games and math puzzles and have made me feel at home in my strange corner of the world.

I'm grateful to many professors I've had at MIT for their dedication to spreading their knowledge. I'm grateful in particular to Michael Sipser for being a paragon of clear lecturing and writing to strive towards, and to Silvio Micali for his zealously Socratic lectures, which forced me to consider the reasons behind every definition. I also thank the authors of math books I've read online for sharing their labors of love for free, like Sergei Winitzki's *Linear algebra via Exterior Products* and Herbert Wilf's *generatingfunctionology*.

I thank my thesis committee for their time and energy. I would also like to acknowledge the professors and organizers of universities and programs who invited and accommodated me for many productive visits, and the funding I've gotten from MIT, Akamai, the NSF, and other sources.

Contents

List of Figures

List of Tables

0.1 Extended abstract

We investigate the computational complexity of quantum computing with identical noninteracting bosons, such as that in a linear optical system. We explore the challenges in building devices that implement this model and in certifying their correctness.

In work done with Scott Aaronson, we introduce BOSONSAMPLING, a computational model of quantum linear optics [1]. A fixed number of identical photons are produced in different modes, pass through a network of beamsplitters and phase-shifters, and are measured in number for each output mode. This model is similar to the Linear Optical Quantum Computing model [27], but has all the measurements performed at once after the network without any adaptiveness. Each run of the experiment produces different random counts. We argue that the statistical distribution of outcomes cannot be reproduced by any classical device in a reasonable time span. This gives hands-on evidence of quantum advantage, that there are quantum phenomena are prohibitive to simulate in the classical world. Moreover, this quantum advantage is already present in limited optical systems, suggesting a lower bar to building devices that exhibit super-classical computation. In particular, this system uses only a limited form of coupling between the photons, and we do not believe it can perform universal quantum computation.

We lay out the computational complexity argument for the classical difficulty of simulating BOSONSAMPLING. At its core is a matrix function called the permanent, which dictates the probability of each outcome. The connection of the permanent to linear optics has been made before, for example by Scheel [38]. In a celebrated result of Valiant [53], computing the permanent is #P-complete and so captures the full power of counting problems like finding the number of satisfying assignments to a Boolean circuit. Using the technique of approximate counting [46] with the help of an NP oracle, one could obtain the probability of an outcome of a classical simulation, and therefore estimate the permanent of a chosen embedded submatrix. This would have unlikely complexity consequences for the polynomial hierarchy PH. We discuss the

difficulties in proving an analogous approximate result, including the conjectures that seem to be needed to push it through: the Permanent of Gaussians Conjecture (PGC), and the Permanent Anti-Concentration Conjecture (PACC). Though unproven, these conjectures are supported by numerical and heuristic evidence.

We further discuss why linear optics in particular is fertile for a result of this nature by comparing similar systems that are nevertheless classically simulable. Though analogues with classical or distinguishable particles are also governed by the matrix permanent, the permanent is of a matrix of non-negative entries. Unlike with complex amplitudes, there cannot be any cancellations between terms of different sign of phase, a uniquely quantum phenomenon. Non-negative permanents can be classically approximated [24]. Fermions are a natural analogue to bosons, with the difference that their probabilities are governed by the determinant rather than the permanent. Unlike the #P-complete permanent, the determinant is polynomial-time computable, and understanding this gulf in difficulty is a central problem in geometric complexity theory (see for example [28]). Moreover, fermions allow an efficient classical sampling algorithm [54] based on computing marginal probabilities.

We then discuss experimental implementations of BOSONSAMPLING. The scalability of current implementations is limited by various sources of noise that accumulate as the problem size grows, including as incorrect input photon counts, inaccuracies in the network components, partial distinguishability, and photon loss in measurement. This includes an overview of theoretical bounds that show certain degrees of noise are prohibitive as in [30] and [36]. We prove a result [5] that pertains to one particular source of inaccuracy, of inexactnesses in the beamsplitters and phaseshifters that comprise the linear optical network, causing it to implement the wrong unitary operation. We prove bounds on their tolerances that suffice to obtain an output distribution close to the ideal one.

Finally, we look at the challenge of certifying a BOSONSAMPLING device. In contrast to efficient quantum factoring [43], a problem that is in NP and can be easily verified by multiplying the factors, the difficulty of computing the permanent that allows us to prove hardness also prevents us from classically confirming the

results. Moreover, BOSONSAMPLING gives outcomes that are probabilistic and not individually checkable. One potential technique is to certify with a network matrix that has been chosen with a submatrix smuggled in whose permanent is so large that its corresponding outcome appears very frequently. We show that this method is vulnerable to a classical adversary detecting such a special submatrix, by proving that these submatrices are marked by exceptionally large entries and a particular structure. We show striking novel connections between unitary matrices with large permanent, near-isometries of the hypercube, sets of strings with assigned Hamming distances, and low-rank decompositions into ± 1-valued matrices.

Joint work with Aaronson [2] argues that the outputs of a BOSONSAMPLING device can be verified to not derive from a uniform distribution. The output probabilities are correlated with the matrix row norms, allowing a weak sanity check on empirically observed photon counts. A statistical test is discussed [50] based on Fourier matrices causing the output states to be concentrated in a small subset. Results on the statistical bunching of bosons obtained with Kuperberg [6] are another way to check the outcome to be consistent with bosonic behavior. We further present a novel weak certification technique based on a classical algorithm to estimate the distribution of integer combinations of the boson counts.

Chapter 1

Computation with identical bosons and BosonSampling

This chapter is based on work done with Scott Aaronson in The Computational Complexity of Linear Optics [1]

1.1 Overview

1.1.1 Quantum advantage

The Extended Church-Turing Thesis (ECT) states that any computational problem that can be efficiently solved in the physical world can also be efficiently solved on a formal model of a computer, that is a probabilistic polynomial-time Turing machine. If we consider that any physical process can be regarded as the computational problem of simulating it, this claim applies broadly to simulating physical phenomena on computers.

However, our understanding of the physical world has expanded to include quantum-mechanical phenomena such as superposition and entanglement, which are not represented in our classical notions of computing. This raises the question of whether these quantum phenomena can be classically simulated in a computational sense, of taking as input a mathematical description of the system and its initial state, and

outputting the observed outcomes.

Evidence has accumulated that the ECT is false. A *quantum advantage* is broadly conjectured, that quantum systems can be computationally more powerful than any classical counterpart. Indeed, harnessing the unique properties of quantum mechanics for computation is central to the research endeavor of quantum computing. We hope to build a practical universal quantum computer that will allow us to efficiently solve problems outside our classical reach.

Proving the existence of a quantum advantage is a central problem of quantum complexity theory, which seeks to categorize the computational hardness of formal problems, as well as the power of formal models of computation. The open problem of BPP vs BQP asks whether quantum algorithms can efficiently solve some problem that classical ones cannot. More strongly, we want to understand what gives quantum computers their apparent advantage. Where exactly does the line between quantum and classical computation lie? What uniquely quantum phenomena are needed to cross this boundary?

In this work, we give evidence for quantum advantage. We state a simple model of quantum computing with identical bosons and prove it cannot be efficiently simulated classically unless we accept certain unlikely complexity-theoretic consequences. Moreover, because our model represents a restricted one-step computation with non-interacting bosons, this suggests that the bar for quantum advantage is lower than might be expected.

1.1.2 Our model

We define a model of quantum computation with identical, noninteracting bosons. Physically, it can be implemented as a linear optical network in which a fixed number of identical photons are produced in different modes, pass through a network of beamsplitters and phaseshifters, and are measured to obtain a photon count for each output mode. This is similar to a standard qubit-based circuit with the particles passing though a series of gates followed by measurements. A key difference is that the photons are identical, so the correspondence of input and output photons is in

superposition even after the measurements are performed.

Our model follows a long history of work in optical interferometry, which investigates interference between waves of light. We compare it to existing similar models.

The Hong-Ou-Mandel dip [23] is an experimentally-observed two-photon interference phenomenon that can be seen as a special case of our model. Two identical photons enter a 50:50 beamsplitter from opposite sides. Individually, each photon would be equally likely to pass through the beamsplitter as to bounce off it, doing so in superposition. When both photons hit the beamsplitter at the same time, they are always observed to exit on the same side.

The Linear Optical Quantum Computing (LOQC) model of Knill, Laflamme, and Milburn [27] shows how to do universal quantum computation on a linear optical device aided by adaptive measurements. Our model is based on a similar construction of a network of photons passing through beamsplitters and phaseshifters and being measured by photodetectors, but does not use adaptiveness and is not believed to be universal. Another scheme by Gottesman, Kitaev, and Preskill [21] instead expresses states in terms of modes of a harmonic oscillator.

Scheel [38] looks at linear-optical networks of identical photons and connects the superposition of paths that the photons may take to computing a matrix function called the *permanent*. We make the same connection, and it is central to our complexity results.

1.1.3 The computational complexity of BosonSampling

We define the computational problem of BosonSampling as randomly sampling from the distribution of a system of identical noninteracting bosons. This is a sampling problem, in that it requires randomly generating an outcome such that each possible outcome occurs with the correct probability. BosonSampling can be performed efficiently on a quantum computer by directly simulating the individual bosons. However, we do not know this problem to be universal for quantum computing, or even for classical computing.

Nevertheless, we give evidence that this problem is outside the reach of efficient

classical computing. Our exact result states that a classical computer cannot perform BOSONSAMPLING exactly unless the polynomial hierarchy collapses.

Theorem 1 (Complexity of exact BOSONSAMPLING). *No polynomial-time classical randomized algorithm can perform* BOSONSAMPLING, *unless* $\mathsf{P}^{\#\mathsf{P}} = \mathsf{BPP}^{\mathsf{NP}}$ *and the polynomial hierarchy collapses.*

An approximate version of the result derives consequences of an algorithm to approximate BOSONSAMPLING by sampling a distribution that's close in variation distance, but it assumes the hardness of a permanent estimate problem $|\mathrm{GPE}|^2_{\pm}$.

Theorem 2 (Complexity of approximate BOSONSAMPLING). *For* $m \geq \frac{n^5}{\delta}\log^2\left(\frac{n}{\delta}\right)$, *there does not exist a classical randomized algorithm* A *that performs approximate* BOSONSAMPLING *in time* $\mathrm{poly}(m, n, 1/\epsilon)$ *unless the* $|\mathrm{GPE}|^2_{\pm}$ *problem is solvable in* $\mathsf{BPP}^{\mathsf{NP}}$.

The $|\mathrm{GPE}|^2_{\pm}$ *problem is: given a matrix* $M \sim \mathcal{G}^{n\times n}$, *a tolerance* ϵ, *and a maximum failure chance* δ, *output* $|\mathrm{Per}\,(M)|^2$ *to within* $\pm\epsilon n!$ *with probability at least* $1-\delta$. *Do this in running time polynomial in* n, $1/\epsilon$, *and* $1/\delta$.

We further show that the hardness of $|\mathrm{GPE}|^2_{\pm}$ can be broken down into two conjectures that together imply it, and so the hardness of approximate boson sampling. These are the Permanent-of-Gaussians Conjecture and the Permanent Anti-Concentration Conjecture.

Conjecture 3 (Permanent-of-Gaussians Conjecture (PGC)). *The following problem called* $\mathrm{GPE}_{\times}$ *is* $\#\mathsf{P}$*-hard: given a matrix* $M \sim \mathcal{G}^{n\times n}$, *a tolerance* ϵ, *and a maximum failure chance* δ, *output* $\mathrm{Per}\,(M)$ *to within* $\pm\epsilon\,|\mathrm{Per}\,(M)|$ *with probability at least* $1-\delta$. *Do this in running time polynomial in* n, $1/\epsilon$, *and* $1/\delta$.

Conjecture 4 (Permanent Anti-Concentration Conjecture (PACC)). *There are constants* C, D *and* $\beta > 0$ *so that for any* n *and* $\epsilon > 0$

$$\Pr_{M\sim\mathcal{G}^{n\times n}}\left[\frac{|\mathrm{Per}\,(M)|^2}{n!} < \epsilon\right] < Cn^D\epsilon^{\beta}$$

1.1.4 Comparison to Shor's algorithm

Shor's algorithm [43] performs factoring in polynomial time on a quantum computer. We view it as evidence for quantum advantage: either this quantum circuit is solving a classically hard problem, or factoring is classically feasible. It is generally conjectured that factoring is not classically feasible, as no efficient algorithm has been found after many years of trying, and the difficulty of factorizing is crucial to the security of cryptographic algorithms such as the RSA public-key cryptosystem. Nevertheless, the problem remains open.

We compare our result to Shor's algorithm as two pieces of evidence for quantum advantage. The style of the results is quite different. For factoring, the main breakthrough is a clever quantum circuit that performs factoring via period-finding. The problem of factoring has existed for centuries, and the new insight shows it to be doable by a quantum algorithm.

BOSONSAMPLING, however, is clearly solvable by a quantum system because it is the problem of simulating a quantum system. The problem is designed to be exactly what a certain kind of quantum computer can do. The work in proving quantum advantage is instead in showing this problem not to be classically solvable.

Unlike factoring, the problem is artificial and specifically designed to demonstrate this quantum advantage. We don't know of any useful problem or task that can be solved by a BOSONSAMPLING device. An advantage of our approach is that the problem can be solved directly by a generic linear-optical system. This suggests that quantum advantage doesn't have to be obtained by cleverly engineering a quantum circuit to do a specific task, but can be obtained from general quantum behavior.

Another point of comparison is the assumptions required for the respective results to demonstrate quantum advantage. For Shor's algorithm, the alternative is that factoring can be done in polynomial time. This would certainly be surprising, but would not change the landscape of complexity classes as the collapse of the polynomial hierarchy would for our exact result. However, our approximate result relies on further unproven conjectures.

1.1.5 Quantum advantage with limited resources

BOSONSAMPLING represents a rather restricted quantum system.

First, the network does not create any explicit coupling on the bosons. Physically, it does not have any photon-photon interactions. Its action on n bosons is characterized completely by its unitary action on a single boson. Given this, it may not be clear how it can do any interesting computation at all. The key is that just by virtue of being identical, the photons and their measured counts can become correlated. The Hong-Ou-Mandel dip [23] is a stark example of this with two photons.

The question of entanglement here is a subtle one. The output state can be factored as a product of single-particles states with a symmetric product that accounts for the particles being interchangeable, but not with the standard tensor product, so it is not separable in the usual sense. The state of a BOSONSAMPLING device before measurement, if allowed to interact with particles encoding qubits, could be used to create entanglement between them. However, in terms of just the bosons, only a limited low-dimensional subspace of states can be generated.

The computation is further limited in its structure as a single unitary operation followed by measuring every particle to obtain counts. The unitary operation is fixed in advance and cannot depend on the results of intermediate measurements as allowed in the LOQC model [27].

That we can get quantum advantage with such a restricted model helps us probe at the dividing line between classical and quantum computation. The BOSONSAMPLING problem and its classical analogue differ only in the use of identical bosons as opposed to classical particles or distinguishable bosons. A key feature missing in the classical or distinguishable case is quantum superposition of possible paths of particles through the network, and the cancellations between amplitudes that can result.

In contrast, the lack of explicit coupling between bosons suggests that this is not an essential feature for quantum advantage. The limited form of entanglement that arises from the particles being identical already suffices. Similarly, adaptive operations are not used and so do not appear to be essential.

Furthermore, the limitedness of the BOSONSAMPLING model makes it easier to implement experimentally than a general quantum computer. A fixed network is used for each run, and the particles interact directly with not network and not with other particles. In fact, the absence of photon-photon interactions outside of particle accelerators is helpful for avoiding unwanted interactions that leak information. So, we get the advantage of the excellent coherent properties of photons without the drawback of them being hard to couple.

Linear optics is an active experimental field and BOSONSAMPLING gives the promise of demonstrating quantum advantage sooner than building a universal quantum computer. This is aiming for the lower hurdle of doing *some* quantum computation we believe to be classically infeasible, rather than building a universal computer than can do *every* possible quantum computation. Progress towards building and certifying BOSONSAMPLING devices is discussed in detail in chapter 2 and chapter 3.

1.1.6 Open problems

This work leaves many open problems for potential future work for both theory and engineering.

- Resolve the Permanent of Gaussians Conjecture (PGC), and the Permanent Anti-Concentration Conjecture (PACC)
- Find a decision problem that is equivalent to BOSONSAMPLING
- Use BOSONSAMPLING to solve a useful problem
- Prove or disprove the existence of intermediate complexity classes between classical computing and BOSONSAMPLING
- Show how to certify the correctness of a BOSONSAMPLING device in way that would convince a skeptic
- Build a scalable linear optical device
- Implement fault-tolerance and error correction in linear optical system

1.2 Identical boson model

We formalize the quantum behavior of identical, noninteracting bosons. The development will be analogous to the definition of qubits and multi-qubit states taught in many introductory textbooks of quantum computing, such as *Nielsen and Chuang* [34]. We will define notions of modes, states, transformations, and measurements for the particles.

1.2.1 Single-boson systems

We start with a simple single-boson system, whose definition will be familiar to those who have studied circuit-based quantum computing. It is a particle whose state belongs to an m-dimensional vectorspace, i.e. a qudit. With only one particle, the notion of identicalness of course does not figure.

Definition 5. *A* single-boson quantum system *is defined as follows:*

- ***State:*** *The quantum state of a boson is a vector v in $\mathbb{C}^m$. Let $x_1, x_2, \ldots, x_m$ be the standard basis states for this vectorspace. A state v is a superposition of these basis states as a linear combination $v = v_1 x_1 + v_2 x_2 + \cdots + v_m x_m$.*
- ***Normalization:*** *A state must by normalized as a unit vector, having $\|v\|^2 = |v_1|^2 + |v_2|^2 + \cdots + |v_m|^2 = 1$.*
- ***Transformation:*** *A state can be acted by a unitary map from $\mathbb{C}^m$ to $\mathbb{C}^m$, i.e. a $m \times m$ unitary matrix. The matrix U transforms the state vector v to Uv.*
- ***Measurement:*** *Measuring the state in the standard basis results in a random mode chosen from $j = 1, 2, \ldots, m$ each with probability $|v_j|^2$.*

Note that the measurement probabilities add to 1 because the state v has a norm of 1, and that transformations preserve this norm because they are unitary. After a measurement results in mode j, the state "collapses" to the basis state x_j where it can be further operated on, though we will not need to do so in this work.

1.2.2 States of identical bosons

Now we consider how individual bosons combine into a joint system of identical bosons.

In a standard qubit-based model, this would be via the tensor product: if v_1 and v_2 are single-particle states, their joint two-particle state is a tensor product $v_1 \otimes v_2$. Note that this operation does not commute: $v_1 \otimes v_2 \neq v_2 \otimes v_1$. This represents that it matters which of the two particles is in which state.

This is not the case for identical particles. Because they cannot be distinguished even in principle, their joint state must act the same if the particles were to be swapped. Hence, the combining operation must commute.[1] In multilinear algebra, the symmetric product, often written $\odot$, is a commuting analogue to the tensor product. We will define this operation in an equivalent but more familiar way in terms of multiplying polynomials, where variables naturally commute as $x_j x_k = x_k x_j$.

If one boson is in a basis state x_1 and another in basis state x_2, their joint state is the product $x_1 x_2$. This is the same as $x_2 x_1$. These x_j's are formal variables and do not represent numerical quantities, as in a generating function. For physicists, these are creation operators $a_j^\dagger$.

A general multi-boson basis state is a monomial $x_1^{s_1} x_2^{s_2} \cdots x_m^{s_m}$ where $s_1, \ldots, s_m$ are natural numbers representing the number of bosons in each mode. The degree of the monomial $s_1 + s_2 + \cdots + s_m$ is the total number of bosons n.

Notation 6. *We will use the following notation throughout to write operations on vectors without listing their components:*

- *Write S for the vector of boson counts $(s_1, s_2, \ldots, s_m)$ and x for the vector of formal variables $(x_1, x_2, \ldots, x_m)$.*
- *Write x^S as shorthand for the monomial $x_1^{s_1} x_2^{s_2} \cdots x_m^{s_m}$*
- *Write $p(x)$ for a multivariate polynomial $p(x_1, x_2, \ldots, x_m)$.*

[1] An alternative is for them to anti-commute, which is the case for fermions. The negation of the amplitude after a swap leaves the probability unchanged.

- *Finally, write $S!$ for the product of factorials $s_1!s_2!\cdots s_m!$ (we will use this soon).*

We will use n for the number of bosons and m as the number of modes throughout.[2] Each monomial corresponds to an unordered selection of n elements from among m choices, allowing repeats. The number of such choices is defined by the multichoose operation

$$M = \left(\binom{m}{n}\right) = \binom{m+n-1}{n}.$$

This value M is the dimension of the vectorspace of n-boson states $\mathbb{C}^M$.

We define an inner product on monomials.

Definition 7. *The inner product of two basis states corresponding to counts $S = (s_1, s_2, \ldots, s_m)$ and $T = (t_1, t_2, \ldots, t_m)$ equals*

$$\left\langle x^S \mid x^T \right\rangle = \begin{cases} S!, & \text{if } S = T \\ 0, & \text{otherwise,} \end{cases}$$

recalling that $S!$ is shorthand for $s_1!s_2!\cdots s_m!$.

Note that though these basis states are orthogonal, they are not always of norm 1. The norm-squared of $S!$ is greater than 1 when any count has $s_j > 1$, or in other words when two bosons occupy the same mode. The reason is combinatorial, that $\left\langle x^S \mid x^T \right\rangle$ represents the number of ways to match equal variables in S and T. When a variable appears $s_j = t_j$ times, there are $s_j!$ ways to perform this matching. In other words, the redundant ways of listing two bosons in the same mode contribute multiple times.

A physical n-boson state with determined boson counts S therefore must equal $x^S/\sqrt{S!}$ so that its norm is 1. For instance, a state of two bosons both in mode 1 is not x_1^2 but $x_1^2/\sqrt{2}$. Nevertheless, it turns out simpler simpler to work with monomials in the unnormalized basis and to remember to include the multiplier when taking inner products.

[2] A general state may be a superposition of different number of bosons, such as for a coherent state, but we will only work with states where the number of bosons is fixed.

A general state may be written a polynomial of degree n. This is known as a *Fock state*. We express is as a linear combination of monomials representing a superposition over basis states with respective coefficients c_S:

$$p = \sum_S c_S x^S.$$

This sum is implicitly taken over all $M = \left(\binom{m}{n}\right)$ partitions S of m non-negative integers whose sum is n. The inner product of two Fock states linearly extends that on monomials. Because it is a Hermitian inner product, the complex coefficients of the first state are conjugated.

Definition 8. *The inner product of two Fock states*

$$p = \sum_S c_S x^S$$

and

$$q = \sum_S d_S x^S$$

is given by

$$\langle p \,|\, q \rangle = \sum_S S! \overline{c_S} d_S.$$

The Fock norm *is therefore*

$$\|p\| = \sqrt{\langle p \,|\, p \rangle} = \sqrt{\sum_S S! \, |c_S|^2}.$$

We measure a Fock state by performing a boson-counting measurement on each mode to obtain a vector of counts, which projects the state into the monomial basis. The probability of measuring a given count is determined by the inner product with a normalized basis state $x^T/\sqrt{T!}$.

Definition 9. *Measuring a Fock state $p = \sum_S c_S x^S$ results in a random count vector*

T with respective probabilities

$$\Pr[T] = \left| \left\langle \frac{x^T}{\sqrt{T!}} \middle| p \right\rangle \right|^2 = \left| \frac{1}{\sqrt{T!}} T! c_T \right|^2 = T! \left| c_T \right|^2 .$$

1.2.3 Transformations on multiple identical bosons

We previously defined how a unitary transformation U acts on a single-boson state v by matrix-vector multiplication Uv. We now extend this to multiboson states.

The matrix U transforms a single-boson basis state x_j into a linear combination $\sum_{k=1}^{m} U_{jk} x_k$. We can think of it as the matrix U acting on the vector of formal variables $x = (x_1, x_2, \ldots, x_m)$ to produce a vector Ux of linear combinations of these formal variables. Its action on multiple bosons is then to perform this variable substitution on every variable in the expression. In other words, it transforms the polynomial $p(x)$ into $p(Ux)$.

Definition 10. *Write $U[p]$ for the action of a matrix U on a Fock state p. Then, the action of U on a basis state x^S is given by the product of linear terms*

$$\begin{aligned} U\left[x^S\right] &= (Ux_1)^{s_1} \cdots (Ux_m)^{s_m} \\ &= (U_{11}x_1 + U_{12}x_2 + \cdots + U_{1m}x_m)^{s_1} \cdots (U_{m1}x_1 + U_{m2}x_2 + \cdots + U_{mm}x_m)^{s_m} \end{aligned} \tag{1.1}$$

This extends linearly to its action on a polynomial, which may also be expressed as $U[p(x)] = p(Ux)$.

The product of linear terms may be expanded as a sum of monomials. In this way, the action U on a single-particle space $\mathbb{C}^m$ defines an action[3] on the n-particle space $\mathbb{C}^M$. Note that this is a homomorphism, in that acting by U then by V is the same as acting via the product VU:

$$V[U[p]] = (VU)[p] .$$

[3] In representation theory, this is a representation of the unitary group given by its n^{th} symmetric power.

The noninteractingness of the bosons is reflected in the fact that the unitary map acts on each formal variable independently. Moreover, because the bosons are indistinguishable, the same map must be applied to each one. If the bosons started in a monomial basis state, a transformation can only take them to a product of linear terms, the analogue of a "separable state", which can be defined via m^2 parameters. Moreover, these linear terms remain orthogonal as vectors. This is only a small subset of the space $\mathbb{C}^M$ of n-boson states. In this way, the homomorphism defines a limited way to act on $\mathbb{C}^M$ via a low-dimensional subgroup.

1.2.4 Multiboson systems

We summarize our formalization of systems of identical bosons.

- **State:** A quantum state of n bosons is a vector v in $\mathbb{C}^M$, where $M = \binom{m}{n} = \binom{m+n-1}{n}$. Its basis elements are monomials x^S, where x^S as shorthand for the monomial $x_1^{s_1} x_2^{s_2} \cdots x_m^{s_m}$. The s_j are natural numbers whose sum is n; they represent the count of bosons in each mode. A state is a linear combination $p = \sum_S c_S x^S$ with complex coefficients c_S.

- **Normalization:** A state p must be normalized by having $\langle p \,|\, p \rangle = \sum_S S! \, |c_S|^2 = 1$, where $S! = s_1! \cdots s_m!$.

- **Transformation:** A state p can be acted by an $m \times m$ unitary map on the single-particle space $\mathbb{C}^m$. The resulting state $U[p]$ may be obtained by acting on the formal variables, that is, replacing each variable x_j with the linear combination $\sum_{k=1}^m U_{jk} x_k$.

- **Measurement:** Measuring the state $p = \sum_S c_S x^S$ results in a random boson count vector S with respective probabilities

$$\Pr[S] = |c_S|^2 \, S!.$$

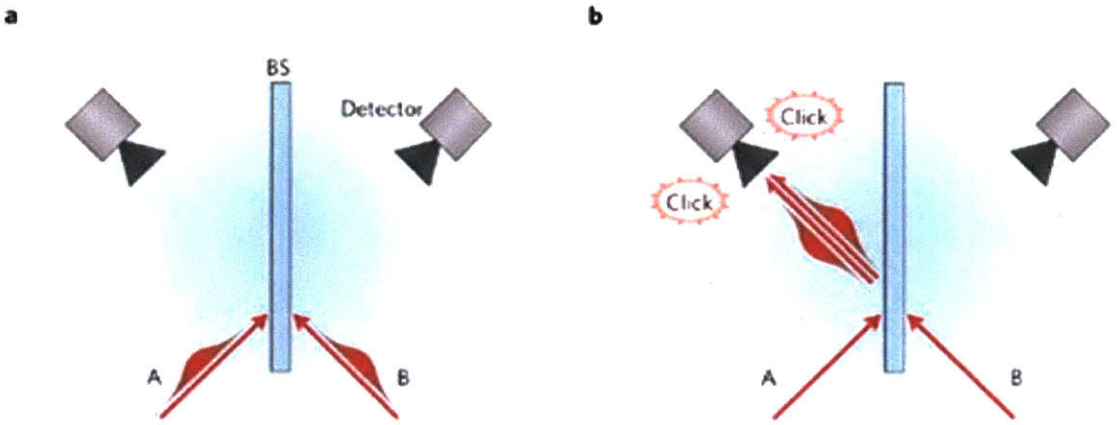

Figure 1-1: A graphical schematic of the Hong-Ou-Mandel dip. Image from *Nature Photonics*

1.2.5 Example: Hong-Ou-Mandel dip

Now would be a good time to see an example of a calculation with Fock states.

The Hong-Ou-Mandel dip [23] is an experimentally-observed two-photon interference phenomenon. Two identical photons enter a 50:50 beamsplitter from opposite sides. Individually, each photon would be equally likely to pass through the beamsplitter as to bounce off it, doing so in superposition. When the two identical photons hit the beamsplitter at the same time, they are always observed to exit on the same side. See the diagrams in fig. 1-1. Even without explicit coupling between the photons, their outcomes managed to become perfectly correlated.

This is a system with $n = 2$ photons and $m = 2$ modes. The initial state is $p = x_1 x_2$, with one photon coming from each side. The 50:50 beamsplitter is a 2×2 Hadamard matrix

$$M = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Acting on an individual photon, this beamsplitter maps x_1 to $\frac{1}{\sqrt{2}}(x_1 + x_2)$, which if measured gives a $1/2$ chance of the photon appearing in each mode. These potential outcomes correspond to counts of $(1, 0)$ and $(0, 1)$. The same is true for the other photon, which maps to $\frac{1}{\sqrt{2}}(x_1 - x_2)$. Now, the action of the beamsplitter on both

photons is

$$\begin{aligned} U[x_1 x_2] &= U[x_1]\, U[x_2] \\ &= \frac{1}{\sqrt{2}}(x_1 + x_2)\frac{1}{\sqrt{2}}(x_1 - x_2) \\ &= \frac{1}{2}\left(x_1^2 - x_2^2\right) \end{aligned} \tag{1.2}$$

Note how in the product, the $-x_1x_2$ and $+x_2x_1$ terms canceled, leading to no term x_1x_2 for the output $(1,1)$ of one photon in each mode. This gives zero probability for the output bosons going different ways. The commutativity of the product was needed for this cancellation to happen.

Measuring afterward gives one of the normalized basis states $x_1^2/\sqrt{2}$ and $x_2^2/\sqrt{2}$. The probability of getting $x_1^2/\sqrt{2}$ for a count vector of $(2,0)$ is $\left|\left\langle \frac{1}{2}x_1^2 \,\middle|\, x_1^2/\sqrt{2}\right\rangle\right|^2 = \left|1/2 \cdot 2 \cdot \sqrt{2}\right|^2 = 1/2$. We similarly get probability $1/2$ for $(0,2)$. These add to 1 as expected.

1.2.6 Analogues with other particles

How does the formalization for systems of multiple bosons summarized in section 1.2.4 change when other kinds of particle take the place of bosons? Boson and fermions are the two classes of elementary particles. Bosons are particles like photons that have integer spin and generally carry forces, whereas fermions are particles that have half-integer spin and include electrons, protons, and neutrons.

The model for fermions is similar except the commutation property $x_jx_k = x_kx_j$ is replaced by anti-commutation $x_jx_k = -x_kx_j$. This type of product is known as the *alternating product* or *wedge product* or *exterior product*, often written $\wedge$. Note that the negative sign appears only in the amplitude. When squaring to get the probability, it disappears. But, the sign matters for relative terms. For example, re-doing the Hong-Ou-Mandel dip in section 1.2.5 with fermions, the terms $-x_1x_2$ and $+x_2x_1$ are not opposite but rather equal, and so they combine constructively for a nonzero probability of the particles splitting ways. In fact, the probability is 1; the particles cannot end in the same mode because the anti-commutation law gives

$x_j^2 = -x_j^2$, which must equal 0. This is the *Pauli exclusion principle*, that two or more identical fermions cannot occupy the same quantum state.

When the particles are classical, such as macroscopic objects, we can developed a similar formalism using probabilities instead of amplitudes. States are now classical mixtures of basis states x^S. Their coefficients represent probabilities, not amplitudes, and measuring uses these coefficients directly without squaring. As probabilities, the coefficients are real numbers between 0 and 1 that add to 1. Transformations still act as linear variable replacements, but are now stochastic matrices so as to preserve the L^1 norm.

In the classical case, nothing changes if the particles are distinguishable, but we simply ignore their differences by only considering the number in each mode after we measure. If we work with distinguishable bosons or fermions, the outcomes are just as if they were classical particles. The amplitudes directly translate to probabilities via their norm-squares, and we may work with these probabilities directly as there is never any superposition that would add two amplitudes. If identical particles enter and exit the system at different times, then it is determined which output particle corresponds to which input, making the particles distinguishable. This also occurs to varying degrees at each optical component if the wavepackets of the photons are shifted in time and so do not overlap perfectly. In [57], Xu develops an intermediate model for partially distinguishable bosons.

1.2.7 Adjoint property

It is not immediately obvious why a transformation $U[p]$ with U unitary is itself a unitary map on $\mathbb{C}^M$. That it is unitary implies that the transformation takes orthogonal states to orthogonal states, and that it preserves that states have a norm of 1. This must be true for the total probability of all possible measurements to sum to 1, and so is implicitly required by the physical interpretation. We prove the following more general property:

Proposition 11. *For any n-boson states p and q and matrix M, the following Fock*

inner products are equal:

$$\langle p \,|\, M[q] \rangle = \left\langle M^\dagger[p] \,\middle|\, q \right\rangle,$$

where $M^\dagger$ is the conjugate-transpose of M.

This is analogous to the familiar property of the inner product that $\langle v \,|\, Mw \rangle = \left\langle M^\dagger v \,\middle|\, w \right\rangle$. The conservation of norm and orthogonality follows from this because

$$\langle U[p] \,|\, U[q] \rangle = \left\langle U^\dagger U[p] \,\middle|\, q \right\rangle = \langle p \,|\, q \rangle,$$

using that $U^\dagger U$ is the identity for U unitary.

We prove proposition 11 by taking a detour to an alternate definition of the Fock inner product as an expectation over Gaussian distributions. This expression treats the formal variables as taking on complex-number values.

Proposition 12. *Let $\mathcal{G}$ be the complex Gaussian distribution $\mathcal{N}(0,1)_{\mathbb{C}}$. The Fock inner product of p and q is given by*

$$\langle p \,|\, q \rangle = \mathbb{E}_{x \sim \mathcal{G}^m} \left[\overline{p(x)} q(x) \right]$$

Proof. Because expectations and inner products are both linear, it suffices to prove the result when p and q are monomials $p = x^S$ and $q = x^T$. The product $\overline{p(x)}q(x)$ is then

$$\overline{p(x)}q(x) = \overline{x^S} x^T = \prod_{j=1}^{m} \overline{x_j^{s_j}} x_j^{t_j}$$

We first show that this has expectation zero when $S \neq T$, matching the inner product. Since the coordinates of $\mathcal{G}^m$ are probabilistically independent, the expectation decomposes into factors as

$$\mathbb{E}_{x \sim \mathcal{G}^m} \left[\overline{p(x)} q(x) \right] = \prod_{j=1}^{m} \mathbb{E}_{x \sim \mathcal{G}} \left[\overline{x_j^{s_j}} x_j^{t_j} \right].$$

Consider one of these terms $\mathbb{E}_{x \sim \mathcal{G}} \left[\overline{x}^a x^b \right]$. We can take out a multiplier of $\overline{x}^a x^a$ to

rewrite this as

$$\mathbb{E}_{x\sim\mathcal{G}} |x|^{2a} x^{b-a}.$$

When $a \neq b$, the distribution of x^{b-a} is radially symmetric, i.e. uniform over phases, so the expectation is 0. Therefore, unless $S = T$ and all the respective counts match, the expectation is 0.

It remains to check that the expectation matches the inner product when $S = T$. We use that

$$\mathbb{E}_{x_j\sim\mathcal{G}} \left[|x|^{2a} = a!\right],$$

which may be obtained from the moment generating function of the χ^2 distribution of $|x|^2$. So, the product of the expectations is $\prod_{j=1}^{m} s_j! = S!$, matching the inner product.

□

We're now ready to prove the adjoint property (proposition 11).

Proof. We want to show that

$$\langle p \,|\, M[q]\rangle = \left\langle M^\dagger[p] \,\middle|\, q\right\rangle.$$

We first show that this holds when M is a unitary matrix U . Rewriting the claim

$$\langle p \,|\, U[q]\rangle = \left\langle U^\dagger[p] \,\middle|\, q\right\rangle$$

as a Gaussian expectation as per proposition 12, we get

$$\mathbb{E}_{x\sim\mathcal{G}^m} \left[\overline{p(x)} q(Ux)\right] = \mathbb{E}_{x\sim\mathcal{G}^m} \left[\overline{p(U^\dagger x)} q(x)\right]. \tag{1.3}$$

We've used here that U acts by variable substitution $U[p(x)] = p(Ux)$. Since the vector Gaussian distribution $\mathcal{G}^m$ is unitarily invariant, we may perform a unitary change of variables $y = Ux$, and equvialently $x = U^\dagger y$, which transforms the left side of the equality into the right side.

Next we show that the claim holds for any diagonal matrix D: $\langle p \,|\, D[q] \rangle = \langle D^\dagger[p] \,|\, q \rangle$. As before, it suffices by linearity to confirm this when p and q are monomials. Note that $D\left[x^S\right]$ multiplies each variable in the monomial by a constant, resulting in the scalar multiple cx^S where $c = d_1^{s_1} d_2^{s_2} \ldots d_m^{s_m}$. So, both sides are 0 unless $p = q$. When $p = q$, we have

$$\langle p \,|\, D[p] \rangle = \langle p \,|\, cp \rangle = c \langle p \,|\, p \rangle ,$$

which turns out to equal

$$\langle D^\dagger[p] \,|\, p \rangle = \langle \overline{c} p \,|\, p \rangle = c \langle p \,|\, p \rangle .$$

Having shown the claim for unitary and diagonal matrices, we now extend it to an arbitrary matrix M. We express M as a product of unitary and diagonal matrices via its singular value decomposition $M = U^\dagger D V$. This lets us disassemble M and pass it from the right side of the $|$ to the left side layer by layer, after which it is reassassembled as $M^\dagger$.

$$\begin{aligned} \langle p \,|\, M[q] \rangle &= \left\langle p \,\middle|\, \left(U^\dagger D V\right)[q] \right\rangle \\ &= \langle U[p] \,|\, (DV)[q] \rangle \\ &= \left\langle D^\dagger U[p] \,\middle|\, V[q] \right\rangle \\ &= \left\langle V^\dagger D^\dagger U[p] \,\middle|\, q \right\rangle \\ &= \left\langle M^\dagger[p] \,\middle|\, q \right\rangle \end{aligned} \tag{1.4}$$

□

1.2.8 Connection to the permanent

We now reformulate identical boson systems in terms of matrix permanents rather than Fock polymomials, in order to facilitate arguments about their computational complexity. The matrix permanent is a function on a square matrix.

Definition 13. *The* permanent *of an $n \times n$ matrix is*

$$\mathrm{Per}(M) = \sum_{\sigma \in S_n} \prod_{j=1}^{n} M_{j,\sigma(j)},$$

where S_n is the set of all permutations of $\{1, 2, \ldots, n\}$.

The $n!$ summands correspond to the $n!$ values to choose n entries entries of the matrix so that every row and column has exactly one entry chosen, and each summand is the product of these entries. The expression is very similar to that of the determinant, but it does not have coefficients of ± 1 depending on the sign of the permutation.

Definition 14. *The* determinant *of an $n \times n$ matrix is*

$$\mathrm{Det}(M) = \sum_{\sigma \in S_n} \mathrm{sign}(\sigma) \prod_{i=j}^{n} M_{j,\sigma(j)},$$

where S_n is the set of all permutations of $\{1, 2, \ldots, n\}$, and $\mathrm{sign}\,\sigma$ equals $+1$ for even permutations (i.e. those composed of an even number of swaps) and -1 for odd permutations.

To understand the connection between permanents and Fock state probabilities, we start by looking at the case where $m = n$ and the initial state has one boson in each mode.

Proposition 15. *Let $m = n$ and let $R = (1, 1, \cdots, 1)$ be the initial state of one boson in each mode, giving $x^R = x_1 x_2 \cdots x_n$. Then, after applying the operation M to obtain $M[x^R]$, the probability of getting a measurement result that is once again R is*

$$\Pr[R \to R] = \left| \left\langle x^R \,\middle|\, M\left[x^R\right] \right\rangle \right|^2 = \left|\mathrm{Per}(M)\right|^2$$

Proof. The transformed state $M[x^R]$ is a product of linear terms

$$M[x^R] = \prod_{j=1}^{n} \left(M_{j,1} x_1 + M_{j,2} x_2 + \cdots + M_{j,n} x_n \right).$$

We can expand this into a sum of monomial terms, each one obtained by choosing for each index j the $\sigma(j)^{\text{th}}$ summand $M_{j,\sigma(j)}x_{\sigma(j)}$. So, the product expands into

$$\sum_{\sigma}\prod_{j=1}^{n} M_{j,\sigma(j)}x_j,$$

where σ ranges among all functions from $\{1, 2, \ldots, n\}$ to itself.

The inner product $\langle x^R \,|\, M\left[x^R\right]\rangle$ extracts the coefficient of the monomial $x^R = x_1x_2\ldots x_n$. This monomial is formed exactly when the $x_{\sigma(j)}$ form a permutation of $x_1, x_2, \ldots, x_n$, which is exactly when σ is a permutation. This gives exactly the expression for the permanent

$$\sum_{\sigma\in S_n}\prod_{j=1}^{n} M_{j,\sigma(j)} = \mathrm{Per}(M).$$

The probability of the outcome is the norm-squared of the amplitude, or $|\mathrm{Per}(M)|^2$.

□

Let's think more about how the permanent came about in the expression for the probability. Each permutation is a matching of the input bosons to the output bosons. Because the bosons are identical, one cannot tell which boson in the initial state corresponds to which boson in the final state, and all $n!$ permutations are potentially possible. The amplitude corresponding to a given path is the product of the corresponding transition amplitudes from each photon's input mode to its output mode. Since the particles do not interact, these transitions are independent events, and we obtain the amplitude for the joint event by multiplying the respective amplitudes to obtain $\prod_{j=1}^{n} M_{j,\sigma(j)}$. The total amplitude is then the sum of amplitudes of all matchings that lead to a given outcome, which gives the permanent.

We will now generalize this expression to an initial state with counts T and the probability of a measurement obtaining counts S. If these counts are collision-free, i.e. are all 0 or 1, it's not hard to see that the result will be like proposition 15 but limited to occupied rows for inputs and columns for outputs. Entries in columns for

modes not containing an initial boson do not figure in the probability, and likewise for rows that do not correspond to a measured boson. We may omit these rows and columns by taking a submatrix. We will need to generalize submatrices to account for more than one boson in a mode of the input or output.

Definition 16. *Let M be an $m \times m$ matrix and let S, T be count vectors that each sum to n. The* generalized submatrix $M_{S,T}$ *is an $n \times n$ matrix obtained by taking S_j copies of each row S and T_j copies of each column T.*

For example, let

$$M = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

and $S = (2, 0, 2)$, $T = (3, 1, 0)$. Here, $m = 3$ and $n = 4$. Then, $M_{S,T}$ is the 4×4 matrix

$$M = \begin{bmatrix} 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 2 \\ 7 & 7 & 7 & 8 \\ 7 & 7 & 7 & 8 \end{bmatrix}.$$

We may obtain a generalized submatrix by applying a linear transformation to the original matrix.

Proposition 17. *The $n \times n$ generalized submatrix $M_{S,T}$ of an $n \times n$ matrix M is given by*

$$M_{S,T} = I_S^\dagger M I_T,$$

where I_T is defined for a count vector T as an $m \times n$ matrix whose columns are unit vectors with x_j appearing t_j times.[4]

Proof. Observe that for a row vector $v^\dagger$, the right-product $v^\dagger I_T$ creates a column vector with t_j copies of each entry v_j. So, $M I_T$ acts the same way duplicating columns

[4]This only defines I_T up to a permutation of its columns, which will suffice for our needs. If we wished to be concrete, we could say the columns are sorted by increasing j.

to the number required for the generalized submatrix. $I_S^\dagger$ then does likewise on the rows of MI_T with counts given by S to create $M_{S,T}$. □

Now we'll generalize proposition 15 using generalized submatrices.

Proposition 18. *If we begin with initial boson counts T, transform by the matrix M, and measure, the probability of obtaining counts S is*

$$\Pr[T \to S] = \frac{1}{S!T!} \left|\mathrm{Per}\left(M_{S,T}\right)\right|^2.$$

Proof. By proposition 17, we can express $\mathrm{Per}\left(M_{S,T}\right) = \mathrm{Per}\left(I_S^\dagger M I_T\right)$. Then, by proposition 15,

$$\mathrm{Per}\left(I_S^\dagger M I_T\right) = \left\langle x^R \,\middle|\, \left(I_S^\dagger M I_T\right)\left[x^R\right]\right\rangle,$$

where $x^R = x_1 x_2 \cdots x_n$. We can apply the adjoint property of proposition 11 to move the $I_S^\dagger$ to the other side:

$$\mathrm{Per}\left(M_{S,T}\right) = \left\langle I_S\left[x^R\right] \,\middle|\, M I_T\left[x^R\right]\right\rangle.$$

Now observe that $I_T\left[x^R\right]$ creates t_1 copies of x_1 in the product, t_2 copies of x_2 and so on, to create x^T. So, $I_T\left[x^R\right] = x^T$ and $I_S\left[x^R\right] = x^S$. This turns the expression into

$$\mathrm{Per}\left(M_{S,T}\right) = \left\langle x^S \,\middle|\, M\left[x^T\right]\right\rangle.$$

Recall once more how to compute $\Pr[T \to S]$. The initial normalized state is $x^T/\sqrt{T!}$, which is transformed to $M[x^T]/\sqrt{T!}$. The probability of measuring S is the norm-squared of the dot product of this with the normalized state $x^S/\sqrt{T!}$, which gives

$$\Pr[T \to S] = \frac{1}{S!T!}\left|\left\langle x^S \,\middle|\, M\left[x^T\right]\right\rangle\right|^2.$$

Using the equality $\mathrm{Per}\left(M_{S,T}\right) = \left\langle x^S \,\middle|\, M\left[x^T\right]\right\rangle$ from before, this gives

$$\Pr[T \to S] = \frac{1}{S!T!}\left|\mathrm{Per}\left(M_{S,T}\right)\right|^2$$

as claimed. □

1.2.9 Physically implementing identical bosonic systems

We'll talk briefly here about the physical implementation of systems of noninteracting identical bosons. See chapter 2 for more on experimental implementations.

Using photons as bosons, such a system can be modeled by a linear-optical network with Fock initial states and photon-counting measurements. The network consists of two types of components, *phaseshifters* and *beamsplitters*, which are optical gates that act locally on 1 or 2 modes. It is easiest to think of their action as a unitary matrix a single photon, which we've shown to determine the unitary action on n identical photons.

A phaseshifter acts on a single mode and applies a phase $e^{i\theta}$, leaving the remaining modes unaffected. A beamsplitter acts on two modes, transforming their amplitudes α_S and α_T by a real rotation matrix for some angle θ:

$$\begin{pmatrix} \alpha'_S \\ \alpha'_T \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \alpha_S \\ \alpha_T \end{pmatrix}. \tag{1.5}$$

Reck et al. [37] prove beamsplitters and phaseshifters to be a universal get set in that any $m \times m$ unitary operation U can be efficiently implemented as a product of them.

Lemma 19. *Any $m \times m$ unitary matrix U can be decomposed into a product of $O(m^2)$ operations each of which is either a beamsplitter or a phaseshifter, i.e. a matrix that is everywhere the identity except for either a single unit phase or a 2×2 real orthogonal block matrix on the diagonal. This decomposition can be found in polynomial time.*

1.3 Exact BosonSampling

1.3.1 Definition of BosonSampling

We define the computational problem of BosonSampling based on our model of systems of noninteracting identical bosons. We start by defining the system in terms of a one-step unitary operation followed by a measurement.

Definition 20. *A* BosonSampling *system is a quantum system implementing the following procedure:*

1. *The initial state consists of n bosons, one in each of the first n modes of m total modes.*
2. *An $m \times n$ column-unitary matrix M is applied.*
3. *The final state is measured to produce a boson count for each of the m output modes.*

For concreteness, we have chosen an initial state $x^T = x_1 x_2 \dots x_n$ with the first n modes occupied by one boson and the remaining $m - n$ empty. That is, $T = (1, 1, \dots, 1, 1, 0, 0, \dots, 0, 0)$ with n ones. As a result, the generalized submatrix uses only the first n columns of the transformation matrix, and the rest are immaterial. We call this $m \times n$ matrix the *network matrix*. As a subset of the columns of a unitary matrix, this matrix M is column-unitary, meaning that $M^\dagger M$ is the identity.

By proposition 18 with T as described, the probability that the BosonSampling system outputs S is given by

$$\Pr[S] = \frac{1}{S!} \left|\mathrm{Per}(M_S)\right|^2 .$$

Here, M_S is a generalized submatrix with rows given by S; the selection of columns was already handled by limiting M to the first n columns.

We define the BosonSampling distribution as the distribution of outputs of a BosonSampling device for a given network matrix M.

Definition 21. *For an $m \times n$ column-unitary matrix M, the* BOSONSAMPLING *distribution $\mathcal{D}_M$ is a distribution over length-m count vectors S of natural numbers that add to n, with*

$$\Pr[S] = \frac{1}{S!} |\mathrm{Per}(M_S)|^2 .$$

Note that though the BOSONSAMPLING distribution is designed to emulate the output distribution of a specific physical system, its definition above is purely mathematical and makes no reference to particles or quantum mechanics. In this way, it can be stated as a black-box computational problem solely in terms of its input and (randomized) output. An algorithm solves this problem by having the correct probabilistic input-output behavior without necessarily simulating the internal behavior of a BOSONSAMPLING system.

Definition 22. *The* BOSONSAMPLING problem *is: given an $m \times n$ column-unitary matrix M, output a random sample of the distribution $\mathcal{D}_M$.*

Let us emphasize here that this is a sampling problem, which means it is solved by a random algorithm whose output probabilities match those of the desired distribution. It is not the problem of computing the permanent of a matrix. Nor is it a decision problem to determine whether or not the input meets some condition, as is typical in complexity theory.

1.3.2 Complexity of exact BOSONSAMPLING

First, note that BOSONSAMPLING can be efficiently performed by a quantum computer. Indeed, it is stated as the behavior of a quantum system of identical bosons, which can be simulated on a standard quantum computer. We define SampBQP as a sampling analogue to BQP (Bounded-Error Quantum Polynomial Time) in which the final measurement produces a sample from a distribution. We require its distribution to be within ϵ of the desired distribution in variation distance, with $0^{1/\epsilon}$ given as an additional input.

Proposition 23. *The problem of* BOSONSAMPLING *is in* SampBQP, *the sampling version of* BQP.

This is implied by simulation of bosonic systems on a qubit-based computer as developed by Feynman [18] and by Abrams and Lloyd [4]. Each boson count is encoded in $\lg n$ qubits. Each 2-mode linear-optical element from the decomposition in [37] is translated to an operation on the corresponding qubits.

Our first main result is the inability of classical computers to efficiently simulate BOSONSAMPLING exactly unless PH collapses. Note that this result does not say anything directly about the computational consequences of a black-box that does BOSONSAMPLING, but rather the consequences of an efficient classical algorithm for it whose random bits we get to control.

Theorem 24 (Complexity of exact BOSONSAMPLING). *No polynomial-time classical randomized algorithm can perform* BOSONSAMPLING, *unless* $\mathsf{P}^{\#\mathsf{P}} = \mathsf{BPP}^{\mathsf{NP}}$ *and the polynomial hierarchy collapses to the third level.*

We gather a few ingredients to use in the proof.

1.3.3 Proof ingredient: Complexity of the permanent

First, we note a celebrated result on the complexity of the permanent.

Theorem 25 (Valiant [53]). *The problem of computing the permanent of an* $n \times n$ *matrix with entries in* $\{0, 1\}$ *is* #P*-complete.*

The class #P is complete for *counting problems*, such as counting how many variable assignments satisfy a Boolean circuit. It is strongly conjectured that such a problem cannot be solved in polynomial time. For one, this would imply that P equals NP. Directly computing the permanent requires summing over $n!$ permutations, an exponential amount. Ryser's formula (stated in the proof of theorem 57) is the best-known classical algorithm for the permanent, using $O(n2^n)$ time with the optimization of processing the summands in Gray code order. This improves over $n!$, but is still of course exponential.

We will need a variant that says that estimating a real permanent-squared is #P-hard.

Theorem 26 (Theorem 28 in [1]). *The following problem is #P-hard: Given a real $n \times n$ matrix M and a multiplicative factor c with $1 \leq c \leq \mathrm{poly}(n)$, compute $|\mathrm{Per}(M)|^2$ to within a multiplicative factor of c.*

We will leverage the close connection between identical bosons outcomes and the complexity of computing the permanent. Keep in mind the BOSONSAMPLING does not allow us to compute permanents directly, which would imply that $\mathsf{BQP} = \mathsf{P}^{\#\mathsf{P}}$, but to sample a distribution described by them. So, we will need a less direct way to derive consequences from a classical simulation. We begin with a way to embed a matrix whose permanent we wish to compute as a submatrix of a unitary network matrix.

Lemma 27. *Let $X \in \mathbb{C}^{n \times n}$. Then for all $m \geq 2n$ and $\varepsilon \leq 1/\|X\|$, there exists an $m \times m$ unitary matrix U that contains εX as a submatrix. Furthermore, U can be computed in polynomial time given X.*

Proof. Let $Y = \varepsilon X$. Then it suffices to show how to construct a $2n \times n$ matrix W whose columns are orthonormal vectors, and that contains Y as its top $n \times n$ submatrix. For such a W can easily be completed to an $m \times n$ matrix whose columns are orthonormal (by filling the bottom $m - 2n$ rows with zeroes), which can in turn be completed to an $m \times m$ unitary matrix in $O(m^3)$ time.

Since $\|Y\| \leq \varepsilon \|X\| \leq 1$, we have $Y^\dagger Y \preceq I$ in the semidefinite ordering. Hence $I - Y^\dagger Y$ is positive semidefinite. So $I - Y^\dagger Y$ has a Cholesky decomposition $I - Y^\dagger Y = Z^\dagger Z$, for some $Z \in \mathbb{C}^{n \times n}$. Let us set $W = \begin{pmatrix} Y \\ \hline Z \end{pmatrix}$. Then $W^\dagger W = Y^\dagger Y + Z^\dagger Z = I$, so the columns of W are orthonormal as desired. □

1.3.4 Proof ingredient: Approximate counting

To show the difficulty of sampling problem, we will need to take a classical circuit that solve the sampling problem and harness it to solve the function problem of estimating the permanent. We use a result of Stockmeyer [46], which introduces the technique of *universal hashing* to take a classical randomized circuit and estimate its acceptance probability with the help of an NP oracle.

Theorem 28 (Stockmeyer [46]). *Let f be a Boolean function from $\{0,1\}^n$ to $\{0,1\}$ and let p be its probability of accepting:*

$$p = \Pr_{x \in \{0,1\}^n} [f(x) = 1] .$$

For any parameter $c \geq 1 + 1/\operatorname{poly}(n)$, there is a $\mathsf{BPP}^{\mathsf{NP}^f}$ machine to approximate p to within a multiplicative factor of g.

So, the machine has access to an NP oracle which itself has oracle access to the function f that it can use deterministically. The NP oracle takes advantage of being able to set the random bits of f. This would not be possible if the circuit f were simply a randomized black box. Nor would it be possible to use a quantum circuit, which does not have "random bits" to set – its randomness is produced on the fly, not read from a pre-computed random tape. In this way, the result derives consequences of the existence of a classical circuit but not a quantum circuit to compute the same thing.

Also note that the probability p of acceptance may be exponentially small. In this situation, one could not simply test the function with many random seeds to estimate the acceptance probability, as chance are that none of the trials would accept. The non-determinism of the NP oracle is used to help find these "needles in the haystack".

1.3.5 Proof of exact hardness

Now we're ready to prove the exact result, which we restate for convenience.

Theorem 29 (Complexity of exact BosonSampling). *No polynomial-time classical randomized algorithm can perform* BosonSampling, *unless $\mathsf{P}^{\#\mathsf{P}} = \mathsf{BPP}^{\mathsf{NP}}$ and the polynomial hierarchy collapses to the third level.*

Proof. Recalling the #P-hard problem from theorem 26, let M be an $n \times n$ real matrix and c a multiplicative factor within which we wish to estimate $|\mathrm{Per}(M)|^2$. Use the embedding of lemma 27 to find a $2n \times n$ column-unitary matrix U whose upper half is ϵM.

Suppose to the contrary that A is a classical randomized algorithm that performs BosonSampling, and consider running it on U to sample the distribution $\mathcal{D}_U$. We consider this algorithm to accept if it produces an output corresponding to M, that is a count of one boson in each of the first n of the $2n$ modes. The probability of this is given by

$$p = |\mathrm{Per}\,(\epsilon M)|^2 = \epsilon^{2n} |\mathrm{Per}\,(M)|^2 .$$

Now, apply the approximate counting result theorem 28 to this classical circuit to estimate the acceptance probability p. In $\mathsf{BPP}^{\mathsf{NP}^A}$, which equals $\mathsf{BPP}^{\mathsf{NP}}$, we estimate the acceptance probability to within a multiplicative factor of $c \geq 1 + 1/\mathrm{poly}(n)$. Dividing by the ϵ^{2n}, we obtain $|\mathrm{Per}\,(M)|^2$ to within a multiplicative factor that easily falls within the required accuracy of theorem 26 to be #P-hard.

So, we've shown that

$$\mathsf{P}^{\#\mathsf{P}} \subseteq \mathsf{BPP}^{\mathsf{NP}}.$$

Toda's Theorem [52] bounds the polynomial hierarchy $\mathsf{PH} \subseteq \mathsf{P}^{\#\mathsf{P}}$, so this would imply $\mathsf{PH} = \mathsf{BPP}^{\mathsf{NP}}$. The Sipser-Lautemann theorem [44] tells us that $\mathsf{BPP} \subseteq \Sigma_2 \cap \Pi_2$, so this is a collapse to the third level.

□

1.3.6 Complexity comparison to other particles

Why bosons for BosonSampling? We show that analogous the problem with fermions or classical particles in place of bosons, as discussed in section 1.2.6, has an efficient classical simulation. So, the hardness result of theorem 24 no longer holds for these particles.

FermionSampling can be defined the same way BosonSampling, except with the determinant taking the role of the permanent. However, the determinant can be computed in polynomial time, for example by Gaussian elimination, much unlike the #P-hardness of the permanent. So, probabilities of FermionSampling outcomes are easy to compute. Understanding the gulf in difficulty between the permanent and determinant is a central problem in geometric complexity theory (see for example

[28]).

Moreover, quantum circuits of identical fermions can be sampled in polynomial time. A class of quantum circuits called *matchgate circuits* can be simulated using the determinant [54], and these circuits were later shown equivalent to systems of identical noninteracting fermions [48, 26]. However, if the fermions are allowed to start in entangled state, then BOSONSAMPLING can be simulating fermions [41].

An alternative approach to simulating FERMIONSAMPLING is to randomly choose the outcome one mode at a time, each time conditioning the distribution for the current mode on the already-decided counts of previous modes. The expression for these conditional marginal probabilities can itself be expressed determinant, making use of linear-algebraic properties of the determinant that do not have an analogue for the permanent. Although BOSONSAMPLING does allow the marginal for a fixed number of modes to be efficiently sampled [22], there is no known efficient way to condition on previous results.

An analogue of BOSONSAMPLING with classical particles also allows for a classical simulation. In fact, this simulation is trivial! Just do the procedure exactly as described – put the particles in their initial modes, move each one randomly as per the stochastic transition matrix, and count how many end in each mode. This can also be done with distinguishable particles by treating them as classical. The unitary network matrix acts as a classical stochastic transition matrix each of whose entries is the norm-square of the corresponding amplitude.

Let's reflect on what gives identical bosons the quantum advantage over classical particles. The key difference is how the overall probability is obtained from the multitude of paths by which the output particles could be matched with the input particles. For classical particles, the probabilities of the respective paths are added, and because these are non-negative real numbers, each additional possibility can only increase the overall probability. In contrast, the paths are in quantum superpositions for identical particles, causing their amplitudes to add. These amplitudes are complex numbers, so different paths may cancel to produce a zero or near-zero overall probability. In this situation, small changes to the summands can cause a large relative change in

the final result. Cancellations like this are at the heart of many quantum algorithms, and we see that the complexity of BOSONSAMPLING relies on them.

Moreover, each probability for the classsical analogue is the permanent of a matrix of nonnegative real numbers, which can be estimated classically in probabilistic polynomial time by an algorithm of Jerrum, Sinclair, and Vigoda [24].

1.4 Approximate BOSONSAMPLING

1.4.1 Approximate versus exact

The exact complexity result theorem 24 is weaker than would be ideal. When it talks about a classical algorithm performing BOSONSAMPLING, it requires that it do so exactly[5], giving an output distribution that precisely equals $\mathcal{D}_M$. This is perhaps an unfair requirement because any actual physical BOSONSAMPLING device must itself have some amount of error, which will cause its output distribution to deviate slightly from the required one. As we would like our result to say something about quantum advantage as seen in real-world devices, it would be best for our result to reflect that.

To make the result robust to error, we only require the output distribution of the device A to be close in variation distance $\|\cdot\|_1$ to the true BOSONSAMPLING distribution. That, is that the total absolute difference of probabilities over all outcomes is small:

$$\|A - \mathcal{D}_M\|_1 = \sum_S \left| \Pr_A [S] - \Pr_{\mathcal{D}_M} [S] \right| < \epsilon$$

The value ϵ is an accuracy parameter that is provided as an input to the algorithm. We say the running time is efficient if its is polynomial in $1/\epsilon$, m, and n.

Definition 30. *The* approximate BOSONSAMPLING problem *is: given an $m \times n$ column-unitary transition matrix M and an error tolerance ϵ, sample from a distribution $\mathcal{D}'_M$ that is within ϵ in variation distance to the* BOSONSAMPLING *distribution*

[5]Or rather, as far as allowed by the limits of machines' ability to represent real numbers, using a polynomial number of bits for exponential accuracy.

$\mathcal{D}_M$:

$$\|\mathcal{D}'_M - \mathcal{D}_M\|_1 < \epsilon$$

We can imagine the classical approximate BOSONSAMPLING algorithm to be designed by an adversary to be useless as possible to us while still satisfying the error bound. This could mean shifting each outcome's probability by a tiny relative amount, or changing a small fraction of outcomes to have very different probabilities, or both. Note that because the total number of outcomes $\left(\binom{m}{n}\right)$ is exponentially large, it is allowed to drastically change an exponentially large number of outcomes that still comprise an exponentially small fraction of them.

The proof in the exact result is not enough on its own to prove the impossibility of approximate BOSONSAMPLING. It relies on embedding a single submatrix whose permanent we want to compute and using approximate counting to estimate its probability of being chosen. But, the device could make it output probability on this one outcome be extremely wrong while still keeping the total error exponentially small. It could even just never output it, giving probability 0. This lets the classical simulation satisfy the approximate contract while preventing us from using it to compute the permanent.

1.4.2 Submatrix hiding and Gaussian matrices

This selective sabotage can be defended against by hiding which submatrix we care about. Instead of just embedding it as the top $n \times n$ submatrix of a $m \times n$ matrix, we can make it a random subset of n rows by shuffling all the rows after embedding. The adversary is only given the network M and is not told which submatrix we started with. That way, we hope they don't know which of the exponentially many possible outcomes to sabotage, and they cannot sabotage most of them without exceeding the total error allowance.

The adversary does know the desired outcome has at most one boson per mode, but we will make the information moot by embedding into a matrix of size $m \geq n^{2+\epsilon}$ rather than $m = 2n$. This is a sparse regime where there are many more modes than

bosons. Here, the fraction of collision-free outcomes $\binom{m}{n} / \left(\!\binom{m}{n}\!\right)$ is nearly 1, as shown in [6].

However, the adversary still may be able to figure out what was our original matrix and what was padding by looking for patterns in the entries. They could succeed if, for example, the norms of entries in our embedded matrix are bigger or smaller on average than those in the padding. Or, perhaps we were interested in the permanent of a matrix with a particular structure, making it easy to detect rows from this matrix.

We can camouflage our embedded matrix perfectly by taking it from the distribution of $n \times n$ submatrices of $m \times n$ Haar-random column-unitary matrices, or equivalently of $m \times m$ Haar-random unitary matrices.

Lemma 31. *Let $\mathcal{S}_{m,n}$ be the distribution of $m \times m$ Haar-random unitary matrices truncated to their top $n \times n$ submatrix. Say we choose a matrix X from $S_{m,n}$ and complete to an $m \times n$ column-unitary matrix U as described in lemma 27 with the padding chosen uniformly at random, then randomly permute its rows. Then, an adversary given only the resulting matrix U cannot find the location of the submatrix X with probability significant better than chance $p > 2/\left(\!\binom{m}{n}\!\right)$ for $m \geq n^{2+\epsilon}$.*

This can be extended to be robust to error: if the submatrix is chosen from a distribution that's ϵ-close to $\mathcal{S}_{m,n}$ in variation distance, then the adversary cannot find X with probability $2(1+\epsilon)/\left(\!\binom{m}{n}\!\right)$.

It behooves us to understand what this distribution of submatrices of unitaries is. It is a standard fact from random matrix theory that an entry of a Haar-random $m \times m$ unitary matrix is distributed nearly as a complex Gaussian with variance $1/m$. In the sparse limit $n \ll m$, where the submatrix is such a small fraction of the unitary matrix that the unitarity constraints apply very weakly, the entries are nearly independent and so tend towards i.i.d. Gaussians.

Lemma 32 (Haar-unitary limiting). *Let $\mathcal{G}^{n \times n}$ be the distribution of $n \times n$ matrices of i.i.d. complex Gaussians $\mathcal{N}(0,1)_{\mathbb{C}}$. Let $\sqrt{m}\mathcal{S}_{m,n}$ be the distribution of $n \times n$ truncations of $m \times m$ Haar-random unitary matrices, scaled up by $\sqrt{m}$ to make each entry have*

variance 1. *Then, for* $m \geq \frac{n^5}{\delta}\log^2\left(\frac{n}{\delta}\right)$,

$$\left\|\sqrt{m}\mathcal{S}_{m,n} - \mathcal{G}^{n\times n}\right\| = O(\delta).$$

So, i.i.d. Gaussian matrices suffice as submatrices to be hidden. Likely $m \geq \frac{n^5}{\delta}\log^2\left(\frac{n}{\delta}\right)$ could be improved to $m \geq \frac{n^{2.01}}{\delta}$ with stronger analysis. We call this the Gaussian Estimation Problem (GPE). Because we will consider different variants of the GPE, we call this one $|\text{GPE}|^2_{\pm}$ for using additive error and norm-squared permanent.

1.4.3 Gaussian permanent estimation and approximate result

The restriction of hiding means we can use an BOSONSAMPLING device only to estimate permanents of random Gaussian matrices.

Definition 33. *The* $|\text{GPE}|^2_{\pm}$ *problem is: given a matrix* $M \sim \mathcal{G}^{n\times n}$, *a tolerance* ϵ, *and a maximum failure chance* δ, *output* $|\text{Per}(M)|^2$ *to within* $\pm\epsilon n!$ *with probability at least* $1-\delta$. *Do this in running time polynomial in* n, $1/\epsilon$, *and* $1/\delta$.

We will show that if this is a hard task, then approximate BOSONSAMPLING cannot be done classically. The proof is similar to that of the exact result *theorem* 24 with accounting for error.

Theorem 34 (Complexity of approximate BOSONSAMPLING). *For* $m \geq \frac{n^5}{\delta}\log^2\left(\frac{n}{\delta}\right)$, *there does not exist a classical randomized algorithm* A *that performs approximate* BOSONSAMPLING *in time* $\text{poly}(m, n, 1/\epsilon)$ *with probability at least* $1-\delta$, *unless the* $|\text{GPE}|^2_{\pm}$ *problem is solvable in* $\mathsf{BPP}^{\mathsf{NP}}$.

Proof. We sketch the proof here. Let $M \sim \mathcal{G}^{n\times n}$ be a random Gaussian matrix whose permanent we wish to approximate to within ϵ. Scale it down by $\sqrt{m}$ and complete it to a $m \times m$ column-unitary matrix U with $m \geq \frac{n^5}{\delta}\log^2\left(\frac{n}{\delta}\right)$. By lemma 31, this matrix cannot be distinguished from a random submatrix significantly better than chance.

Supposed to the contrary there is a classical randomized algorithm A that does approximate BOSONSAMPLING in randomized polynomial time. We will use it to solve the $|\mathrm{GPE}|^2_{\pm}$ problem. Given U as input, the algorithm A samples from a distribution $\mathcal{D}'_U$ that's ϵ-close in variation distance to the true BOSONSAMPLING distribution $\mathcal{D}_U$. Say that A is *close to correct* on an outcome if it chooses it with probability within $\pm\frac{\epsilon}{\delta\binom{m}{n}}$ of its true probability p in $\mathcal{D}_U$. Then, A must be close to correct on for at least $1-\delta$ of the $\binom{m}{n}$ outcomes, or the total error would exceed the allowance of ϵ.

By the Haar-unitary limit and the approximate version of lemma 31, the algorithm cannot locate the submatrix M of U with probability significantly better than chance, so with high probability M is one of matrices that A is close to correct on. That means that if we could obtain its probability of giving M, we could obtain $|\mathrm{Per}\,(M)|^2$ to within $\pm\epsilon n!$ with probability at least $1-\delta$, satisfying the $|\mathrm{GPE}|^2_{\pm}$ problem.

Following the exact result, we can use approximate counting to estimate with the help of an NP oracle the probability that A outputs M. This would therefore solve $|\mathrm{GPE}|^2_{\pm}$ in $\mathsf{BPP}^{\mathsf{NP}}$.

□

1.4.4 Conjectures

We have shown that approximate BOSONSAMPLING is classically hard as long as $|\mathrm{GPE}|^2_{\pm}$ is hard, that is, not in $\mathsf{BPP}^{\mathsf{NP}}$. We would like to prove that approximate BOSONSAMPLING is hard unconditionally, or only relying on the non-collapse of PH as for the exact case. But, the hardness of $|\mathrm{GPE}|^2_{\pm}$ remains open and we lack an alternative proof that does not rely on it.

We proceed by breaking down the hardness of $|\mathrm{GPE}|^2_{\pm}$ into two somewhat more natural conjectures that together imply it, the Permanent-of-Gaussians Conjecture (PGC) and the Permanent Anti-Concentration Conjecture (PACC). Though these are also unproven, we give tentative evidence that they hold true.

For the PGC, we will define $\mathrm{GPE}_{\times}$ as a variant of $|\mathrm{GPE}|^2_{\pm}$. The $\mathrm{GPE}_{\times}$ differs from $|\mathrm{GPE}|^2_{\pm}$ in asking for $\mathrm{Per}\,(M)$ instead of its norm-squared, and in requiring a multiplicative estimate. That is, a estimate that is within $\pm\epsilon\,|\mathrm{Per}\,(M)|$ rather than

$\pm\epsilon n!$. This makes it harder to estimate small permanents. In the extreme case that $\mathrm{Per}(M) = 0$, a solution to the $\mathrm{GPE}_\times$ must also give exactly 0, whereas a $|\mathrm{GPE}|_\pm^2$ solution still has additive leeway.

Definition 35. *The* $\mathrm{GPE}_\times$ *problem is: given a matrix* $M \sim \mathcal{G}^{n\times n}$*, a tolerance* ϵ*, and a maximum failure chance* δ*, output* $\mathrm{Per}(M)$ *to within* $\pm\epsilon\,|\mathrm{Per}(M)|$ *with probability at least* $1-\delta$*. Do this in running time polynomial in* n*,* $1/\epsilon$*, and* $1/\delta$*.*

We then conjecture $\mathrm{GPE}_\times$ to be hard.

Conjecture 36 (Permanent-of-Gaussians Conjecture (PGC))**.** *The* $\mathrm{GPE}_\times$ *problem is* #P*-hard.*[6]

Note that we're conjecturing #P-hardness like the permanent, which is stronger than required. It would suffice for it to be outside of $\mathsf{BPP}^{\mathsf{NP}}$ as we conjecture $|\mathrm{GPE}|_\pm^2$ to be.

We could imagine a state of the world where the permanent is almost always nearly 0 for a random Gaussian matrix. In this case, $|\mathrm{GPE}|_\pm^2$ is trivial – ignore the matrix and just estimate 0. But, the multiplicative estimate of $\mathrm{GPE}_\times$ for the PGC is still hard to obtain because the relative error allowance is small. Since we want $\mathrm{GPE}_\times$ and $|\mathrm{GPE}|_\pm^2$ to be equivalent, we introduce another conjecture to convert between them.

The Permanent Anti-Concentration Conjecture states that the distribution of the permanent is not heavily concentrated around 0. Since $|\mathrm{Per}(M)|^2$ has an expectation of $n!$ for Gaussian $M \sim \mathcal{G}^{n\times n}$, we use $|\mathrm{Per}(M)|^2/n!$ to rescale the distribution to mean 1. We require the probability that $\Pr\left[|\mathrm{Per}(M)|^2/n! < \epsilon\right]$ to be at most $1/\mathrm{poly}$.

Conjecture 37 (Permanent Anti-Concentration Conjecture (PACC))**.** *There are constants* C, D *and* $\beta > 0$ *so that for any* n *and* $\epsilon > 0$

$$\Pr_{M\sim\mathcal{G}^{n\times n}}\left[\frac{|\mathrm{Per}(M)|^2}{n!} < \epsilon\right] < Cn^D\epsilon^\beta$$

[6]Since $\mathrm{GPE}_\times$ is a randomized function problem, it would be more exact to say that if $\mathcal{O}$ solves $\mathrm{GPE}_\times$, then $\mathsf{P}^{\#\mathsf{P}} \subseteq \mathsf{BPP}^{\mathcal{O}}$.

We can define $\mathrm{GPE}_{\pm}$ like $|\mathrm{GPE}|^2_{\pm}$ but without the norm-squared. We show that the PACC implies that $\mathrm{GPE}_{\times}$ and $\mathrm{GPE}_{\pm}$ are interchangeable.

Lemma 38. *Assuming the PACC,* $\mathrm{GPE}_{\times}$ *and* $\mathrm{GPE}_{\pm}$ *are polynomial-time equivalent.*

The direction $\mathrm{GPE}_{\pm} \leq_P \mathrm{GPE}_{\times}$ does not require the PACC and can be proven unconditionally. See Lemmas 46 and 47 of [1] for proofs of this and the further equivalence between $\mathrm{GPE}_{\pm}$ and $|\mathrm{GPE}|^2_{\pm}$. As a result, these conjectures together imply the approximate result.

Proposition 39. *If the Permanent-of-Gaussians Conjecture (PGC) and the Permanent Anti-Concentration Conjecture (PACC) are both true, then there does not exist a classical randomized algorithm A that performs approximate* BosonSampling *in time* $\mathrm{poly}(m, n, 1/\epsilon)$ *with probability at least* $1 - \delta$*, assuming parameters* $m \geq \frac{n^5}{\delta} \log^2\left(\frac{n}{\delta}\right)$.

1.4.5 Evidence for the PGC

The PGC conjectures that computing the permanent is #P hard if given two allowances: only needing to *approximate* the permanent, and only having to be *probably* right, that is with probability $1 - \delta$. We observe that the permanent is known to remain #P hard with either one of the two allowances. However, this combination of "approximately" and "probably" remains open.

Approximately computing the permanent of *any* matrices is #P-hard, as per theorem 26.

Exactly computing the permanent of *most* matrix is #P-hard. This was first shown by Lipton via the self-reducibility of the permanent over finite fields [31]. The idea is to use polynomial interpolation to compute a given permanent using the values of randomly-chosen permanents.

Suppose that we have a oracle to compute the permanent of $n \times n$ matrices over a finite field $\mathbb{F}_p$ that is correct with probability at least $\alpha = 1 - 1/(3n + 3)$. Say we want to find a particular permanent $\mathrm{Per}(M)$. We choose a random matrix N and fix a line of matrices $\mathrm{Per}(M + tN)$ for $t \in \mathbb{F}_p$. Observe that $f(t) = \mathrm{Per}(M + tN)$ is a

degree-n polynomial in t, and that we can query it via the permanent oracle. So, if we evaluate $f(t)$ at $n+1$ distinct values of t, we can determine $f(0) = \mathrm{Per}(M)$ by polynomial interpolation. If N and the values of t are chosen at random, then each oracle query has at most a $1/(3n+3)$ chance of being wrong, so we may union-bound the probability of any mistake at $1/3$. By repeating this protocol, we can amplify to make the failure probability exponentially small.

Better techniques for polynomial interpolation improve this result to require smaller success probabilities α to $\alpha = 3/4 + 1/\mathrm{poly}$ in [19], $\alpha = 1/2 + 1/\mathrm{poly}$ in [20], and $\alpha = 1/\mathrm{poly}$ in [12]. These results can also be applied to $|\mathrm{Per}\,(M)|$ as in $\mathrm{GPE}_\times$ rather than $\mathrm{Per}\,(M)$ using that $|\mathrm{Per}\,(M+tN)|^2$ is a polynomial in t of degree $2n$. The first of these results can be adapted for complex matrix entries from the Gaussian distribution $\mathcal{G}$ rather than a finite field $\mathbb{F}_p$.

However, all these interpolation methods fail when combined with approximating the permanent, as shown in a no-go Section 9.2 of [1]. Small errors at the points being queried can combine to a large error at the target point due to the inherent instability of polynomial interpolation. So, proving the PGC will require a fundamentally different technique from any of those used to show random self-reducibility of the permanent.

1.4.6 Evidence for the PACC

The Permanent Anti-Concentration Conjecture is about the distribution of permanents of random matrices. Unlike the PGC, it is not a question about algorithmic complexity, but one of pure mathematics, specifically random matrix theory. For reference:

Conjecture 40 (Permanent Anti-Concentration Conjecture (PACC)). *There are constants C, D and $\beta > 0$ so that for any n and $\epsilon > 0$*

$$\Pr_{M \sim \mathcal{G}^{n\times n}}\left[\frac{|\mathrm{Per}\,(M)|^2}{n!} < \epsilon\right] < Cn^D\epsilon^\beta$$

The PACC seeks to rule out a state of affairs where the permanent of a random Gaussian matrix is strongly concentrated around 0, trivializing the problem of estimating it additively. We give evidence that this is not the case.

The first piece of evidence is numerical. In fig. 1-2 at the end of this chapter, we plot the distributions of $P_n = |\mathrm{Per}\,(X)|^2/n!$, the norm-squared permanent rescaled to mean 1. We also plot the analogue for the determinant $D_n = |\mathrm{Det}\,(X)|^2/n!$. These plots give an estimate of the distribution for $n = 6$. The numerical evidence up to $n = 10$ is strongly consistent with the PACC. A regression fits around 0 suggests the conjecture is true for exponent β in ϵ^β taking on any value $0 \leq \beta < 1$, and perhaps even $\beta = 1$ itself.

The data also strongly suggests that the permanent and determinant converge to the same distribution. The analogue of the PACC is in fact known to be true for the determinant with exponent $D = \beta(\beta+2)/8$. This follows from an exact characterization of the determinant distribution D_n given by [14]. They show that $|\mathrm{Det}(X)|^2$ is distributed as a product $T_1 T_2 \cdots T_n$ where each $T_k = \chi^2_{2k}$ is a chi-squared variable with $2k$ degrees of freedom, that is the sum of the norm-squares of k independent complex Gaussians $\mathcal{G}$.

Another heuristic piece of evidence is to consider the minor expansion of a permanent as a sum of $n!$ terms $\mathrm{Per}(X) = \sum_{\sigma \in S_n} \prod_{j=1}^n X_{j,\sigma(j)}$. If the summands were probabilistically independent, then we could apply the Central Limit Theorem to show that the permanent approximates a Gaussian distribution with variance $n!$. This would certainly satisfy the PACC. However, some of the summands are correlated by having an overlap of terms. Intuitively, it seems that these correlations should be still weak enough for the bell curve behavior to still apply.

A more refined approach using non-overlapping minors was used by Tao and Vu [47] to show an anti-concentration result for permanent that is unfortunately weaker than what is required for the PACC. Though this result is for Bernoulli matrices, the authors says their methods should apply for Gaussian matrices as well.

Theorem 41 (Tao-Vu [47]).

$$\Pr_{M \sim B^{n \times n}} \left[\frac{|\mathrm{Per}\,(M)|^2}{n!} < \frac{1}{n^{\epsilon n}} \right] < \frac{1}{n^{0.1}}$$

where B is the ± 1-Bernoulli distribution $B = \{+1, -1\}$.

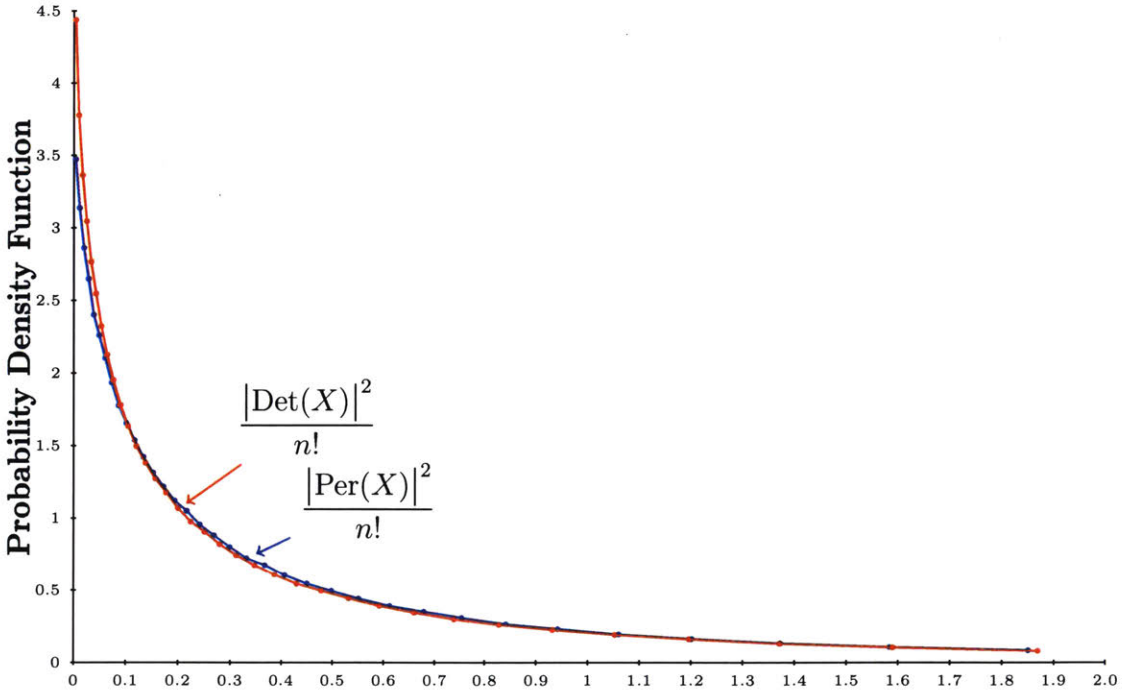

Figure 1-2: Probability density functions of the random variables $P_n = |\mathrm{Per}(X)|^2/n!$ and $D_n = |\mathrm{Det}(X)|^2/n!$ where $X \sim \mathcal{G}^{n\times n}$ is a complex Gaussian random matrix, in the case $n = 6$. Note that $\mathrm{E}[P_n] = \mathrm{E}[D_n] = 1$. As n increases, the bends on the left become steeper. We do not know exactly how the pdfs behave near the origin. This density plot estimate for each of the distributions is produced by generating 10^6 samples and sorting them into 40 equal buckets of $250,000$ points each. So, the first bucket contains the lowest 2.5^{th} percentile of samples, the next bucket the next 2.5^{th} through 5^{th} percentile, and so on. The density is then estimated for each bucket as the fraction of points it contains (0.025) divided by its width, and plotted as a point at the bucket's center. In effect, the plot is a histogram, except rather than using equal intervals on the density axis, the buckets are chosen to contain equal numbers of sample points. As a result, near 0 where the distribution is concentrated, the buckets are narrower and more density estimate points are plotted, allowing the limit behavior for the PACC to be seen more clearly. Thanks to John Watrous for correcting an error in an earlier version of this figure.

Chapter 2

Experimental progress

Parts of this chapter are based on work by the author that appeared as The computational complexity of linear optics [1] *in* Physical Review A *(2015).*

2.1 Experimental background

2.1.1 Motivation for experiments

The complexity results of BosonSampling in chapter 1 have prompted multiple linear optical laboratories to build devices that implement the model. Because the BosonSampling problem is naturally stated in terms of the behavior of a quantum system, it lends well to translating to a physical device. A number of identical photons are generated in distinct input modes (usually spatial), passed through a network of beamsplitters and phaseshifters, and then measured with a photodetector in each mode. The correspondence is very direct and there is little simulation overhead. Each photon in the problem statement corresponds to an actual photon in the experiment, each output of the sampling problem corresponds to a single run of photons through the device, and so forth. (Of course future devices might use a less direct simulation.)

We now consider the motivation for doing such experiments. Perhaps the word "experiment" gives the wrong impression of testing a hypothesis that could be confirmed or disconfirmed by the result of some procedure. That is not the main goal

here. We expect a specific experimental outcome, though it is reassuring when we obtain what we expect.

The statistical distribution of outcomes for the BOSONSAMPLING problem is derived directly from the well-understood quantum-mechanical laws that govern the behavior of identical bosons. It is mathematically expressed in terms of permanents of certain submatrices. Even though we argue that this problem is computationally hard, this is a reflection of our limited computational power rather than confusion over the laws that govern the outcomes. The mathematical status of the results from chapter 1 is not being tested.

One might say that what is being tested is the connection of our model to the reality it is meant to represent. This is true, to an extent. Obtaining experimental results consistent with computed probabilities is a useful check that we haven't made an error, either in deriving formulas or in our understanding of the system. For some, the conceptual implications of quantum mechanics seem so ridiculous that direct observable proof is needed that the world really does work like that. Working with actual devices keeps us grounded. A worthy goal is to get a hands-on demonstration of quantum advantage in contradiction to the Extended Church-Turing thesis. It's one thing to know that there exists a quantum device that can solve a problem beyond the capabilities of existing classical computers, and another thing to have one that does so.

Perhaps what is being tested are our engineering capabilities to build these devices. If an experimental test produces results inconsistent that deviate from the predictions of quantum mechanics, we probably don't conclude that we misunderstood the physics, but that the device is incorrect or suffers from technical limitations, and that we should strive to overcome these with a better device.

2.1.2 Role of linear optics

The challenge towards building scalable BOSONSAMPLING devices is much like the overall quest towards building a quantum computer, but on a more limited scope. Whereas a general qubit-based quantum computer is designed to be universal, i.e. to

perform an arbitrary quantum computation, we set a lower hurdle of merely doing *some* task that exceeds classical capabilities. The BOSONSAMPLING problem is designed to serve as this task, being exactly what a certain quantum device can do. It is not designed to solve a useful problem or do even universal classical computation, much less universal quantum computation.

Moreover, BOSONSAMPLING represents a highly restricted quantum system. It uses noninteracting identical bosons that start with known counts in each mode, passes them through a single-step non-adaptive network with no multi-photon coupling interactions, and measures them in the photon number basis. These limitations make it easier to build a device by narrowing the scope of what it may do. All the quantum effects come from superpositions of identical particles and the interferences of different paths this allows, rather than from explicit coupling as in multiqubit gates. The lack of photon-photon interactions (outside of particle accelerators) is a benefit for a system where noninteraction is part of the design. In contrast, coupling between particles is the point in many implementations of quantum computation where particles represent bits.

2.1.3 Earlier linear optical experiments

Our model follows a long history of work in optical interferometry, which investigates interference between waves of light.

The Hong-Ou-Mandel dip [23] is an experimentally-observed two-photon interference phenomenon. Two identical photons enter a 50:50 beamsplitter from opposite sides. Individually, each photon would be equally likely to pass through the beamsplitter as to bounce off it, doing so in superposition. With both photons hitting the beamsplitter at the same time, they are always observed to exit on the same side. See the diagrams in fig. 1-1 and the calculation in section 1.2.5. Even without explicit coupling between the photons, their outcomes managed to become perfectly correlated. We can view this as a single two-mode BOSONSAMPLING network, with

a 2×2 Hadamard network matrix

$$M = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

With initial state $(1,1)$, the outcome corresponding to final state $S = (1,1)$ has probability $|\mathrm{Per}(M)|^2 = 0$, so the outcome of photons going opposite ways is never observed. The paired-photon outcomes $S = (2,0)$ and $S = (0,2)$ each occur with probability $1/2$.

The Linear Optical Quantum Computing (LOQC) model from Knill, Laflamme and Milburn [27] shows how to do universal quantum computation on a linear optical device aided by adaptive measurements. BosonSampling is based on a similar construction of a network of photons passing through beamsplitters and phaseshifters and being measured by photodetectors, but does not use adaptiveness and is not believed to be universal. Another scheme by Gottesman, Kitaev, and Preskill [21] instead expresses states in terms of modes of a harmonic oscillator. These schemes give the promise that extending the capabilities of devices built to do BosonSampling can lead to building a universal quantum computer.

2.1.4 Experimental tests of BosonSampling

Four independent groups based in Vienna [51], Brisbane [11], Rome [15], and Oxford [45] demonstrated experiments in quick succession in 2012 that implement the BosonSampling setup for small numbers of photons and modes, and checked the results to be close to what was statistically expected. These were later followed by groups in Bristol [13] and Shanghai [56]. These experiments were done with $n = 3$ to $n = 6$ photons with a number of modes ranging between $m = 5$ to $m = 9$. These are summarized in table 2.1. The entries marked * use entangled pairs of photons produced by spontaneous parametric down-conversion (SPDC), whose coupling affects the statistical outcomes.

Table 2.1: Experimental groups performing BOSONSAMPLING experiments with photon and mode counts

Group	Location	# Photons	# Modes
Tillmann et al [51]	Vienna	3	5
Broome et al [11]	Brisbane	3	6
Crespi et al [15]	Rome	3	9
Walmsley et al [45]	Oxford	4*	6
Laing et al [13]	Bristol	6*	6
Lu et al [56]	Shanghai	5	9

2.1.5 Scalability

The implementations of BOSONSAMPLING so far have used a modest number of photons and modes. The computational problem instances that they instantiate can easily be solved classically, and this has been done to verify their results. They serve as proofs of concept for what we hope to be scalable methods to build devices on large numbers of photons and modes.

One may ask what parameters would be large enough to give a convincing demonstration of quantum supremacy. In theory, the computational problem is an asymptotic one and its difficulty relies on solving arbitrarily large instances. In practice, for the given time we may consider $n = 50$ photons and $m = n^2 = 2500$ modes to be just within our classical capabilities to verify. See chapter 3 for discussion on the difficulty of certifying BOSONSAMPLING results and schemes to give weak statistical certification. Still, ideally one would demonstrate with a device that can be scaled up without any hard limits.

2.1.6 Experimental noise

The question of scaling naturally leads to the issue of noise. Real experiments have imperfections that cause them to deviate slightly from the ideal model, and we would like to understand what level of error is tolerable in that it creates only a small deviation in the output distribution.

There are four main sources of noise:

1. Incorrect or correlated initial states

2. Imperfect coding of the unitary U by the linear optical network

3. Partial distinguishability of photons (caused by non-simultaneous arrival), such as mode mismatch within the circuits

4. Photon loss (whether in the network or due to failure to measure)

In section 2.2 and for most of the chapter, we will consider (2), the effect of imperfect coding of the unitary. In current experiments, although individual components are accurate, there is difficulty in either aligning a large number of components or in fabricating precise integrated optics. As a result, inaccurate unitaries remain a significant source of output error in some experiments. The 5-mode and 7-mode experiments in [17] achieved respective fidelities of 0.975 and 0.950, a minority but significant contribution to the variation distance in the output distribution.

2.1.7 Bounds on noise

Many results have proven upper and lower bounds on the amount of noise in various forms that a BosonSampling experiment can withstand in terms of the number of photons n, either in terms of accuracy of the output distribution or in preserving the conjectured computational hardness of BosonSampling.

Leverrier and Patrón [30] demonstrate that to obtain a nearly-correct output distribution, each linear optical element must have fidelity $1 - O\left(1/n^2\right)$ under certain assumptions.

The work of Kalai and Kindler [25] argues that a noise level of additive $\omega\left(1/n\right)$ Gaussian error applied to the overall unitary matrix leads to large deviations in the output distribution, and moreover than this error renders the system classically simulable.

Shchesnovich [40] gives sufficient conditions for an experimental realization of BosonSampling to demonstrate a conflict with the Extended Church-Turing Thesis.

He also proves that for a small distinguishability error, a state fidelity of $O\left(\frac{1}{n}\right)$ is necessary and sufficient to obtain constant distance in the distribution.

Rohde and Ralph [35] give evidence that linear optical systems remain out of reach of classical simulation even in the presence of photon loss and mode mismatch.

Tichy [49] bounds the variation distance in outcomes between partial-distinguishable and perfectly identical photons.

2.2 Robustness to error in the network matrix

We look at the effect caused by imperfections in the linear optical network that cause a deviation in the unitary matrix that it encodes. We assume that the actual network still applies a unitary matrix $\tilde{U}$ (in particular, it takes pure states to pure states), but one that is slightly different from the desired matrix U. We will give an upper bound for the error in the output distribution in terms of the error in U. In particular, we will show that for n photons, an operator distance of $o\left(1/n\right)$ suffices to give $o\left(1\right)$ error in the output distribution.

Our main result is a bound on the error in the BOSONSAMPLING distribution D_U caused by inaccuracy in the single-particle unitary U that encodes the action of the beamsplitters and phaseshifters.

Theorem 42. *For unitary matrices U and $\tilde{U}$, the L_1 distance between the corresponding n-photon* BOSONSAMPLING *distributions $\mathcal{D}_U$ and $\mathcal{D}_{\tilde{U}}$ is bounded as*

$$\left\|\mathcal{D}_{\tilde{U}} - \mathcal{D}_U\right\|_1 \leq n \left\|\tilde{U} - U\right\|_{\text{op}}$$

Note that there is no dependence on the number of modes m. As a result, the accuracy of the unitaries only needs to depend on the number of photons n, with $o\left(\frac{1}{n}\right)$ error sufficing.

Corollary 43. *To obtain vanishingly small error $\left\|\mathcal{D}_{\tilde{U}} - \mathcal{D}_U\right\|_1 = o\left(1\right)$, it suffices for the unitary representing the entire transformation to have $\left\|\tilde{U} - U\right\|_{\text{op}} = o\left(\frac{1}{n}\right)$.*

This can be achieved by having each beamsplitter and phaseshifter in the network be sufficiently accurate. Since such a network can be made with a depth of $O(n \log m)$ components (Theorem 45 of [1]), it suffices to divide the tolerable error by that amount.

Corollary 44. *In order to have* $\|\mathcal{D}_{\tilde{U}} - \mathcal{D}_U\|_1 = o(1)$*, it suffices for every component in the network to have an accuracy of* $\left\|\tilde{A} - A\right\|_{\text{op}} = o\left(\frac{1}{n^2 \log m}\right)$.

2.3 Comparison to previous results

2.3.1 Relation to previous work

The result is comparable to the standard result for qubit-based circuits of Bernstein and Vazirani [8]. To better parallel our main result, we restate this result here assuming identical gates and in terms of particles. We also generalize qubits to m-mode qudits, which does not affect the bound.

Theorem 45. (Bernstein-Vazirani, adapted) *Suppose one applies a noisy unitary matrix* $\tilde{U}$ *to each of* n *distinguishable particles (qudits), then measures each particle to sample an* n*-tuple of measurement outcomes from* $\{1, 2, \ldots, m\}$ *Then, the distance in the outcome distribution* $\mathcal{D}_{\tilde{U}}$ *from that with error-free matrix* U *is bounded as*

$$\|\mathcal{D}_{\tilde{U}} - \mathcal{D}_U\|_1 \leq n \left\|\tilde{U} - U\right\|_{\text{op}}$$

Previous work on BOSONSAMPLING noise sensitivity has given necessary bounds for the required accuracy of the linear optical network. In other words, it's shown that above certain thresholds of noise, one gets large inaccuracies in the distribution of outcomes. Thus, it proves a certain level of noise to be prohibitive for BOSONSAMPLING. This work, in contrast, shows a certain level of accuracy to be sufficient.

The work of Leverrier and Patrón [30] demonstrates that each linear optical element must have fidelity $1 - O(1/n^2)$ by considering a composite experiment in which the network is applied followed by its inverse, with independent noise in each part.

As shown in section 2.4.5, this corresponds to a required single-operator distance of $O\left(1/n^2\right)$, which has a factor of $\log m$ gap from our sufficient bound of $o\left(1/\left(n^2 \log m\right)\right)$ per operator being sufficient. Applying methods in Section 2.4.4 to their result, we obtain an overall distance of $\left\|\tilde{U} - U\right\|_{op} = O\left(\log m/n\right)$, again a factor of $\log m$ off of our result.

Kalai and Kindler [25] argue that a noise level of additive $\omega\left(1/n\right)$ Gaussian error applied to the overall unitary matrix leads to large deviations in the output distribution. Specifically, above such a threshold, one finds vanishingly little correlation between the original and noise permanent of a submatrix, and thus between outcomes of a BOSONSAMPLING experiment. Translating to our error model of unitary noise as in section 2.4.5, a typical such error corresponds to operator distance $\omega\left(1/\sqrt{n}\right)$, significantly above the $O\left(1/n\right)$ distance that we show.

In both cases, once we convert the error measures to a consistent scale, we find the sufficient bound for noise shown in this work is consistent with the necessary bound shown in the previous. Moreover, a gap remains for potential improvement.

Our resulting scaling is similar to that obtained in [42], where for a small distinguishability error, a state fidelity of $O\left(\frac{1}{n}\right)$ is necessary and sufficient to obtain constant distance in the distribution.

2.4 Proof of result

2.4.1 Outline of proof

We give an outline of the proof here, and prove each part in the upcoming sections.

Let Ψ_0 be the initial n-boson state, and let φ be the homomorphism from a unitary acting on one boson to that acting on n identical bosons. Applying unitaries U and $\tilde{U}$ respectively to the initial state Ψ_0 produce:

$$\begin{aligned} \Psi &= \varphi\left(U\right)\Psi_0 \\ \tilde{\Psi} &= \varphi\left(\tilde{U}\right)\Psi_0 \end{aligned}$$

Measuring Ψ and $\tilde{\Psi}$ respectively in the standard basis gives outcome distributions $\mathcal{D}_U$ and $\mathcal{D}_{\tilde{U}}$

The main step is Theorem 46, which states that the distance between the n-boson unitaries is at most a factor of n times that between the 1-boson unitaries

$$\left\|\varphi\left(\tilde{U}\right) - \varphi\left(U\right)\right\|_{\text{op}} \leq n \left\|\tilde{U} - U\right\|_{\text{op}}$$

We then conclude with a standard argument (Lemma 50) that the distance between the output distributions is at most the operator distance between the matrices that produced them

$$\|\mathcal{D}_{\tilde{U}} - \mathcal{D}_U\|_1 \leq \left\|\varphi\left(\tilde{U}\right) - \varphi\left(U\right)\right\|_{\text{op}}$$

2.4.2 Effect of the homomorphism

We first show that close unitaries U and $\tilde{U}$ induce nearby n-boson unitaries $\varphi\left(U\right)$ and $\varphi\left(\tilde{U}\right)$. Thus, if two operations act similarly on single bosons, then they also act similarly on n identical bosons. The blowup is simply a factor of n, the number of bosons.

Theorem 46. *Let φ be the homomorphism that takes an $m \times m$ unitary matrix U acting on a single boson and produces a $N \times N$ unitary matrix acting on n identical bosons with $N = \binom{m}{n}$. Then,*

$$\left\|\varphi\left(\tilde{U}\right) - \varphi\left(U\right)\right\|_{\text{op}} \leq n \left\|\tilde{U} - U\right\|_{\text{op}}$$

In order to prove this, it will be useful to have two lemmas. Lemma 47 expresses the operator distance between two unitary matrices A and B in terms of the eigenvalues of AB^{-1}. Lemma 48 relates the eigenvalues of $\varphi\left(M\right)$ to those of M.

Lemma 47. *If A and B are unitary, their operator distance can be expressed in terms*

of the eigenvalues $\{\lambda_i\}$ *of* AB^{-1} *as*

$$\|A - B\|_{\text{op}} = \max_i |\lambda_i - 1| .$$

Proof. Since AB^{-1} is unitary, it diagonalizes via unitaries as $AB^{-1} = V\text{diag}(\lambda_i)V^*$. Using the operator norm's invariance to left-multiplication or right-multiplication by a unitary, we have

$$\begin{aligned} \|A - B\|_{\text{op}} &= \left\|AB^{-1} - I\right\|_{op} \\ &= \|V\left(\text{diag}\left(\lambda_i\right) - I\right)V^*\|_{\text{op}} \\ &= \|\text{diag}\left(\lambda_i - 1\right)\|_{\text{op}} \\ &= \max_i |\lambda_i - 1| . \end{aligned}$$

□

Lemma 48. *If* M *has eigenvalues* $(\lambda_1, \ldots, \lambda_m)$*, then the eigenvalues of* $\varphi(M)$ *are* $\lambda_1^{s_1} \cdots \lambda_m^{s_m}$ *for each ordered partition* S *of* n *into* m *parts with sizes* $s_1, \ldots, s_m$*.*

Proof. Let v_i be the eigenvector corresponding to λ_i. We will construct eigenvectors of $\varphi(M)$ in terms of the v_i and note that they have the desired eigenvalues.

For each eigenvector v_i, let $v_i(x)$ be the formal polynomial $(v_i)_1 x_1 + \cdots + (v_i)_n x_n$. For each S, let p_S be the degree-n polynomial

$$p_S(x) = v_1^{s_1}(x) \ldots \cdot \cdots \cdot v_m^{s_m}(x)$$

If we consider $\varphi(M)$ as it acts on the Fock basis, we see that each $p_S(x)$ is an eigenvector with eigenvalue $\lambda_1^{s_1} \cdots \lambda_m^{s_m}$:

$$\begin{aligned}\varphi(M)(p_S(x)) &= (Mv_1)^{s_1}(x)\cdots(Mv_m)^{s_m}(x)\\ &= (\lambda_1 v)^{s_1}(x)\cdots(M\lambda_m v)^{s_m}(x)\\ &= \lambda_1^{s_1}\cdots\lambda_m^{s_m}(p_S(x))\end{aligned}$$

Since we have one eigenvalues for each S, the number of which equals the dimension $\left(\binom{m}{n}\right)$ of $\varphi(M)$, this is the full set of eigenvalues. □

Now, we're ready to prove Theorem 46, which we restate here.

Theorem 49. *Let φ be the homomorphism that takes a $m \times m$ unitary matrix U acting on a single boson and produces a $N \times N$ unitary matrix acting on n identical bosons with $N = \left(\binom{m}{n}\right)$. Then,*

$$\left\|\varphi\left(\tilde{U}\right) - \varphi(U)\right\|_{\text{op}} \leq n\left\|\tilde{U} - U\right\|_{\text{op}}$$

Proof. Let $(\lambda_1, \ldots, \lambda_m)$ be the eigenvalues of $\tilde{U}U^{-1}$. From Lemma 48, the eigenvalues of $\varphi\left(\tilde{U}\right)\varphi(U)^{-1}$, which equals $\varphi\left(\tilde{U}U^{-1}\right)$ because φ is a homomorphism, are $\lambda_1^{s_1}\cdots\lambda_m^{s_m}$ for each ordered partition S of n into m parts, which we write as λ^S for brevity.

We now bound the distance of λ^S from 1 in terms of the distances of the λ_i from 1. As eigenvalues of a unitary matrix, the λ_i are complex phases with norm 1, we can inductively apply

$$\begin{aligned}|ab-1| &= |ab-a+a-1|\\ &\leq |a|\,|b-1| + |a-1|\end{aligned}$$

to get

$$\left|\lambda^S - 1\right| \leq \sum_i s_i\,|\lambda_i - 1| \leq n \max_i |\lambda_i - 1| \tag{2.1}$$

From Lemma 47, we have

$$\max_i |\lambda_i - 1| = \left\| \tilde{U} - U \right\|_{\text{op}}$$

and

$$\max_S \left| \lambda^S - 1 \right| = \left\| \varphi\left(\tilde{U}\right) - \varphi\left(U\right) \right\|_{\text{op}},$$

so equation 2.1 gives the desired result

$$\left\| \varphi\left(\tilde{U}\right) - \varphi\left(U\right) \right\|_{\text{op}} \leq n \left\| \tilde{U} - U \right\|_{\text{op}}$$

□

2.4.3 Bounding distance between the output distributions

In Section 2.4.2, we showed that $\tilde{U}$ being close to U implies that the corresponding n-boson transition matrices $\varphi\left(U\right)$ and $\varphi\left(\tilde{U}\right)$ are close. We now argue that applying close transition matrices to the same input produces close measurement distributions.

Let Ψ_0 be the initial n-boson state. For BOSONSAMPLING, this is a Fock basis state $|1_n\rangle$, but this is not necessary for this result. Applying unitaries U and $\tilde{U}$ to Ψ_0 produce states that we call

$$\begin{aligned} \Psi &= \varphi\left(U\right)\Psi_0 \\ \tilde{\Psi} &= \varphi\left(\tilde{U}\right)\Psi_0 \end{aligned}$$

The distributions $\mathcal{D}_U$ and $\mathcal{D}_{\tilde{U}}$ are produced by measuring Ψ and $\tilde{\Psi}$ respectively in the standard basis.

We show that the distance between the distributions is bounded by the operator distance between the respective operators that produced them.

Lemma 50. $\|\mathcal{D}_{\tilde{U}} - \mathcal{D}_U\| \leq \left\| \varphi\left(\tilde{U}\right) - \varphi\left(U\right) \right\|_{\text{op}}$

Proof. We first bound the Euclidian distance of the resulting states from the definition

of the operator norm

$$\begin{aligned}\left\|\tilde{\Psi}-\Psi\right\| &= \left\|\left(\varphi\left(U\right)-\varphi\left(\tilde{U}\right)\right)\Psi_0\right\| \\ &\leq \left\|\varphi\left(\tilde{U}\right)-\varphi\left(U\right)\right\|_{\mathrm{op}}\left\|\Psi_0\right\| \\ &= \left\|\varphi\left(\tilde{U}\right)-\varphi\left(U\right)\right\|_{\mathrm{op}} \end{aligned} \tag{2.2}$$

Now, we show that variation distance between $\mathcal{D}_U$ and $\mathcal{D}_{\tilde{U}}$ is bounded by this distance $\left\|\tilde{\Psi}-\Psi\right\|$.

The variation distance $\left\|\mathcal{D}_{\tilde{U}}-\mathcal{D}_U\right\|_1$ corresponding to the distributions obtained from a standard basis measurement is bounded by the trace distance, the maximum such variation over all projective measurements.

$$\left\|\mathcal{D}_{\tilde{U}}-\mathcal{D}_U\right\|_1 \leq \left\|\tilde{\Psi}-\Psi\right\|_{\mathrm{tr}}$$

We use the expression for trace distance between pure states and bound this expression in terms of $\left\|\tilde{\Psi}-\Psi\right\|$.

$$\begin{aligned}\left\|\mathcal{D}_{\tilde{U}}-\mathcal{D}_U\right\|_1 &\leq \left\|\tilde{\Psi}-\Psi\right\|_{\mathrm{tr}} \\ &= \sqrt{1-\left|\left\langle\tilde{\Psi}\mid\Psi\right\rangle\right|^2} \\ &\leq \sqrt{1-\left(\mathrm{Re}\left\langle\tilde{\Psi}\mid\Psi\right\rangle\right)^2} \\ &= \sqrt{1-\left(1-\frac{1}{2}\left\|\tilde{\Psi}-\Psi\right\|\right)^2} \\ &\leq \left\|\tilde{\Psi}-\Psi\right\| \end{aligned}$$

□

Combining this with Equation 2.2 gives the bound

$$\left\|\mathcal{D}_{\tilde{U}}-\mathcal{D}_U\right\|_1 \leq \left\|\varphi\left(\tilde{U}\right)-\varphi\left(U\right)\right\|_{\mathrm{op}},$$

which, along with Theorem 46

$$\left\|\varphi\left(\tilde{U}\right)-\varphi\left(U\right)\right\|_{\mathrm{op}} \leq n\left\|\tilde{U}-U\right\|_{\mathrm{op}}$$

gives the main result.

2.4.4 Error tolerance of components of the linear optical network

We now investigate the maximum error on components of the linear optical network that still guarantees that the output distribution is vanishingly close to the ideal one. This requires bounding the error of the unitary produced by a linear optical network in terms of that of its components.

Proposition 51. *If each component $\tilde{A}$ of a linear optical network is within operator distance ϵ of the ideal component A*

$$\left\|\tilde{A}-A\right\|_{\mathrm{op}} \leq \epsilon,$$

then the produced unitary U acting on the first n modes has accuracy

$$\left\|\tilde{U}-U\right\|_{\mathrm{op}} = O\left(n\epsilon \log m\right)$$

and the measured output has

$$\left\|\mathcal{D}_{\tilde{U}}-\mathcal{D}_{U}\right\|_{1} = O\left(n^{2}\epsilon \log m\right).$$

Proof. We wish to bound the operator distance error of the network in terms of that of its components. We use two familiar facts about operator distance:

- For components are applied in parallel, the overall operator distance error is at most that of each component, So, if each component has some maximum error, so does each layer in the network.

- For components applied in series, the total operator distance error is at most the sum of the operator distance error of the components.

A linear optical network for n fixed input modes and m output modes can be implemented using $O(mn)$ beamsplitters and phaseshifters in a network of depth $O(n \log m)$ (Theorem 45 of [1]). So, if each optimal element is within operator norm ϵ of the ideal, we are guaranteed the following accuracy for a linear optical network:

$$\left\|\tilde{U} - U\right\|_{op} = O(n \log m) \left\|\tilde{A} - A\right\|_{\text{op}} = O(n\epsilon \log m)$$

Applying the main theorem then gives an overall error of □

$$\|\mathcal{D}_{\tilde{U}} - \mathcal{D}_U\|_1 = O\left(n^2 \epsilon \log m\right).$$

Corollary 52. *In order to have* $\|\mathcal{D}_{\tilde{U}} - \mathcal{D}_U\|_1 = o(1)$*, it suffices for every component in the network to have an accuracy of* $\left\|\tilde{A} - A\right\|_{\text{op}} = o\left(\frac{1}{n^2 \log m}\right)$.

2.4.5 Comparison between noise models

Previous work on BosonSampling noise [25, 30] used different measures of error than we did. In order to put these results on the same scale as ours, we will calculate the amount of operator distance error that corresponds to the errors they prove prohibitive. Note that because these results are optimized for their specific model of error, the converted results are not necessarily the strongest possible.

The work of Leverrier and Patrón [30] demonstrates that each linear optical element must have fidelity $1 - O(1/n^2)$. This corresponds to operator distance $O(1/n^2)$ for each element. From the observation in Section 2.4.4 that the operator distance of the whole network is at most its depth times that of each component, and the result that $O(n \log m)$ depth suffices (Theorem 45 of [1]), this corresponds to necessary error $O(\log m/n)$.

The work of Kalai and Kindler [25] argues that a noise level of additive $\epsilon = \omega(1/n)$ Gaussian error is prohibitive for Boson Sampling. We show that this corresponds to

operator distance

$$\left\|\tilde{U} - U\right\|_{op} = \omega(1/\sqrt{n})$$

so that we may put it on the same scale as out result.

Consider an ϵ-noise of a matrix X. In order to match with operator distance, we consider X to be the entire $m \times m$ unitary matrix, rather than an $n \times n$ submatrix, since we expect the error to affect entries in the whole matrix just as it does the submatrix. Since each entry of a unitary matrix has a norm of $1/\sqrt{m}$ in RMS average, the error should be $\epsilon/\sqrt{m}$.

So, an ϵ-noise of a unitary matrix U is given by

$$\tilde{U} = \sqrt{1-\epsilon}U + \sqrt{\epsilon}G/\sqrt{m},$$

where G is a matrix of i.i.d. complex Gaussians. To first order in ϵ, the difference $\tilde{U} - U$ is given by

$$\tilde{U} - U = -\epsilon U/2 + \sqrt{\epsilon}G/\sqrt{m} + O(\epsilon^2)$$

Since U and $G/\sqrt{m}$ have entries of the same RMS-norm, for small ϵ, the term with coefficient $\sqrt{\epsilon}$ dominates the remaining terms:

$$\tilde{U} - U = \sqrt{\epsilon}G/\sqrt{m} + O(\epsilon)$$

Then, the prohibitive amount of noise $\epsilon = \omega(1/n)$ corresponds to

$$\tilde{U} - U = \omega(1/\sqrt{m})G/\sqrt{m}$$

Finally, with the result from [29] that a random $m \times m$ Gaussian matrix has operator norm $\Theta(\sqrt{m})$ with high probability, $\|G/\sqrt{m}\|_{op} = \Theta(1)$, and so the corresponding operator distance is

$$\left\|\tilde{U} - U\right\|_{op} = \omega(1/\sqrt{n})$$

2.5 Interpretation of results

Note that we do not obtain that a constant error suffices. In fact, constant error does not suffice, as shown in [25, 30], suggesting that fault-tolerance is necessary to perform scalable quantum computing. This is not surprising – we expect that more photons require higher accuracy for the unitary because each photon interaction with the unitary introduces error. Similarly, as the network requires more and more components, each component must have better accuracy to maintain the same overall accuracy.

We conjecture that the requirement we obtain that $\left\|\tilde{U} - U\right\|_{op} = o\left(\frac{1}{n}\right)$ is the best possible. It parallels the Bernstein-Vazirani result for qubit-based circuits [8]. Because each photon passes through the network and experiences its imperfections, it is natural to conjecture that the acceptable error in the network falls inversely with the number of photons. Likewise, since each photon passes through a depth-$O\left(n \log m\right)$ network, one might has guessed that the acceptable error of each component is $O\left(\frac{1}{n \log m}\right)$ time that of the full network, as corresponding to the sufficient bound in 52.

2.5.1 Future work

This result addresses only one type of noise: errors in the beamsplitters and phase-shifters that cause them to implement a slightly erroneous unitary matrix. We would like to extend these results to other sources of noise. The more plausible potential extensions of this approach are those dealing with continuous errors rather than discrete ones like photon losses. One such source is the partial distinguishability of the photons as they pass through the network, a phenomenon that has been mathematically modeled by Tichy [49] and Xu [57].

The gaps between the sufficient bound proven here and the necessary bounds proven in [25, 30] mean that an improvement must be possible to at least one of the sides. Moreover, all the results are fine-tuned for models of noise, so it would be ideal to bound the error under each of the noise models.

Chapter 3

Certification of BosonSampling

3.1 Overview of certification

To give experimental evidence against the Extended Church-Turing Thesis, we would like to demonstrate a BosonSampling device on a scale beyond which a classical computer could simulate its behavior. For such a demonstration to be effective, we must be confident that it is in fact solving the stated computational problem. Moreover, we may wish to convince an observer who does not examine the internals of the device, but treats it as a black box. We call this *certification* of BosonSampling.

Of course, our understanding of quantum mechanics already prescribes the distribution of outcomes a BosonSampling device produces. If one were confident that each component of the optical network implements exactly the quantum transformation its it designed to do, there would be no need to certify the correctness of the outcome. Nevertheless, there are multiple reasons to ask for certification.

First, laboratory components are not perfect implementations of their mathematical idealization. As a result, the physical device performs a noisy version of the computational problem, which may render the computation classically doable. See chapter 2 for an in-depth discussion of the effects of noise. Even if each individual component is relatively accurate, the errors accumulate on each photon for each component it interacts with. The required number of photons and components grows with the size of the instance, and we want to use an instance large enough to be

beyond current capabilities, and to scale beyond that. Verifying small versions of the device does not guarantee that a scaled-up version works correctly. Though we may mathematically bound the total error as in chapter 2, these results assume specific models of noise that do not fully describe actual behavior. Or, there may be further sources of inaccuracy that we have not considered. So, the gold standard for verifying correctness is to check the performance of an actual device.

Next, one might doubt the rules of quantum mechanics that govern the device. A skeptic may question the correctness of the permanent formulas for bosons, or even the notions of superposition and quantum indeterminacy. Or, perhaps the skeptic is convinced of the behavior on a small scale where results can be checked directly, but believes it will break down on a larger scale of, say, 100 photons. Indeed, part of the goal of demonstrating quantum advantage is to show empirically that the quantum world differs from the classical world in a fundamental way, by harnessing quantum phenomena to do computations that cannot efficiently be done classically. Doing so without assuming quantum workings of the device therefore gives direct evidence our for our understanding of quantum mechanics. That it grants additional computational power strongly suggests that it is not simple a recasting of classical principles. A black-box verification that does not rely on physical assumptions is the strongest way to demonstrate quantum advantage.

Finally, one may seek to avoid being tricked by someone maliciously trying to pass off a classical device as doing BosonSampling. Theoretical computer science has a long tradition of asking for protocols to be robust against an adversary who does everything in their power to mislead you. This is an extension of the general worst-case approach to complexity theory that strives to prove results that are strong as possible. By being skeptical of all claims as to the device's working, the verifier comes to rely on only what they can observe during the certification protocol.

3.1.1 Difficulty of certification

BosonSampling is hard to certify for the very reason it is a hard computational problem: the intractability of the permanent. Any outcome of a 100-photon exper-

iment is the permanent of a 100 by 100 matrix, which is beyond our current ability to compute classically. Indeed, that's the very point of the device! To demonstrate quantum advantage, the it must solve a problem whose answer we can't find classically.

Moreover, running a BOSONSAMPLING device doesn't actually tell us the permanent of any matrix. It is solving a sampling problem of producing random photon counts as per a certain distribution. The number of possible photons counts is exponential in the number of photons, so we are unlikely to see any output repeat within a feasible number of trials. Empirically estimating the probability of any one outcome turns out hopeless.

Contrast this with the situation for Shor's algorithm [43]. If a quantum computer outputs the factorization of a number, it is easy to multiply the claimed factors and check the result. Factoring lies within the class NP of problems verifiable in polynomial time. The permanent, however, is complete for #P, which is believed not to lie within NP or even within the polynomial hierarchy. It is not known how one who computes the permanent can write a proof of its value that can be verified without doing a comparably large computation.

Yet, some hope remains for a certification protocol via interactivity. There are examples of proving correctness of a solution for a problem outside of P or even NP via protocols where the prover and verifier communicate a series of messages to each other. The seminal result that IP = PSPACE [39] shows how to perform an interactive proof of any problem solvable in polynomial space. This can be extended to quantum interactive proofs [55] in many powerful ways. Blind quantum computing [10] allows a user to have arbitrary computation done by a quantum server that they can verify correct almost classically.

3.1.2 Weak certification

As of this writing, we are not aware of any protocols that certify BOSONSAMPLING in the adversarial sense. So, we switch our focus to tests that a true BOSONSAMPLING device will pass but a noisy or incorrect one is unlikely to, giving circumstantial

evidence of correctness. We call this *weak certification.*

One approach to weak certification is statistical verification. We find a summary statistic of the BOSONSAMPLING distribution that is a function of the output counts for a trial, and may depend on the network matrix.

Then, we take a polynomial number of output samples from our BOSONSAMPLING device with that network matrix, and compute the statistic for each to get an empirical estimate of its distribution. If it's close to the theoretically computed statistic, we accept the device, and otherwise reject it.

An example of a verifiable statistic is the marginal distribution for a given output mode, that is the probability vector of measuring k photons for each k from 0 to the number of photons n. This may be estimated in polynomial time [22]. Fixing a network matrix and running $\text{poly}(n)$ trials of the device gives us estimates of the probability of each output count in that mode. An honest device will estimate each probability to within $\pm 1/\text{poly}(n)$ with all but exponentially small probability, by the Chernoff Bound. More strongly, any k-mode marginal can be computed in time $n^{O(k)}$ for n photons [22], which is polynomial for fixed k. Each of the possible $\binom{m}{k}$ subsets of output modes of size k can be verified to be consistent.

Does this check guarantee the device is performing BOSONSAMPLING? No. The distribution of output counts could just have the right k-mode marginals. A classical adversary can efficiently make-to-order a joint distribution to achieve any consistent set of polynomially-many marginals [9]. So, the adversary could forge a classical device to produce a distribution specifically to pass this verification scheme.

3.2 Smuggling permanents

Smuggling permanents is an attempt at a hard decision problem that can be solved using a BOSONSAMPLING device, and so can be used for certification. However, we show that such a scheme cannot work.

The idea is to generate rigged network matrices where one output is overwhelmingly likely. Anyone who runs the device can easily check that this is the case. But,

we want it to be classically hard to figure out which outcome is the likely one, or a forger could make a classical device that does the same. We may also formalize this as a search problem to identity the rigged outcome or a decision problem of determining if one exists.

To avoid the rigged outcome being obvious, we need it to have a large permanent (in absolute value) even though its entries look typical, as in the Hiding Lemma (Lemma 42) of [1]. We instead need the largeness of the permanent to via constructive interference among the superposition of paths that contribute to its permanent, in a way that perhaps cannot be detected without being able to compute the permanent itself.

We show that such hiding is not possible.

3.2.1 Row norm bound

We first show that in the sparse regime $m \sim n^{2+\epsilon}$, any network matrix with an outcome with a $1/\operatorname{poly}$ chance of occurring can be distinguished from a Haar-random matrix. In order for a submatrix to have such a large permanent, it must have rows larger than would appear by chance. We will later again discuss the correlation between row norms and outcome probabilities in section 3.4.

Theorem 53. *For any polynomial p, there is an algorithm which distinguishes, with exponentially small error probability, whether a matrix is*

1. *A Haar-random column-unitary $m \times n$ matrix*
2. *An $m \times n$ matrix that contains a $n \times n$ submatrix with large permanent $1/\operatorname{poly}(n)$*

if $m \geq \frac{2}{e^2} n^2$.

Proof. We prove this as follows. A typical row of the matrix has norm-squared n/m, and we will show it is exponentially unlikely that any row exceeds it by a factor of 2. In contrast, we show that any submatrix with a large permanent must have a row with norm-squared much larger than the average. Simply by checking for a row with norm-squared at least $2n/m$, we can distinguish the two cases.

We use the following bound of the permanent of a matrix in terms of the product of the norms of its rows, called the Hadamard Inequality for Permanents. stated and proven in [16].

Proposition 54. *The permanent of an $n \times n$ matrix is bounded by*

$$|\mathrm{Per}\,(M)| \leq \frac{n!}{n^{n/2}} \prod_i |M_i|$$

Corollary 55. *Any $n \times n$ matrix M with permanent of norm α must have a row of norm at least*

$$|M_i| \geq \frac{\alpha^{1/n}\sqrt{n}}{(n!)^{1/n}} \approx \frac{e\alpha^{1/n}}{\sqrt{n}},$$

where we obtain the approximation by Stirling's formula.

We now show that a random row M_i of a $m \times n$ column-unitary matrix is unlikely to have such a large norm. Note that the expected norm-squared of such a row is

$$\mathrm{Exp}\left[\left\|M_i^2\right\|\right] = n\mathrm{Exp}\left[|M_{ij}|^2\right] = n/m.$$

As long as no row exceeds the average norm-squared by more than a factor of c, we have

$$\begin{aligned} |\mathrm{Per}\,(M)| &\leq \frac{n!}{n^{n/2}} \prod_i |M_i| \\ &= \frac{n!}{n^{n/2}} \left(\sqrt{c\frac{n}{m}}\right) \\ &= n!c^{n/2}m^{-n/2} \\ &\approx \left(\frac{n}{\sqrt{m}}\frac{\sqrt{c}}{e}\right)^{n/2}, \end{aligned}$$

where we use Stirling's approximation for the last line. The final expression is exponentially small for $m \geq \frac{c}{e^2}n^2$

It remains to show that the a row of a column-unitary matrix is unlikely to exceed its average norm squared by a large factor. We will do so for the constant factor $c = 2$, though a smaller factor would suffice.

We follow Section 5.1 of [1] in observing that the entries of the rows of a sub-unitary matrix for $m \sim n^2$ tend towards independent complex Gaussians, as the correlations induced by unit norm are weak for such small subsets. The norm-squared is then a sum of $2n$ complex Gaussians, and is therefore is the chi-squared distribution χ^2_{2n} with $2n$ degrees of freedom. The Chernoff bound then implies the following concentration result.

Proposition 56. *The probability that χ^2_{2n} exceeds its mean $2n$ by a factor of d is bounded by*

$$\Pr_{x \sim \chi^2_{2n}} [x \geq 2nd] \leq \left(de^{1-d}\right)^n$$

For $d = 2$, the probability is at most $(2/e)^n$. So, for $m \sim n^2$, by union bound, the probability that any row exceeds this bound is exponentially small. So, the row-norm of a rigged row is exponentially unlikely to be reached by chance.

□

3.2.2 Unitary matrices with large permanents

In the previous section 3.2.1 we showed that a large-permanent matrix cannot be smuggled into a BOSONSAMPLING matrix in the sparse case $m \sim n^{2+\epsilon}$. We now argue that smuggling is not possible when $m = n$, i.e. there is one photon per mode. We show that if one starts with one photon in every mode, there cannot be an overwhelmingly high probability of ending with one photon per mode unless the matrix looks like the identity matrix or a trivial variation of it. We hope to extend these results in the future to other outcomes and cases where $m > n$.

The probability of maintaining the one-photon-per-mode state is the norm-squared permanent of the unitary matrix. We consider the constraints put on the unitary matrix by having a large permanent.

Unitary permanent exactly 1

Recall that $|\mathrm{Per}\, M| \leq \|M\|_2^n$ for an $n \times n$ matrix (see [7] for a proof), so $|\mathrm{Per}\, U| \leq 1$ for U unitary. Indeed, this is confirmed by the physical interpretation that this outcome

cannot have probability greater than 1. We begin by noting that equality only holds for matrices equivalent to the identity by a combination of permutation and phases. Such matrices correspond to optical networks that simply shuffle the photons around without any superposition of paths.

Theorem 57. *If U is a unitary matrix with $|\mathrm{Per}(U)| = 1$, then U equals the identity, up to permuting its rows and multiplying each entry by a phase.*

Proof. By Ryser's formula for computing the permanent, we may express $\mathrm{Per}(U)$ as the sum

$$\mathrm{Per}(U) = \frac{1}{2^n} \sum_{b \in \{-1,+1\}^n} \left(\prod_i b_i \prod_i (Ub)_i \right)$$

First, we will show that to achieve the condition, $|\mathrm{Per}(U)| = 1$, each summand must achieve it maximum possible norm of 1.

The summand corresponding to b has norm $\prod_i |(Ub)_i|$

Because U is unitary, it preserves norm, so

$$\sum_i |(Ub)_i|^2 = \|Ub\|^2 = \|b\|^2 = n$$

By the AM-GM inequality,

$$\prod_i \left(|(Ub)_i|^2 \right) \leq \left(\frac{1}{n} \sum_i |(Ub)_i|^2 \right)^n = 1,$$

so $|\prod (Ub)_i| \leq 1$.

Therefore, to achieve the maximum possible value of $|\mathrm{Per}(U)| = 1$, we must have $|\prod_i (Ub)_i| = 1$ for each $b \in \{-1, +1\}^n$.

In particular, this means the AM-GM inequality is tight, which occurs when the terms are equal, so $|(Ub)_i| = 1$ for each i and b. We write out

$$(Ub)_i = \sum_j U_{i,j} b_j$$

and note that the absolute value of this sum must lie on the unit circle whether each

term is negated or not. A routine verification confirms that the only possibilities are of the form $U_i = (1, 0, \ldots, 0)$ and $U_i = \frac{1}{\sqrt{2}}(1, i, \ldots, 0)$ up to permutation of the entries and overall phase. But the second of these can only appear in a unitary matrix as part of a block $\begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix}$ whose permanent is zero, thus making $\mathrm{Per}(U) = 0$. So, U must only consist of rows with one nonzero entry each, and therefore must have the form described. □

Unitary permanent nearly 1

We now prove a stronger approximate version that applies to unitary matrices with $|\mathrm{Per}(U)| \geq \sqrt{1 - 1/e} + c$ for some constant c. Such matrices must have special identity-like structure in having a large-norm entry in every row.

Theorem 58. *If U is a unitary matrix with $|Perm(U)|^2 = 1 - \epsilon$, then every column j of U has a large entry U_{ij} satisfying $|U_{ij}|^2 \geq 1 - e\epsilon$.*

Proof. Let p_U be the polynomial in formal variables $\vec{x} = (x_1, \ldots, x_n)$ given by

$$p_U(\vec{x}) = U[x_1 x_2 \ldots x_n] = U[x_1] U[x_2] \ldots U[x_n]$$

We may express p_U in the Fock basis as

$$p_U(\vec{x}) = \sum_S \tfrac{1}{S!} \mathrm{Perm}(U_S) \vec{x}^S$$

Let $\alpha = \mathrm{Perm}(U)$, where $|\alpha| = \sqrt{1 - \epsilon}$. Then, α is the coefficient of the monomial $x_1 \ldots x_n$ in p_U. Call this monomial $\vec{x}^{S_0}$. We can then split p_U into this monomial and the remainder

$$p_U(\vec{x}) = \alpha \vec{x}^{S_0} + \beta q(\vec{x})$$

with $\|q(\vec{x})\| = 1$ and $|\beta| = \sqrt{\epsilon}$ so that $\|p_U(\vec{x})\| = |\alpha|^2 + |\beta|^2 = 1$.

Now, consider the effect of setting one of the formal variables $(x_1, \ldots, x_n)$ to 0. Our idea is as follows: When any variable is made 0, the term $\alpha \vec{x}^{S_0}$ vanishes, leaving

$p_U(\vec{x})$ with a term of small Fock norm ϵ, which we show can only happen for U functionally close to the identity.

For ease of notation, say x_1 is the variable to be zeroed. Let $\vec{x}' = (0, x_2 \ldots, x_n)$. Note that $q(\vec{x}')$ is obtained from $q(\vec{x})$ simply by removing all monomial terms containing x_1, so $\|q(\vec{x}')\| \leq \|q(\vec{x})\| = 1$. So, from

$$p_U(\vec{x}') = \beta q(\vec{x}'),$$

we obtain

$$\|p_U(\vec{x}')\|^2 = |\beta|^2 \|q(\vec{x}')\|^2 \leq |\beta|^2 = \epsilon \tag{3.1}$$

Now, let's look at $\|p_U(\vec{x}')\|^2$. We note that $\vec{x}'=(I - |e_1\rangle\langle e_1|)\vec{x}$, so $p_U(\vec{x}') = p_{U'}(\vec{x})$ with $U' = U(I - |e_1\rangle\langle e_1|)$. Therefore,

$$\begin{aligned}
\|p_U(\vec{x}')\|^2 &= \|p_{U'}(\vec{x})\|^2 \\
&= \text{Perm}\left(U'(U')^*\right) \\
&= \text{Perm}\left(U(I - |e_1\rangle\langle e_1|)U^*\right) \\
&= \text{Perm}\left(I - U|e_1\rangle\langle e_1|U^*\right) \\
&= \text{Perm}\left(I - |v\rangle\langle v|\right), \qquad (3.2)
\end{aligned}$$

where in the last line we've written $|v\rangle$ for $U|e_1\rangle$, the first column of U.

So, from equation 3.1 and 3.2, we have

$$\text{Perm}\left(I - |v\rangle\langle v|\right) \leq \epsilon$$

Now, we apply the analogue of Hadamard's Theorem for permanents, proved in [32] that if M is a Hermitian positive semi-definite matrix, then

$$\text{Perm}(M) \geq \prod_i M_{ii}.$$

Since $I - |v\rangle\langle v|$ is clearly Hermitian positive semi-definite, the theorem applies to

show

$$\text{Perm}\left(I - |v\rangle\langle v|\right) \geq \prod_i \left(1 - |v_i|^2\right)$$

and so we have

$$\prod_i \left(1 - |v_i|^2\right) \leq \epsilon$$

We now show that for ϵ to be small, some $|v_i|$ must nearly be 1. Let

$$\delta = \min\left(1 - |v_1|^2, \ldots, 1 - |v_n|^2\right)$$

. We now compute the minimum possible value of $\prod_i \left(1 - |v_i|^2\right)$ subject to this constraint and the constraint that $\sum_i |v_i|^2 = 1$ from v being a column of unitary matrix.

If two of these terms $1 - |v_i|^2$ equal neither the minimum value of δ nor the maximum value of 1, we could achieve a smaller product by increasing one and decreasing the other an equal amount. So, to minimize the product, we must make all but one value extremal.

Suppose we have k values with $|v_i|^2 = 1 - \delta$, one non-extremal value with $|v_j|^2 = a \in [0, 1 - \delta]$, and the rest have $|v_i|^2 = 0$. This gives a product of

$$\prod_i \left(1 - |v_i|^2\right) = (1 - a)\,\delta$$

We observe that $\sum_i |v_i|^2 = 1$ implies $k(1 - \delta) \leq 1$, so $\delta \geq 1 - \frac{1}{k}$, and moreover that $a \in [0, 1 - \delta]$ implies $1 - a \geq \delta$.

$$\begin{aligned}
\prod_i \left(1 - |v_i|^2\right) &= (1 - a)\,\delta^k \\
&\geq \delta\left(1 - \frac{1}{k}\right)^k \\
&\geq \delta e^{-1}
\end{aligned}$$

So, we've found that for a given $\delta = \min\left(1 - |v_1|^2, \ldots, 1 - |v_n|^2\right)$, the lowest product $\prod_i \left(1 - |v_i|^2\right)$ we can achieve is still at least δe^{-1}. Therefore, to satisfy $\prod_i \left(1 - |v_i|^2\right) \leq$

ϵ, we must have $\delta \leq e\epsilon$, so there is some entry v_i of v with $|v_i|^2 \geq 1 - e\epsilon$. Recall that v was the first column of U, and we could repeat this proof for any column to it has such a large entry, which is the result desired. □

Corollary 59. *If U is a Hadamard matrix, meaning that it is unitary and its entries all have norm $1/\sqrt{n}$, then $|Perm(U)|^2 \leq \frac{1}{e}\left(1 - \frac{1}{n}\right)$.*

3.2.3 Unitary permanent $1/\operatorname{poly}$

Stronger results about the permanents of unitary matrices are proven by Aaronson and Nguyen [3] and later improved by Berkowitz and Devlin [7] and by Nguyen [33]. Roughly, they say if that M is a matrix has $\|M\|_2 \leq 1$ and $\mathrm{Per}(M) \geq n^{-C}$ for some constant C, then the matrix must be close to equivalent to the identity matrix in that a significant fraction of the rows of M contain entries that have norm nearly 1. Such matrices can clearly be distinguished from random matrices.

A connection is drawn from orthogonal matrices with high permanent to near-isometries of the hypercube. That is, Euclidian rotations of the hypercube that map a significant fraction of its vertices to other vertices. Clearly any series of right-angle rotations and reflections suffices to do so by preserving the cube's shape, but these correspond to getting a permanent of 1 trivially from a permutation matrix. It is shown that any near-isometry must be close to such a trivial operation.

The translation of the question of unitary permanents to one about rotations of the hypercube is a surprising shift from an algebraic problem to a geometric one. We push these equivalences yet further by recasting it as a combinatorial problem regarding graphs of Hamming distances, and of low-rank decompositions into ± 1-valued matrices. Perhaps these equivalences will allow alternate methods of attack on the problem of submatrices with large permanents. Conversely, it may allow bounds on the permanent to translate to results in other areas. The rest of this section is devoted to developing these equivalences.

Hamming graphs

We start by defining and proving some basic properties about Hamming graphs, which we will then connect to near-isometries of the hypercube.

Definition 60. *For an ordered collection S of binary strings $s_1, \ldots, s_k$ of the same length n, let the Hamming graph G_S give the pairwise Hamming distances between strings*

$$G_S(i,j) = d_H(s_i, s_j)$$

We would like to consider collections S and T with equal Hamming graphs.

Definition 61. *Call two collections S and T* Hamming clones *if they have equal Hamming graphs $G_S = G_T$.*

Note that Hamming clones necessarily have equal numbers of strings. We will also assume they have equal string lengths n.

Of course any S is trivially a Hamming clone to itself. We will be interested in Hamming clones that are not obviously equal, so we define a set of operations that obviously preserve Hamming graph.

Definition 62. *Two Hamming clones are* trivially equal *if they can be obtained by the following operations:*

1. *Permuting the bits in each string by a fixed permutation of the indices*
2. *Complementing the bits in every string corresponding to a fixed subset of the indices*

We will show that this notion of trivial equality has a natural geometric interpretation.

Equivalence of Hamming clones and cube isometries

Now we develop the connection between Hamming graphs and near-isometries of the hypercube. At the heart of this is the one-to-one correspondence between the Hamming distance d_H and Euclidian distance d_E for points in the hypercube.

While we ordinarily think of binary strings over the alphabet $\{0, 1\}$, here we will think of them as points in the hypercube $C_n = \{-1, +1\}^n$. The trivial equivalences of strings are then symmetries of the cube, i.e. rotations and reflections that map the hypercube into itself.

Proposition 63. *Two Hamming clones S and T are trivially equal if they can be mapped to each other (as ordered collection) by an automorphism of the hypercube.*

Proof. The trivial operations for collections of strings are exactly the generators for isometries of the hypercube: Permuting bits correspond to permutations of the coordinates, and complementing bits to reflections over an axis. □

Proposition 64. *S and T are Hamming clones if and only there is an orthogonal matrix M mapping one to the other, i.e. $Ms_i = t_i$ for each i.*

Proof. First, note that the Hamming distance and Euclidian distance are related by $d_E = 2\sqrt{d_H}$. This is a one-to-one map, so equal Hamming graphs correspond exactly to equal pairwise Euclidian distances.

The reverse direction is easy. Since M is orthogonal, it preserves Euclidian distances, and thus Hamming distances.

For the forward direction, a standard result in rigidity theory states that the pairwise Euclidian distances between points specify the set of points globally rigidly, which means up to affine orthonormal transformations $t_i = Ms_i + c$. To show further that such a transformation maps $\vec{0}$ to $\vec{0}$ and is therefore simply orthogonal, extend the sets S and T to contains the negation of every vector in them. Then, their respective pairwise Hamming distances remain equal, as the Hamming distances for complements are fixed as $d_H(a, -b) = n - d_H(a, b)$ and $d_H(-a, -b) = d_H(a, b)$. Since the sum of the average of each set of vectors is $\vec{0}$, it follows the translation constant c must be $\vec{0}$ as well. □

So, non-trivial Hamming clones correspond to ordered subsets of the hypercube that are related by an orthogonal transformation that is not an automorphism of the hypercube. Such a subset that contains a large fraction of the 2^n strings therefore

corresponds to an orthogonal map that maps a large fraction of hypercube points C_n to hypercube points.

Equivalence of Hamming clones and low-rank decompositions

We now push the equivalence further to an interpretation of Hamming clones in terms of low-rank decompositions. We can equate a collection S with a matrix whose rows are the strings in order. Such a matrix is $K \times n$, where n is the length of each string.

Proposition 65. *Two collections S and T have equal Hamming graphs $G_S = G_T$ if and only the corresponding matrices satisfy $SS^t = TT^t$.*

Proof. The (i, j) entry if SS^t is the dot product of strings S_i and S_j. Since each string has fixed norm $\sqrt{n}$ as a vector in $\{-1, +1\}^n$, the inner product $S_i \cdot S_j$ is specified by the Euclidian distance $D_E(s_i, s_j)$. The corresponding distances are equal for S and T, so all entries of SS^t equal those of TT^t. □

Now, we can consider the expression SS^t be a sum of rank-one terms corresponding to the columns c_S of S.

$$SS^t = \sum_{v \in c_S} vv^t$$

Note that SS^t is a $K \times K$ matrix of which a rank-n decomposition is given. In the typical case that we consider where the number of strings K is much greater than the length n, this is a low-rank decomposition. Furthermore, because the c_S only contain entries in ± 1, the same is true for the rank-one summands.

The rank-n decomposition is trivially unaffected by the transformations on a set of strings that trivially don't affect their Hamming graph, which are equivalently the automorphisms of the hypercube C_n. This is evidence that it naturally captures the notion of a Hamming graph.

- Permuting the string indices corresponds to reordering the columns c_S, which clearly doesn't affect their summation.
- Complementing any index flips the column $v \to -v$, which doesn't affect its summand vv^t.

We thus call two rank-n decompositions *nontrivially distinct* if they contain different multisets of rank-one matrices. So, Hamming clones that are not related correspond to distinct rank-one decompositions.

Proposition 66. *Any pair of Hamming clones S and T with K strings each corresponds to two nontrivially distinct rank-n decompositions of a $K \times K$ matrix into n rank-one matrices with entries in ± 1.*

So, the problem of finding large-size Hamming clones that are not trivially equivalent is turned into one of finding non-unique low-rank ± 1-valued decompositions where the rank n is much smaller the matrix size K.

3.3 Collision statistics

This section is based on work with Greg Kuperberg in The bosonic birthday paradox [6]

We look at probabilities of collisions for Haar-random network matrices. We show that they are efficiently computable, and so may be used for statistical verification.

A *collision* is an instance of 2 or more photons being output in the same mode. For classical particles being uniformly independently distributed across m modes, one needs $n \sim \sqrt{2m \ln 2}$ particles until there's a better-than-even chance of at least one collision. This is popularly known as the birthday paradox, that a room with 23 people has two that share a birthday more likely than not.

For identical bosons, the analogous boundary is $n \sim \sqrt{m \ln 2}$, a factor of $\sqrt{2}$ less. This would correspond to needing 16 people until two share a birthday more likely than not. Identical bosons tend to "clump", i.e. appear in the same state more than classically expected. In contrast, identical fermions never experience a collision. The analogue of uniformly distributing bosons across modes is to apply Haar-random uniform network matrix in a BOSONSAMPLING setup. This produces the same maximally-mixed photonic state regardless of the initial state of the bosons. The probabilities for such a state are governed by Bose-Einstein statistics. They

assign an equal probability to each possible output partition, without any correction for multiplicity as in the classical case. For example, two bosons in two different specified modes are equally likely to two in a single specified mode, not twice as likely as classically.

More strongly, we may consider k-fold collisions of k bosons in one mode. We express the asymptotic distribution of k-fold collisions, finding it to be a Poisson distribution.

Theorem 67. *Suppose that there are n photons with m allowed single-particle modes, suppose that they are in the uniform state ρ_{unif} produced by applying a Haar-random unitary matrix, and suppose that $n \sim cm^{(k-1)/k}$ as $n \to \infty$, for some integer $k \geq 2$ and some constant $c > 0$. Then the number of k-fold birthdays converges in distribution to a Poisson random variable with mean c^k, while the number of $(k+1)$-fold-or-more birthdays converges to 0.*

Verifying these statistics for random network matrices confirms that it exhibits the expected identical-boson behavior, letting us distinguish it from one where the particles are classical or distinguishable. Though it does not check for behavior specific to BosonSampling, it does check that the particles exhibit fundamentally bosonic statistical properties.

3.4 Row norm statistics

This section is based on work with Scott Aaronson in BosonSampling Is Far From Uniform [2] .

Row norm verification is based on a simple idea: a larger-norm entry of the network matrix is likelier to cause a photon to be detected in its corresponding output mode. So, rows of the $m \times n$ network matrix with higher norm tend to output more photons on average. Indeed, scaling any row of a matrix scales its permanent in proportion, so multiplying a row by a constant c increases the probability of a photon in that mode by $|c|^2$, if we ignore the possibility of multiple photons and the violation of unitarity.

We define the output statistic R^* as follows.

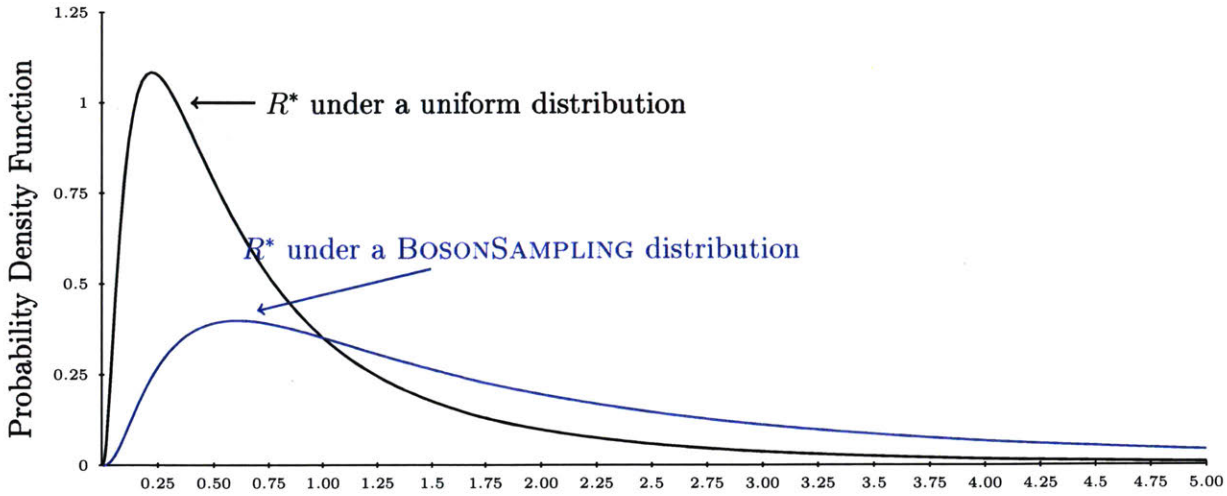

Figure 3-1: Probability density functions for the row-norm estimator $R^*(S)$, when S is drawn either the uniform distribution $\mathcal{U}$ or a Haar-random BosonSampling distribution $\mathcal{D}_M$, in the limits $n \to \infty$ and $m/n \to \infty$. Observe that R^* is typically larger on $\mathcal{D}_M$ than on $\mathcal{U}$. In particular, there's a larger probability of the event E that $R^* \geq 1$

Definition 68. *Let M be the $m \times n$ network matrix truncated to the first n columns, as these are the entries relevant to transitions from the initial state. Let $R_i = \|M_i\|$ be the squared 2-norm of row i of M. Finally, define R^* to be the product of the row R_i corresponding to each photon measured in the output, scaled down by n^n so as to have mean 1. Letting S be the vector of output counts with s_i photons detected in mode i, this is*

$$R^*(S) = \frac{1}{n^n} \prod_{i=1}^{m} R_i^{s_i}.$$

We compare the distribution of $R^*(S)$ on the BosonSampling output distribution $\mathcal{D}_M$ and a uniform distribution over outcomes $\mathcal{U}$. Their respective density functions are plotted in fig. 3-1. These density curves are visibly different, and we will use this to distinguish $\mathcal{D}_M$ from $\mathcal{U}$. Observe that $\mathcal{D}_M$ produces larger R^* on average, corresponding to the observation that larger-norm rows are more likely than chance to omit a photon. For the Haar-uniform distribution $\mathcal{U}$, R^* is a product of independent chi-squared that can be shown to converge to a log-normal distribution.

Now, let $E(S)$ be the event that $R^*(S) \geq 1$. Our main result is that the probability of E differs significantly on $\mathcal{D}_M$ versus $\mathcal{U}$. This lets us distinguish a true

BOSONSAMPLING device from a very simple forgery that produces every outcome with equal likelihood.

For both $\mathcal{D}_M$ and $\mathcal{U}$, we ignore all outcomes that contain any collision. We make collisions unlikely by requiring many more modes than photons: $m \geq n^{5.1}/\delta$.

Theorem 69. *With probability* $1 - O(\delta)$ *over Haar-uniform* M,

$$\Pr_{S \sim \mathcal{D}_M}[E(S)] - \Pr_{S \sim \mathcal{U}}[E(S)] \geq \frac{1}{9}$$

where $\mathcal{D}_M$ *is* BOSONSAMPLING *distribution on* M *and* $\mathcal{U}$ *is uniform over outcomes, both post-selected on no collisions and with* $m \geq n^{5.1}/\delta$.

Estimating the empirical probability of E on polynomially-many runs of the device lets us distinguish BOSONSAMPLING from uniform output with high probability.

3.5 Fourier matrix suppression

We discuss here the idea of certification via Fourier matrix suppression as implemented in [50]. Using a network matrix with a specific highly-symmetric structure, the Fourier matrix, any observed outcome satisfies a certain rule. Any potential outcomes that violate this rule are *suppressed*; they occur with probability 0. This lets us reject any purported BOSONSAMPLING device that ever produces a rule-violating outcome. This rejection is certain rather than merely statistical. However, this certification method has the drawback of applying only to a special structured network matrix, leaving open whether the BOSONSAMPLING device works correctly in general.

The Fourier matrix F is an $m \times m$ unitary matrix whose entries are given by

$$F_{jk} = \frac{1}{\sqrt{m}} \omega^{jk}$$

with $\omega = e^{i\frac{2\pi}{m}}$ the m^{th} root of unity. The BOSONSAMPLING distribution $\mathcal{D}_F$ is unusual in that only a small $\sim 1/m$ fraction of potential output vectors S have a nonzero probability of occurring. The non-suppressed outcomes satisfy a modular-sum law:

$$\sum_{j=0}^{m-1} j s_j \equiv 0 \pmod{m}$$

In other words, if each output photon is weighted by the index of its output mode, counting from 0, the total weight must be a multiple of m.

Theorem 70 ([49]). *For the Fourier matrix F, if outcome S has nonzero probability*

$$\Pr_{\mathcal{D}_F}[S] > 0,$$

then S satisfies

$$\sum_{j=0}^{m-1} j s_j \equiv 0 \pmod{m}.$$

This result is proved in greater generality in [49], but we give a new proof for this specific claim. We find this proof to be nicer and to make more explicit how the symmetries of the Fourier matrix cause the suppression.

Proof. Let S be an output count vector. Its probability of being output equals $|\mathrm{Per}\,(F_S)|^2$, where F_S is a submatrix of F with s_j copies of each row j. We'll show that this permanent must equal 0 unless S satisfies the suppression law.

We use two symmetries of the permanent

1. The permanent is invariant under any permutation of the columns.

2. Scaling any one row by a constant c scales the permanent by c.

By the first symmetry, if we move the first column of F_S to the end to produce F'_S, we still have $\mathrm{Per}(F_S) = \mathrm{Per}(F'_S)$. Let's observe the effect of this rotation on a row of F_S that was row j in F. This row $(1, \omega^j, \omega^{2j}, \ldots, \omega^{(m-1)j})$ is rotated to $(\omega^j, \omega^{2j}, \ldots, \omega^{(m-1)j}, 1)$, which has the effect of multiplying it entrywise by ω^j. Accumulating all these scalings applied to the rows of F_S to make F'_S, we have

$$\mathrm{Per}(F'_S) = \prod_{j=0}^{m-1} \omega^{j s_j} \mathrm{Per}\,(F_S)$$

Since we also have $\mathrm{Per}\,(F'_S) = \mathrm{Per}\,(F_S)$, unless the outcome is suppressed with $\mathrm{Per}\,(F_S) = 0$, we must have $\prod_{j=0}^{m-1} \omega^{js_j} = 1$. Since $\omega^k = 1$ only for $k \equiv 0 \pmod{m}$, this implies that $\sum_{j=0}^{m-1} js_j \equiv 0 \pmod{m}$ as claimed.

□

3.6 Linear statistics

We introduce a new statistic to be used for verification. It is obtained by assigning an integer weight to each mode and summing the weights of all photons in the output.

Definition 71. *The* linear statistic *with integer weights* $w = (w_1, \ldots, w_n)$ *of a* BosonSampling *outcome* $S = (s_1, \ldots, s_m)$ *is the integer combination* $w \cdot S = \sum_{j=1}^{m} w_j s_j$.

Note that the single-mode marginal of mode j is the linear statistic with weights given by the basis vector e_j. We will demonstrate a polynomial-time algorithm to compute the distribution of a linear statistic to arbitrarily high accuracy.

Theorem 72. *There exists a classical algorithm to approximate the distribution of* $w \cdot S$ *for* $S \sim \mathcal{D}_M$ *for integer vectors* w *to within variation distance* ϵ *with high probability. Its running time polynomial in* n, $\max |w_j|$, ϵ^{-1}.

This algorithm is closely tied to computing the moment generating function of the BosonSampling distribution, which we now discuss.

3.6.1 Moment generating functions

Recall that the moment generating function (MGF) $f_X(t)$ for a random variable X equals the expectation $\mathbb{E}\left[e^{tX}\right]$. This generalizes to vector-valued distributions like that of the output count vector $S = (s_1, s_2, \ldots, s_m)$ as

$$f_S(w_1, w_2, \ldots, w_m) = \mathbb{E}\left[e^{w_1 s_1} e^{w_2 s_2} \cdots e^{w_m s_m}\right] = \mathbb{E}\left[e^{w_1 s_1 + w_2 s_2 + \cdots + w_m s_m}\right].$$

If we take w as the vector of weights $(w_1, w_2, \ldots, w_m)$, we can write this more compactly via a dot product

$$f_S(w) = \mathbb{E}\left[e^{w \cdot S}\right].$$

We now express the MGF of the BOSONSAMPLING distribution $\mathcal{D}_M$.

Proposition 73. *The MGF of the* BOSONSAMPLING *distribution $\mathcal{D}_M$ is given by*

$$\mathbb{E}_{S \sim \mathcal{D}_M}\left[e^{w \cdot S}\right] = \mathrm{Per}\left(U^\dagger e^W U\right),$$

where $W = \mathrm{diag}(w)$ is the diagonal matrix of weights w_j.

Proof. We start by expressing the permanent $\mathrm{Per}(U^\dagger e^W U)$ in terms of Fock polynomials. We use a rewording of Corollary 22 of [1] for $m \times n$ matrices.

Proposition 74. *Let A and B be $m \times n$ matrices and $x^{[n]}$ be the monomial $x_1 x_2 \cdots x_n$ of the initial state of one photon in each of the first n modes. Then,*

$$\mathrm{Per}\left(A^\dagger B\right) = \left\langle A\left[x^{[n]}\right], B\left[x^{[n]}\right]\right\rangle$$

Applying this proposition with $A = U$, $B = e^W U$ lets us express the desired permanent as a Fock inner product

$$\mathrm{Per}\left(U^\dagger e^W U\right) = \left\langle U\left[x^{[n]}\right], \left(e^W U\right)\left[x^{[n]}\right]\right\rangle.$$

Now, expand the output state in the Fock basis

$$U\left[x^{[n]}\right] = \sum_S c_S x^S$$

with an amplitude c_S of each outcome S.

The RHS of the inner product is this state with the transformation e^W applied afterward. Since e^W is a diagonal matrix with entries e^{w_j}, it scales each formal variable

x_j into $e^{w_j} x_j$. Applying to x^S and grouping together the scaling factors, we get

$$e^W \left[x^S\right] = \left(\prod_{j=1}^{m} e^{w_j s_j}\right) x^S = e^{w \cdot S} x^S.$$

It follows that

$$\left(e^W U\right) \left[x^{[n]}\right] = \sum_S c_S e^{w \cdot S} x^S.$$

Then, taking the inner product in the Fock basis,

$$\mathrm{Per}\left(U^\dagger e^W U\right) = \left\langle U\left[x^{[n]}\right], \left(e^W U\right)\left[x^{[n]}\right]\right\rangle = \sum_S |c_S|^2 e^{w \cdot S},$$

Recalling that $|c_S|^2$ is the probability of outcome S, this is exactly the desired expectation $\mathbb{E}_{S \sim \mathcal{D}_M}\left[e^{w \cdot S}\right]$.

□

Now we recall Gurvits' additive permanent approximation algorithm.

Theorem 75 (Gurvits [22]). *There exists a classical algorithm to approximate the permanent of an $n \times n$ matrix M to within additive error $\pm\epsilon \|M\|^n$ in spectral norm with high probability with running time polynomial in n and ϵ^{-1}.*

We apply this result to computing the MGF of the permanent for imaginary-valued weight vectors w. In such a situation $U^\dagger e^W U$ has all its spectral values of the form $e^{2\pi i w_j}$ and so spectral has norm 1. Therefore, Gurvits' approximation algorithm let us compute $\mathrm{Per}\left(U^\dagger e^W U\right)$ to within $\pm\epsilon$. Note however that without restriction on w, the MGF $\mathrm{Per}\left(U^\dagger e^W U\right)$ is an arbitrary permanent and so #P-hard to compute.

Corollary 76. *There exists a classical algorithm to approximate the MGF of the* BosonSampling *distribution $\mathbb{E}\left[e^{w \cdot S}\right]$ of a matrix M at imaginary-valued w to within $\pm\epsilon$ additive error with high probability in running time polynomial in n and ϵ^{-1}.*

3.6.2 Linear statistics algorithm

Note the MGF of the BOSONSAMPLING distribution $\mathbb{E}\left[e^{w\cdot S}\right]$ is simply the single-variable MGF of the distribution of the linear statistic $w\cdot S$. We will use our algorithm for the MGF in corollary 76 to solve for the probability distribution of $w\cdot S$. Although corollary 76 applies to imaginary-valued weight vector and linear-statistic weights are integer-valued, this is easy fixed by multiplying the integer-valued weights by a fixed imaginary constant. So, learning the distribution of $\frac{2\pi i}{N}\, w\cdot S$ lets us easily recover that of $w\cdot S$.

Now we're ready to prove our result.

Theorem 77. *There exists a classical algorithm to approximate the distribution of $w\cdot S$ for $S\sim \mathcal{D}_M$ for integer vectors w to within variation distance ϵ with high probability. Its running time is polynomial in n, $\max|w_j|$, ϵ^{-1}.*

Proof. For any integer $N>0$, we may use corollary 76 to approximate the Fourier transform of the distribution of the linear statistic $S\cdot w$:

$$\mathbb{E}_{S\sim D_U}\left[\exp\left(\frac{2\pi i}{N}\, w\cdot S\right)\right]$$

We can expand the expectation as a weighted linear combination

$$\sum_{j=0}^{N-1}\Pr\left[w\cdot S\equiv j \pmod N\right]\exp\left(\frac{2\pi i}{N}j\right)$$

In order to avoid the distribution wrapping back on itself due to the $\pmod N$, we take large enough $N=2n\max|w_i|+1$ so that $|w\cdot s|<N/2$.

$$\sum_{j=-N}^{N}\Pr\left[w\cdot S=j\right]\exp\left(\frac{2\pi i}{N}j\right)$$

So, the expectations are a linear combination of the desired probabilities $\Pr\left[w\cdot S=j\right]$ with coefficients given by a Fourier matrix. Moreover, we can vary the "frequency" ω to ω^k to obtain a set of linear relations, one for each $k\in\{0,1,\ldots,n-1\}$:

$$\mathbb{E}_{S\sim D_U}\left[\exp\left(\frac{2\pi ik}{N}w\cdot S\right)\right]=\sum_{j=-N}^{N}\Pr\left[w\cdot S=j\right]\exp\left(\frac{2\pi ik}{N}j\right)$$

Write x_k for the expectation on the right-hand side and p_j for the desired probabilities $\Pr[w\cdot S=j]$. We have that the x_k and p_j are related by the Fourier matrix F with

$$F_{jk}=\frac{1}{\sqrt{N}}\exp\left(\frac{2\pi i}{N}jk\right)x$$

and so one may solve for the desired probabilities in terms of the efficiently computable expectations $p=N^{-1/2}F^{\dagger}x$ Approximating the expectations x_k as per corollary 76 thus lets us compute the desired distribution p.

It remains now to do an error analysis. Corollary 76 gives a high probability that each approximated expectation $\tilde{x}_k$ is within ϵ of its true value x_k, and so $\|\tilde{x}-x\|_1\leq N\epsilon$.

Then, the approximated probabilities have

$$\begin{aligned}\|\tilde{p}-p\|_1&\leq\sqrt{N}\,\|\tilde{p}-p\|_2\\&=\|F\tilde{x}-Fx\|_2\\&=\|\tilde{x}-x\|_2\\&\leq\|\tilde{x}-x\|_1\\&\leq N\epsilon\end{aligned}\tag{3.3}$$

By taking $\epsilon=\epsilon'/N=\epsilon'/\left(2n\max|w_j|+1\right)$, we obtain $|\tilde{p}-p|_1\leq\epsilon'$ while maintaining a running time polynomial in ϵ', n, and $\max|w_i|$.

□

Bibliography

[1] Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, STOC '11, pages 333–342. ACM, 2011.

[2] Scott Aaronson and Alex Arkhipov. BosonSampling is far from uniform. *Quantum Info. Comput.*, 14(15-16):1383–1423, November 2014.

[3] Scott Aaronson and Hoi Nguyen. Near invariance of the hypercube, 2014.

[4] Daniel S. Abrams and Seth Lloyd. Simulation of many-body Fermi systems on a universal quantum computer. *Phys. Rev. Lett.*, 79:2586–2589, 1997. quant-ph/9703054.

[5] Alex Arkhipov. Boson sampling is robust to small errors in the network matrix. 2014.

[6] Alex Arkhipov and Greg Kuperberg. The bosonic birthday paradox. 2012.

[7] R. Berkowitz and P. Devlin. A stability result using the matrix norm to bound the permanent. *ArXiv e-prints*, June 2016.

[8] E. Bernstein and U.V. Vazirani. Quantum complexity theory. 1997.

[9] S. Bhupatiraju. On the complexity of the marginal satisfiability problem. 2012.

[10] A. Broadbent, J. Fitzsimons, and E. Kashefi. Universal blind quantum computation. *ArXiv e-prints*, July 2008.

[11] Matthew A. Broome, Alessandro Fedrizzi, Saleh Rahimi-Keshari, Justin Dove, Scott Aaronson, Timothy Ralph, and Andrew G. White. Photonic boson sampling in a tunable circuit. 2012.

[12] Jin-Yi Cai, Aduri Pavan, and D. Sivakumar. On the hardness of permanent. In *Proc. Intl. Symp. on Theoretical Aspects of Computer Science (STACS)*, pages 90–99, 1999.

[13] Jacques Carolan, Chris Harrold, Chris Sparrow, Enrique Martin-Lopez, Nicholas J. Russell, Joshua W. Silverstone, Peter J. Shadbolt, Nobuyuki Matsuda, Manabu Oguma, Mikitaka Itoh, Graham D. Marshall, Mark G. Thompson,

Jonathan C. F. Matthews, Toshikazu Hashimoto, Jeremy L. O'Brien, and Anthony Laing. Universal linear optics. 2015.

[14] Kevin P. Costello and Van H. Vu. Concentration of random determinants and permanent estimators. *SIAM J. Discrete Math*, 23(3):1356–1371, 2009. arXiv:0905.1909.

[15] A. Crespi, R. Osellame, R. Ramponi, D. J. Brod, E. F. Galvao, N. Spagnolo, C. Vitelli, E. Maiorino, P. Mataloni, and F. Sciarrino. Experimental boson sampling in arbitrary integrated photonic circuits. 2012.

[16] E. H. Lieb E. Carlen, M. Loss. An Inequality of Hadamard Type for Permanents. *ArXiv e-prints*, 2014.

[17] N. Spagnolo et al. Experimental validation of photonic boson sampling. 2014.

[18] Richard P. Feynman. Simulating physics with computers. *Int. J. Theoretical Physics*, 21(6-7):467–488, 1982.

[19] Peter Gemmell, Richard Lipton, Ronitt Rubinfeld, Madhu Sudan, and Avi Wigderson. Self-testing/correcting for polynomials and for approximate functions. In *Proc. ACM STOC*, pages 32–42, 1991.

[20] Peter Gemmell and Madhu Sudan. Highly resilient correctors for polynomials. *Inform. Proc. Lett.*, 43:169–174, 1992.

[21] Daniel Gottesman, Alexei Kitaev, and John Preskill. Encoding a qubit in an oscillator. *Phys. Rev. A*, (64:012310), 2001. quant-ph/0008040.

[22] L. Gurvits. On the complexity of mixed discriminants and related problems. In *Mathematical Foundations of Computer Science*, pages 447–458, 2005.

[23] C. K. Hong, Z. Y. Ou, and L. Mandel. Measurement of subpicosecond time intervals between two photons by interference. *Physical Review Letters*, 59:2044–2046, November 1987.

[24] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. *J. ACM*, 51(4):671–697, 2004. Earlier version in STOC'2001.

[25] Gil Kalai and Guy Kindler. Gaussian noise sensitivity and BosonSampling, 2014.

[26] Emanuel Knill. Fermionic linear optics and matchgates. quant-ph/0108033, 2001.

[27] Emanuel Knill, Raymond Laflamme, and Gerard J. Milburn. A scheme for efficient quantum computation with linear optics. *Nature*, 409:46–52, 2001. See also quant-ph/0006088.

[28] Joseph M. Landsberg and Nicolas Ressayre. Permanent v. determinant: an exponential lower bound assuming symmetry and a potential path towards Valiant's conjecture, 2015.

[29] R. Latala. Some estimates of norms of random matrices. 2005.

[30] Anthony Leverrier and Raúl García-Patrón. Analysis of circuit imperfections in BosonSampling. 2013.

[31] Richard J. Lipton. New directions in testing. In *Distributed Computing and Cryptography*, pages 191–202. AMS, 1991.

[32] M. Marcus. The permanent analogue of the Hadamard determinant theorem. 1963.

[33] Hoi Nguyen. On matrices of large permanent. personal communication, 2016.

[34] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition.* Cambridge University Press, New York, NY, USA, 10th edition, 2011.

[35] T. C. Ralph P. P. Rohde. Error tolerance of the boson-sampling model for linear optics quantum computing. 2012.

[36] Saleh Rahimi-Keshari, Timothy C. Ralph, and Carlton M. Caves. Sufficient conditions for efficient classical simulation of quantum optics. 2015.

[37] Michael Reck, Anton Zeilinger, Herbert J. Bernstein, and Philip Bertani. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.*, 73(1):58–61, 1994.

[38] Stefan Scheel. Permanents in linear optical networks. quant-ph/0406127, 2004.

[39] A. Shamir. IP = PSPACE. 1992.

[40] V. S. Shchesnovich. Conditions for an experimental Boson-sampling computer to disprove the Extended Church-Turing Thesis. *ArXiv e-prints*, March 2014.

[41] V. S. Shchesnovich. Boson-sampling with non-interacting fermions. *International Journal of Quantum Information*, 13:1550013, March 2015.

[42] V. S. Shchesnovich. Tight bound on the trace distance between a realistic device with partially indistinguishable bosons and the ideal BosonSampling. 2015.

[43] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, 1997. Earlier version in IEEE FOCS 1994. quant-ph/9508027.

[44] M Sipser. A complexity theoretic approach to randomness. 1963.

[45] Justin B. Spring, Benjamin J. Metcalf, Peter C. Humphreys, W. Steven Kolthammer, Xian-Min Jin, Marco Barbieri, Animesh Datta, Nicholas Thomas-Peter, Nathan K. Langford, Dmytro Kundys, James C. Gates, Brian J. Smith, Peter G. R. Smith, and Ian A. Walmsley. Boson sampling on a photonic chip. 2012.

[46] Larry J. Stockmeyer. The complexity of approximate counting. In *Proc. ACM STOC*, pages 118–126, 1983.

[47] Terence Tao and Van H. Vu. On the permanent of random Bernoulli matrices. *Advances in Mathematics*, 220(3):657–669, 2009. arXiv:0804.2362.

[48] Barbara M. Terhal and David P. DiVincenzo. Classical simulation of noninteracting-fermion quantum circuits. *Phys. Rev. A*, 65(032325), 2002. quant-ph/0108010.

[49] M. C. Tichy. Interference of identical particles from entanglement to boson-sampling. *Journal of Physics B Atomic Molecular Physics*, 47(10):103001, May 2014.

[50] Malte C. Tichy, Klaus Mayer, Andreas Buchleitner, and Klaus Mølmer. Stringent and efficient assessment of boson-sampling devices. 2013.

[51] Max Tillmann, Borivoje Dakic, Rene Heilmann, Stefan Nolte, Alexander Szameit, and Philip Walther. Experimental boson sampling. 2012.

[52] Seinosuke Toda. PP is as hard as the polynomial-time hierarchy. *SIAM J. Comput.*, 20(5):865–877, 1991.

[53] Leslie G. Valiant. The complexity of computing the permanent. *Theoretical Comput. Sci.*, 8(2):189–201, 1979.

[54] Leslie G. Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM J. Comput.*, 31(4):1229–1254, 2002. Earlier version in STOC'2001.

[55] T. Vidick and J. Watrous. Quantum proofs. *ArXiv e-prints*, October 2016.

[56] Hui Wang, Yu He, Yu-Huai Li, Zu-En Su, Bo Li, He-Liang Huang, Xing Ding, Ming-Cheng Chen, Chang Liu, Jian Qin, Jin-Peng Li, Yu-Ming He, Christian Schneider, Martin Kamp, Cheng-Zhi Peng, Sven Hofling, Chao-Yang Lu, and Jian-Wei Pan. High-efficiency multiphoton boson sampling. 2017.

[57] C. Xu. Physically realistic formulations of BosonSampling under photon loss or partial distinguishability. Manuscript by private communication, 2013.