Gaussian Boson Sampling Complexity

Andrew Tanggara

Thesis presented for the degree of
Master of Computing (Advanced)

Research School of Computer Science
College of Engineering and Computer Science

and

Center for Quantum Computation and Communication Technology
Research School of Physics
College of Science

Australian National University

Except where otherwise indicated, this thesis is my own original work.

Andrew Tanggara
February 18, 2021

Acknowledgments

This project was initiated by Syed Assad's idea on probing the complexity of the Gaussian Boson Sampling (GBS) quantum computational model by simulating it using a classical algorithm known to be able to reconstruct any quantum states. I knew next to nothing about quantum mechanics, let alone about Gaussian Boson Sampling, at that time. Learning all the intriguing oddities of quantum mechanics, its physical interpretations as well as its mathematics, surely lived up to what I superficially know and heard about quantum mechanics before I started this project. But once things started to make a bit more sense, it surely becomes even more exciting to discover that there are various ways one can probe how a quantum-mechanical computation constitutes a process that is can or cannot be simulated by a classical computation. The whole learning and exploring did not go without errors and persistent mind-bending exercises that sometimes took me days for me to convince myself. This extent of learning about quantum computation and GBS to this point would be impossible to do all on my own. Discussions with Assad and colleagues in the CQC2T group at DQS both have an indispensable contribution towards this learning progression. The supportive learning environment provided by everyone in the group has also been a tremendous motivation, especially in the times I have been working on this project remotely.

Thus, I would like to express my humble gratitude to everyone in the group for having me as part of the group throughout this project. Especially, I would like to attribute Ping Koy as the group leader for creating and guiding an invaluable environment for learning and scientific discussions, Spyros for initially organizing the quantum information discussions and teaching the fundamentals of quantum information, and Hao, Lorcan, Biveen, Ozlem, and Ciaron for bearing with me in discussing quantum information/mechanics concepts that puzzles me. Moreover, my utmost gratitude to Assad for his supervision throughout this project, and his patience and insights in guiding me in learning the physical intuitions, formulating my ideas, and presenting the results in this thesis. In the process of finalizing this thesis, I would also like to sincerely thank Vijey for providing better ways of presenting the results and proofreading most of this thesis contents, Lorcan for providing proofreading and suggestions for the quantum mechanics preliminary, and Dirk for the inputs that produces more rigorous and precise presentations of the complexity theory and computability concepts. Lastly, an inexpressible gratitude to my parents and sister for always supporting my aspiration and efforts throughout my graduate studies to have a career as a research scientist.

Abstract

It remains questionable whether all physically realizable computational model can be simulated by currently available digital computer in an efficient manner. The theory of Quantum Mechanics offers an answer to this question through one of its branches that intersects with Computer Science, known as Quantum Computing. It claims that current computers, which is operating based on classical mechanics cannot efficiently simulate quantum mechanical process. Under this claim, a computer that is operating based on quantum mechanics holds more computational power than classical computers, promising a possibility to computationally solve problems that are classically intractable.

To show that this claim holds, numerous quantum computational models have been proposed and analyzed through the lens of Computational Complexity Theory, providing a rigorous notion of efficiency. Models that are more feasible experimentally, although not necessarily universal, have also been of a particular interests given the numerous complicated challenges surrounding the experimental implementation of a universal quantum computer. One of this non-universal models is called Gaussian Boson Sampling (GBS), which embraces the methods of Quantum Optics by sampling the number of photon outputs from a linear optical network as a means to process quantum information. Compared to its prototype model, Boson Sampling, GBS is more experimentally feasible due to its method of generating photon that is based on the Continuous Variable (CV) paradigm of Quantum Information. Nevertheless as with Boson Sampling, it was also shown that classical simulation of GBS cannot be performed efficiently under some highly plausible conjectures.

In the work presented in this thesis, we probe the complexity of classically constructing the probability distribution induced by GBS using the method of Quantum Homodyne Tomography. Instead of photon number measurement, this method further embraces the CV paradigm in the measurement of GBS by using Homodyne measurements that outputs real numbers. Sample data obtained from the measurement outcomes is then classically processed using Pattern Functions, which we use to approximate the probability distribution of GBS. A series of computer simulations that approximates some probability values of GBS are presented here, which shows that the number of samples needed to decrease the error of the approximation increases rapidly for more accurate approximation. We have derived a bound for the approximation error which can be achieved given that the sampled data satisfies a certain statistical property. This statistical property can eventually be satisfied as the number of Homodyne samples increases.

Having shown another evidence for the complexity of approximating GBS using Homodyne Tomography, we further investigate the nature of the complexity of GBS by analyzing its classical-quantum computational boundary in the spectrum of amount of noise present in

a GBS setup. In doing this, we built our discussions on the measure of non-classicality using the presence of negative values in the Quasi-probability distribution (QPD) representation of states of a quantum system. We then discuss existing results that shows the possibility of efficient classical simulation of a quantum computational model given a sufficient amount of noise. These results impose a necessary criteria of the minimum amount of noise that an experiment has to surpass in order for it to be impossible to simulate classically in an efficient manner. We extend this idea for GBS, where we propose a modified GBS where we consider noise in its measurements that opens a way for a possible formulation of a noise condition that a GBS experiment has to fulfill to rule out possible efficient classical simulation.

To further probe the classical-quantum computational boundary, we also propose a different quantum computational model called the Photon-added Thermal State CV-Sampling (PATS CVS), which can be formulated similarly to our noisy GBS model. We then discuss existing results that analyzes the negativity of the QPD representation of a noisy Photon-added Thermal state (PATS), which constitutes the inputs for PATS CVS. Using the result that shows the quantum-classical transition in terms of the QPD negativity for a PATS state when the amount of noise changes, we discuss the potential of PATS CVS model as a possible avenue to further probe the transition between efficiently simulable and non-efficiently simulable quantum mechanical experiment.

Contents

List of Figures

List of Tables

Notations and Terminologies

Terminologies, and general mathematical notations and definitions will be listed below for readers' convenience.

- Space of Binary Strings $\mathbb{B}^* = \{0,1\}^*$. The length of a binary string $x \in \mathbb{B}^*$ is $|x|$.
- Space of Binary Sequences $\mathbb{B}^{\mathbb{N}} = \{0,1\}^{\infty}$.
- Set of natural numbers $\mathbb{N} = \mathbb{Z}_+ \cup \{0\} = \{0, 1, 2, \dots\}$
- For a set A, number of elements in a set is $|A|$ (e.g: $|\mathbb{B}^n| = 2^n$), and A^c the complement of A.
- Combination $\binom{m}{n} = \frac{m!}{n!(m-n)!}$
- Trigonometric complex exponential $e^{i\theta} = \cos\theta + i\sin\theta$
- For $j \leq k$, $y \in \mathbb{B}^j$, and $x \in \mathbb{B}^k$, if y is the first j bits of x, we write that y is a substring of x or $y \preceq x$. If x is the first j bits of a longer string or a sequence, then we write x^*.
- For $x \in \mathbb{C}$ such that $x = a + bi$ where $a, b \in \mathbb{R}$, then $|x| = \sqrt{a^2 + b^2}$ is the *modulus* of x. The real part of x is indicated by $\mathbf{Re}(x) = a$ and its imaginary part by $\mathbf{Im}(x) = b$.
- We write a matrix $X \in \mathbb{C}^{m,n}$ with its i-th row, j-th column entry $x_{i,j}$ (thus all matrices will be written in capital letters and its entry in small letters with subscripts of its row and column position).

  We write I_m the m dimensional identity matrix or operator, where the subscript is omitted when its dimension is obvious from the context for notational simplicity.
- Kronecker Delta function $\delta_{i,j} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$.
- Dirac Delta function $\delta(z) = \begin{cases} \infty, & \text{if } z = 0 \\ 0, & \text{otherwise} \end{cases}$ for $z \in \mathbb{R}$, such that $\int_{\mathbb{R}} \delta(z)\, dz = 1$.
- For a linear operator or a matrix, we either write in capital letters, such as X or lower-case letters with hat, such as $\widehat{x}$.

- We write sets also in capital letters, distiction between sets and operators or matrices will be made clear from the context.
- For some polynomial of z, we write $\text{poly}(z)$. i.e: $\text{poly}(z) = \sum_{j\in\mathbb{N}} c_j z^j$ for some real or complex numbers $c_0, c_1, \ldots$.
- Probability measure $\mathbb{P} : A \to [0, 1]$ for some (discrete or continuous) set A.
- For independent and identically distributed (i.i.d) random variables $x, y \in \mathbb{R}$, the probability of x and y occurring is $\mathbb{P}(x, y) = \mathbb{P}(x)\ \mathbb{P}(y)$. As for independent events $X, Y \subseteq A$ for some set A, the probability of intersection of X and Y where both events X is $\mathbb{P}(X \cap Y) = \mathbb{P}(X)\,\mathbb{P}(Y)$.
- Big-O notation $\mathcal{O}$. For a real valued function $f : \mathbb{R} \to \mathbb{R}$, if $f(x) = \mathcal{O}(g(x))$, then $\exists c \in \mathbb{N}$ and $\exists x_0 \in X$ such that $\forall x \geq x_0, |f(x)| \leq cg(x)$.

Chapter 1

Introduction

"Machine should work. People should think."

— An excerpt from "Paperwork Explosion" (1967)

The mathematically-motivated notion of computation as a mechanical procedure to formulate a proof of a statement within a formal axiomatic system has been of interests since the 1930s. This is when computational models such as recursive function, Turing Machine, and lambda calculus were proposed by mathematicians Kurt Gödel, Alan Turing, and Alonzo Church, respectively. These three models had since been proven to hold the an equivalent computational power in the sense that, for any statement in a formal axiomatic system that can be proven using a procedure in one model, it is also provable in the other. This gives rise to the Church-Turing thesis, which informally states that any statement provable by these models can be proven by a human mathematician with pens and papers[1].

Since then, these notion of computation has been further exploited beyond abstract mathematical problems to solve problems of more practical concerns with resource constrains. Disciplines in natural sciences has been enjoying this practical applications of computation due to its mechanical nature in solving problems. However, it is the application of computation to model and simulate quantum physics that has given rise to the question whether there exist a physically realizable computation model that can solve some problems more efficiently than the models considered in the Church-Turing thesis.

In a lecture on computationally simulating a quantum system, distinguished physicist Richard Feynman pointed out the difficulties that a classical Church-Turing computer will encounter in simulating a quantum system that can be trivially simulated by a computer that runs on quantum mechanical system[2] [Fey82]. With the formulation of the computation models that harnesses quantum mechanical process by Benioff [Ben80] and later by Deutch [Deu85], the notion of *quantum computation* was born. Not long after, it was proven that there are quantum computational algorithms that can solve problems more efficiently than the best algorithm in the classical sense, most notably are the quantum factoring algorithm

[1] Sometimes, *lots* of them.

[2] At this point, it may be convenient for the reader to interpret "classical" computers as the digital computers as we know it today, whereas a "Quantum" computer will be loosely defined as a computer that exploits the properties of quantum mechanics to perform computation. We will later formalize the notion of computation using Turing Machines and Circuits in [**Chapter 3**].

by Peter Shor [Sho94] and quantum search algorithm by Lov Grover [Gro96]. Along with these advances, Feynman's vision of a quantum computer that outperforms its classical counterpart has since been of a deep scientific, engineering, and technological interests. The task of simulating a quantum system striking physicists and computer scientists alike as an intractable for a classical computer to perform.

Bernstein and Vazirani later in [BV97] rigorously formulated its hardness using Computational Complexity Theory in the form of the class of problems that can be solved with high probability by a universal quantum computer in a polynomial time with bounded error. This class of problems forms a strong theoretical foundation for quantum computation that is known as the **BQP** class. In [BV97], **BQP** class also plays a crucial role in formulating an analog of the Church-Turing thesis in terms of mechanical simulability of any physically realizable computer, the *Complexity-theoretic Church-Turing Thesis*[3]. As opposed to the Church-Turing thesis, it states that all computation model that can be physically realized can simulate each other only with an amount of computation time overhead polynomial in its input size. As it is widely believed that there are problems in **BQP** that cannot be solved classically in polynomial time, an experimental demonstration using a fully working quantum computer is therefore necessary to refute this thesis.

However, it turns out that a physical implementation of quantum computer that can solve useful problems in **BQP** has been persistently hard to construct. The task of experimentally demonstrating a quantum computer's ability to solve a problem that cannot be efficiently computed classically has been thoroughly defined and referred to by John Preskill as "*Quantum Computational Supremacy*" [Pre12] (and which experimental criteria from a Complexity Theoretic perspective listed by Aaronson and Chen [AC16]). Despite the recent claim of a successful experimental demonstration of a universal quantum computer outperforming best known classical algorithm in solving a hard problem by Scientists from Google[4] [Aru+19], it is likely that we are still years away from useful, let alone commercial exploitation of a universal quantum computer.

1.1 Quantum Computation Implementation Challenges and Boson Sampling

With the task of demonstrating quantum supremacy being (arguably) crossed off the list, it remains a challenge for physicists and computer scientists to harness quantum computation power to solve useful problems and to make it more economically available. To bypass persistent difficulties in universal quantum computation and harness its computational speed-up, Noisy Intermediate-Scale Quantum (NISQ) computational models [Pre18] have been proposed for a physical realization of a device that is capable of quantum computational power in the near future. This has been realized in number of quantum computational models that do not concern themselves with universality, but rather on solving a certain class of problems

[3]This is also often referred to as the *Extended Church-Turing Thesis*. Both namings are in spite of any contribution from either Church or Turing.

[4]This claim, however, was being disputed in https://www.ibm.com/blogs/research/2019/10/on-quantum-supremacy/ with more detailed argument in [Ped+19], a contention with interesting argument but will not be further discussed here.

that are otherwise hard for classical computers. Notable examples of NISQ models are Instantaneous Quantum Polynomial-time (IQP) [BMS16], Deterministic Quantum Computation (DQC) [KL98], Quantum Fourier Sampling [FU15], Quantum Approximate Optimization Algorithm [FH16], Random Circuit Sampling [Bou+18], and lastly, Boson Sampling [AA11], which will be the prototypical foundation of the main subject of this thesis.

Since its first inception in the seminal 2011 paper by Scott Aaronson and Alex Arkhipov [AA11], Boson Sampling has been receiving a lot of attention from both Quantum Physics and Theoretical Computer Science communities due to its potential to demonstrate quantum computational supremacy and to be a practically useful problem solver, apart from being both computationally and physically interesting in its own right. Its widespread interest is partly due to the surprising result that shows the impossibility of efficiently[5] simulating the task of sampling photon[6] count from the output distribution from a linear optical network[7] on a classical computer (on average). Aaronson and Arkhipov have elegantly formulated their argument with the aid of Computational Complexity Theory despite also relying on widely believed conjectures.

However before the computational power of Boson Sampling can be experimentally demonstrated, the challenge of reliably generating multiple single photons simultaneously to be fed into each individual input of a linear optical network is still an issue for large scale Boson Sampling due to the probabilistic nature of currently available single photon generators. More precisely, the probability of generating a desired photon inputs decreases exponentially as the size of input grows[8].

This drawback in the physical implementation of Boson Sampling has prompted a modification to its scheme, particularly on the preparation of single-photon inputs. A pioneering work by Lund, et.al. [Lun+14] proposed a Boson Sampling scheme called Scattershot Boson Sampling (SBS) is among the first notable attempts to solve this problem of single-photon inputs generation. The SBS model replaces the single photon inputs with Two-mode Squeezed State input preparations, an input preparation process that yields a state considered as a Gaussian State, which is a subclass of the class of quantum states known as the Continuous Variable (CV) states. As opposed to the Discrete Variable (DV) systems such as Boson Sampling or the more popular Qubit system, CV systems carries information with values over a continuous interval. The CV states, however, when measured using photon-counting device, does act as a probabilistic photon generator. This fact was used by Lund, et.al. to show that the rate of photon generation only decreases polynomially in the input size instead of exponentially, a significant improvement.

The promising result of SBS was further generalized by Kruse, et.al. [Kru+19] with a model called the Gaussian Boson Sampling (GBS) whose inputs are prepared as a Single-

[5]Notion of "efficient" here is in the complexity-theoretic sense, that is a problem can be solved efficiently is there exist an algorithm that outputs a correct answer in polynomial time. The impossibility of efficient simulation "on average" considers all possible linear optical network configurations, which will be discussed in more detail in [**Chapter 4**].

[6]photon is a type of particle under the particle class boson. Considering quantum optics being a prominent quantum technology implementation, we will follow the language of quantum optics and talk in terms of Photons instead of more general Bosons throughout this thesis.

[7]A class of experimental set-up consisting of widely considered readily available apparatus.

[8]If the probability of generating one photon on an input is 0.5, the probability of generating one photon in each input for a system with 10 inputs is 0.5^{10}.

mode Squeezed State (SMSS), a more general input preparation than SBS as proven by Kruse,et.al. The SMSS inputs of GBS can also be thought of as probabilistic photon generators, which were shown to have polynomial decrease of photon generation rate over number of SMSS inputs, similar to SBS. Despite having the CV model of quantum computation not receiving as much attention as its DV counterpart, it is to be noted that the formulation of an "equivalent" CV model to what originally was a DV model has been a growing trend for NISQ models. This is apparent in NISQ models other than Boson Sampling, such as the One Qumode Model [Liu+16] for DQC and CV IQP model [Dou+17] for IQP. In the experimental side for CV NISQ models, a notable claim for quantum computation supremacy using the GBS model was published in [Zho+20] when this thesis is being written. Despite disputes on the supremacy claim[9], this result nevertheless serves as a strong evidence of the growing popularity in the studies of CV NISQ models.

In this thesis, we present a scheme that fully embrace the CV paradigm by formulating a model that is equivalent to GBS, but instead of performing photon counting measurement, uses homodyne detection, a measurement yielding continuous values. An algorithm to construct the probability distribution for GBS using the Homodyne measurement results is also presented along with a computer simulation for simple instances. We also adopted the phase space quantum mechanics formulation and conduct an analysis for necessary and sufficient conditions for hardness of classical simulation of GBS, especially in the presence of noise, which is inevitable in NISQ models in general. In this spirit, we propose two alternative models of GBS that incorporates measurement noises which we hoope would shed new light in the discovery of the "boundary" of the hardness of classical simulation.

1.2 Prerequisites and Thesis Outline

Due to the multidisciplinary nature of this thesis (and Quantum Information Science in general), to cater to readers from both Computer Science and Physics background, we dedicate a chapter [**Chapter 2**] for Quantum Mechanics, Quantum Optics, and Quantum Information preliminaries and a chapter [**Chapter 3**] for Computational Complexity Theory. Nevertheless, knowledge of elementary probability theory, linear algebra, and calculus are assumed from the readers. Some of the most important concepts that will be used throughout the discussions in this thesis include Hilbert space, eigenvalues and eigenvectors, conditional probabilities, norms, and tensor products. For brief preliminary readings in linear algebra and the Hilbert space, the reader may refer to [NC10, Chapter 2.1] or [Ser17, Chapter 2]; both literatures are primary references for the quantum mechanics preliminary chapter in this thesis. However for a more thorough treatment of analysis in Hilbert space, the readers may refer to [SS05]. Additionally, readers may also benefit from elementary knowledge in first order logic for a smooth reading through the sections where polynomial computability and complexity theory constitute a significant part of the discussion. Computability concepts will be, nevertheless, elementarily explained.

After the preliminary parts, we begin the main content of this thesis on the definition and

[9]The scepticism on this claim is nicely summarized in a blog post https://www.scottaaronson.com/blog/?p=5159 by Scott Aaronson, which has also been responded by multiple experts in Boson Sampling. This disputes will be briefly discussed later in [**Chapter 5**].

a discussion of Boson Sampling in [**Chapter 4**] based on the original paper by Aaronson and Arkhipov. In [**Chapter 5**], we introduce the definition of GBS followed by a discussion on the computational complexity of its classical simulation based on the result by Kruse,et.al. The first result of this thesis is presented in [**Chapter 6**] where we derive a formulation of Homodyne Tomography (HT) method and the Pattern Function which lays the foundation for a classical simulation of GBS. We then show the numerical method based on HT to reconstruct an approximated GBS probability distribution, which we built a computer simulation of. The results of this simulation is then presented and discussed. To support the practicality of this simulation method, we present a derivation of a criteria to bound the approximation error, which is based on a well-known results from statistics. Lastly in this chapter, we discuss future improvements that can be made to the HT approximate GBS simulation, including addressing possible negative probabilities in the approximation causing a non-normalized probability distribution.

The next results are discussed in [**Chapter 7**], which focuses on investigating the possibility of efficient classical simulation of GBS based on the amount of noise present in the system. Here, we propose two models that can be understood as noisy GBS that may enable further analysis to better understand the boundary of efficient classical simulation of GBS in the noise spectrum. This chapter will be split into three parts. In the first part, we will discuss about the existing results on how the quasi-probability distributions in phase space quantum mechanics can characterize whether a quantum process can be classically simulated efficiently considering the presence of noise in the system. With the apparent significance of the role that noise plays in the possibility of classical simulation, the second part will be dedicated on the derivation of a GBS model that incurs noise in the photon-counting measurement that is inspired by a similarly formulated model of noisy Boson Sampling. We then discuss its resemblance to the noisy Boson Sampling model, which indicates that it seems plausible that a similar argument can be formulated to show the complexity criteria for our noisy GBS model. We nevertheless, leave these for future work. This noisy GBS model can also be an interesting subject for the Homodyne Tomography simulation that we discussed in the previous chapter as a potential evidence for ease of classical simulation compared to the ideal GBS case.

In the last part, we present another model that uses a Gaussian measurement and a non-Gaussian input state called the Photon-added Thermal State (PATS). This model, which we call the PATS-CVS model, was inspired by a result on the implication of noise to the phase-space quantum mechanics representations. We found that this model can be similarly formulated as the noisy GBS model presented in the previous part of this chapter. Lastly due to the resemblance of this model to both our noisy GBS model and the noisy Boson Sampling model that we previously discussed, we again discuss the plausibility of formulating an argument similar to the noisy Boson Sampling to show a complexity criteria for the PATS CVS model. Once again as we are short of a proof, formulation of a formal argument is left for future work.

Chapter 2

Quantum Mechanics Preliminaries

In this chapter we will discuss concepts in quantum mechanics that are relevant to our discussions later on Boson Sampling (BS) and Gaussian Boson Sampling (GBS). In [**Section 2.1**] we will discuss the representations of states and the evolution of a quantum mechanical system, as well as measurement of the system. Afterwards, these postulates will allow us to have a slightly more advanced discussion on Quantum Optics and Continuous Variable Quantum Information in [**Section 2.2**].

2.1 Postulates of Quantum Mechanics

In this section we will discuss the postulates of quantum mechanics that will be formulated in congruence with our discussions in the later chapters. The postulates are defined here are based on those in [NC10, Chapter 2.2], [Pur01, Chapter 1], and [Ser17, Chapter 2.1]. We will use the state vector form representation in [**Section 2.1.1**] and the density operator form [**Section 2.1.2**] to represent a state of a quantum system and discuss their evolution. Lastly in [**Section 2.1.3**], we will discuss the postulate on measuring a quantum system where a few types of measurements are introduced.

2.1.1 Quantum States and Evolution

In accordance with the first postulate for any given quantum system, we associate it to a complete, norm-induced vector space; namely, a Hilbert Space $\mathcal{S}$, where a vector $|\psi\rangle$ in $\mathcal{S}$ is represents a state of the system.

Postulate 2.1.1 (State Space and State Vector)**.** A quantum system is described by a *State Space* $\mathcal{S}$, where $\mathcal{S}$ is a Hilbert Space, we call $|\psi\rangle \in \mathcal{S}$ its *State Vector* if and only if

- $\langle\psi|\psi\rangle \in \mathbb{R}$ and $\langle\psi|\psi\rangle > 0$, and
- for collection of orthogonal state vectors $\{|\psi_j\rangle\} \subseteq \mathcal{S}$, its linear combination $|\psi\rangle = \sum_j \alpha_j|\psi_j\rangle$, where $\{\alpha_j\} \subseteq \mathbb{C}$, is also a state vector.

If $\sum_j |\alpha_j|^2 = 1$, then $|\psi\rangle$ is a *normalized* state vector. ◀

In fact, for a state vector $|\psi\rangle = \sum_i \alpha_i |\psi_i\rangle$ with mutually orthonormal $\{|\psi_i\rangle\}$, $|\psi\rangle$ is normalized if and only if $\langle\psi|\psi\rangle = 1$. This is clear from

$$\langle\psi|\psi\rangle = \sum_i \alpha_i^* \alpha_i \langle\psi_i|\psi_i\rangle = \sum_i |\alpha_i|^2.$$

It also should be noted that as a consequence of mutually orthonormal $\{|\psi_i\rangle\}$, coefficient α_j can be written as $\langle\psi_j|\psi\rangle$ because

$$\langle\psi_j|\psi\rangle = \sum_i \alpha_i \langle\psi_j|\psi_i\rangle = \alpha_j.$$

Therefore, we can write the state vector as $|\psi\rangle = \sum_i \langle\psi_i|\psi\rangle|\psi_i\rangle$.

Clearly, the state of any interesting quantum system evolves either through interaction with one or more other systems. In quantum mechanics, this is postulated as a linear operator on the Hilbert space $\mathcal{S}$.

Postulate 2.1.2 (Quantum System Evolution [BR97, Chapter 1.2])**.** For Hilbert Space $\mathcal{H}$ and $\mathcal{S}$, a function $A : \mathcal{S} \to \mathcal{H}$ is a *Linear Operator* (or simply, *Operator*) if and only if for $x, y \in \mathcal{S}$ and $c \in \mathbb{C}$,

$$Ax + Ay = A(x + y) \quad \text{and} \quad A(cx) = c(Ax).$$

For Operators $A, B : \mathcal{S} \to \mathcal{H}$ and $c \in \mathbb{C}$, we call $A^\dagger$ and $B^\dagger$ their *Hermitian Conjugate* if and only if

- $(A^\dagger)^\dagger = A$
- $A^\dagger + B^\dagger = (A + B)^\dagger$
- $(AB)^\dagger = B^\dagger A^\dagger$
- $(cA)^\dagger = c^* A^\dagger$

If $A^\dagger = A$, then A is a *Hermitian Operator* (or simply *Hermitian*). If $AA^\dagger = A^\dagger A = I$, then we call A *Unitary*. ◀

In its matrix representation, we can equivalently define a Hermitian operator as a matrix with entries for all i and j, $a_{ij} = a_{ji}^*$. In words, its i-th row, j-th column entry is equal to the complex conjugate of its j-th row i-th column entry.

It is to be noted that because a measurement of a quantum system changes the state of the system, it can be considered as a linear operation as well, as we will see it in the section on the measurement postulate. We will now define the linear algebraic concepts of eigenvalue and eigenstate[1] that are central in discussions throughout this thesis.

Definition 2.1.3 (Eigenvalue and Eigenstate)**.** For an operator $A : \mathcal{H} \to \mathcal{H}$, an ordered collection of complex numbers $\{\lambda_j\}_j$ and an ordered collection of $\{|\lambda_j\rangle\}_j$, both $\lambda \in \mathbb{C}$ is an *Eigenvalue* of A and $|\lambda\rangle$ its corresponding *Eigenstate* if and only if $A|\lambda\rangle = \lambda|\lambda\rangle$, and equivalently $\langle\lambda|A^\dagger = \lambda^*\langle\lambda|$. ◀

[1]The more common concept of *eigenvector* in Linear Algebra texts can be thought as equivalent to eigenstates. Namely, an eigenstate is an eigenvector that satisfies the property of a state vector.

The ordering of the eigenvalues and eigenvectors in the definition above is to indicate the correspondence between eigenvalues and eigenvectors. Thus, we will call λ_j as the corresponding eigenvalue of $|\lambda_j\rangle$ and likewise, $|\lambda_j\rangle$ as the corresponding eigenstate of λ_j.

Now, we will state a property of linear operators that relate them to the eigenvalues and eigenstates.

Proposition 2.1.4 (Spectral Theorem for a Hermitian Operator)**.** *Eigenvalues $\{\lambda_j\}_j$ of Hermitian operator $A : \mathcal{S} \to \mathcal{S}$ are real and their corresponding eigenstates are mutually orthogonal. Moreover, there exist an orthogonal basis of $\mathcal{S}$ consisting of eigenstates of A.*

Proof. Consider eigenvalues λ_k and λ_j of Hermitian operator A. Therefore, we have

$$\langle\lambda_k|A|\lambda_j\rangle = \lambda_k^*\langle\lambda_k|\lambda_j\rangle = \lambda_j\langle\lambda_k|\lambda_j\rangle \tag{2.1.1}$$

because for any eigenvalue λ of A, we have $A|\lambda\rangle = \lambda|\lambda\rangle$ and $\langle\lambda|A^\dagger = \langle\lambda|A = \lambda^*\langle\lambda|$. From (2.1.1) now consider

$$\begin{aligned}\lambda_k^*\langle\lambda_k|\lambda_j\rangle - \lambda_j\langle\lambda_k|\lambda_j\rangle &= 0\\ (\lambda_k^* - \lambda_j)\langle\lambda_k|\lambda_j\rangle &= 0.\end{aligned}$$

Because $|\lambda_k\rangle$ and $|\lambda_j\rangle$ are state vectors, if $k = j$, we have $\langle\lambda_k|\lambda_k\rangle > 0$, thus we have $\lambda_k^* - \lambda_k = 0$, implying that λ_k is real. If $\lambda_k \neq \lambda_j$, then $\lambda_k - \lambda_j = \lambda_k^* - \lambda_j \neq 0$, thus $\langle\lambda_k|\lambda_j\rangle = 0$, that is $|\lambda_k\rangle$ and $|\lambda_j\rangle$ are mutually orthogonal.

We will now prove the existence of the eigenstates orthogonal basis by induction on the dimension of $\mathcal{S}$. In the case if $\mathcal{S}$ is 0 dimension, this is vacuously true. If $\mathcal{S}$ is 1 dimensional, it is trivially true. Now, consider if the dimension of $\mathcal{S}$ is 2, then let λ_1 be an eigenvalue of A and let space V_1 be the space of all vectors $|x\rangle \in \mathcal{S}$ such that $\langle x|\lambda_1\rangle = 0$ (i.e: all $|x\rangle \in V_1$ are orthogonal to $|\lambda_1\rangle$). Therefore we have $A|x\rangle \in V_1$ because $\langle x|A^\dagger|\lambda_1\rangle = \lambda_1\langle x|\lambda_1\rangle$. Now consider operator A_1, a restriction of A to V_1, that is $A_1|x\rangle = A|x\rangle$ for all $|x\rangle \in V_1$. Let λ_2 be an eigenvalue of A_1 with $|\lambda_2\rangle$ as its corresponding eigenvector. Clearly, $|\lambda_1\rangle$ and $|\lambda_2\rangle$ are mutually orthogonal, and because $A|\lambda_2\rangle = A_1|\lambda_2\rangle = \lambda_2|\lambda_2\rangle$, it is clear that both of them are eigenvectors of A.

For the step case, assume that we have orthogonal eigenvectors $|\lambda_1\rangle \dots |\lambda_{k-1}\rangle$ of A which clearly spans $k-1$ dimensions. For a space $V_{k-1} = \{|x\rangle \in \mathcal{S} :\}$ that is orthogonal to $|\lambda_1\rangle \dots |\lambda_{k-1}\rangle$, we again consider restriction A_{k-1} of A such that for all $|x\rangle \in V_{k-1}$, we have $A_{k-1}|x\rangle = A|x\rangle$. Taking an eigenvector $|\lambda_k\rangle$ of A_{k-1}, we get k orthogonal eigenvectors of A, which concludes the proof by induction. □

It therefore follows from [**Proposition 2.1.4**] that a collection of normalized eigenstates $\{|\lambda_j\rangle\}$ (i.e: $\langle\lambda_j|\lambda_j\rangle = 1$) of Hermitian A are mutually orthonormal. It should also be noted that if the geometric multiplicity of an eigenvalue λ of A is greater than 1 (i.e: it has more than 1 corresponding eigenstates), then they must be mutually orthogonal as well because all eigenstates spanning $\mathcal{S}$ are orthogonal.

Lastly, we now define the commutator operation on linear operators on Hilbert space $\mathcal{S}$ which will be important in later discussions because of the non-commutativity in $\mathcal{S}$ in general.

Definition 2.1.5. For *commutator* $[A, B] = AB - BA$ and *anticommutator* $[A, B]_+ = AB + BA$, if $[A, B] = 0$ ($[A, B]_+ = 0$), then A and B *commute* (*anticommute*). If A and B are Hermitian and $[A, B] = 0$, then we call A and B *Compatible*, otherwise they are *Incompatible*. ◀

2.1.2 Quantum System with Mixed State

For a quantum system where we are not certain of its state, it is modeled as what we call a *mixed state*, that is a collection $\{|\psi_j\rangle\}_j$ where we the probability of the system being in state $|\psi_j\rangle$ is p_j. As a consequence, we are no longer able to represent its state as a state vector; thus, require us to introduce the *Density Operator* representation of a state.

Definition 2.1.6 (Density Operator)**.** Consider a quantum system in state space $\mathcal{S}$ and a collection of normalized state vectors $\{|\psi_j\rangle\}_j$ we call an *Ensemble of States* to represent the state of the quantum system where the probability of the system being in state $|\psi_j\rangle$ is $p_j \in [0, 1]$ such that $\sum_j p_j = 1$. We define *Density Operator* $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ as a mixture of ensemble $\{|\psi_j\rangle\}_j$. A system in state ρ transformed with unitary operator $U : \mathcal{S} \to \mathcal{S}$ evolves to $U\rho U^\dagger$.

If the state of the system can be represented by a state vector $|\psi\rangle = \sum_j \alpha_j |\psi_j\rangle$, we say that the system is in a *Pure State*, then we have $\rho = |\psi\rangle\langle\psi|$. If a system is not in a pure state, it is in a *Mixed State*, where the density operator can be more generally written as $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$. ◀

In most of the discussions throughout this thesis, we will refer to a state of quantum system simply as a "state", and purity of the state will be specified where it is relevant to the discussion.

From this definition, therefore, it follows that for a system in a pure state, we have $\rho = |\psi\rangle\langle\psi| = \sum_{j,k} \alpha_j \alpha_k^* |\psi_j\rangle\langle\psi_k|$ and its matrix representation entries $\rho_{j,k} = \alpha_j \alpha_k^*$. Similarly, for a system in a mixed state with ensemble $\{|\psi_u\rangle\}$ where $|\psi_u\rangle = \sum_j \alpha_j^{(u)} |u_j\rangle$, we have

$$\begin{aligned}\rho &= \sum_u p_u |\psi_u\rangle\langle\psi_u| \\ &= \sum_u \sum_{j,k} p_u \alpha_j^{(u)} (\alpha_k^{(u)})^* |u_j\rangle\langle u_k|,\end{aligned}$$

and its matrix representation entries $\rho_{j,k} = \sum_u p_u \alpha_j^{(u)} (\alpha_k^{(u)})^*$.

We also note that for the *Trace* of the density operator $\mathbf{Tr}(\rho) = \sum_j \rho_{j,j}$ where $\rho_{j,j}$ is the j-th diagonal element of matrix representation of ρ, we have

$$\mathbf{Tr}(\rho) = \sum_j \sum_u p_u \alpha_j^{(u)} (\alpha_j^{(u)})^* = \sum_u p_u \sum_j |\alpha_j^{(u)}|^2 = \sum_u p_u = 1, \tag{2.1.2}$$

since $|\psi_u\rangle$ is normalized for all u. For density operator ρ in the Hilbert space $\mathcal{S}$, the trace function $\mathbf{Tr}$ can also be defined as

$$\mathbf{Tr}(\rho) = \sum_j \langle\lambda_j|\rho|\lambda_j\rangle$$

for some basis $\{|\lambda_j\rangle\}_j$ of $\mathcal{S}$. Given a general form $\rho = \sum_u p_u |\psi_u\rangle\langle\psi_u|$ we can also obtain the unity of the trace of ρ as follows

$$\begin{aligned}\mathbf{Tr}(\rho) &= \mathbf{Tr}\left(\sum_u \sum_{j,k} p_u \alpha_j^{(u)} (\alpha_k^{(u)})^* |\lambda_j^{(u)}\rangle\langle\lambda_k^{(u)}|\right) \\ &= \sum_l \sum_u \sum_{j,k} p_u \alpha_j^{(u)} (\alpha_k^{(u)})^* \langle\lambda_l|\lambda_j^{(u)}\rangle\langle\lambda_k^{(u)}|\lambda_l\rangle \\ &= \sum_u \sum_l p_u |\alpha_l^{(u)}|^2 \\ &= 1\end{aligned}$$

where we used the expression of the vectors in terms of the $\{|\lambda_j\rangle\}_j$ basis $|\psi_u\rangle = \sum_j |\lambda_j^{(u)}\rangle$.

Lastly in this section, we state another postulate of Quantum mechanics describing multiple quantum systems or, a composite quantum system. For this, we need the tensor product operation $\otimes$ to compose m Hilbert spaces describing m quantum systems as a single object. Here, we informally define the tensor product as a bilinear map from two Hilbert spaces $\mathcal{S}_a$ and $\mathcal{S}_b$ describing two systems to a product of Hilbert spaces $\mathcal{S}_a \otimes \mathcal{S}_b$ describing the system from their composition. It has the property that for any $|x\rangle_a \in \mathcal{S}_a$ and $|y\rangle_b \in \mathcal{S}_b$, we have a bilinear form $|x\rangle_a \otimes |y\rangle_b = |x, y\rangle \in \mathcal{S}_a \otimes \mathcal{S}_b$ where for $\alpha \in \mathbb{C}$,

- $(\alpha|x\rangle_a) \otimes |y\rangle_b = |x\rangle_a \otimes (\alpha|y\rangle_b) = \alpha(|x\rangle_a \otimes |y\rangle_b)$,
- $(|x_1\rangle_a + |x_2\rangle_a) \otimes |y\rangle_b = (|x_1\rangle_a \otimes |y\rangle_b) + (|x_2\rangle_a \otimes |y\rangle_b)$,
- and $|x\rangle_a \otimes (|y_1\rangle_b + |y_2\rangle_b) = (|x\rangle_a \otimes |y_1\rangle_b) + (|x\rangle_a \otimes |y_2\rangle_b)$.

We do not formally define nor discuss tensor products in detail here, for a thorough discussion in the context of quantum mechanics, readers may refer to [NC10, Chapter 2.1.7].

Postulate 2.1.7. [Composite Quantum System] For a composite quantum system composed of m quantum systems described by m Hilbert spaces $\mathcal{S}_1, \ldots, \mathcal{S}_m$, we describe both the operators and states of the composite quantum system in a tensor product of the individual Hilbert spaces $\bigotimes_{j=1}^m \mathcal{S}_j$. For shorthand, we write the tensor product of m Hilbert space as $\mathcal{S}^{\otimes m} = \bigotimes_{j=1}^m \mathcal{S}_j$. ◀

Although for the rest of this chapter we will be mostly talk in terms of a single quantum system, Hilbert space product representation of a quantum system will be a significant part of the analysis throughout discussions later on BS, GBS, and noise in a quantum system.

2.1.3 Quantum State Measurements

The discovery of quantum mechanics reveals that the observations of identically processed quantum systems obey a certain probability distribution. This leads to a very important postulate in Quantum Mechanics where the probability distribution of an observed value is induced by the state of the quantum system, in which a measurement is represented by a collection of operators. We will call this postulate the *Measurement Postulate* and present

the measurement of a quantum system both in the state vector description $|\psi\rangle$ and density operator description ρ, as well as present a class of measurement that is widely used in quantum mechanics called the *Positive Operator-Valued Measurement*.

Postulate 2.1.8 (General measurement on a quantum state and POVM [NC10, Chapter 2.2.3 and 2.2.6], [Ser17, Chapter 2.1, p.18])**.** Measurement on a quantum system is described by a collection of *Measurement Operators* $\{M_j\}$ each for a corresponding possible measurement outcome $j \in K$ where K is a set of measurement outcomes with respect to a certain basis, such that $\sum_j M_j^\dagger M_j = I$.

- For a system in a pure state $|\psi\rangle$, the probability of having measurement outcome j is $\mathbb{P}_\psi(j) = \langle\psi|M_j^\dagger M_j|\psi\rangle$ and state after performing measurement $\dfrac{M_j|\psi\rangle}{\sqrt{\langle\psi|M_j^\dagger M_J|\psi\rangle}}$, and
- for a system in a mixed state $\rho = \sum_k p_k|\psi_k\rangle\langle\psi_k|$ where $\sum_k p_k = 1$, the probability of obtaining j as $\mathbb{P}_\rho(j) = \sum_k p_k \mathbb{P}_\rho(j|k)$ where $\mathbb{P}_\rho(j|k) = \langle\psi_k|M_j^\dagger M_j|\psi_k\rangle$. State ρ_j after obtaining value j from measurement is a mixture of probability of having state $|\psi_k\rangle$ where j given state of the system before measurement $|\psi_k\rangle$; more formally, $\rho_j = \sum_k \mathbb{P}_\rho(k|j)|\psi_k^{(j)}\rangle\langle\psi_k^{(j)}|$, where $|\psi_k^{(j)}\rangle = \dfrac{M_j|\psi_k\rangle}{\sqrt{\langle\psi_k|M_j^\dagger M_j|\psi_k\rangle}}$.

If we set $M_j^\dagger M_j = E_j$ such that E_j is positive definite for all j, then we have a *Positive Operator-Valued Measurement* or *POVM* on the quantum system with collection of *POVM Elements* $\{E_j\}$ where the state after measurement is not generally defined. For a system with state $|\psi\rangle$, we have $\mathbb{P}_\psi(j) = \langle\psi|E_j|\psi\rangle$; and for a system with state $\rho = \sum_k p_k|\psi_k\rangle\langle\psi_k|$, we have $\mathbb{P}_\rho(j) = \sum_k \mathbb{P}_\rho(j|k)p_k$ where $\mathbb{P}_\rho(j|k) = \langle\psi_k|E_j|\psi_k\rangle$. ◀

Throughout this thesis, where the state inducing the the probability measure is clear from the context, we will omit the subscript of the probability measure $\mathbb{P}_\rho$. Also to put an emphasis on the protocol or model that the outcome probability is associated with, the subscript will instead indicate the name of that protocol or model.

We note that in general, POVM measurements are practically used in a case where the measurement result of interest is done only once (e.g: after performing a certain set of operations). Also, a POVM element E_j is Hermitian since $E_j = M_j^\dagger M_j$ for some operator M_j (although this follows by definition from the positive definiteness of E_j), thus by the spectral theorem it has an orthonormal set of eigenvectors $\{|\lambda_j\rangle\}$ with corresponding eigenvalues $\{\lambda_j\}$ which we can use to write it as $E_j V = V\Lambda$ where $V = (|\lambda_1\rangle, |\lambda_2\rangle, \dots)$ such that $V^\dagger V = V^\dagger V = I$ and $\Lambda = \mathbf{Diag}(\lambda_1, \lambda_2, \dots)$. Consequently, we can write $M_j = \sqrt{E_j}$ because we can diagonalize $\sqrt{E_j} = V\sqrt{\Lambda}V^\dagger$ where $\sqrt{\Lambda} = \mathbf{Diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots)$ and therefore have $M_j^\dagger M_j = \sqrt{E_j}^\dagger \sqrt{E_j} = V\sqrt{\Lambda}V^\dagger V\sqrt{\Lambda}V^\dagger = V\Lambda V^\dagger = E_j$.

To simplify the probabilities in terms of traces, consider set of mutually orthonormal

collection of basis $\{|l\rangle\}$ that includes $|\psi_k\rangle$, then

$$\begin{aligned}\mathbb{P}(j|k) &= \langle\psi_k|M_j^\dagger M_j|\psi_k\rangle \\ &= \sum_l \langle l|M_j^\dagger M_j|\psi_k\rangle\langle\psi_k|l\rangle \\ &= \mathbf{Tr}(M_j^\dagger M_j|\psi_k\rangle\langle\psi_k|),\end{aligned}$$

therefore giving us

$$\mathbb{P}(j) = \mathbf{Tr}\left(\sum_k p_k M_j^\dagger M_j|\psi_k\rangle\langle\psi_k|\right) = \mathbf{Tr}(M_j^\dagger M_j\rho)$$

Then, by Bayes' rule, we have $\mathbb{P}(k|j) = \dfrac{\mathbb{P}(j|k)p_k}{\mathbb{P}(j)} = \dfrac{p_k\langle\psi_k|M_j^\dagger M_j|\psi_k\rangle}{\mathbf{Tr}(M_j^\dagger M_j\rho)}$, and therefore

$$\begin{aligned}\rho_j &= \sum_k \frac{p_k\langle\psi_k|M_j^\dagger M_j|\psi_k\rangle}{\mathbf{Tr}(M_j^\dagger M_j\rho)}\frac{M_j|\psi_k\rangle\langle\psi_k|M_j^\dagger}{\langle\psi_k|M_j^\dagger M_j|\psi_k\rangle} \\ &= \frac{M_j\rho M_j^\dagger}{\mathbf{Tr}(M_j^\dagger M_j\rho)}.\end{aligned}$$

Projective Measurement

Now, we will consider a special case of measurement on a quantum system which is highly relevant to the main content of this thesis and models a lot of important measurements in quantum mechanics called the *Projective Measurement*. The measurement outcome from a Projective measurement is an eigenstate of a Hermitian operator that corresponds to the Projective measurement called the *Observable*, which is central in the concept of Projective measurement. Before proceeding, we will now define both of these concepts.

Definition 2.1.9 (Observables and Projective Measurement [NC10, ch. 2.2.5] [LP07, ch. 1.2.1])**.** For a quantum system with state space $\mathcal{S}$ that is being measured, a Hermitian operator $\widehat{A} : \mathcal{S} \to \mathcal{S}$ is an *Observable* if

- it has a spectral decomposition $\widehat{A} = \sum_j \lambda_j P_j$, where P_j is a projector onto the eigenspace spanned by eigenstate $|\lambda_j\rangle$ that corresponds to eigenvalue λ_j of $\widehat{A}$ such that $\sum_j P_j = I$, and
- the probability of measurement outcome result λ_j from measuring the observable $\widehat{A}$ of a system in a pure state $\rho = |\psi\rangle\langle\psi|$ is $\mathbb{P}(\lambda_j) = \langle\psi|P_j|\psi\rangle$, while its state after this measurement outcome is $\dfrac{P_j|\psi\rangle}{\sqrt{\mathbb{P}(\lambda_j)}}$, and
- if the system is in a mixed state $\rho = \sum_k p_k|\psi_k\rangle\langle\psi_k|$, then the probability of measuring λ_j is $\mathbb{P}(\lambda_j) = \sum_k \mathbb{P}(\lambda_j|k)p_k$ where $\mathbb{P}(\lambda_j|k) = \langle\psi_k|P_j|\psi_k\rangle$, and its state after obtaining

λ_j forms the measurement, $\rho_j = \sum_k \mathbb{P}(k|\lambda_j)\big|\psi_k^{(j)}\big\rangle\big\langle\psi_k^{(j)}\big|$ where $\big|\psi_k^{(j)}\big\rangle = \dfrac{P_j|\psi_k\rangle}{\sqrt{\mathbb{P}(\lambda_j)}}$ and $\mathbb{P}(k|\lambda_j) = \dfrac{\mathbb{P}(\lambda_j|k)p_k}{\mathbb{P}(\lambda_j)}$.

The collection $\{P_j\}_j$ is a measurement called *Projective Measurement* and an operator P_k from spectral decomposition of $\widehat{A}$ is called a *Projective Measurement Operator*. We denote the *Expectation* of $\widehat{A}$ as $\langle\widehat{A}\rangle = \sum_j \lambda_j \mathbb{P}(\lambda_j)$, its *Standard Deviation* (or *Uncertainty*) as $\mathbf{Sd}(\widehat{A}) = \sqrt{\langle\widehat{A}^2\rangle - \langle\widehat{A}\rangle^2}$, and *Variance* $\mathbf{Var}(\widehat{A}) = \mathbf{Sd}^2(\widehat{A})$. ◀

It is to be noted that the state post-measurement for both pure and mixed state can be expressed in terms of $|\lambda_j\rangle$. This is possible because we cna write the state vectors $|\psi\rangle$ and $|\psi_k\rangle$ as a weighted sum of the set of eigenstates of $\widehat{A}$, $|\psi\rangle = \sum_i c_i|\lambda_i\rangle$ giving $P_j|\psi\rangle = \sum_i c_i P_j|\lambda_i\rangle \propto |\lambda_j\rangle$.

Below in [**Corollary 2.1.10**], we note that any observable $\widehat{A}$ always has a set of eigenstates on $\mathcal{S}$ that are mutually orthogonal, which directly follows from [**Proposition 2.1.4**] as a consequence of an Observable being a Hermitian operator.

Corollary 2.1.10. *For an Observable on Hilbert space $\mathcal{S}$, $\widehat{A} : \mathcal{S} \to \mathcal{S}$ with eigenstates $\{|\lambda_j\rangle\}_j$ that spans $\mathcal{S}$, we have $\langle\lambda_k|\lambda_l\rangle = \delta_{k,l}$.*

There are equivalent representations of the projective measurement outcome probability and observable expectation that are more commonly used in literatures on quantum information science and quantum computing that will be useful for discussions in later chapters to talk in a language more familiar within these disciplines. Derivations of these equivalent representations will be made below for both arbitrary pure and mixed states and summarized later in [**Proposition 2.1.11**].

For a system in a state described by vector $|\psi\rangle$, we can equivalently write its probability of obtaining measurement λ_j in terms of $|\lambda_j\rangle$. We do this by considering expressing it in terms of the eigenstates of $\widehat{A}$ and noting that $\langle\lambda_k|\lambda_l\rangle = \delta_{k,l}$ from [**Corollary 2.1.10**], which gives

$$|\psi\rangle = \sum_k \alpha_k|\lambda_k\rangle = \sum_k\sum_l \alpha_l\langle\lambda_k|\lambda_l\rangle|\lambda_k\rangle = \sum_k \langle\lambda_k|\psi\rangle|\lambda_k\rangle. \tag{2.1.3}$$

Then, because P_j is a projector to eigenspace of $|\lambda_j\rangle$, we can write $P_j|\lambda_k\rangle = c_k^{(j)}|\lambda_j\rangle$ where $c_k^{(j)} = 1$ if $j = k$, and $c_k^{(j)} = 0$ otherwise. Again, using $\langle\lambda_j|\lambda_k\rangle = \delta_{j,k}$, we get

$$\mathbb{P}(\lambda_j) = \langle\psi|P_j|\psi\rangle = \sum_{u,v} \alpha_u^*\alpha_v\langle\lambda_u|P_j|\lambda_v\rangle = \sum_{u,v} \alpha_u^*\alpha_v c_v^{(j)}\langle\lambda_u|\lambda_j\rangle = |\alpha_j|^2 = |\langle\lambda_j|\psi\rangle|^2. \tag{2.1.4}$$

Whereas for a system which has its state represented by a density operator ρ, we can write the probability of measurement $\mathbb{P}(\lambda_j|k)$ in terms of Trace and use its cyclic property as follows

$$\langle\psi_k|P_j|\psi_k\rangle = \sum_n \psi_k^{(n)} \sum_m p_j^{(n,m)}\psi_k^{(m)} = \mathbf{Tr}(P_j|\psi_k\rangle\langle\psi_k|) = \mathbf{Tr}(|\psi_k\rangle\langle\psi_k|P_j)$$

where $p_j^{(n,m)}$ is the n-th row, m-th column entry of matrix representation of P_j and $\psi_k^{(n)}$ the n-th entry of state vector $|\psi_k\rangle$. Therefore, we can write $\mathbb{P}(\lambda_j) = \sum_k p_k \mathbf{Tr}(|\psi_k\rangle\langle\psi_k|P_j) = \mathbf{Tr}(\rho P_j)$ by the linearity of Trace function.

Expectation of an observable $\widehat{A}$ on state $|\psi\rangle$ can also be expressed as

$$\langle\widehat{A}\rangle = \sum_j \lambda_j \mathbb{P}(\lambda_j) = \sum_j \lambda_j \langle\psi|P_j|\psi\rangle = \left\langle \psi \middle| \sum_j \lambda_j P_j \middle| \psi \right\rangle = \langle\psi|\widehat{A}|\psi\rangle.$$

Furthermore, expressing $|\psi\rangle$ in terms of eigenstates of $\widehat{A}$ as in (2.1.3) and giving the same argument of orthonormality of $\{|\lambda_j\rangle\}$ as in (2.1.4), we have

$$\begin{aligned}\langle\widehat{A}\rangle = \langle\psi|\widehat{A}|\psi\rangle &= \sum_{u,v} \langle\lambda_u|\psi\rangle^* \langle\lambda_v|\psi\rangle \langle\lambda_u|\widehat{A}|\lambda_v\rangle \\ &= \sum_j \lambda_j \sum_{u,v} \langle\lambda_u|\psi\rangle^* \langle\lambda_v|\psi\rangle \langle\lambda_u|P_j|\lambda_v\rangle \\ &= \sum_j \lambda_j |\langle\lambda_j|\psi\rangle|^2.\end{aligned}$$

Similarly, we can also write the expectation of observable $\widehat{A}$ on state $\rho = \sum_k p_k |\psi_k\rangle\langle\psi_k|$ in terms of Trace function

$$\langle\widehat{A}\rangle = \sum_j \lambda_j \mathbb{P}(\lambda_j) = \sum_j \lambda_j \mathbf{Tr}(\rho P_j) = \mathbf{Tr}(\rho\widehat{A})$$

by linearity property of $\mathbf{Tr}$, and also, equivalently in terms of ensemble $\{|\psi_j\rangle\}$,

$$\langle\widehat{A}\rangle = \sum_j \lambda_j \sum_k \mathbb{P}(\lambda_j|k) p_k = \sum_k \sum_j p_k \lambda_j \langle\psi_k|P_j|\psi_k\rangle = \sum_k p_k \langle\psi_k|\widehat{A}|\psi_k\rangle.$$

Alternatively, $\mathbf{Tr}$ can be evaluated using a basis $\{v_j\}$ of $\mathcal{S}$, giving us a different derivation

$$\begin{aligned}\mathbf{Tr}(\rho\widehat{A}) &= \sum_j \sum_k p_k \langle v_j|\psi_k\rangle \langle\psi_k|\widehat{A}|v_j\rangle \\ &= \sum_k p_k \sum_j \langle\psi_k|\widehat{A}|v_j\rangle \langle v_j|\psi_k\rangle \\ &= \sum_k p_k \langle\psi_k|\widehat{A}|\psi_k\rangle\end{aligned}$$

For convenience, we summarize the equivalent definitions derived above in the following proposition.

Proposition 2.1.11. *For observable $\widehat{A} = \sum_j \lambda_j P_j$ with eigenvalues $\{\lambda_j\}_j$, corresponding eigenstates $\{|\lambda_j\rangle\}_j$, and projection operators $\{P_j\}_j$ of a system in a pure state $|\psi\rangle$, we have*

$$\mathbb{P}(\lambda_j) = |\langle\lambda_j|\psi\rangle|^2, \quad \textit{and} \quad \langle\widehat{A}\rangle = \langle\psi|\widehat{A}|\psi\rangle = \sum_j \lambda_j \langle\lambda_j|\psi\rangle.$$

Similarly, for the same observable $\widehat{A}$ of a system in a mixed state that is described by density operator $\rho = \sum_k p_k |\psi_k\rangle\langle\psi_k|$,

$$\mathbb{P}(\lambda_j) = \mathbf{Tr}(\rho P_j), \quad \text{and} \quad \langle\widehat{A}\rangle = \mathbf{Tr}(\rho\widehat{A}) = \sum_k p_k \langle\psi_k|\widehat{A}|\psi_k\rangle.$$

We now give a short example where observables are used in a physical system, and compare both the classical and quantum models to get a sense of how a system is modeled in terms of observables and how classical and quantum approaches differs.

Example 2.1.12 ([LP07, Ch. 1.2.1])**.** First, we will show the classical model of a one-dimensional Harmonic Oscillator system with a with mass m on position q and frequency of the oscillator ω. Then for momentum p, we have energy of the system

$$e = \frac{1}{2m}p^2 + \frac{m\omega^2}{2}q^2.$$

In the quantum model of the system, we replace position q and momentum p with observables $\widehat{q}$ and $\widehat{p}$, and now the energy of our system is represented by a hermitian operator called *Hamiltonian*

$$H = \frac{1}{2m}\widehat{p}^2 + \frac{m\omega^2}{2}\widehat{q}^2.$$

◇

Lastly in this section we would like to note that most of the discussions in [**Section 2.2**] about Quantum Optics and Continuous Variable quantum information and in later chapters about BS and GBS that involves measurement of the system will involve projective measurement operators in the form of $P_j = |\lambda_j\rangle\langle\lambda_j|$ where $|\lambda_j\rangle$ is an eigenstate of some observable $\widehat{A}$ with set of eigenstates $\{|\lambda_j\rangle\}$. Note that in this case, P_j is also a POVM as $P_j^\dagger P_j = |\lambda_j\rangle\langle\lambda_j|\lambda_j\rangle\langle\lambda_j| = |\lambda_j\rangle\langle\lambda_j|$, thus for any vector $|x\rangle$, we have $\langle x|P_j^\dagger P_j|x\rangle = |\langle x|\lambda_j\rangle|^2 \geq 0$. For this kind of measurement in later discussions, we will refer to it either as a POVM or simply as "measurement".

2.2 Quantum Optics and Continuous Variable Quantum Information

Quantum properties of light in the Quantum Optics sub-discipline were among the first discoveries that led to the established theory of quantum mechanics. The quantum optics interpretation of light as both particles and waves has a natural information-theoretic interpretation that light can carry both discrete and continuous information. The continuous information of system that are manipulated and measured are therefore what we refer to as the *Continuous Variable* (CV) of the quantum system. In this section, we will discuss both discrete and continuous information on an optical quantum system and how they are related. Particular focus will be on discrete quantum system of photonic quantized energy levels, and continuous quantum system of momentum-position values as a quantum Harmonic oscillator, both will be defined in [**Section 2.2.1**] and their relationship through the

quantum harmonic oscillator will be discussed in [**Section 2.2.2**]. We will then define a few states that will be relevant to our discussions later in [**Section 2.2.3**], and lastly in [**Section 2.2.4**], we will then discuss quantum systems in the phase space description and Gaussian states.

We note that discussions on the optical and CV quantum systems will be done in terms of modes as a reference to a composite system. While mathematically, the number of modes in a system is represented in the space that it is described in, physically it has been defined as the collection of solutions to the Maxwell equations of a vacuum space. Discussions on details of this definition will not be pursued here, but readers who were interested should refer to [FT20] or [RW20]. Also, from this section onwards, we will assume that $\hbar = 2$, except where it is stated otherwise.

2.2.1 Bosonic and Quadrature Quantum System

In this subsection, we give two descriptions of an m-mode Bosonic System, namely, the *Bosonic Field* description and the *Quadrature Field* description where the relationship between them reflects the particle-wave duality of quantum mechanics. We will discuss how the latter description defines Continuous Variable Quantum Information in terms of amplitude and phase. While the Hilbert space description of quantum states in both descriptions is defined in a tensor product of m Hilbert spaces, we will later show a different representation of a state of a system in what is called a phase space, instead of a Hilbert space. The analysis of quantum system in the phase space will be highly relevant to our discussion later in [**Chapter 6**] about Homodyne Tomography and Pattern function, which is the main contribution presented in this thesis. While calling the system as "Bosonic" in a general manner, considering the context of quantum optics, we will often use a more specific reference to the discrete quantum excitation level of the system as "Photon".

Definition 2.2.1 (Bosonic Field operators [Wee+12, Section II A])**.** For an m-mode Bosonic System described in a tensor product of Hilbert spaces $\bigotimes_{k=1}^{m} \mathcal{H}_k$, each $\mathcal{H}_k$ is spanned by the *Fock basis* (or *Fock states*, as we will use these terms interchangeably) $\{|n\rangle\}_{n\in m}$ which are the eigenstates of the *Number Operator* observable $\widehat{n}_k = \widehat{a}_k^\dagger \widehat{a}_k$ where $\widehat{a}_k$ and $\widehat{a}_k^\dagger$ are the *annihilation operator* and *creation operator* on $\mathcal{H}_k$, respectively. Both $\widehat{a}_k$ and $\widehat{a}_k^\dagger$ are referred to as the *Bosonic Field Operators*.

The Bosonic field operators obey the commutation relations

$$[\widehat{a}_i, \widehat{a}_j^\dagger] = \delta_{i,j} I, \quad [\widehat{a}_i^\dagger, \widehat{a}_j] = -\delta_{i,j} I, \quad \text{and} \quad [\widehat{a}_i^\dagger, \widehat{a}_j^\dagger] = [\widehat{a}_i, \widehat{a}_j] = 0I$$

and their operation on the fock bases are defined as

$$\widehat{a}_j |n\rangle_j = \begin{cases} 0|0\rangle_j & \text{if } n = 0 \\ \sqrt{n}|n-1\rangle_j & \text{if } n \geq 1 \end{cases}, \quad \text{and} \quad \widehat{a}_j^\dagger |n\rangle = \sqrt{n+1}|n+1\rangle_j$$

where we call $|0\rangle$ a *Vacuum State*. ◀

The number operator, therefore, is an observable with the Fock states as its eigenstates because $\widehat{n}_k |n\rangle_k = n|n\rangle_k$. For a system is in a state ρ, its expectation $\langle \widehat{n}_k \rangle_\rho$ is the average

photon number in the k-th mode and the probability of measuring j photons found is $\mathbb{P}_\rho(j) = \langle j|\rho|j\rangle$. However neither $\widehat{a}_k$ nor $\widehat{a}_k^\dagger$ are Hermitian, hence they are not observables on their own. Letting vector of operators $\widehat{b} = (\widehat{a}_1, \widehat{a}_1^\dagger, \ldots, \widehat{a}_m, \widehat{a}_m^\dagger)$, we can describe the commutation relationship between the annihilation and creation operators of different modes by

$$[\widehat{b}_j, \widehat{b}_k] = \Omega_{j,k} I, \quad \text{where} \quad \Omega = \bigoplus_{j=1}^{m} \Omega^{(j)} = \begin{bmatrix} \Omega^{(1)} & & \\ & \ddots & \\ & & \Omega^{(m)} \end{bmatrix} \quad \text{and} \quad \Omega^{(j)} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

which is consistent with the commutation relations $[\widehat{a}_i, \widehat{a}_j^\dagger] = \delta_{i,j} I$. It is also worth noting that $\mathcal{H}$ spanned by Fock basis $\{|n\rangle\}_{n\in\mathbb{N}}$ has infinite dimensions. A space with finite dimensions where quantum systems can be described as completely, called the phase space will be discussed in [**Section 2.2.4**] later, but now we will first discuss the quadrature field description of a quantum system that

Definition 2.2.2 (Quadrature Field)**.** For an m-mode Bosonic System described in a tensor product of Hilbert spaces $\bigotimes_{k=1}^{m} \mathcal{H}_k$, for each Hilbert space $\mathcal{H}_k$ corresponding to mode k, we have *Quadrature Field Operators*, which are the *position operator* $\widehat{q}_k = \widehat{a}_k^\dagger + \widehat{a}_k$ and *momentum operator* $\widehat{p}_k = i(\widehat{a}_k^\dagger - \widehat{a}_k)$. We also have $|q\rangle_k, |p\rangle_k \in \mathcal{H}$ as the eigenstates of $\widehat{q}_k$ and $\widehat{p}_k$ with eigenvalues $q, p \in \mathbb{R}$, respectively (i.e: $\widehat{q}_k|q\rangle_k = q|q\rangle_k$ and $\widehat{p}_k|p\rangle_k = p|p\rangle_k$). We call $|q\rangle_k$ and $|p\rangle_k$ the *Quadrature Eigenstates* of $\widehat{q}_k$ and $\widehat{p}_k$, respectively. ◀

The equivalence relationship between $\widehat{p}_k$, $\widehat{q}_k$, and $\widehat{a}_k$ can also be written as

$$\begin{aligned} \widehat{p}_k &= i(\widehat{q}_k - \widehat{a}_k - \widehat{a}_k) \\ -2\widehat{a}_k &= -\widehat{q}_k - i\widehat{p}_k \\ \widehat{a}_k &= \frac{\widehat{q}_k + i\widehat{p}_k}{2} \end{aligned} \tag{2.2.1}$$

where the first equivalence is obtained by by substituting in $\widehat{a}_k^\dagger = \widehat{q}_k - \widehat{a}_k$ to definition of $\widehat{p}_k$. Therefore, $\widehat{q}_k$ and $\widehat{p}_k$ can be interpreted as a cartesian decomposition of $\widehat{a}_k$; that is the real and imaginary part of annihilation operator $\widehat{a}_k$. A similar decomposition can be done for the creation operator, giving us

$$\widehat{a}_k^\dagger = \frac{1}{2}(\widehat{q}_k - i\widehat{p}_k). \tag{2.2.2}$$

Analogous to bosonic field description, we now consider the commutation relation between $\widehat{p}_j$ and $\widehat{q}_k$:

$$\begin{aligned} [\widehat{p}_j, \widehat{q}_k] &= i(\widehat{a}_j^\dagger - \widehat{a}_j)(\widehat{a}_k^\dagger + \widehat{a}_k) - i(\widehat{a}_k^\dagger + \widehat{a}_k)(\widehat{a}_j^\dagger - \widehat{a}_j) \\ &= i[\widehat{a}_j^\dagger, \widehat{a}_k] + i[\widehat{a}_k^\dagger, \widehat{a}_j] \\ &= -2i\delta_{j,k} I \end{aligned} \tag{2.2.3}$$

and

$$\begin{aligned} [\widehat{q}_j, \widehat{p}_k] &= i(\widehat{a}_j^\dagger + \widehat{a}_j)(\widehat{a}_k^\dagger - \widehat{a}_k) - i(\widehat{a}_k^\dagger - \widehat{a}_k)(\widehat{a}_j^\dagger + \widehat{a}_j) \\ &= i[\widehat{a}_j, \widehat{a}_k^\dagger] + i[\widehat{a}_k, \widehat{a}_j^\dagger] \\ &= 2i\delta_{j,k} I. \end{aligned} \tag{2.2.4}$$

With a vector of quadrature field operators $\widehat{x} = (\widehat{q}_1, \widehat{p}_1, \ldots, \widehat{q}_m, \widehat{p}_m)$ and the Ω matrix used for the commutation relation of bosonic field description, we can write $[\widehat{x}_j, \widehat{x}_k] = 2i\Omega_{j,k}I$.

As the upcoming discussions will focus on individual modes, for notational simplicity, the subscript indicating which mode of the system that the operator or state belongs to will be suppressed for the rest of this section. The generalization to the multimode[2] case is straightforward by considering the tensor product of operators or vectors.

Now, notice that both $\widehat{q}$ and $\widehat{p}$ are Hermitian operators, as can be seen from

$$\begin{aligned}\widehat{q}^\dagger &= (\widehat{a}^\dagger + \widehat{a})^\dagger = \widehat{a} + \widehat{a}^\dagger = \widehat{q} \\ &\text{and} \\ \widehat{p}^\dagger &= i^*(\widehat{a}^\dagger - \widehat{a})^\dagger = i(-\widehat{a} + \widehat{a}^\dagger) = \widehat{p}.\end{aligned}$$

Additionally, their eigenstates satisfy

$$\int_{\mathbb{R}} |x\rangle\langle x| \, dx = I$$

as it was proven in [BR97, Chapter 3.2] by showing that for $k, l \in \mathbb{N}$

$$\int_{\mathbb{R}} \langle k|x\rangle\langle x|l\rangle \, dx = \delta_{k,l}.$$

Furthermore, by looking at the expectation of the quadrature operator $z \in \{q, p\}$ of a system in an arbitrary state ρ,

$$\begin{aligned}\langle\widehat{z}\rangle_\rho &= \mathbf{Tr}(\rho\widehat{z}) \\ &= \int_{\mathbb{R}} \langle z|\rho\widehat{z}|z\rangle \, dz \\ &= \int_{\mathbb{R}} z\langle z|\rho|z\rangle \, dz, \end{aligned} \tag{2.2.5}$$

it is clear that the $\langle z|\rho|z\rangle$ term inside the integral induces a probability measure to the real values z and allows us to define the probability of an outcome $z \in \mathbb{R}$ from a system in a state ρ as

$$\mathbb{P}(z) = \langle z|\rho|z\rangle. \tag{2.2.6}$$

This integral expression of the trace and the expectation is a generalization of the sum expression of an expectation of an observable with discrete eigenstates stated in [**Definition 2.1.9**]. Thus, the position operator $\widehat{q}$ and momentum operator $\widehat{p}$ for a single mode are observables that can be written as

$$\widehat{q} = \int_{\mathbb{R}} q|q\rangle\langle q| \, dq \quad \text{and} \quad \widehat{p} = \int_{\mathbb{R}} p|p\rangle\langle p| \, dp.$$

Therefore, the quadrature field operators are observables where the probability of obtaining measurement outcome $q, p \in \mathbb{R}$ for a given quantum system in state ρ are $\mathbb{P}_\rho(q) = \langle q|\rho|q\rangle$

[2]Short for "multiple mode".

and $\mathbb{P}_\rho(p) = \langle p|\rho|p\rangle$, respectively. Also, given the continuous eigenstates of $\widehat{q}$ and $\widehat{p}$, we can represent the trace of an operator in quadrature basis as

$$\mathbf{Tr}(F) = \int_{\mathbb{R}} \langle r|F|r\rangle \, dr$$

for some operator F.

Also, it is worth noting that quadrature eigenstates of $\widehat{p}$ and that of $\widehat{q}$ are *improper* eigenstates as $\langle v|v\rangle \neq 1$ for any eigenvalue $v \in \mathbb{R}$ of $\widehat{p}$ or $\widehat{q}$ [3]. Additionally in each mode, $\widehat{p}$ and $\widehat{q}$ each has a continuous spectrum of eigenstates $\{|p\rangle\}_{p\in\mathbb{R}}$ and $\{|q\rangle\}_{q\in\mathbb{R}}$ (and thus a continuous corresponding eigenvalues). This does not abide by the usual linear algebraic definition of eigenvectors in Hilbert space, however, they will be treated as eigenvector-eigenvalue pairs fo the quadrature operators as in other CV quantum information literatures. Lastly, we note that the eigenstates of both quadrature form the following relation

$$|p\rangle = \frac{1}{2\sqrt{\pi}} \int e^{-iqp/2}|q\rangle \, dq \quad \text{and} \quad |q\rangle = \frac{1}{2\sqrt{\pi}} \int e^{-iqp/2}|p\rangle \, dp.$$

2.2.2 Bosonic and Quadrature Field Representations of Quantum Harmonic Oscillator

To further discuss the relationship and equivalence between quadrature and bosonic definitions of a quantum state of a system, we recall the quantum harmonic oscillator system in [**Example 2.1.12**] with observables $\widehat{q}$ and $\widehat{p}$, and Hamiltonian

$$H = \frac{1}{2m}\widehat{p}^2 + \frac{m\omega^2}{2}\widehat{q}^2.$$

We will derive its definition in terms of the bosonic field operators of a single mode, as in [Lou00, Chapter 4.3], consistent with the wave-particle duality of light, with quadrature observables $\widehat{q}$ and $\widehat{p}$ for wave definition, and bosonic observables $\widehat{a}^\dagger$ and $\widehat{a}$ for particle definition. In a similar manner to (2.2.1), we define the creation and annihilation operators $\widehat{a}$ and $\widehat{a}^\dagger$ in terms of our quantum harmonic oscillator system

$$\widehat{a} = \frac{1}{\sqrt{2m\hbar\omega}}(m\omega\widehat{q} + i\widehat{p}) \quad \text{and} \quad \widehat{a}^\dagger = \frac{1}{\sqrt{2m\hbar\omega}}(m\omega\widehat{q} - i\widehat{p}). \tag{2.2.7}$$

Considering the number operator

$$\begin{aligned}
\widehat{a}^\dagger\widehat{a} &= \frac{1}{2m\hbar\omega}(\widehat{p}^2 + (m\omega\widehat{q})^2 + im\omega\widehat{q}\widehat{p} - im\omega\widehat{p}\widehat{q}) \\
&= \frac{1}{\hbar\omega}\left(\frac{1}{2m}\widehat{p}^2 + \frac{m\omega^2}{2}\widehat{q}^2 + \frac{i\omega}{2}[\widehat{q},\widehat{p}]\right) \\
&= \frac{1}{\hbar\omega}\left(H - \frac{\hbar\omega}{2}I\right)
\end{aligned} \tag{2.2.8}$$

[3] This is due to the commutation relation between $\widehat{q}$ and $\widehat{p}$ as shown in [Ser17, Chapter 2.1].

where the last equality is by $[\widehat{q},\widehat{p}] = \hbar i$ in comparison with (2.2.3) and (2.2.4) where we set $\hbar = 2$. Rearranging the terms in (2.2.8), we can now write the Hamiltonian in terms of number operator

$$H = \hbar\omega(\widehat{a}^\dagger\widehat{a} + I/2). \tag{2.2.9}$$

In showing the increase and decrease of energy in the oscillator (or equivalently, of number of photons) caused by applying the creation and annihilation operators on a fock state $|n\rangle$, let eigenvalue λ_n of H and its corresponding fock eigenstate $|n\rangle$, then we have

$$\begin{aligned}
\hbar\omega(\widehat{a}^\dagger\widehat{a} + I/2)|n\rangle &= \lambda_n|n\rangle \\
\hbar\omega(\widehat{a}^\dagger\widehat{a}^\dagger\widehat{a} + \widehat{a}^\dagger/2)|n\rangle &= \lambda_n\widehat{a}^\dagger|n\rangle \\
\hbar\omega(\widehat{a}^\dagger(\widehat{a}\widehat{a}^\dagger - I) + \widehat{a}^\dagger/2)|n\rangle &= \lambda_n\widehat{a}^\dagger|n\rangle \\
H\widehat{a}^\dagger|n\rangle &= (\lambda_n + \hbar\omega)\widehat{a}^\dagger|n\rangle
\end{aligned} \tag{2.2.10}$$

where the third equality was obtained using commutation relation $[\widehat{a}^\dagger,\widehat{a}] = -I$. From (2.2.10), it can be seen that $(\lambda_n + \hbar\omega)$ is an eigenvalue of H with eigenstate $\widehat{a}^\dagger|n\rangle$. We now define $c_n|n+1\rangle = \widehat{a}^\dagger|n\rangle$ for some constant $c_n \in \mathbb{C}$ and $\lambda_{n+1} = (\lambda_n + \hbar\omega)$, which is consistent with equivalence between the interpretations of an increase in the number of photons in the (single mode) system and an increase of energy level by amount of $\hbar\omega$. Similarly for the annihilation operator $\widehat{a}$,

$$\begin{aligned}
\hbar\omega(\widehat{a}\widehat{a}^\dagger\widehat{a} + \widehat{a}/2)|n\rangle &= \lambda_n\widehat{a}|n\rangle \\
\hbar\omega((\widehat{a}^\dagger\widehat{a} + I)\widehat{a} + \widehat{a}/2)|n\rangle &= \lambda_n\widehat{a}|n\rangle \\
H\widehat{a}|n\rangle &= (\lambda_n - \hbar\omega)\widehat{a}|n\rangle.
\end{aligned}$$

Then, defining $a_n|n-1\rangle = \widehat{a}|n\rangle$ for some constant $a_n \in \mathbb{C}$ and $\lambda_{n-1} = (\lambda_n - \hbar\omega)$ gives equivalence of interpretations between a decrease of the photon number and a decrease of energy level by $\hbar\omega$.

However, because the number of photons as well as the amount of energy in our oscillator must be non-negative, we must consider the case of the vacuum state $n = 0$. To model this physical event, we define $\widehat{a}|0\rangle = 0$ indicating the *vacuum state condition*, and therefore

$$H|0\rangle = \hbar\omega(\widehat{a}^\dagger\widehat{a} + I/2)|0\rangle = \frac{\hbar\omega}{2}|0\rangle = \lambda_0|0\rangle$$

and consequently, $\lambda_0 = \hbar\omega/2$. Recalling that $\lambda_{n+1} = \lambda_n + \hbar\omega$, we can now rewrite it as $\lambda_n = \hbar\omega(n + 1/2)$ for $n \in \mathbb{N}$. Putting together our new definition of the eigenvalues with H, we get

$$\begin{aligned}
H|n\rangle &= \lambda_n|n\rangle \\
\hbar\omega(\widehat{a}^\dagger\widehat{a} + I/2)|n\rangle &= \hbar\omega(n + 1/2)|n\rangle \\
\widehat{a}^\dagger\widehat{a}|n\rangle &= n|n\rangle,
\end{aligned}$$

that is, the number operator $\widehat{n} = \widehat{a}^\dagger\widehat{a}$, which is an observable of the number of photons in, or equivalently, the energy level of the oscillator.

We now impose the normalization property of Fock eigenstates $\{|n\rangle\}_{n\in\mathbb{N}}$ where $\langle n|n\rangle = 1$ for all $n \in \mathbb{N}$ to derive the value of constants c_n and a_n by considering $\widehat{a}^\dagger|n\rangle = c_n|n+1\rangle$ and $\widehat{a}|n\rangle = a_n|n-1\rangle$, then by

$$|a_n|^2\langle n+1|n+1\rangle = \langle n|\widehat{a}\widehat{a}^\dagger|n\rangle = \langle n|(\widehat{a}^\dagger\widehat{a} + I|n\rangle = (n+1)\langle n|n\rangle$$

and

$$|a_n|^2\langle n-1|n-1\rangle = \langle n|\widehat{a}^\dagger\widehat{a}|n\rangle = n\langle n|n\rangle$$

we have $|c_n|^2 = n+1$ and $|a_n|^2 = n$. Then, rewriting $\widehat{q}$ and $\widehat{p}$ in terms of $\widehat{a}^\dagger$ and $\widehat{a}$, we first consider

$$\widehat{q} = \frac{\sqrt{2m\hbar\omega}\widehat{a}^\dagger + i\widehat{p}}{m\omega} \quad \text{and} \quad \widehat{p} = i(m\omega\widehat{q} - \sqrt{2m\hbar\omega}\widehat{a})$$

by rearranging (2.2.7). Substituting $\widehat{q}$ into $\widehat{a}$ and $\widehat{p}$ into $\widehat{a}^\dagger$ of (2.2.7) give us

$$\sqrt{2m\hbar\omega}\widehat{a} = \sqrt{2m\hbar\omega}\widehat{a}^\dagger + 2i\widehat{p}$$

$$\widehat{p} = \frac{i\sqrt{2m\hbar\omega}}{2}(\widehat{a}^\dagger - \widehat{a})$$

and

$$\sqrt{2m\hbar\omega}\widehat{a}^\dagger = m\omega\widehat{q} - m\omega\widehat{q} - \sqrt{2m\hbar\omega}\widehat{a}$$

$$\widehat{q} = \sqrt{\frac{\hbar}{2m\omega}}(\widehat{a}^\dagger + \widehat{a}).$$

To further simplify our oscillator by eliminating dimensions m and ω, we define the dimensionless quadrature operators [BL05, Section II A] [Lou00, Chapter 4.3] as

$$\widehat{q} = \sqrt{\frac{2m\omega}{\hbar}}\widehat{q} \quad \text{and} \quad \widehat{p} = \sqrt{\frac{2}{m\hbar\omega}}\widehat{p} \tag{2.2.11}$$

giving us $\widehat{q} = \widehat{a}^\dagger + \widehat{a}$ and $\widehat{p} = i(\widehat{a}^\dagger - \widehat{a})$ where inversely,

$$\widehat{a}^\dagger = \frac{1}{2}(\widehat{q} - i\widehat{p}) \quad \text{and} \quad \widehat{a} = \frac{1}{2}(\widehat{q} + i\widehat{p}).$$

Now writing the Hamiltonian H of our quantum harmonic oscillator in terms of $\widehat{q}$ and $\widehat{p}$ using (2.2.11), we get

$$H = \frac{1}{2m}\left(\sqrt{\frac{m\hbar\omega}{2}}\widehat{p}\right)^2 + \frac{m\omega^2}{2}\left(\sqrt{\frac{m\hbar}{2m\omega}}\widehat{q}\right)^2 = \frac{\hbar\omega}{4}(\widehat{p}^2 + \widehat{q}^2)$$

while writing it in terms of creation and annihilation operator

$$H = \frac{\hbar\omega}{4}((i(\widehat{a}^\dagger - \widehat{a}))^2 + (\widehat{a}^\dagger + \widehat{a})^2) = \frac{\hbar\omega}{4}(2\widehat{a}^\dagger\widehat{a} + 2\widehat{a}\widehat{a}^\dagger) = \hbar\omega(\widehat{a}^\dagger\widehat{a} + I/2)$$

as in (2.2.9), where $\widehat{a}^\dagger\widehat{a}|n\rangle = n|n\rangle$ with

$$\widehat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle \quad \text{and} \quad \widehat{a}|n\rangle = \begin{cases} 0|n\rangle & \text{if } n = 0 \\ \sqrt{n}|n-1\rangle & \text{if } n \neq 0 \end{cases}$$

consistently with [**Definition 2.2.1**] and [**Definition 2.2.2**].

2.2.3 Quantum States of Light

With the introductions of the Fock states and quadrature states, we are now ready to define other quantum states of light that will be relevant to our discussions later. Definitions of displacement and squeezing operators will also be stated, to facilitate the definitions of these states, as other operations on an optical CV quantum system will be defined as well. We will state their definitions in what in quantum mechanics are called the Schrödinger picture and the Heisenberg picture. They can be understood as the description of the evolution of the system in terms of a state ρ and in terms of an observable A, respectively. The operators are defined differently for the bosonic description and the quadrature description of a system, the same letter will be used to indicate both operators, but with the operators on the bosonic system written in a boldface style. First, we will introduce the displacement operator and the coherent state.

Definition 2.2.3 (Displacement Operator (or Weyl Operator) [Wee+12, Section II.A.1], [BL07, Chapter 3, p.44], [Wan+07, Chapter 1.1, p.7], [FOP05, Chapter 1.4.1])**.** In an m-mode quantum system, we define the *Displacement Operator* (or *Weyl Operator*) in terms of the quadrature operators as

$$D(\xi) = \mathbf{exp}(i\widehat{x}^\top \Omega \xi)$$

where $\xi \in \mathbb{R}^{2m}$ and vector of quadrature operators $\widehat{x} = (\widehat{q}_1, \widehat{p}_1, \ldots, \widehat{q}_m, \widehat{p}_m)$.

Whereas in terms of the bosonic operators and $\alpha_1, \ldots, \alpha_m \in \mathbb{C}$, we define the displacement operators as

$$\boldsymbol{D}(\alpha) = \mathbf{exp}\left(\sum_{j=1}^{m} \alpha_j \widehat{a}_j^\dagger - \alpha_j^* \widehat{a}_j\right)$$

which is equivalent to the definition in terms of the quadrature operators given that $\alpha_j = \frac{1}{2}(\xi_{2j} + i\xi_{2j+1})$ and $\widehat{a}_j = \frac{1}{2}(\widehat{q}_{2j} + i\widehat{p}_{2j+1})$. ◀

Note that the operator exponential $\mathbf{exp}(i\widehat{x}\Omega\xi)$ can be expanded as

$$\begin{aligned}
\mathbf{exp}(i\widehat{x}^\top \Omega \xi') &= \mathbf{exp}\left(i(\widehat{q}_1, \widehat{p}_1, \ldots, \widehat{q}_m, \widehat{p}_m) \begin{bmatrix} 0 & 1 & & & \\ -1 & 0 & & & \\ & & \ddots & & \\ & & & 0 & 1 \\ & & & -1 & 0 \end{bmatrix} \begin{bmatrix} \xi_1' \\ \vdots \\ \xi_{2m}' \end{bmatrix} \right) \\
&= \mathbf{exp}(i(\xi_2'\widehat{q}_1 - \xi_1'\widehat{p}_1 + \xi_4'\widehat{q}_2 - \xi_3'\widehat{p}_2 + \cdots + \xi_{2m}'\widehat{q}_m + \xi_{2m-1}'\widehat{p}_m)) \\
&= \sum_{k\in\mathbb{N}} \frac{1}{k!} \left(\sum_{j=1}^{m} \xi_{2j}'\widehat{q}_j - \xi_{2j-1}'\widehat{p}_j \right)^k .
\end{aligned}$$

Where in the simpler case of a single mode system, we have

$$D(\xi) = \mathbf{exp}\left(i(\widehat{q}, \widehat{p}) \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} (\xi_1', \xi_2')^\top \right) = \mathbf{exp}(i\xi_2'\widehat{q} - i\xi_1'\widehat{p}) = \sum_{k\in\mathbb{N}} \frac{1}{k!} (i\xi_2'\widehat{q} - i\xi_1'\widehat{p})^k.$$

The following result shows how the annihilation operator and the quadrature operators each behave when the displacement operator is applied to it, which will be useful later in our discussion.

Proposition 2.2.4 (Bosonic and Quadrature operator displacement)**.** *For $\alpha \in \mathbb{C}$, we have the displacement operator operating on the annihilation operator $\widehat{a}$ as*

$$\boldsymbol{D}^\dagger(\alpha)\widehat{a}\boldsymbol{D}(\alpha) = \widehat{a} + \alpha I.$$

While for $x = (q, p)^\top \in \mathbb{R}^2$ and the quadrature operators $\widehat{q}$ and $\widehat{p}$, we have

$$D^\dagger(x) \begin{bmatrix} \widehat{q} \\ \widehat{p} \end{bmatrix} D(x) = \begin{bmatrix} \widehat{q} \\ \widehat{p} \end{bmatrix} + \begin{bmatrix} qI \\ pI \end{bmatrix}.$$

The proof of this proposition will be omitted and interested readers should refer to [BR97, Chapter 3.6] or [Ser17, Chapter 3.2]. We are now ready to define an important state in quantum optics that will be relevant in our discussions later.

Definition 2.2.5 (Coherent state)**.** Considering a single mode system, we call $|\alpha\rangle = \boldsymbol{D}(\alpha)|0\rangle$ a *coherent state*. ◀

The coherent state can therefore be thought of as a displaced vacuum state, given its preparation by operating the the displacement operator on a vacuum state $\boldsymbol{D}(\alpha)|0\rangle = |\alpha\rangle$. For $\alpha \in \mathbb{C}$, coherent state $|\alpha\rangle$ is an eigenstate of annihilation operator $\widehat{a}$ with α as its corresponding eigenvalue, as can be seen from

$$\begin{aligned} \widehat{a}|\alpha\rangle &= \widehat{a}\boldsymbol{D}(\alpha)|0\rangle \\ &= \boldsymbol{D}(\alpha)\boldsymbol{D}^\dagger(\alpha)\widehat{a}\boldsymbol{D}(\alpha)|0\rangle \\ &= \boldsymbol{D}(\alpha)(\widehat{a} + \alpha)|0\rangle \\ &= \alpha\boldsymbol{D}(\alpha)|0\rangle \\ &= \alpha|\alpha\rangle \end{aligned}$$

where we have used the identity between displacement and annihilation operator above and $\widehat{a}|0\rangle = 0$.

Now we define a squeezing operator that decreases the variance of one of the quadratures in a quantum system described in terms of quadratures. This however, will be clear later when we define Gaussian states, which can be described by its statistical moments. Now we will define it in terms of the bosonic operators. Definition of this state will make way for us to define a squeezed state, which intuitively is a vacuum state with one of its quadrature variance decreased and the other increased.

Definition 2.2.6 (Squeezing operator and Squeezed state)**.** For a *Squeezing parameter* $r \in \mathbb{R}$, we define the *Squeezing operator* on a quantum system described on a Hilbert space $\mathcal{H}$ as

$$\boldsymbol{S}(r) = \mathbf{exp}\left(\frac{r}{2}(\widehat{a}^2 - \widehat{a}^{\dagger 2})\right).$$

Operating on the annihilation operator $\widehat{a}$ gives $S^\dagger(r)\widehat{a}S(r) = \widehat{a}\cosh r - \widehat{a}^\dagger \sinh r$. Having it operating on the quadrature operators require us to define

$$S(r) = \begin{bmatrix} e^{-r} & 0 \\ 0 & e^{r} \end{bmatrix}$$

and thus squeezing the quadrature operators gives

$$S(r)\begin{bmatrix} \widehat{q} \\ \widehat{p} \end{bmatrix} = \begin{bmatrix} \widehat{q}e^{-r} \\ \widehat{p}e^{r} \end{bmatrix}.$$

For $r \in \mathbb{R}$, we define the *Squeezed State* of a single mode system as

$$|r\rangle = S(r)|0\rangle$$

or in the more general density matrix notation as $\rho_r = S(r)|0\rangle\langle 0|S^\dagger(r)$. ◀

A derivation of the operation of the squeezing operator on the annihilation operator $\widehat{a}$ is omitted here, although interested readers may refer to [BR97, Chapter 3.7] or [Leo97, Chapter 2.3]. The squeezed state will be highly relevant in our discussions on GBS later where we will have Gaussian states of the system described as their statistical moments. Thus, we will later define both squeezing operator and the squeezed state in the form of statistical moments of a quantum state.

Another important state in quantum optics and CV quantum information is the Thermal state, we define its density operator as follows.

Definition 2.2.7 (Thermal state)**.** For expected photon number $\bar{n} \in \mathbb{R}_{\geq 0}$, the density operator state description in a single mode system

$$\rho_{\bar{n}}^{\text{Th}} = \frac{1}{\bar{n}+1}\sum_{n\in\mathbb{N}}\left(\frac{\bar{n}}{\bar{n}+1}\right)^n |n\rangle\langle n|$$

is called a *Thermal state*. ◀

Lastly, we define another important optical operation called the *Beamsplitter*, which "mixes" two modes in a system by interfering the two light sources.

Definition 2.2.8. Given a two-mode system in a state $\rho_1 \otimes \rho_2$ we define the *beamsplitter* that interferes the two modes as

$$B(\theta) = \mathbf{exp}(\theta(\widehat{a}_1^\dagger \widehat{a}_2 - \widehat{a}_1\widehat{a}_2^\dagger))$$

where subscripts 1 and 2 on the bosonic operators indicate their corresponding mode.

Considering two-mode system annihilation operators $\widehat{a}_1$ and $\widehat{a}_2$, we define a beamsplitter with *transmissivity* parameter $\tau = \cos^2\theta \in [0,1]$ as

$$\boldsymbol{B}(\tau) = \begin{bmatrix} \sqrt{\tau} & -\sqrt{1-\tau} \\ \sqrt{1-\tau} & \sqrt{\tau} \end{bmatrix},$$

which operation on these bosonic modes gives

$$\boldsymbol{B}(\tau)\begin{bmatrix}\widehat{a}_1\\ \widehat{a}_2\end{bmatrix} = \begin{bmatrix}\widehat{a}_1\sqrt{\tau} - \widehat{a}_2\sqrt{1-\tau}\\ \widehat{a}_1\sqrt{1-\tau} + \widehat{a}_2\sqrt{\tau}\end{bmatrix}.$$

Analogous to the beamsplitter operations on the annihilation operators we define the beamsplitter on the quadrature operators $\widehat{q}$ and $\widehat{p}$ as

$$B(\tau) = \begin{bmatrix}\sqrt{\tau}I & -\sqrt{1-\tau}I\\ \sqrt{1-\tau}I & \sqrt{\tau}I\end{bmatrix},$$

for 2×2 identity matrix I, while operating it on the quadrature operators gives

$$B(\tau)\begin{bmatrix}\widehat{q}_1\\ \widehat{p}_1\\ \widehat{q}_2\\ \widehat{p}_2\end{bmatrix} = \begin{bmatrix}\widehat{q}_1\sqrt{\tau} - \widehat{q}_2\sqrt{1-\tau}\\ \widehat{p}_1\sqrt{\tau} - \widehat{p}_2\sqrt{1-\tau}\\ \widehat{q}_1\sqrt{1-\tau} + \widehat{q}_2\sqrt{\tau}\\ \widehat{p}_1\sqrt{1-\tau} + \widehat{p}_2\sqrt{\tau}\end{bmatrix}.$$

◀

Phase Shifted Bosonic and Quadrature Operators

We first need to define phase shifting with a *Phase Shifter* operator $U(\theta) = \mathbf{exp}(i\theta\widehat{a}^\dagger\widehat{a})$ on creation and annihilation operators; that is by $U^\dagger(\theta)\widehat{a}^\dagger U(\theta)$ and $U^\dagger(\theta)\widehat{a}U(\theta)$. We evaluate them by solving their differential equations calculated from their first derivative as follows

$$\begin{aligned}\frac{d}{d\theta}U^\dagger(\theta)\widehat{a}^\dagger U(\theta) &= \left(\frac{d}{d\theta}U^\dagger(\theta)\right)\widehat{a}^\dagger U(\theta) + U^\dagger\widehat{a}^\dagger\left(\frac{d}{d\theta}U(\theta)\right)\\ &= -i(U^\dagger(\theta)\widehat{a}^\dagger\widehat{a}\widehat{a}^\dagger U(\theta) - U^\dagger(\theta)\widehat{a}^\dagger U(\theta)\widehat{a}^\dagger\widehat{a})\\ &= -iU^\dagger(\theta)(\widehat{a}^\dagger\widehat{a}\widehat{a}^\dagger - \widehat{a}^\dagger\widehat{a}^\dagger\widehat{a})U(\theta)\\ &= -iU^\dagger(\theta)(\widehat{a}^\dagger[\widehat{a},\widehat{a}^\dagger])U(\theta)\\ &= -iU^\dagger(\theta)\widehat{a}^\dagger U(\theta)\end{aligned}$$

and similarly

$$\begin{aligned}\frac{d}{d\theta}U^\dagger(\theta)\widehat{a}U(\theta) &= -iU(\theta)(\widehat{a}^\dagger\widehat{a}\widehat{a} - \widehat{a}\widehat{a}^\dagger\widehat{a})U(\theta)\\ &= iU(\theta)([\widehat{a},\widehat{a}^\dagger]\widehat{a})U(\theta)\\ &= iU^\dagger(\theta)\widehat{a}U(\theta)\end{aligned}$$

because $[\widehat{a},\widehat{a}^\dagger] = I$ and $[\widehat{a}^\dagger,\widehat{a}] = -I$. Considering $U^\dagger\widehat{a}^\dagger U$ as a function f such that $\frac{d}{d\theta}f(\theta) = -if(\theta)$, then we must have $f(\theta) = e^{-i\theta}$, but if $\theta = 0$, then $U^\dagger(\theta)\widehat{a}^\dagger U(\theta) = \widehat{a}^\dagger$. Thus, we can deduce that $U^\dagger(\theta)\widehat{a}^\dagger U(\theta) = \widehat{a}^\dagger e^{-i\theta}$, and with a similar argument of having $U^\dagger\widehat{a}U = g$ and $\frac{d}{d\theta}g(\theta) = ig(\theta)$, we also have $U^\dagger(\theta)\widehat{a}U(\theta) = \widehat{a}e^{i\theta}$.

Proposition 2.2.9 (Phase Shift)**.** *For a Phase Shifter Operator $U(\theta) = \mathbf{exp}(i\theta\widehat{a}^\dagger\widehat{a})$ on a single mode bosonic system, we have*

$$U^\dagger(\theta)\widehat{a}U(\theta) = \widehat{a}e^{i\theta} \quad \textit{and} \quad U^\dagger(\theta)\widehat{a}^\dagger U(\theta) = \widehat{a}^\dagger e^{-i\theta}.$$

For the purpose of later derivations, we now list a few important properties of $U(\theta)$ that follows from the properties of operator exponential. Expanding the complex operator exponential with Taylor Expansion, we get

$$\begin{aligned} U(\theta) &= \mathbf{exp}(i\theta\widehat{n}) \\ &= \cos(\theta\widehat{n}) + i\sin(\theta\widehat{n}) \\ &= \left(\frac{(\theta\widehat{n})^0}{0!} - \frac{(\theta\widehat{n})^2}{2!} + \frac{(\theta\widehat{n})^4}{4!} - \dots\right) - i\left(\frac{(\theta\widehat{n})^1}{1!} - \frac{(\theta\widehat{n})^3}{3!} + \dots\right); \end{aligned}$$

secondly, by noting that $\mathbf{Det}(\mathbf{exp}(i\theta\widehat{n})) = \mathbf{exp}(\mathbf{Tr}(i\theta\widehat{n})) \neq 0$, we can see that $U(\theta)$ is invertible for any θ; and lastly, that $U(\theta)$ is unitary (i.e: $U^\dagger(\theta) = U^{-1}(\theta)$), which is clear from $U(\theta)U^\dagger(\theta) = U^\dagger(\theta)U(\theta) = \mathbf{exp}(i\theta\widehat{n} - i\theta\widehat{n}) = I$.

We now consider phase shifting operation on the quadrature operators $U^\dagger(\theta)\widehat{q}U(\theta)$ and $U^\dagger(\theta)\widehat{p}U(\theta)$. Expressing them in terms of $\widehat{a}^\dagger$ and $\widehat{a}$ with $\hbar = 2$, we get

$$\begin{aligned} U^\dagger(\theta)\widehat{q}U(\theta) &= U^\dagger(\theta)(\widehat{a}^\dagger + \widehat{a})U(\theta) \\ &= \widehat{a}^\dagger e^{-i\theta} + \widehat{a}e^{i\theta} \\ &= \frac{1}{2}((\widehat{q} - i\widehat{p})(\cos\theta - i\sin\theta) + (\widehat{q} + i\widehat{p})(\cos\theta + i\sin\theta)) \\ &= \widehat{q}\cos\theta - \widehat{p}\sin\theta \end{aligned} \tag{2.2.12}$$

where relations (2.2.1) and (2.2.2) were used to obtain the third equality. Similarly applying the phase shifter operator to the momentum operator,

$$\begin{aligned} U^\dagger(\theta)\widehat{p}U(\theta) &= iU^\dagger(\theta)(\widehat{a}^\dagger - \widehat{a})U(\theta) \\ &= i(\widehat{a}^\dagger e^{-i\theta} - \widehat{a}e^{i\theta}) \\ &= \frac{i}{2}((\widehat{q} - i\widehat{p})(\cos\theta - i\sin\theta) - (\widehat{q} + i\widehat{p})(\cos\theta + i\sin\theta)) \\ &= \widehat{q}\sin(\theta) + \widehat{p}\cos(\theta). \end{aligned}$$

We will, more compactly write these phase-shifted quadrature operators as follows.

Definition 2.2.10 (Phase-shifted Quadrature Operator)**.** We write the *Phase-shifted Quadrature Operators* for an m-mode system as $\widehat{q}_\theta = (\widehat{q}_{1,\theta}, \dots, \widehat{q}_{m,\theta})^\top$ and $\widehat{p}_\theta = (\widehat{p}_{1,\theta}, \dots, \widehat{p}_{m,\theta})^\top$ where for each individual modes, we have

$$\widehat{q}_{k,\theta} = \widehat{q}_k\cos(\theta) - \widehat{p}_k\sin(\theta) \quad \text{and} \quad \widehat{p}_{k,\theta} = \widehat{q}_k\sin(\theta) + \widehat{p}_k\cos(\theta).$$

If $\theta = 0$, then $\widehat{q}_k = \widehat{q}_{k,\theta}$ and $\widehat{p}_k = \widehat{p}_{k,\theta}$.

Thus, phase shifter operation on quadrature operators in a single mode system can be defined as

$$\bar{U}(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

which gives

$$\bar{U}(\theta)\begin{bmatrix} \widehat{q} \\ \widehat{p} \end{bmatrix} = \begin{bmatrix} \widehat{q}\cos(\theta) - \widehat{p}\sin(\theta) \\ \widehat{q}\sin(\theta) + \widehat{p}\cos(\theta) \end{bmatrix}.$$

◀

As $\cos(\theta + \frac{\pi}{2}) = -\sin(\theta$ and $\sin(\theta + \frac{\pi}{2}) = \cos\theta$, it is clear that $\widehat{q}_{\theta-\frac{\pi}{2}} = \widehat{p}_\theta$. Also, as with the quadrature operator with no phase shifts in [**Section 2.2.1**], for each mode we have both $\widehat{q}_{k,\theta}$ and $\widehat{p}_{k,\theta}$ as observables with eigenvalues $q_{k,\theta}$ and $p_{k,\theta}$ and eigenstate $|q_{k,\theta}\rangle$ and $|p_{k,\theta}\rangle$. Moreover, eigenvalues of $\widehat{q}$ are eigenvalues of $\widehat{q}_\theta$ with phase-shifted eigenstate $|q\rangle$, as it is the same for $\widehat{p}$.

Proposition 2.2.11. *For $z \in \{q, p\}$, if $z \in \mathbb{R}$ is an eigenvalue of $\widehat{z}$ with eigenstate $|z\rangle$, then z is an eigenvalue of $\widehat{z}_\theta$ with eigenstate $U^\dagger(\theta)|z\rangle$.*

This fact can easily be shown using the definition of an eigenvalue - eigenstate below,

$$\begin{aligned}\widehat{z}_\theta U^\dagger(\theta)|z\rangle &= U^\dagger(\theta)\widehat{z}U(\theta)U^\dagger(\theta)|z\rangle \\ &= U^\dagger(\theta)\widehat{z}|z\rangle \\ &= zU^\dagger(\theta)|z\rangle.\end{aligned}$$

Where we have also used the unitary property of $U(\theta)$ and the definition of $\widehat{z}$.

Now, as we had done for the quadrature operators, we look into the expectation of the phase shifted quadrature operator, which evaluates to an integral as follows.

$$\begin{aligned}\langle\widehat{z}_\theta\rangle_\rho &= \mathbf{Tr}(\rho\widehat{z}_\theta) \\ &= \mathbf{Tr}(\rho U^\dagger(\theta)\widehat{z}U(\theta)) \\ &= \mathbf{Tr}(U(\theta)\rho U^\dagger(\theta)\widehat{z}) \\ &= \int_{\mathbb{R}} z\langle z|U(\theta)\rho U^\dagger(\theta)|z\rangle\, dz\end{aligned}$$

where we have used the cyclic property of the **Tr** function to obtain the third equality. The $\langle z|U(\theta)\rho U^\dagger(\theta)|z\rangle$ term in the integral induces a probability measure over all eigenvalues $z \in \mathbb{R}$ using phase shifted state ρ of the system as opposed to the non-phase-shifted state ρ for the expectation of $\widehat{z}$ in (2.2.5). Thus we now define a notation for the phase-shifted quadrature operators.

Definition 2.2.12 (Phase-shifted Quandrature operator.)**.** For a phase-shifted quadrature operators on a single mode system $\widehat{q}_\theta$ and $\widehat{p}_\theta$ we denote $\{|q_\theta\rangle\}$ and $\{|p_\theta\rangle\}$ as their eigenstates and $\{q_\theta\}$ and $\{p_\theta\}$ as their eigenvalues. We used the shorthand $|z_\theta\rangle = U^\dagger(\theta)|z\rangle$ for $z \in \{q, p\}$ and write the probability of obtaining measurement outcome $z \in \mathbb{R}$ from observable $\widehat{z}_\theta$ as $\mathbb{P}(z, \theta) = \langle z|U(\theta)\rho U^\dagger(\theta)|z\rangle = \langle z_\theta|\rho|z_\theta\rangle$. ◀

This phase-shifted quadrature operators will be relevant in one of the main parts of his thesis, which is on the application of Homodyne Tomography method on GBS. This method uses a measurement uniquely for CV systems called the Homodyne measurement, which measures the observable $\widehat{q}_\theta$ as we will see in [**Section 2.2.5**].

2.2.4 Quantum Mechanics in Phase Space

In this section, we will discuss the description of a quantum system in *Phase Space* as an alternative to infinite-dimensional Hilbert space description. First, we will state the definition of a phase space.

Definition 2.2.13 (Phase Space [ARL14, Chapter 2.2])**.** For an m mode system and a *Symplectic Form* matrix $\Omega \in \mathbb{R}^{2m,2m}$ where

$$\Omega = \bigoplus_{j=1}^{m} \Omega^{(j)} = \begin{bmatrix} \Omega^{(1)} & & \\ & \ddots & \\ & & \Omega^{(m)} \end{bmatrix} \quad \text{and} \quad \Omega^{(j)} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

we call $\Gamma = (\mathbb{R}^{2m}, \Omega)$ a *Phase Space*. ◀

We note that the Symplectic form matrix Ω is based on the commutation relation of the bosonic and quadrature operators as described in [**Section 2.2.1**]. Also, instead of having a product of infinite dimensional Hilbert space to describe the state of an m mode system, we can instead describe the state of the system in $2m$ dimensional real number space. As we proceed, we will later see that for a class of quantum states called the Gaussian states, which are particularly relevant in later discussions on GBS, can be very conveniently described only by a $2m$ dimensional real vector and a $2m \times 2m$ real matrix in the phase space. But, now we must first define the functions on the phase space that we use to describe a quantum state, called the quasiprobability distribution functions.

Quasiprobability Distribution State Representation

In the phase space, the corresponding state representation of a density operator ρ in the Hilbert space can be represented by a quasiprobability distribution (QPD) function on the phase space. This function is defined in terms of the *Characteristic Function* which is central in showing the relation between the density operator state representation and the quasiprobability distribution state representation.

Definition 2.2.14 (Characteristic Function [Wee+12, Section I.A.1], [BL07, Chapter 3.3.1], [BR97, Chapter 4.4], [ARL14, Chapter 2.2])**.** For an m dimensional system with operator F on Hilbert space product $\mathcal{H}^{\otimes m}$ and ordering parameter $s \in [-1, 1]$, we have the *s-parametrized Characteristic function* that describes it

$$\chi_F(\xi; s) = \mathbf{Tr}(FD(\xi))\, \mathbf{exp}(s||\xi||^2/2)$$

for $\xi \in \mathbb{R}^{2m}$. If $s = 0$, we write $\chi_F(\xi) = \chi_F(\xi; 0)$ and simply call it the *Characteristic function*. Where the displacement operator is expressed in terms of the bosonic operators $\boldsymbol{D}(\alpha)$, for $\alpha \in \mathbb{C}^m$ the characteristic function is also defined as $\chi_F(\alpha; s)$. ◀

This is defined in a more general manner in terms of a general operator on Hilbert space as we will often discuss the characteristic function χ_ρ as a representation of quantum state density operator ρ. We note as well that this definition contains another generality in the parameter $s \in [-1, 1]$ that indicates the ordering of the annihilation and creation operators in the bosonic displacement operator $\boldsymbol{D}(\alpha)$. The characteristic function is an important concept in phase space description of a quantum system as it provides a direct relation between it and any operator in the Hilbert space. This fact is stated more precisely in the following proposition.

Proposition 2.2.15 (Fourier-Weyl relation [Ser17, Chapter 4.2], [FOP05, Section 1.4.1, eq.(1.37)], [Wan+07, Section 1.2, p.7], [LPD95, Section 2, eq.(3)], [Nav15, Chapter 4.3.3, eq.(4.54)]). *For an operator $F : \mathcal{S} \to \mathcal{S}$ on Hilbert space $\mathcal{S}$ and $\xi \in \mathbb{R}^{2m}$, we have*

$$F = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^{2m}} \chi_F(\xi) D^\dagger(\xi)\, d\xi.$$

As F corresponds to a general operator, this relationship can be stated more specifically for density operator state

$$\rho = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^{2m}} \chi_\rho(\xi) D^\dagger(\xi)\, d\xi. \tag{2.2.13}$$

Considering the fact that the displacement operators forms a complete orthogonal basis of the Hilbert space of linear operators on $\mathbb{R}^{2m}$, denoted as $L^2(\mathbb{R}^{2m})$ [4] , this relationship can be interpreted as showing that a density operator state description ρ (or any operator in general) on Hilbert space $\mathcal{H}^{\otimes m}$ can be expressed as an integral of the displacement operators over $\xi \in \mathbb{R}^{2m}$ that is induced by the measure χ_ρ.

Having discussed about the characteristic function, we are now ready to begin the discussion on the QPDs. First, consider an n dimensional Fourier Transform $\mathcal{F}$ on an arbitrary function $f : \mathbb{R}^n \to \mathbb{R}^n$

$$\mathcal{F}f(\xi) = \int_{\mathbb{R}^n} \mathbf{exp}(-2\pi i x^\top \xi) f(x)\, dx.$$

To obtain the s parametrized quasiprobability distribution (QPD) function, we perform a $2m$ dimensional symplectic Fourier Transform on $\chi_\rho(x; s)$ as follows.

Definition 2.2.16 (s parametrized Quasiprobability Distribution [ARL14, Section 2.2], [Wee+12, Chapter II.A.1], [BL07, Chapter 3], [Leo97, Chapter 3.2]). For an m mode quantum system with a density operator state ρ, we define its *s parametrized Quasiprobability Distribution (QPD)* state representation as

$$W_\rho(x; s) = \frac{1}{(2\pi)^{2m}} \int_{\mathbb{R}^{2m}} \mathbf{exp}(-i x^\top \Omega \xi) \chi_\rho(\xi; s)\, d^{2m}\xi$$

where $s \in [-1, 1]$ and $x = (q_1, p_1, \ldots, q_m, p_m)$. For $s = 0$, we write $W_\rho(x) = W_\rho(x; 0)$, which we call the *Wigner Function*. For $s = -1$, we have the *Q Function* $Q(x) = W(x; -1) = \frac{1}{2\pi}\langle\alpha|\rho|\alpha\rangle$ where $|\alpha\rangle = \bigotimes_j |\alpha_j\rangle$ and $\alpha_j = \frac{q_j + ip_j}{2}$. For $s = 1$, we have the *P Function* $P(x) = W(x; 1)$ that is defined as the measure of the integral expansion of a density operator ρ in terms of the coherent states as

$$\rho = \int_{\mathbb{R}} \int_{\mathbb{R}} P(x) |\alpha\rangle\langle\alpha|\, dq\, dp$$

again for $|\alpha\rangle = \bigotimes_j |\alpha_j\rangle$ and $\alpha_j = \frac{q_j + ip_j}{2}$. ◀

[4] This orthogonality is induced by the inner product $\mathbf{Tr}(D(x)D(-y)) = (2\pi)^m \delta^{2m}(x - y)$, which can be shown by applying the relation in [**Proposition 2.2.15**] to the displacement operator itself as mentioned in [Ser17, Chapter 4.2].

The three instances of the s parametrized QPD that were mentioned above for $s \in \{-1, 0, 1\}$, each corresponds to the ordering of bosonic operators for the displacement operator. Furthermore, each QPD behaves quite differently. It can be thought that the lower the value of parameter s is, the better behaved the QPD is. A notable difference in behavior is that the Q function always non-negative for all $x \in \mathbb{R}^{2m}$, while the Wigner function and P function may have negative values for some $x \in \mathbb{R}^{2m}$. However in addition to having negative values, the P function may also become singular for some $x \in \mathbb{R}^{2m}$. Among three of them, the Wigner function is the most relevant in making way for the discussions on Gaussian states and also in our discussions later on BS and GBS. Therefore, we will now state some properties of the Wigner function.

Proposition 2.2.17 ([VR89])**.** *For a quantum system in state ρ, we can write the probability of measuring the position quadrature of ρ after a phase shift by θ as*

$$\mathbb{P}(q, \theta) = \int_{\mathbb{R}} W_\rho(q\cos\theta - p\sin\theta, q\sin\theta + p\cos\theta)\, dp.$$

The proposition above will prove to be central in the method of Homodyne Tomography and in the construction of the Pattern Function that we will discuss later in [**Chapter 6**]. But for now recalling that $\mathbb{P}(q, \theta) = \langle q|U(\theta)\rho U^\dagger(\theta)|q\rangle$, it is intuitive to understand this fact as stating that marginalizing the momentum quadrature p from the Wigner function $W_\rho(q\cos\theta - p\sin\theta, q\sin\theta + p\cos\theta)$ will give us the probability of obtaining q from a phase-shifted ρ. Lastly, we now will state another improtant property of the Wigner function which is highly relevant in the measurement of a CV system.

Proposition 2.2.18 (Overlap Property [CG69], [BL07, Chapter 3.3.1], [Leo97, eq.(3.22)])**.** *For a single mode system with state ρ and an operator F, we have the overlap formula*

$$\mathbf{Tr}(\rho F) = 2\pi \int\int W_\rho^{(s)}(q, p) W_F^{(-s)}(q, p)\, dq\, dp$$

where for the specific case of $s = 0$, we have in terms of Wigner functions

$$\mathbf{Tr}(\rho F) = 2\pi \int\int W_\rho(q, p) W_F(q, p)\, dq\, dp.$$

Gaussian States

We are now ready to define the Gaussian states, an important class of states of a quantum system that will be the main subject of discussion in the GBS model as it performs photon-counting measurement on a Gaussian state. They can be completely described by its first two statistical moments, namely its mean and variance as we will see in its definition below.

Definition 2.2.19 (Gaussian states)**.** For an m mode state of a quantum system with density matrix representation ρ, we say that ρ is a *Gaussian state* if its corresponding Wigner function can be written in a Gaussian form, namely

$$W_\rho(x) = \frac{1}{(2\pi)^m \sqrt{\mathbf{Det}(\sigma)}} \mathbf{exp}\left(-\frac{1}{2}(x - \bar{x})^\top \sigma^{-1}(x - \bar{x})\right)$$

where $\bar{x} = (\langle\widehat{q}_1\rangle, \langle\widehat{p}_1\rangle, \ldots, \langle\widehat{q}_m\rangle, \langle\widehat{p}_m\rangle)$ is the mean of each quadrature and σ is the covariance matrix of ρ where its i-th row and j-th column entry is $\sigma_{i,j} = \frac{1}{2}\langle[\widehat{x}_j, \widehat{x}_k]_+\rangle - \langle\widehat{x}_j\rangle\langle\widehat{x}_k\rangle$. ◀

Given our definition on the variance of a measurement of a quantum state in [**Definition 2.1.9**], we will now discuss briefly about its relationship between to the covariance matrix σ. Recalling our definition of Variance $\mathbf{Var}(\widehat{x}) = \langle\widehat{x}^2\rangle - \langle\widehat{x}\rangle^2$, which equals to uncertainty $\mathbf{Sd}^2(\widehat{x})$, we can rewrite it as

$$\mathbf{Var}(\widehat{x}) = \langle\widehat{x}^2\rangle - 2\langle\widehat{x}\rangle^2 + \langle\widehat{x}\rangle^2 = \langle\widehat{x}^2 - 2\langle\widehat{x}\rangle\widehat{x} + \langle\widehat{x}\rangle^2 I\rangle = \langle(\widehat{x} - \langle\widehat{x}\rangle I)^2\rangle.$$

Therefore, we can write variance between quadrature operators $\widehat{x}_j$ and $\widehat{x}_k$ as

$$\begin{aligned}\sigma_{j,k} &= \frac{1}{2}\langle[\Delta(\widehat{x}_j), \Delta(\widehat{x}_k)]_+\rangle \\ &= \frac{1}{2}\langle[\widehat{x}_j - \langle\widehat{x}_j\rangle I, \widehat{x}_k - \langle\widehat{x}_k\rangle I]_+\rangle\end{aligned}$$

where its diagonal elements $\sigma_{k,k}$ can be expressed as the variance of $\widehat{x}_k$

$$\sigma_{k,k} = \frac{1}{2}\langle 2\widehat{x}_k^2 - 4\widehat{x}_k\langle\widehat{x}_k\rangle + 2\langle\widehat{x}_k\rangle^2 I\rangle = \langle\widehat{x}_k^2\rangle - \langle\widehat{x}_k\rangle^2 = \mathbf{Var}(\widehat{x}_k)$$

using the linearity of $\mathbf{Tr}$ and because $\langle I\rangle = \mathbf{Tr}(\rho) = 1$, which is consistent with the definition of variance in [**Definition 2.1.9**]. Moreover, we can write $\sigma_{j,k}$ in a simpler expression

$$\begin{aligned}\sigma_{j,k} &= \frac{1}{2}\langle\widehat{x}_j\widehat{x}_k + \widehat{x}_k\widehat{x}_j - 2\widehat{x}_k\langle\widehat{x}_j\rangle - 2\widehat{x}_j\langle\widehat{x}_k\rangle + 2\langle\widehat{x}_j\rangle\langle\widehat{x}_k\rangle I\rangle \\ &= \frac{1}{2}\langle[\widehat{x}_j, \widehat{x}_k]_+\rangle - \langle\widehat{x}_k\langle\widehat{x}_j\rangle + \widehat{x}_j\langle\widehat{x}_k\rangle - \langle\widehat{x}_j\rangle\langle\widehat{x}_k\rangle I\rangle \\ &= \frac{1}{2}\langle[\widehat{x}_j, \widehat{x}_k]_+\rangle - \langle\widehat{x}_j\rangle\langle\widehat{x}_k\rangle.\end{aligned}$$

which is precisely what was stated in [**Definition 2.2.19**].

We would like to also note that the covariance Matrix σ must satisfy the Heisenberg uncertainty relation

$$\sigma + i\Omega \geq 0$$

as described in [BL07, Chapter 3.3]and [Wee+12, Section II.A.1]. The Heisenberg uncertainty relation is a fundamental consequence of quantum mechanics that intuitively states that simultaneous measurement of position and momentum cannot be done accurately.

The class of Gaussian state includes the states that we discussed about in the previous sections, the vacuum state, coherent state, squeezed state, and thermal state. The samples taken from the Gaussian probability distribution with mean vector μ and covariance matrix σ that are induced by the Gaussian states can provide an intuitive visual explanation on how different Gaussian states behaves, as in [**Figure 2.1**]. All of the Gaussian states in the figure can be thought of a vacuum state undergoing a certain transformation. Whereas the vacuum state can be described solely by the covariance matrix that coincides with the 2×2 identity matrix $\sigma_v = I_2$, the thermal state with expected photon number $\bar{n}$ can be thought

of as a vacuum state with additional variance of $2\bar{n}$ imposed on each quadrature, producing convariance matrix $\sigma_{\text{Th}} = \sigma_v + (2\bar{n})I_2$.

Also, a squeezed state can be thought of as having the covariance matrix of a vacuum state with a squeezing operation $S(r)$ applied to it (see [**Definition 2.2.6**]) where the variance of one quadrature is decreased while the other increased, resulting in $\sigma_r = S(r)\sigma_v S^\dagger(r)$. This variance due to the Heisenberg uncertainty relation which imposes a lower bound restriction on the variance of both quadratures. Further, a phase shift can be seen as "rotating" the Gaussian state representation in [**Figure 2.1**]. This is clear from observing the samples taken from the probabilty distribution of a phase shifted squeezed state where it is shifted by the amount of phase shift θ, producing a state with covariance matrix $\sigma_\theta = \bar{U}(\theta)\sigma_r\bar{U}^\dagger(\theta)$.

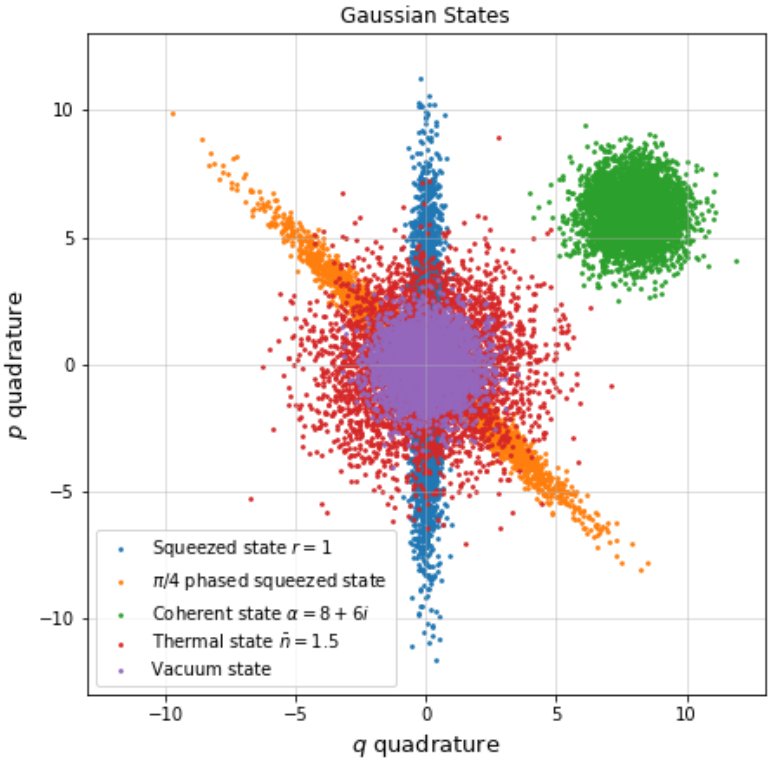

Figure 2.1: Scatter plot of 4000 samples from each of the Gaussian distribution induced by various Gaussian states mentioned in [**Section 2.2.3**] for a single-mode system.

Lastly as discussed in [Ser17], a Gaussian state in an m mode system can also be equivalently defined using its density operator representation in Hilbert space $\mathcal{S}$ as

$$\rho = \frac{\mathbf{exp}(-\beta H)}{\mathbf{Tr}(\mathbf{exp}(-\beta H))}$$

where $\beta > 0$ and $H = \frac{1}{2}\widehat{x}^T W \widehat{x} + x^\top \widehat{x}$ for symmetric and positive-definite[5] matrix $W \in \mathbb{R}^{2m,2m}$, real numbers $x \in \mathbb{R}^{2m}$, and $\widehat{x} = (\widehat{q}_1, \widehat{p}_1, \ldots, \widehat{q}_m, \widehat{p}_m)$. It is from here that the "displacement"

[5]Meaning that, for all $z \in \mathbb{R}^{2m}$, we have $z^\top W z > 0$.

to the values measured from the observables that was stated in [**Proposition 2.2.4**] can be easily seen, where displacement $D(q,p)$ operating on $(\widehat{q},\widehat{p})$ adds the mean of the measured value from them by (q,p). This can clearly be seen from [**Figure 2.1**] where a coherent state $|\alpha\rangle$ – which is a displaced vacuum state, induces a probability distribution with the same variance as the vacuum, but a mean that is shifted by $\mathbf{Re}(\alpha)$ on the q quadrature and by $\mathbf{Im}(\alpha)$ on the p quadrature. Thus, for a Gaussian state ρ with mean $\bar{x}$, the mean of its displacement $D^\dagger(x)\rho D(x)$ is $\bar{x}+x$. This is noted in the discussion in [Ser17, Chapter 3.3] on the relationship between the density operator representation of a Gaussian state with its mean vector and covariance matrices by taking its first and second statistical moment.

2.2.5 Continuous-Variable Measurements

In this last section of the preliminary chapter on CV quantum information, we will cover two most prominently used CV measurements scheme, the Homodyne measurement and the Heterodyne measurement. Both will be an important concept in our discussions later on different quantum computation models.

Homodyne Measurement

Restating what was discussed in [**Section 2.2.3**], we have the probability of obtaining measurement outcome $z \in \mathbb{R}$ from observable $\widehat{z}_\theta$ as

$$\mathbb{P}(q,\theta) = \langle q_\theta|\rho|q_\theta\rangle,$$

which is the probability of obtaining measurement outcome of q from θ phase shifted state ρ. This measurement can obtained by a measurement scheme on a CV quantum system, which we call the *Homodyne measurement* which method will be defined and discussed below.

Definition 2.2.20 (Homodyne Measurement)**.** For a single mode system in state ρ, a Homodyne measurement on the system with local oscillator phase θ will result on a probability of obtaining a measurement result q from observable $\widehat{q}_\theta$ as $\mathbb{P}(q,\theta) = \langle q_\theta|\rho|q_\theta\rangle$ where $|q_\theta\rangle = U^\dagger(\theta)|q\rangle$. ◀

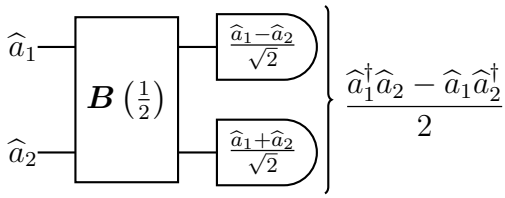

Figure 2.2: Homodyne Measurement illustration – Homodyne outcome observable $\frac{\widehat{a}_1^\dagger\widehat{a}_2-\widehat{a}_1\widehat{a}_2^\dagger}{2}$ is obtained by subtracting photon number observable of the second mode $\frac{1}{2}(\widehat{a}_1+\widehat{a}_2)^\dagger(\widehat{a}_1+\widehat{a}_2)$ by the photon number observable of the first mode $\frac{1}{2}(\widehat{a}_1-\widehat{a}_2)^\dagger(\widehat{a}_1-\widehat{a}_2)$ after the balanced beamsplitter.

In an experimental setup, Homodyne measurement is obtained by interfering the output mode with a *Local Oscillator* with large amplitude using a *balanced beamsplitter*, and then

subtracting the photodetection measurement results between the two output modes of the beamsplitter. Consistent with definition of a beamsplitter in [**Section 2.2.3**], we describe the balanced beamsplitter as

$$\boldsymbol{B}(1/2) = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

thus producing the annihilation operators of the output modes as a mixture of the annihilation operators of both modes before the beamsplitter

$$\boldsymbol{B}(1/2) \begin{bmatrix} \widehat{a}_1 \\ \widehat{a}_2 \end{bmatrix} = \begin{bmatrix} (\widehat{a}_1 - \widehat{a}_2)/\sqrt{2} \\ (\widehat{a}_1 + \widehat{a}_2)/\sqrt{2} \end{bmatrix}.$$

Considering the photon number observable for each mode

$$\widehat{n}_1 = \frac{1}{2}(\widehat{a}_1 - \widehat{a}_2)^\dagger(\widehat{a}_1 - \widehat{a}_2)$$

and

$$\widehat{n}_2 = \frac{1}{2}(\widehat{a}_1 + \widehat{a}_2)^\dagger(\widehat{a}_1 + \widehat{a}_2)$$

we obtain

$$\widehat{h} = \widehat{n}_2 - \widehat{n}_1 = \widehat{a}_1^\dagger \widehat{a}_2 + \widehat{a}_1 \widehat{a}_2^\dagger$$

as can be seen in the Homodyne measurement scheme representation in [**Figure 2.2**]. Using the physical interpretation of the local oscillator being in a coherent state $|\alpha\rangle$ where $|\alpha|$ is large, we can then rewrite

$$\widehat{h} = \alpha \widehat{a}_1^\dagger + \alpha^* \widehat{a}_1,$$

then further rewriting $\alpha \in \mathbb{C}$ in its polar form $\alpha = |\alpha| e^{-i\theta}$ with $\theta \in [0, 2\pi]$ being the phase of the local oscillator, we have

$$\widehat{h} = |\alpha|(e^{-i\theta}\widehat{a}_1^\dagger + e^{i\theta}\widehat{a}_1) = |\alpha|\widehat{q}_\theta$$

where we have used the relation between the phase-shifted quadrature operator $\widehat{q}$ and the bosonic operators in (2.2.12).

A more rigorous treatment of the derivation on the relation between $\widehat{h}$ and $\widehat{q}_\theta$ can be done by showing that for a two-mode system in a state $\rho \otimes |\alpha\rangle_{22}\langle\alpha|$, the expectation of the observable of one photodetector subtracted with the other coincides with the expectation of $\widehat{q}_\theta$ in the limit of large local oscillator amplitude, namely

$$\mathbf{Tr}_1(\rho \otimes |\alpha\rangle_{22}\langle\alpha|(\widehat{a}_1^\dagger \widehat{a}_2 + \widehat{a}_1 \widehat{a}_2^\dagger)) = \mathbf{Tr}_1(\rho\ {}_2\langle\alpha|\widehat{a}_1^\dagger \widehat{a}_2 + \widehat{a}_1 \widehat{a}_2^\dagger|\alpha\rangle_2) = \mathbf{Tr}_1(\rho \widehat{q}_\theta).$$

This can be achieved by considering a moment generating function

$$h(x) = \mathbf{Tr}_1\left(\rho \otimes |\alpha\rangle_{22}\langle\alpha| \mathbf{exp}\left(x \frac{\widehat{a}_1^\dagger \widehat{a}_2 + \widehat{a}_1 \widehat{a}_2^\dagger}{|\alpha|}\right)\right)$$

and showing that the derivative of $h(x)$ when $x = 0$ in the limit $|\alpha| \to \infty$ is precisely the expectation of $\widehat{q}_\theta$. In analytical notations,

$$\frac{d}{dx}\left(\lim_{|\alpha|\to\infty} h(x)\right)\Big|_{x=0} = \mathbf{Tr}_1(\rho\widehat{q}_\theta).$$

The technical details of the proof is omitted here, although interested readers may refer to [Ser17, Chapter 5.4.1].

Heterodyne Measurement

As another important CV measurement method, we will briefly discuss the Heterodyne measurement, which will have an essential part in our later discussion on a computation model we call the CV-sampling in [**Section 7.3**]. Heterodyne measurement can be thought of as Homodyne measurements on each of the two outputs from a beamsplitter with transmissivity τ that interferes the mode subject to being measured with a vacuum, as illustrated in [**Figure 2.3**] which we will base our discussion upon. Here however, we will only discuss the physical interpretation of the Heterodyne measurement setup. A derivation and more detailed discussion of the Heterodyne measurement can be found in [Leo97, Chapter 6] and [Ser17, Chapter 5.4.2]. Now, we define its outcome probability.

Definition 2.2.21 (Heterodyne Measurement)**.** For a system in state ρ, Heterodyne measurement with interfering beamsplitter with transmissivity $\tau \in [0, 1]$ will result in a measurement outcome $\alpha = \frac{1}{2}(q + ip)$ with probability

$$\mathbb{P}(\alpha) = \frac{1}{\pi}\mathbf{Tr}(\rho|\alpha, r\rangle\langle\alpha, r|) \tag{2.2.14}$$

where $|\alpha, r\rangle = \boldsymbol{D}(\alpha)\boldsymbol{S}(r)|0\rangle$ with $r = \ln\frac{\sqrt{\tau}}{\sqrt{1-\tau}}$. Quadrature measurement outcomes q and p are the eigenvalues of the observables

$$\widehat{q} = \sqrt{\tau}\widehat{q}_1 - \sqrt{1-\tau}\widehat{q}_0 \quad \text{and} \quad \widehat{p} = \sqrt{1-\tau}\widehat{p}_1 + \sqrt{\tau}\widehat{p}_0,$$

respectively, where $\widehat{q}_1$ and $\widehat{p}_1$ are the quadrature operators of the system and $\widehat{q}_0$ and $\widehat{p}_0$ are the quadrature operators of the vacuum input. ◀

All Homodyne measurements use a single local oscillator which comes from the balanced beamsplitter acting on $\widehat{a}_v$ and $\widehat{a}_2$ in [**Figure 2.3**] that interferes a strong local oscillator $\widehat{a}_2$ with a vacuum $\widehat{a}_v$. The outcome from each Homodyne measurement is therefore represented by observables $\sqrt{\tau}\widehat{q}_1 - \sqrt{1-\tau}\widehat{q}_0$ and $\sqrt{1-\tau}\widehat{p}_1 + \sqrt{\tau}\widehat{p}_0$, which essentialy means that due to the interference with the vacuum mode $\widehat{a}_0$, the measurement of the quadratures corresponding to $\widehat{a}_1$ is corrupted by noise from the vacuum, represented by the $\sqrt{1-\tau}\widehat{q}_0$ and $\sqrt{\tau}\widehat{p}_0$ terms. The squeezing parameter $r = \ln\frac{\sqrt{\tau}}{\sqrt{1-\tau}}$ of the displaced squeezed state $|\alpha, r\rangle$ that the state is projected to in [**Definition 2.2.21**] can therefore be interpreted as an indication of which quadrature the noise is more allocated to as a consequence of Heisenberg uncertainty relation. As τ gets closer to 1, measurement of quadrature q will be more accurate, as measurement of p will be more noisy, while having τ closer to 0 will result in the opposite.

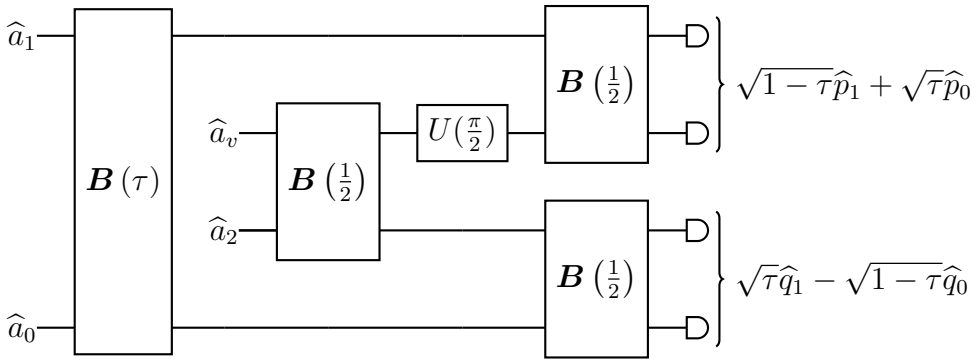

Figure 2.3: Heterodyne Measurement illustration – with the mode subject to the measurement represented with $\widehat{a}_1$, beamsplitter $\boldsymbol{B}(\tau)$ interferes $\widehat{a}_1$ with vacuum mode with observable $\widehat{a}_0$. Homodyne measurement outcome of the two output modes from $\boldsymbol{B}(\tau)$ corresponds to observables $\sqrt{\tau}\widehat{q}_1 - \sqrt{1-\tau}\widehat{q}_0$ and $\sqrt{1-\tau}\widehat{p}_1 + \sqrt{\tau}\widehat{p}_0$. Beamsplitter $\boldsymbol{B}(\frac{1}{2})$ operating on $\widehat{a}_v$ and $\widehat{a}_2$ outputs local oscillator for the Homodyne measurements where the phase shift $U(\frac{\pi}{2})$ is operating on one of them to provide a Homodyne measurement outcomes of different quadratures (i.e: $\widehat{q}_{\theta-\frac{\pi}{2}} = \widehat{p}_\theta$).

Chapter 3

Computational Complexity Theory Preliminaries

Whether a problem can be efficiently computed by a certain computational model is a question relevant in determining the complexity of a problem that is characterized by computation. While this is a question that is interesting in its own right from a computational perspective, this is central in understanding the power of a computational model, particularly to Quantum Computation, being the subject area of this thesis. As mentioned in the introduction section, *Computational Complexity Theory* (may be referred to later simply as "complexity theory") has been used as a rigorous mathematical tool to characterize the complexity of a problem using polynomial computability in the size of input. The notion of polynomial computability is therefore used as an indication whether a problem can be efficiently solved by a certain computational model. Thus, it can be thought that one of the essence of complexity theory is that all problems solvable in polynomial time by a machine that belongs to a computational model constitutes a class of problem. As a consequence, a significant part of the study of computational complexity theory focuses on how one class of problem relate to another. The analysis of this relationship can be done very conveniently in complexity theory using a concept from computability theory using the notion of Turing Machines that relativizes computation to an *Oracle* machine, which is an imaginary machine that can provide a yes/no answer to a certain question[1]. This interesting concept of computation relative to an oracle will be discussed in more detail and defined more precisely later in this chapter.

This chapter is dedicated to computational complexity theory concepts relevant to arguments formulated for the complexity for Boson Sampling (BS) and Gaussian Boson Sampling (GBS) discussed in later chapters. Readers who are familiar with Computational Complexity Theory may skip going through this whole chapter and refer to [**Table 3.1**] which lists complexity classes relevant to BS and GBS along with their brief definitions, also links to more detailed definitions throughout this chapter. An illustration of important relationships between the complexity classes is also shown in [**Figure 3.1**]. The table and figure may serve to all readers as references to the definitions of complexity classes used for the rest of

[1]The concept of an oracle machine is a very interesting topic relevant to both computational complexity and foundations of mathematics. It can be traced back to Alan Turing's doctoral thesis on ordinal logic [Tur39].

this thesis.

We will first define Turing Machines in [**Section 3.1**], which defines the notion of classical computation and then in [**Section 3.2**] define Oracle Turing Machines to define computation relative to a class problem, which is central to the polynomial hierarchy complexity class. Later in [**Section 3.3**], we will discuss counting problems and #**P** complexity class which will be of primary interest in our complexity argument of BS and GBS. Then, in [**Section 3.4**] we will discuss probabilistic classical computation, which made up a complexity class of problems that can be efficiently solved by classical computers. We will also discuss its relationship with quantum computation, where we will also briefly touch on postselection, an interesting cocnept that increases the computational of a quantum computer. Lastly in [**Section 3.5**], we will define different types of problems, namely search and sampling problems, which BS and GBS are both formulated as. Formal definitions for complexity classes in this chapter closely follow those in [AB09] and [KSV02].

Computational Complexity Classes

A list of brief definitions of complexity classes that are discussed in this chapter is shown in [**Table 3.1**], where links to more precise definitions are also present. Relationships between Computational Complexity Classes are shown in [**Figure 3.1**] below along with a reference to their formal definition in this chapter. For conciseness, we will refer to classical Turing Machine as `TM`, non-deterministic Turing Machine as `nTM`, probabilistic Turing Machine as `pTM`, and quantum Turing Machine as `qTM` throughout this chapter.

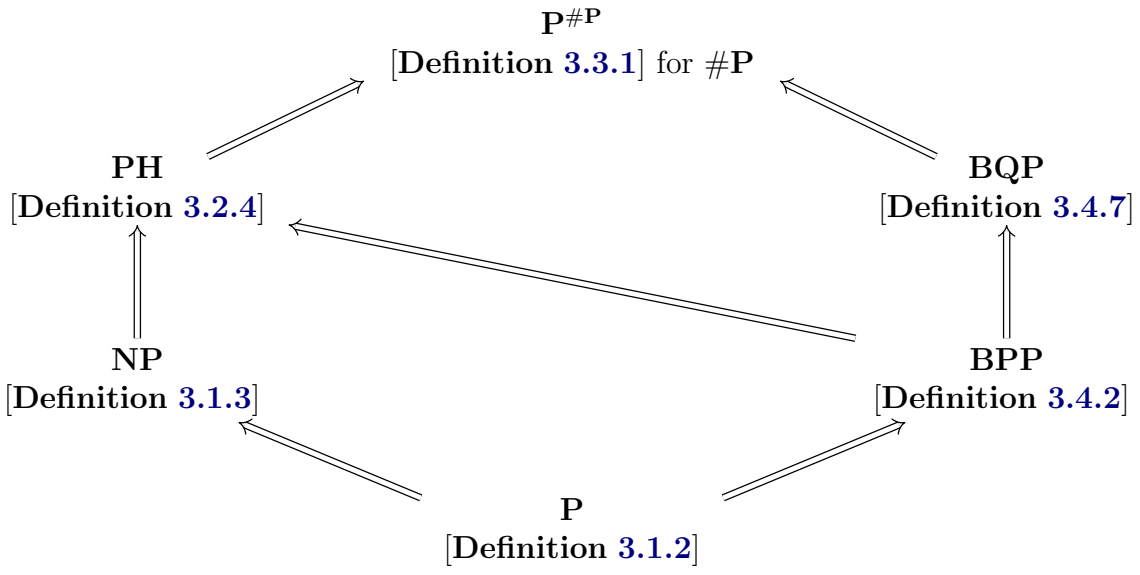

Figure 3.1: Relationship between Complexity Classes. For classes **A** and **B**, we denote $\mathbf{A} \Rightarrow \mathbf{B}$ if **B** contains **A** (i.e: $\forall A \in \mathbf{A}$, we have $A \in \mathbf{B}$).

Complexity Class	Contains
P [**Definition 3.1.2**]	Sets that are decidable in polynomial time by a `TM`.
NP [**Definition 3.1.3**]	Sets which there exist a polynomial-time `TM` that decides each of its element x given some second input string $y \in \mathbb{B}^{\text{poly}(\lvert x \rvert)}$. $\Leftrightarrow$ Sets decidable in polynomial time by a `nTM`.
BPP [**Definition 3.4.2**]	Sets that are decidable by a polynomial-time `pTM` with bounded error.
BQP [**Definition 3.4.7**]	Sets that are decidable by a polynomial-time `qTM` with bounded error.
#P [**Definition 3.3.1**]	Natural number valued functions that for any string x counts the number of strings that can serve as a second input for a `TM` M such that M accepts x. $\Leftrightarrow$ Natural number valued functions that outputs the number of accepting paths of an `nTM` for any string x.
FP [**Definition 3.3.3**]	Functions that are computable by a polynomial `TM`.
Σ_j^p [**Proposition 3.2.8**]	Sets that are decidable by are decidable by an `nTM` equipped with $j-1$ nested **NP**-complete oracles (e.g: $\Sigma_4^p = \mathbf{NP}^{\mathbf{NP}^{\mathbf{NP}^{\mathbf{NP}}}}$).
PH [**Definition 3.2.4**]	Union of all classes in polynomial hierarchy (i.e: $\mathbf{PH} = \bigcup_{n\in\mathbb{N}} \Sigma_n^p$).

Table 3.1: Brief definitions of complexity classes.

3.1 Turing Machine and Decision Problem

In defining the notion of computation in a more general sense to make way for a more specific notion of polynomial computability that is central in characterizing complexity of a problem, we will resort to the Turing Machine as our computational model of choice. We define it as follows.

Definition 3.1.1 (Turing Machine `TM`)**.** A *Turing Machine* `TM` M is a machine with a bi-directional infinite tape with a read/write head that is defined by the tuple (Q, Γ, δ) where

- $Q = \{q_1, \ldots, q_k\}$ is a set of k *states* that M can be in. There are *initial state* and *final state* $q_s, q_f \in Q$ where M always begins in state $q_s \in Q$ and halts if it is in state q_f.
- A finite set $\Gamma = \{0, 1, \Lambda\}$ of all possible symbols on the tape we call an *alphabet* where Λ is an *empty symbol*.
- $\delta : Q \times \Gamma \to Q \times \Gamma \times \{L, R, S\}$ is a *transition function* that takes the symbol $\sigma \in \Gamma$ on the tape that the read/write head is currently on and the current state $q \in Q$, then write a symbol on the tape where the head is at, move to a certain direction the tape, and make a state transition.

All but finitely many symbols on the tape are empty symbols Λ. When M halts on *input* x, we write the symbols left on the tape of M as $M(x)$ and call it the *output* of M on input x. The output $M(x)$ consists of all symbols from the leftmost non-Λ symbol on the tape to the rightmost non-Λ. We may further specify the number of steps $k \in \mathbb{N}$ that M took before outputting $M(x)$ by writing $M(x)[k]$. ◀

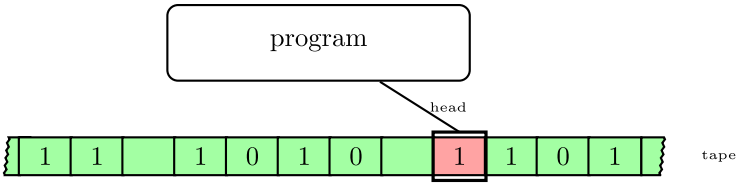

Figure 3.2: Turing Machine illustration where its "Program" is comprised of a description the transition δ between its states Q.

An illustration of the Turing machine can be seen in [**Figure 3.2**] where a "program" which is comprised of a description of the transition function δ between states Q. Using the set of rules according to δ and states Q, the program dictates how the read-write head of the `TM` moves around and changes the symbols on the tape given a binary string input. The set of rules that governs the transition function δ can be described by a transition diagram or a table showing which symbol $\{0, 1, \Lambda\}$ does the machine head write on the tape position its currently on and which direction $\{L, R, S\}$ does it go, given that the machine is in a particular state.

An example table description of a `TM` can be found in [**Table 3.2**]. It shows the transition rules for a `TM` that outputs a 1 if the length of input string is odd and 0 otherwise. The first column indicating the state that the `TM` is currently in, and the other columns with $x \in \{0, 1, \Lambda\}$ at the top indicates the state the `TM` transition to, the symbol it writes on the tape position the head is currently on, and the direction it moves after writing that symbol. This is described by a 3 tuple $Q \times \Gamma \times \{L, R, S\}$. As we assume that the read-write head is initially somewhere on the input string on the tape and expect for the input string, all symbols are Λ, the state q_s moves the head all the way to the left-most non Λ symbol on the tape then move to state q_1 after encountering the first Λ. State q_1 indicates a state where the machine has seen even number of symbols so far, if it sees another symbol, it erases it by replacing it with Λ and move to the right, then transitioning to q_2. Thus, q_2 indicates that the machine has seen an odd number of symbols. When the machine is in either q_1 or q_2 and sees a Λ, it then transitions to q_f, changes the symbol of the position its head is currently on to 0 if it is in state q_1 or 1 if it is in state q_2 and halts.

We also note that it is possible to have empty symbols between binary strings on the tape for the output of the `TM`. However for simplicity, here we will assume that all output consists only of a single binary string, i.e: there is only one single block of binary string on the output tape when the machine halts.

Note that it is possible that for a given input x, TM M does not halt; that is, M keeps operating on its tape forever. Determining whether a `TM` M halts on an arbitrary input, which is referred to as the *Halting Problem* is an incomputable problem[2]. Although the

[2]An "Incomputable" problem means that there is no Turing Machine in existence that can solve in a

State	0	1	Λ
q_s	$q_s, 0, L$	$q_s, 1, L$	q_1, Λ, R
q_1	q_2, Λ, R	q_s, Λ, R	$q_f, 0, S$
q_2	q_1, Λ, R	q_1, Λ, R	$q_f, 1, S$

Table 3.2: Turing Machine Transition Table that describes a TM that outputs a 1 if the input string has an odd length and 0 if the input string has an even length. The first column indicating the state that the TM is currently in, and the other columns with $x \in \{0, 1, \Lambda\}$ at the top indicates the state the TM transition to, the symbol it writes on the tape position the head is currently on, and the direction it moves after writing that symbol. We assume that the read-write head is initially somewhere on the input string on the tape, where all symbols on the tape are Λ except for the input string.

halting problem introduces a notion of incomputability, which is closely related to Gödel's Incompleteness theorem and foundation of mathematics, it has no significant implication to our later discussions on BS and GBS. Readers who are interested may refer to [AB09, Chapter 1.5.1].

With a precise definition of computation given by the Turing Machine formalism, we are now ready to define our first type of problems that are simply in the form of a question of whether a binary string x is in a set $A \subseteq \mathbb{B}^*$, called a *decision problem*. Now applying the notion of computation using the Turing Machine to a decision problem and imposing a constraint on the number of steps it takes to outputs an answer using a polynomial in the length of binary string input x, we define the most basic class in complexity theory consists of decision problems solvable in polynomial time[3].

Definition 3.1.2 (Polynomial-time Turing Machine, **P** class, and Decidability)**.** Let M denote a TM that for any input $x \in \mathbb{B}^*$ halts in polynomial time in the length of x. We call this TM a *polynomial-time Turing Machine* (or simply *polynomial Turing Machine*). A set $A \subseteq \mathbb{B}^*$ is in class **P** iff there exist a polynomial TM M such that A is of the form

$$A = \{x \in \mathbb{B}^* : \exists c \geq 1, M(x)[\mathcal{O}(|x|^c)] = 1\}.$$

Equivalently in words, x is in A if and only if there exists a polynomial TM M such that for all input strings $x \in A$, there exists $c \geq 1$ independent of x such that M halts in $\mathcal{O}(|x|^c)$ steps with symbol 1 on its tape. For such set A, we say that A is *decidable* by M or M *decides* A. ◀

For a set $A \subseteq \mathbb{B}^*$, we will refer to the task of deciding whether $x \in \mathbb{B}^*$ is in A as a *decision problem*. If we further generalize our definition of Turing Machine for it to take in two binary string inputs that are distinguishable (by a certain encoding), we can define another interesting and important class of decision problems called the **NP** class. This class can be intuitively understood as the class of problems that can be decided in polynomial

finite amount of computational steps.

[3]To provide a less verbose discussion, we will often refer to a computation that takes a polynomial time in input size/length, such as poly (n) when we have a binary string $x \in \mathbb{B}^n$ as an input to machine M, simply as "polynomial time".

time on whether a certain binary string x is in a set $A \subseteq \mathbb{B}^*$ given an additional binary string information y as its second input. This is defined more precisely below.

Definition 3.1.3 (**NP** class)**.** A set $A \subseteq \mathbb{B}^*$ is in class **NP** iff there exists a polynomial $p : \mathbb{N} \to \mathbb{N}$ and polynomial-time `TM` M such that

$$A = \{x \in \mathbb{B}^* : \exists y \in \mathbb{B}^{p(|x|)}, M(x, y) = 1\}.$$

Or in words, A is a set of strings x that can be decided by a polynomial time `TM` M if there is another string y with length at most polynomial function of x that acts as second input to M. ◀

The string y can be seen as the "answer" to a decision problem, while `TM` M "checks" the correctness of the answer in polynomial time. Clearly $\mathbf{P} \subseteq \mathbf{NP}$, as if $|y| = 0$, then whether x is in set A can be solely decided by a `TM` in polynomial time (i.e: $M(x) = 1$ as in [**Definition 3.1.2**]). However, whether if $\mathbf{P} \subsetneq \mathbf{NP}$ is not known, despite being widely believed to be true. The strict containment of **P** in **NP**, can be understood intuitively as that there is a problem which solution can be efficiently machine-checked, despite it being impossible to efficiently compute the solution itself. We also note that it is possible to encode multiple inputs onto the `TM` such that the `TM` can distinguish these inputs in polynomial time. Now to further the discussion on the **NP** class, we define the *Non-deterministic Turing Machine*.

Definition 3.1.4 (Non-deterministic Turing Machine `nTM`)**.** A Non-deterministic Turing Machine `nTM` is defined by $(Q, \Gamma, \delta_0, \delta_1)$ similar to `TM` but with the transition at each step performed simultaneously with δ_0 or δ_1. A *Non-deterministic Computation Path* (or simply *Computation Path* or *Path* where context is clear) made by `nTM` Φ is a string $c \in \mathbb{B}^*$ where at the i-th step, Φ performs the transition δ_{c_i}. If there exists a path $c \in \mathbb{B}^*$ of Φ such that it halts in state q_f given an input x, then we say that $\Phi(x)$ is the output from the computation path c, or that Φ outputs $\Phi(x)$ from path c. ◀

Intuitively, the `nTM` have "multiple programs" that runs at the same time, each indicated by the computation path transition functions δ_0 and δ_1 where each of them does operation on the tape differently with different computation time. This will result in the `nTM` having two transition tables as shown in [**Table 3.2**]. In a polynomial-time `nTM`, Φ with input x, there can be at most $2^{p(|x|)}$ many computation paths for some polynomial p and each of them may have different output $\Phi(x)$. For example, for a computational path 1010 (performing transition using δ_1, δ_0, δ_1, then δ_0), Φ halts and outputs $\Phi(x) = y$, however for path 1011, Φ still does not halt, and therefore the computation will continue and "branches" using simultaneously both δ_0 and δ_1.

From this definition of `nTM` we can formulate a definition of the **NP** class in terms of `nTM`. The class **NP** can therefore be intuitively understood as a class of decision problems that can be decided by an `nTM` in a polynomial time. Here, we define a polynomial-time Non-deterministic Turing Machine as an `nTM` that for all input $x \in \mathbb{B}^*$, every computational path c given input x is of length $p(|x|)$ for some polynomial p. A set A is decidable by a polynomial `nTM` Φ if for all $x \in A$, there exist a computational path $c \in \mathbb{B}^{p(|x|)}$ with output $\Phi(x) = 1$.

Theorem 3.1.5 (nTM characterization of **NP** Class)**.** *Set A is in* **NP** *iff there exist a polynomial* nTM Φ *that decides A.*

Proof. ($\Leftarrow$) Let Φ be an nTM that decides A, then there exists a polynomial p where for all input $x \in A$, Φ halts in $p(|x|)$ steps and outputs $\Phi(x) = 1$. Thus, for $x \in A$ there exists a path $y \in \mathbb{B}^{p(|x|)}$ that outputs $\Phi(x) = 1$. Therefore, there exist a TM M that can simulate both transition functions δ_0 and δ_1 of $\Phi(x)$ given path y (for example, by using y to determine which transition to use at a given time step) giving $M(x, y) = 1$ iff $x \in A$.

($\Rightarrow$) For a polynomial $p : \mathbb{N} \to \mathbb{N}$ and polynomial TM M, suppose $x \in A$ iff $\exists y \in \mathbb{B}^{p(|x|)}$ such that $M(x, y) = 1$. Let Φ be an nTM such that on input x it writes all string y on its output tape of length $|y| = p(|x|)$. It then runs $M(y, x)$ after writing every string y and checks if it returns 1. Then, Φ halts if $M(y, x) = 1$ and returns 1 for $|y| = p(|x|)$, or halts and returns 0 if there is no $|y| = p(|x|)$ such that $M(y, x) = 1$. □

3.2 Oracle Machines and Polynomial Hierarchy

In this section we define a variant of the Turing Machine with an access to solution of a certain class of decision problem, known as an *Oracle Turing Machine*, or simply *Oracle Machine*. In later section we will define an oracle machine for a class of different type of problem, called the Counting Problem, that is an important part of later discussions. The concept of oracles is central in defining the Polynomial Hierarchy complexity class, which is a significant component of the argument for the complexity of both Boson Sampling and Gaussian Boson Sampling.

Definition 3.2.1 (Oracle Machine [AB09, Definition 3.4])**.** An *Oracle Turing Machine* M^A is a Turing machine with an additional tape, we call the *Oracle* (or, Oracle tape) and additional states q_c and q_n. After M writes a string z to the Oracle tape, M transitions to state q_c if $z \in A$ or to q_n if $z \notin A$. We call A as the Oracle of Oracle machine M^A. If machine M^A with input x on its read-write tape halts, we write its output on the read-write tape as $M^A(x)$. ◀

It can be understood that the a machine in possession of an Oracle A has the complete knowledge of the membership of all binary strings in A, thus able to decide whether an arbitrary string z is in A. This feature of the Oracle machine enable it to serve as a tool in relativizing computation of a certain decision problem with respect to another decision problem, which are captured in the concept of reduction, and problem class hardness and completeness.

Definition 3.2.2 (Reduction, Hardness, and Completeness)**.** For sets of strings A and A', A is *reducible* to A', which is written as $A \leq_p A'$, if there exist a polynomial TM $M : \mathbb{B}^* \to \mathbb{B}^*$ such that for all $x \in \mathbb{B}^*$, $x \in A$ iff $M(x) \in A'$. For a complexity class $\mathcal{C}$, if there exist A' such that for all $A \in \mathcal{C}$ we have $A \leq_p A'$, we call A' as *$\mathcal{C}$-hard*. Furthermore, if $A' \in \mathcal{C}$, then it is *$\mathcal{C}$-complete*. ◀

We will now show an example from [AB09, Theorem 2.9] that showcases a complete problem from the complexity class **NP**, or an **NP**-complete problem.

Example 3.2.3. Consider set $S = \{(i, x, n, t) : \exists y \in \mathbb{B}^n, M_i(x, y)[\leq t] = 1\}$ of tuples consisting of $i \in \mathbb{B}^*$, an encoding of `TM` M_i, $x \in \mathbb{B}^*$ an its input, and $n, t \in \mathbb{N}$ where there exist a string y of length n such that $M_i(x)$ halts in t steps or less and outputs 1. To show that S is **NP**-complete, first, let $A \in \mathbf{NP}$. By definition of **NP** class,

$$A = \{x \in \mathbb{B}^* : \exists y \in \mathbb{B}^{p(|x|)}, M_e(x, y) = 1\}$$

for a polynomial p and a `TM` M_e running in polynomial-time q w.r.t the length of its input. For all $x \in A$, let $z_x = (e, x, p(|x|), q(m))$ where e is encoding of M_e and $m = |x|+p(|x|)+c$ is the length of input encoding of (x, y) with c as a constant from the overhead of encoding/decoding of (x, y). Now consider a `TM` T_A such that $T_A(x) = z_x$, which encodes $e, x, p(|x|), q(m)$ together in time-steps polynomial in $|z_x|$ (a task that can be done by putting a binary string in between the encodings of each number that is not a prefix to any of them). The runtime of M_e is within $q(m)$ as $|x| + p(|x|) + c$ is the time-steps that M_e takes to read and decode both inputs, that are of length $|x|$ and $p(|x|)$. Therefore, we can construct a polynomial `TM`, T_A for each $A \in \mathbf{NP}$ such that $T_A(x) \in S$, implying that for all $A \in \mathbf{NP}, A \leq_p S$, thus showing that S is **NP**-complete. ◇

With the concept of the completeness of a problem, we are now ready to state the definition of the *Polynomial Hierarchy* class **PH** that is central in the argument for the complexity of both Boson Sampling and Gaussian Boson Sampling.

Definition 3.2.4 (Polynomial Hierarchy and the Σ_i^p)**.** A set $A \subseteq \mathbb{B}^*$ is in Σ_i^p class iff there exists a polynomial time `TM` M and a polynomial q such that A is of the form

$$\left\{x \in \mathbb{B}^* : \exists y_1 \in \mathbb{B}^{q(|x|)}, \forall y_2 \in \mathbb{B}^{q(|x|)}, \ldots, Q_i y_i \in \mathbb{B}^{q(|x|)}, M(x, y_1, \ldots, y_i) = 1\right\}$$

where Q_i is the quantifier '$\exists$' if i is odd and '$\forall$' if i is even. We now define the class *Polynomial Hierarchy* as $\mathbf{PH} = \bigcup_{i\in\mathbb{N}} \Sigma_i^p$. ◀

Note that for A in Σ_0^p, it is of the form $\{x \in \mathbb{B}^* : M(x) = 1\}$; in words, A is decidable by a polynomial `TM` M, and therefore belong in class **P**. Similarly, for A in Σ_1^p, it is of the form

$$\left\{x \in \mathbb{B}^* : \exists y \in \mathbb{B}^{q(|x|)}, M(x, y) = 1\right\}$$

for a polynomial `TM` M and polynomial $q : \mathbb{N} \to \mathbb{N}$, which is precisely the definition of the **NP** class [**Definition 3.1.3**].

Definition 3.2.5. Sets that can be decided by a polynomial `TM` equipped with oracle D is in the class $\mathbf{P}^D$. That is, for polynomial-time `TM` M, we have

$$\mathbf{P}^D = \{A \subseteq \mathbb{B}^* : \exists M, \forall x \in \mathbb{B}^*, x \in A \Leftrightarrow M^D(x) = 1\}.$$

In words, the class $\mathbf{P}^D$ consists of all set A such that there exists a polynomial oracle Turing Machine M^D that decides A in polynomial time. Similarly, if we replace M with a polynomial `nTM` Φ equipped with oracle D, we have the class $\mathbf{NP}^D$. ◀

Consider a polynomial Oracle Machine equipped with a SAT oracle $M^{\mathbf{SAT}}$. Since **SAT** is in **NP**-complete class, for all D in **NP**, we have $D \leq_p \mathbf{SAT}$, and so D can be decided by $M^{\mathbf{SAT}}$ and $\mathbf{P}^{\mathbf{SAT}}$ which contains all set A in **NP**. To characterize this, we now define a class of sets that are decidable by a polynomial `TM` or `nTM` equipped with complete set of complexity classes.

Definition 3.2.6 (Complete problems)**.** For a class $\mathcal{C}$, a $\mathcal{C}$-complete oracle D, and polynomial-time `TM` M, we define the class $\mathbf{P}^{\mathcal{C}}$ as

$$A \in \mathbf{P}^{\mathcal{C}} \Leftrightarrow (\exists M^D, x \in A \Leftrightarrow M^D(x) = 1).$$

In words, $\mathbf{P}^{\mathcal{C}}$ contains sets that are decidable by a polynomial `TM` equipped with a $\mathcal{C}$*-complete* oracle. If we replace `TM` M with an `nTM` Φ, then we have $\mathbf{NP}^{\mathcal{C}}$. ◀

As **SAT** is **NP**-complete[4], the class of all sets that are decidable by polynomial `TM` and are equipped with a **SAT** oracle can be written as

$$\mathbf{P}^{\mathbf{NP}} = \left\{A \subseteq \mathbb{B}^* : \exists M^{\mathbf{SAT}}, x \in A \Leftrightarrow M^{\mathbf{SAT}}(x) = 1\right\}$$

where $M^{\mathbf{SAT}}$ denotes a **SAT** oracle machine. We must note that the oracle equipped by a polynomial `TM` M can be thought as part of the input of M in the following sense

$$\mathbf{P}^{\mathbf{NP}} = \{A \subseteq \mathbb{B}^* : \exists y_{\mathbf{SAT}}, x \in A \Leftrightarrow M(x, y_{\mathbf{SAT}}) = 1\}$$

where $y_{\mathbf{SAT}}$ is a special input on M's tape that encodes the information that allows M to decide whether z is in **SAT** in a single step. Thus, for all $x \in A \in \mathbf{P}^{\mathbf{NP}}$, M outputs $M(x, y_{\mathbf{SAT}}) = 1$ within polynomial time $p(|x|)$.

Now, to obtain a further understanding about the polynomial hierarchy, we need to define a class which contains complement of all sets in Σ_i^p classes, which we call the Π_i^p classes.

Definition 3.2.7 (Π_k^p classes)**.** A set $A \subseteq \mathbb{B}^*$ is in Π_i^p class iff there exists a polynomial time `TM` M and polynomial q such that A is of the form

$$\left\{x \in \mathbb{B}^* : \forall y_1 \in \mathbb{B}^{q(|x|)}, \exists y_2 \in \mathbb{B}^{q(|x|)}, \ldots, Q_i y_i \in \mathbb{B}^{q(|x|)}, M(x, y_1, \ldots, y_i) = 1\right\}$$

where Q_i is quantifier '$\forall$' if i is odd and '$\exists$' if i is even. ◀

The polynomial hierarchy can also be written in terms of nested complete-oracle relativization on an `nTM` machine as follows.

Proposition 3.2.8. *For all $k \geq 1$, $\Sigma_{k+1}^p = \mathbf{NP}^{\Sigma_k^p}$.*

Noting that $\Sigma_1^p = \mathbf{NP}$, we have $\Sigma_2^p = \mathbf{NP}^{\mathbf{NP}}$, $\Sigma_3^p = \mathbf{NP}^{\mathbf{NP}^{\mathbf{NP}}}$, and so on, where we have "nested" **NP**-complete oracles. This is mainly due to $\forall y_1, \exists y_2, \ldots, Q_k y_k, M(x, y_1, \ldots, y_k) = 1$ being the negation of a statement which can be queried to an **NP**-complete oracle equipped with $k-1$ nested **NP**-complete oracles equivalent to the k-th hierarchy $\mathbf{NP}^{\mathbf{NP}^{\cdot^{\cdot}}} = \Sigma_k^p$.

[4]We omit the rather technical proof of **NP**-completeness of **SAT**. Interested readers should refer to [AB09, Chapter 2.3].

This nested oracle is then used as an oracle to a polynomial nTM, giving $\mathbf{NP}^{\Sigma_k^p}$ to decide the statement $\exists y_1, \forall y_2, \ldots, Q_{k+1} y_{k+1}, M(x, y_1, \ldots, y_{k+1}) = 1$. Conversely, we can construct an nTM Φ equipped with a $\Sigma_k\mathbf{SAT}$ oracle[5] such that $\Phi(x) = 1$ if $x \in A$ for $A \in \Sigma_{k+1}^p$ with path $p \in \mathbb{B}^j$ and answers to the oracle queries $a \in \mathbb{B}^l$. Readers who are interested in a formal proof for $k = 2$ may refer to [AB09, Theorem 5.12].

An illustration of the Polynomial Hierarchy can be found in [**Figure 3.3**] where the relationship between the Σ_j^p classes and Π_j^p classes are shown. In this illustration, we also define the class $\Delta_j^p = \mathbf{P}^{\Sigma_{j-1}^p}$ for the sake of completeness of the illustration although it will not particularly be relevant in our discussion later.

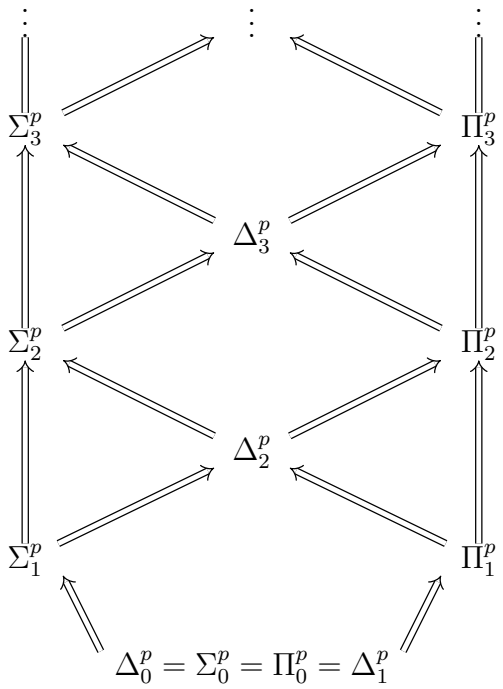

Figure 3.3: Illustration of the Polynomial Hierarchy. For classes **A** and **B** in this illustration, we denote $\mathbf{A} \Rightarrow \mathbf{B}$ if **B** contains **A** (i.e: $\forall A \in \mathbf{A}$, we have $A \in \mathbf{B}$). We define the class Σ_j^p as in [**Definition 3.2.4**] and the class Π_j^p as in [**Definition 3.2.7**]. For the sake of the completeness of the illustration of the Polynomial Hierarchy, we also define the class $\Delta_j^p = \mathbf{P}^{\Sigma_{j-1}^p}$.

To conclude this section, we now define a notion of *Collapse* of the polynomial hierarchy which will be a significant part of the complexity-theoretic hardness arguments for both Boson Sampling and Gaussian Boson Sampling.

Proposition 3.2.9 (Polynomial Hierarchy Collapse)**.** *If there exist* $i \in \mathbb{N}$ *such that* $\Sigma_i^p = \Pi_i^p$*, then for all* $j > i$*,* $\Sigma_j^p = \Sigma_{j-1}^p$*.*

[5] $\Sigma_k\mathbf{SAT}$ is a Σ_k^p-complete problem of the form $\exists u_1, \forall u_2, \ldots, Q_k u_k, \varphi(u_1, \ldots, u_k) = 1$ where φ is a boolean function.

The proof of the polynomial hierarchy collapse essentially considers the quantified statement $Q_k y_k Q_{k-1} y_{k-1} \dots Q_1 z_1, M(x, y_k, \dots, y_1)$ and noting that if both Q_j and Q_{j-1} are the same quantifier (i.e: both being either $\forall$ or $\exists$), then we can construct a polynomial-time pairing function g so both can be expressed as one quantifier, either as $\forall g(y_j, y_{j-1})$ or $\exists g(y_j, y_{j-1})$. This proof is given in more detail below.

Proof. Let $A \in \Sigma_j$ for $j > i$. Then there exists `TM` M such that

$$A = \{x : \exists y_j \forall y_{j-1} \dots Q_1 y_1, M(x, y_j, \dots, y_1) = 1\}$$

where Q_j alternates between $\exists$ and $\forall$. For the same `TM` M, consider

$$\tilde{A} = \{(x, y_j, \dots, y_{i+1}) : Q_i y_i \dots Q_1 y_1, M(x, y_j, \dots, y_1) = 1\}.$$

where quantifier $Q_i = \exists$ if i is even and $Q_i = \forall$ if i is odd, therefore $\tilde{A} \in \Sigma_i^p$ if i is even and $\tilde{A} \in \Pi_i^p$ if i is odd. By our assumption, we have $\tilde{A} \in \Sigma_i^p = \Pi_i^p$, thus there exists `TM` $\tilde{M}$

$$\begin{aligned}\tilde{A} &= \{(x, y_j, \dots, y_{i+1}) : Q_i y_i \dots Q_1 y_1, M(x, y_j, \dots, y_1) = 1\} \\ &= \{(x, y_j, \dots, y_{i+1}) : P_i z_i \dots P_1 z_1, \tilde{M}(x, y_j, \dots y_{i+1}, z_i, \dots, z_1) = 1\}\end{aligned}$$

where Q_k and P_k opposite quantifiers. We can now write A in terms of $\tilde{A}$

$$\begin{aligned}A &= \{x : \exists y_j, \dots, Q_{i+1} y_{i+1}, (x, y_j, \dots, y_{i+1}) \in \tilde{A}\} \\ &= \{x : \exists y_j, \dots, Q_{i+1} y_{i+1}, P_i z_i, \dots, P_1 z_1, \tilde{M}(x, y_j, \dots y_{i+1}, z_i, \dots, z_1) = 1\}\end{aligned}$$

where Q_{i+1} and P_i are the same quantifier as $Q_i \neq P_i$. This implies that we can have a polynomial-time pairing function to combine Q_{i+1} and P_i, leaving A with only $j - 1$ quantifiers, thus allowing us to infer that $A \in \Sigma_{j-1}^p$. □

Collapse of Polynomial Hierarchy [**Proposition 3.2.9**] clearly implies that $\mathbf{PH} \subseteq \Sigma_i^p$ because for all $j > i$ we have $\Sigma_j^p = \Sigma_{j-1}^p$ and thus $\Sigma_j^p = \Sigma_i^p$. It is widely believed that the Polynomial Hierarchy does not collapse, i.e: $\Sigma_j^p \neq \Sigma_i^p$ for all $j \neq i$ despite the lack of formal proof. This belief is partly due to the Kleene's Arithmetic Hierarchy, the Recursion Theory analogue of the **PH**, having its classes in lower hierarchy level strictly contained in classes in higher hierarchy level[6]. Another reason is that there is no known polynomial-time algorithm for **NP**-complete problems. A substantial amount of examples of **NP**-complete problems are discussed by Richard Karp in [Kar72] where he mentions that the polynomial hierarchy would collapse if $\mathbf{P} = \mathbf{NP}$ as a consequence of a provable existence of a polynomial-time algorithm that would solve an **NP**-complete problem. A similar discussion on polynomial hierarchy collapse as a consequence of having some $j \geq 1$ such that $\Sigma_j^p = \Pi_j^p$ has also been presented in [MS72]. We will not pursue further discussion about other properties of **PH**. Interested readers may find more properties of **PH** in [Sto77].

[6]Classes in the Kleene's Arithmetic Hierarchy can be defined similarly as [**Definition 3.2.4**] by replacing polynomial-time Turing Machines M in the definition by Turing Machines, replacing polynomial-time computability with computability, and write Σ_j^0 instead of Σ_j^p. In the Arithmetic Hierarchy we have $\Sigma_j^0 \subsetneq \Sigma_i^0$ for all $j < i$.

3.3 Counting Problems and #P Class

So far we have only consider decision problems and their corresponding complexity classes which contain sets. However, to be able to talk about a different type of problem, namely the ones which involves classes of numerical functions, we need to define a different complexity class, the #**P** class. Our definitions closely follows [AB09, Chapter 17]. We now state a definition for #**P** class in terms of polynomial-time TM.

Definition 3.3.1 (#**P** Class [AB09, Definition 17.5])**.** We denote $f : \mathbb{B}^* \to \mathbb{N}$, a polynomial p, and polynomial time TM M such that

$$f \in \#\mathbf{P} \Leftrightarrow \exists p, \exists M, \forall x \in \mathbb{B}^*, f(x) = \left|\left\{y \in \mathbb{B}^{p(|x|)} : M(x, y) = 1\right\}\right|.$$

In words, f is in #**P** class if and only if there exists a polynomial p and a TM M such that for all $x \in \mathbb{B}^*$, $f(x)$ outputs the number of binary strings y of length $p(|x|)$, where y is the second output to M when $M(x, y) = 1$. ◀

We can equivalently, characterize #**P** class in terms of nTM by interpreting the second input y to TM $M(x, y)$ in [**Definition 3.3.1**] as a path in the computation of nTM that leads to an accepting state. This leads us to the following equivalent definition of the #**P** class.

Proposition 3.3.2 (nTM characterization of #**P** class)**.** *For polynomial* nTM $\Phi : \mathbb{B}^* \to \mathbb{B}^*$ *and function* $f : \mathbb{B}^* \to \mathbb{N}$

$$\#\mathbf{P} = \{f : f(x) = [\#\ \textit{of accepting paths for}\ \Phi(x)]\}.$$

We call the type of class #**P** as a class of *counting problems*; that is, a class of functions $f : \mathbb{B}^* \to \mathbb{N}$, as opposed to class of decision problems. However, noting the existence of bijection between $\mathbb{B}^*$ and $\mathbb{N}$, we will talk about $\mathbb{B}^*$ valued and $\mathbb{N}$ valued functions interchangeably as $\mathbb{B}^*$ values gives more convenience in terms of denoting a function that can be computed by a TM. Therefore, in order to talk about computing such function with a TM, we now define a complexity class that contains functions that can be simulated by a polynomial TM given any input, analogous to **P**. in the sense that it contains all functions computable by a polynomial TM.

Definition 3.3.3 (**FP** class)**.** Let us denote function $f : \mathbb{B}^* \to \mathbb{B}^*$ and polynomial-time TM M such that

$$\mathbf{FP} = \{f : \exists M, \forall x \in \mathbb{B}^*, f(x) = M(x)\}.$$

For a function f such that $f(x) = M(x)$ for all $x \in \mathbb{B}^*$, we say that f is computable by M, or M computes f. ◀

We note that in comparison with a TM that *decides* $A \in \mathbf{P}$ which outputs only either 0 or 1, a TM that *computes* $f \in \mathbf{FP}$ may output more than one binary string.

Analogous to how we treat the relativization of a computation with respect to a a decision problem oracle, we do the same with counting problem, that is, equipping a machine with an oracle that can be queried on a solution of a counting problem.

Definition 3.3.4 (Oracle Machine and completeness for Counting Problems)**.** For a function $f : \mathbb{B}^* \to \mathbb{B}^*$, oracle machine M^f is a TM equipped with oracle $\{(x, i) : f(x)_i = 1\}$ where $f(x)_i$ is the i-th bit of $f(x)$. To put it into words, M^f can decide in a single step whether the i-th bit of the output of the function $f(x)$ is 1. We say that M^f is an oracle machine equipped with *oracle function* f.

For a function $g : \mathbb{B}^* \to \mathbb{B}^*$ and denoting M^g as a polynomial oracle machine with oracle function g, let

$$\mathbf{FP}^g = \{f : \exists M^g, \forall x \in \mathbb{B}^*, f(x) = M^g(x)\},$$

that is, the set of all functions $f : \mathbb{B}^* \to \mathbb{B}^*$ that are computable by a polynomial machine that can query the oracle $\{(x, i) : g(x)_i = 1\}$ the i-th bit of output of any input x to g. For a class of counting problems $\mathcal{F}$, if all $f \in \mathcal{F}$, we have $f \in \mathbf{FP}^g$, then g is $\mathcal{F}$-*hard*. Furthermore, if $g \in \mathcal{F}$, then g is $\mathcal{F}$-*complete*. ◀

Lastly in this section, we would like to state an important result by Toda [Tod91] about the relationship between the class of #**P**-complete problems $\mathbf{P}^{\#\mathbf{P}}$ and the polynomial hierarchy **PH**.

Theorem 3.3.5 ([Tod91])**.** *It holds that* $\mathbf{PH} \subseteq \mathbf{P}^{\#\mathbf{P}}$.

This relation between $\mathbf{P}^{\#\mathbf{P}}$ and **PH**, along with the widely believed conjecture that **PH** does not collapse will be an important argument for the complexity of BS and GBS. We will see later that polynomial hierarchy collapse can be obtained from equipping a certain oracle to a machine that solves a class inside the polynomial hierarchy.

3.4 BPP, BQP, and Postselection (Probabilistic Computation and Quantum Computation)

To define a notion of randomized computation, we now need to define a Turing Machine that captures a notion of randomness.

Definition 3.4.1 (Probabilistic Turing Machine pTM)**.** A Probabilistic Turing Machine pTM, $\Phi : \mathbb{B}^* \to \mathbb{B}$ is defined by the tuple $(S, \Gamma, \delta_0, \delta_1)$ as a Non-deterministic Turing Machine is, but with two transition functions $\delta_0, \delta_1 : Q \times \Gamma \to Q \times \Gamma \times \{L, R, S\}$ where in each step, the probability of transitioning using δ_0 and δ_1 each is $1/2$. ◀

The pTM can be thought as similar to an nTM, except that it does not perform both transitions δ_0 and δ_1 simultaneously in the sense of [**Definition 3.1.4**], but it rather randomly uses either δ_0 or δ_1 at any given time-step. In each time-step, the probability it uses δ_0 is $1/2$ and the probability of it using δ_1 is $1/2$.

Analogous to **P** class of decision problems that are computable in polynomial time by a TM, it is natural to define a class of decision problems that are "solvable" by a polynomial pTM (i.e: a pTM that for all input $x \in \mathbb{B}^*$ halts in polynomial-time in $|x|$). This however leaves the question about how does the random transition function influence the outcome of the computation. First of all, considering a fixed input $x \in \mathbb{B}^*$ to a pTM Φ, its output $\Phi(x)$ has a random element to it. Two instances of computation may give a different output $\Phi(x)$. This

requires us to include the concept of probability of a $\Phi(x)$ outputting a certain outcome. Thus in solving a problem, one would require that the `pTM` chosen to solve that problem will output the correct answer to the problem with a reasonably high probability, in addition to the requirement of polynomial-time computation. We capture this notion more precisely in the definition of the class of problems that can be solved by a `pTM` efficiently with a high probability.

Definition 3.4.2 (**BPP** class)**.** A set A is in class **BPP** iff there exist a polynomial `pTM` Φ such that for all $x \in A$, $\Phi(x) = 1$ with probability at least $2/3$. Equivalently in analytical form, $A \in$ **BPP** iff

$$x \in A \Leftrightarrow \exists \Phi, (\mathbb{P}(\Phi(x) = 1) \geq 2/3)\,.$$

As Φ always halts in polynomial-time in input length $|x|$, implicitly $\forall y \notin A$ the probability of an error is $\mathbb{P}(\Phi(y) = 1) \leq 1/3$. ◀

As in many literatures and textbook on computational complexity theory, we note that the choice of $1/3$ and $2/3$ are arbitrary. Thus, the probability of an error can be arbitrarily close to 0 in a complexity-theoretic analysis of a probabilistic computation. It can be generally stated that a set in **BPP** class is decidable with some small error $\varepsilon > 0$, which when multiple `pTM` is ran on input x, then a vote can be taken from all k `pTM` to decide if $x \in A$. The accuracy therefore increase as k increases[7].

Now, we will state an equivalent definition of the **BPP** class in terms of `TM`. This definition intuitively states that a probabilistic computation can be interpreted as a `TM` that receives two inputs: a string subject to the decision problem and a "random" input string that acts as a source of randomness for the `TM`. We state this more precisely below.

Definition 3.4.3 (**BPP** class in terms of `TM`)**.** A set A is in class **BPP** iff there exist a polynomial `TM` M and polynomial p such that there are at least $2/3$ many strings y of length $p(|x|)$ where $M(x, y) = 1$ and for all $x \notin A$, there are at most $1/3$ many strings y of length $p(|x|)$ where $M(x, y) = 1$. Equivalently in analytical form, there exist a polynomial TM M and polynomial p such that

$$\forall x \in A, \mathbb{P}\left\{y \in \mathbb{B}^{p(|x|)} : M(x, y) = 1\right\} \geq 2/3$$
$$\text{and}$$
$$\forall x \notin A, \mathbb{P}\left\{y \in \mathbb{B}^{p(|x|)} : M(x, y) = 1\right\} \leq 1/3$$

for uniform random probability distribution, i.e: $\mathbb{P}(y) = \frac{1}{2^{p(|x|)}}$ for all $y \in \mathbb{B}^{p(|x|)}$. ◀

This is an equivalent definition to [**Definition 3.4.2**] that can be easily seen by interpreting that the second input y is a string from the set $\mathbb{B}^{p(|x|)}$ that is somehow chosen using a certain "random" procedure that is independent from the deterministic `TM`.

Class **BPP** is going to be an important concept that is essential to the complexity theoretic argument later for the sampling problems BS and GBS that are intrinsically random computations. To support this, we will now state a fact that relates the **BPP** class to the polynomial hierarchy that was shown by Lautemann.

Theorem 3.4.4 ([Lau83])**.** **BPP** $\subseteq \Sigma_2^p$

[7] For a detailed discussion, see [KSV02, Chapter 4, pp.36-37].

3.4.1 Quantum Computation

Quantum computation can either be defined as a Quantum Turing Machine (as classical computation is usually defined),or a Quantum Circuit as both are equivalent with only polynomial overhead as shown in [Yao93]. As we have discussed in [**Chapter 2**] that a composite quantum system is described by a tensor product of Hilbert spaces, here we will omit the notion of composite quantum system and loosely write the tensor product of Hilbert spaces using a single symbol $\mathcal{S}$. We will also at times use a more general lowercase letter in representing a vector in the Hilbert space and use the Dirac notation in others where the status of a vector needs to be emphasized. This is done for notational simplicity for the rest of this chapter to focus on the discussions which emphasizes more on computational instead of physical aspects of quantum computation.

Now as a definition of quantum computation that is more widely used, we will first state the definition of a quantum circuit, which is essentially a series of unitary operations on a vector in a Hilbert space describing the evolution of the quantum system. This is stated more precisely below.

Definition 3.4.5 (Uniform Quantum Circuit [AB09, Chapter 10.3], [KSV02, Section 9.3, p.88], [Vaz02, p.12])**.** A *Uniform Quantum Circuit* Z of size k operating on a quantum system described by a Hilbert Space $\mathcal{S}$ is a sequence of unitary operators $Z_1, \ldots, Z_k$ where $Z_i : \mathcal{S} \to \mathcal{S}$ such that there exist a polynomial `TM` M which given 0^k, outputs $M(0^k) \in \mathbb{B}^*$ that encodes a description of Z. Thus, if $Z_1, \ldots, Z_k$ is in the matrix form operating on $|x\rangle \in \mathcal{S}$, we have $Z(x) = Z_k \ldots Z_1 |x\rangle$. ◀

We note that the Hilbert space $\mathcal{S}$ that the unitary operators are operating on can be generalized to a tensor product fo Hilbert spaces, however as each operator may operate on different Hilbert spaces, we write in the definition only in terms of a single Hilbert space $\mathcal{S}$ for simplicity. As of the description of the circuit Z generated by the polynomial `TM` M, it was shown in [KSV02, Chapter 8] that the description of any unitary operator Z can be generated by a polynomial time `TM`.

As we have mentioned earlier about a composite quantum system that is described by a tensor product of Hilbert spaces, it is convenient to think the unitary operators as operating on at most two subsystems, each represented by a single Hilbert Space in the tensor product Hilbert Space. This can be seen more clearly in an example of a quantum circuit illustrated in [**Figure 3.4**].

For the Quantum Turing Machine `qTM` definition, intuitively described, is essentially a `TM` which (1) the symbol written on its tape, (2) the position of its read-write head, and (3) state can be in superposition of multiple binary strings. We call this a *configuration* of the `qTM` at a certain point of time, and for the set of all possible configurations C_Υ, we its superposition is a linear combination of all $c \in C_\Upsilon$. This is described more precisely below.

Definition 3.4.6 (Quantum Turing Machine `qTM` [BV97], [Vaz02])**.** A *Quantum Turing Machine* `qTM` that is defined by the tuple (Q, Γ, δ) where for

- description of symbols on the tape $a \in \Gamma$,
- current q state of Υ,

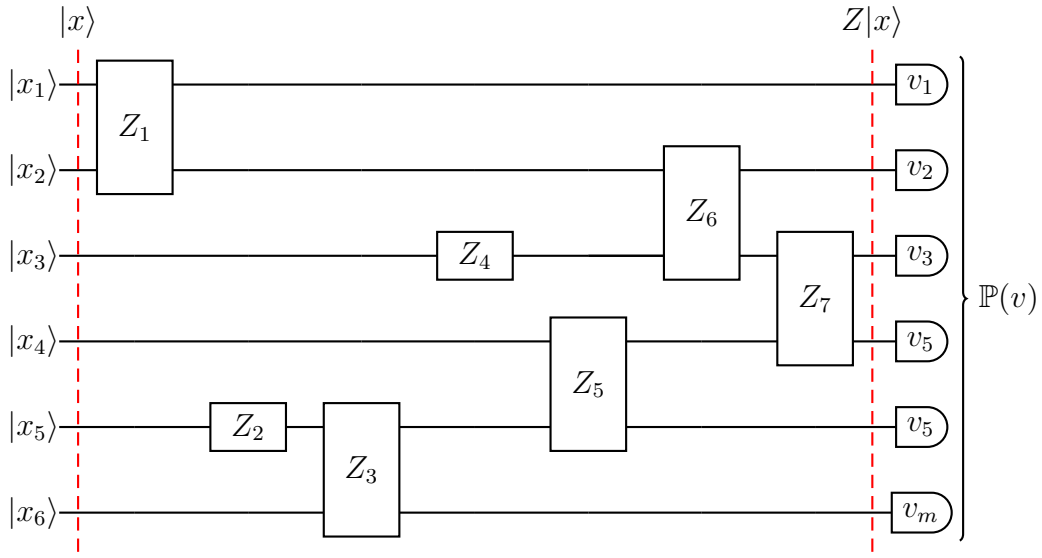

Figure 3.4: Illustration of a quantum circuit of size 7 with a composition of 6 quantum systems, which input is represented by the tensor product state $|x\rangle = \bigotimes_{j=1}^{6} = |x_j\rangle_j$. Unitary transformations Z_i that is often thought as "gates" operate on at most two subsystems, and the whole circuit operating that is represented as $Z = Z_6 \dots Z_1$ operates on the input $|x\rangle$, giving the output state $Z|x\rangle$. We also assume a measurement $\{M_v\}_v$ in some basis.

- position of the head on the tape $m \in \mathbb{Z}$,
- the subset of complex numbers $\tilde{\mathbb{C}}$ such that $x + yi \in \tilde{\mathbb{C}}$, the n-th bit of x and y can be computed in polynomial time in input length, and
- complete Hilbert space $\mathcal{S}$,

we have *configuration* $|c\rangle = |a, q, m\rangle$ of Υ where Hilbert space $\mathcal{S}$ is spanned by the set $C_\Upsilon = \{|c\rangle\}$ of all possible configurations of Υ. Transition function $\delta : \Gamma \times S \to \tilde{\mathbb{C}}^{\Gamma \times S \times \{L,R,S\}}$ tranforms $|v\rangle \in \mathcal{S}$ to $\delta(|v\rangle) = \sum_{c \in C_\Upsilon} \alpha_c |c\rangle \in \mathcal{S}$ which we call the *superposition of* Υ. For conciseness, we write the output state of Υ given an input $x \in \mathbb{B}^*$ as $|\Upsilon(x)\rangle$. ◀

This apparently complicated definition requires a few remarks. The `qTM` definition may consider only a finite number of possible configurations to prevent having an infinite dimensional Hilbert space, such as by considering only a tape that can only contain a finite number of symbols that is polynomial in the length of input. Nevertheless, with infinite dimensions we still have a perfectly valid Hilbert space. Additionally, this `qTM` definition that seems to be completely distinct from the circuit definition can be formulated to be more similarly. This is done by representing the transition between configurations of Υ as a unitary operator $U_\Upsilon : \mathcal{S} \to \mathcal{S}$ that modifies the amplitude α_c for each configuration $|c\rangle = |a, q, m\rangle$, which corresponds to a transition $\delta(a, q)$ with the read-write head on position m. Lastly, we note that the measurement on the resulting state from the `qTM` is then measured using the measurement on the basis of all possible configurations $\{M_c\}_{c \in C_\Upsilon}$ as defined in [**Section 2.1.3**]

although the part of c that describes the position of the read-write head on the tape is generally not considered, similar to the case of TM.

Although both definitions have been proposed, one should note that most literatures on quantum computation uses the quantum circuit model. More specifically for non-universal computation models, which is the primary subject of this thesis, the choice of circuit representation is much more convenient due to mere restriction to the types of unitary operations allowed in the computational model. Thus, the circuit formulation of the computation model will be used in most of discussions throughout this thesis.

Analogous to the **P** and **BPP** classes, we now define a class of problems that can be decided in polynomial time by a uniform quantum circuit where as the definitions before, we indicate an output of 1 as the input being accepted.

Definition 3.4.7 (Class **BQP**)**.** A set $A \subseteq \mathbb{B}^*$ is in **BQP** iff there exist a polynomial qTM Υ such that for all input $x \in A$ on the read-write tape of Υ, the probability of measuring an output of 1 from its read-write tape is at least $2/3$. ◀

The above definition follows the original definition of **BQP** from [BV97, Section 8.1] and a later review [Vaz02], which formulates it in terms of qTM. However as we mentioned above that the quantum circuit description is more commonly used in the literatures on NISQ models, we can also define the class **BQP** in terms of quantum circuits as shown in [AB09, Definition 10.9] and [KSV02, Chapter 9.4].

Definition 3.4.8 (Quantum circuit class **BQP**)**.** A function $f : \mathbb{B}^* \to \mathbb{B}$ is in the class **BQP** iff for all $n \in \mathbb{N}$, there exist a polynomial p and a polynomial TM M that given an input 0^n outputs a binary string $M(0^n)$ that encodes the description of a uniform quantum circuit Z of size $k \leq p(n)$ (in the sense of [**Definition 3.4.5**]) such that for all $x \in \mathbb{B}^n$, we have $Z(x) = f(x)$ with probability at least $2/3$. ◀

Note that for both definitions of **BQP**, we implicitly assume a proper measurement of the quantum mechanical system within each models. With the notations defined in [**Section 2.1.3**], we can more explicitly define a measurement $\{M_0, M_1\}$ and define $\mathbb{P}(\Upsilon(x) = 1) = \langle \Upsilon(x)|M_1^\dagger M_1|\Upsilon(x)\rangle$ and $\mathbb{P}(Z(x) = 1) = \langle x|Z_1^\dagger \dots Z_k^\dagger M_1^\dagger M_1 Z_k \dots Z_1|x\rangle$. The probability of rejecting input x is similarly defined by replacing M_0 instead of M_1.

Now we state an important concept in both probabilistic and quantum computation called the *Postselection*. This method essentially selects a set of outcomes from measurements of probabilistic or quantum computation based on a certain criteria. It was interestingly shown by Aaronson in [Aar05] that performing postselection on probabilistic or quantum computation increases its computational power.

Definition 3.4.9 (**PBPP** and **PBQP** [AA11, Definition 8])**.** Set A is in the class **PBPP** iff there exist a polynomial pTM Φ such that for all $x \in \mathbb{B}^*$ we have

- $\mathbb{P}(\Phi(x) = 1^*) > 0$,
- If $x \in A$, then $\dfrac{\mathbb{P}(\Phi(x) = 11^*)}{\mathbb{P}(\Phi(x) = 1^*)} \geq 2/3$, and
- If $x \notin A$, then $\dfrac{\mathbb{P}(\Phi(x) = 11^*)}{\mathbb{P}(\Phi(x) = 1^*)} \leq 1/3$.

If we replace Φ with qTM Υ or a uniform polynomial size quantum circuit Z, then A is in **PBQP**. ◀

The expression $\frac{\mathbb{P}(\Phi(x) = 11^*)}{\mathbb{P}(\Phi(x) = 1^*)}$ in the definition above can be interpreted as a conditional probability, that is, the probability of the second symbol of the string $\Phi(x)$ being 1 given that the first symbol being 1. The first symbol being 1 can be further interpreted as an indication that the computation succeeds (according to a certain criteria) and the second symbol being 1 indicating that the input x is accepted.

3.5 Search and Sampling Problems

Lastly in this complexity theory preliminary chapter, we would like to define sampling and search problem, which are proven to be equivalent by Aaronson in [Aar10b]. This general definition of sampling problem will be relevant to our discussions on BS and GBS later, as they too are sampling problems. But first, to be able to talk about how far a probability distribution is from another probability distribution we define the Total Variation distance between distribution $\mathcal{D}$ and $\mathcal{D}'$ as $||\mathcal{D} - \mathcal{D}'||_{\text{TV}} = \frac{1}{2}\sum_x |\mathbb{P}_{\mathcal{D}}(x) - \mathbb{P}_{\mathcal{D}'}(x)|$ where $\mathbb{P}_{\mathcal{D}}(x)$ is the probability of x according to distribution $\mathcal{D}$.

Definition 3.5.1 (**SampP** and **SampBQP**)**.** For *Sampling Problem* $S = \{\mathcal{D}_x\}_{x\in\mathbb{B}^*}$ where $\mathcal{D}_x$ is a probability distribution over $\mathbb{B}^*$, we have S is in **SampP** if there exist a polynomial pTM Φ such that for some $\varepsilon > 0$ and input $(x, 0^{\lceil \frac{1}{\varepsilon} \rceil})$ outputs a $z \in \mathbb{B}^*$ with probability $\mathbb{P}_{\mathcal{D}'}(z)$ for a distribution $\mathcal{D}'_x$ where $||\mathcal{D}_x - \mathcal{D}'_x||_{\text{TV}} \leq \varepsilon$. If Φ is replaced with qTM Υ, then S is in **SampBQP**. ◀

The string $x \in \mathbb{B}^*$ that identify each distribution in $\{\mathcal{D}_x\}_{x\in\mathbb{B}^*}$ can be understood as a parameter of the probabilistic process that that determines the probability distribution that the pTM induces given input x. Additionally, we can formulate an equivalent definition in terms of quantum circuit instead of qTM. In this case, we again consider a polynomial TM M. For all $n \in \mathbb{N}$, $M(0^n)$ uniformly outputs a description of quantum circuit Z of size $k \leq p(n)$ for some polynomial p such that for $x \in \mathbb{B}^n$, $Z(x, 0^{\lceil \frac{1}{\varepsilon} \rceil})$ outputs a sample from a probability distribution $\mathcal{D}'_x$ such that $||\mathcal{D}_x - \mathcal{D}'_x||_{\text{TV}} \leq \varepsilon$.

Although we do not give the proof of equivalence between sampling problems and search problems, we will show the definition of search problems below. Readers are referred to [Aar10a] for the proof of equivalence between these two problems.

Definition 3.5.2 (**FBPP** and **FBQP**)**.** For a *Search Problem* $R = \{U_x\}_{x\in\mathbb{B}^*}$ with nonempty sets $U_x \subseteq \mathbb{B}^*$, R is in **FBPP** if there exist a polynomial pTM Φ such that for any $\varepsilon > 0$, we have $\Phi(x, 0^{\lceil \frac{1}{\varepsilon} \rceil}) = y$ such that $\mathbb{P}(y \in U_x) \geq 1 - \varepsilon$. We call such pTM machine Φ an **FBPP** *Machine*. If we replace Φ with qTM Υ, then R is in **FBQP** and we call Υ an **FBQP** *Machine*. ◀

Once again, we note that the definition in terms of quantum circuit instead of qTM is possible. As in the case for the sampling problem, this is done by considering a TM M that for any $n \in \mathbb{N}$, $M(0^n)$ uniformly outputs a circuit Z of size $k \leq p(n)$. Given input $(x, 0^{\lceil \frac{1}{\varepsilon} \rceil})$ the circuit Z outputs $y \in U_x$ with probability at least $1 - \varepsilon$.

Chapter 4

Boson Sampling

The Boson Sampling problem that was first proposed by Aaronson and Arkhipov in [AA11] (which we will refer to as "AABS" from this point onwards and refer to Aaronson and Arkhipov as AA) is a sampling problem of a vector v from the set $C_{m,n} = \{d \in \mathbb{B}^m : \sum_{i=1}^{m} d_i = n\}$ with a probability distribution where the probability of measuring photons at the output $\mathbb{P}(v|u, A) = |\langle v|A|u\rangle|^2$ is determined by a unitary transformation A operating on initial state $|u\rangle$ for $u \in C_{m,n}$. Physically, AABS can be interpreted as shooting n non-interacting photons into the input modes of an m-mode linear interferometer that is followed by measurements to determine whether there is a photon at each of the output modes of the interferometer. This is considered as analogous to a Galton Board (or "Bean Machine") in [AA11], which illustration is shown in [**Figure 4.1**]. The Galton Board is a box with a hole on top of it and pegs (represented by grey circles in [**Figure 4.1**]) that are arranged inside it such that when a ball is put into the hole, it can fall to either side of each peg it feel onto until it reaches the bottom of the box where there are holes that the balls can end up in. This is similar to AABS where a photon that is fed into the linear interferometer will "randomly" end up in any one of the possible output modes. As in the Galton board, in an actual Boson Sampling experiment, it is possible to have more than one photon measured in one output mode. However, considering only the cases where there are n photons measured at the outputs with at most one photon in each output mode suffices for the computational complexity argument for Boson Sampling. It is also assumed that each of the n single photons are fed into the first n input modes of the interferometer. This set of states $|d\rangle$ for $d \in C_{m,n}$ where there are n photons in total, but with no more than one on each mode is called the set of *Collision-free states*. This is a subset of the Fock basis discussed in [**Section 2.2.1**] as the Fock basis spans an infinite dimensional Hilbert space due to considerations of having from 0 to arbitrarily many photons in each mode in the system.

This chapter will be separated into two parts. We will first discuss in [**Section 4.1**] three equivalent definitions of Boson Sampling to make the model clear to the reader. Then in [**Section 4.2**], we will discuss the complexity results for both the exact and approximation cases of classical simulation of AABS, where their proofs will also be outlined.

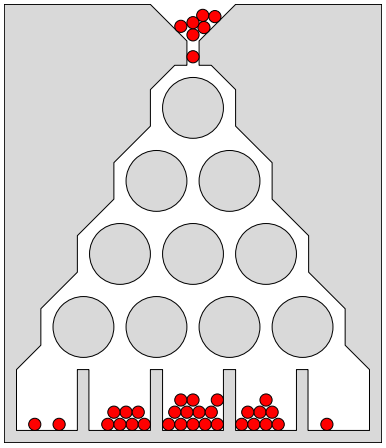

Figure 4.1: Galton Board illustration.

4.1 Boson Sampling Definitions

In this section, we will discuss about the three equivalent definitions of the Boson Sampling model presented in [AA11] based on three different paradigms: (1) Quantum Computation, (2) Fock inner product polynomial, and lastly (3) Permanent of submatrix of unitary interferometer description. The latter definition leads to the argument of the complexity of classically simulating Boson Sampling that depends on the hardness of the computation of its output probability that can be expressed in terms of the Permanent function, which was shown by Valiant in [Val79] to be #**P**-complete, or equivalently, in the class $\mathbf{P}^{\#\mathbf{P}}$. For all three definitions, we consider the quantum system of the AABS as an m-mode system, where m needs to be in the order of polynomial of n, i.e: $m = \mathcal{O}(n^c)$ for some $c \in \mathbb{N}$ and n single photon inputs. This is a requirement of the minimum number of modes of the interferometer in relation to the number of input photons.

4.1.1 Quantum Mechanics Definition

We first denote that the state of the system $|\psi\rangle$ at any point of the computation will generally be in a quantum superposition

$$|\psi\rangle = \sum_{d \in C_{m,n}} \alpha_d |d\rangle \tag{4.1.1}$$

which is a mixture of state $|d\rangle$ for $d \in C_{m,n}$ each with an amplitude[1] α_d for each corresponding d such that $\sum_{d \in C_{m,n}} |\alpha_d|^2 = 1$. This aligns with the quantum mechanics description in [**Section 2.1**] as the probability of a measurement outcome $d = (d_1, \ldots, d_m) \in C_{m,n}$ of state $|\psi\rangle$ in (4.1.1) is $\mathbb{P}(d) = |\langle d|\psi\rangle|^2 = |\alpha_d|^2$. In formalizing the measurement scheme in accordance with the projective measurement postulate [**Postulate 2.1.8**] and [**Definition 2.1.9**], we may define the measurement as $\{P_d\}_{d \in C_{m,n}}$ where each of the measurement operator is a projection to the state $|d\rangle$, that is $P_d = |d\rangle\langle d|$.

[1] the "amplitude" of each $|d\rangle$ can be thought as a weight assigned for each $|d\rangle$

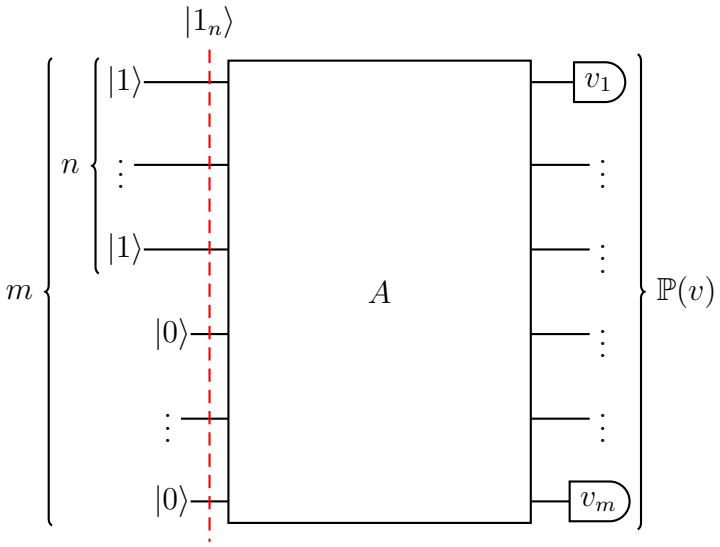

Figure 4.2: Boson Sampling circuit illustration where $|1_n\rangle$ indicates the input of single photons in the first n modes, A being the unitary matrix representation of the linear interferometer, $v_j \in \mathbb{B}$ being the measurement outcome at the j-th mode, and $\mathbb{P}(v)$ being the probability of measurement outcome $v \in D_{m,n}$.

As in [AA11], we consider the initial (prepared) state of our system as $|1_n\rangle$ where 1_n is the binary string $(1, \ldots, 1, 0, \ldots, 0)^\top$, which indicates the presence of a single photon in each of the first n out of m modes. We also have unitary transformation $A = A_k \ldots A_1$ where $A_i \in \mathbb{C}^{m,m}$ is either a beamsplitter B or a phaseshifter U that are defined in [**Section 2.2**]. By the evolution using the linear interferometer unitary representation A and photon counting measurement we then can finally formulate the probability outcome of AABS as

$$\mathbb{P}(d) = \mathbf{Tr}(|d\rangle\langle d|A|1_n\rangle\langle 1_n|A^\dagger) = |\langle d|A|1_n\rangle|^2$$

in alignment with quantum mechanics postulates discussed in [**Section 2.1**].

4.1.2 Polynomial Definition

In this second definition, a state of the AABS system is described by an m-variable (the number of modes in the system) polynomial with a degree governed by the number of photons in the modes at that point of time. Its evolution is obtained by applying a unitary operator to the variables of the polynomial where the probability distribution of the measurement outcome $s \in C_{m,n}$ is determined by the coefficients of the polynomial.

We describe the system using a a Hilbert space $\mathcal{P}_{n,m}$ of multivariate polynomials in the form of $p(z) = \sum_{s \in C_{m,n}} \gamma_s z^s$ for $z \in \mathbb{R}^m$ [2] and coefficients $\gamma_s \in \mathbb{C}$ for $z^s = z_1^{s_1} \ldots z_m^{s_m}$ equipped with Fock norm and inner product defined as follows:

[2] As in the AABS paper, it is convenient to interpret the variables $z_1, z_2, \ldots$ of the polynomial as real numbers. They are, in fact the creation operators $\hat{a}^\dagger$ that was discussed in [**Section 2.2.1**]. Therefore, having s_j photons in the j-th mode corresponds to applying the creation operator of the j-th mode $\hat{a}_j^\dagger$ s_j times, hence its exponential value.

Definition 4.1.1 (Fock norm and inner-product [AA11, Section 3.2])**.** For polynomials $p, q \in \mathcal{P}_{n,m}$ with coefficients γ_s and β_s respectively, we have *Fock Inner Product*

$$\langle p, q \rangle_f = \sum_{s \in C_{m,n}} \gamma_s^* \beta_s (s_1! \dots s_m!)$$

and *Fock Norm* $||p||_f = \sqrt{\langle p, p \rangle_f}$. ◀

We consider initial state of our system with single photons in the first n input modes of the interferometer as the polynomial $p_{1_n}(z) = z_1 \dots z_n$, where applying a unitary transformation A will result in $A[p_{1_n}](z) = p_{1_n}(Az) = \prod_{i=1}^{n} \sum_{j=1}^{m} a_{i,j} z_j$ where $a_{i,j}$ is the i-th row, j-th entry of A [3]. This is obtained by considering matrix representation of unitary transformation $A \in \mathbb{C}^{m,m}$, and

$$Az = \begin{bmatrix} \sum_j a_{1,j} z_j \\ \vdots \\ \sum_j a_{m,j} z_j \end{bmatrix},$$

where taking product of the first n terms of Az gives us precisely the polynomial $A[p_{1_n}](z)$. Lastly, a measurement outcome of s from the system in a state represented by a polynomial $p_\psi(z) = \sum_{s \in C_{m,n}} \gamma_s z^s$ will yield the probability

$$\mathbb{P}(s) = |\gamma_s|^2 s!$$

where $s! = \prod_{j=1}^{m} s_j!$.

For better clarity, below we provide an example of a system in a Hilbert space state representation $|\psi\rangle = \sum_{s \in C_{m,n}} \gamma_s |s\rangle$ in terms of the polynomial p_ψ.

Example 4.1.2. Given states $|\psi\rangle, |\varphi\rangle$ with amplitudes $\gamma_s, \beta_s \in \mathbb{C}$ respectively, we write their polynomial state representation as

$$p_\psi(z) = \sum_{s \in D_{m,n}} \frac{\gamma_s z^s}{\sqrt{s_1! \dots s_m!}} \quad \text{and} \quad p_\varphi(z) = \sum_{s \in D_{m,n}} \frac{\beta_s z^s}{\sqrt{s_1! \dots s_m!}}$$

and equivalently for $p_{|\varphi\rangle}$. Thus, we have their Fock inner product

$$\langle p_\psi, p_\varphi \rangle_f = \sum_{s \in C_{m,n}} \left(\frac{\gamma_s}{\sqrt{s_1! \dots s_m!}} \right)^* \left(\frac{\beta_s}{\sqrt{s_1! \dots s_m!}} \right) s_1! \dots s_m! = \sum_{s \in C_{m,n}} \gamma_s^* \beta_s.$$

◇

The following facts on the polynomial state representation formulation was shown by AA to argue that the polynomial definition of AABS is equivalent to the quantum mechanical definition.

[3] Here, we loosely use matrix A for both general unitary transformation and its matrix representation. A more rigorous treatment that shows the homomorphism between operators acting on a single mode and operators acting on multiple modes is discussed in detail in [AA11].

Theorem 4.1.3 ([AA11]). *For states $|\psi\rangle, |\varphi\rangle$ in the Hilbert space, we have*

$$\langle A[p_\psi], A[p_\varphi]\rangle_f = \langle p_\psi, p_\varphi\rangle$$
$$\textit{and}$$
$$||A[p_\psi]||_f^2 = 1$$

These facts ensures that the probability induced by the states described by the polynomial definition is normalized, i.e: the sum of probabilities of all measurement outcomes is equal to one, as stated using the Fock norm. Further noting that the Fock inner product $\langle p_\psi, p_\varphi\rangle$ where in [**Example 4.1.2**] coincides with $\langle\psi|\varphi\rangle$, we can see that the outcome probability of both the polynomial and the quantum mechanics definitions are equivalent[4].

4.1.3 Matrix Permanent Definition

The last definition that we need to discuss about now is the AABS definition in terms of the *Permanent function*. As before, we base this definition on single photon inputs on the first n input modes of an m mode interferometer A with outputs $d \in C_{m,n}$. The probability of obtaining measurement outcome d is therefore defined as

$$\mathbb{P}(d) = \mathbf{Per}(A_d) = \sum_{\sigma\in P_n}\prod_{i=1}^{n} A_{d\,i,\sigma(i)}$$

where P_n is all symmetric group of size n (i.e: elements of $n \times n$ matrix in different rows and columns in symmetric locations), and A_d is the submatrix of A with entries that are in the the first n rows of A and in the j-th column if $d_j = 1$. We give an example of the calculation of the permanent as demonstrating the symmetric group indexing of a matrix as well.

Example 4.1.4 (Calculation of Matrix Permanent)**.** Single photon inputs at mode 1, 3, and 4 to interferometer A that result in an output pattern d with single photon in output modes 3, 6, and 8, will result in the submatrix

$$A_d = \begin{bmatrix} t_{1,3} & t_{3,3} & t_{4,3} \\ t_{1,6} & t_{3,6} & t_{4,6} \\ t_{1,8} & t_{3,8} & t_{4,8} \end{bmatrix}$$

where the permanent is

$$\mathbf{Per}(A_d) = t_{1,3}t_{3,6}t_{4,8} + t_{1,3}t_{3,8}t_{4,6} + t_{3,3}t_{1,6}t_{4,8} + t_{3,3}t_{4,6}t_{1,8} + t_{4,3}t_{1,6}t_{3,8} + t_{4,3}t_{3,6}t_{1,8}.$$

◇

In showing the equivalence of the permanent definition to the first two definitions, the following relation between the permanent and the fock inner product can be established

$$\mathbf{Per}(A_d) = \langle p_d, A[p_{1_n}]\rangle.$$

[4]The equivalence is also due to the equivalence of representations of unitary transformations in both definitions that was shown in [AA11], although due to its small relevance to the overall discussion in this thesis, is omitted here due to its technicalities.

This relation can be established by mapping the coefficients of polynomials p_d and $A[p_{1_n}]$ to the entries of A_d and further mapping variable terms $z_1 \dots z_n$ in the polynomials to a single variable as done by AA. With this equivalence between the AABS definitions in terms of the permanent and in terms of the polynomial, the equivalence of all three definitions follows directly.

4.2 Complexity of Boson Sampling

The AABS scheme was shown to be hard to simulate on a classical computer, both exactly and approximately assuming that the Polynomial Hierarchy does not collapse (as discussed briefly in [**Section 3.2**]. This hardness is in the computational complexity theoretic sense that was discussed in [**Chapter 3**], which states that there does not exist any algorithm that runs in a polynomial time that can either exactly or approximately sample from the AABS circuit. The complexity of approximate sampling were shown by AA through a series of arguments involving two more widely considered plausible conjectures that we will discuss later in this section.

The first result for the complexity of AABS states that

> "Exact Boson Sampling cannot be performed in polynomial time by a Classical Computer, unless $\mathbf{P}^{\#\mathbf{P}} = \mathbf{BPP}^{\mathbf{NP}}$."

This is due to the fact that the problem of approximating $\mathbf{Per}^2(X)$ is $\#\mathbf{P}$-hard because given an oracle V such that $\mathbf{Per}^2(X)/y \leq V(X) \leq y\mathbf{Per}^2(X)$, there exists a polynomial time algorithm that queries v polynomial-many times that can exactly compute the exact value of $\mathbf{Per}(X)$. This fact is then used to show that having being able to sample from an oracle $O(A)$ where the top left $n \times n$ entries of interferometer A is εX for some perturbation $\varepsilon > 0$, we can then approximate $\mathbf{Per}^2(X)$ with a $\mathbf{BPP}^{\mathbf{NP}^O}$ algorithm that was proposed by Stockmeyer in [Sto83]. Now we note that if there exist a polynomial time algorithm which is able to simulate an exact AABS given a source of randomness, then there is a $\mathbf{BPP}^{\mathbf{NP}^{\mathbf{P}}} = \mathbf{BPP}^{\mathbf{NP}}$ algorithm that can compute the $\#\mathbf{P}$-Hard problem of approximating $|\mathbf{Per}(X)|^2$. Since $\mathbf{BPP} \subseteq \Sigma_2^p$ by [Lau83], we then have $\mathbf{BPP}^{\mathbf{NP}} \subseteq \Sigma_3^p$. Therefore, having both $\Sigma_3^p \subseteq \mathbf{PH} \subseteq \mathbf{P}^{\#\mathbf{P}}$ and $\mathbf{P}^{\#\mathbf{P}} \subseteq \mathbf{BPP}^{\mathbf{NP}}$ implies that $\mathbf{P}^{\#\mathbf{P}} = \mathbf{BPP}^{\mathbf{NP}}$, in words, the Polynomial Hierarchy has collapsed to the third level.

4.2.1 Approximate Boson Sampling Complexity

In this subsection, we will discuss the hardness of the approximate Boson Sampling by outlining the arguments presented by AA. It is structured by starting with the problem of additive estimation of $|\mathbf{Per}(X)|^2$, called the $|\mathbf{GPE}|^2_{\pm}$, that was shown to be solvable in $\mathbf{BPP}^{\mathbf{NP}^O}$ for an approximate Boson Sampling oracle O. For unitary matrix interferometer $A \in \mathbb{C}^{m,n}$ [5], error bound β, and random string r that are fed into the oracle O, a machine that is equipped with oracle O can sample from a distribution $\mathbb{P}_{\texttt{ABS}}$ that is β away from the exact

[5] Note that, this is due to only having single photon states in the first n input modes, thus only selecting the first n columns of an $m \times m$ unitary matrix.

AABS distribution $\mathbb{P}_{\mathtt{BS}}$ in total variation distance, that is $||\mathbb{P}_{\mathtt{ABS}} - \mathbb{P}_{\mathtt{BS}}||_{\mathrm{TV}} = \frac{1}{2}\sum_s |\mathbb{P}_{\mathtt{ABS}}(s) - \mathbb{P}_{\mathtt{BS}}(s)|$. Note that the only source of randomness for the oracle comes from the binary string r, thus rendering oracle O as entirely deterministic. Then, to show that $|\mathbf{GPE}|^2_{\pm}$ is #**P**-Hard, we will need to establish a polynomial equivalence[6] between $|\mathbf{GPE}|^2_{\pm}$ and the problem of multiplicative estimation of $|\mathbf{Per}(X)|$, called the $\mathbf{GPE}_{\times}$. This is shown under the assumption that a conjecture called the Permanent Anti-concentration conjecture holds. Lastly to relate to the #**P**-Hardness of approximate Boson Sampling, it was argued that the conjecture that $\mathbf{GPE}_{\times}$ is #**P**-Hard holds, which renders $|\mathbf{GPE}|^2_{\pm}$ to be #**P**-Hard and hence $\#\mathbf{P} \subseteq \mathbf{BPP}^{\mathbf{NP}}$ if Boson Sampling can be simulated classically in a polynomial time. The majority of the details of the argument that involves technical proofs will not be presented here, and readers who are interested should refer to the original AABS paper [AA11]. However, key parts of the arguments that will be relevant to later discussions will be presented with reasonable details. We first define the $|\mathbf{GPE}|^2_{\pm}$ problem.

Problem 4.2.1 (Gaussian Permanent Additive Estimation $|\mathbf{GPE}|^2_{\pm}$)**.** For a complex Matrix $X \in \mathbb{C}^{n,n}$ whose entries are i.i.d complex numbers drawn from a complex Gaussian distribution with mean 0 and variance 1, and $\varepsilon, \delta \in (0,1)$, compute the function $g(X)$ such that $g(X) \in [p - \varepsilon n!, p + \varepsilon n!]$ for $p = |\mathbf{Per}(X)|^2$ with probability at least $1 - \delta$ with respect to X, in polynomial time in n, $1/\varepsilon$, and $1/\delta$. ◀

Outlining the proof that $|\mathbf{GPE}|^2_{\pm} \in \mathbf{FBPP}^{\mathbf{NP}^O}$ for an approximate AABS oracle O, we first note a trick that AA uses to encode a matrix $X \in \mathbb{C}^{n,n}$ and embed it in the interferometer A using a randomized $\mathbf{BPP}^{\mathbf{NP}}$ algorithm we call E. The algorithm E does this task given a particular X, with a certain success probability to output a unitary matrix $A \in \mathbb{C}^{m,n}$ for $m \geq n$ with zX being a submatrix of it for some number z that is used to encode X. The probability distribution of where in A the encoding zX is located is shown to be uniform over all X that has a nonzero probability of success encoding by E, i.e: zX as a submatrix of A is uniformly distributed across the entries of A. For example, let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}.$$

Then, the probability of $E(X)$ outputting matrix A with encoded X as its submatrix

$$A_1 = \begin{bmatrix} zx_{11} & zx_{12} \\ a_{21} & a_{22} \\ zx_{21} & zx_{22} \\ a_{41} & a_{42} \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} zx_{11} & zx_{12} \\ zx_{21} & zx_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix} \quad \text{and} \quad A_3 = \begin{bmatrix} a_{11} & a_{12} \\ zx_{11} & zx_{12} \\ a_{31} & a_{32} \\ zx_{21} & zx_{22} \end{bmatrix}$$

and other possible arrangements of submatrix are equal, as is given a matrix A, the probability of obtaining any submatrix A_d given a uniform distribution over $d \in C_{4,2}$ is equal. It is important to note that the probability of uniformly choosing two out of four rows in the

[6] this equivalence is in the sense that one is polynomial-time reducible to the other as described in [**Section 3.2**].

example above and the probability of the encoding outcome is equal. This will be important when considering the output binary string from the oracle later. Further, the probability of both drawing X from the $n \times n$ gaussian-distribution and having $E(X)$ successfully outputs a unitary A that contains zX as its submatrix is shown to coincide with the probability of drawing the same unitary matrix A from Haar distribution of unitary matrices.

Due to the property of this encoding scheme, the oracle can only sees the A matrix but not the information on which submatrix of A is the encoding zX. We now consider the the probability p_d of the oracle O that is β in total variation distance from the exact AABS drawing an outcome photon pattern d given unitary matrix A as

$$p_d = \mathbb{P}\left\{r \in \mathbb{B}^{\mathrm{poly}(m)} : O(A, \beta, r) = d\right\}.$$

Note that prior to drawing $d \in C_{m,n}$ from O, we first draw a Gaussian-random X, then constructs A that contains the encoded X from $E(X)$.

Now, an essential fact in showing that $|\mathbf{GPE}|^2_{\pm} \in \mathbf{FBPP}^{\mathbf{NP}^O}$ is that generating the Haar-random matrix A that has zX embedded to it can be equivalently done in two ways:

1. Generate Gaussian-random X, then use algorithm E to generate A such that for some $d^* \in C_{m,n}$, we have $A_{d^*} = zX$.
2. Generate a Haar-random A, then set $X = A_{d^*}/z$.

Therefore, for $d^* \in C_{m,n}$ we obtain the following equality

$$\mathbb{P}\{d \in C_{m,n} : S(p_d)\} = \mathbb{P}\{X : S(p_{d^*})\} \tag{4.2.1}$$

for any statement S. In fact, we can consequently bound the probability of the sample from O being an error as

$$\mathbb{P}\{d \in C_{m,n} : |p_d - \mathbb{P}_{\mathtt{BS}}(A_d)| > \tilde{\varepsilon}\} = \mathbb{P}\{X : |p_{d^*} - \mathbb{P}_{\mathtt{BS}}(A_{d^*})| > \tilde{\varepsilon}\} < \tilde{\delta}.$$

Further using Stockmeyer's result in [Sto83] to obtain a distribution p'_{d^*} that approximates p_{d^*} using a $\mathbf{BPP}^{\mathbf{NP}^O}$ algorithm, we obtain another bound to the probability of error for the p'_{d^*} approximation

$$\mathbb{P}\{X : |p_{d^*} - p'_{d^*}| > \gamma p_{d^*}\} < \frac{1}{2^m}$$

in polynomial time with respect to $1/\gamma$ and m. Along with another bound

$$\mathbb{P}\{X : p_{d^*} > \varepsilon'\} < \delta'$$

that can easily be shown using Markov inequality and the argument on the equivalence between distribution of $d \in C_{m,n}$ and X we have used to obtain (4.2.1), we can therefore use all three bounds to obtain a bound on the errors of the Stockmeyer algorithm in approximating $\mathbb{P}_{\mathtt{BS}}$ as follows

$$\begin{aligned}
&\mathbb{P}\{X : |p_{d^*} - \mathbb{P}_{\mathtt{BS}}(A_{d^*})| < \varepsilon n!\} \\
&\leq \mathbb{P}\{X : p_{d^*} > \varepsilon'\} + \mathbb{P}\{X : |p_{d^*} - p'_{d^*}| > \gamma p_{d^*}\} + \mathbb{P}\{X : |p_{d^*} - \mathbb{P}_{\mathtt{BS}}(A_{d^*})| > \tilde{\varepsilon}\} \\
&\leq \delta
\end{aligned}$$

where we choose the bounds such that $\delta \geq \tilde{\delta} + \delta' + 1/2^m$ and $\varepsilon n! \leq \varepsilon' + \gamma p_{d^*} + \tilde{\varepsilon}$, therefore solving $|\mathbf{GPE}|^2_{\pm}$ in $\mathbf{BPP}^{\mathbf{NP}^O}$.

What is now left is to present the argument that $|\mathbf{GPE}|^2_{\pm}$ is #**P**-Hard. We now state the definition of multiplicative estimation of Gaussian matrices permanent problem.

Problem 4.2.2 (Gaussian Permanent Multiplicative Estimation $\mathbf{GPE}_{\times}$)**.** For a complex Matrix $X \in \mathbb{C}^{k,k}$ whose entries are i.i.d complex numbers drawn from a complex Gaussian distribution with mean 0 and variance 1, and $\varepsilon, \delta \in (0,1)$, compute the function $g(X)$ such that $g(X) \in [p - \varepsilon p, p + \varepsilon p]$ for $p = \mathbf{Per}(X)$ with probability at least $1 - \delta$ with respect to X, in polynomial time with respect to k, $1/\varepsilon$, and $1/\delta$. ◀

First, as in the argument by AA, the Permanent Anti-Concentration conjecture (PACC) is used to show that $|\mathbf{GPE}|^2_{\pm}$ and $\mathbf{GPE}_{\times}$ are polynomial time equivalent. However, reducing $|\mathbf{GPE}|^2_{\pm}$ to $\mathbf{GPE}_{\times}$ can be shown directly without using PACC. PACC is used rather to reduce $\mathbf{GPE}_{\times}$ to $|\mathbf{GPE}|^2_{\pm}$. The necessity of the PACC conjecture is essential in the reduction of $\mathbf{GPE}_{\times}$ to $|\mathbf{GPE}|^2_{\pm}$, intuitively because compared to reduction of $|\mathbf{GPE}|^2_{\pm}$ to $\mathbf{GPE}_{\times}$ that uses the more robust result of the $\mathbf{GPE}_{\times}$ problem to compute $|\mathbf{GPE}|^2_{\pm}$, using result of $|\mathbf{GPE}|^2_{\pm}$ problem to estimate $\mathbf{GPE}_{\times}$ is more "prone" to error. Thus, the PACC provides an assumption that there are a small number of $X \in \mathbb{C}^{n,n}$ such that $\mathbf{Per}(X)$ is small, avoiding a high probability of sampling a matrix X such that the error in estimating $\mathbf{Per}(X)$ multiplicatively is large relative to the small value of $\mathbf{Per}(X)$. Although we do not discuss PACC in detail, this is nonetheless an important property of intermediate quantum computation models as it is a common requirement for complexity proofs in many of these models. A discussion as well as results that show how PACC property is used for computational models other than Boson Sampling can be found in [Han+18].

With the polynomial equivalence between $|\mathbf{GPE}|^2_{\pm}$ and $\mathbf{GPE}_{\times}$, under the assumption of the Permanent of Gaussian Conjecture (PGC) asserting that $\mathbf{GPE}_{\times}$ is #**P**-Hard, the argument for the complexity of Approximate Boson Sampling is therefore concluded. The shortcoming on a proof that $\mathbf{GPE}_{\times}$ is #**P**-Hard (or not) was very much noted by AA as a noteworthy gap in their argument, however, they do provide evidences of other problems that are "similar" to $\mathbf{GPE}_{\times}$ being #**P**-Hard, supporting the PGC.

Chapter 5

Gaussian Boson Sampling

Although Aaronson and Arkhipov's Boson Sampling (AABS) is motivated by the difficulties of implementing a full-scale Quantum Computer, it still poses practical implementational challenges. One of them is the absence of a device to generate a single photon deterministically, which causes an exponential decrease in the probability of successfully generating the desired of number of single photons as the input size incrases. This was being briefly mentioned in the introduction of this thesis and is evident in the case where the probability of generating a single photon in one mode p becomes p^k when k single-photons are desired.

In response to this challenge, other input preparation schemes have been proposed to bypass this shortcoming. One of the earliest of this being a scheme called the Scattershot Boson Sampling (SBS) was proposed in [Lun+14], where single photon inputs are replaced with Two-mode Squeezed State (TMSS), which essentially is a beamsplitter applied to two Single-mode Squeezed State (SMSS) that are squeezed in different axes (i.e: one being $S(r)$, the other $S(-r)$). One of the output from the beamsplitter is then fed into a linear interferometer Z where the number of photon-count is then measured at its output. However, the other output that is being measured by a photon-counting measurement is considered as a way of heralding[1] the inputs. It was shown that the complexity arguments of AABS also hold for both exact and approximate SBS.

Pushing further the attempt to create a more experimentally plausible model that is able to demonstrate quantum computational supremacy while being more general than SBS has been proposed in [Ham+17] and further analyzed in [Kru+19]. This model is referred to as the Gaussian Boson Sampling (GBS). We will use the latter literature as our main reference in this chapter. The GBS model replaces the single photon inputs of AABS with Single-mode Squeezed States (SMSS) which belongs in the class of Gaussian States discussed in [**Section 2.2.4**]. Having SBS also being a state in the class of Gaussian States, it was then shown that SBS is a special case of GBS where the unitary transformation is considered as a tensor product of the identity and the unitary representation of the interferometer. Besides having the same line of complexity argument as AABS (despite differences due to properties that are unique to GBS), it is possible in GBS to adjust parameters that are not present in

[1] The term "heralding" here means that the modes from one of the outputs of the TMSS are being observed to find out photon number occurrence in the other. These numbers are always the same due to quantum entanglement. This can be considered as a "trick" to observe the number of photons successfully prepared. This observation can be then used to discard the experiment instances where the input is not desired.

AABS which allow one to optimize the probability of the occurrence of a certain input. This input preparation of GBS that is inherently different from AABS is shown to scale better in comparison to the exponential decrease of the single-photon generation in AABS, remedying the input preparation challenge.

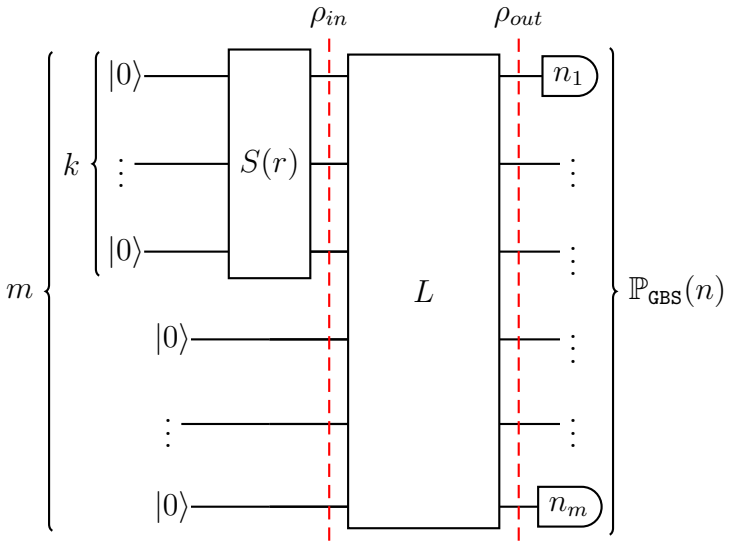

Figure 5.1: An m-mode GBS circuit where $S(r)$, $r \in \mathbb{R}^k$ is the product of l single mode squeezers, L is an m mode interferometer, and $n_1, \ldots, n_m \in \mathbb{B}$ being a Collision-free measured photon number in each of the output mode of the interferometer. Here, we have input $\rho_{in} = (\bigotimes_{j=1}^{k} |r_j\rangle_{jj}\langle r_j|) \otimes (\bigotimes_{j=k+1}^{m} |0\rangle_{jj}\langle 0|)$.

The invention of GBS that is motivated in providing a model that is hard to simulate classically, yet more feasible to experimentally implement compared to AABS, eventually bears fruit. In late 2020 during the time this thesis was being written, a group of Physicists at University of Science and Technology China announced and published in [Zho+20] their quantum supremacy demonstration using a 100-mode GBS setup. This claim, however, is not without dispute. As summarized by Scott Aaronson in [Aar20], a high photon loss incurred in the experiment and a lack of rigorous quantitative benchmarking criteria for GBS are among the reasons that weakens the quantum computational supremacy claim. Nevertheless, it has been considered as one of the most successful experimental results for the family of Boson Sampling computation model. As exciting as this discussion is, this experimental achievement will not be discussed in detail here due to the publication time of this result being very close to the end of this project, posing a strict time constraint to thoroughly assess this claim.

Similar to discussion on AABS in [**Chapter 4**], in this chapter we will first state the definition of GBS in [**Section 5.1**]. Later in [**Section 5.2**] we will then outline the complexity argument of GBS and discuss other important requirements and properties of GBS that distinguish it from AABS.

5.1 Gaussian Boson Sampling and Squeezing

As noted above, what distinguishes GBS from AABS in terms of experimental apparatus is the use of Single-mode Squeezed State (SMSS) as inputs to the linear interferometer instead of single photons as shown in [**Figure 5.1**]. Describing the referenced figure, in GBS, we consider having SMSS inputs in the first k modes of the input of the linear interferometer L and have them represented as the k-mode squeezing operator $S(r) = \bigotimes_{j=1}^{k} S_j(r_j)$ for S_j being the squeezing operator of the j-th mode operating on the first k vacuum states, resulting in $|r_j\rangle_j = S_j(r_j)|0\rangle_j$. (For definition of squeezing operator, see [**Definition 2.2.6**].) Explicitly writing this input in the density operator representation, we get

$$\rho_{in} = \left(\bigotimes_{j=1}^{k} |r_j\rangle_{jj}\langle r_j| \right) \otimes \left(\bigotimes_{j=k+1}^{m} |0\rangle_{jj}\langle 0| \right).$$

It is known that the single photon state $|1\rangle$ does not belong to the class of Gaussian states, whereas SMSS $|r\rangle$ does. Recalling the definition of Gaussian state [**Definition 2.2.19**], this fact primarily implies that the use of phase-space quasiprobability distribution (QPD) is far more convenient in GBS compared to AABS due to the Gaussian repserentation of the QPD of the output state from the GBS interferometer, which we can write as $\rho_{out} = L\rho_{in}L^\dagger$.

As mentioned in [**Section 2.2.4**], Gaussian states can be completely described using its first and second moment, namely the mean vector and the covariance matrix. However, it is convenient and can be easily generalized to consider the mean as 0, leaving only the covariance matrix as a description of the state of the system in the definition and analysis of GBS as done in [Kru+19]. This is due to the fact that the mean value being merely a "displacement" of the mean from the observables through the displacement operator that was stated in [**Proposition 2.2.4**] and briefly discussed at the end of [**Section 2.2.4**].

Now, using the QPD representation, we can use the overlap property [**Proposition 2.2.18**] to write down the measurement outcome probability

$$\mathbb{P}_{\texttt{GBS}}(n) = \pi^m \int_{\mathbb{C}^m} W^{(s)}_{\rho_{out}}(\alpha) W^{(-s)}_{|n\rangle\langle n|}(\alpha)\, d\alpha$$

for POVM element $|n\rangle\langle n| = \bigotimes_{j=1}^{m} |n_j\rangle_{jj}\langle n_j|$. Here, once again we restrict our attention to the Collision-free outcome measurement $n \in C_{m,k}$ as in the AABS discussion and in [Kru+19]. However, we note that the QPD representations here are over m complex numbers α instead $2m$ real numbers q and p with the usual relation $\alpha_j = \frac{1}{2}q_j + ip_j$. As in [Kru+19], we choose $s = -1$ resulting in the Q function and P function representations inside the integral with the Q function of ρ_{out} in the gaussian form

$$Q_{\rho_{in}}(\alpha) = \frac{1}{(2\pi)^m \sqrt{\mathbf{Det}(\sigma_Q)}} \mathbf{exp}\left(-\frac{\alpha^\dagger \sigma_Q^{-1} \alpha}{2} \right)$$

where $\sigma_Q = \frac{1}{2}I + \sigma$, and the P function being

$$P_{|n\rangle\langle n|}(\alpha) = \prod_{j=1}^{m} \frac{e^{|\alpha_j|^2}}{n!} \delta(\alpha_j)\delta(\alpha_j^*) \left(\frac{\partial^2}{\partial\alpha_j \partial\alpha_j^*} \right)^n.$$

Evaluating the integral that involve a series of rather technical derivations whose details were specified in [Kru+19], leave us with the output photon-count probability in terms of the Hafnian function

$$\mathbb{P}_{\texttt{GBS}}(n) = \frac{1}{n!\sqrt{\mathbf{Det}(\sigma_Q)}}\mathbf{Haf}(A_n).$$

where $n! = n_1! \dots n_m!$ and A_n being the submatrix of the $2m \times 2m$ matrix

$$A = \begin{bmatrix} 0 & I_m \\ I_m & 0 \end{bmatrix} (I_{2m} - \sigma_Q^{-1})$$

where we take the j-th and $j+m$-th column and $j-th$ and $j+m$-th row for every j such that $n_j = 1$. The Hafnian function defined as

$$\mathbf{Haf}(X) = \sum_{\mu \in PMP} \prod_{(i,j)\in\mu} x_{i,j}$$

where PMP is the Perfectly Matching Permutation set of $2k$ numbers for total number of detected photons k, which consists of all possible arrangement of splitting $\{1, \dots, 2k\}$ into two disjoint subsets of equal size, in which there will be $(2k-1)!!$ possible arrangements.

An important fact to note is that the Hafnian matrix is a generalization of the Permanent, as

$$\mathbf{Per}(X) = \mathbf{Haf}\left(\begin{bmatrix} 0 & X \\ X^\dagger & 0 \end{bmatrix}\right).$$

Due to the structure of σ, we can further express A as a block matrix

$$A = \left[\begin{array}{c|c} B & C \\ \hline C^\dagger & B^* \end{array}\right]$$

where B^* is a matrix with its individual elements being the complex conjugates of B. Then, as it was shown that for SMSS inputs, we have $B \neq 0$ and $C = 0$, it is clear that the outcome probability $\mathbb{P}(n)$ cannot be represented in terms of the Permanent. However, for a thermal state input (for definition of Thermal state, see [**Section 2.2.3**]), we have $C \neq 0$ and $B = 0$, leaving us with $\mathbf{Per}(C)$ for positive-definite Hermitian matrix C, which as discussed in [RLR15] allows a possible approximation using the $\mathbf{BPP^{NP}}$ Stockmeyer algorithm [Sto83].

5.2 Complexity of Gaussian Boson Sampling

This section will focus on discussing the complexity of the approximate GBS. But beforehand, we note that due to the possible formulation of Hafnian in terms of the Permanent, the complexity of exact GBS is ensured. Outlining briefly the argument for the complexity of exact GBS, we can easily follow the same reasoning as the argument for the complexity of AABS where we use the Stockmeyer's algorithm and the exact GBS oracle O, giving us a $\mathbf{BPP^{NP^O}}$ algorithm which can approximate the Hafnian function. Considering the special

cases where the Hafnian can be written as the Permanent, it follows that we can approximate the permanent in $\mathbf{BPP}^{\mathbf{NP}^O}$, thus implying that if there is a polynomial algorithm that simulates O, then the Polynomial Hierarchy collapses to the third level due to $\mathbf{P}^{\#\mathbf{P}} \subseteq \mathbf{BPP}^{\mathbf{NP}}$.

Now in arguing for the complexity of the approximate GBS, we give the similar problem formulation of additive approximation the Hafnian function $|\mathbf{GHE}|^2_{\pm}$ for a GBS setup with $m = k^2$ modes, restricting to collision-free measurement outcomes with a total of k photons, and k input SMSS with identical squeezing parameter r.

Problem 5.2.1 (Gaussian Hafnian Additive Estimation $|\mathbf{GHE}|^2_{\pm}$)**.** For a complex Matrix $X \in \mathbb{C}^{k,k}$ whose entries are i.i.d complex numbers drawn from a complex Gaussian distribution with mean 0 and variance 1, and $\varepsilon, \delta \in (0,1)$, compute the function $g(X)$ such that $g(X) \in [p-\xi, p+\xi]$ for $\xi = \frac{\varepsilon k!}{(cm)^k}$, $c = \frac{\tanh r}{\cosh r}$, and $p = |\mathbf{Haf}(XX^\dagger)|^2$ with probability at least $1-\delta$ with respect to X, in polynomial time in k, $1/\varepsilon$, and $1/\delta$. ◀

Simplifyng the expression of the outcome probability for a SMSS inputs and collision-free output $n \in C_{m,k}$ as in [Kru+19], we have the outcome probability as

$$\mathbb{P}_{\mathtt{GBS}}(n) = \frac{|\mathbf{Haf}(B_n)|^2}{\sqrt{\mathbf{Det}(\sigma_Q)}} \tag{5.2.1}$$

for $B = L(\bigoplus_{j=1}^m \tanh(r))L^\dagger$, we can use the same trick as the AABS proof for $|\mathbf{GPE}|^2_{\pm} \in \mathbf{BPP}^{\mathbf{NP}^\Xi}$ for Boson Sampling oracle Ξ. This time, we hide the encoding of the Gaussian-random matrix X into interferometer L, which by considering (5.2.1) implies that we can have the submatrix B_n of a Haar-random B to be proportional to the $XX^\dagger$. Analogous to the AABs argument, this shows that given the Gaussian-random matrix X, we can use a $\mathbf{BPP}^{\mathbf{NP}}$ algorithm to generate a Haar-random matrix B that encodes $XX^\dagger$. The rest of the argument, which involve upper-bounding the probability

$$\mathbb{P}\left\{X : |p_{n^*} - \mathbb{P}_{\mathtt{GBS}}(n^*)| > \frac{\varepsilon k!}{m^k}\right\} < \delta$$

for p_{n^*} being the probability of outcome $n^* \in C_{m,k}$ from the Stockmeyer's approximation algorithm with oracle O, closely follows the argument for AABS that was outlined in [**Section 4.2**].

To support the argument on the experimental feasibility of GBS compared to AABS, it was further shown in [Kru+19] that the probability of an invalid GBS experiment instance due to the occurrence of more than one photons at any one of the output mode is shown to be $\frac{1}{l}$ for l SMSS inputs with identical r such that $\sinh^2 r = 1$ for a Haar-random interferometer L. To further support the experimental feasibility of GBS, a formula to maximize the probability of obtaining the desired number of photon outputs in terms of squeezing parameter r and number of SMSS inputs l was also presented in [Kru+19].

Chapter 6

Homodyne Tomography Gaussian Boson Sampling

The intrinsically probabilistic nature of a measurement on a quantum mechanical system requires one to know the probability for each output to obtain a complete information about a state. Although the probabilities are not directly observable, one can construct them using Quantum State Tomography method that uses measurement outcomes from identical states (with phase variations). In fully embracing the Continuous Variable (CV) process beyond state preparation in GBS, we turn to Homodyne measurement (for more detailed discussion on Homodyne measurement, see [**Section 2.2.5**]) in place of photon counting measurement, which allow us to use the method of Homodyne Tomography that was first proposed in [Smi+93] to compute the probability distribution of the measurement outcomes. Tomographic computations on the Homodyne measurement outcomes can be done classically (although not necessarily efficiently, as we will see) by constructing what we call a Pattern Function, which will be derived in [**Section 6.1**]. As a universal tomography method, Homodyne Tomography can simulate any measurement scheme, as we will see shortly. Later in [**Section 6.2**] we will discuss the numerical method for the probability distribution reconstruction and a simulation implementing it, and lastly in [**Section 6.3**], we present and discuss a complete algorithm to perform approximate GBS that is reconstructed using Homodyne Tomography.

6.1 Homodyne Tomography and Pattern Function

The method of Homodyne Tomography can be elegantly derived from quantum mechanics concepts that we have seen in [**Chapter 2**], especially using the phase space formulation of quantum mechanics discussed in [**Section 2.2.4**] and the Homodyne measurement defined in [**Section 2.2.5**]. This is a rather universal method due to its ability to reconstruct the probability $\mathbb{P}(a) = \langle a|\rho|a\rangle$ of an arbitrary measurement outcome a (from an arbitrary set of basis $\{|a\rangle\}$), which can also be thought as one of the diagonal entries of density operator ρ. This is done by obtaining Homodyne measurement outcomes from a system in a state ρ which are then fed into a function we call the *Pattern Function* whose output values are then averaged. We now state the following theorem that shows this relation more precisely.

Theorem 6.1.1 (Homodyne Tomography)**.** *For an m mode system in state ρ and a projective measurement $\{|a\rangle\langle a|\}_a$, we can write the probability of obtaining measurement outcome of a as*

$$\mathbb{P}(a) = \langle a|\rho|a\rangle = \int_{\mathbb{R}^m} \int_{[0,2\pi]} \mathbb{P}(q,\theta) F_a(q,\theta)\ d^m\theta\ d^m q$$

*for phase-shifted quadrature probability defined in [**Definition 2.2.12**] and Pattern Function*

$$F_a(q,\theta) = \int_{\mathbb{R}^m} \frac{1}{(2\pi)^m} \langle a|\mathbf{exp}\left(i\zeta^\top (q - \widehat{q}_\theta)\right)|a\rangle \prod_{j=1}^m |\zeta_j|\ d\zeta_j.$$

A more general formulation of the Homodyne Tomography does exist, as discussed in [LPD95] and [Leo97, Chapter 5.3] where the efficiency of the Homodyne detectors being used to obtain the quadrature measurements is taken into account. However, this generality will not be pursued here as the Homodyne Tomography formulation that we employ suffices to conduct relevant discussions in this thesis.

Deriving the Homodyne Tomography Method and the Pattern Function

Now, we will show the Homodyne Tomography formula in [**Theorem 6.1.1**] in this subsection. First consider an m mode quantum system in a state ρ and recall the representation of density operator in terms of the m mode displacement operator D that was defined in [**Definition 2.2.3**] for $2m$ real numbers $\xi = (\xi_1, \ldots, \xi_{2m})$. This is restated below for convenience,

$$\rho = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^{2m}} \chi_\rho(\xi) D^\dagger(\xi)\ d\xi.$$

Rewriting it in polar coordinate $\zeta_\theta = (\zeta_1 \sin\theta_1, \zeta_1 \cos\theta_1, \ldots, \zeta_m \sin\theta_m, \zeta_m \cos\theta_m)$ gives us

$$\rho = \frac{1}{(2\pi)^m} \int_{[0,\pi]^m} \int_{\mathbb{R}^m} \chi_\rho(\zeta_\theta)\ D^\dagger(\zeta_\theta) \prod_{j=1}^m |\zeta_j|\ d\zeta_j\ d\theta_j, \tag{6.1.1}$$

where we have $\prod_{j=1}^m |\zeta_j|\ d\zeta_j\ d\theta_j$ because of m-dimensional polar coordinate variable change in the integration (i.e: $\int d\xi_j d\xi_{j+1}$ to $\int |\zeta_j| d\zeta_j d\theta_j$), and vector of phase-shifted quadrature operators $\widehat{q}_\theta = (\widehat{q}_{\theta,1}, \ldots, \widehat{q}_{\theta,m})^\top$ with $\widehat{q}_{\theta,j} = \widehat{q}_j \cos\theta_j - \widehat{p}_j \sin\theta_j$ as before. Now, to express ρ in terms of the phase-shifted quadrature probability, we must show the following lemma

Lemma 6.1.2. *For state ρ of an m mode system, we have the following relation between the characteristic function and the phase-shifted quadrature probability*

$$\chi(\zeta_\theta) = \int_{\mathbb{R}^m} \mathbf{exp}(i\zeta^\top q) \mathbb{P}(q,\theta)\ dq.$$

Proof. We first expand the characteristic function explicitly as

$$\begin{aligned}\chi_\rho(\zeta_\theta) &= \mathbf{Tr}(\rho D_k(\zeta_\theta)) \\ &= \mathbf{Tr}(\rho \mathbf{exp}(i\widehat{x}^\top \Omega \zeta_\theta)) \\ &= \mathbf{Tr}\left(\rho \mathbf{exp}\left(i\sum_{j=1}^{m} \zeta_j(\widehat{q}_j \cos\theta_j - \widehat{p}_j \sin\theta_j\right)\right) \\ &= \mathbf{Tr}\left(\rho \mathbf{exp}(iU^\dagger(\theta)\widehat{q}^\top \zeta U(\theta))\right)\end{aligned}$$

where in the fourth equality we used the phase shifted position quadrature operator as in equation (2.2.12) with an m mode phase shift operator $U(\theta) = U_1(\theta_1)\dots U_m(\theta_m)$. Further considering the fact that each $U_k(\theta_k)$ is a unitary operator, namely $U_k^\dagger(\theta_k) = U_k^{-1}(\theta_k)$, we use the linearity property of exponential operator $\mathbf{exp}(AXA^{-1}) = A\mathbf{exp}(X)A^{-1}$ for any X and invertible A and get

$$\begin{aligned}\chi_\rho(\zeta_\theta) &= \mathbf{Tr}\left(\rho U^\dagger(\theta)\mathbf{exp}(i\widehat{q}^\top \zeta)U(\theta)\right) \\ &= \int_{\mathbb{R}^m} \langle q|U(\theta)\rho U^\dagger(\theta)\mathbf{exp}(i\widehat{q}^\top \zeta)|q\rangle \; d^m q \end{aligned} \tag{6.1.2}$$

where to obtain the second equality the cyclic property of $\mathbf{Tr}$[1] was used. Now, consider the eigenvalue-eigenstate relation of the quadrature operator of each mode $\widehat{q}_k|q_k\rangle_k = q_k|q_k\rangle_k$ and apply them in the Taylor expansion expression of the linear operator exponential as

$$\begin{aligned}\mathbf{exp}(i\widehat{q}_k\zeta_k)|q_k\rangle_k &= \sum_{l\in\mathbb{N}} \frac{1}{l!}\left(\zeta_k\widehat{q}_k\right)^l |q_k\rangle_k \\ &= \sum_{l\in\mathbb{N}} \frac{1}{l!}\left(\zeta_k\right)^l q_k^l|q_k\rangle_k \\ &= \mathbf{exp}(i\zeta_k q_k)|q_k\rangle_k\end{aligned}$$

where we take the above example in the single mode case for clarity, although generalization to multimode case should be straightforward due to the k-th mode quadrature operator only operating on the k-th mode eigenstate. Applying this eigenvalue-eigenstate relation to (6.1.2) gives us

$$\begin{aligned}\chi_\rho(\zeta_\theta) &= \int_{\mathbb{R}^m} \langle q|U(\theta)\rho U^\dagger(\theta)|q\rangle \mathbf{exp}(iq^\top \zeta) \; d^m q \\ &= \int_{\mathbb{R}^m} \mathbb{P}(q,\theta)\mathbf{exp}(iq^\top \zeta) \; d^m q.\end{aligned}$$

where we also use the definition of $\mathbb{P}(q,\theta)$ [**Definition 2.2.12**] where $\mathbb{P}(q,\theta) = \langle q|U(\theta)\rho U^\dagger(\theta)|q\rangle$, concluding the proof. □

[1] That is, $\mathbf{Tr}(ABC) = \mathbf{Tr}(CAB)$.

With this relationship between the characteristic function and the phase-shifted quadrature probability, we can now substitute it into (6.1.1) to obtain

$$\begin{aligned}\rho &= \frac{1}{(2\pi)^m}\int_{[0,2\pi]^m}\int_{\mathbb{R}^m}\int_{\mathbb{R}^m}\mathbb{P}(q,\theta)\mathbf{exp}(iq^\top\zeta)\ D^\dagger(\zeta_\theta)\prod_{j=1}^m|\zeta_j|\ dq_j\ d\zeta_j\ d\theta_j \\ &= \int_{[0,2\pi]^m}\int_{\mathbb{R}^m}\mathbb{P}(q,\theta)\left(\int_{\mathbb{R}^m}\frac{1}{(2\pi)^m}\mathbf{exp}(iq^\top\zeta)\mathbf{exp}\left(-i\zeta^\top\widehat{q}_\theta\right)\prod_{j=1}^m|\zeta_j|\ d\zeta_j\right)\ d^mq\ d^m\theta \\ &= \int_{[0,2\pi]^m}\int_{\mathbb{R}^m}\mathbb{P}(q,\theta)\left(\int_{\mathbb{R}^m}\frac{1}{(2\pi)^m}\mathbf{exp}\left(i\zeta^\top(q-\widehat{q}_\theta)\right)\prod_{j=1}^m|\zeta_j|\ d\zeta_j\right)\ d^mq\ d^m\theta.\end{aligned}$$

Now, let us define the *Kernel function* K as

$$K(q-\widehat{q}_\theta) = \int_{\mathbb{R}^m}\frac{1}{(2\pi)^m}\mathbf{exp}\left(i\zeta^\top(q-\widehat{q}_\theta)\right)\prod_{j=1}^m|\zeta_j|\ d\zeta_j$$

which allow us to conveniently write the density operator as

$$\rho = \int_{[0,2\pi]^m}\int_{\mathbb{R}^m}\mathbb{P}(q,\theta)K(q-\widehat{q}_\theta)\ d^mq\ d^m\theta.$$

Now writing the probability of measurement outcome a from an arbitrary projective measurement $\{|a\rangle\langle a|\}_a$, we can obtain from the above formula

$$\mathbb{P}(a) = \langle a|\rho|a\rangle = \int_{[0,2\pi]^m}\int_{\mathbb{R}^m}\mathbb{P}(q,\theta)F_a(q,\theta)\ d^mq\ d^m\theta$$

for the Pattern Function F defined in [**Theorem 6.1.1**] and at the same time also proving the theorem.

6.2 Homodyne Tomography for GBS

Due to its universality, the method of Homodyne Tomography (HT) can be used to perform GBS by computing the probability distribution of its measurement outcomes as discussed in [**Chapter 5**]. However, solving GBS cannot be done classically in a polynomial time as the outcome probability of GBS $\mathbb{P}_{\texttt{GBS}}(n)$ is proportional to the Hafnian function whose classical approximation is a #**P**-hard problem. Nevertheless, the method of HT to GBS is an exploratory attempt to provide an alternative to applying approximate GBS using Homodyne Detection, which we will discuss in detail in [**Section 6.3**]. In this section, we will provide the supporting experimental demonstration of numerical simulations in reconstructing the GBS probability distribution by the way of HT method.

Recalling the exact GBS scheme, we can directly relate the hafnian function and the HT formula for the GBS probability distribution by [**Theorem 6.1.1**], producing

$$\mathbb{P}_{\texttt{GBS}}(n) = \frac{|\mathbf{Haf}(B_n)|^2}{\sqrt{\mathbf{Det}(\sigma_Q)}} = \int_{[0,2\pi]^m}\int_{\mathbb{R}^m}\mathbb{P}_{\texttt{HG}}(q,\theta)F_{|n\rangle}(q,\theta)\ d^mq\ d^m\theta \tag{6.2.1}$$

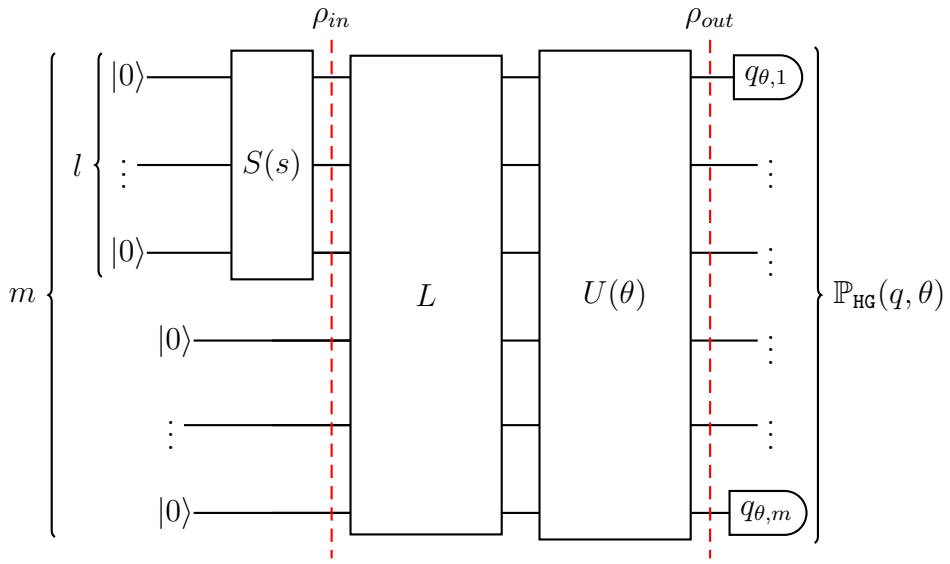

Figure 6.1: m-mode Homodyne GBS circuit where $S(s)$, $s \in \mathbb{R}^m$ is the product of l single mode squeezers, L is an m mode interferometer, $U(\theta)$ is a product of m phase shifters, each with a random θ_j uniformly distributed in the interval $[0, 2\pi]$, and output Homodyne measurement from each of the m modes $q \in \mathbb{R}^m$. Each $q_{\theta,j}$ signifies the Homodyne measurement on the interferometer output mode j after some random phase shift θ_j.

where we have $\mathbb{P}_{\mathtt{HG}}$ as the probability distribution for the measurement outcome of GBS with Homodyne measurement in place of photon counting as illustrated in [**Figure 6.1**]. We will refer to this alternate setup of GBS as *Homodyne GBS*. Having the HT integral to be proportional to the hafnian function, it directly follows that computationally evaluating this integral is a hard problem for a classical computer.

However in an experimental setup, this integral can only be evaluated exactly in the limit of infinitely many homodyne data. Hence, a numerical computation that approximates the probabilities is a more practical approach. Fortunately as discussed in [DPS03, Chapter 3.3] and [PŘ04, Chapter 2], we can employ the Monte Carlo statistical method that allows us to compute the approximation of the probabilities using finite data along with their error bounds. In this section, we will first describe the Monte Carlo method for numerically computing the GBS probability distribution and its errors due to finite data, and then discuss the computer simulation of the GBS HT. Since only Gaussian operations are used throughout the circuit that consists of squeezers, phase-shifters, and beamsplitters, the output Homodyne samples can be efficiently simulated classically by constructing the covariance matrix of the output state from the interferometer ρ_{out}.

Monte Carlo Method for Homodyne Tomography

The Monte Carlo method for Homodyne Tomography uses the property that the HT integral of the Pattern Function over quadrature values $q \in \mathbb{R}^m$ and phase $\theta \in [0, 2\pi]^m$ in (6.2.1) is equal to the average of the pattern function in the limit of $N \to \infty$ many Homodyne data,

that is

$$\mathbb{P}_{\texttt{GBS}}(k) = \lim_{N\to\infty} \frac{1}{N} \sum_{j}^{N} F_{|k\rangle}(q_j, \theta_j). \tag{6.2.2}$$

Therefore for some finitely many N Homodyne data, we have a probability that approximates the exact probability $\mathbb{P}_{\texttt{GBS}}(k)$ for each measurement outcome k,

$$\mathbb{P}^{(N)}_{\texttt{PG}}(k) = \frac{1}{N} \sum_{j}^{N} F_{|k\rangle}(q_j, \theta_j)$$

where the probability distribution is over collision-free states for l outcome photons $C_{m,l}$. We call $\mathbb{P}^{(N)}_{\texttt{PG}}(k)$ the Pattern Function GBS probability. The distribution $\mathbb{P}^{(N)}_{\texttt{PG}}$ therefore approximates $\mathbb{P}_{\texttt{GBS}}(k)$ for an output pattern k within certain error $\beta > 0$ given a sufficiently large N.

The probabilistic nature of quantum mechanics that governs the probabilistic Homodyne measurement outcomes implies that the error β can only be achieved within a certain non-unity probability. For $\mathbb{P}^{(N)}_{\texttt{PG}}(k)$, this will depend on the obtained set of N homodyne data. This is written as $Q_N = \{(q_j, \theta_j)\}_{j=1}^{N}$ where here we define the phase θ_j for each homodyne data as distributed uniformly over $[0, 2\pi]$, as in HT literatures such as [Mun+95], [Mun+96], and [Ban99]. To compute the approximation error due to finite data, we adopt the statistical sample standard deviation as the notion of error which has been used thoroughly in numerical methods for HT, such as in [DPS03], [Dar02], and [PŘ04]. We now state a general definition of the standard deviation error.

Definition 6.2.1. For some real valued function $f : \mathbb{R}^n \to \mathbb{R}$ and N, independent and identically distributed samples (or data) $\{x_j\}_{j=1}^{N}$, and its *sample mean* $\bar{f} = \frac{1}{N}\sum_{j=1}^{N} f(x_j)$, we have the *Sample Standard Deviation*

$$\mathbf{SSd}_N(f) = \sqrt{\mathbf{SVar}_N(f)}$$

for *Sample Variance*

$$\mathbf{SVar}_N(f) = \frac{1}{N-1} \sum_{j=1}^{N} \left(f(x_j) - \bar{f}\right)^2.$$

For the true standard deviation and variance, we denote them by $\mathbf{Sd}$ and $\mathbf{Var}$ where $\mathbf{Var}(f) = \int \mathbb{P}(x) \left(f(x) - \bar{f}\right)^2 \, dx$. ◀

However, computing the standard deviation of $\mathbb{P}^{(N)}_{\texttt{PG}}$ is not possible due to its mean being the probability distribution of exact GBS $\mathbb{P}_{\texttt{GBS}}$, where computing its value is an $\#\mathbf{P}$-hard problem. However, it is possible to express the standard deviation of $\mathbb{P}^{(N)}_{\texttt{PG}}$ in terms of standard deviation of $F_{|k\rangle}$, which applies to the sample standard deviation as well. We state this more precisely in the proposition below which will be subsequently proven.

Proposition 6.2.2. *For the Pattern Function GBS, we can write its sample standard deviation as*

$$\mathbf{SSd}_N\left(\mathbb{P}^{(N)}_{\mathtt{PG}}(k)\right) = \sqrt{\frac{1}{N(N-1)}\sum_{j=1}^{N}\left(F_{|k\rangle}(q_j,\theta_j) - \mathbb{P}^{(N)}_{\mathtt{PG}}(k)\right)^2}.$$

However, before showing this proposition, we will need a well known result in statistics commonly found in any standard statistics textbook, the Central Limit Theorem.

Lemma 6.2.3 (Central Limit Theorem)**.** *For a set of N independent data $\{x_j\}_{j=1}^N$ that are sampled from an arbitrary distributions, each with a mean μ_j and variance σ_j^2, then the distribution for the random variable $\bar{x}_N = \frac{1}{N}\sum_{j=1}^N x_j$ will converge to a Gaussian distribution with mean and variance*

$$\bar{\mu} = \sum_{j=1}^{N}\frac{\mu_j}{N} \quad \textit{and} \quad \bar{\sigma}^2 = \frac{1}{N^2}\sum_{j=1}^{N}\sigma_j^2,$$

respectively, as $N \to \infty$.

We simply state the central limit theorem and omit the proof as interested readers can easily find it in any standard statistics textbook. We will now give the proof for an expression for the sample standard deviation of Pattern Function GBS approximation in [**Proposition 6.2.2**].

*proof of [**Proposition 6.2.2**].* We first note that both the expected values of $\mathbb{P}^{(N)}_{\mathtt{PG}}(k)$ and $F_{|k\rangle}(q_j,\theta_j)$ are equal to the GBS probability $\mathbb{P}_{\mathtt{GBS}}(k)$. Further, we note the relationship between their variances

$$\begin{aligned}\mathbf{Var}\left(\mathbb{P}^{(N)}_{\mathtt{PG}}(k)\right) &= \mathbf{Var}\left(\frac{1}{N}\sum_{j=1}^{N}F_{|k\rangle}(q_j,\theta_j)\right)\\ &= \frac{1}{N^2}\sum_{j=1}^{N}\mathbf{Var}\left(F_{|k\rangle}(q,\theta)\right)\\ &= \frac{1}{N}\mathbf{Var}\left(F_{|k\rangle}(q,\theta)\right) \qquad (6.2.3)\end{aligned}$$

where we have used the linearity of variance in the first equality, and then used the fact that for all j, (q_j,θ_j) are distributed according to the same probability measure. For clarity, we state the definition of the variance for the pattern function below

$$\begin{aligned}\mathbf{Var}\left(F_{|k\rangle}(q,\theta)\right) &= \left(\int_{\mathbb{R}}\int_{[0,2\pi]}\mathbb{P}(q,\theta)F^2_{|k\rangle}(q,\theta)\,dq\,d\theta\right) - \mathbb{P}^2_{\mathtt{GBS}}(k)\\ &= \int_{\mathbb{R}}\int_{[0,2\pi]}\mathbb{P}(q,\theta)\left(F_{|k\rangle}(q,\theta) - \mathbb{P}_{\mathtt{GBS}}(k)\right)^2\,dq\,d\theta\end{aligned}$$

for any Homodyne sample $q_\theta \in \{(q_j, \theta_j)\}_{j=1}^N$. This integral therefore can be approximated using N Homodyne data $\{(q_j, \theta_j)\}_{j=1}^N$ by the sample variance sum

$$\mathbf{SVar}_N\left(F_{|k\rangle}(q_j, \theta_j)\right) = \frac{1}{N-1}\sum_{j=1}^{N}\left(F_{|k\rangle}(q, \theta) - \mathbb{P}_{\texttt{PG}}^{(N)}(k)\right)^2$$

where we have the sample mean $\mathbb{P}_{\texttt{PG}}^{(N)}(k)$ in place of the true probability $\mathbb{P}_{\texttt{GBS}}(k)$ where sample mean converges to as $N \to \infty$. Consequently, $\lim_{N\to\infty} \mathbf{SVar}_N\left(F_{|k\rangle}(q, \theta)\right) = \mathbf{Var}\left(F_{|k\rangle}(q, \theta)\right)$

This relationship between the variance of the approximate probability of sampled output photon pattern $\mathbb{P}_{\texttt{PG}}^{(N)}(k)$ and the variance of the value of pattern function from the Homodyne samples $F_{|k\rangle}(q, \theta)$ shows that the error of the approximate probability decreases at the rate of $\mathcal{O}(N^{\frac{1}{2}})$. Further, by the Central Limit Theorem, the set of approximation values obtained from evaluating the Pattern Function on the N Homodyne samples $\{(q_j, \theta_j)\}_{j=1}^N$ is normally distributed with $\mathbb{P}_{\texttt{GBS}}(k)$ being its mean. Now, as stated in [DPS03, Section 3.3.1]and [PŘ04, p. 2.4.4], for some $v_k \in \mathbb{R}_{>0}$ and standard deviation

$$\begin{aligned}
\mathbf{SSd}_N\left(\mathbb{P}_{\texttt{PG}}^{(N)}(k)\right) &= \sqrt{\mathbf{SVar}_N\left(\mathbb{P}_{\texttt{PG}}^{(N)}(k)\right)} \\
&= \sqrt{\frac{1}{N}\mathbf{SVar}_N\left(F_{|k\rangle}(q, \theta)\right)} \\
&= \sqrt{\frac{1}{N(N-1)}\sum_{j=1}^{N}\left(F_{|k\rangle}(q_j, \theta_j) - \mathbb{P}_{\texttt{PG}}^{(N)}(k)\right)^2}.
\end{aligned}$$

□

From here on, we will write the standard deviation of approximated GBS probability as $\varepsilon_{N,k} = \mathbf{SSd}_N\left(\mathbb{P}_{\texttt{PG}}^{(N)}\right)$. Further noting that the $\mathbb{P}_{\texttt{PG}}^{(N)}$ gets increasingly more distributed according to a Gaussian distribution with a mean $\mathbb{P}_{\texttt{GBS}}$ as N grows, we thus have the probability of $\mathbb{P}_{\texttt{GBS}}(k)$ being less than $v_k\varepsilon_{N,k}$ away from $\mathbb{P}_{\texttt{PG}}^{(N)}(k)$ for a $v_k \in \mathbb{R}$ as

$$\mathbb{P}\left\{Q_N : \Delta_k^{(N)} < v_k\varepsilon_{N,k}\right\} = \mathbf{erf}\left(\frac{v_k}{\sqrt{2}}\right) \tag{6.2.4}$$

where **erf** is the Gauss error function, and $\Delta_k^{(N)} = |\mathbb{P}_{\texttt{PG}}^{(N)}(k) - \mathbb{P}_{\texttt{GBS}}(k)|$. Now, combining [**Proposition 6.2.2**] and (6.2.4), we can numerically calculate the probability of $\mathbb{P}_{\texttt{PG}}^{(N)}(k)$ incurring a certain error for some N homodyne data Q_N.

In a numerical simulation, it will be computationally expensive to compute the sample standard deviation in the form of [**Proposition 6.2.2**] in every step (i.e: everytime a homodyne data point is generated) as the sum will need to be evaluated for all N datapoints which qrows quickly. Therefore, we will now derive another expression for $\mathbf{SSd}_N\left(\mathbb{P}_{\texttt{PG}}^{(N)}(k)\right)$ in terms of $\mathbf{SSd}_N\left(\mathbb{P}_{\texttt{PG}}^{(N-1)}(k)\right)$ to avoid evaluating the sum at every step.

Proposition 6.2.4. *We can write the sample standard deviation of* $\mathbb{P}_{\texttt{PG}}^{(N)}(k)$ *as*

$$\mathbf{SSd}_N\left(\mathbb{P}_{\texttt{PG}}^{(N)}(k)\right) = \sqrt{\frac{1}{N-1}\left(\mathbf{SSd}_{N-1}\left(\mathbb{P}_{\texttt{PG}}^{(N-1)}\right) + F_{|n\rangle}^2(q_N) - N(\mathbb{P}_{\texttt{PG}}^{(N)}(k))^2 + (N-1)(\mathbb{P}_{\texttt{PG}}^{(N-1)}(k))^2\right)}.$$

Proof. First, we write for short $\upsilon_N = \sum_{j=1}^{N}\left(F_{|k\rangle}(q_j) - P_N\right)^2$ and $P_N = \mathbb{P}_{\texttt{PG}}^{(N)}(k)$, then

$$\begin{aligned}
\upsilon_N &= \sum_{j=1}^{N}\left(F_{|k\rangle}(q_j) - P_N\right)^2 \\
&= \sum_{j=1}^{N} F_{|k\rangle}^2(q_j) - 2P_N F_{|k\rangle}(q_j) + P_N^2 \\
&= \sum_{j=1}^{N} F_{|k\rangle}^2(q_j) - 2P_N N\sum_{k=1}^{N}\frac{F_{|k\rangle}(q_k)}{N} + NP_N^2 \\
&= \sum_{j=1}^{N} F_{|k\rangle}^2(q_j) - NP_N^2
\end{aligned}$$

where we distributed the sum to obtain the third equality and used the definition of P_N to obtain the fourth equality. Now, consider

$$\begin{aligned}
\upsilon_N - \upsilon_{N-1} &= \sum_{j=1}^{N} F_{|k\rangle}^2(q_j) - \sum_{k=1}^{N-1} F_{|k\rangle}^2(q_k) - NP_N^2 + (N-1)P_{N-1}^2 \\
\upsilon_N - \upsilon_{N-1} &= F_{|k\rangle}^2(q_N) - NP_N^2 + (N-1)P_{N-1}^2 \\
\upsilon_N &= \upsilon_{N-1} + F_{|k\rangle}^2(q_N) - NP_N^2 + (N-1)P_{N-1}^2.
\end{aligned}$$

Thus, we now have obtained the expression of υ_N in terms of υ_{N-1} and the mean of the first $N-1$ values P_{N-1}. We are now ready to present the sample variance $\mathbf{SVar}_N\left(\mathbb{P}_{\texttt{PG}}^{(N)}\right)$ in terms of the newly obtained expression of υ_N,

$$\begin{aligned}
\mathbf{SVar}_N\left(\mathbb{P}_{\texttt{PG}}^{(N)}(k)\right) &= \frac{1}{N-1}\sum_{j=1}^{N}\left(F_{|k\rangle}(q_j) - P_N\right)^2 \\
&= \frac{1}{N-1}\left(\upsilon_{N-1} + F_{|k\rangle}^2(q_N) - NP_N^2 + (N-1)P_{N-1}^2\right).
\end{aligned}$$

Lastly, using the definition of the sample standard deviation $\mathbf{SSd}_N\left(\mathbb{P}_{\texttt{PG}}^{(N)}(k)\right) = \sqrt{\mathbf{SVar}_N\left(\mathbb{P}_{\texttt{PG}}^{(N)}(k)\right)}$, we obtain the desired form. □

Fock State Pattern Function

Before we proceed to the discussion on the result of computer simulations of Homodyne Tomography of GBS, we first need to show a closed form expression for the Fock state

pattern function in (6.2.1)

$$F_{|n\rangle}(q,\theta) = \int_{\mathbb{R}^m} \frac{1}{(2\pi)^m} \langle n|\mathbf{exp}\left(i\zeta^\top(q-\widehat{q_\theta})\right)|n\rangle \prod_{j=1}^{m} |\zeta_j|\, d\zeta_j.$$

We first note the phase invariance of the pattern function

$$\begin{aligned} F_{|n\rangle}(q,\theta) &= \int_{\mathbb{R}^m} \frac{1}{(2\pi)^m} \langle n|\mathbf{exp}\left(i\zeta^\top(q-\widehat{q_\theta})\right)|n\rangle \prod_{j=1}^{m} |\zeta_j|\, d\zeta_j \\ &= \int_{\mathbb{R}^m} \frac{1}{(2\pi)^m} \langle n|U^\dagger(\theta)\mathbf{exp}\left(i\zeta^\top(q-\widehat{q})\right)U(\theta)|n\rangle \prod_{j=1}^{m} |\zeta_j|\, d\zeta_j \\ &= \int_{\mathbb{R}^m} \frac{1}{(2\pi)^m} \langle n|\mathbf{exp}\left(i\zeta^\top(q-\widehat{q})\right)|n\rangle \prod_{j=1}^{m} |\zeta_j|\, d\zeta_j \end{aligned}$$

using the fact that $e^{i\theta\widehat{a}^\dagger\widehat{a}}|n\rangle = e^{i\theta n}|n\rangle$ for the number operator observable that is defined in [**Definition 2.2.1**] and that $e^{i\theta(n-n)} = 1$. Thus we will more compactly rewrite the phase-invariant pattern Fock state function by dropping the phase θ parameter, as follows

$$F_{|n\rangle}(q) = \int_{\mathbb{R}^m} \frac{1}{(2\pi)^m} \langle n|\mathbf{exp}\left(i\zeta^\top(q-\widehat{q})\right)|n\rangle \prod_{j=1}^{m} |\zeta_j|\, d\zeta_j.$$

Consequently, we can rewrite the GBS outcome probability in (6.2.1) as

$$\mathbb{P}_{\texttt{GBS}}(n) = \int_{\mathbb{R}^m}\int_{[0,2\pi]^m} \mathbb{P}_{\texttt{HG}}(q,\theta)\; d^m\theta\; F_{|n\rangle}(q)\; d^m q.$$

However, we still have the pattern function $F_{|n\rangle}$ written in terms of an integral, which may cause difficulties in implementing a computer simulation of HT. For simplicity, we will consider the pattern function for a single mode system and use the result discussed in [Leo97, Chapter 5.2] and [Leo+96] where we can express the Fock state pattern function as

$$F_{|n\rangle}(q) = \frac{d}{dq}\psi_n(q)\;\varphi_n(q)$$

where we have the *wave function* $\psi_n(q)$ and *irregular wave function* $\varphi_n(q)$. The wave function can be defined using the recurrence relation

$$\psi_0(q) = \pi^{-\frac{1}{4}} e^{-\frac{q^2}{2}},$$

and

$$\psi_n(q) = \frac{1}{\sqrt{2n+2}}(q\psi_{n-1}(q) - \psi'_{n-1}(q)),$$

whereas the irregular wave function can be defined by the recurrence relation

$$\varphi_0(q) = \pi^{\frac{3}{4}} e^{-\frac{q^2}{2}}\,\mathbf{erfi}\,(q)$$

and

$$\varphi_n(q) = \frac{1}{\sqrt{2n+2}}(q\varphi_{n-1}(q) - \varphi'_{n-1}(q))$$

for the imaginary Gauss error function $\textbf{erfi}\,(x) = 2\pi^{-\frac{1}{2}} \int_0^x e^{z^2}\, dz$ and derivative $f' = \frac{d}{dx} f$. These relations are the solutions from the stationary Schrodinger equation stated in [Leo97, eqn.(5.72)] and [Leo+96, eqn.(11)]

$$\left(-\frac{1}{2}\frac{d^2}{dq^2} + \frac{q^2}{2}\right)\xi_n = \left(n + \frac{1}{2}\right)\xi_n$$

solving for ξ_n. For their derivatives, we can similarly define a recurrence relation

$$\psi_0'(q) = \pi^{-\frac{1}{4}}(-q)e^{-\frac{q^2}{2}},$$
and
$$\varphi_0'(q) = 2\pi^{\frac{1}{4}} e^{\frac{q^2}{2}} - q\,\textbf{erfi}\,(q)$$
and
$$\xi_n'(q) = \frac{1}{\sqrt{2n+2}}((2n+2)\xi_{n-1}(q) + q\xi_{n-1}'(q) - q^2\xi_{n-1}(q))$$

for $\xi_n' \in \{\psi_n', \varphi_n'\}$, which all can be directly obtained by taking the first derivative of the recurrence relations of ψ_n and φ_n. The derivation of these recurrence relations are discussed in [Leo97, Chapter 5.2] and [Leo+96]. These relation are used to compute the pattern function in the implementation of the HT for GBS simulation, whose results are discussed in the next section.

6.2.1 Computer Simulation for Pattern Function GBS Probability Approximation

Here, we present the results of the simulation in approximating the GBS probability by computing the Pattern Function GBS (PF GBS) probability for N phase-randomized homodyne data $\{(q_j, \theta_j)\}_{j=1}^N$ using Homodyne Tomography method. The size of the GBS system we use are relatively small with only up to 20 modes. The computation of the Hafnian function uses the Hafnian function from The Walrus library [GIQ19] in Python 3.5. The complete Python code for this PF-GBS computer simulation can be found in the Github repository `https://github.com/ATanggara/GBS-PF-sim`.

There are three types of simulation runs based on the choice of interferometer: Haar-random unitary matrix, Haar-random orthogonal matrix, and a predefined interferometer we call the "square" interferometer. We do not specify the actual experimental setup that can be realized from the random unitary and orthogonal matrix, although it is to be noted that any unitary operator can be decomposed into beamsplitters and phase shifters as shown in [Rec+94]. Our simulation runs that use the square interferometer is clearly illustrated for a 6 mode Homodyne GBS in [**Figure 6.2**]. The square interferometer consists only balanced beamsplitters where for a $2m$ mode GBS, we have at first m beamsplitters for each pair of inputs, then further having $m-1$ beamsplitters after the first m beamsplitter, this is then repeated by m times, producing an interferometer with m beamsplitters between the input and the output. For the simulation runs using square interferometers, we have squeezed state input in each of the input modes of the interferometer.

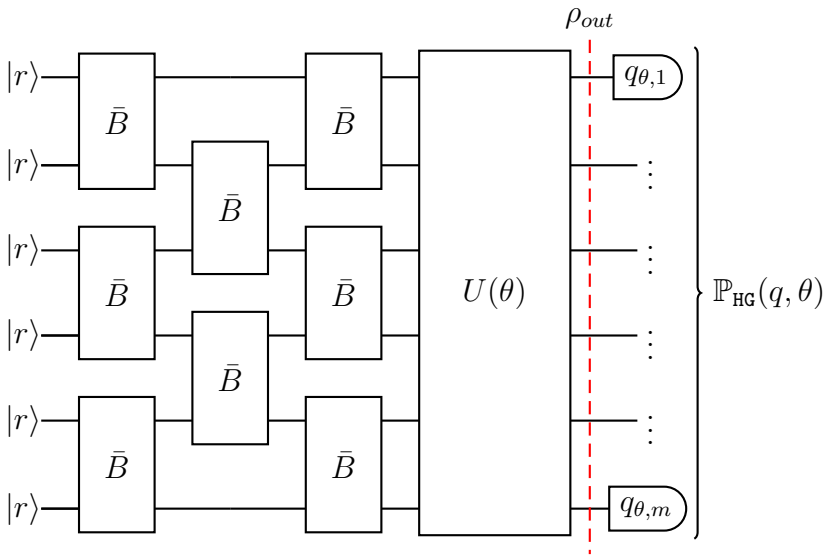

Figure 6.2: Six-mode Homodyne GBS circuit with a "square" interferometer. We have Single Mode Squeezed State (SMSS) $|r\rangle = S(r)|0\rangle$, balanced beamsplitter $\bar{B} = B(1/2)$, product of m phase shifters $U(\theta)$ each with a random phase θ_j uniformly distributed in the interval $[0, 2\pi]$, and lastly output Homodyne measurement from each of the six phase-shifted output modes $q \in \mathbb{R}^6$. Each $q_{\theta,j}$ signifies the Homodyne measurement on the interferometer output mode j after some random phase shift θ_j.

For the rest of this section, we will showcase our results as line plots of standard deviation errors and true errors of the GBS-PF approximation over the number of Homodyne data. This plots will be shown for setups with different number of modes for each types of interferometer (random orthogonal, random unitary, and square). Additionally for the square interferometer, we present a plot that shows the number of data required to achieve a certain standard deviation error. Plots presented in this chapter summarize the more detailed full results, which can be found in [**Appendix A**] either by showing the average of several simulation runs or by showing a typical case. Where necessary, we will at times refer to the detailed plots in [**Appendix A**].

Simulations using Random Interferometer

We first present and discuss the results of the PF GBS simulation for orthogonal interferometers [**Figure 6.3**] and for unitary interferometers [**Figure 6.4**]. We ran simulations with different number of modes $m \in \{4, 6, 8, 10\}$, where for each mode number m, we randomly generate 10 orthogonal interferometers and 10 unitary interferometers[2]. For each of the mode numbers, we also have a particular input single mode squeezed state arrangement and output probability of interest $\mathbb{P}_{\mathtt{PG}}(k)$:

- For the 4 mode simulations, we have $k = (0, 1, 1, 0)$ and 2 SMSS inputs with $r = 0.5$.

[2]These randomly generated interferometers are in the form of $2m \times 2m$ sympletic matrices.

- For 6 mode simulations, we have $k = (0, 0, 1, 1, 0, 0)$ and 4 SMSS inputs with $r = 0.5$.
- For 8 mode simulations, $k = (0, 0, 1, 1, 0, 0, 1, 1)$ and 4 SMSS inputs with $r = 0.5$.
- Lastly, for 10 mode simulations, $k = (0, 0, 1, 1, 0, 0, 1, 1, 0, 0)$ and 4 SMSS inputs with $r = 0.5$.

As in the original GBS problem discussed in [**Chapter 5**], we have all of the l SMSS inputs in the first l input modes of the interferometer.

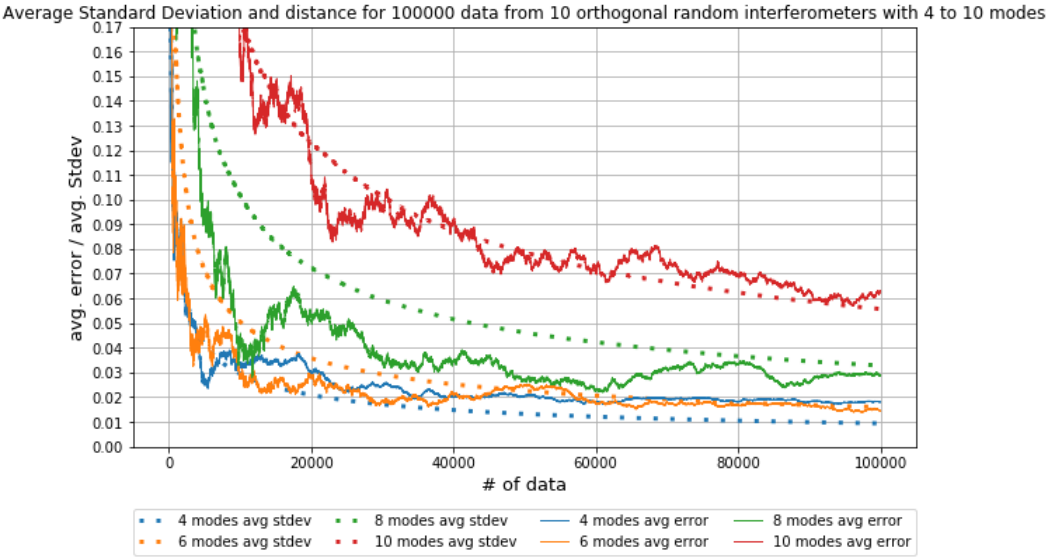

Figure 6.3: Plot of average standard deviation and average true error of PF-GBS simulation runs with $m \in \{4, 6, 8, 10\}$ modes, each with 10 orthogonal random interferometers. Each color represents the average true error $|\mathbb{P}_{\texttt{PG}}(k) - \mathbb{P}_{\texttt{GBS}}(k)|$ and average standard deviation over 10 simulation runs, each with an orthogonal random interferometer. The solid lines represent the average of the true error and the dotted lines represents the average standard deviation. For the 4 mode simulations, we have $k = (0, 1, 1, 0)$ and 2 SMSS inputs with $r = 0.5$. For 6 mode simulations, we have $k = (0, 0, 1, 1, 0, 0)$ and 4 SMSS inputs with $r = 0.5$. For 8 mode simulations, $k = (0, 0, 1, 1, 0, 0, 1, 1)$ and 4 SMSS inputs with $r = 0.5$. Lastly, for 10 mode simulations, $k = (0, 0, 1, 1, 0, 0, 1, 1, 0, 0)$ and 4 SMSS inputs with $r = 0.5$.

It can be seen from both plots in [**Figure 6.3**] and [**Figure 6.4**] that the true error, i.e: how much does an approximation $\mathbb{P}_{\texttt{PG}}(k)$ deviates from a true value $\mathbb{P}_{\texttt{GBS}}(k)$, calculated by $|\mathbb{P}_{\texttt{PG}}(k) - \mathbb{P}_{\texttt{GBS}}(k)|$ are mostly below the standard deviation throughout the simulation for 100000 Homodyne data. To relate this result with the standard deviation of the approximation $\mathbb{P}_{\texttt{PG}}(k)$, we must recall the Central Limit Theorem [**Lemma 6.2.3**], which states that the distribution of the approximation $\mathbb{P}_{\texttt{PG}}(k)$ is converging towards a Gaussian distribution with a mean $\mathbb{P}_{\texttt{GBS}}(k)$ and standard deviation presented in [**Proposition 6.2.2**]. This is due to $\mathbb{P}_{\texttt{PG}}(k)$ being the average of the pattern functions $F_{|k\rangle}$ of the Homodyne data.

Having the distribution of $\mathbb{P}_{\texttt{PG}}$ being closer to a Gaussian distribution as we have more Homodyne samples, we can therefore say that the probability of the true error $|\mathbb{P}_{\texttt{PG}}(k) -$

$\mathbb{P}_{\texttt{GBS}}(k)|$ exceeding the standard deviation is governed by the Gauss error function $\mathbf{erf}\left(\frac{1}{\sqrt{2}}\right)$ where

$$\mathbf{erf}\,(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt.$$

Thus, the probability of the true error exceeding the standard deviation is $\mathbf{erf}\left(\frac{1}{\sqrt{2}}\right) \approx 0.68$. This proeprty is also noted in literatures on Homodyne Tomography, such as [DPS03] and [PŘ04]. Both simulation results that uses orthogonal internerometer and unitary interferometer appear to be in agreement as observed from plots in [**Figure 6.3**] and [**Figure 6.4**].

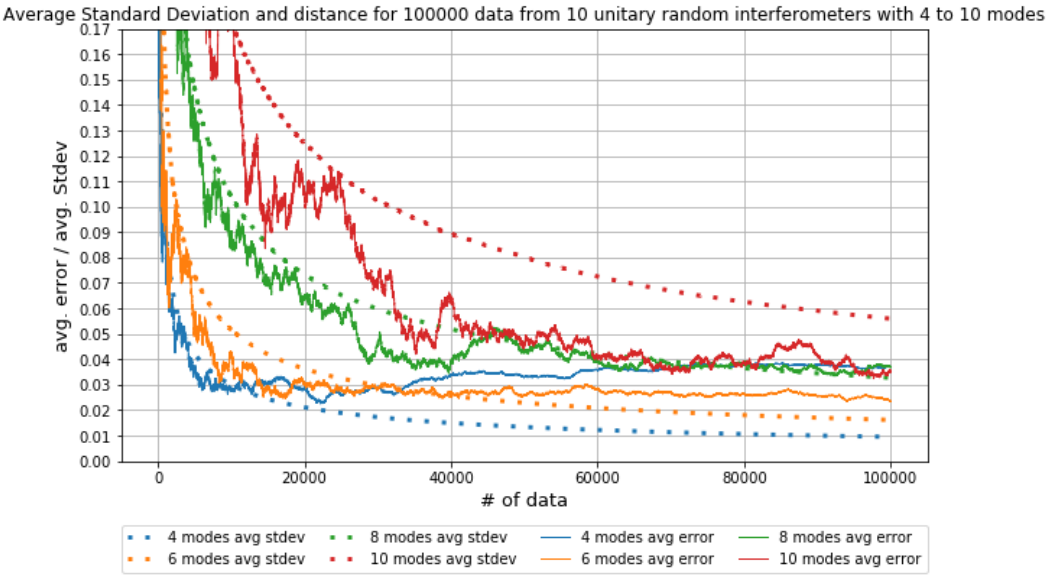

Figure 6.4: Plot of average standard deviation and average true error of PF-GBS simulation runs with $m \in \{4, 6, 8, 10\}$ modes, each with 10 unitary random interferometers. Each color represents the average true error $|\mathbb{P}_{\texttt{PG}}(k) - \mathbb{P}_{\texttt{GBS}}(k)|$ and average standard deviation over 10 simulation runs, each with an orthogonal random interferometer. The solid lines represent the average of the true error and the dotted lines represents the average standard deviation. For the 4 mode simulations, we have $k = (0, 1, 1, 0)$ and 2 SMSS inputs with $r = 0.5$. For 6 mode simulations, we have $k = (0, 0, 1, 1, 0, 0)$ and 4 SMSS inputs with $r = 0.5$. For 8 mode simulations, $k = (0, 0, 1, 1, 0, 0, 1, 1)$ and 4 SMSS inputs with $r = 0.5$. Lastly, for 10 mode simulations, $k = (0, 0, 1, 1, 0, 0, 1, 1, 0, 0)$ and 4 SMSS inputs with $r = 0.5$.

In comparing the average standard deviation and average true error of the simulations using orthogonal and unitary interferometer for different number of modes $m \in \{4, 6, 8, 10\}$, we can infer some similarities and differences. The standard deviations of simulations for both types of interferometers are clearly similar in terms of their decreasing rate and final value after the first 100000 Homodyne data. However, the average true error seems to notably differ for the case of 10 modes, where the average true error for the unitary interferometer are significantly lower from its standard deviation compared to the average true error of the

orthogonal interferometer. This can be seen from the red lines in both [**Figure 6.3**] and [**Figure 6.4**].

Another notable difference can also be seen in the case of 4 modes, represented by the blue lines in [**Figure 6.3**] and [**Figure 6.4**]. Although the difference does not seem to be as extreme as the case of 10 modes, it is apparent that the average true error for the unitary interferometer is somewhat higher compared to those for the orthogonal interferometer. Although the cause of these differences may be a mere product of the probabilistic nature of the interferometers, it can also be further analyzed by decomposing the interferometers into beamsplitters and phase shifters and looking into their arrangement which may imply a stronger quantum property such as entanglement compared to other arrangement. This analysis is not pursued here, but readers who are interested in more details may refer to the plots of both standard deviation and true error for individual interferometers that are presented and discussed in [**Appendix A**].

Simulations using Square Interferometer

Similar to the case of randomly generated interferometer, in this section we will show three simulation results that uses square interferometers.

1. Standard deviation and true error

We show the standard deviation of a PF GBS approximation $\mathbb{P}_{\texttt{PG}}(k)$ and its true error $|\mathbb{P}_{\texttt{PG}}(k) - \mathbb{P}_{\texttt{GBS}}(k)|$ as the number of Homodyne data increases up to 100000 samples in the cases of $m \in \{2, 4, 6, 8, 10\}$ modes. This result is shown in the plot in [**Figure 6.5**]. As for the case of random interferometers, it can be observed that the true errors rarely exceed their standard deviation. The behavior of the errors seems to behave even better compared to the cases for random interferometer shown in [**Figure 6.3**] and [**Figure 6.4**] where true errors exceed standard deviations more often throughout the 100000 Homodyne data.

It is also apparent that the rate at which the standard deviation decreases is shlower as the Homodyne data increases for approximations of GBS with more modes. Their corresponding true errors are also apparently worse with more fluctuations for 10 and 8 modes, that are represented by the purple line and the red line in [**Figure 6.5**]. Further details on how the actual approximation values of $\mathbb{P}_{\texttt{PG}}$ behaves throughout the 100000 Homodyne data relative to the true values $\mathbb{P}_{\texttt{GBS}}$ can be seen in [**Figure A.7**].

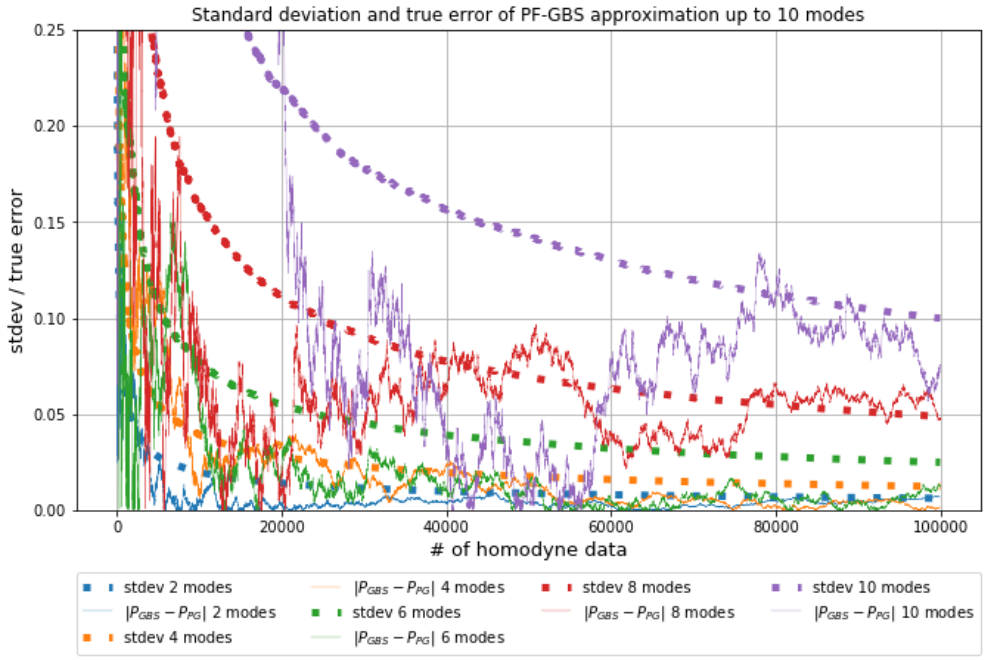

Figure 6.5: Plot of standard deviation error and true error of GBS-PF simulation run with $m \in \{2, 4, 6, 8, 10\}$ modes for the probability of measurement outcome $k = (1, \ldots, 1)$. We use a square interferometer with m input SMSS. Each color represents the true error $|\mathbb{P}_{\texttt{PG}}(k) - \mathbb{P}_{\texttt{GBS}}(k)|$ (solid lines) and standard deviation (dotted lines) of a simulation run with m modes.

2. Number of Homodyne data required to achieve a standard deviation

Secondly, we present plots that show the number of data required on average in order for the approximation $\mathbb{P}_{\texttt{PG}}(k)$ to achieve a certain standard deviation for the cases of even modes from 2 to 20 modes. Here, we are interested in the probability for outcome $k = (0, 0, \ldots, 0)$. This choice of outcome pattern is deliberate to have probabilities that are significantly larger than 0 to avoid numerical errors and have standard deviation values that are also considerably larger. This result is presented in [**Figure 6.6**]. For each mode $m \in \{2, 4, \ldots, 20\}$, we run 20 simulations, which we then take the average of the resulting standard deviations. From these average standard deviations, we then search for the minimum number of data such that the average standard deviation is less than or equal to a certain value δ.

In [**Figure 6.6**], we show this for $\delta \in \{0.01, 0.0175, 0.025, 0.05, 0.1, 0.25, 0.4, 0.6, 0.8\}$. It can be clearly seen that as we require a smaller δ, the number of Homodyne data required on average are larger for PF GBS with more modes. For a large δ, it can be seen that the number of data required to achieve δ increases very rapidly in what seems to be an exponential increase as the number of mode increases. As an example, the number of data required on average to achieve a standard deviation of 0.05 for an 8 mode PF GBS is about 27000, while a 10 mode PF GBS requires about 77000 data. For PF GBS of more than 10 modes, it is beyond 100000 data, thus is not shown in the plot.

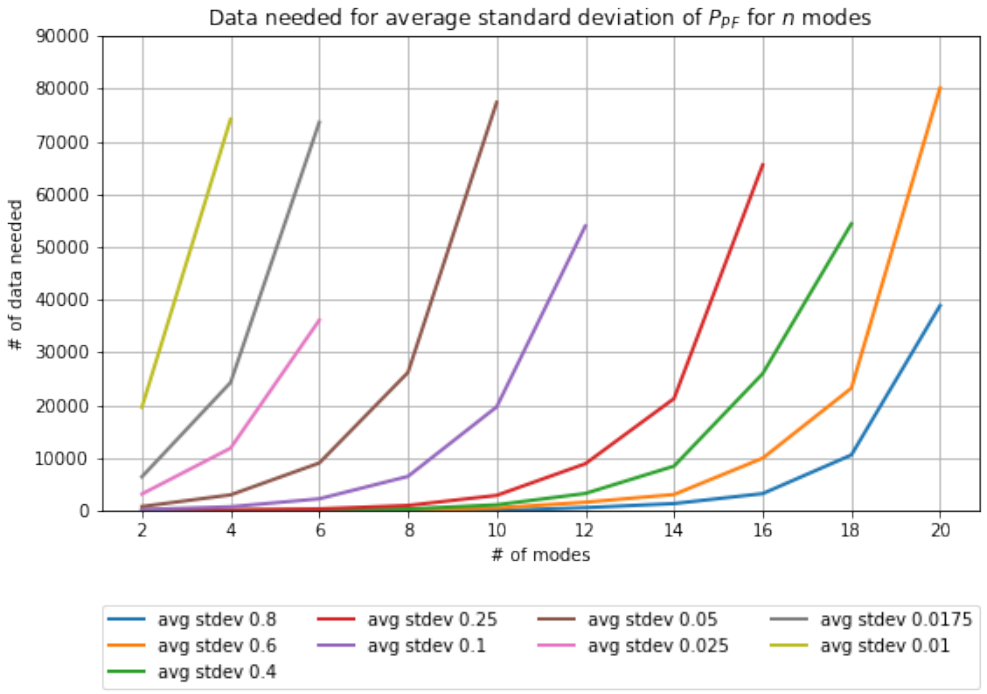

Figure 6.6: The number of Homodyne data required on average for a PF-GBS to achieve a certain standard deviation as the number of modes increases. Shown are the average over 20 runs of PF-GBS simulations from two to 20 modes.

3. Standard deviation and true errors as mode number increases

The final result that we present for the simulations using square interferometers shows the average standard deviation, average true error $|\mathbb{P}_{\texttt{PG}}(k) - \mathbb{P}_{\texttt{GBS}}(k)|$, and average difference $\mathbb{P}_{\texttt{PG}}(k) - \mathbb{P}_{\texttt{GBS}}(k)$ across the 20 simulation runs for each $m \in \{2, 4, \ldots, 20\}$. This is shown in the plot in [**Figure 6.7**]. The three averages are shown as the solid lines in the plot, while the dotted lines are the difference $\mathbb{P}_{\texttt{PG}}(k) - \mathbb{P}_{\texttt{GBS}}(k)$ for each of the 20 simulation runs for all $m \in \{2, 4, \ldots, 20\}$ modes.

The average standard deviation and average true errors are particularly relevant in relation to the previous result in [**Figure 6.6**] that shows the number of Homodyne data required to achieve a certain standard deviation. It is to be noted that both the average standard deviation and the true error of the approximation increases with the number of modes, where both reaches to about 0.5 for the case of 20 modes. In [**Figure 6.7**], the average standard deviation is represented by the blue solid line and the average true error is represented by the orange solid line. The rate which the average true error increases also appear to be in agreement with the rate which the standard deviation increases. This rapid increase of standard deviation is also apparent in the difference for each of the 20 individual simulation runs that either tends to have a larger positive values or negative values as the number of modes increases. It can be seen in the behavior of the dotted lines in [**Figure 6.7**] that are more spread away from 0 as the number of modes increases.

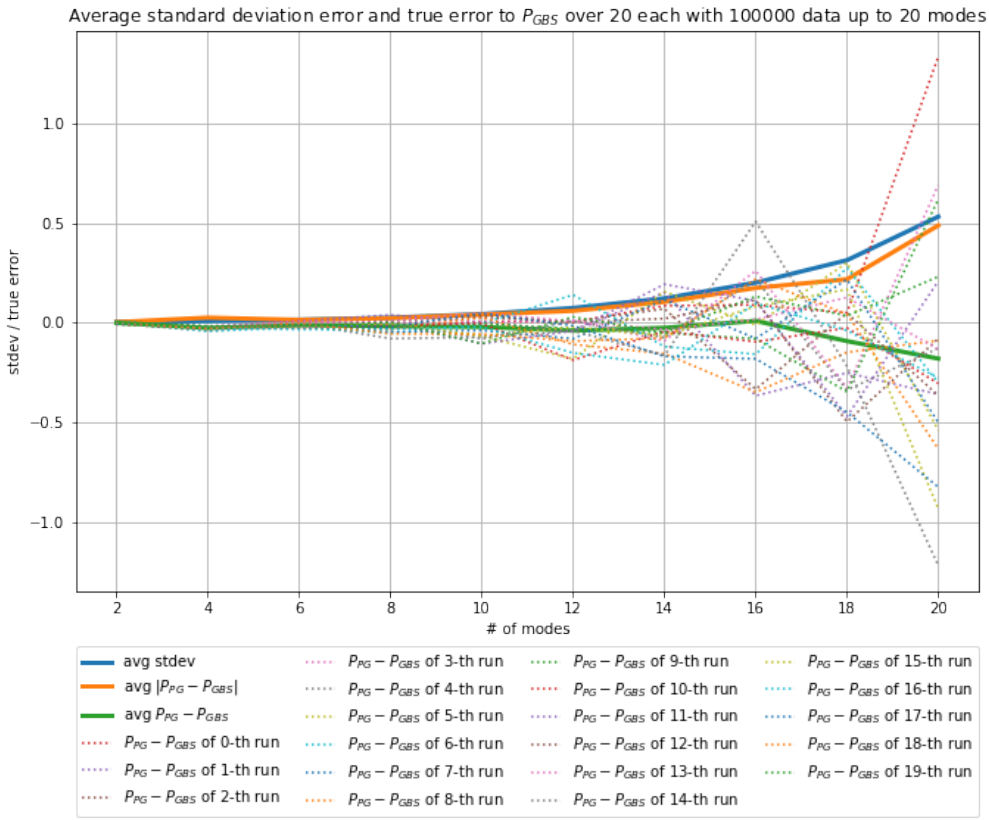

Figure 6.7: Average of standard deviation of PF-GBS and true error as the number of modes grows. Shown are the average over 20 runs of PF-GBS simulations from two to 20 modes.

6.3 Homodyne Tomography Approximate Gaussian Boson Sampling

In this section, we construct an algorithm that performs Approximate Gaussian Boson Sampling consists of l single mode squeezed state input and m-modes interferometer using the Monte Carlo method discussed in the previous section to approximate the probability for each outcome $k \in C_{m,l}$. As evident from the simulation results presented in the previous section, a computer program that implements the Homodyne Tomography method for GBS can be implemented. We call this approximation algorithm the N *Pattern Function GBS* (or, simply N PF GBS) and write the distribution induced by it as $\mathbb{P}_{\mathtt{PG}}^{(N)}$ and the probability of having an output pattern k as $\mathbb{P}_{\mathtt{PG}}^{(N)}(k) = \frac{1}{N}\sum_j^N F_{|k\rangle}(q_j, \theta_j)$ as in the previous section. This Homodyne data induced distribution $\mathbb{P}_{\mathtt{PG}}^{(N)}$ can therefore be used as an approximation to GBS with the same preparation, evolution, and measurement operations. This approximation is built on the results on the errors of $\mathbb{P}_{\mathtt{PG}}^{(N)}$ being discussed in the previous section. The accuracy of this approximation is ensured by the result [**Proposition 6.2.2**] and (6.2.4) discussed and shown in the previous section that enables one to compute the probability of having the error of $\mathbb{P}_{\mathtt{PG}}^{(N)}$ being bounded by a certain value. We now define this approximation in terms of an oracle for approximate GBS, as in the discussions on the GBS and BS.

Definition 6.3.1 (Pattern Function Approximate GBS)**.** Given a Homodyne GBS circuit that is described by the distribution $\mathbb{P}_{\texttt{HG}}$ with:

1. $m \times m$ matrix B,
2. desired error bound β, and
3. random string $r \in \{0,1\}^{\text{poly}(m)}$,

we have the *Pattern Function Approximate GBS* oracle machine $O_{\texttt{PG}}(B, \beta, r)$ that outputs $k \in \{0,1\}^m$ with probability

$$\mathbb{P}^{(N)}_{\texttt{PG}}(k) = \mathbb{P}\left\{r \in \{0,1\}^{\text{poly}(m)} : O_{\texttt{PG}}(B,\beta,r) = k\right\} = \frac{1}{N}\sum_{j=1}^{N} F_{|k\rangle}(q_j, \theta_j)$$

where $Q_N = \{q_{\theta,j}\}_{j=1}^m$ is a set of N Homodyne samples that is distributed according to $\mathbb{P}_{\texttt{HG}}$ that satisfies $\left|\left|\mathbb{P}^{(N)}_{\texttt{PG}} - \mathbb{P}_{\texttt{GBS}}\right|\right|_{\text{TV}} \leq \beta$. ◀

As shown in the previous section, we have

$$\mathbb{P}_{\texttt{GBS}}(k) = \lim_{N\to\infty} \mathbb{P}^{(N)}_{\texttt{PG}}(k) = \int_{\mathbb{R}^m}\int_{[0,2\pi]^m} \mathbb{P}_{\texttt{HG}}(q_\theta) F_{|k\rangle}(q_\theta)\, d\theta\, dq = \frac{1}{\sqrt{\textbf{Det}(\sigma_Q)}}|\textbf{Haf}(B_k)|^2 \quad (6.3.1)$$

where for $q_{\theta,1}, \ldots, q_{\theta,N}$ that are distributed according to the distribution $\mathbb{P}_{\texttt{HG}}$ induced by a Homodyne GBS circuit, as stated in the previous section,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} F_{|k\rangle}(q_j, \theta_j) = \mathbb{P}_{\texttt{GBS}}(k)$$

where for each j, $q_{\theta,j}$ is dependent on the Homodyne GBS circuit (see circuit representation [**Figure 6.1**]) where the i-th output of the interferometer is phase shifted by θ_i that is uniformly distributed in $[0, 2\pi]^m$ as proposed in [Mun+95]. Building on the results shown in [**Section 6.2**], we will construct the algorithm $O_{\texttt{PG}}$ that will serve as an approximation of $\mathbb{P}_{\texttt{GBS}}$ that can act as an approximate GBS oracle to the $\textbf{BPP}^{\textbf{NP}}$ Stockmeyer's algorithm [Sto83] to solve $|\textbf{GHE}|^2_{\pm}$ as discussed in [**Section 5.2**] given a finite amount of Homodyne data.

One might note that for some $k \in C_{m,l}$, the value of $\mathbb{P}^{(N)}_{\texttt{PG}}(k)$ may not be non-negative for some $\{q_{\theta,j}\}_j$ due to the Pattern Function $F_{|k\rangle}$ being generally not non-negative everywhere. This will pose a problem, preventing one from sampling from the approximated distribution $\mathbb{P}^{(N)}_{\texttt{PG}}$ if there if a k such that $\mathbb{P}^{(N)}_{\texttt{PG}}(k) < 0$. Moreover, this will also cause the approximated distribution to be unnormalized, namely $\sum_k \mathbb{P}^{(N)}_{\texttt{PG}}(k) \neq 1$. However in this chapter we will put our sole focus on putting a bound for the approximation error and not trouble ourselves with the sampling from the approximated distribution, thus the problem of negative values will not pose an issue in the analysis later. We will get back to address this issue later.

6.3.1 Error Bound for the PF GBS Approximation

First, we restate the definition of the set of Collision-Free photon states which was considered in both AABS [AA11] and GBS [Kru+19] papers in their complexity argument restricting the sampled photon pattern to those with only either 0 or 1 photon in each mode as we have discussed previously in [**Chapter 4**] and [**Chapter 5**].

Definition 6.3.2 (Collision-Free States)**.** Let the set of *Collision-Free* states of l photons in an m-mode system

$$C_{m,l} = \left\{ x \in \mathbb{B}^m : \sum_{j=1}^{m} x_j = l \right\}.$$

◀

Similar to in AABS and GBS papers, we also aim to bound the probability of getting a uniformly distributed $k \in C_{m,l}$ such that $\Delta_k^{(N)} = \left| \mathbb{P}_{\mathtt{PG}}^{(N)}(k) - \mathbb{P}_{\mathtt{GBS}}(k) \right| > \frac{\varepsilon}{2} \frac{l!}{m^l}$ to be smaller than $\frac{\delta}{4}$. The GBS result shows that this bound can be achieved by deterministically having $||\mathcal{D} - \mathcal{D}'||_{\mathrm{TV}} = \frac{1}{2} \sum_{k \in C_{m,l}} |\mathbb{P}_{\mathcal{D}}(k) - \mathbb{P}_{\mathcal{D}'}(k)| \leq \beta$ by setting an appropriate $\beta > 0$, where $\mathcal{D}$ and $\mathcal{D}'$ are the exact GBS distribution and the approximate GBS oracle distribution, respectively. More precisely, the approximate GBS oracle is given input $(X, 0^{\frac{1}{\beta}}, r)$ where X is the Gaussian random matrix which is to be fed into the Hafnian function, $\beta > 0$ is the upper bound of the Total Variation Distance between $\mathbb{P}_{\mathtt{PG}}^{(N)}$ and $\mathbb{P}_{\mathtt{GBS}}$, and $r \in \mathbb{B}^{\mathrm{poly}(m)}$ is the only randomness source for the oracle to achieve

$$\mathbb{P}\left\{ k : \Delta_k^{(N)} > \frac{\varepsilon}{2} \frac{l!}{m^l} \right\} < \frac{\delta}{4}$$

where we have $\frac{\varepsilon}{2} = \frac{12\beta}{\delta}$.

Now, using the PF GBS machine acting as our approximate GBS oracle, we are interested in the distance between the distribution from N Homodyne data from a Homodyne GBS circuit and the exact GBS distribution

$$\left|\left| \mathbb{P}_{\mathtt{PG}}^{(N)} - \mathbb{P}_{\mathtt{GBS}} \right|\right|_{\mathrm{TV}} \leq \beta. \tag{6.3.2}$$

However, this cannot be practically achieved deterministically because of the approximating distribution $\mathbb{P}_{\mathtt{PG}}^{(N)}$ that is induced by finitely $N < \infty$ many random Homodyne data. To proceed, we first need to define

$$\mathcal{Q}_\sigma^N = \left\{ \{q_{\theta,j}\}_{j=1}^N : q_{\theta,j} \sim \mathcal{N}(0, \sigma_{\theta,j}) \right\}$$

which is the set of N Homodyne samples from the output of the Homodyne GBS circuit whose interferometer output is described by covariance matrix $\sigma = LS(s)S^\dagger(s)L^\dagger$ where for each j, we have $\sigma_{\theta,j} = U(\theta_j)\sigma U^\dagger(\theta_j)$ with random $\theta_j \in [0, 2\pi]$ that is uniformly distributed. Note that the PF GBS machine $O_{\mathtt{PG}}$ needs to reconstruct the covariance matrix σ of the output state from the interferometer from the matrix B that is being fed into the hafnian

function. This task however, can be done easily by noting the relations between them as it was defined in [Kru+19],

$$B \oplus B^* = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} (I - (\sigma + I/2)^{-1}).$$

Now to conduct the analysis on bounding the probability of the algorithm incurring a large error, we resort to model the probability measure on having a sufficiently large error $\Delta_k^{(N)}$

$$\begin{aligned} &\mathbb{P}\left\{(k, Q_N) \in C_{m,l} \times \mathcal{Q}_\sigma^N : \Delta_k^{(N)} > \frac{\varepsilon}{2}\frac{l!}{m^l}\right\} \\ &= \int_{\mathcal{Q}_\sigma^N} \mathbb{P}\left\{k : \Delta_k^{(N)} > \frac{\varepsilon}{2}\frac{l!}{m^l}\right\} \mathbb{P}(Q_N)\, dQ_N \end{aligned}$$

on the product space between the no-collision output photon sample $k \in C_{m,l}$ and the sampled Homodyne data $Q_N \in \mathcal{Q}_\sigma^N$. The reason here is because, for each k, $\Delta_k^{(N)}$ depends on Q_N because the distribution $\mathbb{P}_{\texttt{PG}}^{(N)}$ is defined by Q_N. Thus, in addition to bounding the probability measure of $k \in C_{m,l}$ such that the error is sufficiently large, we need to find both the probability measure of $Q_N \in \mathcal{Q}_\sigma^N$ that satisfy the condition (6.3.2).

It is however, relieving to note that $\left|\left|\mathbb{P}_{\texttt{PG}}^{(N)} - \mathbb{P}_{\texttt{GBS}}\right|\right|_{\text{TV}} = \sum_{k \in C_{m,l}} \Delta_k^{(N)}$ is a sum with finite number of terms, which implies that computing whether a given Q_N satisfies the condition (6.3.2) can be done in a finite amount of time. Nevertheless, $\Delta_k^{(N)}$ cannot generally be efficiently computed in practice because it requires the computation of probability measure $\mathbb{P}_{\texttt{GBS}}$, which in turn requires the computation of the Hafnian function that is known to be #**P**-hard [Kru+19]. Therefore, we need to compute the error using the collection of Pattern Functions on the N Homodyne data $\{F_{|k\rangle}(q_j, \theta_j)\}_{j=1}^N$ and its sample mean, namely $\mathbb{P}_{\texttt{PG}}^{(N)}$. To do this, we need to use the results from [**Section 6.2**] that relates the standard deviation error of $\mathbb{P}_{\texttt{PG}}^{(N)}$ to those of $\{F_{|k\rangle(q_j,\theta_j)}$. For convenience, we restate them below,

$$\varepsilon_{N,k} = \sqrt{\frac{1}{N(N-1)} \sum_{j=1}^{N} \left(F_{|k\rangle}(q_j, \theta_j) - \mathbb{P}_{\texttt{PG}}^{(N)}(k)\right)^2} \tag{6.3.3}$$

and

$$\mathbb{P}\left\{Q_N : \Delta_k^{(N)} < v_k \varepsilon_{N,k}\right\} = \textbf{erf}\left(\frac{v_k}{\sqrt{2}}\right)., \tag{6.3.4}$$

where $v_k \in \mathbb{R}$ and $\textbf{erf}()$ is the Gauss error function.

These formulae will be the main tool for the PF GBS algorithm to calculate the sample standard deviation error ε_N, k of the approximated probability $\mathbb{P}_{\texttt{PG}}^{(N)}(k)$ and the probability of achieving an error within a certain value in terms of ε_N, k. However, note that we now need to consider these formulae for all $k \in C_{m,l}$ in constructing the bound. Therefore for shorthand, for $v = (v_1, \ldots, v_{|C_{m,l}|})$ we write the multivariate Gauss error function as

$$\textbf{erf}\left(\frac{v}{\sqrt{2}}\right) = \prod_{k \in C_{m,l}} \textbf{erf}\left(\frac{v_k}{\sqrt{2}}\right), \tag{6.3.5}$$

which we define rather loosely, depending on the dimension of the input. We are now ready to state the main result of this chapter on the criteria to bound the PF GBS approximation.

Theorem 6.3.3 (PF GBS Sample Error Bound)**.** *For error of the approximate PF GBS machine* $\Delta_k^{(N)} = |\mathbb{P}_{\texttt{PG}}^{(N)}(k) - \mathbb{P}_{\texttt{GBS}}(k)|$*, where* $k \in C_{m,l}$ *is a collision-free measured outcome and for* $\xi, \gamma \in (0,1)$*, we have*

$$\mathbb{P}\left\{(k, Q_N) : \Delta_k^{(N)} > \xi\right\} < \gamma$$

if we have sufficiently small $\{\varepsilon_{N,k}\}_{k\in C_{m,l}}$ *such that*

$$\sum_{k\in C_{m,l}} v_k \varepsilon_{N,k} \leq \xi |C_{m,l}| \left(\frac{\textbf{erf}\left(\frac{v}{\sqrt{2}}\right) - (1-\gamma)}{\textbf{erf}\left(\frac{v}{\sqrt{2}}\right)} \right),$$

for $\{v_k\}_{k\in C_{m,l}}$ *chosen such that* $\textbf{erf}\left(\frac{v}{\sqrt{2}}\right) > 1-\gamma$*.*

Now, we give a proof for this error bound criteria.

Proof of [Theorem 6.3.3]

First, we consider the bound (6.3.4) for all $k \in C_{m,l}$, which allow us to compute the probability of sampling a Q_N such that the Total Variation distance of $\mathbb{P}_{\texttt{PG}}^{(N)}$ from $\mathbb{P}_{\texttt{GBS}}$ being smaller than a value that is dependent on $\{v_k\}_{k\in C_{m,l}}$ and $\{\varepsilon_{N,k}\}_{k\in C_{m,l}}$. This can be achieved by simply taking the intersection of the event where $\Delta_k^{(N)} < v_k \varepsilon_{N,k}$ is satisfied for all k, giving us the probability

$$\mathbb{P}\left\{Q_N : \forall k, \Delta_k^{(N)} < v_k \varepsilon_{N,k}\right\} = \prod_{k\in C_{m,l}} \textbf{erf}\left(\frac{v_k}{\sqrt{2}}\right), \tag{6.3.6}$$

which is a product of (6.3.4) over all $k \in C_{m,l}$. This probability measure is computable and will be of much use later to determine the probability of obtaining a Q_N from the subset of $\mathcal{Q}_\sigma^N$ where condition (6.3.2) is satisfied.

Now, recalling the definition of the Total Variation distance, we have

$$\left|\left|\mathbb{P}_{\texttt{PG}}^{(N)} - \mathbb{P}_{\texttt{GBS}}\right|\right|_{\text{TV}} = \frac{1}{2} \sum_{k\in C_{m,l}} \Delta_k^{(N)}.$$

It is therefore clear from this definition that if we have the set of values $\{\varepsilon_{N,k}\}_{k=1}^m$ and $\{v_k\}_{k\in C_{m,l}}$ such that

$$\sum_{k\in C_{m,l}} v_k \varepsilon_{N,k} \leq 2\beta \tag{6.3.7}$$

we can obtain

$$K = \left\{Q_N : \forall k, \Delta_k^{(N)} < v_k \varepsilon_{N,k}\right\} \subseteq \left\{Q_N : \left|\left|\mathbb{P}_{\texttt{PG}}^{(N)} - \mathbb{P}_{\texttt{GBS}}\right|\right|_{\text{TV}} \leq \beta\right\}. \tag{6.3.8}$$

Which means that as long as the condition (6.3.7) is satisfied, we can be sure that any Q_N in the set K will satisfy the Total Variation bound (6.3.2), and the probability of getting a Q_N that satisfy it, that is, having a Q_N in K, is dictated by (6.3.6) where for all k, we can tweak the value v_k and have $\varepsilon_{N,k}$ arbitrarily small as N gets arbitrarily large. The set K is the union of the sets represented by the green and yellow regions in [**Figure 6.8**].

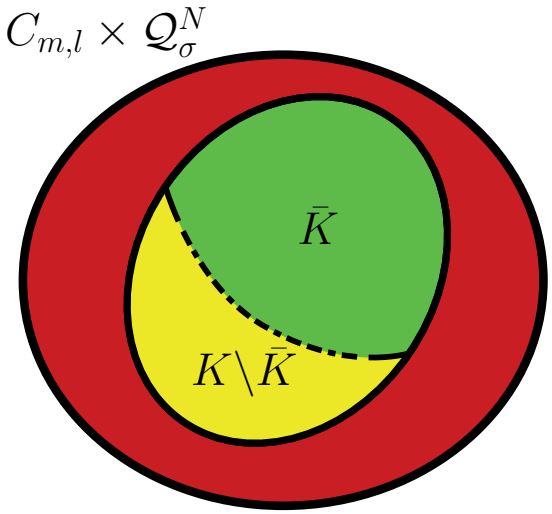

Figure 6.8: Product space of $C_{m,l} \times \mathcal{Q}_\sigma^N$. The region in red contains all N Homodyne data Q_N such that there exists a $k \in C_{m,l}$ such that $\Delta_k^{(N)} > v_k \varepsilon_{n,k}$. Its complement K as defined in (6.3.8) contains the set $\bar{K} = \left\{(k, Q_n) : (\Delta_k^{(N)} \leq v_k \varepsilon_{N,k}) \wedge (\Delta_k^{(N)} \leq \frac{2a\beta}{|C_{m,l}|})\right\}$, in words, the set of Q_N such that for all $k \in C_{m,l}$ both conditions $\Delta_k^{(N)} \leq v_k \varepsilon_{N,k}$ and $\Delta_k^{(N)} \leq \frac{2a\beta}{|C_{m,l}|}$ are satisfied. Thus, the difference between K and $\bar{K}$, namely $K \backslash \bar{K}$ is the set where we have Q_N such that $\forall k, \Delta_k^{(N)} < v_k \varepsilon_{N,k}$ is satisfied, but there is a k such that $\Delta_k^{(N)} > \frac{2a\beta}{|C_{m,l}|}$.

We now direct our attention to the set of no-collision output patterns $C_{m,l}$ of l photons in an m modes system. Now, given an arbitrary set of N Homodyne samples $Q_N \in \mathcal{Q}_\sigma^N$, let the Expectation of error $\Delta_k^{(N)}$ over all $k \in C_{m,l}$

$$\mathbb{E}_{C_{m,l}}\left(\Delta_k^{(N)}\right) = \frac{1}{|C_{m,l}|} \sum_{k \in C_{m,l}} \Delta_k^{(N)}.$$

Further let $a > 1$, then by Markov inequality we have the probability of k with error larger than $a\mathbb{E}_{C_{m,l}}\left(\Delta_k^{(N)}\right)$

$$\mathbb{P}\left\{k : \Delta_k^{(N)} > a\mathbb{E}_{C_{m,l}}\left(\Delta_k^{(N)}\right)\right\} < \frac{1}{a}.$$

To be able to bound the expectation, we therefore only consider the Homodyne samples Q_N that are in K as defined in (6.3.8) so that

$$\mathbb{E}_{C_{m,l}}\left(\Delta_k^{(N)}\right) \leq \frac{2\beta}{|C_{m,l}|}.$$

Thus we can write the probability of obtaining k that is in $\left\{k : \Delta_k^{(N)} > \frac{2a\beta}{|C_{m,l}|}\right\}$ with $Q_N \in K$ as

$$\begin{aligned}
&\mathbb{P}\left(\left\{k : \Delta_k^{(N)} > \frac{2a\beta}{|C_{m,l}|}\right\} \wedge \left\{Q_N : \forall k, \Delta_k^{(N)} \leq v_k \varepsilon_{N,k}\right\}\right) \\
&< \mathbb{P}\left(\left\{k : \Delta_k^{(N)} > a\mathbb{E}_{C_{m,l}}\left(\Delta_k^{(N)}\right)\right\} \wedge \left\{Q_N : \forall k, \Delta_k^{(N)} \leq v_k \varepsilon_{N,k}\right\}\right) \\
&= \int_K \mathbb{P}\left\{k : \Delta_k^{(N)} > a\mathbb{E}_{C_{m,l}}\left(\Delta_k^{(N)}\right)\right\} \mathbb{P}(Q_N)\, dQ_N \\
&< \frac{1}{a}\mathbb{P}\left\{Q_N : \forall k, \Delta_k^{(N)} \leq v_k \varepsilon_{N,k}\right\} \\
&= \frac{1}{a}\prod_{k \in C_{m,l}} \mathbf{erf}\left(\frac{v_k}{\sqrt{2}}\right).
\end{aligned} \tag{6.3.9}$$

In other words, this is the probability measure on the product space $C_{m,l} \times \mathcal{Q}_\sigma^N$ of the values of Q_N that are in the set K and values of k for those $Q_N \in K$ such that the error exceeds the lower bound $\Delta_k^{(N)}$. This set is represented by the yellow region in [**Figure 6.8**] and can be described as the set where there are some k such that the error of probability $\mathbb{P}_{\mathtt{PG}}^{(N)}(k)$ are "large".

However, we still need to consider all Q_N where the condition $\forall k, \Delta_k^{(N)} \leq v_k \varepsilon_{N,k}$ is not satisfied, namely those such that $\exists k, \Delta_k^{(N)} > v_k \varepsilon_{N,k}$, which is represented by the red region in [**Figure 6.8**]. In this region of Q_N, we cannot be sure if the bound on the Total Variation Distance (6.3.2) can be satisfied. As $K^c = \left\{Q_N : \exists k, \Delta_k^{(N)} > v_k \varepsilon_{N,k}\right\}$ is the complement of $K = \left\{Q_N : \forall k, \Delta_k^{(N)} \leq v_k \varepsilon_{N,k}\right\}$, we can write its probability as

$$\begin{aligned}
\mathbb{P}\left\{Q_N : \exists k, \Delta_k^{(N)} > v_k \varepsilon_{N,k}\right\} &= 1 - \mathbb{P}\left\{Q_N : \forall k, \Delta_k^{(N)} \leq v_k \varepsilon_{N,k}\right\} \\
&= 1 - \prod_{k \in C_{m,l}} \mathbf{erf}\left(\frac{v_k}{\sqrt{2}}\right).
\end{aligned}$$

Thus, we can use this to bound the probability measure on the product space $C_{m,l} \times \mathcal{Q}_\sigma^N$ of k with $\Delta_k^{(N)} > \frac{2a\beta}{|C_{m,l}|}$ for all Q_N in K^c, namely

$$\begin{aligned}
&\mathbb{P}\left(\left\{k : \Delta_k^{(N)} > \frac{2a\beta}{|C_{m,l}|}\right\} \wedge \left\{Q_N : \exists k, \Delta_k^{(N)} > v_k \varepsilon_{N,k}\right\}\right) \\
&= \int_{K^c} \mathbb{P}\left\{k : \Delta_k^{(N)} > \frac{2a\beta}{|C_{m,l}|}\right\} \mathbb{P}(Q_N)\, dQ_N \\
&< \mathbb{P}\left\{Q_N : \exists k, \Delta_k^{(N)} > v_k \varepsilon_{N,k}\right\} \\
&= 1 - \prod_{k \in C_{m,l}} \mathbf{erf}\left(\frac{v_k}{\sqrt{2}}\right).
\end{aligned} \tag{6.3.10}$$

We note that this bound is not tight as we are not able to put the upper bound $\frac{2a\beta}{|C_{m,l}|}$ on the expectation $\mathbb{E}_{C_{m,l}}\left(\Delta_k^{(N)}\right)$ as we did previously, this is because we are unable bound the Total

Variation Distance (6.3.2) for the Q_N in the complement of K. This can be interpreted as considering all Homodyne sample $Q_N \in K^c$ to be erroneous despite their corresponding k. Thus, we can bound the probability measure of k with large error $\Delta_k^{(N)}$, which is comprised of the red and yellow regions in [**Figure 6.8**], as follows

$$\begin{aligned}
&\mathbb{P}\left\{(k, Q_N) : \Delta_k^{(N)} > \frac{2a\beta}{|C_{m,l}|}\right\} \\
&= \int_{\mathcal{Q}_\sigma^N} \mathbb{P}\left\{k : \Delta_k^{(N)} > \frac{2a\beta}{|C_{m,l}|}\right\} \mathbb{P}(Q_N)\, dQ_N \\
&= \int_{K} \mathbb{P}\left\{k : \Delta_k^{(N)} > \frac{2a\beta}{|C_{m,l}|}\right\} \mathbb{P}(Q_N)\, dQ_N + \int_{K^c} \mathbb{P}\left\{k : \Delta_k^{(N)} > \frac{2a\beta}{|C_{m,l}|}\right\} \mathbb{P}(Q_N)\, dQ_N \\
&< \frac{1}{a} \prod_{k \in C_{m,l}} \mathbf{erf}\left(\frac{v_k}{\sqrt{2}}\right) + \left(1 - \prod_{k \in C_{m,l}} \mathbf{erf}\left(\frac{v_k}{\sqrt{2}}\right)\right)
\end{aligned} \tag{6.3.11}$$

where we have used (6.3.9) and (6.3.10) to obtain the inequality.

Now we look into

$$\mathbb{P}\left\{(k, Q_N) : \Delta_k^{(N)} > \xi\right\} < \gamma \tag{6.3.12}$$

for some $\xi, \gamma \in (0, 1)$. First, let $\mathbf{erf}\left(\frac{v}{\sqrt{2}}\right) = \prod_{k \in C_{m,l}} \mathbf{erf}\left(\frac{v_k}{\sqrt{2}}\right)$ for short, and

$$\xi = \frac{2a\beta}{|C_{m,l}|} \quad \text{and} \quad \gamma = \frac{1}{a}\mathbf{erf}\left(\frac{v}{\sqrt{2}}\right) + \left(1 - \mathbf{erf}\left(\frac{v}{\sqrt{2}}\right)\right)$$

as in (6.3.11). Noting that $a > 1$ must be true and that

$$a = \frac{\mathbf{erf}\left(\frac{v}{\sqrt{2}}\right)}{\mathbf{erf}\left(\frac{v}{\sqrt{2}}\right) - (1 - \gamma)},$$

we therefore must have sufficiently large $\{v_k\}_{k \in C_{m,l}}$ such that $\mathbf{erf}\left(\frac{v}{\sqrt{2}}\right) > 1 - \gamma$ to make the denominator greater than 0. This, therefore imposes a condition for the value $\{v_k\}_{k \in C_{m,l}}$ for a desired upper bound γ on the probability of error $\Delta_k^{(N)}$ in (6.3.12). Further, substituting a into $\xi = \frac{2a\beta}{|C_{m,l}|}$ gets us

$$\beta = \frac{\xi |C_{m,l}|}{2} \frac{\mathbf{erf}\left(\frac{v}{\sqrt{2}}\right) - (1 - \gamma)}{\mathbf{erf}\left(\frac{v}{\sqrt{2}}\right)}$$

which we use to determine the condition of how small the values $\{\varepsilon_{N,k}\}_{k \in C_{m,l}}$ need to be to obtain $\mathbb{P}\left\{k : \Delta_k^{(N)} > \xi\right\} < \gamma$, namely

$$\sum_{k \in C_{m,l}} v_k \varepsilon_{N,k} \leq 2\beta = \xi |C_{m,l}| \left(\frac{\mathbf{erf}\left(\frac{v}{\sqrt{2}}\right) - (1 - \gamma)}{\mathbf{erf}\left(\frac{v}{\sqrt{2}}\right)}\right). \tag{6.3.13}$$

Thus, the proof for [**Theorem 6.3.3**] is concluded. □

Note that the smaller γ is, the larger $\{v_k\}_{k\in C_{m,l}}$ needs to be to satisfy $\mathbf{erf}\left(\frac{v}{\sqrt{2}}\right) > 1-\gamma$. In turn, as $\{v_k\}_{k\in C_{m,l}}$ gets larger, we need smaller value of $\{\varepsilon_{N,k}\}_{k\in C_{m,l}}$ to satisfy the condition (6.3.13). Because for each k, the value $\varepsilon_{N,k}$ decreases at the rate of $\mathcal{O}(N^{-\frac{1}{2}})$, increasing the accuracy of the approximation will require substantially more Homodyne data as N gets larger.

Lastly, we come back to the problem of having $\mathbb{P}^{(N)}_{\mathtt{PG}}(k) < 0$ for some k given a sampled homodyne data Q_N. Imposing this condition for the set of Q_N such that $\mathbb{P}^{(N)}_{\mathtt{PG}}(k)$ is non-negative for all $k \in C_{m,l}$ will further restrict the set of "valid" homodyne data Q_N, which we represent above as the set

$$K = \left\{Q_N : \forall k, \Delta^{(N)}_k < v_k\varepsilon_{N,k}\right\}.$$

Thus requiring the non-negativity condition will force us to take its subset by taking its intersection with another set that contains only all Q_N such that $\mathbb{P}^{(N)}_{\mathtt{PG}}(k)$ is non-negative for all $k \in C_{m,l}$, namely

$$W = \left\{Q_N : \forall k, \Delta^{(N)}_k < v_k\varepsilon_{N,k}\right\} \wedge \left\{Q_N : \forall k, \mathbb{P}^{(N)}_{\mathtt{PG}}(k) \geq 0\right\}.$$

Having W as the subset of K, the complement of W will therefore be a set with at least the same measure as the complement of K, and thus will cause the bound in [**Theorem 6.3.3**] to no longer hold. Restricting the valid set of homodyne samples from K to W can be understood as "increasing" the size of the set of undesired homodyne samples Q_N. However, this issue can be practically overcame by keep taking more homodyne data after satisfying the bounds in [**Theorem 6.3.3**] and "naively" checking at every step whether $\mathbb{P}^{(N)}_{\mathtt{PG}}(k) \geq 0$ for all $k \in C_{m,l}$ until it is satisfied. After having $\mathbb{P}^{(N)}_{\mathtt{PG}}(k) \geq 0$ for all $k \in C_{m,l}$, we can then normalize the probabilities by taking

$$\tilde{\mathbb{P}}^{(N)}_{\mathtt{PG}}(k) = \frac{\mathbb{P}^{(N)}_{\mathtt{PG}}(k)}{\sum_{k\in C_{m,l}} \mathbb{P}^{(N)}_{\mathtt{PG}}(k)}.$$

Nevertheless, this may incur at most an additional deviation of $|\tilde{\mathbb{P}}^{(N)}_{\mathtt{PG}}(k) - \mathbb{P}^{(N)}_{\mathtt{PG}}(k)|$ from the true probability $\mathbb{P}_{\mathtt{GBS}}(k)$ for each k, further invalidating the bounds in [**Theorem 6.3.3**].

We note that the absence of a tighter bound for the set of undesired Q_N (i.e: complement of W), including those that cause $\mathbb{P}^{(N)}_{\mathtt{PG}}(k) < 0$ for some k is rather a substantial shortcoming of this approximation method. The issue of having an unnormalized probability distribution caused by the unnormalized values of pattern function $F_{|k\rangle}$ over $k \in C_{m,l}$ is also a rather significant shortcoming that prevents the use of this algorithm with the assurance of the error bound shown in [**Theorem 6.3.3**]. Further analysis to obtain this bound and to normalize the PF GBS probability distribution is therefore, a good starting point for any future plans of further improving the PF GBS algorithm.

6.3.2 PF GBS probability distribution approximation simulation with error bound

Despite the shortcoming of having some negative values in the approximated probability distribution $\mathbb{P}_{\texttt{PG}}$ which leaves it unnormalized, we will display computer simulation results for the PF GBS approximation for the complete probability distribution $\mathbb{P}_{\texttt{GBS}}$ for collision-free states $k \in C_{m,l}$. Here, we will focus on the relatively simple case with $m = 6$ modes and $l = 2$ photons at the output with 2 SMSS inputs with $r = 0.25$. This simple case is deliberately chosen to ensure feasibility in computing the Hafnian functions to calculate the true probabilities. The simulations were ran until the conditions specified in [**Theorem 6.3.3**] are satisfied. We restate these conditions below:

> For $\xi, \gamma \in (0,1)$, if we have standard deviations $\{\varepsilon_{N,k}\}_{k\in C_{6,2}}$ such that
>
> $$\sum_{k\in C_{m,l}} v_k \varepsilon_{N,k} \leq \xi |C_{m,l}| \left(\frac{\mathbf{erf}\left(\frac{v}{\sqrt{2}}\right) - (1-\gamma)}{\mathbf{erf}\left(\frac{v}{\sqrt{2}}\right)} \right), \tag{6.3.14}$$
>
> then we have
>
> $$\mathbb{P}\left\{(k, Q_N) : \Delta_k^{(N)} > \xi\right\} < \gamma, \tag{6.3.15}$$
>
> for $\{v_k\}_{k\in C_{6,2}}$ such that $\prod_{k\in C_{6,2}} \mathbf{erf}\left(\frac{v_k}{\sqrt{2}}\right) > 1-\gamma$.

We consider the approximation errors bounds $(\xi, \gamma) \in \{(0.05, 0.4), (0.04, 0.3), (0.03, 0.3)\}$. The simulation result for the first pair of bounds is shown in [**Figure 6.9**], the second pair in [**Figure 6.10**], and the last pair in [**Figure 6.11**]. We will refer to each simulation by their (ξ, γ) pair values.

It is clear that in the simulation $(0.05, 0.4)$ shown in [**Figure 6.9**], none of the error $\Delta_k^{(N)}$ for any $k \in C_{6,2}$ represented by the red bars exceeds $\xi = 0.05$, despite the criteria only guaranteeing that at least 0.6 of the collision free sets $C_{6,2}$ have an error below $\xi = 0.05$. This is because the nature of the criteria [**Theorem 6.3.3**] that is not a tight bound, which means that inequality (6.3.14) with $\xi = 0.05$ and $\gamma = 0.4$ was satisfied way before $\{\varepsilon_{N,k}\}_{k\in C_{6,2}}$ satisfy the inequality (6.3.15). Similar case occurs in the other two simulation results $(0.04, 0.3)$ [**Figure 6.10**] and $(0.03, 0.3)$ [**Figure 6.11**] where none of the errors exceeds their respective bound ξ.

However, as the number of Homodyne data N are larger for $(0.04, 0.3)$ and $(0.03, 0.3)$, the total of errors from all $k \in C_{6,2}$ that are represented by the red bars in the plots are smaller. Graphically, the total area of the red bars for $(0.03, 0.3)$ is smaller than $(0.04, 0.3)$, and the total area of the red bars for $(0.04, 0.3)$ is smaller than $(0.05, 0.4)$. This is due to the standard deviation $\varepsilon_{N,k}$ of $\mathbb{P}_{\texttt{PG}}^{(N)}(k)$ for each $k \in C_{6,2}$ are closer to 0 as $N \to \infty$. Thus we should note that

- for $(0.03, 0.3)$ we have $N = 4875000$,
- for $(0.04, 0.3)$, we have $N = 2750000$, and

- for $(0.05, 0.4)$, we have $N = 895000$.

As a consequence of the Central Limit Theorem [**Lemma 6.2.3**] which shows that for each k, the distribution of $\mathbb{P}_{\texttt{PG}}^{(N)}(k)$ is Gaussian with standard deviation $\varepsilon_{N,k}$, we know that the probability of the error $\Delta_k^{(N)}$ exceeding $\varepsilon_{N,k}$ is bounded and stays constant as N increases. Therefore the total error $\sum_{k \in C_{6,2}} \Delta_k^{(N)}$ is also smaller for larger N. This can be seen from the Total Variation distance (as defined in the beginning of [**Section 6.3.1**]) of the resulting approximated distribution $\mathbb{P}_{\texttt{PG}}^{(N)}$ from the true distribution $\mathbb{P}_{\texttt{GBS}}$ for each simulation, where we have

- $\left|\left|\mathbb{P}_{\texttt{PG}}^{(N)} - \mathbb{P}_{\texttt{GBS}}\right|\right|_{\text{TV}} \leq \beta \approx 0.1285$ for $(0.03, 0.3)$,
- $\left|\left|\mathbb{P}_{\texttt{PG}}^{(N)} - \mathbb{P}_{\texttt{GBS}}\right|\right|_{\text{TV}} \leq \beta \approx 0.1714$ for $(0.04, 0.3)$, and
- $\left|\left|\mathbb{P}_{\texttt{PG}}^{(N)} - \mathbb{P}_{\texttt{GBS}}\right|\right|_{\text{TV}} \leq \beta \approx 0.2818$ for $(0.05, 0.4)$.

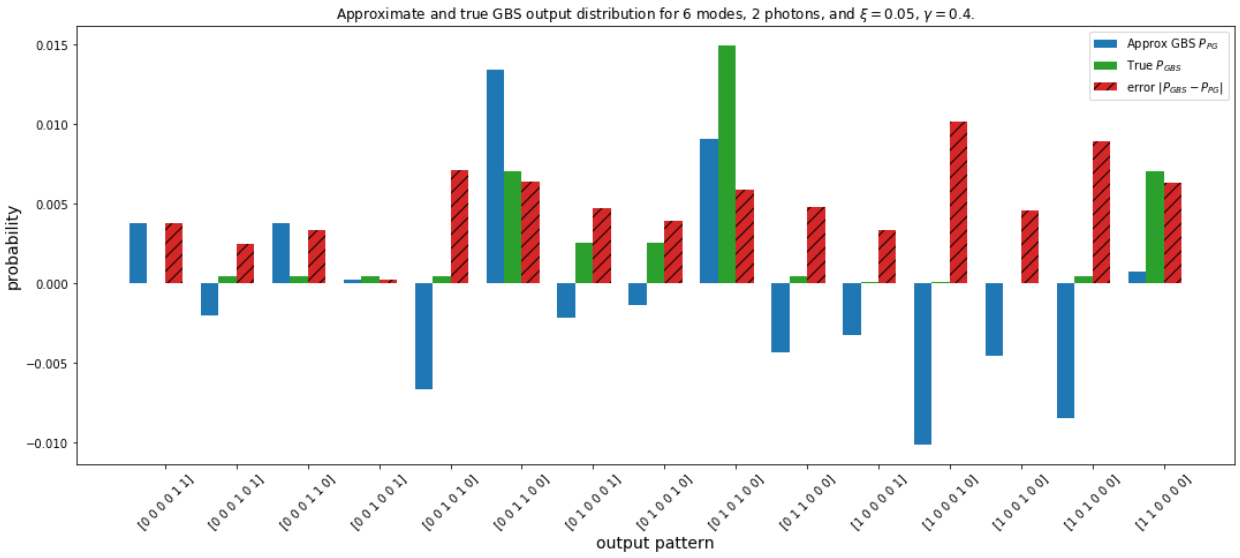

Figure 6.9: Bar plot of PF GBS approximation $\mathbb{P}_{\texttt{PG}}^{(N)}(k)$, true probability $\mathbb{P}_{\texttt{GBS}}(k)$, and error $|\mathbb{P}_{\texttt{PG}}^{(N)}(k) - \mathbb{P}_{\texttt{GBS}}(k)|$ for all collision free outputs of two photons from a 6 mode interferometer $k \in C_{6,2}$. We use 2 SMSS inputs with $r = 0.25$. Homodyne samples are taken until the condition in [**Theorem 6.3.3**] is satisfied for $\xi = 0.05$ and $\gamma = 0.4$. For each $k \in C_{6,2}$ across the x axis, we have three bars: a blue bar representing the approximation $\mathbb{P}_{\texttt{PG}}^{(N)}(k)$, a green bar representing the true probability $\mathbb{P}_{\texttt{GBS}}(k)$, and a striped red bar representing error $\Delta_k^{(N)} = |\mathbb{P}_{\texttt{PG}}^{(N)}(k) - \mathbb{P}_{\texttt{GBS}}(k)|$. This condition was satisfied with $N = 895000$ and $||\mathbb{P}_{\texttt{PG}} - \mathbb{P}_{\texttt{GBS}}||_{\text{TV}} \leq \beta \approx 0.2818$.

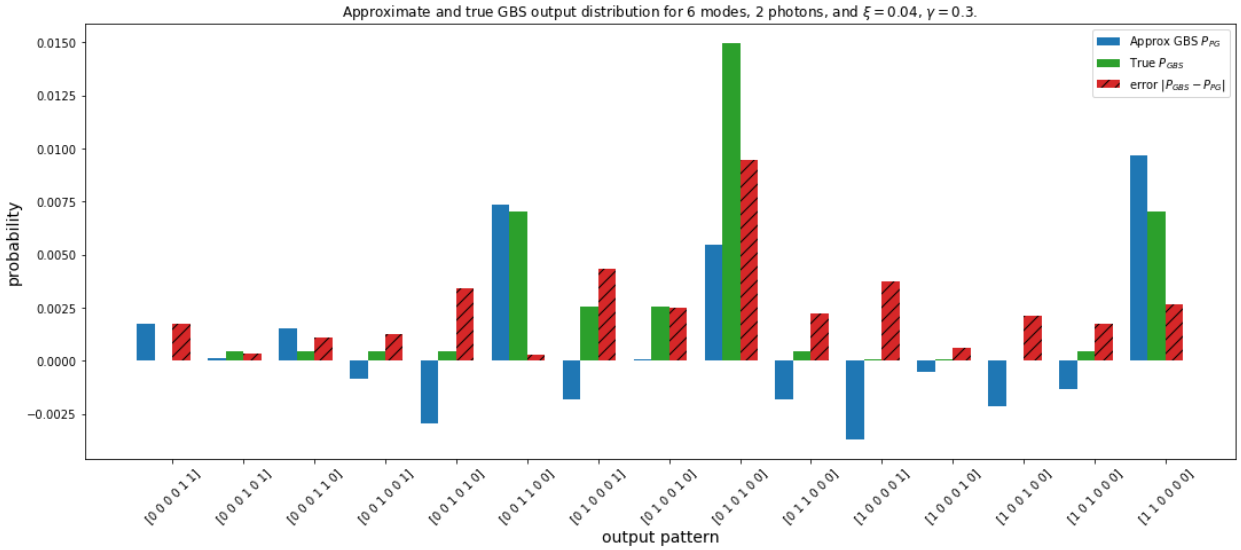

Figure 6.10: Bar plot of PF GBS approximation $\mathbb{P}_{\mathtt{PG}}^{(N)}(k)$, true probability $\mathbb{P}_{\mathtt{GBS}}(k)$, and error $|\mathbb{P}_{\mathtt{PG}}^{(N)}(k) - \mathbb{P}_{\mathtt{GBS}}(k)|$ for all collision free outputs of two photons from a 6 mode interferometer $k \in C_{6,2}$. We use 2 SMSS inputs with $r = 0.25$. Homodyne samples are taken until the condition in [**Theorem 6.3.3**] is satisfied for $\xi = 0.04$ and $\gamma = 0.3$. For each $k \in C_{6,2}$ across the x axis, we have three bars: a blue bar representing the approximation $\mathbb{P}_{\mathtt{PG}}^{(N)}(k)$, a green bar representing the true probability $\mathbb{P}_{\mathtt{GBS}}(k)$, and a striped red bar representing error $\Delta_k^{(N)} = |\mathbb{P}_{\mathtt{PG}}^{(N)}(k) - \mathbb{P}_{\mathtt{GBS}}(k)|$.This condition was satisfied with $N = 2750000$ and $||\mathbb{P}_{\mathtt{PG}} - \mathbb{P}_{\mathtt{GBS}}||_{\mathrm{TV}} \leq \beta \approx 0.1714$.

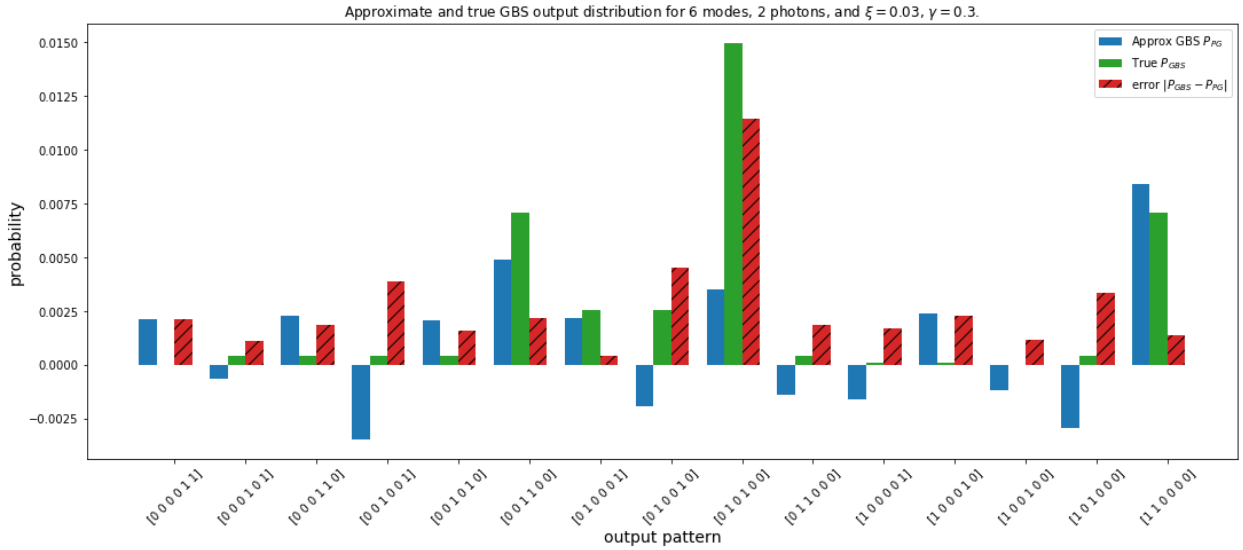

Figure 6.11: Bar plot of PF GBS approximation $\mathbb{P}_{\mathtt{PG}}^{(N)}(k)$, true probability $\mathbb{P}_{\mathtt{GBS}}(k)$, and error $|\mathbb{P}_{\mathtt{PG}}^{(N)}(k) - \mathbb{P}_{\mathtt{GBS}}(k)|$ for all collision free outputs of two photons from a 6 mode interferometer $k \in C_{6,2}$. We use 2 SMSS inputs with $r = 0.25$. Homodyne samples are taken until the condition in [**Theorem 6.3.3**] is satisfied for $\xi = 0.03$ and $\gamma = 0.3$. For each $k \in C_{6,2}$ across the x axis, we have three bars: a blue bar representing the approximation $\mathbb{P}_{\mathtt{PG}}^{(N)}(k)$, a green bar representing the true probability $\mathbb{P}_{\mathtt{GBS}}(k)$, and a striped red bar representing error $\Delta_k^{(N)} = |\mathbb{P}_{\mathtt{PG}}^{(N)}(k) - \mathbb{P}_{\mathtt{GBS}}(k)|$.This condition was satisfied with $N = 4875000$ and $||\mathbb{P}_{\mathtt{PG}} - \mathbb{P}_{\mathtt{GBS}}||_{\mathrm{TV}} \leq \beta \approx 0.1285$.

Chapter 7

Quasiprobability Distribution Sampling and Noisy GBS

The negativity of a quasi-proability distribution (QPD) representation of a state[1] as a characterization of non-classicality has been a nearly century-old subject of interest, dating back to the inception of the proto-conceptualization of a QPD by Eugene Wigner [Wig32] which now known as the celebrated Wigner function (or Wigner QPD). Another attempt in probing the quantum-classical boundary by representing a general state ρ in terms of coherent states $|\alpha\rangle$ by Glauber [Gla63] also gave rise to a QPD representation which is later known as the P function (or P QPD) that can be understood as the measure in representing a state ρ as an integral of Coherent states $|\alpha\rangle$, thus any state in which corresponding P function exist and is non-negative everywhere is considered as "classical". A generalization that include these QPDs functions (including another QPD introduced by Husimi [Hus40]) by Cahill and Glauber [CG69] later shows that these QPDs can be related to one another by means of convolution, consequently giving a continuous family of QPDs that may result in a range of non-classicality measure of a state.

The importance of Wigner QPD towards the characterization of distinction between quantum and classical physical processes is later emphasized through the Hudson's Theorem [Hud74] which states that for every pure state $|\psi\rangle$, its Wigner function representation W_ψ is non-negative everywhere if and only if $|\psi\rangle$ is a Gaussian state. This fact in a sense supports the notion that the more experimentally accessible Gaussian states can be considered as "classical". However, in a later discussion we will see that considering the more general s parametrized QPD defined in [**Definition 2.2.16**] and the overlap formula in [**Proposition 2.2.18**] between QPD representations of two operations can make possible an efficient classical computational simulation of a quantum system.

To put the QPD characterization of non-classicality into a wider perspective, we note that it may be distinct from other characterization of non-classicality, such as quantum discord [FP12], and may have certain problems, such as mentioned in [Wun04]. However, we will be mainly consider QPD negativity as a characterization for the existence of efficient classical simulation on sampling from a measurement of a state, which on its own, may also be interpreted as a characterization of non-classicality of a state (obtained from a certain process

[1] As discussed in [**Section 2.2.4**].

or computation). Thus, one can draw a certain correspondence between the computational complexity characterization and QPD negativity characterization for the non-classicality of a state.

This other quantum-classical characterization notion in computation that suggest the complexity of quantum computation for a classical computer has also been increasingly shown using the QPDs, both in its s parametrized general form and in the more widely studied Wigner function form. This QPD characteization of computational complexity materialize in groudbreaking results such as [ME12], [Vei+13], and [RRC16] that brought the QPD non-classicality characterization closer to computational complexity characterization. As the former two results focuses on analyzing the non-negativity of the Wigner function as a sufficient condition for an efficient classical simulation of a quantum system, the latter result took this attempt further by generalizing it to a more general s parametrized QPD.

In addition, the result by Rahimi-Keshari, Ralph, and Caves [RRC16] (we refer to as RRC) further provides an analysis on this QPD negativity characterization on the Boson Sampling model, which we are dearly familiar with by showing the possibility of efficient classical simulation of Boson Sampling given a certain amount of noise. More precisely, this result uses the overlap formula presented in [**Proposition 2.2.18**], which we restate below

$$\mathbf{Tr}(\rho F) = 2\pi \int\int W_{\rho}^{(s)}(q,p) W_{F}^{(-s)}(q,p)\, dq\, dp.$$

The result by RRC replaces the general operator F with a POVM, therefore expressing the measurement outcome probability $\mathbb{P}(n) = \mathbf{Tr}(\rho P_n)$. In their result, the overlap formula however was presented in a different form where a term corresponding to the linear interferometer is also present and with a POVM that considers noise sources from detector efficiency η and dark count d_c. Then, it was shown by RRC that if there exists an ordering parameter s that satisfies an inequality in terms of η and d_c, then all of the QPDs in the overlap integral are non-negative everywhere, which in turn implies that an efficient classical algorithm that simulates the system exists.

This noise bound can also be seen as a necessary (but not sufficient) condition in order for one to carry out a Boson Sampling experiment that cannot be efficiently simulated using a classical computer. For a Boson Sampling experiment $\mathcal{B}$ and some number $c \in [0,1]$, we can write this more formally as

$$\begin{aligned}&(\mathcal{B} \text{ incurs noise } \eta \geq c) \Rightarrow (\exists \text{ classical algorithm } \mathcal{A} \text{ that simulates } \mathcal{B} \text{ efficiently.})\\ &\equiv (\nexists \text{ classical algorithm } \mathcal{A} \text{ that simulates } \mathcal{B} \text{ efficiently.}) \Rightarrow (\mathcal{B} \text{ incurs noise } \eta < c).\end{aligned}$$

The result by RRC is therefore complements other results that shows a sufficient condition on the amount of noise for a Boson Sampling experiment to be hard for a classical computer to perform. This can be more formally written as

$$\begin{aligned}&(\mathcal{B} \text{ incurs noise } \eta < c) \Rightarrow (\nexists \text{ classical algorithm } \mathcal{A} \text{ that simulates } \mathcal{B} \text{ efficiently})\\ &\equiv (\exists \text{ classical algorithm } \mathcal{A} \text{ that simulates } \mathcal{B} \text{ efficiently}) \Rightarrow (\mathcal{B} \text{ incurs noise } \eta \geq c).\end{aligned}$$

A list and discussion on results of both necessary and sufficient conditions to conduct a classically hard Boson Sampling experiment can be found in [Shc19], which also itself shows

a sufficient condition for the noise on a Boson Sampling experiment building on a previous result by Aaronson and Brod [AB16].

We now note that this goal of analyzing the "boundary" on the noise spectrum between classical and quantum computation in terms of whether a system can be efficiently simulated classically within the GBS paradigm is the primary subject of this chapter. This both physically and computationally motivated goal will be approached from the phase-space quantum mechanics formulation of QPDs, particularly from the results in [Qi+20] and [RRC16], and also from a more conventional analysis in the Hilbert space quantum mechanics representation closer to the results from [AB16] and [Shc19]. A particular focus on noise models will be put on the noise incurring in the photodetector, the photodetector efficiency and dark count. Lastly, we will present a possible approach in formulating a sufficient noise condition to have a classically hard GBS using a scheme called the CV sampling and photon-added thermal states, which is motivated by the QPD negativity non-classicality characterization analysis.

Throughout this chapter, we will conduct discussions and investigations on the classical-quantum computational complexity boundary by using the GBS model. We will first discuss the aforementioned results by RRC [RRC16] and Qi, et.al [Qi+20] in [**Section 7.1**] about the role of QPD and their non-negativity in efficient classical simulation of a quantum computational model, especially when there is noise in the system. Following this path further in [**Section 7.2**], we will discuss the lossy Boson Sampling result in [AB16] that shows if a certain number of loss incurs in a Boson Sampling experiment, then it is still classically hard to simulate. We then propose a GBS model that considers noise in the photodetection measurement, which outcome distribution can be formulated similarly to the lossy Boson Sampling of [AB16].

Lastly in [**Section 7.3**], we propose a different linear optical model that is closely related to the previously discussed noisy GBS model we call the Photon-added Thermal State Continuous-Variable sampling (PATS CV sampling). This model notably uses the Photon-added Thermal state (PATS) as an input, which is prepared by feeding a single photon (using the creation operator $\widehat{a}^\dagger$ as in [**Definition 2.2.1**]) to a thermal state [**Definition 2.2.7**]. The consideration on the use of PATS is based on results indicating that a thermal state is classical in both physical and computational sense: (1) first, a Boson-Sampling-type computational model that feeds thermal state inputs instead of single-photon inputs can be efficiently simulated in $\mathbf{BPP}^{\mathbf{NP}}$ [RLR15], and (2) secondly, the P-function of the thermal state is Gaussian, thus non-negative everywhere. Adding a non-classical single-photon to a "classical" thermal state can therefore be intuitively understood as moving it towards the non-classical spectrum. Further imposing noise on it, then move it back towards the classical spectrum. This is shown in [Vei+13] that although the Wigner function of a noiseless PATS has negative values, it can be non-negative everywhere after going through a noisy channel[2]. Although we come short of providing proofs for a sufficient noise condition for both models that are proposed in this chapter, it is hoped that these formulations can provide a clearer path for further investigation of the classical-quantum computational complexity boundary

[2]A noisy channel is modeled by operating a beamsplitter to a subject state ρ with a vacuum $|0\rangle$ on the other input of the beamsplitter, then tracing out an output of the beamsplitter. The noise amount is determined by the transmissivity τ of the beamsplitter. More precisely, $\mathbf{Tr}_b\left(B(\tau)\rho\otimes|0\rangle_{bb}\langle 0|B^\dagger(\tau)\right)$.

using the GBS model, or Continuous Variable (CV) linear-optics more generally.

Before we discuss the QPD negativity analysis to characterize hardness of classical simulation, we will first in a more general sense discuss about different types of sampling algorithms.

Two Types of Sampling Simulation

The Pattern Function Gaussian Boson Sampling (PF GBS) algorithm we presented in [**Chapter 6**] that simulates an approximate GBS performs an explicit construction of the probability distribution $\mathbb{P}_{\texttt{PG}}$ that approximates the exact GBS probability distribution $\mathbb{P}_{\texttt{GBS}}$ using the method of Homodyne Tomography that approximates the probability $\mathbb{P}_{\texttt{PG}}(k)$ for all collision-free measurement output $k \in C_{m,l}$. This explicit construction, however is not necessary in order to perform sampling from a distribution. In the case of GBS, where the probability $\mathbb{P}_{\texttt{GBS}}(k)$ can be expressed in terms of the hafnian function that is #**P**-hard to estimate, this explicit construction of the approximated probability distribution can be seen as an approach that should be avoided. Therefore, the results from [ME12], [Vei+13], and [RRC16] do present a method such that one does not need to explicitly computationally construct the probability distribution, but instead sample from the QPDs in the overlap integral of the outcome probability. We now define terms to refer to these two methods.

1. **Strong Simulation**: An algorihtm that computes the (approximated) probability of measurement outcomes $\mathbb{P}(n)$ for all $n \in C_{m,l}$, therefore reconstructing the probability distribution $\mathbb{P}$. This is the type of simulation that is employed in the PF GBS algorithm proposed in [**Chapter 6**] which approximates the probability distribution of measurement outcome from a GBS setup $\mathbb{P}_{\texttt{GBS}}$.

2. **Weak Simulation**: Obtain a collision-free measurement outcome $n \in C_{m,l}$ drawn from (approximately) the same probability distribution as $\mathbb{P}(n)$ using a classical probabilistic algorithm without needing to explicitly compute $\mathbb{P}(n)$ for all $n \in C_{m,l}$. Employing the QPD formulation of phase space quantum mechanics as discussed in [ME12, p.3], [Vei+13], and [RRC16], this is done by taking samples from the QPDs in the overlap integral of the output probability.

We will now discuss the weak simulation of GBS which will employ the method of sampling from the QPDs in the overlap integral of its outcome probability $\mathbb{P}_{\texttt{GBS}}$ and analyze the noise criteria for a necessary condition to perform a classically hard GBS experiment.

7.1 Weak Classical Sampling Simulation of GBS

We first recall the motivation of Gaussian Boson Sampling (GBS) model that is centered around the implementational feasibility of demonstrating the hardness of sampling from a linear optical system, it is only natural to expand the Boson Sampling result from RRC to the GBS regime although it may pose a challenge given the additional squeezing parameter r that is likely to come into play in determining possible efficient classical simulation. The attempt on analyzing the possible classical simulation of GBS due to the non-negativity of its corresponding QPD has been done by Qi, et.al in [Qi+20] by directly building on the

result by RRC. This was done thoroughly by considering the noise in different stages of the computation, first the noisy channel before the interferometer η_L, then the loss (quantum efficiency) η_c and dark count d_c of the photon-counting photodetectors. Qi et.al further derived an inequality in terms of all three noise models, squeezing parameter r, and number of input squeezed states l where as a consequence of the inequality being satisfied, an efficient classical algorithm can simulate this noisy GBS setup using the scheme proposed by RRC using the QPDs. We will now start the discussion by presenting the weak sampling scheme proposed in [RRC16], which laid the foundations for the result from Qi, et.al.

Sufficient Conditions for Efficient Classical weak sampling of GBS

We will first state an unsurprising, yet interesting result stating that the QPD representations in the overlap of the outcome probability of a noiseless GBS cannot be efficiently sampled classically. But first we need to define an important element that we need in expressing the overlap of $\mathbb{P}_{\texttt{GBS}}$.

Definition 7.1.1 (Transfer Function)**.** For $s, t \in [-1, 1]$ where for $s \leq t$, and $\alpha, \beta \in \mathbb{C}^m$, we have *Transfer function* for linear interferometer L

$$T_L^{(s,t)}(\beta|\alpha) = \frac{2^m}{\pi^m \mathbf{Det}(\sigma)} \mathbf{exp}(-2(\beta - L\alpha)\sigma^{-1}(\beta - L\alpha)^\dagger)$$

where $\sigma = I - L^\dagger L - sI - tL^\dagger L$. ◀

Now we first restate the measurement outcome probability of GBS in terms of the overlap integral of product of QPD as in [RRC16]

$$\mathbb{P}(n) = \int_{\mathbb{C}^m} \int_{\mathbb{C}^m} W_{|n\rangle\langle n|}^{(-s)}(\beta)\, T_L^{(s,t)}(\beta|\alpha)\, W_{\rho_{in}}^{(t)}(\alpha)\, d^{2m}\alpha\, d^{2m}\beta \tag{7.1.1}$$

for $s, t \in [-1, 1]$, where for $s \leq t$, we have transfer function for linear interferometer L. It is from these QPDs inside the integral that we intend to sample from, if each of them are non-negative and can be sampled from efficiently using a classical computer. Further, we note that this expression was derived from the more common form of the overlap as defined in [**Proposition 2.2.18**]. This derivation is left out here, although interested reader may refer to [RRC16]. Now we state our result on the impossibility of efficiently sampling from all of these QPDs using a classical computer.

Proposition 7.1.2. *For a GBS setup with unitary linear interferometer L and squeezing $r > 0$ where overlap of probability of outcome measurement $n \in C_{m,l}$*

$$\mathbb{P}(n) = \int_{\mathbb{C}^m} \int_{\mathbb{C}^m} W_{|n\rangle\langle n|}^{(-s)}(\beta)\, T_L^{(s,t)}(\beta|\alpha)\, W_{\rho_{in}}^{(t)}(\alpha)\, d^{2m}\alpha\, d^{2m}\beta,$$

there is no $s, t \in [-1, 1]$ such that $s \leq t$ where the QPDs $W_{|n\rangle\langle n|}^{(-s)}$, $T_L^{(s,t)}$, and $W_{\rho_{in}}^{(s)}$ can all be classically sampled efficiently.

Proof. Further considering an ideal unitary L (that incurs no loss) and $s = t$, we have $\sigma = 0$, and thus

$$T_L^{(s,t)}(\beta|\alpha) = \delta^{2m}(\beta - L\alpha).$$

Using the property of δ, where $\int f(x)\delta(x-y)\,dx = f(y)$, we can rewrite $\mathbb{P}(n)$ as

$$\mathbb{P}(n) = \int_{\mathbb{C}^m} W^{(-s)}_{|n\rangle\langle n|}(L\alpha)\, W^{(s)}_{\rho_{in}}(\alpha)\, d^{2m}\alpha.$$

We further note that $s = t$ is the optimal condition in order to get the most well behaved $W^{(-s)}$ and $W^{(t)}$, because $W^{(-s)}$ behaves better as $s \to 1$ and $W^{(t)}$ is more ill-behaving as $t \to 1$ (i.e: $W^{(-s)}$ gets more like the Q function while $W^{(t)}$ gets more like the P function, thus for $s \leq t$, if $W^{(t)}_\rho$ is non-negative everywhere, then it must be the case that $W^{(s)}_\rho$ is non-negative everywhere). Particularly for $W^{(-s)}_{|n\rangle\langle n|}$ as in Cahill and Glauber [CG69, eqn.(7.24) and (7.27)], we have

$$W^{(-s)}_{|n\rangle\langle n|}(\alpha) = \frac{2}{1-s}\left(\frac{1-s}{-s-1}\right)^n \mathbf{exp}\left(-\frac{2|\alpha|^2}{1+s}\right) L_n\left(\frac{4|\alpha|^2}{1-s^2}\right)$$

for generalized Laguerre Polynomial L_n, and for $s = 1$,

$$W^{(-1)}_{|n\rangle\langle n|}(\alpha) = Q_{|n\rangle\langle n|}(\alpha) = \frac{1}{n!}|\alpha|^{2n}\mathbf{exp}(-|\alpha|^2).$$

This implies that for $s < 1$, there are values of n and α that make $W^{(-s)}_{|n\rangle\langle n|}(\alpha) < 0$, for instance when we have $n = 1$ and $|\alpha| < \sqrt{(1-s^2)/4}$. Thus for $s < 1$, we cannot sample $W^{(-1)}_{|n\rangle\langle n|}$ classically due to its negative values at some points.

As a consequence, we need to have the QPD for ρ_{in} in the form of P function because we have $W^{(1)}_{\rho_{in}}$. However, this is known to be more singular than the δ function, as explicit formulation can be seen in [Kie+11, eqn.(3)]. Thus for squeezed state ρ_{in} with squeezing $r > 0$, sampling from $W^{(s)}_{\rho_{in}}$ is not possible for $s = 1$ as we require, rendering efficient sampling from the QPDs impossible. □

This result is however a more special case of the result presented in [Qi+20] where they consider the cases where there are noise in the GBS setup. More explicitly, they derived a condition that expresses a sufficient amount of noise in a GBS circuit for its quasi-probability representations to be non-negative, implying a possible efficient classical simulation. We restate their main result below.

Theorem 7.1.3 (Sufficient criteria for GBS efficient classical simulation [Qi+20])**.** *For a GBS setup with k squeezed state inputs, each with squeezing parameter r, photodetector dark count rate d_c, and overall transmission rate η, if the following inequality is satisfied*

$$\operatorname{sech}\left(\frac{1}{2}\Theta\left(\ln\left(\frac{1-d_c}{\eta e^{-2r}+1-\eta}\right)\right)\right) > \mathbf{exp}(-\frac{\varepsilon^2}{4k})$$

then the GBS setup can be efficiently simulated classically.

Given an ideal GBS setup, i.e: with $d_c = 0$ and $\eta = 1$, and also $r > 0$, we can then write the left hand side of the inequality as

$$\operatorname{sech}\left(\ln \frac{1}{e^{-2r}}\right)$$

and evaluates to less than 1 for $r > 0$. Observing that the right hand side term $\mathbf{exp}(-\frac{\varepsilon^2}{4k})$ always evaluates to 1 or larger for for any ε and k, we can conclude that this condition cannot be satisfied for a noiseless GBS.

From our result and from the more general result of [Qi+20], it is therefore impossible to perform a classical sampling using the QPDs of noiseless GBS. Thus it is necessary to consider noise of some sort in order to advance the investigation to the "boundary" between efficient and hard computation of a quantum system. We have seen from the result of [Qi+20] the necessary condition for the most amount of noise can be present for a GBS experiment for it to be hard to simulate classically. However, to close in on the boundary on the noise spectrum that determines whether it is possible to efficiently simulate a GBS experiment, we analyze a possible path towards a sufficient noise condition in order for a GBS experiment to be classically hard. We will discuss this attempt in the following section by formulating a model of GBS with noisy photodetectors.

7.2 Lossy Gaussian Boson Sampling

In their 2016 paper, Aaronson and Brod [AB16] employed a loss model for Boson Sampling that considers a fixed number s of lost photons at the preparation stage of the protocol, where there are l single photon inputs prepared. We will call this model as the *Fixed Loss Boson Sampling*. However, the Fixed Loss Boson Sampling have mentioned that this loss model is closely related to a model where there is a certain probability p_j of photon loss in input mode j. This is due to that the fixed loss photon can be considered as the average of the total photon loss $s = \sum_j p_j$. Assuming the same probability of photon loss in each mode $p = p_j$ for all j, we have $s = lp$ for l being the number of single photon inputs. It was shown that for s photons lost out of l input photons, if a classical algorithm $\mathcal{A}$ can approximate $\mathbb{P}_{s\text{-BS}}$ sufficiently close with sufficiently high probability, then we have $|\mathbf{GPE}|_{\pm}^2 \in \mathbf{BPP}^{\mathcal{A}}$. Then, we can follow closely the argument for the complexity of Boson Sampling outlined in [**Section 4.2**] but using an oracle that samples from the Fixed Loss Boson Sampling oracle instead. The proof that the Fixed Loss Boson Sampling cannot be efficiently simulated classically hence follows from the argument that $|\mathbf{GPE}|_{\pm}^2$ is #**P**-Hard.

Additionally, [AB16] also proposes another addition of noise to the AB fixed loss Boson Sampling model by imposing a *dark count* onto the photon counting measurements. Dark count measurement noise incurs some probability of the photon-counting measurement, or *photodetector*, randomly measuring a single photon when there is none present. They consider the case where there is a maximum number s of photon loss and dark count. Thus, the photon count being registered at the measurement consists of $l - s$ photons that went through the interferometer photon counting measurement and s dark counts. In which for a collision free outcome $k \in C_{m,l}$, we can consider a number of different ways of having the l single photons

It was then argued that the AB fixed loss Boson Sampling an approximation of $z|\mathbf{Per}(X)|^2$ where z is some number that is dependent on the probability of a dark count measurement.

More recently, [Shc19] has also proposed a Boson Sampling with a noise model imposed on its linear interferometer that is quantified in ε, we will refer to this model as the ε-interferometer BS. This result has shown that this interferometer noise model can be reformulated as a noise at the photodetection stage in the form of losses and a particular model of dark count. It was then shown that the probability of a measurement outcome from this noisy photodetection model is close to the probability $\mathbb{P}_{s\text{-LBS}}$ from the noise model of fixed photon loss presented in [AB16] if $\varepsilon = \mathcal{O}(1/k)$ for k single photon inputs. Consequently, it follows that ε-interferometer BS is as hard as AB fixed loss Boson Sampling.

Now, considering again the GBS paradigm, despite results analyzing necessary noise conditions for a classically hard GBS, such as the result in [Qi+20] we discussed in the previous section, there has been very little attention towards the analysis of sufficient noise conditions for GBS to be classically hard. To further the investigation of the "boundary" of the hardness of classical simulation for GBS, it is therefore necessary and natural to analyze a possible noise criteria for a sufficient condition for a classically hard GBS experiment. In this section, we therefore lay down a reasonable formulation of a noisy GBS with an outcome probability that is in a similar form to the results of [AB16] and [Shc19]. This model focuses on modeling the noise of its photodetectors, specifically the efficiency and dark counts.

7.2.1 GBS with Imperfect Photodetection

To later compare our formulation of noisy GBS with, we will now state the measurement outcome probability of the AB Boson Sampling [AB16], including the model that considers dark counts.

Definition 7.2.1 (Fixed Loss Boson Sampling and Shuffled Boson Sampling [AB16])**.** For an Boson Sampling with m mode interferometer $L \in \mathbb{C}^{l,m}$ and s photons loss before going into the interferometer out of a total of l prepared single photon inputs, we have the measurement outcome of $k \in C_{m,l-s}$ as

$$\mathbb{P}(k) = \frac{1}{|C_{l,l-s}|} \sum_{v \in C_{l,l-s}} |\mathbf{Per}(L_{v,k})|^2$$

where the sum is over all no-collision states with a total of $l-s$ photons in the first l modes of the m input modes of the interferometer. We call this model as the *Fixed Loss Boson Sampling*.

If have instead an m-mode Boson Sampling with interferometer $L \in \mathbb{C}^{l,m}$ that suffers from a maximum of s photon losses and dark count noise, we now have the outcome probability of $k \in C_{m,l}$ as

$$\mathbb{P}(k) = \sum_{j=0}^{s} \frac{d_j}{|C_{l,l-j}|^2} \sum_{v \in C_{l,l-j}} \sum_{u \in C^{(k)}_{l,l-j}} |\mathbf{Per}(L_{v,u})|^2 \tag{7.2.1}$$

where the probability of j dark counts is $d_j \in [0,1]$ and $C^{(k)}_{l,l-j} = \{u \in C_{m,l-j} : \forall j, u_j \leq k_j\}$. We call this model the *Shuffled Boson Sampling*. ◀

The set $C^{(k)}_{l,l-j}$ can be understood as the set of collision free states that represents the actual photon outputs from the interferometer, with k being the measured outcome considering j dark counts that compensate j photon losses at the single-photon preparations. For example, if we have photon loss of $j = 2$ and $k = (1, 1, 0, 0, 1, 1)$ (thus $l = 4$), we may have the dark count occurring in mode 1 and 6, in which the actual output from the interferometer is $u = (0, 1, 0, 0, 1, 0)$. We would also like to note that as in [AB16] for the Fixed Loss Boson Sampling, we can conveniently consider the interferometer L as a $l\times(l-s)$ matrix, considering that we have l single-photon states prepared, but measure only $(l - s)$ single photon states due to s losses. The $(l - s)$ rows are selected from the j-th row of the full $l \times m$ matrix where there is a photon measured in the j-th mode, i.e. $k_j = 1$. But for the Shuffled Boson Sampling, we can consider L to be a $l \times l$ matrix instead because we prepare l single-photon states and measure l single-photon states, although only $(l - s)$ rows and columns will be selected in calculating the permanent due to s photon losses and s dark counts. Again, here we select the rows according to measurement outcome of interest $k \in C_{m,l}$.

Now we will begin to derive our noisy GBS model. First, we note that we model dark count and photon loss of the photodetection as in [BK85], identically for the measurement in each of the m output modes of the GBS circuit. As the loss and dark count model in [BK85] considers the sets of measurement outcome beyond the collision-free states, we will adopt this more general model faithfully and put aside the collision-free measurement outcome requirement for now. We will now introduce the following notations, which we will use throughout the derivation and discussion of this noisy GBS model:

1. For $u \in \mathbb{N}^m$, we have $\mathbb{P}_{\mathrm{dc}}(u) = \prod_{j=1}^{m} \mathbb{P}_{\mathrm{dc}}(u_i)$ with $\mathbb{P}_{\mathrm{dc}}(u_j)$ being the probability of registering u_j dark count in the j-th mode.

2. For loss probability $(1 - \eta)$ on each photodetector and $j, n \in \mathbb{N}^m$, the probability of detecting j photon pattern when having pattern n actually present is $\mathbb{P}^{(\eta)}_{\mathrm{ef}}(j|n) = \eta^{|j|}(1 - \eta)^{|n-j|}\binom{n}{j}$, where we have 1-norm $|j| = \sum_{i=1}^{m} |j_i|$, multidimensional binomial coefficient $\binom{n}{j} = \prod_{i=1}^{m} \frac{n_i!}{j_i!(n_i-j_i)!}$. We also define the multidimensional Photon Loss probability $\mathbb{P}^{(\eta)}_{\mathrm{ef}}(j|n) = \eta^{|j|}(1-\eta)^{|n-j|}\binom{n}{j}$, and the determinant coefficient $d = (\mathbf{Det}(\sigma_Q))^{-1/2}$ as in [**Section 5.1**].

With probability of having $u \in \mathbb{N}^m$ dark count $\mathbb{P}_{\mathrm{dc}}(u)$, where u_j is the dark count of the j-th mode and with loss probability $(1 - \eta)$ on each photodetector, probability of detecting j photon pattern when having n actually present $\mathbb{P}^{(\eta)}_{\mathrm{ef}}(j|n) = \eta^{|j|}(1-\eta)^{|n-j|}\binom{n}{j}$ as in [BK85], we have the POVM of the photodetector $\{\Pi_k(\eta)\}_{k\in\mathbb{N}}$ with POVM element

$$\Pi_k(\eta) = \sum_{j\in\Gamma_k} \mathbb{P}_{\mathrm{dc}}(k - j) \sum_{n\in\Theta_j} \mathbb{P}^{(\eta)}_{\mathrm{ef}}(j|n)|n\rangle\langle n|$$

where we define the sets

$$\Gamma_n = \{j \in \mathbb{N}^m : \forall i, j_i \leq n_i\}$$
$$\text{and}$$
$$\Theta_n = \{j \in \mathbb{N}^m : \forall i, j_i \geq n_i\}.$$

Now, considering the outcome probability of the noiseless GBS $\mathbb{P}_{\texttt{GBS}}(k) = \mathbf{Tr}(\rho_{out}|k\rangle\langle k|) = \frac{1}{\sqrt{\mathbf{Det}(\sigma_Q)}}|\mathbf{Haf}(B_k)|^2$ as in [**Section 5.1**], we can easily see that the probability of the measured output photon pattern using our noisy photodetectors is

$$\begin{aligned}\mathbb{P}^{(B)}_{\eta\texttt{GBS}}(k) &= \mathbf{Tr}(\rho_{out}\Pi_k(\eta)) \\ &= \sum_{j\in\Gamma_k} \mathbb{P}_{\text{dc}}(k-j) \sum_{n\in\Theta_j} \mathbb{P}^{(\eta)}_{\text{ef}}(j|n)\; d|\mathbf{Haf}(B_n)|^2 \qquad (7.2.2) \\ &= \sum_{j\in\Gamma_k} \mathbb{P}_{\text{dc}}(k-j) \left(\mathbb{P}^{(\eta)}_{\text{ef}}(j|k)\; d|\mathbf{Haf}(B_k)|^2 + \sum_{n\in\Theta_j\setminus\{k\}} \mathbb{P}^{(\eta)}_{\text{ef}}(j|n)\; d|\mathbf{Haf}(B_n)|^2 \right).\end{aligned}$$

Where clear from the context, the superscript "(B)" indicating the input matrix to the Hafnian function will be surpressed for notational simplicity. We can alternatively switch the outer and inner sums and write the probability of samples from η-GBS as a mixture of the probability of all possible sampled patterns from the ideal GBS, obtaining

$$\mathbb{P}_{\eta\texttt{GBS}}(k) = \sum_{n\in\mathbb{N}^m} d|\mathbf{Haf}(B_n)|^2 \sum_{j\in\Gamma_n} \mathbb{P}_{\text{dc}}(k-j)\mathbb{P}^{(\eta)}_{\text{ef}}(j|n). \qquad (7.2.3)$$

Further, we can separate the terms in the sum with Hafnians with the corresponding matrix B for output measurement k from the terms with Hafnians of matrices for the other outputs, as follows

$$\mathbb{P}_{\eta\texttt{GBS}}(k) = \tilde{\mathbb{P}}_{\eta\texttt{GBS}}(k) + \bar{\mathbb{P}}_{\eta\texttt{GBS}}(k),$$

where we define

$$\tilde{\mathbb{P}}_{\eta\texttt{GBS}}(k) = d|\mathbf{Haf}(B_k)|^2 \sum_{j\in\Gamma_k} \mathbb{P}_{\text{dc}}(k-j)\mathbb{P}^{(\eta)}_{\text{ef}}(j|k)$$

and

$$\begin{aligned}\bar{\mathbb{P}}_{\eta\texttt{GBS}}(k) &= \sum_{j\in\Gamma_k} \mathbb{P}_{\text{dc}}(k-j) \sum_{n\in\Theta_j\setminus\{k\}} \mathbb{P}^{(\eta)}_{\text{ef}}(j|n)\; d|\mathbf{Haf}(B_n)|^2 \\ &= \sum_{n\in\mathbb{N}\setminus\{k\}} d|\mathbf{Haf}(B_n)|^2 \sum_{j\in\Gamma_n} \mathbb{P}_{\text{dc}}(k-j)\mathbb{P}^{(\eta)}_{\text{ef}}(j|n).\end{aligned}$$

where we define $\mathbb{P}_{\text{dc}}(k) = \mathbb{P}_{\text{dc}}(0)$ for all negative integer $k < 0$.

It therefore apparent that we can express $\mathbb{P}_{\eta\texttt{GBS}}$ in terms of nested sum of Hafnians for different submatrices of B as the expression of the probability of the measurement outcome (7.2.2) resembles that of the Shuffled Boson Sampling in (7.2.1). Moreover as in [AB16], we can express in the sum a term that is proportional to the Hafnian of a matrix that we are interested in, $\tilde{\mathbb{P}}_{\eta\texttt{GBS}}(k)$.

Despite this seemingly promising formulation, we admit that we fell short of providing a proof of values of η and $\mathbb{P}_{\text{dc}}$ that can be sufficiently achieved to guarantee that an experiment based on out noisy GBS model is hard to simulate classically. However, given the resemblance of this model to the AB lossy Boson Sampling, we do think that a similar argument as one

given in [AB16] is a promising avenue that can be pursued to achieve the hardness proof for our noisy GBS model. This can be done by showing that for there exist an oracle $\mathcal{Q}$ approximating $\mathbb{P}_{\eta\texttt{GBS}}$ with a high probability, there is an algorithm $\mathcal{K}$ in the polynomial hierarchy such that when it equips $\mathcal{Q}$, it can approximate the $|\mathbf{GHE}|^2_{\pm}$ problem. Given this possible approximation of $|\mathbf{GHE}|^2_{\pm}$ by $\mathcal{K}^{\mathcal{Q}}$, we can then show that $|\mathbf{GHE}|^2_{\pm} \in \Sigma^p_u$ for the u-th level of the polynomial hierarchy that the algorithm $\mathcal{K}$ is in.

However, there is another technicality issue where the sets Θ_j and Γ_n need to be modified such that only collision-free measurement outcomes will be considered. This is a technicality that needs to be addressed to pursue this proof technique so that we can embed the random matrix X in the Haar-random matrix B. Given this shortcoming, these possible pathways will be left for future work.

Lastly, we also note that a Homodyne Tomography (HT) computer simulation that was discussed in [**Chapter 6**] that runs this noisy GBS model is also an interesting expansion to both the PF GBS algorithm and the noisy GBS model. The rate of how the approximation error is decreasing in the HT method may be one interesting and important aspect to be observed in this kind of simulation, especially in comparing between different loss probability and dark count of the model. Although this is surely an interesting avenue to pursue in investigating how noise influences a computer simulation of GBS, we will leave it for future work. In the following section, we will again give a different model that was more motivated by the QPD negativity analysis that seems to also show promise for a proof of a sufficient noise condition for a hard GBS experiment.

7.3 Photon-Added Thermal State Continuous-Variable Sampling

In this very last section of this thesis, we again propose a model in an attempt to formulate a sufficient noise condition for a classically hard GBS experiment to close in on the quantum-classical computational boundary. We call this model the Photon-added Thermal state Continuous-Variable sampling (PATS CV sampling). This model in a sense extends the model proposed in [Cha+17] called the CV-Sampling (CVS) and make changes to it by using the *Single Photon-added Thermal state* (which we will often refer to omitting the word "single", and abbreviate as PAT state) as an input. It was shown in [Vei+13] that a PAT state passing through a noisy channel with transmissivity η has a Wigner function representation W_{PAT} that is non-negative everywhere for $\eta \leq \frac{1}{2}$, and negative for some values for $\eta > \frac{1}{2}$. This result [Vei+13] on a precise value of η where its Wigner function transitions from being all non-negative to be negative for some values can be seen as a promising approach in closing in on the "boundary" of the hardness of classical simulation for a quantum system.

Recalling our discussions in [**Section 7.1**] on sampling of QPD representations inside the overlap of a measurement outcome probability as an efficient classical simulation of GBS, the choice of CVS model and PAT state can seem to be promising. However, we must first define the CVS model.

Definition 7.3.1 (Continuous Variable Sampling (CVS) [Cha+17])**.** For an m mode Continuous-Variable Sampling (CVS) system, we have

1. photon added squeezed states $\widehat{a}^\dagger|r\rangle$ or photon subtracted squeezed states $\widehat{a}|r\rangle$ as the inputs for the first l modes and single mode squeezed states $|r\rangle$ for the last $m-l$ modes,
2. unitary evolution matrix $Q = He^{-i\theta L}$ where H is an orthogonal matrix and L is a symmetric orthogonal matrix, and
3. Heterodyne measurement at all m outputs.

Thus the probability of the measurement outcome $\alpha \in \mathbb{C}^m$ is $\mathbb{P}_{\mathsf{cvs}}(\alpha) = \mathbf{Tr}(|\alpha\rangle\langle\alpha|\, Qb|s\rangle\langle s|b^\dagger Q^\dagger)$ for $b \in \{\widehat{a}, \widehat{a}^\dagger\}$. ◀

It is clear therefore that $\mathbb{P}_{\mathsf{cvs}}$ can be formulated in the overlap form of (7.1.1), giving

$$\mathbb{P}_{\mathsf{cvs}}(\alpha) = \int_{\mathbb{C}^m}\int_{\mathbb{C}^m} W^{(-s)}_{|\alpha\rangle\langle\alpha|}(\beta)\; T^{(s,t)}_Q(\beta|\gamma)\; W^{(t)}_{\rho_{in}}(\gamma)\; d^{2m}\gamma\; d^{2m}\beta. \tag{7.3.1}$$

As in the original formulation of CVS, ρ_{in} is either photon added squeezed states or photon subtracted squeezed state, the negativity behavior of $W_{\rho_{in}}$ is unclear. However, it was proven in [Cha+17] that CVS is no more than a special case of GBS, which measurement outcome probability can be expressed in terms of the #**P**-hard hafnian function. Thus by the same argument outlined in [**Section 5.2**], CVS cannot be efficiently simulated classically given that the polynomial hierarchy does not collapse. Keeping this fact in mind and at the same time referring back to (7.3.1), we can see that we can follow the same argument in the proof of [**Proposition 7.1.2**] where we can rewrite (7.3.1) as

$$\mathbb{P}_{\mathsf{cvs}}(\alpha) = \int_{\mathbb{C}^m} W^{(-s)}_{|\alpha\rangle\langle\alpha|}(Q\beta)\; W^{(s)}_{\rho_{in}}(\beta)\; d^{2m}\beta$$

because Q is unitary. Further noting that $W^{(-s)}_{|\alpha\rangle\langle\alpha|}$ is Gaussian for $s = 0$ and that the Wigner function of PAT state is non-negative for transmissivity $\eta \leq 1/2$ by [Vei+13], classically sampling from the overlap (7.3.1) is possible if we have the PAT state with transmissivity $\eta \leq 1/2$ as input $\rho_{in} = \rho^{\text{PAT}}_\eta$. However, given that $s = 0$, having the PAT transmissivity $\eta > 1/2$ will make the classical sampling from the overlap not possible because of the negativity of W_{PAT} for some values. With the considerations towards QPD negativities as a characterization of non-classicality, the investigation into the classical hardness boundary of the CVS model with input PAT states with different transmissivity η is therefore seems to be a promising path. We call the CVS model with PAT state as its input as *PATS CV Sampling*.

Now, the reader might ask the relevance of this model towards the complexity of GBS. The answer to this would be that given that CVS is a special case of GBS, we will further slow later that PAT sampling is none other than GBS with particularly formulated noisy photodetectors.

7.3.1 Photon Added Thermal State and Thermalized Number State

In this subsection, we will derive some expressions and properties of the PAT state, which we simply define as a thermal state with expected photon number $\bar{n}$ with bosonic creation operator applied to it $\rho^{\text{PAT}}_{\bar{n}} = \widehat{a}^\dagger \rho^{\text{Th}}_{\bar{n}} \widehat{a}$. Most importantly here, we will derive the

relationship between the *Thermalized Single-photon state* and the PAT state, presented in [**Theorem 7.3.4**]. This will be an important relation that indicates a physical realization of a normalized PAT state, which we will require in order to analyze the PAT CV sampling model later in [**Section 7.3.2**]. This relationship furthermore interestingly from a physical perspective shows the relationship between adding a single photon into a thermal state and imposing a "thermal noise" onto a single-photon state. This "thermal noise" is represented by a Thermalizing operator which essentially is a squeezing operation on two modes with one output mode traced out, as we will discuss in more detail later. We will firstly derive this *Thermalizing operator* that generates a thermal state from a vacuum with the proposition below.

Proposition 7.3.2 (Thermal State Decomposition)**.** *For $r \in \mathbb{R}$, unitary operator $T(r) = \mathbf{exp}(r(\widehat{a}_c\widehat{a}_d - \widehat{a}_c^\dagger\widehat{a}_d^\dagger))$ on two mode system $\mathcal{H}_c \otimes \mathcal{H}_d$, Thermal State $\rho_{\bar{n}}^{\text{Th}}$ in system $\mathcal{H}_c$ with mean photon number $\bar{n} = \sinh^2(r)$ can be expressed as*

$$\rho_{\bar{n}}^{\text{Th}} = \mathbf{Tr}_d\left(T(r)|0,0\rangle\langle 0,0|T^\dagger(r)\right)$$

where $|a,b\rangle = |a\rangle_c \otimes |b\rangle_d$.

We call T the *Thermalizing operator* as in [BK85]; [FC88]; [FF98]. We note that the Thermalizing Operator T coincides with the *Two-mode Squeezing* operation $S_2(r) = \mathbf{exp}\left(r(\widehat{a}_c\widehat{a}_d - \widehat{a}_c^\dagger\widehat{a}_d^\dagger)\right)$ [Wee+12]. Thus, tracing one mode of a two-mode squeezed state with squeezing r will produce a thermal state with mean photon number $\bar{n} = \sinh^2(r)$.

*proof of [**Proposition 7.3.2**].* For simplicity, we write a vector in a single Hilbert space $\mathcal{H}_c$ as $|x\rangle$ without subscript where there is no apparent ambiguity. Now, expanding the Thermal state in $\mathcal{H}_c$ in terms of number state $|n\rangle$, we have

$$\begin{aligned}\rho_{\bar{n}}^{\text{Th}} &= \frac{1}{\bar{n}+1}\sum_{n\in\mathbb{N}}\left(\frac{\bar{n}}{\bar{n}+1}\right)^n |n\rangle\langle n| \\ &= \left(1-\tanh^2(r)\right)\sum_{n\in\mathbb{N}}\tanh^{2n}(r)|n\rangle\langle n|\end{aligned}$$

where the second equality was obtained by noting that $\frac{1}{\bar{n}+1} = 1-\frac{\bar{n}}{\bar{n}+1}$ and using the hyperbolic function identity $\tanh^2(r) = \frac{\sinh^2(r)}{1+\sinh(^2r)}$. We now consider the Hilbert space $\mathcal{H}_d$ and using the orthonormality of the number state basis $\langle n|m\rangle = \delta_{n,m}$ in $\mathcal{H}_d$ we get

$$\begin{aligned}\rho_{\bar{n}}^{\text{Th}} &= \left(1-\tanh^2(r)\right)\sum_{n,j\in\mathbb{N}}(-\tanh(r))^{n+j}\sum_{k\in\mathbb{N}} {}_d\langle k|n\rangle_d\,|n\rangle_{cc}\langle j|\,{}_d\langle j|k\rangle_d \\ &= \sum_{k\in\mathbb{N}} {}_d\langle k|\left(1-\tanh^2(r)\right)\sum_{n,j\in\mathbb{N}}(-\tanh(r))^n\,|n,n\rangle\langle j,j|\,(-\tanh(r))^j\,|k\rangle_d \\ &= \mathbf{Tr}_d\left(\left(1-\tanh^2(r)\right)^{\frac{1}{2}}\sum_{n,j\in\mathbb{N}}(-\tanh(r))^n\,|n,n\rangle\langle j,j|\,(-\tanh(r))^j\left(1-\tanh^2(r)\right)^{\frac{1}{2}}\right).\end{aligned}$$

Noting that the expression inside the trace is precisely the two-mode squeezed state $T(r)|0,0\rangle = S_2(r)|0,0\rangle = \sqrt{1-\lambda^2}\sum_j(-\lambda)^j|j,j\rangle$ with $\lambda = \tanh(r)$ as in [Wee+12], we finally get

$$\rho_{\bar{n}}^{\mathrm{Th}} = \mathbf{Tr}_d\left(T(r)|0,0\rangle\langle 0,0|T^\dagger(r)\right).$$

□

Now, with the concept of Thermalizing operator in hand, we can further look at the case where it is applied to a single-photon state $|1\rangle$ instead of vacuum $|0\rangle$ to probe into the relation between states where a single photon is added onto a thermal state and where a thermalization operator is applied to a single-photon state. We must first introduce the latter where thermalizing operation is done on a single photon state, producing a state we call the *Thermalized Single-photon State* (TSP state).

Definition 7.3.3. For $|a,b\rangle = |a\rangle_c \otimes |b\rangle_d$ and $\bar{n} = \sinh^2(r)$ let *Thermalized Single-photon State (*TSP*)*

$$\rho_{\bar{n}}^{\mathrm{TSP}} = \mathbf{Tr}_d\left(T(r)\widehat{a}_c^\dagger|0,0\rangle\langle 0,0|\widehat{a}_c T^\dagger(r)\right).$$

◀

We now ready to state the relation between TSP state and the state in which thermalization is done on a two mode vacuum before adding a single photon to it, namely the Single-Photon Added Thermal state (PAT state).

Theorem 7.3.4. *For $\bar{n} = \sinh^2(r)$, Thermal state $\rho_{\bar{n}}^{\mathrm{Th}}$, Photon-added Thermal state $\rho_{\bar{n}}^{\mathrm{PAT}} = \widehat{a}^\dagger\rho_{\bar{n}}^{\mathrm{Th}}\widehat{a}$, and Thermalized Single Photon state $\rho_{\bar{n}}^{\mathrm{TSP}}$, we have the relation*

$$\rho_{\bar{n}}^{\mathrm{TSP}} = \left(1-\tanh^2(r)\right)\rho_{\bar{n}}^{\mathrm{PAT}}.$$

Equivalently by the hyperbolic function identity $1+\sinh(\theta) = \frac{1}{1-\tanh(\theta)}$, we can write the above relation as

$$\rho_{\bar{n}}^{\mathrm{PAT}} = (1+\sinh^2(r))\rho_{\bar{n}}^{\mathrm{TSP}} = (1+\bar{n})\rho_{\bar{n}}^{\mathrm{TSP}}.$$

Clearly, when there is no squeezing, i.e: $r=0$, (since $\tanh^2(0)=0$), we have $\rho_{\bar{n}}^{\mathrm{TSP}} = \rho_{\bar{n}}^{\mathrm{PAT}}$. Otherwise, as $|r|\to\infty$, we have $\rho_{\bar{n}}^{\mathrm{PAT}} = 0$.

*proof of [**Theorem 7.3.4**].* First we consider applying a thermalization operation $T(r)$ on a two-mode state with a single photon in the first mode and a vacuum in the second mode $|1,0\rangle$, resulting in the state

$$\tilde{\rho}_{\bar{n}}^{\mathrm{TSP}} = T(r)\widehat{a}_c^\dagger|0,0\rangle\langle 0,0|\widehat{a}_c T^\dagger(r),$$

thus $\rho_{\bar{n}}^{\mathrm{TSP}} = \mathbf{Tr}_d(\tilde{\rho}_{\bar{n}}^{\mathrm{TSP}})$. Because T is unitary, as $T^\dagger(r) = T(-r)$ and $T(-r)\ T(r) = I$, we can write

$$\tilde{\rho}_{\bar{n}}^{\mathrm{TSP}} = \widehat{a}_c^\dagger(r)\ T(r)|0,0\rangle\langle 0,0|T^\dagger(r)\ \widehat{a}_c(r).$$

where we have the Thermalized Bosonic Operators $\widehat{a}_c^\dagger(r) = T(r)\widehat{a}_c^\dagger T^\dagger(r)$ and $\widehat{a}_c(r) = T(r)\widehat{a}_c T^\dagger(r)$. Using the identity of thermalized bosonic in terms of hyperbolic functions presented in [**Appendix B**], we have

$$\begin{aligned}\widehat{a}_c^\dagger(r) &= \cosh(r)\widehat{a}_c^\dagger + \sinh(r)\widehat{a}_d \\ \widehat{a}_c(r) &= \cosh(r)\widehat{a}_c + \sinh(r)\widehat{a}_d^\dagger\end{aligned}$$

as in [BR97, Ch.3.7] and [BK85]. Letting $\tilde{\rho}_{\bar{n}}^{\mathrm{Th}} = T(r)|0,0\rangle\langle 0,0|T^\dagger(r)$ for conciseness, we take the partial trace on $\mathcal{H}_d$ space of $\tilde{\rho}_{\bar{n}}^{\mathrm{TSP}}$, obtaining

$$\begin{aligned}&\mathbf{Tr}_d(\tilde{\rho}_{\bar{n}}^{\mathrm{TSP}}) \\ &= \mathbf{Tr}_d\left(\left(\cosh(r)\widehat{a}_c^\dagger + \sinh(r)\widehat{a}_d\right)\tilde{\rho}_{\bar{n}}^{\mathrm{Th}}\left(\cosh(r)\widehat{a}_c + \sinh(r)\widehat{a}_d^\dagger\right)\right) \\ &= \mathbf{Tr}_d\left(\cosh^2(r)\widehat{a}_c^\dagger\tilde{\rho}_{\bar{n}}^{\mathrm{Th}}\widehat{a}_c + \sinh^2(r)\widehat{a}_d\rho_{\bar{n}}^{\mathrm{Th}}\widehat{a}_d^\dagger + \sinh(r)\cosh(r)\left(\widehat{a}_c^\dagger\tilde{\rho}_{\bar{n}}^{\mathrm{Th}}\widehat{a}_d^\dagger + \widehat{a}_d\tilde{\rho}_{\bar{n}}^{\mathrm{Th}}\widehat{a}_c\right)\right).\end{aligned} \tag{7.3.2}$$

We will now consider each of the three terms in (7.3.2) separately as the partial trace $\mathbf{Tr}_d$ is a linear mapping. First, we have a closer look into the density operators in last term, each containing the bosonic operators of Hilbert space $\mathcal{H}_c$ and $\mathcal{H}_d$ in either sides of $\tilde{\rho}_{\bar{n}}^{\mathrm{Th}}$. We define $t_r = \tanh(r)$ for shorthand, then expand $\tilde{\rho}_{\bar{n}}^{\mathrm{Th}}$ in terms of outer product of the Fock basis and get

$$\begin{aligned}\mathbf{Tr}_d(\widehat{a}_c^\dagger\tilde{\rho}_{\bar{n}}^{\mathrm{Th}}\widehat{a}_d^\dagger) &= \left(1-t_r^2\right)\sum_{n,m\in\mathbb{N}}(-t_r)^{n+m}\sum_{k\in\mathbb{N}}{}_d\langle k|n\rangle_d\,\widehat{a}_c^\dagger|n\rangle_{cc}\langle m|_d\langle m|\widehat{a}_d^\dagger|k\rangle_d \\ &= \left(1-t_r^2\right)\sum_{n\in\mathbb{N}}\sum_{m=1}^{\infty}(-t_r)^{n+m}\,\delta_{n,m-1}|n+1\rangle_{cc}\langle m|m(n+1) \\ &= \left(1-t_r^2\right)\sum_{m=1}^{\infty}(-t_r)^{2m-1}\,|m\rangle_{cc}\langle m|m^2 \\ &= \left(1-t_r^2\right)\sum_{m\in\mathbb{N}}(-t_r)^{2m+1}\,\widehat{a}_c^\dagger|m\rangle_{cc}\langle m|\widehat{a}_c\end{aligned}$$

and

$$\begin{aligned}\mathbf{Tr}_d(\widehat{a}_d\tilde{\rho}_{\bar{n}}^{\mathrm{Th}}\widehat{a}_c) &= \left(1-t_r^2\right)\sum_{n,m\in\mathbb{N}}(-t_r)^{n+m}\sum_{k\in\mathbb{N}}{}_d\langle k|\widehat{a}_d|n\rangle_d|n\rangle_{cc}\langle m|\widehat{a}_c\;{}_d\langle m|k\rangle_d \\ &= \left(1-t_r^2\right)\sum_{n=1}^{\infty}\sum_{m\in\mathbb{N}}(-t_r)^{n+m}\,\delta_{n-1,m}|n\rangle_{cc}\langle m+1|n(m+1) \\ &= \left(1-t_r^2\right)\sum_{n=1}^{\infty}(-t_r)^{2n-1}\,|n\rangle_{cc}\langle n|n^2 \\ &= \left(1-t_r^2\right)\sum_{n\in\mathbb{N}}(-t_r)^{2n+1}\,\widehat{a}_c^\dagger|n\rangle_{cc}\langle n|\widehat{a}_c.\end{aligned}$$

Now we note that they are multiplicatively equal to the photon added thermal state. More precisely

$$\mathbf{Tr}_d(\widehat{a}_c^\dagger\tilde{\rho}_{\bar{n}}^{\mathrm{Th}}\widehat{a}_d^\dagger) = \mathbf{Tr}_d(\widehat{a}_d\tilde{\rho}_{\bar{n}}^{\mathrm{Th}}\widehat{a}_c) = (-t_r)\widehat{a}^\dagger\rho_{\bar{n}}^{\mathrm{Th}}\widehat{a}.$$

Now, the second term of (7.3.2) can be rewritten as follows

$$\begin{aligned}
\mathbf{Tr}_d\left(\widehat{a}_d\tilde{\rho}_{\bar{n}}^{\mathrm{Th}}\widehat{a}_d^\dagger\right) &= (1-t_r^2)\sum_{n,m\in\mathbb{N}}(-t_r)^{n+m}\sum_{k\in\mathbb{N}}{}_d\langle k|\widehat{a}_d|n\rangle_d|n\rangle_{cc}\langle m|\,{}_d\langle m|\widehat{a}_d^\dagger|k\rangle_d \\
&= (1-t_r^2)\sum_{n=1}^{\infty}(-t_r)^{2n}|n\rangle_{cc}\langle n|n \\
&= (1-t_r^2)(-t_r)^2\sum_{n\in\mathbb{N}}(-t_r)^{2n}\widehat{a}_c|n\rangle_{cc}\langle n|\widehat{a}_c^\dagger \\
&= (-t_r)^2\widehat{a}^\dagger\rho_{\bar{n}}^{\mathrm{Th}}\widehat{a}.
\end{aligned}$$

Lastly for the first term of (7.3.2), we simply have

$$\begin{aligned}
\mathbf{Tr}_d\left(\widehat{a}_c^\dagger\tilde{\rho}_{\bar{n}}^{\mathrm{Th}}\widehat{a}_c\right) &= (1-t_r^2)\sum_{n,m\in\mathbb{N}}(-t_r)^{n+m}\sum_{k\in\mathbb{N}}{}_d\langle k|n\rangle_d\,\widehat{a}_c^\dagger|n\rangle_{cc}\langle m|\widehat{a}_c\,{}_d\langle m|k\rangle_d \\
&= (1-t_r^2)\sum_{n\in\mathbb{N}}(-t_r)^{2n}\widehat{a}_c^\dagger|n\rangle_{cc}\langle n|\widehat{a}_c \\
&= \widehat{a}^\dagger\rho_{\bar{n}}^{\mathrm{Th}}\widehat{a}.
\end{aligned}$$

Therefore noting that each of the terms in (7.3.2) includes $\rho_{\bar{n}}^{\mathrm{PAT}} = \widehat{a}^\dagger\rho_{\bar{n}}^{\mathrm{Th}}\widehat{a}$ and using the hyperbolic function identity $\cosh^2(r) = \frac{1}{1-\tanh^2(r)}$ and $\sinh^2(r) = \frac{\tanh^2(r)}{1-\tanh^2(r)}$, we end up with

$$\begin{aligned}
\rho_{\bar{n}}^{\mathrm{TSP}} &= \mathbf{Tr}_d(\tilde{\rho}_{\bar{n}}^{\mathrm{TSP}}) \\
&= \left(\cosh^2(r) + 2\sinh(r)\cosh(r)(-t_r) + \sinh^2(r)(-t_r)^2\right)\rho_{\bar{n}}^{\mathrm{PAT}} \\
&= \left(\sqrt{\frac{1}{1-t_r^2}} + (-t_r)\sqrt{\frac{t_r^2}{1-t_r^2}}\right)^2\rho_{\bar{n}}^{\mathrm{PAT}} \\
&= (1-t_r^2)\rho_{\bar{n}}^{\mathrm{PAT}},
\end{aligned}$$

which concludes the proof. □

With this discussion resulting in the relationship between thermalized single-photon state and the photon-added thermal state, we are now ready for a more detailed discussion on the PATS CV model.

7.3.2 Time-Reversed PATS CV Sampling

In this section, we will present the time-reversed PATS CV sampling model that is inspired by the similarly named model in [Cha+17] that "reverses" the CVS model by considering the input density operator as a measurement operator and the measurement operator as an input density operator to the conjugated linear intrferometer. We similarly formulate the time-reversed PATS CV sampling model by formulating the normalized PAT state $\rho_{\bar{n}}^{\mathrm{nPAT}}$, which we will define shortly, as a POVM element. Similar approach to formulate a POVM element that coincides with a noisy PAT state $\rho_{\bar{n},\eta}^{\mathrm{nPAT}}$ will be done as well. We will show that

these POVM elements, can be interpreted as photodetection POVMs with particular dark count and loss noise models.

First, to have a better sense of the "time reversal" of the PATS CV sampling model, we define the time-reversed CVS model proposed in [Cha+17]. Illustrated in [**Figure 7.1**], the time-reversed CVS model is comprised of single-mode squeezed state (SMSS) inputs, which density operator coincides with the POVM element of the Heterodyne measurement, and on/off photodetection after a squeezing operation, which coincides with the density operator of the photon-added squeezed state.

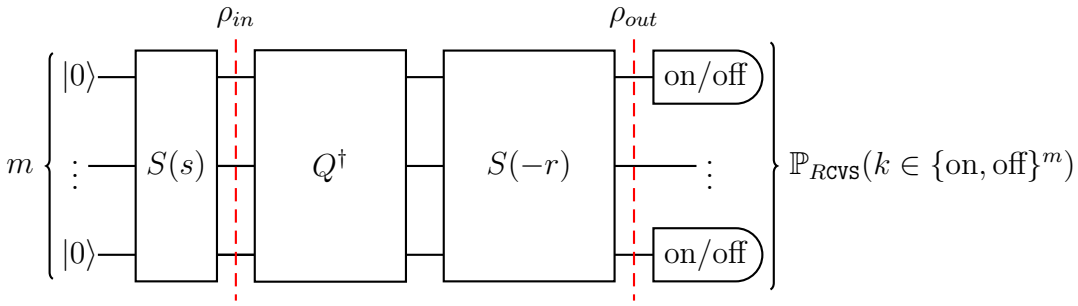

Figure 7.1: Reversed CV sampling (RCVS) circuit which coincides with GBS with an additional squeezing before the measurement.

Analogous what has been done to the time-reversed CVS, we can equivalently define the input to the PATS CV sampling model as SMSS inputs as we also have the Heterodyne measurements. However, we will need to find a POVM element that coincides with the density operator of the PAT state. But first, we need to address a problem with the PAT state, that is the fact that it is unnormalized, that is $\mathbf{Tr}(\rho_{\bar{n}}^{\text{PAT}}) \neq 1$ as a consequence of having a non-unitary creation operator in it. This is apparent from [**Definition 2.2.1**] that $\widehat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle$. To explicitly calculate $\mathbf{Tr}(\rho_{\bar{n}}^{\text{PAT}})$, consider an expansion of the PAT state in the Fock basis from the definition of the Thermal state [**Definition 2.2.7**] with $1+\bar{n} = 1+\sinh^2(r) = \frac{1}{1-\tanh^2(r)}$

$$\begin{aligned}\rho_{\bar{n}}^{\text{PAT}} &= (1-\tanh^2(r))\sum_{n\in\mathbb{N}} \tanh^{2n}(r)\, \widehat{a}^\dagger|n\rangle\langle n|\widehat{a} \\ &= (1-\tanh^2(r))\sum_{n\in\mathbb{N}} \tanh^{2n}(r)(n+1)|n+1\rangle\langle n+1|.\end{aligned}$$

Therefore, we can write its trace as

$$\begin{aligned}\mathbf{Tr}(\rho_{\bar{n}}^{\text{PAT}}) &= (1-\tanh^2(r))\sum_{n\in\mathbb{N}} \tanh^{2n}(r)\sum_{k\in\mathbb{N}}\langle k|n+1\rangle\langle n+1|k\rangle(n+1) \\ &= (1-\tanh^2(r))\sum_{n\in\mathbb{N}} \tanh^{2n}(r)(n+1) \\ &= (1-\tanh^2(r))^{-1}. \qquad (7.3.3)\end{aligned}$$

Using the hyperbolic function relation, we can rewrite the trace of $\rho_{\bar{n}}^{\text{PAT}}$ as $\mathbf{Tr}(\rho_{\bar{n}}^{\text{PAT}}) = \frac{1}{1-\tanh^2(r)} = 1+\sinh^2(r) = 1+\bar{n}$. Thus, for expected photon number $\bar{n} > 0$ or equivalently

for $r > 0$, we clearly have $\mathbf{Tr}(\rho_{\bar{n}}^{\mathrm{PAT}}) > 1$. Measurement using this POVM element Over a certain basis states $\{|\psi_j\rangle\}_j$ (for example, the fock states $\{|n\rangle\}_{n\in\mathbb{N}}$), it can be thought that the sum of the probabilities over that basis is larger than one $\sum_j \mathbf{Tr}(|\psi_j\rangle\langle\psi_j|\rho_{\bar{n}}^{\mathrm{PAT}}) > 1$, thus cannot induce a proper probability distribution.

Hence, we need a normalized PAT state to proceed to formulate the time-reversed PAT CV samping model. We define this normalized PAT state as $\rho_{\bar{n}}^{\mathrm{nPAT}} = \frac{1}{\mathbf{Tr}(\rho_{\bar{n}}^{\mathrm{PAT}})}\rho_{\bar{n}}^{\mathrm{PAT}}$ and claim that we have seen this in the previous section, as it coincides with the thermalized single-photon state $\rho_{\bar{n}}^{\mathrm{TSP}}$. This is stated with the proposition below.

Proposition 7.3.5. *For a mean photon number $\bar{n}$, we have $\rho_{\bar{n}}^{\mathrm{nPAT}} = \rho_{\bar{n}}^{\mathrm{TSP}}$.*

Proof. This follows directly from the normalizing coefficient of $\rho_{\bar{n}}^{\mathrm{PAT}}$ shown in (7.3.3). Therefore we can rewrite the normalized PAT state as

$$\rho_{\bar{n}}^{\mathrm{nPAT}} = \frac{1}{\mathbf{Tr}(\rho_{\bar{n}}^{\mathrm{PAT}})}\rho_{\bar{n}}^{\mathrm{PAT}} = (1-\tanh^2(r))\rho_{\bar{n}}^{\mathrm{PAT}}.$$

From [**Theorem 7.3.4**], the equivalence between the normalized PAT state and the thermalized single-photon state directly follows. □

This implies that the physical realization of a normalized PAT state can be done by applying thermal noise (represented by the thermalization operator T) to the state $|1,0\rangle$ and tracing out the second mode. Alternatively, we can also express the normalized PAT state as

$$\begin{aligned}\rho_{\bar{n}}^{\mathrm{nPAT}} &= (1-\tanh^2 r)^2 \sum_{n\in\mathbb{N}} (\tanh^2 r)^n (n+1)|n+1\rangle\langle n+1| \\ &= \frac{(1-\tanh^2 r)^2}{\tanh^2 r} \sum_{n\in\mathbb{N}} (\tanh^2 r)^n n|n\rangle\langle n| \\ &= \frac{1}{\bar{n}(1+\bar{n})} \sum_{n\in\mathbb{N}} \left(\frac{\bar{n}}{1+\bar{n}}\right)^n n|n\rangle\langle n| \end{aligned} \tag{7.3.4}$$

where we again use the relation $\frac{1}{\tanh^2 r} = 1+\sinh^2 r = 1+\bar{n}$ to obtain the last equality.

We will now state an imperfect on-off photodetector with some probability of dark count and photon-loss as the Photodetector modeled in [BPP98] that is closely related to a measurement which POVM element coincides with $\rho_{\bar{n}}^{\mathrm{nPAT}}$. Now, the POVM element of the imperfect on-off photodetector in [BPP98] is

$$\tilde{\Pi}_k(v,\eta) = \sum_{n=0}^{k} \mathbb{P}_{\mathrm{dc}}(k-n) \sum_{j=n}^{\infty} \mathbb{P}_{\mathrm{ef}}(n|j)|j\rangle\langle j| \tag{7.3.5}$$

where $\mathbb{P}_{\mathrm{dc}}(k-n)$ is the probability of having $k-n$ dark counts, and $\mathbb{P}_{\mathrm{ef}}(n|j) = \eta^n(1-\eta)^{j-n}\binom{j}{n}$ the probability of detecting n photons given actually having j photons present, i.e: losing $j-n$ photons. This POVM is defined for each photon number count, producing a POVM element for each $k \in \mathbb{N}$, although we need a measurement that have two outputs, either

"on" or "off". Therefore to only have two different states of the photodetector, representing a click and no click[3], we define the following POVM.

Definition 7.3.6 (Imperfect d-photon On-Off Photodetector POVM)**.** For *Imperfect d-photon On-Off Photodetector* POVM $\Pi^{(d)} = \{\Pi_k^{(d)}\}_{k\in\{\text{on,off}\}}$, we have the POVM elements

$$\Pi_{\text{on}}^{(d)} = \mathbb{P}_{\text{dc}}(0)\sum_{j=d}^{\infty}\mathbb{P}_{\text{ef}}(d|j)|j\rangle\langle j|$$

and

$$\Pi_{\text{off}}^{(d)} = I - \Pi_d,$$

where $\mathbb{P}_{\text{dc}}(0)$ is the probability of registering no dark count. ◀

The imperfect d-photon on-off photodetection can be physically interpreted as a photodetector where one can be sure that there are d photons or more when it is in the "on" state[4]. This is apparent by noticing that

$$\langle j|\Pi_{\text{on}}^{(d)}|j\rangle = \begin{cases} 0 & \text{, if } j < d \\ \mathbb{P}_{\text{dc}}(0)\mathbb{P}_{\text{ef}}(d|j) & \text{, if } j \geq d \end{cases}$$

and

$$\langle j|\Pi_{\text{off}}^{(d)}|j\rangle = \begin{cases} 1 & \text{, if } j < d \\ 1 - (\mathbb{P}_{\text{dc}}(0)\mathbb{P}_{\text{ef}}(d|j)) & \text{, if } j \geq d \end{cases}$$

where if the photodetector is in the "on" state, then, there is 0 probability of having less than d photons.

However, it does not seem to be obvious how $\Pi^{(d)}$ is related to $\tilde{\Pi}(v,\eta)$. To reveal this relationship, first notice that when $d = 2$ we have

$$\tilde{\Pi}_1(v,\eta) = \mathbb{P}_{\text{dc}}(1)\left(\sum_{j=0}^{\infty}\mathbb{P}_{\text{ef}}(0|j)|j\rangle\langle j|\right) + \mathbb{P}_{\text{dc}}(0)\left(\sum_{j=1}^{\infty}\mathbb{P}_{\text{ef}}(1|j)|j\rangle\langle j|\right),$$

where the second term is precisely $\Pi_{\text{on}}^{(1)}$. This easily generalizes to the fact that $\Pi_{\text{on}}^{(d)}$ for all $d \in \mathbb{N}$ is the d-th term of the expansion of the first sum of $\tilde{\Pi}_d(v,\eta)$ in (7.3.5). We note that this imperfect on-off photodetection model is also adopted in [RRC16] and more recently in [Qi+20] for the case of $\Pi_{\text{on}}^{(0)}$. It was also noted in [RRC16] that the POVM element $\Pi_{\text{on}}^{(0)}$ coincides with the density operator of the Thermal state.

We now consider the case of Imperfect 1-photon On-Off Photodetector Π_1, in which relevance to PATS CV-sampling will be discussed later in more detail. Let $\mathbb{P}_{|k\rangle}^{(1)}(\text{on}) =$

[3]We will, at times refer to the event of our photodetector registers the "on" state as a "click" and the "off" state as a "no click" according to the POVM subject to the discussion, regardless if there is actually a photon in the mode being measured or if it is a dark count.

[4]To say this in a more statistical terms: The chance of a "False Positive", or the chance of falsely measuring d photons or more when there is actually less than d, is 0.

$\langle k|\Pi_{\text{on}}^{(1)}|k\rangle$, then we have

$$\mathbb{P}_{|k\rangle}^{(1)}(\text{on}) = \begin{cases} 0 & \text{, if } k = 0 \\ \mathbb{P}_{\text{dc}}(0)\eta & \text{, if } k = 1 \\ \mathbb{P}_{\text{dc}}(0)\eta(1-\eta)^{k-1}k & \text{, if } k \geq 2 \end{cases}$$

and

$$\mathbb{P}_{|k\rangle}^{(1)}(\text{off}) = \begin{cases} 1 & \text{, if } k = 0 \\ 1 - \mathbb{P}_{\text{dc}}(0)\eta & \text{, if } k = 1 \\ 1 - \mathbb{P}_{\text{dc}}(0)\eta(1-\eta)^{k-1}k & \text{, if } k \geq 2 \end{cases}.$$

Clearly, when we have an ideal photodetector $\Pi_{\text{on}}^{(1)}$, i.e. $\mathbb{P}_{\text{dc}}(0) = 1$ and $\eta = 1$, we have a photodetector that only clicks when there is a single photon, as $\mathbb{P}_{|k\rangle}^{(1)}(\text{on})$ equals to 0 for all $k \in \mathbb{N}$ except for $k = 1$. We now state the relation of this $\Pi^{(1)}$ measurement and the PAT state.

Proposition 7.3.7. *For normalized Photon-added Thermal State $\rho_{\bar{n}}^{\text{nPAT}}$ where $\bar{n} = \sinh^2 r$, Imperfect On-Off Photodetector POVM element*

$$\Pi_{\text{on}}^{(1)} = \mathbb{P}_{\text{dc}}(0) \sum_{j=1}^{\infty} \mathbb{P}_{\text{ef}}(1|j)|j\rangle\langle j|l$$

coincides with $\rho_{\bar{n}}^{\text{nPAT}}$, where we have $\mathbb{P}_{\text{dc}}(0) = 1 - \tanh^2 r$ and $\mathbb{P}_{\text{ef}}(1|j) = (1 - \tanh^2 r)(\tanh^2 r)^{j-1}j$.

Proof. For shorthand, let $t_r = \tanh r$. We first restate the definition of normalized PAT state,

$$\rho_{\bar{n}}^{\text{nPAT}} = (1 - t_r^2) \sum_{n \in \mathbb{N}} (1 - t_r^2)(t_r^2)^{n-1} n|n\rangle\langle n|.$$

Then, setting $\eta = 1 - t_r^2$ and $\mathbb{P}_{\text{dc}}(0) = 1 - t_r^2$ will give us

$$\rho_{\bar{n}}^{\text{nPAT}} = \mathbb{P}_{\text{dc}}(0) \sum_{n=1}^{\infty} \eta(1-\eta)^{n-1} n|n\rangle\langle n|$$

where the coefficients in the sum is precisely $\mathbb{P}(1|n)$, concluding the proof. □

We note that the POVM $\Pi_{\text{on}}^{(1)}$ can be interpreted as a photodetector that clicks on detecting a single photon with detection efficiency $\eta = 1 - t_r^2$ when there is no dark count. Therefore one can be certain that the system will have 1 or more photons present when it clicks because a measurement $\Pi_{\text{on}}^{(1)}$ on a state $|0\rangle$ has 0 probability of clicking, as can be seen from $\langle 0|\Pi_{\text{on}}^{(1)}|0\rangle = 0$. It is also worth noticing that POVM $\tilde{\Pi}_{\text{on}}^{(0)} = \mathbb{P}_{\text{dc}}(0) \sum_{n \in \mathbb{N}} (1-\eta)^n |n\rangle\langle n|$ coincides with thermal state $\rho_{\bar{n}}^{\text{Th}} = (1 - \tanh^2 r) \sum_{n \in \mathbb{N}} (\tanh^2 r)^n |n\rangle\langle n|$, for $\mathbb{P}_{\text{dc}}(0) = 1 - \tanh^2 r$ and $\eta = 1 - \tanh^2 r$.

Now, we consider an m-mode system where POVM measurement $\Pi_j^{(1)} = \{\Pi_{\text{on},j}^{(1)}, \Pi_{\text{off},j}^{(1)}\}$ is done on the j-th mode, and with POVM element for a click on the j-th mode as $\Pi_{\text{on},j}^{(1)}$

and for the whole system as $\Pi_{\text{on}}^{(1)} = \bigotimes_{j=1}^{m} \Pi_{\text{on},j}^{(1)}$. We can further express $\Pi_{\text{on}}^{(1)}$ in terms of the m-mode number state operator $|n\rangle\langle n| = \bigotimes_{j=1}^{m} |n_j\rangle\langle n_j|$ for $n_j \in \mathbb{N}$ and $n = (n_1, \ldots, n_m)$ for shorthand as follows

$$\begin{aligned}
\Pi_{\text{on}}^{(1)} &= \bigotimes_{j=1}^{m} \sum_{n_j \in \mathbb{N}} (1 - t_{r_j}^2)^2 (t_{r_j}^2)^{n_j - 1} n_j |n_j\rangle\langle n_j| \\
&= \sum_{n_1 \in \mathbb{N}} (1 - t_{r_1}^2)^2 (t_{r_1}^2)^{n_1 - 1} n_1 \left(|n_1\rangle\langle n_1| \otimes \left(\bigotimes_{j=2}^{m} \sum_{n_j \in \mathbb{N}} (1 - t_{r_j}^2)^2 (t_{r_j}^2)^{n_j - 1} n_j |n_j\rangle\langle n_j| \right) \right) \\
&= \sum_{n \in \mathbb{N}^m} \left(\prod_{j=1}^{m} (1 - t_{r_j}^2)^2 (t_{r_j}^2)^{n_j - 1} n_j \right) |n\rangle\langle n|
\end{aligned}$$

by the tensor product property $(\sum_j c_j |a_j\rangle) \otimes |b\rangle = \sum_j c_j (|a_j\rangle \otimes |b\rangle)$ and $|a\rangle \otimes (\sum_j c_j |b_j\rangle) = \sum_j c_j (|a\rangle \otimes |b_j\rangle)$ for squeezing parameter for each mode $r = (r_1, \ldots, r_m)$. Therefore, the probability of a click from POVM measurement $\Pi_{\text{on}}^{(1)}$ on an arbitrary state ρ is

$$\mathbb{P}_{\rho}^{(1)}(\text{on}) = \mathbf{Tr}(\rho \Pi_{\text{on}}^{(1)}) = \sum_{n \in \mathbb{N}^m} \left(\prod_{j=1}^{m} \xi_j(r, n) \right) \mathbf{Tr}(\rho |n\rangle\langle n|) \tag{7.3.6}$$

where $\xi_j(r, n) = (1 - t_{r_j}^2)^2 (t_{r_j}^2)^{n_j - 1} n_j$ is the probability of measuring n_j photons in the j-th mode, and as $\mathbf{Tr}(\rho |n\rangle\langle n|) =: \mathbb{P}_{\rho}(n)$ is the probability of measuring a photon pattern n on ρ, we can interpret $\mathbf{Tr}(\rho \Pi^{(1)})$ as a mixture of probability of measuring all possible photon pattern on ρ. However, it is worth noting that $\prod_j \xi_j(r, n) = 0$ if $\exists j$ such that $n_j = 0$, indicating that the probability of measuring 0 photons in any of the m modes is 0, thus does not contribute to the mixture. We are able to more precisely write the probability of a click as

$$\mathbf{Tr}(\rho \Pi_{\text{on}}^{(1)}) = \sum_{n \in \mathbb{N}_{>0}^m} \left(\prod_{j=1}^{m} \xi_j(r, n) \right) \mathbf{Tr}(\rho |n\rangle\langle n|)$$

by exchanging $\mathbb{N}$ with $\mathbb{N}_{>0}^m = \{n \in \mathbb{N}^m : \forall j, n_j > 0\}$.

Finally, reversing our PAT CV sampling model as it was done for the CVS model, we will obtain its output probability as

$$\mathbb{P}_{\text{PAT-CV}}(0) = \mathbf{Tr}(Q \rho_{\vec{n}}^{\text{nPAT}} Q^{\dagger} |0, r\rangle\langle 0, r|) = \mathbb{P}_{\rho}^{(1)}(\text{on}) = \mathbf{Tr}(\rho \Pi_{\text{on}}^{(1)}) \tag{7.3.7}$$

where $|0, r\rangle\langle 0, r| = |r\rangle\langle r|$ is the POVM element of the Heterodyne measurement[5] projecting to a vacuum state (equivalently, coherent state with amplitude $\alpha = 0$), and $\rho = Q^{\dagger} |r\rangle\langle r| Q$. This is illustrated more clearly in [**Figure 7.2**] where we indicate the measurement outcome probability as $\mathbb{P}_{\mathit{R}\text{PAT-CV}}(\text{k})$ where k is a vector of m on/off values indicating outputs from all m detectors.

[5]For the definition of the POVM element of Heterodyne measurement, see [**Definition 2.2.21**].

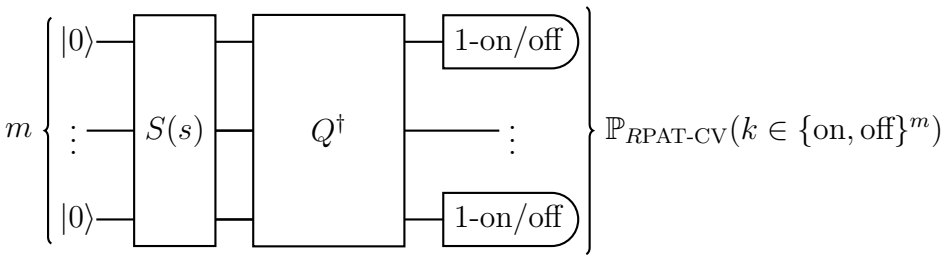

Figure 7.2: Reversed Photon-added Thermal State Continuous-Variable sampling (RPAT-CV) circuit which coincides with GBS with 1-photon on-off photodetector, indicated by "1-on/off" in picture.

Noisy normalized Photon-Added Thermal State

We now consider imperfection in the preparation of $\rho_{\bar{n}}^{\text{nPAT}}$ state by modelling it as an ideal $\rho_{\bar{n}}^{\text{nPAT}}$ going through a noisy channel, i.e: $\rho_{\bar{n}}^{\text{nPAT}}$ as the first input to a beamsplitter with transmissivity $\eta \in [0,1]$ with vacuum $|0\rangle$ going into the second input port, and then taking only the first output port. More precisely, we write it as

$$\rho_{\bar{n},\eta}^{\text{nPAT}} = \mathbf{Tr}_d \left(B(\eta) \rho_{\bar{n}}^{\text{nPAT}} \otimes |0\rangle_{dd}\langle 0| B^\dagger(\eta) \right)$$

and say that $\rho_{\bar{n},\eta}^{\text{nPAT}}$ is the nPAT state with efficiency η.

Proposition 7.3.8. *Given mean photon number $\bar{n} = \sinh^2 r$, we can write the* nPAT *state with efficiency η as*

$$\rho_{\bar{n},\eta}^{\text{nPAT}} = \mathbb{P}_{\text{dc}}(0) \sum_{n=1}^{\infty} \mathbb{P}_{\text{ef}}(1|n) \sum_{j=0}^{n} b_{n,j}(\eta) |j\rangle\langle j|$$

where $\mathbb{P}_{\text{dc}}(0)$ is the probability of no dark count, $\mathbb{P}_{\text{ef}}(1|n) = (1-t_r^2)(t_r^2)^{n-1} n$ is the probability of detecting one photon given having n actually present, and $b_{n,j}(\eta) = \binom{n}{j} \eta^j (1-\eta)^{n-j}$ is the binomial distribution.

The proof of this proposition can be directly obtained by noticing that the tensor product with $|0\rangle_{dd}\langle 0|$, the beamsplitter $B(\eta)$, and partial trace $\mathbf{Tr}_d$ only affects the $|n\rangle\langle n|$ terms in the expression of $\rho_{\bar{n}}^{\text{nPAT}}$ shown in (7.3.4). Further recalling that $\mathbf{Tr}(B(\eta)|n,0\rangle\langle n,0|B^\dagger(\eta)) = \sum_{j=0}^{n} b_{n,j}(\eta)|j\rangle\langle j|$ gives us the desired expression above.

Now, we consider the reversal of the noisy PAT CV sampling, which is essentially the PAT CV sampling model with efficiency η imposed on the input PAT states. This will get us to a relation that is very similar to (7.3.7), which is by simply replacing the $\Pi_{\text{on}}^{(1)}$ by a POVM element that coincides with the PAT state with efficiency η, that is $\Pi_{\text{on},\eta}^{(1)} = \rho_{\bar{n},\eta}^{\text{nPAT}}$. For $\eta > 0$, we will no longer have the comfort of the guarantee of a false "on" measurement due to

$$\langle 0|\Pi_{\text{on},\eta}^{(1)}|0\rangle = \mathbb{P}_{\text{dc}}(0) \sum_{n=1}^{\infty} \mathbb{P}_{\text{ef}}(1|n) b_{n,0}(\eta). \tag{7.3.8}$$

Thus, we claim that this is a more realistic model, one which may reveal to us the quantum-classical computational complexity boundary in the spectrumof η for this noisy PAT CV sampling model.

Reflecting back to [**Section 7.2**], we again have the outcome probability measurement in terms of a sum of the probabilities for the noiseless case. In the above formulation, given that ρ is an GBS output, that is squeezed states going through a unitary interferometer, we can write $\mathbf{Tr}(\rho_{out}|n\rangle\langle n|) = \frac{1}{\sqrt{\mathbf{Det}(\sigma_Q)}}|\mathbf{Haf}(B_n)|^2$ as in [**Section 5.1**], again giving us a weighted sum of hafnians. Due to this fact, we again restate the similarity with proven results of sufficient noise conditions for Boson Sampling, such as in [AB16] and [Shc19] where the outcome probability is analogously to what we have, a weighted sum of permanents.

However, we again admit that we fall short of a proof that shows a sufficient noise condition for the PATS CV sampling (or its GBS equivalent), and we again note that a natural approach will be to use an oracle that approximates the probability of the measurement outcome of the PATS CV sampling circuit, then construct an algorithm that sits somewhere inside the polynomial hierarchy that can approximate the $|\mathbf{GHE}|^2_{\pm}$ by using the oracle. This argument that is rather ubiquitous will then show the classical hardness under the assumption that the polynomial hierarchy does not collapse. However, in the case of PATS CV sampling, a difficulty from having a rather non-standard formulation of the POVM may pose a challenge, especially with its ability to perfectly distinguish a state with 0 and 1 photon in the noiseless case. These possible avenues in showing a sufficient noise condition for PATS CV sampling may therefore be an interesting future work.

Chapter 8

Conclusion

The exhilarating search for an answer on whether computers as we know it today can feasibly simulate any physical system remains a task for Computer Scientists and Physicists alike. However, remarkable advances in both theoretical and experimental research of quantum computation in the past decade has been raising the stakes. With the emergence of theoretical formulation of sub-universal quantum computation model from the persistent challenges in the implementation of a universal quantum computer, the sub-universal Gaussian Boson Sampling (GBS) model stands out as one of the most promising avenue to demonstrate the impossibility of efficiently simulating a quantum system classically. Moreover, given its sub-universal nature, GBS may also act as a tool in investigating the boundary between quantum computational models that cannot be efficiently simulated classically and those that can be efficiently simulated.

In [**Section 6.2**] of this thesis, we have discussed and presented the method of quantum Homodyne Tomography to construct a classical algorithm that can simulate an approximate GBS by reconstructing the GBS probability distribution using samples taken from a Gaussian distribution. We call this algorithm the Pattern Function GBS (PF GBS) algorithm from the name of a special function we use to process the samples. In [**Section 6.3**], we then derive a statistical criteria for the samples that guarantees an error bound for the PF GBS approximation to support its practical use. Nevertheless, a persistent problem of negative probability values in the approximated distribution remains unsolved, and is left for future work in improving the PF GBS algorithm. Also, computer simulation results evidently shows a persistent boundary to simulate the approximation with an indefinite precision due to sharp increase in the amount of data needed to obtain better accuracy.

With the apparent challenges in feasibly simulating GBS classically using the PF GBS algorithm, in [**Chapter 7**] we resort to investigating the case of classically simulating GBS that incurs noise to further probe the quantum-classical computational complexity boundary within the GBS paradigm. Based on existing results, in [**Section 7.1**] we have discussed the existence of efficient classical simulation of a noisy GBS model from analyzing the presence of negativity in the quasi-probability distribution (QPD) representations of the GBS system, resulting in a noise condition that must be satisfied for a GBS experiment to be hard to classically simulate. The presence of negativity in the QPD representation of a quantum state has been an extensively studied measure of non-classicality of a quantum state and has been considered as indication of resource for quantum computation. Then insipired by an

existing noise model for Boson Sampling, the prototypical model of GBS, in [**Section 7.2**] we have proposed a model of GBS with noises imposed on its measurement to attempt to show a sufficient noise condition that a GBS experiment can satisfy for it to be hard to simulate classically. Although we came short on showing a sufficient noise condition for this model, it bears a resemblance to the noisy Boson Sampling model. Thus, we conjecture that a similar proof technique to the noisy Boson Sampling may show a sufficient noise condition for the noisy GBS. However, further attempts in constructing this proof for our proposed model is left for future work.

Lastly in [**Section 7.3**], to further investigate the quantum-classical computational boundary, we propose a quantum computational model that involves a more distinct quantum states compared to those used in GBS. We call this model the Photon-added Thermal State Continuous-Variable sampling (PATS CVS) from the use of Photon-added Thermal (PAT) states as inputs in this model. Our considerations in using PAT state were based on the non-classicality analysis of the noisy PAT state that shows the transition between negative and non-negative QPD of PAT state in the noise spectrum, showing us a promise of a path to the quantum-classical computational boundary for the PATS CVS model. We have derived a formulation of PATS CVS that is closely related to the noisy GBS model that we proposed and the noisy Boson Sampling model that we have discussed. Despite coming short of a proof in showing the quantum-classical computational boundary for the PATS CVS model within the noise spectrum of the PAT state, we again conjecture that a similar argument to the noisy Boson Sampling model can provide a closer view to this boundary. Though, further investigations into the construction of this proof is left for future work.

Along with the rigorous argument for the complexity of GBS, our evidences that substantially many data are required to simulate GBS with the HT method supports the impossibility of efficiently simulating GBS classically. It is therefore natural to expand this investigation to the realm of noisy GBS to probe the quantum-classical computational complexity of GBS. With the proposal of these models to classically simulate GBS and for a GBS-related quantum computation, it is hoped that more work can be done in the future to improve these models, which holds the potential to obtain better understanding of the nature of sub-universal quantum computation and the possibility of classically simulating it in the presence of noise. The further development of these models can be beneficial for experimentalists alike to validate and benchmark experiment setups suffering unavoidable noise, especially, given the reasonable feasibility of the implementation of these models (such as squeezed states of GBS and PAT states). Ultimately from the entirety of this work, it is hoped that it will bring inspirations for experienced scientists and students of science alike to pursue the task of probing into the unseen limit of computation in the proximity of its physical realization to extend the limit of our computational and physical understanding.

Appendix A

Complete PF GBS Computer Simulation Results

In this appendix, we present more detailed plots of PF GBS computer simulations, which are presented and discussed more concisely in [**Section 6.2**]. Here, we also group them based on the type of interferometer (Random Orthogonal, Random Unitary, or Square Interferometer) and by the type of value they plot. For the simulation runs that use random unitary and random orthogonal we show:

1. in [**Section A.1**], plots of the $\mathbb{P}_{\texttt{PG}}$ values as they change with increasing amount of data, in comparison with the "true" probability $\mathbb{P}_{\texttt{GBS}}$ calculated using the hafnian,

2. in [**Section A.2**], plots of the standard deviation errors over number of homodyne data,

3. in [**Section A.3**], plots of the "actual" error $|\mathbb{P}_{\texttt{PG}}(k) - \mathbb{P}_{\texttt{GBS}}(k)|$ over the number of homodyne data.

In all of these plots, we show how the value changes as the number of data increases for each of the 10 runs for both randomly generated unitary or orthogonal interferometer.

Finally in [**Section A.4**], for the simulations using square interferometers, we present a typical convergence plot of $\mathbb{P}_{\texttt{PG}}$ to $\mathbb{P}_{\texttt{GBS}}$ and its standard deviation value as the number of homodyne data increases. We also show a plot of convergence of $\mathbb{P}_{\texttt{PG}}$ to $\mathbb{P}_{\texttt{GBS}}$ for simulation runs with different number of modes.

A.1 Convergence of $\mathbb{P}_{\mathtt{PG}}$ to $\mathbb{P}_{\mathtt{GBS}}$ for Random Interferometer GBS

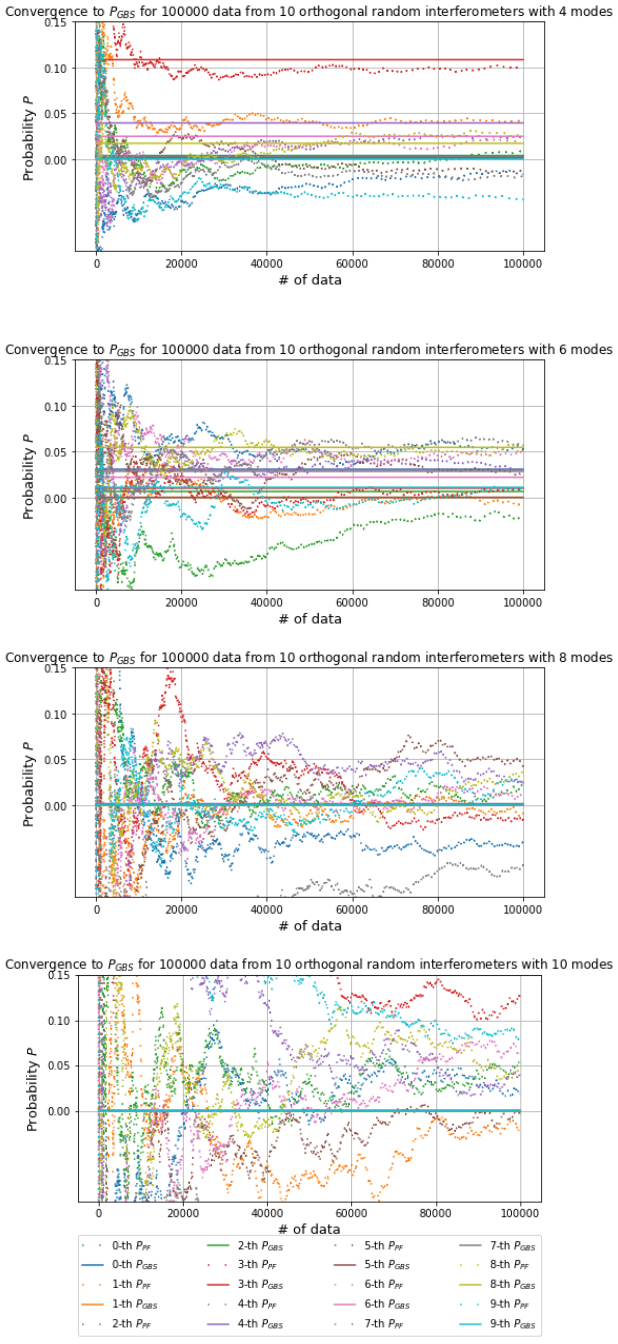

Figure A.1: Convergence of approximation $\mathbb{P}_{\mathtt{PG}}$ to $\mathbb{P}_{\mathtt{GBS}}$ for random orthogonal interferometers for GBS with 4, 6, 8, and 10 modes. Legend for all plots is located at the very bottom.

For the four plots showing the convergence of $\mathbb{P}_{\mathtt{PG}}$ to $\mathbb{P}_{\mathtt{GBS}}$ for GBS with orthogonal interferometer [**Figure A.1**] and GBS with unitary interferometer [**Figure A.2**], we show the true probability $\mathbb{P}_{\mathtt{GBS}}(k)$ for a randomly generated interferometer as a solid line, whereas its approximation $\mathbb{P}_{\mathtt{PG}}(k)$ is shown as the dotted line with the same color in each plots. The number of mode m of the simulated GBS setup in each plot differs for $m \in \{4, 6, 8, 10\}$. The plots located lower in both [**Figure A.1**] and [**Figure A.2**] are the plots for simulations with more modes as can be seen from the plot titles.

Following are the input arrangement and output probability for each simulation according to mode number:

- For the 4 mode simulations, we have $k = (0, 1, 1, 0)$ and 2 single mode squeezed state (SMSS) inputs with $r = 0.5$.
- For 6 mode simulations, we have $k = (0, 0, 1, 1, 0, 0)$ and 4 SMSS inputs with $r = 0.5$.
- For 8 mode simulations, we have $k = (0, 0, 1, 1, 0, 0, 1, 1)$ and 4 SMSS inputs with $r = 0.5$.
- Lastly for 10 mode simulations, we have $k = (0, 0, 1, 1, 0, 0, 1, 1, 0, 0)$ and 4 SMSS inputs with $r = 0.5$.

We generated 10 random orthogonal and 10 random unitary interferometers, thus there are 10 pairs of lines in each plot with each pair representing the convergence of $\mathbb{P}_{\mathtt{PG}}$ to $\mathbb{P}_{\mathtt{GBS}}$.

Despite having the average standard deviation and average true error $|\mathbb{P}_{\mathtt{PG}}(k) - \mathbb{P}_{\mathtt{GBS}}(k)|$ in [**Figure 6.3**] and [**Figure 6.4**]

for the orthogonal and unitary interferometers not behaving very differently, the top two plots in both [**Figure A.1**] and [**Figure A.2**] for 4 and 6 mode simulations seems to slightly differ. The deviations of the PF-GBS approximation $\mathbb{P}_{\mathtt{PG}}$ form the true value $\mathbb{P}_{\mathtt{GBS}}$ seems to be more in the case of with unitary interferometer. Some of the approximation $\mathbb{P}_{\mathtt{PG}}$ for the probabilities $\mathbb{P}_{\mathtt{GBS}}$ that are relatively larger for the orthogonal interferometers in [**Figure A.1**] seems to be converging to the true probabilities $\mathbb{P}_{\mathtt{GBS}}$ as the data reaches 100000. This is clear in the top-most plot of [**Figure A.1**] where the approximations for the 3rd interferometer (red line), which the true value is a little over 0.1, and other three interferometers with values between 0.025 and 0.05 (purple, pink, and yellow) are less deviated from the true values, while the ones with true value very close to 0 experiencing worse deviations.

In contrast, approximations $\mathbb{P}_{\mathtt{PG}}$ for the unitary interferometers in the top two plots of [**Figure A.2**] does not seem to be converging to the true probabilities $\mathbb{P}_{\mathtt{GBS}}$ after the first 100000 Homodyne data. Nevertheless, in the cases with larger number of modes $m \in \{8, 10\}$ where the true probability values $\mathbb{P}_{\mathtt{GBS}}$ are very close to 0, approximations for both simulations with orthogonal and unitary interferometers seems to be similarly deviant throughout the first 100000 Homodyne data.

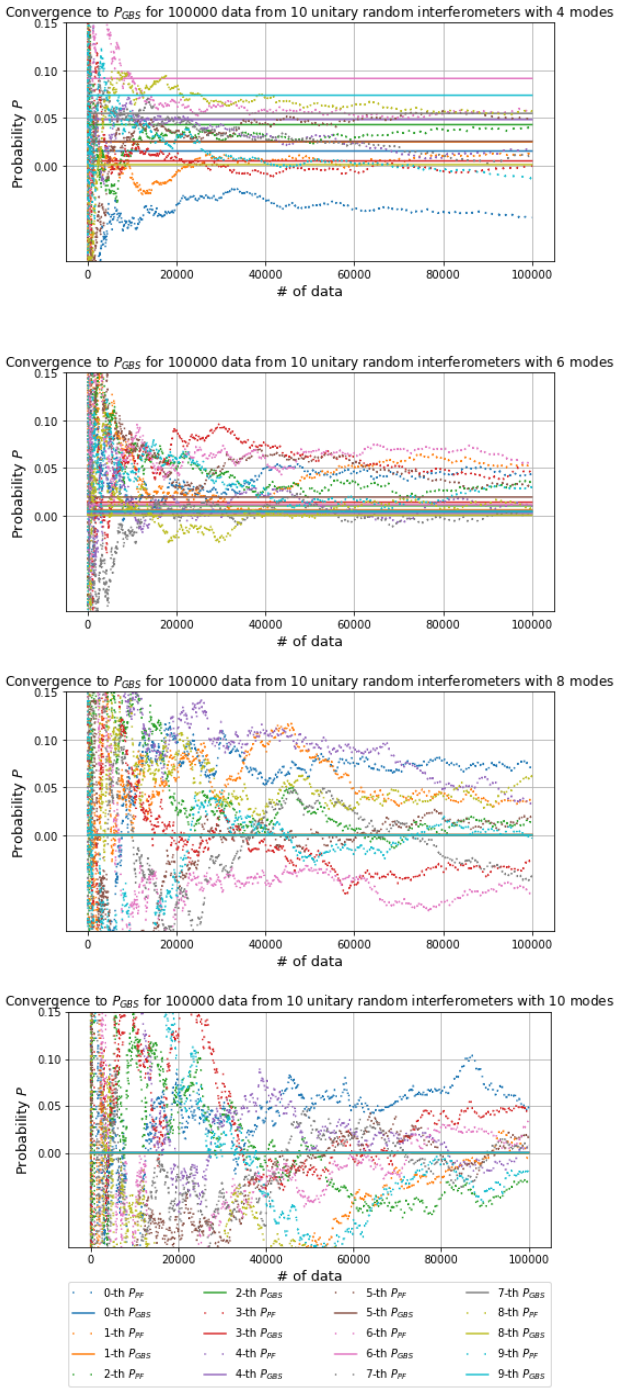

Figure A.2: Convergence of approximation $\mathbb{P}_{\mathtt{PG}}$ to $\mathbb{P}_{\mathtt{GBS}}$ for random unitary interferometers for GBS with 4, 6, 8, and 10 modes. Legend for all plots is located at the very bottom.

A.2 Standard deviation of $\mathbb{P}_{\mathtt{PG}}$ for Random Interferometer GBS

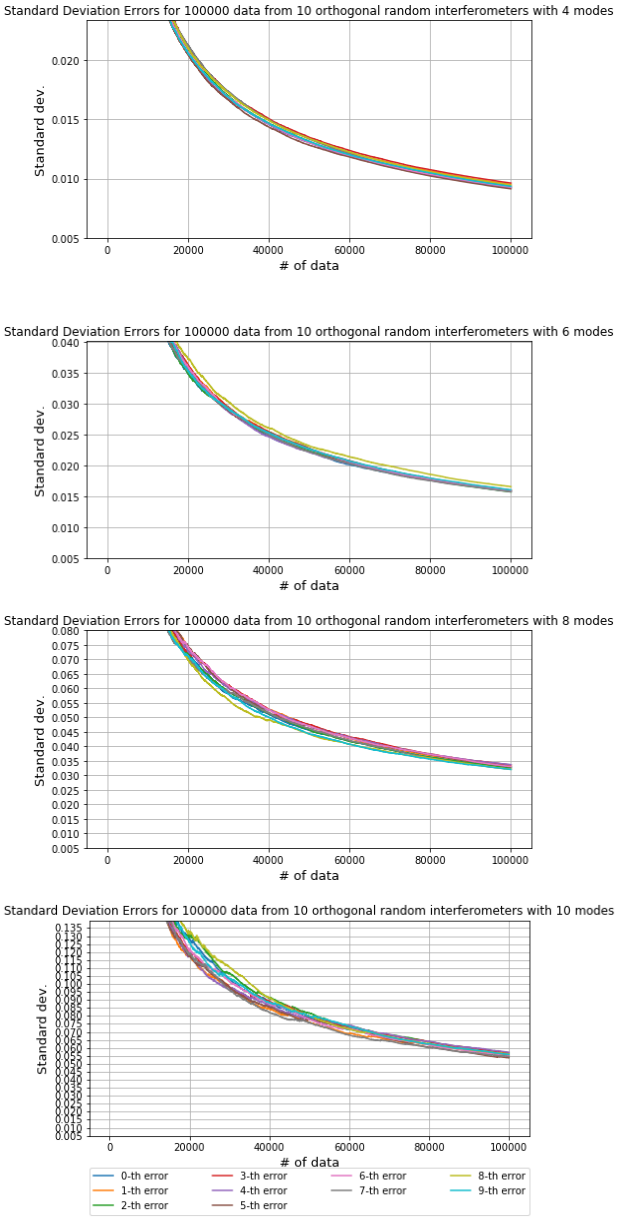

Figure A.3: Standard deviation of approximation $\mathbb{P}_{\mathtt{PG}}$ for random orthogonal interferometers in 4, 6, 8, and 10 mode GBS.Legend for all plots is located at the very bottom.

In this section, we show the change in standard deviation of approximation $\mathbb{P}_{\mathtt{PG}}(k)$ for GBS with orthogonal interferometer [**Figure A.3**] and GBS with unitary interferometer [**Figure A.4**]. As in [**Section A.1**], the PF-GBS simulations were ran with different number of modes m for $m \in \{4, 6, 8, 10\}$. Here, we also have the following input arrangement and output probability according to number of modes:

- For the 4 mode simulations, we have $k = (0, 1, 1, 0)$ and 2 single mode squeezed state (SMSS) inputs with $r = 0.5$.
- For 6 mode simulations, we have $k = (0, 0, 1, 1, 0, 0)$ and 4 SMSS inputs with $r = 0.5$.
- For 8 mode simulations, we have $k = (0, 0, 1, 1, 0, 0, 1, 1)$ and 4 SMSS inputs with $r = 0.5$.
- Lastly for 10 mode simulations, we have $k = (0, 0, 1, 1, 0, 0, 1, 1, 0, 0)$ and 4 SMSS inputs with $r = 0.5$.

Recalling the plots of the average standard deviation of $\mathbb{P}_{\mathtt{PG}}(k)$ for the simulations using orthogonal interferometers [**Figure 6.3**] and simulations using unitary interferometers [**Figure 6.4**], there does not seem to be any significant difference in terms of the rate they decreases. Similarly, standard deviations for all 10 randomly generated interferometers between the plots in [**Figure A.3**] and the plots in [**Figure A.4**] do not seem to for each node number $m \in \{4, 6, 8, 10\}$, as for each m, the standard deviations seems to end up in about the same value after the first 100000 Homodyne data.

This similarity can be clearly seen by observing that for the cases with $m = 4$, the top-most plots in [**Figure A.3**] and in [**Figure A.4**] show that the standard deviations end up at about 0.009. Also, the second plots with $m = 6$ show that the standard deviations end up at about 0.016. The third plots with $m = 8$ show that the standard deviations end up at about 0.033. Lastly, the bottom plots in both [**Figure A.3**] and [**Figure A.4**] with $m = 10$ show the standard deviation to end up at about 0.057.

However, one can observe that the bottom two plots in [**Figure A.4**] seems to vary more for different interferometers compared to the bottom two plots in [**Figure A.3**] within the region of 10000 Homodyne data to about 60000 Homodyne data. The cause of this considerably minor anomaly is unknown, thus require further analysis. Although it is always possible to obtain an "interesting" set of sampled random unitary interferometers that are very distinct from each other, one may intuitively infer that given the complex number entries of the unitary interferometer that consist of real and imaginary part, the space where one can sample a unitary interferometer is larger, causing more uncertainty in the sampled interferometer.

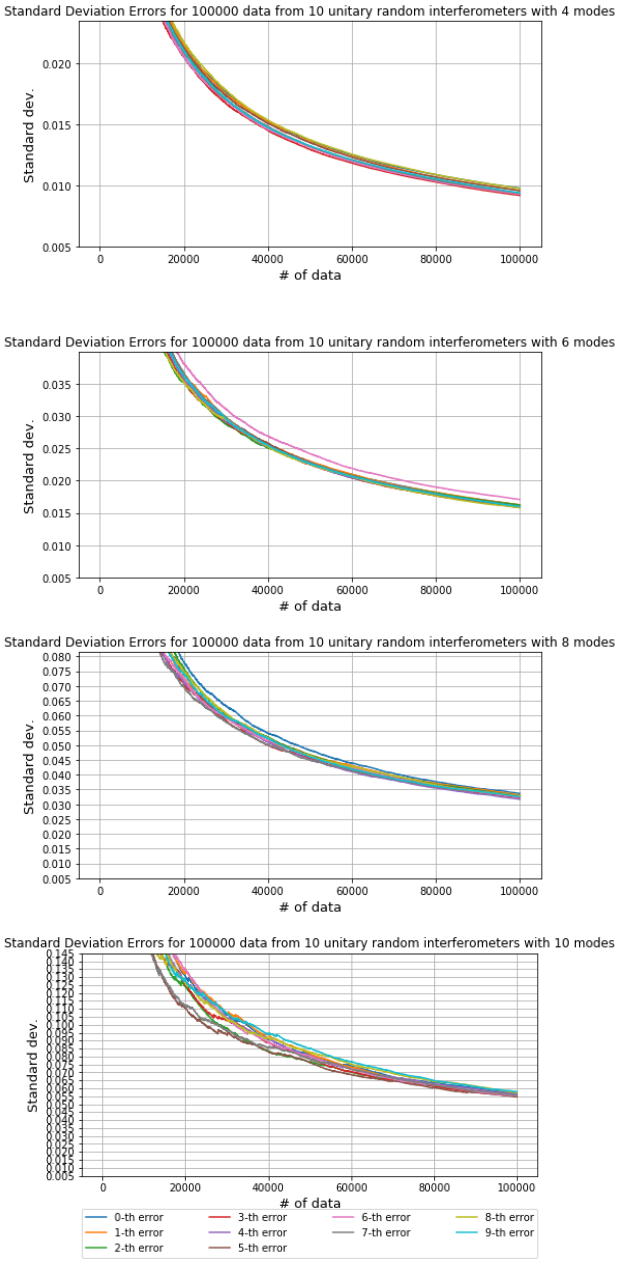

Figure A.4: Standard deviation of approximation $\mathbb{P}_{\mathtt{PG}}$ for random unitary interferometers in 4, 6, 8, and 10 mode GBS.Legend for all plots is located at the very bottom.

A.3 True error of $\mathbb{P}_{\mathtt{PG}}$ with Random Interferometers

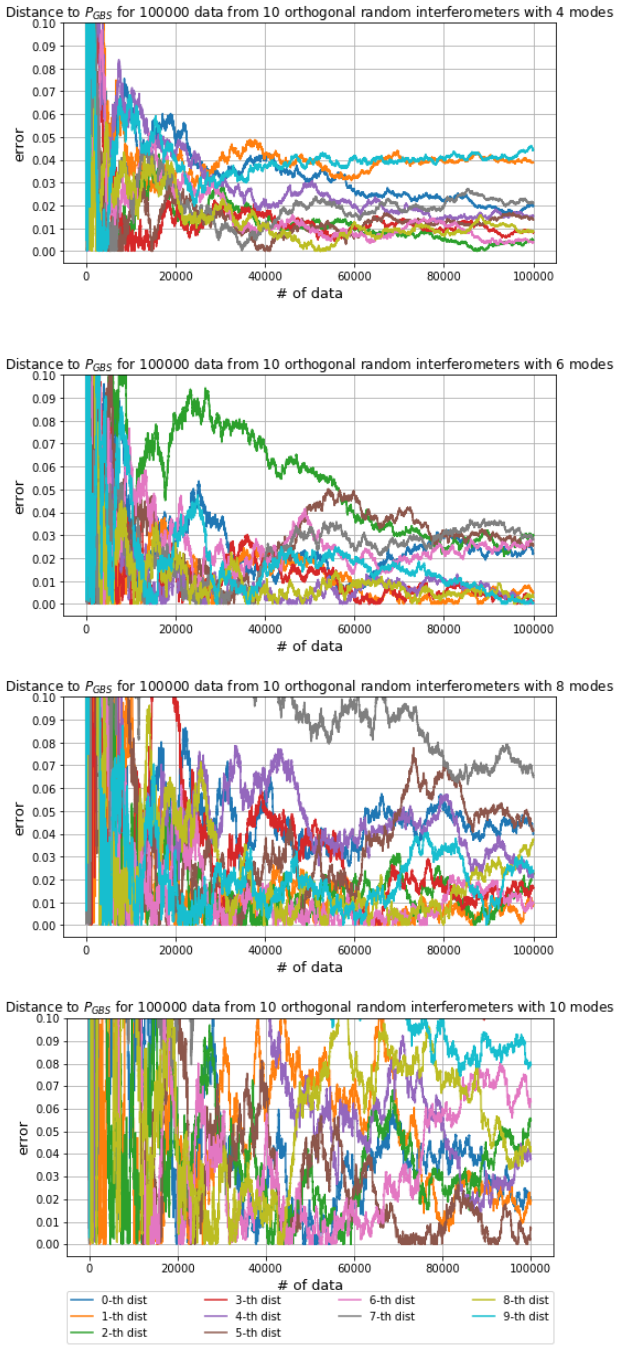

Figure A.5: Actual error $|\mathbb{P}_{\mathtt{PG}} - \mathbb{P}_{\mathtt{GBS}}|$ for random orthogonal interferometers in 4, 6, 8, and 10 mode GBS.Legend for all plots is located at the very bottom.

The plots we present in this section show the true error of the approximation $\mathbb{P}_{\mathtt{PG}}$, which is calculated by its Euclidean distance to $\mathbb{P}_{\mathtt{GBS}}$, that is $|\mathbb{P}_{\mathtt{PG}}(k) - \mathbb{P}_{\mathtt{GBS}}(k)|$. Similar to the previous sections, here we show the plots of the true errors for each of the simulation runs for 10 randomly generated interferometer for each different number of modes $m \in \{4, 6, 8, 10\}$. We show the true error plots for the simulation runs that uses orthogonal interferometer in [**Figure A.5**] and for simulation runs that uses unitary interferometer in [**Figure A.6**]. Also similarly, we are interested in the following input arrangement and output probability according to number of modes:

- For the 4 mode simulations, we have $k = (0, 1, 1, 0)$ and 2 single mode squeezed state (SMSS) inputs with $r = 0.5$.
- For 6 mode simulations, we have $k = (0, 0, 1, 1, 0, 0)$ and 4 SMSS inputs with $r = 0.5$.
- For 8 mode simulations, we have $k = (0, 0, 1, 1, 0, 0, 1, 1)$ and 4 SMSS inputs with $r = 0.5$.
- Lastly for 10 mode simulations, we have $k = (0, 0, 1, 1, 0, 0, 1, 1, 0, 0)$ and 4 SMSS inputs with $r = 0.5$.

Once again, we recall that the plot for the average of all 10 randomly generated interferometers for each number of modes $m \in \{4, 6, 8, 10\}$ for both orthogonal interferometer [**Figure 6.3**] and unitary interferometer [**Figure 6.4**] which seems to slightly differ in their average true errors. This is particularly apparent in the case of 4 modes (blue lines in [**Figure 6.3**] and

[**Figure 6.4**]) where the average true error for the unitary interferometers is significantly higher compared to the orthogonal interferometers. A distinction is also apparent in the 10 mode case (red lines in [**Figure 6.3**] and [**Figure 6.4**]) where the average true error for the unitary interferometers is significantly lower than the average true error of the orthogonal interferometers.

These notable differences in the average plots is apparent in the plots in [**Figure A.5**] and [**Figure A.6**]. The significantly higher average true error for the 4 mode unitary interferometers can be seen in the top plot of [**Figure A.6**] where some of the true error values does not seem to be decreasing within the first 100000 Homodyne data. This is contrary to the true errors for the 4 mode unitary interferometers in the top plot of [**Figure A.5**] where significant decrease of true error can be seen to happen in the first 20000 Homodyne data.

The second notable distinction between the average true errors of the 10 mode case of the orthogonal interferometers and unitary interferometers can also be observed in the bottom plots of [**Figure A.5**] and [**Figure A.6**]. The lines in the bottom plot of [**Figure A.5**] seems to be more wildly fluctuating throughout the 100000 Homodyne samples. In contrast, the bottom plot of [**Figure A.6**] fluctuates as first and seems to be decreasing after about the first 60000 Homodyne data. This distinction can also be seen from the true error value of the simulations in [**Figure A.6**] that falls withint the interval of $(0, 0.05)$, while the true errors of the simulations in [**Figure A.5**] falling within a wider interval of $(0, 0.08)$.

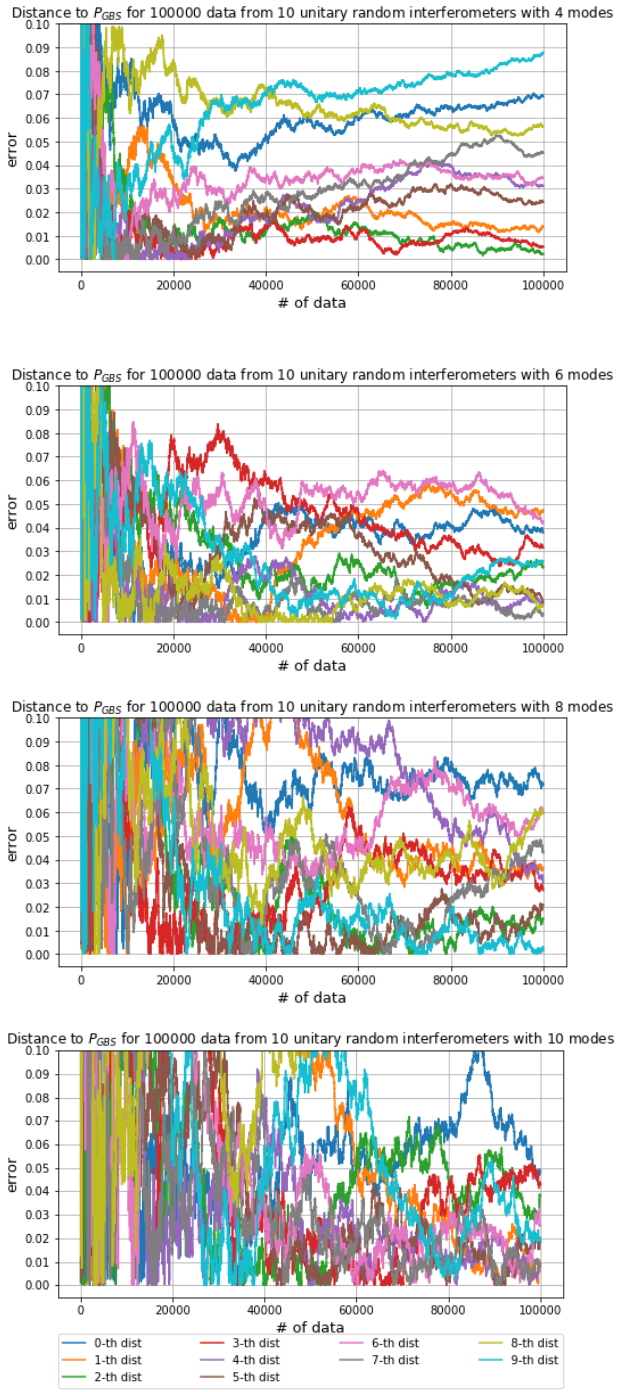

Figure A.6: Actual error $|\mathbb{P}_{\mathtt{PG}} - \mathbb{P}_{\mathtt{GBS}}|$ for random complex unitary interferometers in 4, 6, 8, and 10 mode GBS.Legend for all plots is located at the very bottom.

A.4 Simulations using Square Interferometers

In this section, we will present the plots describing our simulations of PF GBS approximation using a square interferometer as described in [**Section 6.2.1**] and illustrated in [**Figure 6.2**]. First, we present the plot in [**Figure A.7**] that shows the convergence of the approximation $\mathbb{P}_{\texttt{PG}}(k)$ to the true probability $\mathbb{P}_{\texttt{GBS}}(k)$ for different number of modes $m \in \{2, 4, 6, 8, 10\}$ and output $k = (1, 1, \ldots, 1)$. This plot is a complement to the plot in [**Figure 6.5**] that shows the corresponding standard deviation and true error $|\mathbb{P}_{\texttt{PG}}(k) - \mathbb{P}_{\texttt{GBS}}(k)|$ from the same 100000 Homodyne samples. In [**Figure A.7**], we show the approximations $\mathbb{P}_{\texttt{PG}}$ over the number of Homodyne data as solid lines and the true probabilities $\mathbb{P}_{\texttt{GBS}}$ as the dotted lines. As we have 5 different mode numbers m, there are 5 pairs of lines in the plot. However, it is difficult to distinguish in the plot the probabilities $\mathbb{P}_{\texttt{GBS}}(k)$ for 4 or more modes which are very close to 0 with

$$\begin{aligned}
\mathbb{P}_{\texttt{GBS}}(1,1,1,1) &\approx 7.95 \times 10^{-3} \\
\mathbb{P}_{\texttt{GBS}}(1,1,1,1,1,1) &\approx 1.13 \times 10^{-6} \\
\mathbb{P}_{\texttt{GBS}}(1^8) &\approx 1.44 \times 10^{-8} \\
\mathbb{P}_{\texttt{GBS}}(1^{10}) &\approx 4.35 \times 10^{-11}
\end{aligned}$$

where $\mathbb{P}_{\texttt{GBS}}(1^m)$ is the probability of measuring one photons in each of the m output modes.

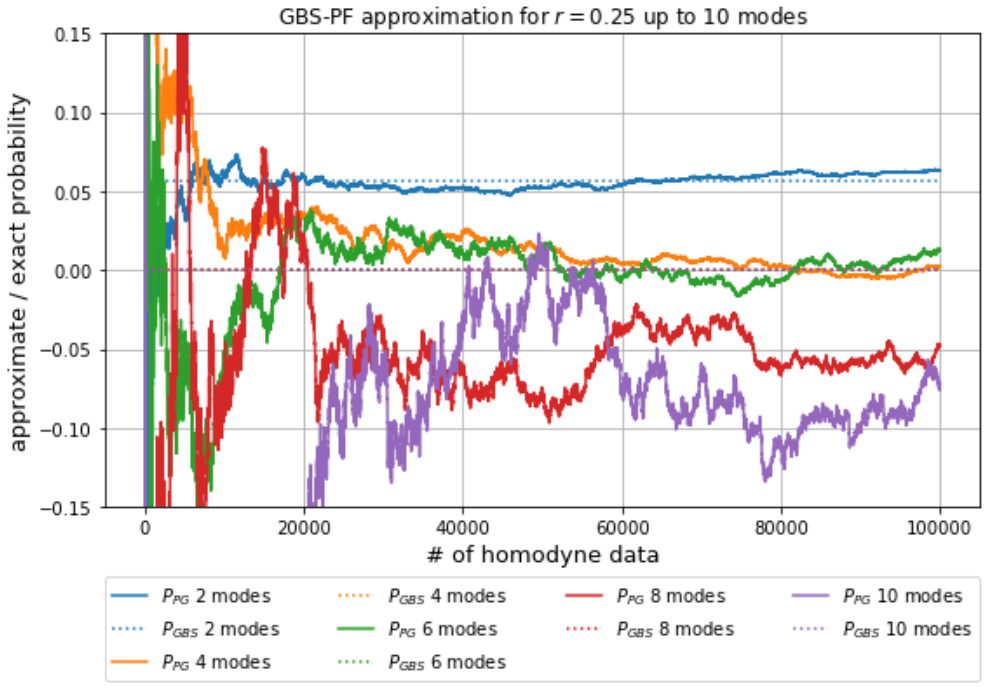

Figure A.7: Convergence of $\mathbb{P}_{\texttt{PG}}$ to $\mathbb{P}_{\texttt{GBS}}$ for PF-GBS with square interferometer for measurement $(1, \ldots, 1)$ at the output.

Now, recall that the plot in [**Figure 6.5**] showing that the standard deviations for larger number of modes decreases slower as the number of Homodyne data grows, and also showing

that the true errors are larger for simulations with larger more modes. This trend can be observed from the convergence plot [**Figure A.7**] where we can see that the approximations $\mathbb{P}_{\texttt{PG}}(k)$ for $m \in \{2, 4, 6\}$ modes are significantly close to the true probabilities $\mathbb{P}_{\texttt{GBS}}(k)$ despite having the true probabilities for $m \in \{4, 6\}$ indistinguishably close to each other in the plot. It is clear that the approximations $\mathbb{P}_{\texttt{PG}}(k)$ for $m \in \{8, 10\}$ modes deviates way more compared to the approximations for less modes as they end up being more than 0.05 away from the true values $\mathbb{P}_{\texttt{GBS}}(k)$ after the first 100000 Homodyne samples.

This observation also agrees with the simulation results displayed in the plots in [**Figure 6.6**] and [**Figure 6.7**] which show that as we require the approximation $\mathbb{P}_{\texttt{PG}}$ to have a smaller standard deviation, the increase of the number of Homodyne data required is more rapid for GBS with more modes.

Typical case of a square interferometer PF GBS

We now present the standard deviation, true error, and the convergence of approximation $\mathbb{P}_{\texttt{PG}}(k)$ to the true probability $\mathbb{P}_{\texttt{GBS}}(k)$ of a typical case of a PF-GBS simulation with square interferometer. In these results, we are interested in the probability for outcome $k = (1, 1, 1, 1)$. These results are shown in the plots in [**Figure A.8**]. We would like to point out in these results that the true errors $|\mathbb{P}_{\texttt{PG}}(k) - \mathbb{P}_{\texttt{GBS}}(k)|$ rarely exceeds the standard deviation in agreement with the probability of it exceeding the standard deviation $1 - \textbf{erf}\left(\frac{1}{\sqrt{2}}\right) \approx 1 - 0.68$ as discussed in [**Section 6.2.1**]. This is apparent in the left-hand-side plot of [**Figure A.8**] where the orange line is rarely above the green line or below the red line (which does not seem to occur at all in this case. Moreover, the right-hand-side plot further shows that the true error values (green line) only exceed two standard deviation values (orange line) very rarely. This seems to agree with the probability of the true error exceeding the standard deviation, which is $1 - \textbf{erf}\left(\frac{2}{\sqrt{2}}\right) \approx 1 - 0.95$.

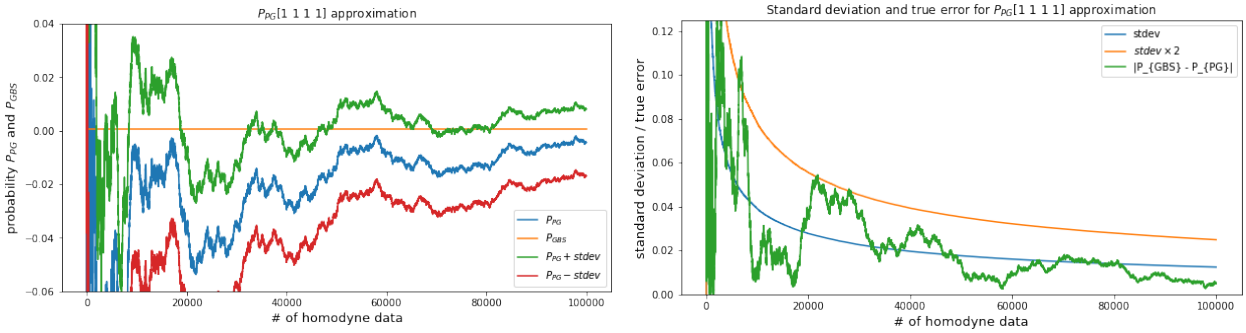

Figure A.8: Results of PF GBS approximation with square interferometer and outcome measurement of $k = (1, 1, 1, 1)$. (1) Left hand side plot: Convergence of $\mathbb{P}_{\texttt{PG}}(k)$ to $\mathbb{P}_{\texttt{GBS}}(k)$ along with upper and lower bound from a single standard deviation. (2) Right hand side plot: one standard deviation (blue line), two standard deviations (orange line), and true error $|\mathbb{P}_{\texttt{PG}}(k) - \mathbb{P}_{\texttt{GBS}}(k)|$ (green line).

Appendix B

Hyperbolic Function Representation of Thermalized Bosonic Operators

Consider the Baker-Campbell-Hausdorff (BCH) Formula

$$e^{tA}Be^{-tA} = \sum_{n\in\mathbb{N}} \frac{t^n}{n!}[A,B]^{(n)}$$

where we define the commutator term recursively as follows

$$\begin{aligned}
[A,B]^{(0)} &= B \\
[A,B]^{(1)} &= [A,B] \\
[A,B]^{(n)} &= [A,[A,B]^{(n-1)}].
\end{aligned}$$

Thus, applying this formula to the Bosonic creation operator Thermalized by the operation $T(r) = \mathbf{exp}(r(\widehat{a}_c\widehat{a}_d - \widehat{a}_c^\dagger\widehat{a}_d^\dagger))$ gives us

$$T(r)\widehat{a}_c^\dagger T^\dagger(r) = \sum_{n\in\mathbb{N}} \frac{r^n}{n!}[\widehat{a}_c\widehat{a}_d - \widehat{a}_c^\dagger\widehat{a}_d^\dagger, \widehat{a}_c^\dagger]^{(n)}.$$

For shorthand, let $G = \widehat{a}_c\widehat{a}_d - \widehat{a}_c^\dagger\widehat{a}_d^\dagger$. thus we have

$$\begin{aligned}
[G,\widehat{a}_c^\dagger]^{(0)} &= \widehat{a}_c^\dagger \\
[G,\widehat{a}_c^\dagger]^{(1)} &= [G,\widehat{a}_c^\dagger] = (\widehat{a}_c\widehat{a}_c^\dagger\widehat{a}_d - \widehat{a}_c^\dagger\widehat{a}_c^\dagger\widehat{a}_d^\dagger) - (\widehat{a}_c^\dagger\widehat{a}_c\widehat{a}_d - \widehat{a}_c^\dagger\widehat{a}_c^\dagger\widehat{a}_d^\dagger) = \widehat{a}_d \\
[G,\widehat{a}_c^\dagger]^{(2)} &= [G,\widehat{a}_d] = (\widehat{a}_c\widehat{a}_d\widehat{a}_d - \widehat{a}_c^\dagger\widehat{a}_d^\dagger\widehat{a}_d) - (\widehat{a}_c\widehat{a}_d\widehat{a}_d - \widehat{a}_c^\dagger\widehat{a}_d\widehat{a}_d^\dagger) = \widehat{a}_c^\dagger \\
&\vdots
\end{aligned}$$

Thus we have $[G,\widehat{a}_c^\dagger]^{(n)} = \widehat{a}_c^\dagger$ when n is even, and $[G,\widehat{a}_c^\dagger]^{(n)} = \widehat{a}_d$ when n is odd. Therefore the BCH expansion of the Thermalized creation operator becomes

$$\begin{aligned}
T(r)\widehat{a}_c^\dagger T^\dagger(r) &= \left(\sum_{n\in\mathbb{N}_{\text{Even}}} \frac{r^n}{n!}\widehat{a}_c^\dagger\right) + \left(\sum_{m\in\mathbb{N}_{\text{Odd}}} \frac{r^m}{m!}\widehat{a}_d\right) \\
&= \cosh(r)\widehat{a}_c^\dagger + \sinh(r)\widehat{a}_d.
\end{aligned}$$

We note that this is in agreement with [BR97, Ch.3.7, eqn.(3.7.58)]. Its Hermitian conjugate is simply

$$T(r)\widehat{a}_c T^{\dagger}(r) = \cosh(r)\widehat{a}_c + \sinh(r)\widehat{a}_d^{\dagger},$$

where the distributivity of the Hermitian conjugate is used, and which is in agreement with [BR97, Ch.3.7, eqn.(3.7.56)] and [BK85, eqn.(2.25)].

Bibliography

[AA11] 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* (2011). DOI: 10.1145/1993636.1993682.

[Aar05] Scott Aaronson. "Quantum computing, postselection, and probabilistic polynomial-time". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 461.2063 (May 2005), pp. 3473–3482. DOI: 10.1098/rspa.2005.1546.

[Aar10a] Scott Aaronson. "BQP and the polynomial hierarchy". In: *Proceedings of the 42nd ACM symposium on Theory of computing - STOC 10* (2010). DOI: 10.1145/1806689.1806711.

[Aar10b] Scott Aaronson. *The Equivalence of Sampling and Searching*. 2010. arXiv: 1009.5104 [quant-ph].

[Aar20] Scott Aaronson. *Chinese BosonSampling experiment: the gloves are off*. Dec. 2020. URL: https://www.scottaaronson.com/blog/?p=5159.

[AB09] Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach*. Cambridge Univ. Press, 2009.

[AB16] Scott Aaronson and Daniel J. Brod. "BosonSampling with lost photons". In: *Physical Review A* 93.1 (2016). DOI: 10.1103/physreva.93.012335.

[AC16] Scott Aaronson and Lijie Chen. *Complexity-Theoretic Foundations of Quantum Supremacy Experiments*. 2016. arXiv: 1612.05903 [quant-ph].

[ARL14] Gerardo Adesso, Sammy Ragy, and Antony R. Lee. "Continuous Variable Quantum Information: Gaussian States and Beyond". In: *Open Systems and Information Dynamics* 21 (Dec. 2014). DOI: 10.1142/s1230161214400010.

[Aru+19] Frank Arute et al. "Quantum Supremacy using a Programmable Superconducting Processor". In: *Nature* 574 (2019), pp. 505–510. URL: https://www.nature.com/articles/s41586-019-1666-5.

[Ban99] Konrad Banaszek. "Statistical uncertainty in quantum-optical photodetection measurements". In: *Journal of Modern Optics* 46.4 (1999), pp. 675–692. DOI: 10.1080/09500349908231294.

[Ben80] Paul Benioff. "The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines". In: *Journal of Statistical Physics* 22.5 (1980), pp. 563–591. DOI: 10.1007/bf01011339.

[BK85] S. M. Barnett and P. L. Knight. "Thermofield analysis of squeezing and statistical mixtures in quantum optics". In: *Journal of the Optical Society of America B* 2.3 (1985), p. 467. DOI: 10.1364/josab.2.000467.

[BL05] Samuel L. Braunstein and Peter Van Loock. "Quantum information with continuous variables". In: *Reviews of Modern Physics* 77.2 (2005), pp. 513–577. DOI: 10.1103/revmodphys.77.513.

[BL07] Dagmar Bruß and Gerd Leuchs. *Lectures on Quantum Information*. Wiley-VCH, 2007.

[BMS16] Michael J. Bremner, Ashley Montanaro, and Dan J. Shepherd. "Average-Case Complexity Versus Approximate Simulation of Commuting Quantum Computations". In: *Physical Review Letters* 117.8 (2016). DOI: 10.1103/physrevlett.117.080501.

[Bou+18] Adam Bouland et al. *Quantum Supremacy and the Complexity of Random Circuit Sampling*. 2018. arXiv: 1803.04402 [quant-ph].

[BPP98] Stephen M. Barnett, Lee S. Phillips, and David T. Pegg. "Imperfect photodetection as projection onto mixed states". In: *Optics Communications* 158.1-6 (1998), pp. 45–49. DOI: 10.1016/s0030-4018(98)00511-2.

[BR97] Stephen M. Barnett and Paul M. Radmore. *Methods in theoretical quantum optics*. Clarendon Press, 1997.

[BV97] Ethan Bernstein and Umesh Vazirani. "Quantum Complexity Theory". In: *SIAM Journal on Computing* 26.5 (1997), pp. 1411–1473. DOI: 10.1137/s0097539796300921.

[CG69] K. E. Cahill and R. J. Glauber. "Density Operators and Quasiprobability Distributions". In: *Physical Review* 177.5 (1969), pp. 1882–1902. DOI: 10.1103/physrev.177.1882.

[Cha+17] U. Chabaud et al. "Continuous-variable sampling from photon-added or photon-subtracted squeezed states". In: *Physical Review A* 96.6 (2017). DOI: 10.1103/physreva.96.062307.

[Dar02] Giacomo Mauro D'ariano. "Tomographic methods for universal estimation in quantum optics". In: *In International School of Physics Enrico Fermi*. IOS Press, 2002.

[Deu85] David Deutch. "Quantum theory, the Church–Turing principle and the universal quantum computer". In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 400.1818 (1985), pp. 97–117. DOI: 10.1098/rspa.1985.0070.

[Dou+17] T. Douce et al. "Continuous-Variable Instantaneous Quantum Computing is hard to sample". In: *Quantum Information and Measurement (QIM) 2017* (2017). DOI: 10.1364/qim.2017.qt5b.1.

[DPS03] G. Mauro D'Ariano, Matteo G. A. Paris, and Massimiliano F. Sacchi. *Quantum Tomography*. 2003. arXiv: quant-ph/0302028 [quant-ph].

[FC88] H. Fearn and M.j. Collett. "Representations of Squeezed States with Thermal Noise". In: *Journal of Modern Optics* 35.3 (1988), pp. 553–564. DOI: 10.1080/09500348814550571.

[Fey82] Richard P. Feynman. "Simulating physics with computers". In: *International Journal of Theoretical Physics* 21.6-7 (1982), pp. 467–488. DOI: 10.1007/bf02650179.

[FF98] Hong-yi Fan and Yue Fan. "New representation of thermal states in thermal field dynamics". In: *Physics Letters A* 246.3-4 (1998), pp. 242–246.

[FH16] Edward Farhi and Aram W Harrow. *Quantum Supremacy through the Quantum Approximate Optimization Algorithm*. 2016. arXiv: 1602.07674 [quant-ph].

[FOP05] Alessandro Ferraro, Stefano Olivares, and Matteo G. A. Paris. *Gaussian states in continuous variable quantum information*. 2005. arXiv: quant-ph/0503237 [quant-ph].

[FP12] Alessandro Ferraro and Matteo G. A. Paris. "Nonclassicality Criteria from Phase-Space Representations and Information-Theoretical Constraints Are Maximally Inequivalent". In: *Physical Review Letters* 108.26 (2012). DOI: 10.1103/physrevlett.108.260403.

[FT20] C. Fabre and N. Treps. "Modes and states in quantum optics". In: *Reviews of Modern Physics* 92.3 (Sept. 2020). ISSN: 1539-0756. DOI: 10.1103/revmodphys.92.035005. URL: http://dx.doi.org/10.1103/RevModPhys.92.035005.

[FU15] Bill Fefferman and Chris Umans. *The Power of Quantum Fourier Sampling*. 2015. arXiv: 1507.05592 [cs.CC].

[GIQ19] Brajesh Gupt, Josh Izaac, and Nicolás Quesada. "The Walrus: a library for the calculation of hafnians, Hermite polynomials and Gaussian boson sampling". In: *Journal of Open Source Software* 4.44 (2019), p. 1705. DOI: 10.21105/joss.01705. URL: https://doi.org/10.21105/joss.01705.

[Gla63] Roy J. Glauber. "Coherent and Incoherent States of the Radiation Field". In: *Physical Review* 131.6 (1963), pp. 2766–2788. DOI: 10.1103/physrev.131.2766.

[Gro96] Lov K. Grover. "A fast quantum mechanical algorithm for database search". In: *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing - STOC 96* (1996). DOI: 10.1145/237814.237866.

[Ham+17] Craig S. Hamilton et al. "Gaussian Boson Sampling". In: *Conference on Lasers and Electro-Optics* (2017). DOI: 10.1364/cleo_qels.2017.ftu1f.2.

[Han+18] Dominik Hangleiter et al. "Anticoncentration theorems for schemes showing a quantum speedup". In: *Quantum* 2 (May 2018), p. 65. ISSN: 2521-327X. DOI: 10.22331/q-2018-05-22-65. URL: https://doi.org/10.22331/q-2018-05-22-65.

[Hud74] R.l. Hudson. "When is the wigner quasi-probability density non-negative?" In: *Reports on Mathematical Physics* 6.2 (1974), pp. 249–252. DOI: 10.1016/0034-4877(74)90007-x.

[Hus40] Kodi Husimi. "Some Formal Properties of the Density Matrix". In: *Proceedings of the Physico-Mathematical Society of Japan. 3rd Series* 22.4 (1940), pp. 264–314. DOI: 10.11429/ppmsj1919.22.4_264.

[Kar72] Richard M. Karp. "Reducibility among Combinatorial Problems". In: *Complexity of Computer Computations* (1972), pp. 85–103. DOI: 10.1007/978-1-4684-2001-2_9.

[Kie+11] T. Kiesel et al. "Direct Sampling of Negative Quasiprobabilities of a Squeezed State". In: *Physical Review Letters* 107.11 (2011). DOI: 10.1103/physrevlett.107.113604.

[KL98] E. Knill and R. Laflamme. "Power of One Bit of Quantum Information". In: *Physical Review Letters* 81.25 (1998), pp. 5672–5675. DOI: 10.1103/physrevlett.81.5672.

[Kru+19] Regina Kruse et al. "Detailed study of Gaussian boson sampling". In: *Physical Review A* 100.3 (2019). DOI: 10.1103/physreva.100.032326.

[KSV02] A. Yu. Kitaev, A. Shen, and M. N. Vyalyi. *Classical and quantum computation*. American Mathematical Society, 2002.

[Lau83] Clemens Lautemann. "BPP and the polynomial hierarchy". In: *Information Processing Letters* 17.4 (1983), pp. 215–217. DOI: 10.1016/0020-0190(83)90044-3.

[Leo+96] U Leonhardt et al. "Sampling of photon statistics and density matrix using homodyne detection". In: *Optics Communications* 127.1-3 (1996), pp. 144–160. DOI: 10.1016/0030-4018(96)00061-2.

[Leo97] Ulf Leonhardt. *Measuring the quantum state of light*. Cambridge Univ. Press, 1997.

[Liu+16] Nana Liu et al. "Power of one qumode for quantum computation". In: *Physical Review A* 93.5 (2016). DOI: 10.1103/physreva.93.052304.

[Lou00] Rodney Loudon. *The quantum theory of light*. Oxford University Press, 2000.

[LP07] Peter Lambropoulos and David Petrosyan. *Quantum optics and quantum information*. Springer, 2007.

[LPD95] U. Leonhardt, H. Paul, and G. M. D'Ariano. "Tomographic reconstruction of the density matrix via pattern functions". In: *Physical Review A* 52.6 (1995), pp. 4899–4907. DOI: 10.1103/physreva.52.4899.

[Lun+14] A. P. Lund et al. "Boson Sampling from a Gaussian State". In: *Physical Review Letters* 113.10 (May 2014). DOI: 10.1103/physrevlett.113.100502.

[ME12] A. Mari and J. Eisert. "Positive Wigner Functions Render Classical Simulation of Quantum Computation Efficient". In: *Physical Review Letters* 109.23 (2012). DOI: 10.1103/physrevlett.109.230503.

[MS72] A. R. Meyer and L. J. Stockmeyer. "The equivalence problem for regular expressions with squaring requires exponential space". In: *13th Annual Symposium on Switching and Automata Theory (swat 1972)* (1972). DOI: 10.1109/swat.1972.29.

[Mun+95] M. Munroe et al. "Photon-number statistics from the phase-averaged quadrature-field distribution: Theory and ultrafast measurement". In: *Physical Review A* 52.2 (1995). DOI: 10.1103/physreva.52.r924.

[Mun+96] M. Munroe et al. "High-Efficiency, Ultrafast Photon-Number Statistics from Phase-Averaged Homodyne Detection". In: *Coherence and Quantum Optics VII* (1996), pp. 53–61. DOI: 10.1007/978-1-4757-9742-8_10.

[Nav15] Carlos Navarrete-Benlloch. *An introduction to the formalism of quantum information with continuous variables*. Morgan and Claypool Publishers, 2015.

[NC10] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2010.

[Ped+19] Edwin Pednault et al. *Leveraging Secondary Storage to Simulate Deep 54-qubit Sycamore Circuits*. 2019. arXiv: 1910.09534 [quant-ph].

[PŘ04] Matteo Paris and Jaroslav Řeháček. *Quantum state estimation*. Springer, 2004.

[Pre12] John Preskill. *Quantum computing and the entanglement frontier*. 2012. arXiv: 1203.5813 [quant-ph].

[Pre18] John Preskill. "Quantum Computing in the NISQ era and beyond". In: *Quantum* 2 (2018), p. 79. DOI: 10.22331/q-2018-08-06-79.

[Pur01] Ravinder R. Puri. *Mathematical methods of quantum optics*. Springer-Verlag, 2001.

[Qi+20] Haoyu Qi et al. "Regimes of Classical Simulability for Noisy Gaussian Boson Sampling". In: *Physical Review Letters* 124.10 (2020). DOI: 10.1103/physrevlett.124.100502.

[Rec+94] Michael Reck et al. "Experimental realization of any discrete unitary operator". In: *Physical Review Letters* 73.1 (1994), pp. 58–61. DOI: 10.1103/physrevlett.73.58.

[RLR15] Saleh Rahimi-Keshari, Austin P. Lund, and Timothy C. Ralph. "What Can Quantum Optics Say about Computational Complexity Theory?" In: *Physical Review Letters* 114.6 (Nov. 2015). DOI: 10.1103/physrevlett.114.060501.

[RRC16] Saleh Rahimi-Keshari, Timothy C. Ralph, and Carlton M. Caves. "Sufficient Conditions for Efficient Classical Simulation of Quantum Optics". In: *Physical Review X* 6.2 (2016). DOI: 10.1103/physrevx.6.021039.

[RW20] Michael Raymer and Ian Walmsley. "Temporal modes in quantum optics: then and now". In: *Physica Scripta* 95.6 (Mar. 2020). URL: https://iopscience.iop.org/article/10.1088/1402-4896/ab6153.

[Ser17] Alessio Serafini. *Quantum continuous variables: a primer of theoretical methods*. CRC Press, Taylor and Francis Group, 2017.

[Shc19] V. S. Shchesnovich. "Noise in boson sampling and the threshold of efficient classical simulatability". In: *Physical Review A* 100.1 (2019). DOI: 10.1103/physreva.100.012340.

[Sho94] P.w. Shor. "Algorithms for quantum computation: discrete logarithms and factoring". In: *Proceedings 35th Annual Symposium on Foundations of Computer Science* (1994). DOI: 10.1109/sfcs.1994.365700.

[Smi+93] D. T. Smithey et al. "Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum". In: *Physical Review Letters* 70.9 (Jan. 1993), pp. 1244–1247. DOI: 10.1103/physrevlett.70.1244.

[SS05] Elias M. Stein and Rami Shakarchi. *Real analysis: measure theory, integration, and Hilbert spaces*. Princeton Univ. Press, 2005.

[Sto77] Larry J. Stockmeyer. "The polynomial-time hierarchy". In: *Theoretical Computer Science* 3.1 (1977), pp. 1–22. DOI: 10.1016/0304-3975(76)90061-x.

[Sto83] Larry Stockmeyer. "The complexity of approximate counting". In: *Proceedings of the fifteenth annual ACM symposium on Theory of computing - STOC 83* (1983). DOI: 10.1145/800061.808740.

[Tod91] Seinosuke Toda. "PP is as Hard as the Polynomial-Time Hierarchy". In: *SIAM Journal on Computing* 20.5 (1991), pp. 865–877. DOI: 10.1137/0220053.

[Tur39] A. M. Turing. "Systems of Logic Based on Ordinals". In: *Proceedings of the London Mathematical Society* s2-45.1 (1939), pp. 161–228. DOI: 10.1112/plms/s2-45.1.161.

[Val79] Leslie G. Valiant. "The complexity of computing the permanent". In: *Theoretical Computer Science* 8.2 (1979). URL: https://doi.org/10.1016/0304-3975(79)90044-6.

[Vaz02] Umesh V. Vazirani. "A survey of quantum complexity theory". In: *Proceedings of Symposia in Applied Mathematics Quantum Computation* (2002), pp. 193–217. DOI: 10.1090/psapm/058/1922899.

[Vei+13] Victor Veitch et al. "Efficient simulation scheme for a class of quantum optics experiments with non-negative Wigner representation". In: *New Journal of Physics* 15.1 (2013), p. 013037. DOI: 10.1088/1367-2630/15/1/013037.

[VR89] K. Vogel and H. Risken. "Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase". In: *Physical Review A* 40.5 (1989), pp. 2847–2849. DOI: 10.1103/physreva.40.2847.

[Wan+07] X Wang et al. "Quantum information with Gaussian states". In: *Physics Reports* 448.1-4 (2007), pp. 1–111. DOI: 10.1016/j.physrep.2007.04.005.

[Wee+12] Christian Weedbrook et al. "Gaussian quantum information". In: *Reviews of Modern Physics* 84.2 (2012), pp. 621–669. DOI: 10.1103/revmodphys.84.621.

[Wig32] E. Wigner. "On the Quantum Correction For Thermodynamic Equilibrium". In: *Physical Review* 40.5 (1932), pp. 749–759. DOI: 10.1103/physrev.40.749.

[Wun04] Alfred Wunsche. "About the nonclassicality of states defined by nonpositivity of theP-quasiprobability". In: *Journal of Optics B: Quantum and Semiclassical Optics* 6.2 (2004), pp. 159–171. DOI: 10.1088/1464-4266/6/2/006.

[Yao93] A. Chi-Chih Yao. "Quantum circuit complexity". In: *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science* (1993). DOI: 10.1109/sfcs.1993.366852.

[Zho+20] Han-Sen Zhong et al. "Quantum computational advantage using photons". In: *Science* (2020). ISSN: 0036-8075. DOI: 10.1126/science.abe8770. eprint: https://science.sciencemag.org/content/early/2020/12/02/science.abe8770.full.pdf. URL: https://science.sciencemag.org/content/early/2020/12/02/science.abe8770.