Numerical Finite Key Analysis

by

Ian George

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Physics (Quantum Information)

Waterloo, Ontario, Canada, 2020

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

While all results in this thesis are in effect my own, beyond the advisement from my advisor, Norbert Lütkenhaus, I would like to acknowledge the following contributions which influence this thesis:

1. Some preliminary results were found to have small errors by Jie Lin which I then corrected. Furthermore, Jie Lin in some sense acted as an advisor on the numerical framework I was working within.

2. With the permission of my coauthors, Norbert Lütkenhaus and Jie Lin, Chapters 4 and 5 are largely taken from our paper [26], which reported the results of my research under the advisement of my coauthors. While I wrote the majority of the paper and the results of these sections are my own work, I believe the conciseness and clarity of the presentation in Section 4.3.3 is the result of Jie Lin's contributions to the writing of the paper.

3. At one point I didn't know I could write the trace norm as a linear SDP for Hermitian matrices in the manner I needed and Jamie Sikora showed me how to do this. Without that, none of this thesis would have happened.

4. Appendix A are results which were improved in their formulation thanks to comments from Jamie Sikora and Jie Lin. Jamie Sikora also brought it to my attention that the results of Appendix A were not necessary for Lemma 16.

Abstract

Quantum key distribution (QKD) [7, 22] is a cryptographic protocol in which two legitimate parties, Alice and Bob, establish an information-theoretically secure secret key using the properties of quantum mechanics. As such, the goal of this subfield is to design a QKD protocol, prove its information-theoretic security, and, if the design is practical, build an implementation. In this process, a major contribution of the theorist is to find a good middle ground between an accurate representation of the implemented protocol and the simplicity of the analysis when proving the information-theoretic security of the implementation. For some time this middle ground was neglected by instead assuming that Alice and Bob can send an infinite number of quantum signals to each other, which no implementation can ever achieve. This point was resolved in Renner's PhD thesis [51] in which he developed the general framework of security of QKD with finite resources– now known as finite key analysis. Unfortunately, the calculations of finite analysis are difficult without nice symmetries and so most works since Renner's thesis have focused on simplified protocols with nice symmetries. In this thesis, we begin by both expounding and improving the theory of finite key analysis. Our improvement of the analysis in turn improves the amount of key that can be generated for protocols without nice symmetries. Following this improvement, we present a numerical method for the finite key analysis of QKD protocols that can be represented in finite-dimensional Hilbert spaces without requiring specific symmetries. Lastly, we present the finite key analysis for variations of the BB84 protocol [7] for both better understanding of the finite key analysis and proof of the general applicability of our numerical method.

Acknowledgements

First I would like to thank my advisor, Norbert Lütkenhaus. I am deeply indebted to him for his continuous insight, generosity with his time, general good-naturedness, and patience. I am very thankful he was willing to give me the opportunity to do my Master's work in his Optical Quantum Communication Theory (OQCT) lab.

Many thanks to Jie Lin who advised me on the numerical QKD framework, always pushed me to be more exact, and has answered a million of my questions in the middle of his own research. I'd also like to thank the rest of the OQCT lab for their support, ideas, and many fun meals. I hope we can all go back to Lanzhou's someday.

Thanks to the other people around Waterloo who have been wonderful to me: Jamie Sikora who taught me so much about semidefinite programming and optimization theory and showed me how fun of a topic it is, my housemates, Brad, Lane, and Vinodh, who have been a blast to live with and are always encouraging, and the entire University of Waterloo Ballroom Dance team for being welcoming and providing me something to do when not thinking about quantum information theory.

Finally, I would like to thank my parents and sister for their love, encouragement, and support.

Table of Contents

List of Figures

Chapter 1

Introduction

Quantum key distribution (QKD) [7, 22] is a cryptographic protocol in which two legitimate parties, Alice and Bob, establish an information-theoretically secure key using the properties of quantum mechanics. This protocol differs from historical cryptographic protocols because its security is derived from physical principles rather than assumptions on Eve's computational power. Furthermore, unlike many classical cryptographic protocols in which an adversary may store all of the data exchanged during the protocol indefinitely, the data exchanged during the QKD protocol is secure against later advancements in quantum hacking. In other words, the secret key generated in QKD is as secure as it was at the time of generation, or, formally speaking, QKD has the property of forward secrecy. The information-theoretic security and forward secrecy of QKD makes it a desirable protocol to have in one's cryptographic toolbox.

For these reasons, in the past three decades since this protocol was invented, there has been rapid advancement in both theory and implementation of QKD protocols. For a QKD implementation to be useful, it must have (1) a feasible (though, ideally, practical) implementation in the real world and (2) an information-theoretic security proof which satisfactorily represents the feasible implementation.[1] The job/goal of those who work on QKD is to make at least one of these two requirements easier to satisfy. This thesis is no different as it is interested in constructing a numerical method of determining information-theoretic security proofs for feasible and practical implementations of QKD protocols.

Historically, many proofs of information-theoretic security of QKD protocols rely on the assumption that Alice and Bob exchange an infinite number of signals to establish

[1] We say 'satisfactorily' as no model perfectly models reality, but if we believe they are nearly the same, we will be happy– until someone shows us why we should not have been.

their secret key. This is known as infinite or asymptotic key analysis and is in principle impossible to implement.[2] Furthermore, the more signals Alice and Bob want to send between each other, the more time the protocol will take. It follows that from an implementation standpoint it is important to know what length of secret key they can establish using a reasonable amount of time/signals. Finite key analysis, first presented by Renner [51], solves this gap between the real world implementation and the security proof by determining an upper bound on the key length of an ε-secure key using finite resources where ε represents the probability that Alice and Bob do not abort the protocol *and* the output of the protocol is not information-theoretically secure.

Unfortunately, because of both the mathematical and conceptual difficulty of finite key analysis, most works following Renner's thesis only perform the analysis for simple symmetric protocols with a simplified analysis. This is an issue as finite resources is a reality of all protocols and many protocols are neither simple nor symmetric. Therefore we need methods for performing finite key analysis for such protocols. In this thesis, we present tools to remedy this gap by introducing a numerical method for determining the ε-secure key length of general (device-dependent) QKD protocols that can be represented in finite-dimensional Hilbert spaces without forcing any required symmetries. In Chapter 2, we present the necessary background in mathematics and physics to understand the thesis. In Chapter 3, we present the theory of finite key analysis. This chapter expands upon the points made in Renner's thesis [51], documents important results in finite key analysis since Renner's thesis, and presents new results. Specifically, it presents an improvement to the analysis of the parameter estimation subprotocol (Theorem 5), which can improve the key rate for asymmetric observations in protocols, as well as a new discussion on the completeness of device-dependent QKD protocols (Section 3.1.3). In Chapter 4, after a review of the numerical method for asymptotic key rates [17, 68] which we are extending, we present our tight and reliable numerical method for determining ε-secure finite key lengths of general device-dependent QKD protocols. In Chapter 5, we use our numerical method to showcase both the generality of our numerical method as well as provide insights on current finite key proof methods. Finally, in Chapter 6, we take stock of what we have shown and look at what important topics remain to be solved in the field of finite key analysis.

[2] We say in principle because, as we will we see in Chapter 5, one should be able to achieve the asymptotic result for a sufficiently large finite number of signals.

Chapter 2

Background

In this chapter, we will review the mathematical framework of quantum mechanics necessary for quantum key distribution. We then review semidefinite programming which is the quintessential tool used in this thesis. Much of the notation is based on [66].

2.1 Quantum Mechanics

2.1.1 Operator Theory

Hilbert Spaces

Quantum mechanics fundamentally is concerned with describing physical systems which fail to be properly predicted by the methods from classical mechanics and electrodynamics. All such systems are represented in Hilbert spaces. This is not a trivial choice but rather is the mathematical object that allows for unifying the matrix mechanics formalism and the Schrödinger wave mechanics formalism [49].

Hilbert spaces are Euclidean inner product spaces which are complete and separable. In the case of quantum mechanics, the Hilbert space in question is an inner product space over the field of complex numbers. We note that the demands of the space being complete and separable are non-trivial for infinite-dimensional Hilbert spaces. Conveniently, this thesis is exclusively concerned with finite-dimensional Hilbert spaces where these nuances need not be worried about. Throughout this thesis we denote the Hilbert space of a quantum system A by $\mathcal{H}_A$ as is popular among physicists. Furthermore, as we will only be concerned

with finite-dimensional Hilbert spaces over complex numbers, for every Hilbert space $\mathcal{H}$ defined in this thesis, there will exist a finite set, or alphabet, Σ, which is the dimension of the Hilbert space. That is to say, for any Hilbert space $\mathcal{H}$, there exists Σ such that $\mathcal{H} = \mathbb{C}^{|\Sigma|}$. This equivalence will be used often. For simplicity, we will write $\mathbb{C}^{|\Sigma|}$ as $\mathbb{C}^{\Sigma}$ throughout this thesis. It follows from this observation of finite-dimensional Hilbert spaces that a vector $v \in \mathcal{H}$ is a $|\Sigma|$-dimensional vector and we could index denote its entries by v_a where $a \in \Sigma$. We now summarize all of these points.

Hilbert Space

In this thesis, $\mathcal{H}_A$ denotes a Hilbert space of a quantum system A. The Hilbert space $\mathcal{H}_A$ is equivalent to $\mathbb{C}^{\Sigma}$ for some alphabet Σ. Let $u, v \in \mathcal{H}$. The inner product of $\mathcal{H}$ is defined as

$$\langle u, v \rangle = \sum_{a \in \Sigma} \overline{u_a} v_a = u^\dagger v$$

where $\overline{u_a}$ is the conjugate of complex u_a and $\dagger$ represents the Hermitian conjugate.

Linear Operators

As one might expect from being formalized in Hilbert spaces, quantum mechanics is generally taken to be linear. Consequently, we will be greatly concerned with linear operators. Given Hilbert spaces $\mathcal{H}_A = \mathbb{C}^{\Sigma}$, $\mathcal{H}_B = \mathbb{C}^{\Lambda}$, the space of linear operators from $\mathcal{H}_A$ to $\mathcal{H}_B$ will be denoted $\mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$.

By fixing a basis, we can associate a linear operator $X \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ with a matrix, M_X, by the following relation:

$$M_X = \sum_{a \in \Lambda, b \in \Sigma} \langle e_a, X e_b \rangle$$

where e_i denotes the vector with a 1 in the i^{th} entry and 0 elsewhere for the basis defined by the relevant alphabet (i.e. it fixes an orthonormal basis). We will from now on refer to both the matrix and the operator by X for simplicity.

When we are considering $\mathrm{L}(\mathcal{H}_A, \mathcal{H}_A)$, we will simply denote it $\mathrm{L}(\mathcal{H}_A)$. Note that $\mathrm{L}(\mathcal{H}_A)$ corresponds to the set of square matrices of dimension $|\Sigma|$.

Trace

Given $u, v \in \mathcal{H}_A$, the outer product of u and v is $uv^\dagger \in \mathrm{L}(\mathcal{H}_A)$. The trace, denoted Tr, is

the unique linear map defined by $\mathrm{Tr} : \mathrm{L}(\mathcal{H}_A) \to \mathbb{C}$ such that $\mathrm{Tr}\big(uv^\dagger\big) = \langle v, u\rangle$. One can then use this map to define an inner product on $\mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$. Given $X, Y \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$,

$$\langle A, B\rangle = \mathrm{Tr}\big(A^\dagger B\big)$$

Note that as one expresses matrices using an orthonormal basis, it follows from the definition of Tr that, in the case of $\mathrm{L}(\mathcal{H}_A)$, it is equivalent to the definition

$$\mathrm{Tr}(X) = \sum_{a\in\Sigma} \langle e_a, Xe_a\rangle \text{ for } X \in \mathrm{L}(\mathcal{H}_A) \ .$$

In other words, for a square matrix, the trace is the sum of the diagonal entries. We can now summarize the sets of linear operators which will matter in this thesis.

Important Subsets of Linear Operators

Consider $\mathcal{H}_A$. We can then define the following subsets of $\mathrm{L}(\mathcal{H}_A)$:

1. *Normal Operators*:
$$\{X \in \mathrm{L}(\mathcal{H}_A) : XX^\dagger = X^\dagger X\}$$
2. *Hermitian Operators*:
$$\mathrm{Herm}(\mathcal{H}_A) = \{X \in \mathrm{L}(\mathcal{H}_A) : X = X^\dagger\}$$
3. *Positive Semidefinite Operators*:
$$\mathrm{Pos}(\mathcal{H}_A) = \{X \in \mathrm{L}(\mathcal{H}_A) : X = Y^\dagger Y, Y \in \mathrm{L}(\mathcal{H}_A)\}$$
4. *Density Operators*
$$\mathrm{D}(\mathcal{H}_A) = \{X \in \mathrm{Pos}(\mathcal{H}_A) : \mathrm{Tr}(X) = 1\}$$
5. *Isometries*
$$U(\mathcal{X}, \mathcal{Y}) = \{V \in \mathrm{L}(\mathcal{X}, \mathcal{Y}) : V^\dagger V = \mathbb{1}_{\mathcal{X}}\}$$

Note that $\mathrm{D}(\mathcal{H}_A) \subset \mathrm{Pos}(\mathcal{H}_A) \subset \mathrm{Herm}(\mathcal{H}_A)$. Furthermore note $\forall V \in U(\mathcal{X}, \mathcal{Y})$, $\|Vx\|_2 = \|x\|_2 \, \forall x \in \mathcal{X}$ which is what makes them isometries.

As we will see, this thesis is largely interested in positive semidefinite operators, and so we note the following facts about them.

Positive Semidefinite Operators

Consider an arbitrary finite dimensional Hilbert space, $\mathcal{H}$.

1. *Equivalent Definitions*: There exist equivalent definitions of positive semidefinite operators. The following two are particularly useful:

$$\begin{aligned}\text{Pos}(\mathcal{H}) &= \{X \in L(\mathcal{H}) : \forall v \in \mathcal{H}, \langle X, vv^\dagger \rangle \geq 0\} \\ &= \{X \in L(\mathcal{H}) : \lambda_{\min}(X) \geq 0\}\end{aligned}$$

   where $\lambda_{min}(\cdot)$ denotes the minimum eigenvalue.

2. *Convex Cone*: The set of positive semidefinite operators, $\text{Pos}(\mathcal{H})$, is a convex cone. That is to say if $\lambda \geq 0$ and $X \in \text{Pos}(\mathcal{H})$, $\lambda X \in \text{Pos}(\mathcal{H})$. Furthermore, if $X, Y \in \text{Pos}(\mathcal{H})$, then for $\lambda \in (0, 1)$, $\lambda X + (1 - \lambda)Y \in \text{Pos}(\mathcal{H})$.

3. *Partial Ordering Induced by Cone*: There exists a partial ordering defined using the positive semidefinite cone known as the Loewner ordering. Given $X, Y \in L(\mathcal{H})$, $X \succeq Y$ if and only if $X - Y \in \text{Pos}(\mathcal{H})$. This is the only partial ordering used in this thesis, so we use $\succeq$ to denote it.

Linear Maps

Just like linear operators take vectors to other vectors in a linear fashion, one can use linear maps to take one Hilbert space to another Hilbert space in a linear fashion. Given $\mathcal{H}_A$, $\mathcal{H}_B$, we denote the set of linear maps from $\mathcal{H}_A$ to $\mathcal{H}_B$ by $T(\mathcal{H}_A, \mathcal{H}_B)$. Just as in the case of linear operators, we define $T(\mathcal{H}_A) \equiv T(\mathcal{H}_A, \mathcal{H}_A)$.

Important Subsets of Linear Maps

1. *Hermitian-Preserving Maps*

$$\{\Phi \in T(\mathcal{H}_A, \mathcal{H}_B) : \forall H \in \text{Herm}(\mathcal{H}_A), \Phi(H) \in \text{Herm}(\mathcal{H}_B)\}$$

2. *Positive Maps*:

$$\{\Phi \in T(\mathcal{H}_A, \mathcal{H}_B) : \forall X \in \text{Pos}(\mathcal{H}_A), \Phi(X) \in \text{Pos}(\mathcal{H}_B)\}$$

3. *Completely Positive (CP) Maps*:

$$\{\Phi \in T(\mathcal{H}_A, \mathcal{H}_B) : \forall \mathcal{H}_C, \forall X \in \mathrm{Pos}(\mathcal{H}_A \otimes \mathcal{H}_C), (\Phi \otimes \mathrm{id}_{\mathrm{L}(\mathcal{H}_C)})(X) \in \mathrm{Pos}(\mathcal{H}_B \otimes \mathcal{H}_C)\}$$

4. *Trace Non-Increasing (TNI) Maps*:

$$\{\Phi \in T(\mathcal{H}_A, \mathcal{H}_B) : \forall X \in \mathrm{L}(\mathcal{H}_A), \mathrm{Tr}(\Phi(X)) \leq \mathrm{Tr}(X)\}$$

   (a) *Trace Preserving (TP) Maps*: Trace non-increasing maps which satisfy the equality for all $X \in \mathrm{L}(\mathcal{H}_A)$.

5. *Unital Maps*:

$$\{\Phi \in T(\mathcal{H}_A, \mathcal{H}_B) : \Phi(\mathbb{1}_{\mathcal{H}_A}) = \mathbb{1}_{\mathcal{H}_B}\}$$

   where $\mathrm{id}_{\mathrm{L}(\mathcal{H}_A)}$ is the map which maps every element of $\mathrm{L}(\mathcal{H}_A)$ to itself and $\otimes$ is the tensor product defined below.

Adjoint Map

Given a map $\Phi \in T(\mathcal{H}_A, \mathcal{H}_B)$, the adjoint map, denoted $\Phi^\dagger$ is uniquely defined by

$$\langle \Phi(X), Y \rangle = \langle X, \Phi^\dagger(Y) \rangle$$

where $X \in \mathcal{H}_A$, $Y \in \mathcal{H}_B$.

Tensor Product & Kronecker Product

The tensor product, denoted $\otimes$, is a way of constructing a Hilbert space out of other Hilbert spaces. It is an interesting mathematical object but for our work all we need is the following points:

1. Given $\mathcal{H}_A = \mathbb{C}^\Sigma$, $\mathcal{H}_B = \mathbb{C}^\Lambda$, $\mathcal{H}_A \otimes \mathcal{H}_B = \mathbb{C}^{\Sigma \times \Lambda}$ where $\times$ is the Cartesian product.
2. Given $\Phi \in T(\mathcal{H}_A, \mathcal{H}_B)$, $\Psi \in T(\mathcal{H}_C, \mathcal{H}_D)$, one can construct the tensor product of the two linear maps, $\Phi \otimes \Psi \in T(\mathcal{H}_A \otimes \mathcal{H}_C, \mathcal{H}_B \otimes \mathcal{H}_D)$, by the action, $(\Phi \otimes \Psi)(X \otimes Y) = \Phi(X) \otimes \Psi(Y)$.
   (a) As operators are also maps, an identical claim can be made for the tensor product of operators.

3. Given $X \in \mathcal{H}_A = \mathbb{C}^\Sigma$, $Y \in \mathcal{H}_B = \mathbb{C}^\Lambda$, as the basis is already implicitly fixed by the alphabets, we can write the operators in the following manner:

$$X = \sum_{a,b \in \Sigma} X(a,b) e_a e_b^\dagger \qquad\qquad Y = \sum_{c,d \in \Lambda} Y(c,d) e_c e_d^\dagger$$

then the operator $X \otimes Y$ corresponds to the Kronecker product of the matrix representations. That is $X \otimes Y$ corresponds to the matrix

$$X \otimes Y = \sum_{\substack{a,b \in \Sigma \\ c,d \in \Lambda}} X(a,b) Y(c,d) e_a e_b^\dagger \otimes e_c e_d^\dagger$$

In summary, the first two points tell us how the tensor product works, the last point tells us how we are going to represent tensor products in the numerics.

With this underlying mathematical framework, we can now expound the theory of quantum information.

2.1.2 Quantum Probability Theory

> Quantum probability theory is
> an operator theory with a soul.
>
> *Quantum Channels and their Capacities*
> Alexander Holevo

We can now breathe life into operator theory by using it to construct an extension of classical probability theory induced by quantum mechanics.

Quantum States, Preparations, and Measurements

Physically, there is a quantum system A which is measured by a measurement apparatus M which results in an outcome (a read-out on the measurement apparatus) of O_i where $i \in \Sigma$. Generally, to run an experiment or an information processing scheme, the physicist prepares the quantum system A many times and performs their measurement M on the quantum system each time. This will always lead to an observed probability distribution

(often called a frequency distribution) over the outcomes $\{O_i\}_{i\in\Sigma}$. One might denote this frequency distribution by $f \in \mathcal{P}(\Sigma)$ where $\mathcal{P}(\Sigma)$ denotes the set of probability distributions over the finite alphabet Σ.

This empirical reality leads to the following formalism:

- A quantum system is represented by a density matrix $\rho_A \in \mathrm{D}(\mathcal{H}_A)$.
- A measurement M is represented by a positive-operator valued measure (POVM) which is a set of positive semi-definite operators that add up to identity

$$\{\Gamma_i\}_{i\in\Sigma} \subseteq \mathrm{Pos}(\mathcal{H}_A) \text{ such that } \sum_{i\in\Sigma} \Gamma_i = \mathbb{1}$$

This choice of representation follows from Born's rule, which states that given such a representation one finds then that $p(i) = \mathrm{Tr}(\rho_A\Gamma_i)$ where $p(i)$ is the probability that one's measurement M results in outcome i given one is measuring ρ_A.

Furthermore, we note that if one takes a probability distribution $p \in \mathcal{P}(\Sigma)$ which is a vector, and write it as the diagonal of a matrix, denoted $\mathrm{diag}(p)$,one will note $\mathrm{diag}(p) \in \mathrm{D}(\mathcal{H}_B)$ where $\mathcal{H}_B = \mathbb{C}^\Sigma$. This tells us probability distributions can be viewed as a subset of density matrices. Alternatively, this tells us that quantum states are themselves a generalization of probability distributions which gives rise to quantum probability theory, and thereby quantum information theory.

Joint Quantum Probability Distributions and Entanglement

If quantum states are generalizations of probability distributions, we may wish to give an account of the generalization of joint probability distributions, so that we might have a notion of correlations. Fundamentally, a quantum joint probability distribution over Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ would be any density matrix $\rho_{AB} \in \mathcal{H}_{AB} \equiv \mathcal{H}_A \otimes \mathcal{H}_B$. More traditionally ρ_{AB} would be referred to as a bipartite state as it is defined over two quantum registers. We will use the two terms interchangeably depending on whichever is clearer. There are two special classes of joint quantum probability distributions which we will be interested in for this thesis: entangled states and classical-quantum states.

Separable and Entangled States

Consider two Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B$. Consider a density matrix $\rho_{AB} \in \mathrm{D}(\mathcal{H}_{AB})$. The operator is separable if there exists Σ such that $\{\sigma_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{H}_A)$, $\{\tau_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{H}_B)$, $p \in \mathcal{P}(\Sigma)$ such that $\rho_{AB} = \sum_{a\in\Sigma} p(a)\, \sigma_a \otimes \tau_a$.

A density matrix ρ_{AB} is entangled if it is not separable.

Entangled states are a special class of joint quantum probability distributions because they satisfy certain properties:

- An entangled state ρ_{AB} can allow for non-local correlations between the two subsystems, which is not possible for classical systems.
- Entanglement (the property of being an entangled state) may be a resource for certain quantum information processing tasks.
- The set of separable states form a convex set, and so the entanglement of a system may be detected via testing.

In other words, entangled states, a class of quantum joint probability distributions, can have specific properties that classical joint probability distributions don't, and these properties at times can be utilized for information processing, as we will see in Section 2.2.

Classical-Quantum States

Consider two Hilbert spaces $\mathcal{H}_X, \mathcal{H}_E$ where $\mathcal{H}_X = \mathbb{C}^\Sigma$. Consider a density matrix $\rho_{XE} \in \mathrm{D}(\mathcal{H}_{X\otimes E})$. The operator is a classical-quantum state if there exists an orthonormal basis $\{|x\rangle\}_{x\in\Sigma}$ for $\mathcal{H}_X$, a probability distribution $p \in \mathcal{P}(\Sigma)$ and set of density matrices $\{\rho_E^x\}_{x\in\Sigma} \subset \mathrm{D}(\mathcal{H}_E)$ such that

$$\rho_{XE} = \sum_{x\in\Sigma} p(x)\, |x\rangle\langle x| \otimes \rho_E^x \ .$$

As one can see from the definition of classical-quantum registers, one could trivially extend the definition for any number of classical and quantum registers. Furthermore, classical quantum states are a subset of the separable states, which gives us an upper bound on how strongly correlated the systems may be. The use of the classical-quantum states is that sometimes one will perform an operation that takes a (multipartite) quantum state

ρ_{AE} and outputs a joint state with a classical register X which is correlated with another register E' which remains a quantum state. This sort of procedure, which is done in QKD, results in classical-quantum states. Such an operation is modeled using a quantum channel.

Quantum Channels

So far we have treated quantum states as static quantum probability distributions. However, we know quantum systems are physical and they change, and these changes must be modeled. This is done with quantum channels. Quantum channels are the most general description of the evolution of (closed) quantum systems. This is easy to see in that Schrödinger's equation is a specific example of a unitary channel.

Quantum Channel

Given systems A and B represented by Hilbert spaces $\mathcal{H}_A$,$\mathcal{H}_B$, a quantum channel is a completely-positive trace-preserving (CPTP) map $\Phi \in T(\mathcal{H}_A, \mathcal{H}_B)$.

A quantum channel is a good physical choice as its positivity along with its trace-preserving property guarantees us it takes quantum probabilities to quantum probabilities. That a quantum channel is completely positive tells us that even if there exists another quantum state, say ρ_C, then $\Phi \otimes \mathrm{id}_{\mathrm{L}(\mathcal{H}_C)}(\rho_{AC})$ will also be a proper quantum state, so that quantum systems are well behaved 'globally.'

While the definition of quantum channels is quite nice, one might want a more direct way of representing them for calculations. There are a few such representations. However, it will be sufficient to consider the Kraus representation of linear maps (and quantum channels) for this thesis.

Kraus Representation

For any completely positive non-zero map, $\Phi \in T(\mathcal{X}, \mathcal{Y})$, there exists a set $\{K_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X}, \mathcal{Y})$ such that

$$\Phi(X) = \sum_{a \in \Sigma} K_a X K_a^\dagger \qquad \forall X \in \mathrm{L}(\mathcal{X})$$

Furthermore, if Φ is also trace-preserving, $\sum_{a \in \Sigma} K_a^\dagger K_a = \mathbb{1}_{\mathcal{X}}$ and consequently the adjoint map $\Phi^\dagger$ is unital.

There are all sorts of classes of channels, but for the needs of this thesis, these are the classes we will need and the notation we will use.

Relevant Classes of Channels

Let $\mathcal{H}_A = \mathbb{C}^\Sigma, \mathcal{H}_B = \mathbb{C}^\Lambda$. All definitions will be for $\Phi \in T(\mathcal{H}_A, \mathcal{H}_B)$.

1. A *state preparation channel* is a channel of the form $\Phi(X) = \sum_{a\in\Sigma} \langle X, e_a e_a^\dagger \rangle \rho_a$ where $\rho_a \subseteq \mathrm{D}(\mathcal{H}_B)$. Note in the case that X is a density matrix, the channel prepares ρ_a with probability $X(a,a)$ which is why it is a preparation channel.

2. A *measurement channel* is a channel of the form $\Phi_{\mathcal{P}}(X) = \sum_{j\in\Lambda} \langle X, \widetilde{\Gamma}_j \rangle e_j e_j^\dagger$ where $\{\widetilde{\Gamma}_j\}_{j\in\Lambda}$ is a POVM. We will also refer to this channel as a *probability map* because when X is a density matrix, by Born's rule, the channel outputs the probability distribution of the measurement outcomes given that input state.

3. A *classical-to-classical channel* [67] is a channel of the form $\Phi_C(X) = \sum_{a\in\Sigma} p(b|a) \langle X, e_a e_a^\dagger \rangle$ where $p(b|a)$ is a conditional probability distribution. The channel is given this name because if X is a density matrix, it is treated like a classical probability distribution (only its diagonal elements matter) and the output of the channel will be a probability distribution (and thus a classical object).

2.1.3 Measures and Entropic Quantitites of Quantum Probabilities

So far we have defined quantum probabilities, joint quantum probabilities, and how quantum probabilities can be evolved over time. Information theory is concerned with quantifying information. As probabilities are about what we *expect* to happen, information is about what we *learn* when something happens given what we expected to happen. For this reason, information theory quantifies information using measure-like functions over probability distributions. Identically, quantum information theory is concerned with measure-like functions over density matrices. If the function is comparing the density matrices themselves, these are (generally) considered measures. If the functions are quantifying the information stored in the quantum state, then these are considered entropic quantities, or entropies. There is an endless amount of literature on entropies and measures, but here we just discuss the specific ones we will need. Generally both the measures and entropies

have operational interpretations which provide a language for the math, and so we include these as they help us talk about QKD in the later section. Note for the rest of this thesis $\log \equiv \log_2$. We begin with the measures.

Measures

A measure here simply means that it is a means to compare two quantum states, it does not necessarily mean that it is a metric.

Trace Distance (Variational Distance)

Let $\rho \in \mathrm{L}(\mathcal{H}_A)$ where $\mathcal{H}_A = \mathbb{C}^\Sigma$. The *trace distance* is:

$$\|\rho\|_1 = \sum_{a \in \mathrm{Spec}(\rho)} |\lambda_a(\rho)| \tag{2.1}$$

which is denoted in this manner as it is the trace norm of the difference between ρ and σ (see 2.13).

Operationally, when ρ, σ are quantum states, the trace distance of $\frac{1}{2}\|\rho - \sigma\|_1$ equals the optimal probability one can successfully distinguish the two states using the best choice of POVM [32].

For historical reasons, when comparing two classical probability distributions, the trace distance is also referred to as the total variational distance, and we will use this phrase in the thesis.

Diamond Norm [34]

Let $\Phi \in T(\mathcal{H}_A, \mathcal{H}_B)$. The *diamond norm* is defined as:

$$\|\Phi\|_\diamond \equiv \max_{\sigma \in \mathrm{Pos}(\mathcal{H}_A \otimes \mathcal{H}_B) : \|\sigma\|_1 \leq 1} \|(\Phi \otimes \mathrm{id}_{\mathcal{H}_A})(\sigma)\|_1 \ . \tag{2.2}$$

Operationally, if one has two channels, Φ and Ψ, it has been shown the diamond norm $\frac{1}{2}\|\Phi - \Psi\|_\diamond$ characterizes how well one can successfully distinguish the two channels using the best choice of input state and POVM [34].

Quantum Relative Entropy

Let $\rho, \sigma \in \mathrm{Pos}(A)$. The *quantum relative entropy* is:

$$D(\rho||\sigma) = \begin{cases} \mathrm{Tr}(\rho(\log(\rho) - \log(\sigma))) & \mathrm{im}(\sigma) \subseteq \mathrm{im}(\rho) \\ \infty & \text{otherwise} \end{cases} \tag{2.3}$$

Entropies

The von Neumann Entropy

Let $\rho_A \in \mathcal{H}_A = \mathbb{C}^\Sigma$. The *von Neumann Entropy* is defined as

$$H(A) \equiv - \sum_{a \in \Sigma : \lambda_a(\rho) > 0} \lambda_a \log \lambda_a$$

where λ_a is the a^{th} eigenvalue of ρ.

This can be viewed roughly as how much one learns about the state ρ_A when one performs a projective measurement in its eigenbasis.

While the von Neumann entropy is fascinating in and of itself, in this thesis we are only interested in entropic quantities for bipartite quantum states which in some sense follow from it. We now define these.

Measures of Quantum Informational Content

Let $\rho_{AB} \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$,

- The *quantum mutual information* of the state is

  $$I(A:B) \equiv H(A) + H(B) - H(A,B) = D(\rho_{AB}||\rho_A \otimes \rho_B)$$

  which can be viewed as measuring the information stored in ρ_A that is also stored in ρ_B.

- The *quantum conditional entropy* of A given B of the state ρ_{AB} is

  $$H(A|B) \equiv H(A,B) - H(B)$$

which can be viewed as the amount of information in A which is not contained B.

Smoothed Entropies

To be able to do finite key analysis, Renner introduced the min- and max-entropies as well as their smoothed versions in [51].

Min- and Max-entropy

Let $\rho_{AB} \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$.

1. The *min-entropy* is defined as

$$H_{\min}(A|B) \equiv -\log \min\{\mathrm{Tr}(\sigma_B) : \mathbb{1}_A \otimes \sigma_B \succeq \rho_{AB}\} \ .$$

   Operationally, this is the maximum amount of uniform randomness which one can guarantee to extract from the register A given the side information of register B.

2. The *max-entropy* is defined as

$$H_{\max}(A|B) \equiv \log \| \mathrm{Tr}_A \, \Pi^{\rho_{AB}} \|_\infty$$

   where $\|\cdot\|_\infty$ is the spectral norm and $\Pi^{\rho_{AB}}$ is the projection onto the support of ρ_{AB}. Operationally, this is the maximum amount one can compress the register A, without any risk of failure, given the side information of register B.

Note: Here we have used the definitions as written in [23] as they are more condensed than those given in [51]. We also note the max-entropy written here is the quantum version of the Hartley entropy which was proposed in Renner's thesis [51] and has the presented operational interpretation. This is defined using the $\alpha = 0$ Petz quantum Renyi divergence. These days people *generally* define the max-entropy using the $\alpha = 1/2$ sandwiched (or 'minimal') quantum Renyi divergence as that max-entropy then satisfies a duality relation with the min-entropy [64]. See Section 5.2 of [62] for these definitions and Pg. 56 of [61] for the discussion of the difference in choice of max-entropy.

The definitions given above can be seen as reasonable definitions of min- and max-entropy in the sense that if one considers a classical probability distribution and lets the

register B be trivial. Then $H_{\min}(A|B) = H_{\min}(A) = -\log \|\rho_A\|_\infty$ which is the entropy of the most likely outcome and thus will contribute the least entropy, and $H_{\max}(A|B) = H_{\max}(A) = \log \operatorname{rank} \rho_A$ which is the entropy of the uniform distribution on the support of ρ_A which would maximize the entropy on the support.

Unfortunately, the min- and max-entropy may change rapidly when the quantum state is barely changed under the trace norm,[1] and so they need to be 'smoothed.'

Smoothed Min- and Max-Entropies

Let $\rho_{AB} \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, the *smoothed min-entropy* is defined as

$$H_{\min}^{\varepsilon}(A|B) \equiv \max_{\rho \in \mathcal{B}^{\varepsilon}(\rho)} H_{\min}(A|B) \tag{2.4}$$

and the *smoothed max-entropy* is defined as

$$H_{\max}^{\varepsilon}(A|B) \equiv \min_{\rho \in \mathcal{B}^{\varepsilon}(\rho)} H_{\max}(A|B) \tag{2.5}$$

where for both $\mathcal{B}^{\varepsilon}(\rho) \equiv \{\overline{\rho} \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B) : \|\rho - \overline{\rho}\|_1 \leq \varepsilon\}$.

Both functions have the operational interpretations of their respective unsmoothed versions except one views ε as the accepted probability of the task failing.

I note that since 2009, $\mathcal{B}^{\varepsilon}(\rho)$ is generally defined in terms of the purified distance [64] rather than the trace norm. However, for our purposes this will be sufficient, as we only need results from Renner's thesis from 2005 [51].

While the smooth entropy calculus has a myriad of interesting results, the only one we will need for this thesis is the following:[2]

[1] There is a good example of this in terms of max-entropy at the beginning of Section 3.2 of [51]. It is harder to construct an example for min-entropy, but it must be the case by the duality of min- and max-entropy (Definition 2 of [35])

[2] The well-read reader who knows where we are going in this thesis will ask why we use this result which may be looser than the Quantum Asymptotic Equipartion (QAEP) Theorem [63]. In effect this is because the statement of that theorem relies on a term, $\Upsilon(A|B)_{\rho|\rho}$, which depends on $H_{\min}(A|B)_\rho$ and $H_{\max}(A|B)_\rho$. This would require solving two extra SDPs, which may have numerical error, for a correction term that may not be that much better.

Variation of Corollary 3.3.7 of [51]

Let $\rho_{XB} \in \mathrm{D}(\mathcal{H}_X \otimes \mathcal{H}_B)$ be a classical-quantum state. Then for any $\varepsilon \geq 0$,

$$\frac{1}{n} H_{\min}^{\varepsilon}(\rho_{XB}^{\otimes n} | \rho_B^{\otimes n}) \geq H(XB) - H(B) - \delta \tag{2.6}$$

where $\delta = 2\log_2(\mathrm{rank}(\rho_X) + 3)\sqrt{\frac{\log_2(2/\varepsilon)}{2}}$.

The 'variation' is simply that there is a well-documented mistake in the argument of the square root in Theorem 3.3.6 of [51] which has been corrected in this statement and that we take a less loose bound presented in the proof of Corollary 3.3.7 at the expense of it not being in terms of entropies.

2.1.4 de Finetti Reductions

The last tool quantum information theoretic tool we will need in this thesis are de Finetti theorems. Historically, information theory has been interested in identically and independently distributed (i.i.d.) random variables. This means each random variable in a sequence $(X_1, X_2, ..., X_n)$ has been sampled from the same probability distribution p and that the outcome of each random variable is independent of the others, $p(X_i = x_1 \wedge X_j = x_2) = p(X = x_1) * p(X = x_2)$ for all $i, j \in [n]$. In the language of quantum mechanics, this is when one considers a quantum state of the form $\rho^{\otimes n}$.

To calculate properties of large systems (such as entropies) it is normally easiest to consider i.i.d. states as they are of a very nice form, with nice behaviour under measures (such as in Eqn. 2.6). However, assuming i.i.d. structure is quite unreasonable (if we send a laser pulse that is prepared the same way every time through a fiber optic, it will not come out of the fiber optic identically every time). Therefore we need an argument for how we can reduce more general quantum states to i.i.d. quantum states. These always require some promise on the symmetry of the general quantum state. We present two such theorems here after defining a few classes of operators related to the symmetries necessary. The first theorem is a statement of the Quantum de Finetti theorem from [13]. The second is a statement of the Finite Quantum de Finetti theorem from [51]. Those who may want/need a further discussion of de Finetti reductions (as well as other reductions to i.i.d.) in terms of quantum information processing tasks, we suggest they peruse [2].

Permutation Invariant and Exchangeable States

Consider $\mathcal{H}_A$. Consider $|\psi\rangle \in \mathcal{H}_A^{\otimes n}$ and $\rho_n \in \mathrm{D}(\mathcal{H}_A^{\otimes n})$. For each permutation of n elements, $\pi \in \mathcal{S}_n$, let the unitary operator which permutes the n Hilbert systems be W_π.

A pure state $|\psi\rangle$ is permutation invariant if it satisfies

$$|\psi\rangle = W_\pi |\psi\rangle \quad \forall \pi \in \mathcal{S}_n \ .$$

Permutation invariant pure states lie in the symmetric subspace which is defined as:

$$\mathrm{Sym}(\mathcal{H}_A^{\otimes n}) \equiv \{u \in \mathcal{H}_A^{\otimes n} : W_\pi u = u \ \ \forall \pi \in \mathcal{S}_n\} \ .$$

Generalizing from pure states, a density matrix ρ_n is *exchangeable* if

$$\rho_n = W_\pi \rho_n W_\pi^\dagger \ \ \forall \pi \in \mathcal{S}_n \ .$$

A state ρ_n is *n+k-exchangeable* if for some fixed $k \geq 0$ there exists $\rho_{n+k} \in \mathrm{D}(\mathcal{H}_A^{\otimes (n+k)})$ such that ρ_{n+k} is exchangeable and $\mathrm{Tr}_{A^{\otimes k}} \rho_{n+k} = \rho_n$.

A state ρ_n is *infinitely exchangeable* if for all $k \in \mathbb{N}$ there exists $\rho_{n+k} \in \mathrm{D}(\mathcal{H}_A^{\otimes (n+k)})$ such that ρ_{n+k} is exchangeable and $\mathrm{Tr}_{A^{\otimes k}} \rho_{n+k} = \rho_n$.

One might note that i.i.d. states are infinitely exchangeable as $\forall k \in \mathbb{N}$, $\mathrm{Tr}_{A^{\otimes k}} \rho^{\otimes n} \otimes \rho^{\otimes k} = \rho^{\otimes n}$. However, i.i.d. states are not the only exchangeable states. The de Finetti theorems tell us relationships between these types of states and (almost) i.i.d. states.

Quantum de Finetti Theorem [13, 33]

Let $\rho_n \in \mathrm{D}(\mathcal{H}_A^{\otimes n})$ be an infinitely exchangeable state. Then there exists a measure ν on $\mathrm{D}(\mathcal{H}_A)$ such that

$$\left\| \rho_n - \int_{\sigma \in \mathrm{D}(\mathcal{H}_A)} \sigma^{\otimes n} \nu(\sigma) \right\|_1 = 0 \tag{2.7}$$

One can see this tells us that infinitely exchangeable states are in effect just mixtures of i.i.d. states. Of course infinitely exchangeable states are also a rather untenable demand in experiment and so there exist Finite Quantum de Finetti theorems. We present the one from [51] which tells us how well one can decompose an $n+k$-exchangeable state into

states that are *almost like* mixtures of i.i.d. states.

Space of Almost i.i.d. States [51]

Consider the Hilbert space $\mathcal{H}_A$. Let $0 \leq m \leq n$. Fix a pure state $|\theta\rangle \in \mathcal{H}_A$. Consider the set of pure states which are partially i.i.d. pure states with respect to $|\theta\rangle$:

$$\mathcal{V}(\mathcal{H}_A^{\otimes n}, |\theta\rangle^{\otimes m}) \equiv \{W_\pi(|\theta\rangle^{\otimes m} \otimes |\Psi\rangle) : \pi \in S_n, |\Psi\rangle \in \mathcal{H}_A^{\otimes(n-m)}\}$$

We can then consider the subspace of symmetric pure states which are composed of partially i.i.d. states with respect to $|\theta\rangle$:

$$\mathrm{Sym}(\mathcal{H}_A^{\otimes n}, |\theta\rangle^{\otimes m}) \equiv \mathrm{Sym}(\mathcal{H}_A^{\otimes n}) \bigcap \mathrm{span}\left(\mathcal{V}(\mathcal{H}_A^{\otimes n}, |\theta\rangle^{\otimes m})\right)$$

we call this space the space of almost i.i.d. states as they are symmetric states made up of linear combinations of partially i.i.d. states and if $m = n$ one recovers the space of i.i.d. states.

Finite Quantum de Finetti Theorem [51]

Let $\rho_{n+k} \in \mathrm{Sym}(\mathcal{H}_A^{\otimes(n+k)})$. Let $0 \leq r \leq n$. Then there exists a measure ν on the unit sphere $\mathcal{S}_1(\mathcal{H}_A)$ and a pure state $\sigma^{|\theta\rangle} \in \mathrm{Sym}(\mathcal{H}_A^{\otimes n}, |\theta\rangle^{\otimes n-r})$ for each $|\theta\rangle \in \mathcal{S}_1(\mathcal{H}_A)$ such that

$$\left\| \mathrm{Tr}_{A^{\otimes k}}(\rho_{n+k}) - \int_{|\theta\rangle \in \mathcal{S}_1(\mathcal{H}_A)} \sigma^{|\theta\rangle} \nu(|\theta\rangle) \right\|_1 \leq 2e^{-\frac{k(r+1)}{2(n+k)} + \frac{1}{2}\dim(\mathcal{H}_A)\ln k} \tag{2.8}$$

2.2 Quantum Key Distribution

Quantum key distribution (QKD) is the quantum information processing task of establishing a secret key between two parties, traditionally named Alice and Bob. Fundamentally, a QKD protocol is a protocol in which Alice and Bob try to generate a pair of keys (S_A, S_B) which are the same (*correct*) and unknown to Eve (*secret*). To implement this protocol, Alice and Bob are given access to a classical channel which Eve can listen to but not tamper with (*authenticated classical channel*) and a quantum channel which Eve may tamper with how she pleases (*insecure quantum channel*). The subprotocols involved in QKD can be partitioned into the ones which use the quantum channel, referred to as

the *quantum phase* of the protocol, and the subprotocols involving the classical channel, referred together as the *classical phase* of the protocol. The type of quantum state used to implement the quantum phase is often used to partition QKD protocols into two types: prepare and measure (PM) QKD protocols and entanglement-based (EB) QKD protocols. In prepare and measure protocols, Alice sends signals to Bob from an ensemble $\{p_x, |\varphi_x\rangle\}$ where p_x is the a priori probability she sends $|\varphi_x\rangle$. In entanglement-based protocols Alice (or Eve!) prepares a joint quantum state ρ_{AB} which both Alice and Bob receive a half of to measure. In this thesis we are interested in security proofs, and it turns out one can always prove security of a PM QKD protocol by means of an equivalent EB QKD protocol using the source-replacement scheme [18, 24]. This allows us, without loss of generality, to present the general QKD protocol as being entanglement-based and then design our numerics to prove security for EB QKD protocols. In this section, we present the generic QKD protocol, the source-replacement scheme, and the definition of security for a QKD protocol.

2.2.1 General QKD Protocol

Following Fig. 2.1, without loss of generality,[3] we now present the general implementation of a QKD protocol.

1. *State Preparation and Transmission*: Alice prepares an entangled quantum state ρ_{AB} and sends half of it to Bob. Alice does this N times.

2. *Measurement and Data Partitioning*: Alice and Bob measure each of the N entangled quantum states ρ_{AB} and store the data pertaining to each measurement. In view of future communication, they partition their respective data from each measurement, indexed by i, into private information, $\overline{A}_i, \overline{B}_i$, and public information $\widetilde{A}_i$, $\widetilde{B}_i$ which they announce publicly.

3. *Parameter Estimation*: Alice and Bob announce their fine-grained data about some random subset of the N signals of size m to construct the frequency distribution $f(a, b)$. If $f(a, b)$ is in a set of pre-agreed upon accepted statistics, $\mathcal{Q}$, Alice and Bob proceed. Otherwise, they abort the protocol.

4. *Announcements and General Sifting*: Alice and Bob throw out results of some subset of the $N - m$ signals based on public information made available in Step 2. The

[3] While the description here is referred to as entanglement-based, as we will review later, one can mathematically describe prepare-and-measure-based protocols also as entanglement-based protocols.

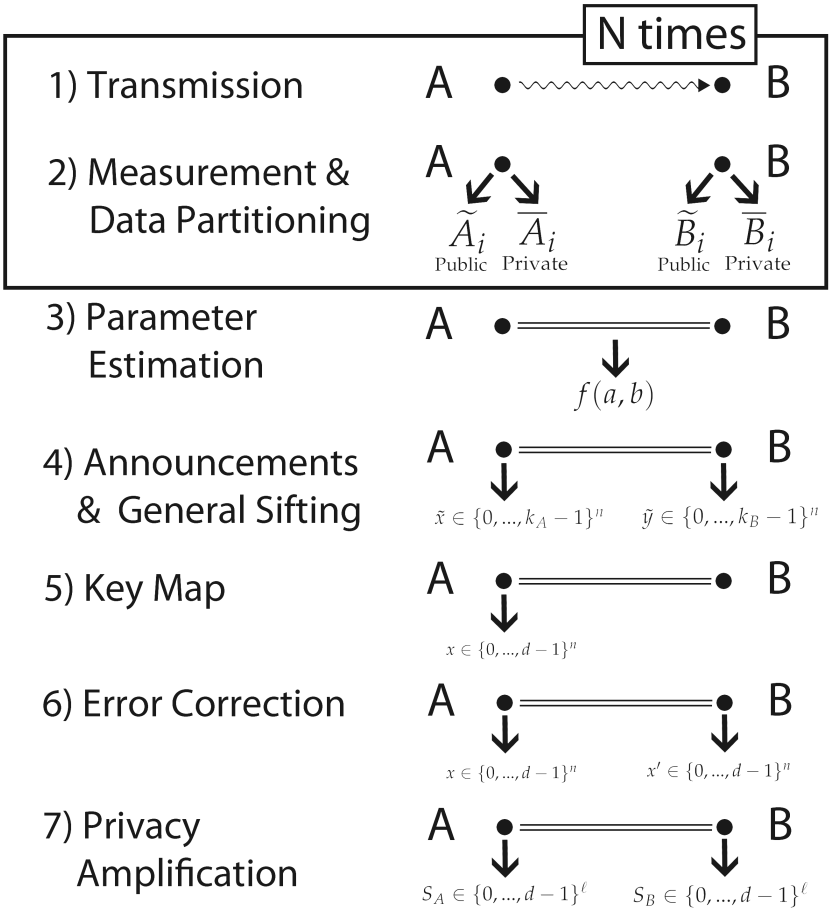


Figure 2.1: General QKD Protocol.

remaining private information forms their raw keys $\tilde{x} \in \{0,...,k_A-1\}^n$ and $\tilde{y} \in \{0,...,k_B-1\}^n$ where k_A and k_B are the number of possible outcomes for Alice and Bob's measurements respectively.

5. *Key Map*: Alice computes the key map,[4] a function of both her private data and the public data of both parties to obtain a key, $x \in \{0,1,...,d-1\}^n$.

6. *Error Correction*: Alice and Bob publicly communicate to try and get $\tilde{y}$ and x to agree and thus Bob obtains $x' \in \{0,1,...,d-1\}^n$.

[4]Alternatively, Bob can compute the key map. This is commonly referred to as reverse reconciliation, and in this case Alice and Bob's roles are reversed in steps 5. and 6.

7. *Privacy Amplification:* Alice and Bob produce their final keys by using a two-universal hash function on the key map result x (Theorem 5.5.1 of [51]). Privacy amplification ends with Alice and Bob having keys S_A and S_B respectively.

2.2.2 Source-Replacement Scheme

As mentioned earlier, the source replacement scheme is a formulation of the prepare-and-measure protocol in the language of entanglement-based protocols, which allows the description of the QKD protocol in the previous section to be general. It was first made use of in the analysis of BB84 [8] and Gaussian CV-QKD [28]. The general method for the equivalence was then expounded in [18, 24]. By formulating the prepare and measure protocol in the language of entanglement-based protocols, whatever the key rate is for the entanglement-based protocol is also the key rate for the original prepare and measure protocol.

Imagine a prepare and measure protocol in which Alice sends the ensemble $\{p_x, |\varphi_x\rangle\}$ where p_x is the a priori probability of sending the signal state $|\varphi_x\rangle$. By the source-replacement scheme, it is equivalent for Alice to prepare the entangled state:

$$|\Phi\rangle_{AS} = \sum_x \sqrt{p_x}\, |x\rangle_A |\varphi_x\rangle_S \ .$$

Alice first sends Bob's portion of the state, the signal space S, to Bob through a quantum channel $\mathcal{E} : \mathcal{H}_S \to \mathcal{H}_B$ leading to the resulting joint state:

$$\rho_{AB} = (\mathrm{id}_{\mathcal{H}_A} \otimes \mathcal{E})(|\Phi\rangle \langle\Phi|_{AS})$$

where $\mathrm{id}_{\mathcal{H}_A}$ is the identity channel on the A space. After Alice performs a local projective measurement on the A space, she effectively sends $|\varphi_x\rangle$ to Bob with probability p_x just like in the prepare-and-measure scheme. Consequently Bob receives the conditional state

$$\rho_B^x = \frac{1}{p_x} \mathrm{Tr}_A[\rho_{AB}(|x\rangle \langle x| \otimes \mathbb{1}_B)]$$

Assume that in the original prepare-and-measure protocol Alice and Bob ended up with a joint frequency distribution $f(x, b)$ where $b \in \Sigma$ and $|\Sigma|$ is the number of POVM elements for Bob's POVM $\{\Gamma_b^B\}_{b\in\Sigma}$. It follows by the source-replacement scheme that asymptotically it is equivalent for us to constrain ρ_{AB} by

$$\mathrm{Tr}\big(\rho_{AB}(|x\rangle \langle x| \otimes \Gamma_b^B)\big) = f(x, b) \ \ \forall x, b \ .$$

However, these observations alone will not be sufficient under source-replacement when optimizing over a set of states as these correlations can be satisfied by a separable state, and the whole point of source-replacement is in effect to make all QKD security proofs about entanglement.

That a separable state can satisfy the above observations can be seen as follows. Let $|\Psi'\rangle_{AA'} = \sum_k |k\rangle\langle k| \otimes |\phi_k\rangle\langle\phi_k|$. Let $\rho'_{AB} = \mathrm{id}_{\mathcal{H}_A} \otimes \mathcal{E}_{A'\to B'}(|\Psi'\rangle\langle\Psi'|)$. Then we see that:

$$\mathrm{Tr}\left(\rho'_{AB}\,|k\rangle\langle k| \otimes \Gamma_b^B\right) = p(k,b) = \mathrm{Tr}\left(\rho_{AB}\,|k\rangle\langle k| \otimes \Gamma_b^B\right) .$$

To solve this issue one needs constraints which determine the marginal of the joint state, ρ_A. A simple manner of doing this is to add an additional set of observations of the form $\{\mathrm{Tr}\left(\rho_A \Gamma_i^A\right) = \gamma_i\}_{i\in\Sigma}$ which fix the marginal ρ_A.

2.2.3 Security

The history of QKD security is very interesting. The original security definition was about the mutual information between Alice's key S_A and the outcome, O, of Eve measuring the quantum state(s) she used to eavesdrop: $I(S_A : O) \leq \varepsilon$. (It is sufficient to show the mutual information with Alice's key is small as ideally Bob's key is identical.) However, it was shown that while the key S_A is secure, it is only secure so long as it is not used [36]. This meant that while the *output* of the protocol was secure under this security definition, you could not *use* the key securely. The ability to use the key securely after the protocol is referred to as *universal composability* and it is the goal of QKD implementations to produce composably secure keys.

There exist a few frameworks for determining the security definition of composably secure keys and in that sense all are sufficient for the security definition in this thesis. However, of particular interest to the author of the thesis is the Abstract Cryptography framework [47] which, by its construction, is a framework which *derives* only composable security definitions from abstract notions. The security definition for QKD has been derived in [50] using the Abstract Cryptography framework which is perhaps its most rigorous defense of being a composably secure definition. We refer any reader to [47, 50] for further justification of the security definition.

Composable ε-Secure Key

A key is ε-secure (sometimes referred to as ε-secret) with respect to an adversary

Eve if the joint state, $\rho_{S_A E'}$, between Alice's key, S_A, and Eve's knowledge, E' is if

$$\frac{1}{2}\|\rho_{S_A E'} - \tau_{S_A} \otimes \rho_{E'}\|_1 \leq \varepsilon \tag{2.9}$$

where $\tau_{S_A} = \frac{1}{|\mathcal{S}_A|}\sum_{s\in\mathcal{S}_A} |s\rangle\langle s|$ and $\rho_{S_A E'}$ is the output of the protocol conditioned on not aborting.

1st Security Definition of a QKD Protocol

Definition 1. *A QKD protocol is ε-secure if for all input states, the output $\rho_{S_A S_B E'}$ satisfies*

$$\frac{1}{2}\|\rho_{S_A S_B E'} - \tau_{S_A S_B} \otimes \rho_{E'}\|_1 \leq \varepsilon \tag{2.10}$$

where $\tau_{S_A S_B} = \frac{1}{|\mathcal{S}_A|}\sum_{s\in\mathcal{S}_A} |s\rangle\langle s| \otimes |s\rangle\langle s|$, $\rho_{S_A S_B E'}$ is the output of the protocol conditioned on not aborting, and $\rho_{E'} \equiv \mathrm{Tr}_{S_A S_B}(\rho_{S_A S_B E'})$.

It is crucial to note three aspects of the definition of ε-secure key and ε-secure QKD protocol. The first is that because these definitions are in terms of trace distance, by the operational interpretation of trace distance, we can see that the ε-security may be interpreted as the demand that the optimal amount the output key and the ideal key could ever be distinguished is ε from random guessing. The second is that these definitions were defined on the output of the state conditioned on not aborting.[5] This means that one could re-write the condition of ε-secure key as

$$(1 - p_{abort})\frac{1}{2}\|\hat{\rho}_{S_A E'} - \tau_{S_A} \otimes \rho_{E'}\|_1 \leq \varepsilon$$

where $\mathrm{Tr}(\rho_{S_A E'}) = (1 - p_{abort})$ and $\hat{\rho}_{S_A E'} = \frac{1}{\mathrm{Tr}\left(\rho_{S_A E'}\right)}\rho_{S_A E'}$. Note that the value of p_{abort} will be dependent on the input state given this definition. The equivalent equation for an ε-secure QKD protocol is straightforward. This re-writing makes it clear that in some sense you can build some very impractical but ε-secure protocols. For example, if I just have a QKD device that always aborts, then it is 0-secure as $(1 - p_{abort}) = 0$. Furthermore you could construct impractical QKD protocols which will abort with high probability to drive down the ε-term in exchange for a better key rate when the protocol outputs a key.

[5] We note one could define the security not conditioned on passing. However, this immediately implies the definition conditioned on passing as when the QKD protocol aborts the distance between the ideal output and the implementation output is zero. See the discussion leading up to Eqn. 3.2 of [4].

We will return to this point in Section 4.2.1. The final point is to note that the security definition requires a fixed output key space $\mathcal{S}_A$. This means that for any input, either the protocol must abort or the output must be a key in this key space. This will ultimately imply that the output key of the protocol will need to be of fixed length. We will return to this point in Section 3.5.2.

With these observations we can note there are in fact alternative, but related, security definitions for QKD protocols.

2nd Security Definition of a QKD Protocol

Definition 2. *A QKD protocol is ε-secure where $\varepsilon = \varepsilon' + \varepsilon''$ if the QKD protocol outputs a ε'-secret key and is ε''-correct (i.e. $\Pr[S_A \neq S_B \wedge \neg abort] \leq \varepsilon''$).*

3rd Security Definition of a QKD Protocol

Definition 3. *Let $\Phi^{QKD} : (\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes N} \to \mathcal{H}_{S_A} \otimes \mathcal{H}_{S_B}$ be the CPTNI map which represents the implemented QKD protocol. Let Ψ^{QKD} represent the ideal QKD protocol. The implemented QKD protocol is ε-secure if*

$$\frac{1}{2}\|\Phi^{QKD} - \Psi^{QKD}\|_\diamond \leq \varepsilon \ .$$

Operationally, one may view this as the deviation from random guessing one could achieve in discriminating between the ideal and implemented QKD protocols simply by testing on inputs.

It is worth showing quickly how these security definitions relate as these relations will be used.

Proposition 1. *The first QKD security definition is implied by the second.*

Proof. Assume a QKD protocol outputs a ε'-secret and ε''-correct key. Consider Eqn. 2.10:

$$\begin{aligned}
&\frac{1}{2}\|\rho_{S_AS_BE'} - \tau_{S_AS_B} \otimes \rho_{E'}\|_1 \\
=&\frac{1}{2}\|\rho_{S_AS_BE} - \rho^{S_A=S_B}_{S_AS_BE} + \rho^{S_A=S_B}_{S_AS_BE} - \tau_{S_AS_B} \otimes \rho_E\|_1 \\
\leq&\frac{1}{2}\|\rho_{S_AS_BE} - \rho^{S_A=S_B}_{S_AS_BE}\|_1 + \frac{1}{2}\|\rho^{S_A=S_B}_{S_AS_BE} - \tau_{S_AS_B} \otimes \rho_E\|_1 \\
\leq&\Pr[S_A \neq S_B \wedge \neg\text{abort}] + \varepsilon' \\
\leq&\varepsilon' + \varepsilon'' \equiv \varepsilon
\end{aligned}$$

□

Proposition 2. *The third QKD security definition is always implied by the first and they are equivalent in the case that one finds the smallest ε which satisfies the first definition.*

Proof. Let the first security definition be satisfied for a given QKD protocol. By the definition of the diamond norm (Eqn. 2.2),

$$\begin{aligned}
&\frac{1}{2}\|\Phi^{\text{QKD}} - \Psi^{\text{QKD}}\|_\diamond \\
=&\max_{\sigma \in \mathcal{H}^{\otimes N}_{ABE} : \|\sigma\|_1 \leq 1} \|(\Phi^{\text{QKD}} \otimes \text{id}_{\mathcal{H}_E})(\sigma) - (\Psi^{\text{QKD}} \otimes \text{id}_{\mathcal{H}_E})(\sigma)\|_1 \\
=&\max_{\sigma \in \mathcal{H}^{\otimes N}_{ABE} : \|\sigma\|_1 \leq 1} \|(\Phi^{\text{QKD}} \otimes \text{id}_{\mathcal{H}_E})(\sigma) - \tau_{S_AS_B} \otimes \rho_{E'}\|_1 \\
\leq&\varepsilon
\end{aligned}$$

Where the inequality becomes an equality if ε is the tightest security parameter under the first security definition. □

2.2.4 Classes of Security Proofs

To prove security, we now know in effect one wants to prove that their QKD protocol sufficiently maps to some mathematical promise that the output of the protocol, which includes Eve's output, is of the form in Eqn. 2.10. This cannot be done without any assumptions. Therefore there is a trade-off between how few assumptions about the protocol one makes and how easily they can implement the protocol and generate key from the protocol. Assumptions come in roughly two forms: assumptions about the devices used

and assumptions about what Eve does.

There are many choices one can make in the assumptions on the devices. Here we define the two types of assumptions (device-dependent and measurement-device-dependent) which we consider in Chapter 5 along with a third which we will use for comparison at times.

(Subset of) Assumptions on Devices

- *Device-Dependent (DD) QKD*: Alice and Bob's devices are both promised to behave just as they did when they were constructed and tested prior to using them for the protocol.
- *Measurement-Device-Independent (MDI) QKD*: Alice and Bob's devices both send signals and are promised to behave just as they did when they were constructed and tested. However, their signals are both sent to a third untrusted party who performs the measurement and makes announcements. This third party's (measurement) device is completely uncharacterized. Conceptually, MDI QKD may be seen as a subset of DD QKD protocols as Alice and Bob's devices are still characterized.
- *Device-Independent (DI) QKD*: Alice and Bob's devices are not trusted, and they must rely on a proof of quantum correlation between Alice and Bob to generate a secret key.

Beyond our assumptions on the devices, to prove security against Eve, one must first state what Eve can do. The power of Eve is normally broken down into two types. In both cases Eve has as large of a quantum memory available to her as she wants. Under the *collective attack* assumption, Eve is allowed to have a new ancillary system interact with each signal Alice sends to Bob. She may then store all of her ancillary systems and measure them at any point (even once the key has been generated or even used). Under the *coherent attack* assumption, Eve may interact with all of the signals Alice sends to Bob in a coherent fashion, which may lead to the total state Alice and Bob use to generate their key being highly entangled across different 'signals.' Lastly, while people generally assume collective attacks lead to signal state structures that are i.i.d. (i.e. the total quantum state of the protocol is of the form $\rho^{\otimes N}$), the definition of collective attacks does not seem to imply this, so we refer to this special case as *i.i.d. collective* attacks.

Coherent attacks are what one would want to prove security against as it is clearly

the least restrictive demand on Eve (in fact it tells Eve she can do whatever she pleases according to quantum mechanics or any theory which retains quantum mechanics [15]). In this thesis we will show that our numerical solver can calculate key rates for coherent attacks, but we will focus on i.i.d. collective attacks as the belief is there will soon exist a proof method for finite key analysis which gets nearly the i.i.d. collective attack secret key rate under the coherent attack assumption.

2.3 Semidefinite Programming

Semidefinite programming is a specific subfield of convex optimization which optimizes over the positive semidefinite cone. First we give a short review the standard form of a semidefinite programming. We then introduce the specific tools which are useful in this thesis.

Semidefinite Program [66]

Let $\Psi \in \mathrm{T}(\mathcal{H}_A, \mathcal{H}_B)$ be a Hermitian-preserving map, $A \in \mathrm{Herm}(\mathcal{H}_A)$, and $B \in \mathrm{Herm}(\mathcal{H}_B)$. A semidefinite program is a triple (Ψ, A, B), with the following associated optimization problems:

$$\begin{aligned} &\text{minimize} && \langle A, X\rangle \\ &\text{subject to} && \Psi(X) = B \\ &&& X \in \mathrm{Pos}(\mathcal{X}) \end{aligned} \tag{2.11}$$

$$\begin{aligned} &\text{maximize} && \langle B, Y\rangle \\ &\text{subject to} && \Psi^\dagger(Y) \preceq A \\ &&& Y \in \mathrm{Herm}(\mathcal{Y}) \end{aligned} \tag{2.12}$$

where $\Psi^\dagger$ is the adjoint map of Ψ; that is, $\Psi^\dagger$ is the unique linear map that satisfies the adjoint equation $\langle Y, \Psi(X)\rangle = \langle \Psi^\dagger(Y), X\rangle$ for every $X \in \mathrm{L}(\mathcal{H}_A)$ and $Y \in \mathrm{L}(\mathcal{H}_B)$. Eqn.(2.11) is referred to as the *primal problem* and Eqn.(2.12) is referred to as the *dual problem*. We define $\mathcal{A} = \{X \in \mathrm{Pos}(\mathcal{X}) | \Psi(X) = B\}$ and $\mathcal{B} = \{Y \in \mathrm{Herm}(\mathcal{Y}) | \Psi^\dagger(Y) \preceq A\}$. These sets are referred to as the *feasible set* of the primal problem and dual problem, respectively.

Note: The definitions in Eqns. 2.11, 2.12 are not the exact definitions from [66] but they are equivalent.

By *weak duality*, for all semidefinite programs, the optimal value of the primal problem, denoted by α, is always greater than or equal to the optimal value to the dual problem, denoted by β. If a semidefinite program has that $\alpha = \beta$, it is said to have *strong duality*.

A sufficient condition to show strong duality for SDP is Slater's condition for the standard from presented here.

Theorem 3. *(Slater's Condition) For a semidefinite program (Ψ, A, B), if $\mathcal{A} \neq \emptyset$ and there exists a Hermitian operator Y which strictly satisfies the dual problem, that is, $\Psi^\dagger(Y) \prec A$, then $\alpha = \beta$ and the optimal value is obtained in the primal problem.*

2.3.1 Trace Norm Semidefinite Program

Semidefinite programming is a powerful tool for quantum information theory as most everything in quantum information theory is about positive semidefinite or Hermitian matrices. For the use of this thesis we will be primarily interested in the semidefinite program for the trace norm, or Schatten 1-norm, which was implicitly defined in Eqn. 2.1.

Trace Norm

Let $X \in \mathrm{L}(\mathcal{H}_A)$. The *trace norm* of X is:

$$\|X\|_1 = \mathrm{Tr}\left(\sqrt{X^\dagger X}\right) = \sum_{\lambda_a \in \mathrm{Spec}(X)} |\lambda_a| \tag{2.13}$$

One might have an intuition that this should have an SDP as the trace norm is a convex function. There in fact exists an SDP for calculating the trace norm for arbitary matrices $X \in L(\mathcal{H}_A, \mathcal{H}_B)$ [66], but we will only have to consider $X \in \mathrm{Herm}(\mathcal{H}_A)$ and so we present an arguably more intuitive primal and dual that work in this case. For the general case, we refer the reader to [66].

Trace Norm SDP for Hermitian Operators

Let $X \in \mathrm{Herm}(\mathcal{H}_A)$. The trace norm of $\|X\|_1$ is achieved by the following SDPs:

$$\begin{aligned} &\text{minimize} && \mathrm{Tr}(P) + \mathrm{Tr}(Q) \\ &\text{subject to} && P \succeq X \\ &&& Q \succeq -X \\ &&& P, Q \in \mathrm{Pos}(\mathcal{H}_A) \end{aligned} \tag{2.14}$$

$$\begin{aligned} &\text{maximize} && \langle X, R\rangle - \langle X, S\rangle \\ &\text{subject to} && 0 \preceq R \preceq \mathbb{1} \\ &&& 0 \preceq S \preceq \mathbb{1} \\ &&& R, S \in \mathrm{Herm}(\mathcal{H}_A) \end{aligned} \tag{2.15}$$

By definition of primal and dual problem (Eqns. 2.11 & 2.12 respectively), Eqn. 2.14 is the primal problem and Eqn. 2.15 is the dual problem. As we have stated the two problems

both obtain the trace norm for Hermitian X, i.e. the SDP satisfies strong duality, we prove that now for completeness by constructing the optimal solution for both SDPs.

Proof. Let $X \in \text{Herm}(\mathcal{H}_A)$ where $\mathcal{H}_A \cong \mathbb{C}^\Sigma$. By the spectral theorem, we can write $X = \sum_{k\in\Sigma} \lambda_k \Pi_k$ where $\{\Pi_k\}_{k\in\Sigma}$ are unique projectors and $\{\lambda_k\}_{k\in\Sigma} \subset \mathbb{R}$ are the eigenvalues of X. Define the following operators

$$\overline{P} \equiv \sum_{k:\lambda_k \geq 0} \lambda_k \Pi_k \qquad \overline{Q} \equiv \sum_{k:\lambda_k < 0} -\lambda_k \Pi_k$$
$$\overline{R} \equiv \sum_{k:\lambda_k \geq 0} \Pi_k \qquad \overline{S} \equiv \sum_{k:\lambda_k < 0} \Pi_k \ .$$

One might notice that $\overline{R}$ is the projector onto the positive eigenspace of X and $\overline{S}$ is the projector onto the negative eigenspace of X. Furthermore, $\overline{P}$ is the projection of X onto its positive eigenspace and $\overline{Q}$ is the projection of X onto its negative eigenspace.

Then it is clear by construction that $\overline{P}$, $\overline{Q}$ are feasible for the SDP in Eqn. 2.14 and by definition of the trace will have the objective function obtain the value of $\|X\|_1$ by Eqn. 2.13. Furthermore, any other feasible operator can only increase the objective function as both P, Q must be positive semidefinite.

Similarly, it is clear $\overline{R}, \overline{S}$ are feasible for the SDP in Eqn. 2.15 and that they lead to the objective function taking the value $\|X\|_1$. We therefore have constructed solutions to Eqns. 2.14 and 2.15 and shown them to be equal. As weak duality holds for all SDPs, and we have shown the primal and dual solutions may be equal, strong duality by definition must hold. □

It was important that we see this SDP has strong duality as ultimately we will be interested in another SDP which includes trace norms, and we might not expect it have strong duality if the trace norm itself does not.

Chapter 3

Theory of Finite Key Analysis

In this chapter we present the theory of finite key analysis. In some sense, as we will see, once one has fixed their assumptions about the output of the quantum phase of the protocol, finite key analysis is only about the analysis of the classical phase. For this reason, this chapter focuses on the subprotocols of the classical phase in sufficient detail for this thesis. Many of the ideas in this section are from Renner's PhD thesis [51]. However, one major result of this thesis is an improvement to the analysis of the parameter estimation subprotocol of quantum key distribution (Theorem 5), which is presented in this chapter. Furthermore, we present a discussion on calculating the completeness of device-dependent QKD protocols, which the author does not know to be in any other work. Lastly, so that the thesis is self-contained, we present a security proof for i.i.d. collective attacks which is a simplification of the proof for Theorem 6.5.1 of [51]. It is the security proof implicitly used in [53, 54] and works which followed from it, but it is not presented anywhere as far as we know. This security proof is what allows us to guarantee our numerical method will determine secure keys.

Subprotocol Security

Recall from Section 2.2.3 that the security of QKD can be reduced to making sure the protocol is ε'-secret and ε''-correct. As one would expect, the security of the subprotocols of QKD are what determine the security of the protocol. With a finite number of signals, no subprotocol can be implemented ideally. These imperfections are represented with ε-terms which contribute to the ε'-secrecy and ε''-correctness of the QKD protocol. We give these ε-terms now:

1. ε_{PE} can be viewed as the probability of the parameter estimation protocol not aborting and the state generated in the state transmission, which Alice and Bob tested m times, not being included in their security analysis.

2. $\bar{\varepsilon}$ is the probability of Eve knowing the key because for each state feasible according to parameter estimation, ρ_{AB}, Alice and Bob a priori consider the min-entropy of the state $\overline{\rho}_{AB}$ that maximizes the min-entropy over the set of states $\bar{\varepsilon}$-similar to ρ_{AB}. In other words, this is the probability of failure due to considering the smooth min-entropy (Eqn. 2.4) of the state in the security analysis rather than the min-entropy.

3. ε_{EC} is the probability that Alice and Bob do not abort the protocol and obtain outputs that differ, i.e. $x \neq x'$.

4. ε_{PA} is the probability that Alice and Bob do not abort the protocol and that the key is known to Eve because the privacy amplification failed.

One may view the rest of this chapter as giving an account of these ε-terms and how they arise in the security proof used in our numerics.

3.1 Parameter Estimation

Recall from the description of a generic QKD protocol (see Fig. 2.1) that Alice and Bob use their respective devices to perform measurements on the shared quantum state, ρ_{AB}, N times.[1] The goal of parameter estimation is to determine the density matrices which need to be considered in the security analysis given Alice and Bob sacrifice m signals for sampling. To make this clear, we first describe the subprotocol of parameter estimation in detail and then explain how the security statement put forth in [51] follows from the description of the protocol.

The protocol is as follows. It is always assumed Alice and Bob's measurement devices are separable. That is to say, if Alice's and Bob's measurements are decribed with POVMs $\{\widetilde{\Gamma}_i^A\}_{i\in\Sigma_A}$ and $\{\widetilde{\Gamma}_k^B\}_{k\in\Sigma_B}$, then their joint measurement is of the form $\{\widetilde{\Gamma}_{(i,k)}\}$ where $\widetilde{\Gamma}_{(i,k)} \equiv \Gamma_i^A \otimes \Gamma_k^B$. This can be viewed as an assumption on the fact that Alice and Bob's measurement devices are non-signaling.[2] By this assumption, for each signal, Alice gets

[1] Recall we are using the entanglement-based framework without loss of generality.

[2] This is a security assumption that is implied by device-dependent QKD. When one *only* makes this assumption without characterizing the device, one has something referred to as non-signalling QKD [1, 2, 46].

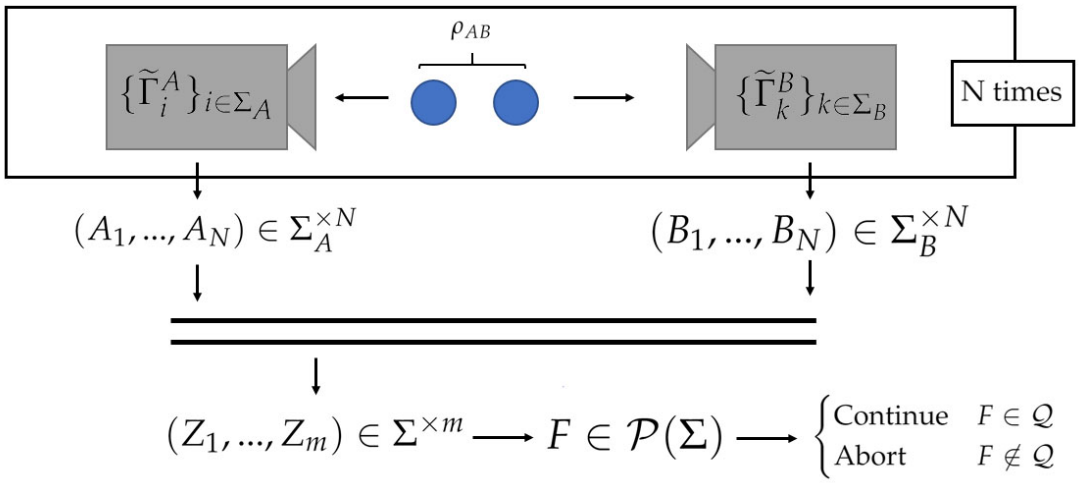


Figure 3.1: Diagram of the parameter estimation subprotocol. Alice and Bob perform their measurements resulting in their respective sequences. They use classical communication to construct a sequence of joint outcomes. From this they construct a frequency distribution F which they then use to decide whether or not to abort the protocol.

a classical read-out from her measurement apparatus which is unaffected by Bob's measurement apparatus and likewise for Bob. We can therefore define $\Sigma \equiv \Sigma_A \times \Sigma_B$ as the alphabet for the joint outcome of Alice and Bob, Z as the corresponding random variable, and $\{\widetilde{\Gamma}_j\}_{j\in\Sigma}$ as the corresponding POVM using the implicit mapping $\Sigma \to \Sigma_A \times \Sigma_B$. Following this construction, when Alice and Bob classically announce their respective outcomes for a subset of the N measurements of size m, they are able to construct a sequence $\vec{z} = (Z_1, ..., Z_m) \in \Sigma^{\times m}$. From this sequence they are able to construct a corresponding frequency distribution over Σ, sometimes referred to as a *type*:

$$F_{\vec{z}} \equiv \sum_{z\in\Sigma} \frac{|\{i \in [m] : Z_i = z\}|}{m} |z\rangle\langle z| \ . \tag{3.1}$$

Then if F is included in a set of frequency distributions Alice and Bob have agreed to accept, $\mathcal{Q}$, they continue the protocol. Otherwise they abort. See Fig. 3.1 for a diagrammatic depiction of this procedure.

Given the subprotocol, we would like to construct a set of states to consider in our security proof and a corresponding security claim about this set of states. This requires two observations.

First, we note that we cannot control what sort of attack Eve performs, and so we can

never apply a probability distribution over the set of density matrices tested. Instead, we need a way of determining the set of states our security analysis should consider purely on the relationship between the input density matrix, Alice and Bob's joint POVM $\{\widetilde{\Gamma}_j\}_{j\in\Sigma}$, and the set of accepted frequencies, $\mathcal{Q}$. For this reason, our security definition must be interested in determining the probability Alice and Bob will continue the protocol given they tested a specific density matrix.

The second observation is that we need to ignore (at least a subset of) the states which have sufficiently low probability of being accepted given Alice and Bob testing them. The need to ignore this set is worth explaining with an example. Imagine that Alice and Bob let Eve construct the joint qubit state to send it to Alice and Bob N times. Furthermore, Alice and Bob decide that they are only ever going to accept the parameter estimation if the resulting F is the same as if they had measured the maximally entangled state being sent to them all m times. Certainly for large (but finite) m most of the time Alice and Bob will abort if Eve does not send the maximally entangled state N times. However it is also the case that there is a non-zero probability that Alice and Bob each receive a maximally mixed state N times and yet their statistics from testing m times are as if they are ideally correlated. Therefore, it is necessary to ignore the states which have some very small probability of giving rise to F. A state which will lead to the parameter estimation subprotocol being aborted except with probability ε_{PE} is said to be ε_{PE}-securely filtered [51].

It follows from these observations that we need a quantitative method of determining the states that are ε_{PE}-securely filtered so as to throw out the states which are ε_{PE}-securely filtered by parameter estimation. This will construct a set one can consider in the security proof. We will now present in detail how this is done.[3]

Given the description of parameter estimation above, the parameter estimation subprotocol can be described using a CPTNI map constructed in the following manner. First, let $\Phi_{\text{PE}}^{\text{Step 1}} \in T((\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes m}, \mathbb{C}^{\Sigma\times m})$ be defined by

$$\Phi_{\text{PE}}^{\text{Step 1}}(X) = \sum_{\vec{j}\in\Sigma^{\times m}} \text{Tr}\Big(X\widetilde{\Gamma}_{\vec{j}}\Big) \Big|\vec{j}\Big\rangle\Big\langle\vec{j}\Big|$$

where $\widetilde{\Gamma}_{\vec{j}} \equiv \widetilde{\Gamma}_{j_1} \otimes ... \otimes \widetilde{\Gamma}_{j_m}$. In words, assuming X is a quantum state, this map takes the input for the entire parameter estimation subprotocol and maps it to the probability distribution over the possible observed sequences $\vec{j} \in \Sigma^{\times m}$ (i.e. it is the measurement

[3] While the result of these ideas was presented in [51], this discussion was not, and the detail is necessary to understand our new result (Theorem 5).

channel for the whole subprotocol). Next let $\Phi_{\text{PE}}^{\text{Step 2}} \in T(\mathbb{C}^{\Sigma^{\times m}}, \mathbb{C})$ be defined by:

$$\Phi_{\text{PE}}^{\text{Step 2}}(X) = \langle \Pi_{\mathcal{Q}}, X \rangle \text{ where } \Pi_{\mathcal{Q}} = \sum_{\vec{j} \in \Sigma^{\times m} : f_{\vec{j}} \in \mathcal{Q}} \left| \vec{j} \right\rangle \left\langle \vec{j} \right|$$

In words, $\Phi_{\text{PE}}^{\text{Step 2}}$ takes a probability distribution over the possible sequences $\vec{j} \in \Sigma^{\times m}$ for an input state and determines the probability that Alice and Bob will accept given the definition of $\mathcal{Q}$. Therefore, $\Phi_{\text{PE}} \equiv \Phi_{\text{PE}}^{\text{Step 2}} \circ \Phi_{\text{PE}}^{\text{Step 1}}$ takes a state tested in parameter estimation to the probability of Alice and Bob accepting.

If a total of N signals are exchanged, without any assumptions on the signals sent, the total protocol signal state is any $\rho_N \in \text{D}((\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes N})$. It follows that the marginal which was tested by the parameter estimation is any state $\sigma \in \text{D}((\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes m})$. We can therefore define the set of ε_{PE}-securely filtered states:

$$\mathbf{S}_{\leq \varepsilon_{\text{PE}}} \equiv \{\sigma \in \text{D}((\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes m}) : \Phi_{\text{PE}}(\sigma) \leq \varepsilon_{\text{PE}}\} \tag{3.2}$$

Unfortunately, m is going to be quite large and so in effect there is no hope of determining this set, which implies there is no way to determine the key rate over the complement of this set as we would like to do. Instead, we do the following trick. In the security proof (Section 3.5), we are going to promise ρ_N is an infinitely exchangeable state which will allow us to approximate ρ_N by a convex combination of i.i.d. states using the Quantum de Finetti Theorem (Eqn. 2.7). Therefore, rather than worrying about what arbitrary states will be ε_{PE}-securely filtered, we just worry about the set of i.i.d. states which will be ε_{PE}-securely filtered, which is clearly a subset of $\mathbf{S}_{\varepsilon_{\text{PE}}}$.[4]

3.1.1 Securely Filtered i.i.d. States

Assume that the input to the parameter estimation subprotocol, $\overline{\sigma}$, is of the form $\sigma^{\otimes m}$. We can then note the following:

$$\begin{aligned}
\Phi_{\text{PE}}^{\text{Step 1}}(\sigma^{\otimes m}) &= \sum_{\vec{j} \in \Sigma^{\times m}} \text{Tr}\left(\sigma^{\otimes m} \widetilde{\Gamma}_{\vec{j}}\right) \left| \vec{j} \right\rangle \left\langle \vec{j} \right| \\
&= \sum_{(j_1, j_2, \ldots, j_m) \in \Sigma^{\times m}} \text{Tr}\left(\sigma \widetilde{\Gamma}_{j_1}\right) \text{Tr}\left(\sigma \widetilde{\Gamma}_{j_2}\right) \ldots \text{Tr}\left(\sigma \widetilde{\Gamma}_{j_m}\right) |j_1, j_2, \ldots, j_m\rangle\langle j_1, j_2, \ldots, j_m| \ .
\end{aligned}$$

[4] This trick was first done in [51] where it was instead assumed the state would be $n+k$-exchangeable, so that Renner could use his Finite Quantum de Finetti theorem (Eqn. 2.8) to reduce to a convex combination of a set of states which are more complex than i.i.d. states.

This tells us that, since both the state and the measurement are in tensor product form, the probability of a given sequence of measurement outcomes is the same as the probability of the sequence from i.i.d. sampling from the probability distribution, $P \in \mathcal{P}(\Sigma)$, determined by Born's rule. Formally, the distribution P can be determined using the probability map $\Phi_{\mathcal{P}}$ using the POVM for the QKD protocol.

This reduces the question of ε_{PE}-securely filtering $\sigma^{\otimes m}$ to a sampling problem: If one constructs a frequency distribution F by sampling m times from a probability distribution P, except with probability at most ε_{PE}, how different from P could F be? If we can quantify this 'how different,' then any σ which induces a probability distribution P which is too different from any $F \in \mathcal{Q}$ must be ε_{PE}-filtered as no frequency distribution Alice and Bob would accept would arise from P except with ε_{PE} probability.

The quantitative answer to this sampling problem is as follows, which is a simplification of what is done in [51]:

Theorem 4. *If* $\|P - F\|_1 > \mu \equiv \sqrt{2}\sqrt{\frac{\ln(1/\varepsilon_{PE}) + |\Sigma| \ln(m+1)}{m}}$*, then, except with probability* ε_{PE}*,* F *did not arise from i.i.d. sampling from* P*.*

Proof. By Theorem 11.2.1 of [60], given a frequency distribution F constructed from sampling i.i.d. random variables from a probability distribution P which has $|\Sigma|$ outcomes,

$$\text{Prob}[D(F||P) > \epsilon] \leq 2^{-m(\epsilon - |\Sigma| \frac{\log_2(m+1)}{m})}$$

Furthermore, Lemma 11.6.1 of [60] states:

$$\sqrt{2 \ln 2 D(F||P)} \geq \|F - P\|_1$$

Therefore,

$$\begin{aligned}
&\text{Prob}\left[\|F - P\|_1 > \sqrt{2 \ln 2 \epsilon}\right] \\
\leq &\text{Prob}\left[\sqrt{2 \ln 2 D(F||P)} > \sqrt{2 \ln 2 \epsilon}\right] \\
\leq &2^{-m(\epsilon - |\Sigma| \frac{\log_2(m+1)}{m})} \\
\equiv &\varepsilon_{\text{PE}}
\end{aligned}$$

Then except with probability ε_{PE}, $\|F - P\|_1 \leq \sqrt{2 \ln 2 \varepsilon} \equiv \mu$ We now just solve for μ using arithmetic:

$$\begin{aligned}
\varepsilon_{\text{PE}} &= 2^{-m(\frac{\mu^2}{2 \ln 2} - |\Sigma| \log_2(m+1)/m} \\
\Rightarrow \mu &= \sqrt{2}\sqrt{\frac{\ln(1/\varepsilon_{\text{PE}}) + |\Sigma| \ln(m+1)}{m}}
\end{aligned}$$

□

It therefore follows from Theorem 4 that, the set of i.i.d. states which are ε_{PE}-securely filtered by the parameter estimation subprotocol are defined by:

$$\mathbf{S}^{\text{i.i.d.}}_{\leq\varepsilon_{\text{PE}}} = \{\sigma \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B) : \min_{F\in\mathcal{Q}} \|\Phi_{\mathcal{P}}(\sigma) - F\|_1 > \mu\}$$

and therefore, under the assumption of i.i.d. collective attack, one could optimize the key rate over:

$$\mathbf{S}^{\text{simple}}_{\mu} \equiv \{\sigma \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B) : \min_{F\in\mathcal{Q}} \|\Phi_{\mathcal{P}}(\sigma) - F\|_1 \leq \mu\} \tag{3.3}$$

as all other states would only be accepted with probability ε_{PE}.

3.1.2 Multiple Coarse-Grainings

While the previous section is sufficient for obtaining key rates, we are interested in the optimal key rates within the proof method. One would expect the previous section to not be sufficient in general as one would notice that $\{\sigma^{\otimes m} : \sigma \in \mathbf{S}^{\text{i.i.d.}}_{\leq\varepsilon_{\text{PE}}}\} \subset \mathbf{S}_{\leq\varepsilon_{\text{PE}}}$ and so we might expect there to be cases where getting a better approximation of $\mathbf{S}_{\leq\varepsilon_{\text{PE}}}$ could improve the key rate. This argument is furthered by the following simple intuition. The variation bound μ in Theorem 4 is a function of the size of Σ. Therefore, μ's size is directly determined by the size of Σ. It follows that if in the parameter estimation subprotocol were defined for the 'fine-grained' data which has a large alphabet, the variation bound μ may be large. However, if the parameter estimation subprotocol were defined over 'coarse-grained' (i.e. 'data processed') data, the alphabet is smaller and so the variation bound μ is smaller if all other terms are as before. This implies defining the parameter estimation subprotocol over coarse-grained data may improve the keyrate. We will in fact see in Chapter 5 that coarse-graining the fine-grained data will improve the keyrate. However, this is an issue as this would suggest that Alice and Bob throwing out information about the protocol can improve their security against Eve. We now solve this issue. First we introduce the mathematical framework of coarse-graining, and then we solve the issue with Theorem 5 and Corollary 6. Theorem 5 is the most general way of stating the resolution. Corollary 6 is what will be useful for the numerics and is (a slight generalization of) what is stated in our paper [26].

Framework of Coarse-Grainings

Formally, one coarse-grains the data $F \in \mathcal{P}(\Sigma)$ using a conditional probability distribution $p_{\Lambda|\Sigma}$. As we represent everything in terms of linear operators, the action of coarse-graining is a classical-to-classical channel [67]. Using this framework of coarse-graining, one can define a generalization of Eqn. 3.3:

$$\begin{aligned}\mathbf{S}_{\mu}^{\mathcal{Q}} \equiv &\{\sigma \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B) : \min_{F\in\mathcal{Q}} \|\mathcal{N}(\Phi_{\mathcal{P}(\Sigma)}(\sigma)) - \mathcal{N}(F)\|_1 \leq \mu\} \\ = &\{\sigma \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B) : \min_{F\in\mathcal{Q}} \|\Phi_{\mathcal{P}}(\sigma) - \mathcal{N}(F)\|_1 \leq \mu\} \end{aligned} \tag{3.4}$$

where we have let $\Phi_{\mathcal{P}(\Sigma)}$ be the probability map for the fine-grained data and $\Phi_{\mathcal{P}}$ be the effective probability map defined using the effective POVM $\{\widetilde{\Gamma}_i^C\}$ where $\widetilde{\Gamma}_i^C = \sum_{j\in\Sigma} p_{\Lambda|\Sigma}(i,j)\widetilde{\Gamma}_j$. We use Eqn. 3.4 because in coding the optimization problem it will simplify discussing the numerics.

Security of Multiple Coarse-Grainings

Theorem 5. *Let Ξ be a finite alphabet for indexing. Let $\varepsilon_{PE} > 0$. Let $\mathrm{D}((\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes m})$ be treated as the universal set. For all $k \in \Xi$, let $\mathbf{S}_k \subseteq \mathrm{D}((\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes m})$ such that for all $\sigma \notin \mathbf{S}_k$, σ is ε_{PE}-securely filtered. Define $\mathbf{S}_{multi} \equiv \bigcap_{k\in\Xi} \mathbf{S}_k$. Then $\forall \sigma \notin \mathbf{S}_{multi}$, σ is ε_{PE}-securely filtered.*

Proof. For any set $X \subseteq \mathrm{D}((\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes m})$, let $\overline{X} \equiv \{x \in \mathrm{D}((\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes m}) | x \notin X\}$. We know by the definition of the sets $\mathbf{S}_k$ that $\forall k \in \Xi$, $\forall \sigma \in \overline{\mathbf{S}_k}$, σ is $\varepsilon_{\mathrm{PE}}$-securely filtered. It immediately follows that $\forall \sigma \in \bigcup_k \overline{\mathbf{S}_k}$, σ is $\varepsilon_{\mathrm{PE}}$-securely filtered. Note that $\overline{\bigcap_k \mathbf{S}_k} = \bigcup_k \overline{\mathbf{S}_k}$. Therefore $\forall \sigma \notin \bigcap_k \mathbf{S}_k$, σ is $\varepsilon_{\mathrm{PE}}$-securely filtered. □

This tells us that given a set of sets where each set's complement is a subset of $\mathbf{S}_{\leq\varepsilon_{\mathrm{PE}}}$, the complement of the intersection of these sets is also a subset of $\mathbf{S}_{\leq\varepsilon_{\mathrm{PE}}}$. This implies optimizing over the intersection retains the same level of security. This proves we can consider the sets defined for different coarse-grainings.

While the previous theorem is sufficient, it would be good to have an explicit statement of its corollary in terms of i.i.d. states, which we now present.

Corollary 6. *Fix $\varepsilon_{PE} > 0$. Let Ξ be a finite alphabet indexing these multiple coarse-grainings. For each $k \in \Xi$, let*

$$\mathbf{S}_{\mu_k}^{\mathcal{Q}} = \{\sigma \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B) : \min_{F\in\mathcal{Q}} \|\Phi_{\mathcal{P}_k}(\sigma) - \mathcal{N}_k(F)\|_1 \leq \mu_k\}$$

where, $\Phi_{\mathcal{P}_k}(\rho) = \sum_{i\in\Lambda_k} \mathrm{Tr}\left(\rho\widetilde{\Gamma}_i^{C_k}\right)|i\rangle\langle i|$ *is the corresponding probability map,* $\mathcal{N}_k(X) = \sum_{i,j} p_{\Lambda_k|\Sigma}(j|i)\langle i|X|i\rangle|j\rangle\langle j|$ *is the corresponding coarse-graining channel, and* μ_k *is determined using Theorem 4 so that* $\forall\sigma \notin \mathbf{S}_{\mu_k}$, $\sigma^{\otimes m}$ *is* ε_{PE}*-securely filtered. Define* $\mathbf{S}_{multi} = \bigcap_k \mathbf{S}_{\mu_k}$. *If* $\sigma \notin \mathbf{S}_{multi}$, *then* $\sigma^{\otimes m}$ *is* ε_{PE}*-securely filtered.*

The proof is identical to Theorem 5 except having to move between statements about σ and $\sigma^{\otimes m}$.

Using Corollary 6, we can define the general set to optimize over under the assumption of i.i.d. collective attack:

$$\mathbf{S}^{\mathcal{Q}}_{\varepsilon_{\mathrm{PE}}} = \{\sigma \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B)\,|\,\forall k \in \Xi, \exists F \in \mathcal{Q} : \|\Phi_{\mathcal{P}_k}(\sigma) - \mathcal{N}_k(F_k)\|_1 \leq \mu_k\} \tag{3.5}$$

where Ξ is an alphabet for indexing the number of coarse-grainings. Thus, we can consider multiple coarse-grainings.

There are two important observations to be made. The first is that for $\rho \in \mathbf{S}^{\mathcal{Q}}_{\varepsilon_{\mathrm{PE}}}$ one does not need a single $F \in \mathcal{Q}$ which satisfies all k variation bounds with respect to ρ but rather $(F_1, ..., F_{|\Xi|}) \in \mathcal{Q}^{\times|\Xi|}$ so that F_k satisfies the k^{th} variation bound with respect to ρ. This is a property of the proof method we have used as we intersect the sets. An alternative proof method that only considers one F that satisfies all constraints at the same time remains an open problem. The second observation is that to define $\mathbf{S}^{\mathcal{Q}}_{\varepsilon_{\mathrm{PE}}}$, each $\mathbf{S}^{\mathcal{Q}}_{\mu_k}$ being intersected must be defined using a coarse-graining which acts on fine-grained statistics over the same alphabet Σ. Otherwise more testing would be necessary which would relate to a different set and a different security claim. To visualize Eqn. 3.5, see Figure 3.2, which presents Eqn. 3.5 for a protocol which accepts only a single distribution F.

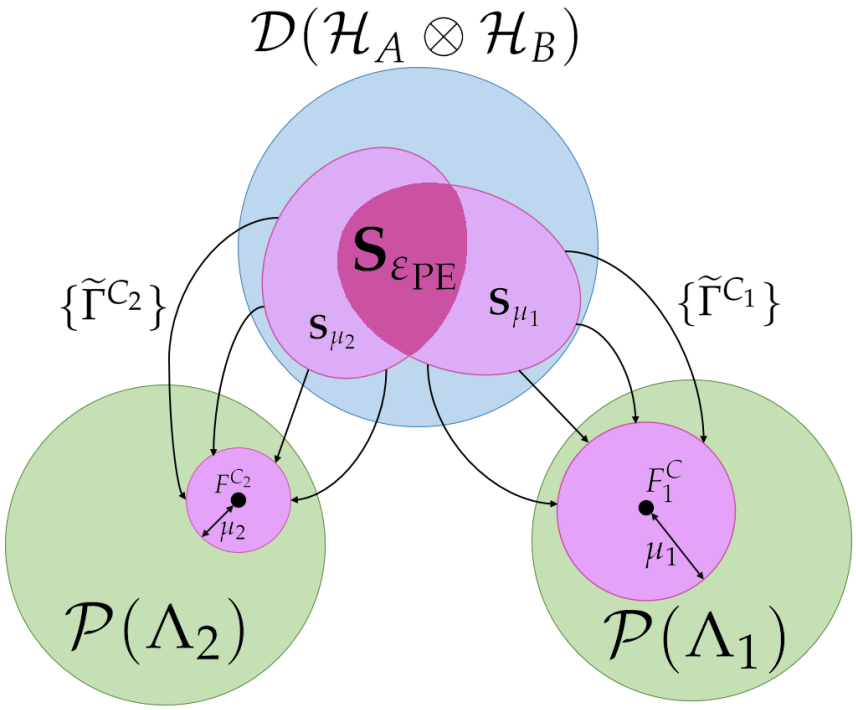


Figure 3.2: Here we see a visualization of using multiple coarse-grainings to improve the key rate by decreasing the number of states which we need to optimize over. Here we let $\mathcal{Q} = F$. We then consider two coarse-grainings which result in the frequency distributions F^{C_1} and F^{C_2}. Any probability distribution within μ_k of F^{C_k} must then be considered. One can then use the corresponding coarse-grained POVM, $\{\widetilde{\Gamma}_k^C\}$, to determine what set of states, $\mathbf{S}_{\mu_k}$, will map into the set of distributions around F_k^C. The intersection of $\mathbf{S}_{\mu_1}$ and $\mathbf{S}_{\mu_2}$, denoted $\mathbf{S}_{\varepsilon_{\text{PE}}}$, is the set of states that must be considered when proving security. Here we have drawn $\mathbf{S}_{\mu_1}$ and $\mathbf{S}_{\mu_2}$ such that neither is a subset of the other. This is the case where multiple coarse-grainings will be useful, though it is not necessarily always the case.

3.1.3 Completeness of QKD protocols

The preceding approach to constructing the set $\mathbf{S}_\mu^{\mathcal{Q}}$ relies heavily on the method of types (the theorems used in the proof of Theorem 4 are results from the method of types). In this section we show one could also use the method of types to calculate the completeness of device-dependent QKD protocols. We also show how this relates to the previous analysis

in [51].

Any (sub)protocol is defined as $\varepsilon^C_{\text{(sub)protocol}}$-complete if on the honest implementation of the (sub)protocol, the probability of aborting the (sub)protocol is at most $\varepsilon^C_{\text{(sub)protocol}}$. In [51], a (sub)protocol was defined as ε-robust for a given input $\overline{\sigma}$, if the probability of the subprotocol aborting on that input was at most ε. Therefore one can see a (sub)protocol being $\varepsilon^C_{\text{(sub)protocol}}$-complete as being $\varepsilon^C_{\text{(sub)protocol}}$-robust on the honest implementation input. In the case of device-dependent QKD, the completeness of the protcol, $\varepsilon^C_{\text{QKD}}$, is upper bounded by $\varepsilon^C_{\text{PE}} + \varepsilon^C_{\text{EC}} + \varepsilon^C_{\text{PA}}$. It is noted in [51] that in effect one may assume that the error correction and privacy amplification are chosen in applications such that, for any input on which the parameter estimation is $\varepsilon^{Rob}_{\text{PE}}$-robust, the error correction and privacy amplification subprotocols are 0-robust. It would follow $\varepsilon^C_{\text{QKD}} = \varepsilon^C_{\text{PE}}$. We would therefore like a method for determining the completeness of the parameter estimation subprotocol so that we might determine the completeness of the QKD protocol as a whole.

Old Approach

In [51], for fixed $\mathcal{Q}$ and $\varepsilon_{\text{PE}} > 0$, Renner claims to construct a set of ε_{PE}-robust i.i.d. states (Eqn. 6.2 of [51]):[5]

$$\mathbf{S}^{\text{rob}}_{\varepsilon_{\text{PE}}} = \{\sigma \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B) \;\mid\; \min_{F \notin \mathcal{Q}} \|\Phi_{\mathcal{P}}(\sigma) - \mathcal{N}(F)\|_1 > \mu\}$$

where μ is as defined in Theorem 4. Now one sees that if $\sigma \in \mathbf{S}^{\text{rob}}_{\varepsilon_{\text{PE}}}$, then, by definition of the set, $\forall F \notin \mathcal{Q}, \|\Phi_{\mathcal{P}}(\sigma) - \mathcal{N}(F)\|_1 > \mu$. By Theorem 4, it follows that when measuring $\sigma^{\otimes m}$, each $F \notin \mathcal{Q}$ will not arise except with probability ε_{PE}. This however *does not* prove that $\sigma \in \mathbf{S}^{\text{rob}}_{\varepsilon_{\text{PE}}}$ is ε_{PE}-robust. This is because while the probability of Alice and Bob obtaining a sequence with a frequency distribution $F \notin \mathcal{Q}$ is less than or equal to ε_{PE} for all $F \notin \mathcal{Q}$, the probability of Alice and Bob obtaining *any* frequency distribution $F \notin \mathcal{Q}$ would be $\sum_{F \notin \mathcal{Q}} \Pr[F|\sigma^{\otimes m}]$ where $\Pr[F|\sigma^{\otimes m}]$ is the probability of getting the frequency distribution F given testing $\sigma^{\otimes m}$. In other words, a state $\sigma^{\otimes m}$ is ε_{PE}-robust only if $\Phi_{\text{PE}}(\sigma^{\otimes m}) \geq 1 - \varepsilon_{PE}$, but $\mathbf{S}^{\text{rob}}_{\varepsilon_{\text{PE}}}$ cannot guarantee this.

Calculation of Robustness

Given this, we propose a different approach for calculating the robustness of input i.i.d. states. This in turn allows us to, in principle, determine the completeness of device-

[5] In Eqn. 6.2 of [51], Renner uses $\geq$, however given Theorem 2, it ought to be $>$ as is used in this definition of the set.

dependent QKD protocols as well their robustness on the worst case-scenario not ε_{PE}-filtered i.i.d. collective attack. For this we need the following facts from the method of types.

Summary of Method of Types (Section 11.1 of [60])

For every definition, assume one sampled from the probability distribution $P \in \mathcal{P}(\Sigma)$ m times in an i.i.d. fashion.

1. The frequency distribution $F_{\vec{z}} \in \mathcal{P}(\Sigma)$ constructed from a sequence $\vec{z} \in \Sigma^{\times m}$ (Eqn. 3.1) is known as a *type*.
2. The set of sequences which have the same type, $F_{\vec{z}}$, is called the *type class* and is denoted $T(F_{\vec{z}})$.
3. (Eqn. 11.17 of [60]) The number of sequences that give rise to the same type, i.e. the size of the type class, is $|T(F_{\vec{z}})| = \binom{m}{mf(1)\ mf(2)\ \ldots\ mf(|\Sigma|)}$.
4. (Theorem 11.1.2 of [60]) The probability of obtaining the sequence $\vec{z}$ is $2^{-m(D(F_{\vec{z}}||P)+H(F_{\vec{z}}))}$.

Proposition 7. *Given $\sigma \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, the parameter estimation protocol is ε_{PE}^{Rob}-robust on $\sigma^{\otimes m}$ where*

$$\begin{aligned}\varepsilon_{PE}^{Rob} &= 1 - \sum_{F \in \mathcal{Q}} |T(F)| 2^{-m(D(F||\Phi_{\mathcal{P}}(\sigma))+H(F))} \\ &= \sum_{F \notin \mathcal{Q}} |T(F)| 2^{-m(D(F||\Phi_{\mathcal{P}}(\sigma))+H(F))} \ .\end{aligned}$$

Proof. Follows immediately from the definition of robustness, and Items 3 and 4 of the preceding summary of method of types. □

Corollary 8. *The parameter estimation protocol is ε_{PE}^{C}-complete where*

$$\begin{aligned}\varepsilon_{PE}^{C} &= 1 - \sum_{F \in \mathcal{Q}} |T(F)| 2^{-m(D(F||\Phi_{\mathcal{P}}(\xi))+H(F))} \\ &= \sum_{F \notin \mathcal{Q}} |T(F)| 2^{-m(D(F||\Phi_{\mathcal{P}}(\xi))+H(F))}\end{aligned}$$

and $\xi \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is the output of the honest implementation (i.e. ξ is the state which is a result of the (noisy) channel when Eve does not attack).

In general, determing the robustness of the protocol on a given state will be a combinatorial nightmare. This is because, given $\mathcal{Q} \subset \mathcal{P}(\Sigma)$, one needs to efficiently iterate over the (finite) subset of the probability simplex $\mathcal{P}(\Sigma)$ that makes up the acceptance set $\mathcal{Q}$. However, in the case of $\mathcal{Q} \subset \mathcal{P}(\{0,1\})$, it is easy (albeit not efficient) to iterate over the types and so the robustness can be calculated.

However, in the case where such an approximation is untenable (as may be the case even for $\mathcal{Q} \subset \mathcal{P}(\{0,1\})$ as the number of tests grows, under the assumption that $\mathcal{Q}$ is constructed in a specific but reasonable manner (as will be argued in Section 4.2), we may use the same tools as Theorem 4 to determine completeness of the device-dependent QKD protocol:

Proposition 9. *Let $\rho \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be the prototype of the output i.i.d. state in the honest implementation.*[6] *Let $\{\tilde{\Gamma}_j\}_{j\in\Sigma}$ be Alice and Bob's joint POVM and $\mathcal{Q} = \{F : \|\Phi_{\mathcal{P}}(\rho) - F\|_1 \leq t\}$ where $t \in (0,1)$. Let Alice and Bob use m signals for parameter estimation. It follows by the same argument as in Theorem 4 that*

$$\varepsilon_{PE}^{C} \leq 2^{-m\left(\frac{t^2}{2\ln 2} - |\Sigma| \log_2(m+1)/m\right)}.$$

3.2 Announcements, General Sifting, and the Key Map

Announcements, general sifting, and the key map (Steps 4 and 5 of Fig. 2.1), sometimes known as blockwise processing, are rather straight forward but deserve of a bit of explanation.

Announcements & General Sifting

In principle, Alice and Bob may have more information about the run of the protocol than the outcome of their measurements. For example, in entanglement-based BB84, Alice and Bob may choose their measurement apparatuses to measure in the X-basis or Z-basis. Therefore they have information about the basis as well as the measurement of the outcome. Alice and Bob then choose to decide to tell each other whether they measured in the X or Z-basis because if they didn't use the same basis, the result won't be correlated. This means they have chosen to make their basis choice public information, but they have kept the measurement outcomes as private data. Using this public information they can throw out, or *sift*, the a priori poorly correlated data where they didn't measure in the same

[6] We use the output state rather than the state prior to transmission as the honest implementation may include noise due to the channel.

basis. This general idea can be abstracted to the idea that Alice and Bob partition the information from running the transmission and measurement steps of the QKD protocol into public and private data and then announcing the public data to throw out portions of the private data they don't want. As it is a generalization of the sifting step of BB84, we refer to this as announcements and general sifting.

Key Map

At this point in the protocol, Alice and Bob still both have some subset of their private data, $\{0, ..., k_A-1\}^n$, $\{0, ..., k_B-1\}^n$. They would like to convert that data into a key. To do this, one of the parties, let's say Alice, has a function $f^A_{KM} : \{0, ..., k_A-1\}^n \to \{0, ..., d-1\}^{n'}$ which maps her remaining private data to the raw key $x \in \{0, ..., d-1\}^{n'}$. In general, f^A_{KM} does not have to be deterministic. Furthermore, $n' = n/b$ if it maps blocks of private data. For this thesis it will be sufficient to assume $n' = n$. An identical description can be made in the case Bob performs the key map. In that case it is referred to as 'reverse reconciliation' whereas when Alice performs the key map it is referred to as 'direct reconciliation.'

3.3 Error Correction

Once Alice has the raw key, x, from performing the key map, the goal is for Bob to acquire the same key.[7] This is known as error correction (Step 6 of Fig. 2.1). In general, for Bob to successfully guess Alice's raw key using his data, Alice is going to have to provide some information about her raw key to Bob. Unfortunately, any information she gives Bob, she is also giving Eve. Therefore the goal is for Alice to give the minimum amount of information about her raw key to Bob necessary for him to guess x' such that $x' = x$.

One can view this problem from a slightly different angle. Bob has fine-grained data $\tilde{y} \in \mathcal{Y} \equiv \{0, k_B - 1\}^n$ which is, to some degree, correlated with Alice's raw key, $x \in \mathcal{X} \equiv \{0, ..., d-1\}^n$. This correlation means that Bob has some guess already as to what Alice's key is, i.e. Bob has some side information Y. One can therefore imagine Alice needs to encode (i.e. compress) her information about the raw key such that it is enough information for Bob to guess the raw key (i.e. decode the encoded message) and such that it also encodes no more information than that so as to avoid leaking more information to Eve than necessary. Therefore the problem of error correction in QKD can be seen as the problem of encoding (by Alice) with decoding side information (held by Bob), which is a special case of distributed source coding for correlated sources. The Slepian-Wolf coding

[7] Remember, Bob could also do the key map, in which case one can just flip the names of the parties in this section.

theorem (Section 15.4 of [60]) tells us that fundamentally this task can be done at a rate of $H(X|Y)$ where X and Y are random i.i.d. random variables.[8] However, this rate only holds for asymptotic behaviour and our interest is in bounding the information needed for error correction in the finite case.

In the finite case, one can simply imagine that for each $x \in \mathcal{X}$ and $\tilde{y} \in \mathcal{Y}$ there is a transcript $c \in \mathcal{C}$ for the error correction protocol [51]. It follows that one can think of a probability distribution over the classical communication conditioned on Alice's raw key and Bob's fine-grained data: $P_{C|x\in\mathcal{X},\tilde{y}\in\mathcal{Y}}$. Then the amount of information leaked to Eve fundamentally is [51]:

$$\text{leak}_F \equiv \log|\mathcal{C}| - \inf_{x,\tilde{y}} H_{\min}(P_{C|x\in\mathcal{X},\tilde{y}\in\mathcal{Y}}) \ . \tag{3.6}$$

However, modeling the information leaked to Eve during error correction using this definition is not feasible. Therefore one needs a simpler method. There are multiple ways, but we will consider two common choices.

1. (*Corollary 6.3.5 of [51]*) Given Alice's raw key $x \in \{0, ..., d-1\}^n$ and Bob's fine-grained data y can be treated as being obtained from sampling from a joint probability distribution P_{XY} in an i.i.d. fashion, there exists a method of error correction using two-universal hash functions with failure probability ε_{EC} such that

$$\text{leak}_{\varepsilon_{\text{EC}}} \leq nH(X|Y) + \sqrt{n}\sqrt{3\log(2/\varepsilon_{\text{EC}})}\log(d+3) \tag{3.7}$$

2. Given an error correction scheme applied to fixed blocks of $x \in \{0, ..., d-1\}^n$, the error correction scheme will not achieve the efficiency given by the Slepian-Wolf coding rate. This inefficiency can be characterized by $f_{\text{EC}} > 1$ [31]. Then if f_{EC} is appropriately chosen,

$$\text{leak}_{\varepsilon_{\text{EC}}} \leq nf_{\text{EC}}H(X|Y) + \log(2/\varepsilon_{EC}) \tag{3.8}$$

   where the second term is due to the tolerated failure probability, ε_{EC}, of using a 2-universal hash function to verify the success of the error correction scheme.[9]

[8] It is not a problem that this is under the assumption of i.i.d. sampling as asymptotically QKD can always assume i.i.d. behaviour [51].

[9] Given a family of functions from set A to set B, $\mathcal{F}$, and a probability distribution over $\mathcal{F}$, P_F, $(P_F, \mathcal{F})$ is two-universal if $\Pr_{f\in\mathcal{F}}[f(x) = f(x') \wedge x \neq x'] \leq \frac{1}{|B|}$.

As one can see, both methods of bounding the information leaked to Eve in performing error correction are methods of correcting for the difference between the encoding that satisfies the Slepian-Wolf coding theorem. Furthermore, assuming a good choice of error correction code, both will achieve the Slepian-Wolf coding rate in the limit of large block length. Lastly, we note that while Eqn. 3.7 above provides better bounds, one would expect more realistic bounds using Eqn. 3.8, and so this is what we will use throughout this thesis.

3.3.1 Error Detection

In Eqn. 3.8 we have a term that is a function of the error correction failure probability, ε_{EC}. This is not due to the error correction scheme itself, though it may fail, but rather it is the error probability of *detecting* that the error correction scheme has failed. In terms of procedure, the basic idea is rather than making the error correction work except with small probability, one can just use an easier to implement error correction scheme, and once Bob has his guess, x', Bob can use a two-universal hash function, g, calculate $g(x')$ and then send $g(x')$ and g to Alice. Then, by the choice of the two-universal hash function, the probability Alice finds $g(x) = g(x')$ and $x \neq x'$ is less than or equal to the tolerated failure probability ε_{EC} and the amount of information leaked by $g(x)$ and g is $\log(2/\varepsilon_{\text{EC}})$.[10]

3.4 Privacy Amplification

Once Alice and Bob have the same raw key except with some small probability, all that is left is for them to make sure it is secure from Eve. Recall from Definition 2 that a protocol is $\varepsilon' + \varepsilon''$-secure if the QKD protocol outputs an ε'-secret and ε''-correct key. For the key to be ε'-secret, it is sufficient to demand (Eqn. 2.9)

$$\frac{1}{2}\|\rho_{S_A E'} - \tau_{S_A} \otimes \rho_{E'}\|_1 \leq \varepsilon$$

where $\tau_{S_A} = \frac{1}{|\mathcal{S}_A|}\sum_{s\in\mathcal{S}_A} |s\rangle\langle s|$. It was shown by Renner [51] how this could be achieved using two-universal hash functions.

[10] The proof is effectively the same as the one used for Lemma 6.3.3. of [51].

Leftover Hashing Lemma (Corollary 5.6.1 of [51])

Let $\rho_{XE} \in \mathrm{D}(\mathcal{H}_X \otimes \mathcal{H}_E)$ be a classical-quantum state where $\mathcal{H}_X \cong \mathbb{C}^{\Sigma}$. Let $\mathcal{F}$ be a family of two-universal hash functions from $\Sigma \to \{0,1\}^{\ell}$ and $\varepsilon > 0$. Then

$$\frac{1}{2}\|\rho_{F(X)EF} - \frac{1}{|\{0,1\}^{\ell}|}\mathbb{1}_{F(X)} \otimes \rho_{EF}\|_1 \leq \varepsilon + 2^{-\frac{1}{2}\left(H_{\min}^{\varepsilon}(X|E)-\ell-2\right)} \tag{3.9}$$

One can see from the leftover hashing lemma (Eqn. 3.9) that Alice and Bob can use a two-universal hash function to make their raw keys secret and, furthermore, it is a function of the smooth min-entropy of the entire raw key conditioned on Eve's state. This gives us the following corollary.

Corollary 10. *(Eqn. 11 of [54]) Given $\rho_{XE} \in \mathrm{D}(\mathcal{H}_X \otimes \mathcal{H}_E)$ be a classical-quantum state. Let $\varepsilon_{sec} \equiv \varepsilon' + 2^{-\frac{1}{2}\left(H_{\min}^{\varepsilon'}(X|E)-\ell-2\right)}$, $\varepsilon' > 0$. Let $\varepsilon_{PA} \equiv \varepsilon_{sec} - \varepsilon' > 0$. Then for a ε_{sec}-secret key, the length of the key, ℓ, must satisfy the following bound:*

$$\ell \leq H_{\min}^{\varepsilon'}(X|E) - 2\log(1/\varepsilon_{PA}) \tag{3.10}$$

Note that the operational interpretation of $\varepsilon_{\mathrm{PA}}$ as the probability of privacy amplification failing is straightforward from this corollary as the goal of privacy amplification is to generate a ($\varepsilon_{\mathrm{sec}}-$)secret key and $\varepsilon_{\mathrm{PA}}$ represents the distance from achieving this secrecy goal.

The statement of this corollary versus other texts may be helpful. In [51] and [53], by using careful choices for the relationships between the ε-terms, the $\varepsilon_{\mathrm{PA}}$-term is not written out itself. Here we have stated Corollary 10 in the manner of [54] where the smoothing term on the min-entropy is in a sense decoupled from the privacy amplification failure probability, $\varepsilon_{\mathrm{PE}}$.

3.5 Security Proof

With the different subprotocols of QKD and their failure probabilities accounted for, we can use them to derive the security statement against i.i.d. collective attacks. We will prove this for the slightly more general assumption that the output of the transmission step of the protocol is infinitely exchangeable which as a corollary is the security statement for i.i.d. collective attacks. The result and proof are a simplification of Theorem 6.5.1 of

[51]. It is also implicitly the security proof used in [53, 54]. Lastly, once we have shown security against i.i.d. collective attacks, we show how to extend these results to coherent attack security in manners achievable with our numerical method in Chapter 4.

Theorem 11. *(Security for Protocols with Infinitely Exchangeable Output of Transmission Step) Let $\varepsilon_{PE}, \bar{\varepsilon}, \varepsilon_{EC}, \varepsilon_{PA} > 0$. Assuming the output of the quantum phase of the QKD protocol is infinitely exchangeable, the QKD protocol is $\varepsilon = \varepsilon_{PE} + \bar{\varepsilon} + \varepsilon_{EC} + \varepsilon_{PA}$-secure given that, when the protocol does not abort, the output key is of length ℓ where*

$$\ell \leq n(H_\mu(X|E) - \delta(\bar{\varepsilon})) - leak_{\varepsilon_{EC}} - 2\log_2(1/\varepsilon_{PA}) \tag{3.11}$$

with the following definitions:

$$H_\mu(X|E) = \min_{\rho \in \mathbf{S}^{\mathcal{Q}}_{\bar{\varepsilon}_{PE}}} H(X|E)_{\mathcal{G}(\rho)} \tag{3.12}$$

$$leak_{\varepsilon_{EC}} = n f_{EC} H(X|Y) + \log_2(2/\varepsilon_{EC})$$

$$\delta(\bar{\varepsilon}) = 2\log_2(d+3)\sqrt{\frac{\log_2(2/\bar{\varepsilon})}{n}} \tag{3.13}$$

$$\mu = \sqrt{2}\sqrt{\frac{\ln(1/\varepsilon_{PE}) + |\Sigma|\ln(m+1)}{m}}\,, \tag{3.14}$$

$\mathbf{S}^{\mathcal{Q}}_{\bar{\varepsilon}_{PE}}$ is as defined in Eqn. 3.5, n is the number of signals used to generate key after parameter estimation and blockwise processing, and $\mathcal{G}$ is a CPTNI map which represents a round of the QKD protocol on (an arbitrary) purification of input ρ.

Proof. Denote the result of the output of the transmission step of the QKD protocol by $\rho_{A^N B^N} \in \mathrm{D}((\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes N})$. Let the purification of $\rho_{A^N B^N}$ be $\rho_{A^N B^N E^N} \in \mathrm{D}((\mathcal{H}_{ABE})^{\otimes N})$ where $\mathcal{H}_{ABE} \equiv \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E$ and $\mathcal{H}_E \cong \mathcal{H}_A \otimes \mathcal{H}_B$. Assume $\rho_{A^N B^N E^N}$ is an infinitely exchangeable operator. By the Quantum de Finetti Theorem (Eqn. 2.7), we know there exists a measure ν such that

$$\|\rho_{A^N B^N E^N} - \int_{\sigma \in \mathrm{D}(\mathcal{H}_{ABE})} \sigma^{\otimes N} \nu(\sigma)\|_1 = 0$$

We now apply the parameter estimation map, $\Phi^{\mathrm{PE}} \otimes \mathrm{id}_E : (\mathcal{H}_{ABE})^{\otimes N} \to \mathcal{H}_{ABE}^{\otimes(N-m)}$, to both states. As the the trace distance cannot increase under the action of a CPTNI map, the trace distance will stay 0. Thus,

$$\|\rho^{\mathrm{PE}} - \int_{\sigma \in \mathrm{D}(\mathcal{H}_{ABE})} \sigma^{\mathrm{PE}} \nu(\sigma)\|_1 = 0$$

where the notation is to make things more clean. We now define the set $\mathcal{V}^\mu = \{\overline{\sigma} \in \mathrm{D}(\mathcal{H}_{ABE}) : \mathrm{Tr}_E(\overline{\sigma}) \in \mathbf{S}^{\mathcal{Q}}_{\bar{\varepsilon}_{\mathrm{PE}}}\}$. Therefore:

$$\begin{aligned}
0 =& \|\rho^{\mathrm{PE}} - \int_{\sigma\in\mathrm{D}(\mathcal{H}_{ABE})} \sigma^{\mathrm{PE}}\nu(\sigma)\|_1 \\
\leq& \|\rho^{\mathrm{PE}} - \int_{\sigma\in\mathcal{V}^\mu} \sigma^{PE}\nu(\sigma)\|_1 + \| \int_{\sigma\in\mathcal{V}^\mu} \sigma^{PE}\nu(\sigma) - \int_{\sigma\in\mathrm{D}(\mathcal{H}_{ABE})} \sigma^{\mathrm{PE}}\nu(\sigma)\|_1 \\
\leq& \|\rho^{\mathrm{PE}} - \int_{\sigma\in\mathcal{V}^\mu} \sigma^{PE}\nu(\sigma)\|_1 + \int_{\sigma\in\overline{\mathcal{V}^\mu}} \|\sigma^{PE}\|_1\nu(\sigma) \\
\leq& \|\rho^{\mathrm{PE}} - \int_{\sigma\in\mathcal{V}^\mu} \sigma^{PE}\nu(\sigma)\|_1 + \varepsilon_{\mathrm{PE}}
\end{aligned}$$

where the second inequality comes from the fact $\mathrm{D}(\mathcal{H}_{ABE}) = \mathcal{V}^\mu + \overline{\mathcal{V}^\mu}$ and the third comes from the fact ν is a measure and $\|\sigma^{\mathrm{PE}}\|_1 \leq \varepsilon_{\mathrm{PE}}$ for all $\sigma \in \overline{\mathcal{V}^\mu}$ by construction of $\mathcal{V}^\mu$.

As announcements, general sifting, and the key map can also be represented by (the n-fold application of) CPTNI maps and the trace norm is non-increasing under the action of CPTNI maps, we can conclude

$$\|\rho_{XYE'} - \int_{\sigma\in\mathcal{V}^\mu} \sigma_{XYE'} \quad \nu(\sigma)\|_1 \leq \varepsilon_{\mathrm{PE}} \tag{3.15}$$

where E' is Eve's register along with any classical communication from these steps, X is the register of Alice's raw key, and Y is the data available to Bob.

Using Corollary 10, where we fix $\varepsilon' = \bar{\varepsilon} + \varepsilon_{\mathrm{PE}}$ and replace the register E by $E'C$ where C is a register that contains the communication from error correction,

$$\ell \leq H^{\varepsilon'}_{\min}(X^n|E'C) - 2\log_2(1/\varepsilon_{\mathrm{PA}}) \ .$$

Note that n is the number of signals in the raw key. It excludes signals used for parameter estimation or that were consumed as part of the announcements, general sifting, and key map (i.e. blockwise processing).

Lemma 2 of [54] tells us we may remove the register C from the min-entropy term by subtracting the fundamental amount of information leaked during error correction, $\mathrm{leak}_{\mathrm{F}}$ defined in Eqn. 3.6,

$$\ell \leq H^{\varepsilon'}_{\min}(X^n|E') - \mathrm{leak}_{\mathrm{F}} - 2\log_2(1/\varepsilon_{\mathrm{PA}})$$

We then replace leak_{F} with the looser but manageable term, $\text{leak}_{\varepsilon_{\text{EC}}}$ from Eqn. 3.8:

$$\ell \leq H_{\min}^{\varepsilon'}(X^n|E') - \text{leak}_{\varepsilon_{\text{EC}}} - 2\log_2(1/\varepsilon_{\text{PA}}) \ .$$

We now need a method for showing we can convert this to a minimization problem. To do this, realize one could consider an auxiliary (possibly infinitely-large) classical register and define the classical-quantum (sub-normalized) state

$$\rho^{dF} \equiv \int_{\sigma \in \mathcal{V}^\mu} \nu(\sigma) \, |\sigma\rangle\langle\sigma|_{\mathcal{V}_\mu} \otimes \sigma_{X^n E'} \ ,$$

where it is like a classical-quantum state as $|\sigma\rangle\langle\sigma|_{\mathcal{V}_\mu}$ keeps track of which state in the set $\mathcal{V}_\mu$ is being considered, which is why we label that register with the subscript of the set. We can then construct the following chain of inequalities:

$$\begin{aligned}
& H_{\min}^{\varepsilon'}(X^n|E')_{\rho_{XE'}} \\
\geq & H_{\min}^{\bar{\varepsilon}}(X^n|E')_{\rho^{dF}} \\
\geq & H_{\min}^{\bar{\varepsilon}}(X^n|\mathcal{V}_\mu E')_{\rho^{dF}} \\
\geq & \min_{\sigma \in \mathcal{V}_\mu} H_{\min}^{\bar{\varepsilon}}(X^n|E')_{\sigma_{X^n E'}} \ ,
\end{aligned}$$

where the first inequality follows from the definition of smooth min-entropy and Eqn. 3.15, the second by the strong subadditivity of smooth min-entropy (Eqn. 3.19 of [51]), and the third by a property of how smooth min-entropy behaves when conditioned on a classical register (Eqn. 3.20 of [51]).

Then by using our variation of Corollary 3.3.7 of [51] (Eqn. 2.6) to replace the smooth min-entropy term with a von Neumann entropy term, we find:

$$\ell \leq n(\min_{\sigma \in \mathcal{V}_\mu} H(X|E)_{\sigma_{X^n E'}} - \delta(\bar{\varepsilon})) - \text{leak}_{\varepsilon_{\text{EC}}} - 2\log_2(1/\varepsilon_{\text{PA}}) \ .$$

Finally, we note that $\sigma_{ABE} \in \mathcal{V}_\mu$ if and only if $\sigma_{AB} \in \mathbf{S}_{\varepsilon_{\text{PE}}}^{\mathcal{Q}}$ and $\sigma_{X^n E'}$ is simply the output of a round of the QKD protocol on (an arbitrary) purification of σ_{AB}. Therefore, we may equivalently write

$$\ell \leq n(H_\mu(X|E) - \delta(\bar{\varepsilon})) - \text{leak}_{\varepsilon_{\text{EC}}} - 2\log_2(1/\varepsilon_{\text{PA}}) \ .$$

Furthermore we can verify the security parameter as $\varepsilon_{\text{PA}} \equiv \varepsilon_{\text{sec}} - \varepsilon' \Rightarrow \varepsilon_{\text{sec}} = \varepsilon_{\text{PE}} + \bar{\varepsilon} + \varepsilon_{\text{PA}}$ and the key is ε_{EC}-correct, so, by Definition 2 of QKD security, the protocol is $\varepsilon = \varepsilon_{\text{PE}} + \bar{\varepsilon} + \varepsilon_{\text{EC}} + \varepsilon_{\text{PA}}$-secure. □

First, we note that (our choice of) $\mathcal{G}$ which represents the action of a round of the QKD protocol on a state ρ_{AB} is presented in Section 4.1. We also note that under the assumption of collective attacks, one would not need to use the series of inequalities to obtain the minimization over the smooth min-entropy as the argument for the minimization would simply follow from the fact there would be multiple collective attacks Alice and Bob may not abort on. We now present the collective attack security as a corollary.

Corollary 12. *(Security for i.i.d. Collective Attacks) Let $\varepsilon_{PE}, \bar{\varepsilon}, \varepsilon_{EC}, \varepsilon_{PA} > 0$. Assuming i.i.d. collective attack, the QKD protocol is $\varepsilon = \varepsilon_{PE} + \bar{\varepsilon} + \varepsilon_{EC} + \varepsilon_{PA}$-secure given that, when the protocol does not abort, the output key is of length ℓ where*

$$\ell \leq n(H_\mu(X|E) - \delta(\bar{\varepsilon})) - leak_{\varepsilon_{EC}} - 2\log_2(1/\varepsilon_{PA}) \tag{3.16}$$

where the definitions are the same as from Theorem 11.

Proof. By definition, an i.i.d. collective attack results in the output of the transmission step of the QKD protocol being in i.i.d. form, $\rho_{ABE}^{\otimes N}$. All i.i.d. states are infinitely exchangeable. Thus Theorem 11 completes the proof. □

3.5.1 Coherent Attack Security Proofs

So far we have presented a way of calculating the security for i.i.d. collective attacks which imposes special limitations on the power of Eve. However, the goal of QKD ultimately is to provide information-theoretic security against coherent attacks so that Eve is only limited by the laws of quantum mechanics. While in Chapter 5 we present results using the key rate calculation for i.i.d. attacks, it is important to prove that in principle the numerical method we present in Chapter 4 can be used for coherent attack analysis. To do this we present two methods of 'lifting' the i.i.d. collective attacks, the Finite Quantum de Finetti method [51] and the post-selection technique [14], and the extra requirements to apply these methods.

Finite Quantum de Finetti Method

The Finite Quantum de Finetti Method is Theorem 11 where one replaces the Quantum de Finetti theorem in the proof with the Finite Quantum de Finetti theorem (Eqn. 2.8). This requires a few replacements and additional parameters. This was all done in [51] after which this chapter is based. For one to use the Finite Quantum de Finetti theorem, Alice

and Bob must randomly permute their signals after state transmission so as to force the state into the symmetric subspace, but in principle one can always do this. As this method was worked out in [51], we simply state a variation of the security result where instead of fixing a specific relation between the security parameters of the subprotocol for a clean result, we state the theorem so that it is straightforward to apply in generality.

Theorem 6.5.1 [51] Given a general QKD protocol as defined in Figure 2.1 where a total of N signals are transmitted, m of the signals are used for parameter estimation, and n of the signals are used for key generation, let $k \in \mathbb{N}$ and $bn + m + k = N$ where b accounts for block-wise processing. Let $\bar{\varepsilon}, \varepsilon_{\text{EC}}, \varepsilon_{\text{PA}}, \varepsilon_{\text{PE}}, \varepsilon_{\text{QdF}} > 0$. Then the QKD protocol is $(\varepsilon_{\text{QdF}} + \varepsilon_{\text{PE}} + \bar{\varepsilon} + \varepsilon_{\text{EC}} + \varepsilon_{\text{PA}})$-secure if the error correction is ε_{EC}-secure and if

$$\ell \leq n[H_\mu(X|E) - \delta(\bar{\varepsilon})] - 2(m+k)\log_2(\dim(\mathcal{H}_A \otimes \mathcal{H}_B)) - \text{leak}_{\varepsilon_{\text{EC}}} - 2\log_2(\frac{1}{\varepsilon_{\text{PA}}}) \quad (3.17)$$

where

$$\mu \equiv 2\sqrt{h\left(\frac{r}{m}\right) + \frac{\log_2(1/\varepsilon_{\text{PE}}) + |\Sigma|\log_2(\frac{m}{2}+1)}{m}} \quad (3.18)$$

$$\delta(\bar{\varepsilon}) \equiv (\frac{5}{2}\log_2(d) + 4)\sqrt{h(r/n) + \frac{2}{n}\log_2(4/\bar{\varepsilon})} \quad (3.19)$$

$$r \equiv \left(\frac{bn+m}{k} + 1\right)\left(2\ln\left(\frac{2}{\varepsilon_{\text{QdF}}}\right) + \dim(\mathcal{H}_A \otimes \mathcal{H}_B)^2 \ln(k)\right) - 1 \leq N \quad (3.20)$$

Proof. See [51]. □

As one can see from the theorem statement, one will expect the key rate to be quite bad unless one can make $r \ll n$ and $r \ll m$ or $r = m$. As such, people have remained interested in ways of determining ε-security under coherent attacks. One such method is the following technique which gets tighter bounds.

Post-Selection Technique

The post-selection technique [14] stems from a similar intuition to that of a de Finetti theorem. Physically, the use of the Finite Quantum de Finetti theorem in the security proof tells us that if the total signal state of the QKD protocol, composed of many subsystems, is properly symmetric (i.e. invariant under permutation of the subsystems), one can decompose the state into a mixture of almost i.i.d. states. This allows one to reduce

the security to something like that of the protocol's behaviour on (a convex combination of) i.i.d. states. Rather than focusing on the signal state, the post-selection technique tells us if the quantum channel which describes the protocol is permutation invariant, then the behaviour of the protocol on general inputs can be characterized as being polynomial in the security of the protocol on (convex combinations of) i.i.d. states.

Post-Selection Technique for Coherent Attack Analysis of QKD [14]

Let $\mathcal{E}^{\mathrm{QKD}} : \mathrm{D}(\mathcal{H}_{AB}^{\otimes N}) \to \mathrm{D}(\mathcal{H}_{S_A} \otimes \mathcal{H}_{S_B})$ be the CPTP map that models the physical implementation of the QKD protocol. Let $\mathcal{F}^{\mathrm{QKD}}$ be the ideal operation of the QKD protocol. Define $\Delta \equiv \mathcal{E}^{\mathrm{QKD}} - \mathcal{F}^{\mathrm{QKD}}$. Let Δ be such that for any permutation $\pi \in \mathcal{S}_N$ of the subsystems there exists a CPTP map K_π such that $\Delta \circ \pi = K_\pi \circ \Delta$. Let $\mathcal{E}^{\mathrm{QKD}}$ to be ε'-secure on i.i.d. collective attacks. Then

$$\|\Delta\|_\diamond \leq \varepsilon'(N+1)^{d_{AB}^2-1} \equiv \varepsilon \ . \tag{3.21}$$

Given the relationship between the diamond norm definition of QKD security and the trace norm definition of QKD security (Proposition 2), this implies $\mathcal{E}^{\mathrm{QKD}}$ is ε-secure on any input.

Note: Here we stated the post-selection technique (Theorem 1 of [14]) just in terms of QKD, which is observed as a corollary (Eqn. 4 of [14]).

It follows from the post-selection technique that if the theorem is satisfied, one simply has to perform more privacy amplification than under the i.i.d. collective attack assumption to acquire the security parameter they want for general attacks.

As one will note the theorem implicitly requires that $\mathcal{E}^{\mathrm{QKD}}$ is a permutation-invariant map. This is not a trivial requirement and introduces further ε-terms to approximate the protocol by a permutation-invariant one. It was shown in Section 3.4.3 of [4] how this could be done. Our interest for this thesis is not to focus on these coherent attack proof techniques, but simply to justify that they can be rigorously applied to lift the calculation in Corollary 12, which guarantees our numerical method has not lost this important generality. We do not focus on the results of these proof methods as we expect in the future there will be better proof methods for coherent attack analysis which will achieve much closer key rates to those found using the i.i.d. collective attacks.

3.5.2 Adaptive Security

It is worthwhile to now discuss what these security proofs have promised and what they have not. Specifically we must note that these security proofs are the security for *fixed length* protocols. This means that either the QKD protocol aborts or it will produce a key of some length ℓ. The reason for this is that the ε-security definition we used (Eqn. 2.10) demands that, when the protocol does not abort, the output is ε-close to uniform distribution over the key space. This is why in the security proof we must determine the worst case of all accepted observations to perform the privacy amplification.

This is in fact not what the normal theorist or experimentalist thinks ought to be done, and, understandably, they don't like this seemingly unnecessary cost to the key when the observations are better than the worst case accepted observations. As is perhaps natural, we all think we should run our protocol, see our observations, and then perform privacy amplification that, except with probability $\varepsilon_{\mathrm{PE}} + \bar{\varepsilon} + \varepsilon_{\mathrm{PA}}$, will be sufficient to obtain the secrecy we would like conditioned on what we saw. It follows that such a procedure would lead to adapting the length of the output key given our observations. This in turn requires a different ε-security definition. The ε-security of an adaptive QKD protocol was first given in [6].

Composable ε-Secure QKD Protocol with Adaptive Key Length [6]

A QKD protocol with adaptive key length is ε-secure if for all input states, the output state satisfies

$$\sum_{m=0} \Pr(m) \frac{1}{2} \| \hat{\rho}^m_{S_A S_B E'} - \tau^m_{S_A S_B} \otimes \rho_{E'} \|_1 \leq \varepsilon \tag{3.22}$$

where m denotes a specific length of the output key and $\tau^m_{S_A S_B} = \frac{1}{|\mathcal{S}^m_A|} \sum_{s^m \in \mathcal{S}^m_A} |s^m\rangle\langle s^m| \otimes |s^m\rangle\langle s^m|$.

Our goal in noting this, unfortunately, is not to fix the break between fixed length and adaptive security, but to make it clear why it exists as the author of this thesis, almost everyone the author has ever spoken to about this, and most publications on finite key analysis have seemed explicitly or implicitly to think one could make the following connection between the adaptive key length and fixed key length security.

Naive Adaptive Key Protocol from Fixed Key Protocol

Strategy:

1. Take a given device-dependent QKD protocol.
2. Run the QKD protocol up through obtaining the observed frequency distribution F in the parameter estimation step.
3. Determine the length ℓ of a ε-secure key for the QKD protocol if it were used in the fixed length manner and only accepted the frequency distribution F. (We will later refer to such a QKD protocol as a QKD protocol with unique acceptance. See Section 4.2.1.)
4. Proceed to perform error correction and privacy amplification to achieve that key length ℓ.

Belief: If you always follow these steps, by using the fixed length analysis, one is running an adaptive key length protocol with the same security parameter.

When written out in this manner, perhaps it is obvious this is not likely to be true. Specifically, recall that for the fixed key length the ε-security depends on how often the QKD protocol aborts. Imagine originally I consider an observed frequency distribution F such that $\ell \leq m'$. Writing the fixed length protocol security in terms of the adaptive key length security we have $\Pr(m') = (1 - p_{abort})$ and $\Pr(0) = p_{abort}$. However, I am now going to accept other observations than just F. For this new protocol, we may conclude $\Pr(m') \geq 1 - p_{abort}$. However we don't know about the distinguishability of $\hat{\rho}^{m'}_{S_A S_B E'}$, although realistically it should only stay the same or increase as no new procedure for obfuscating information from Eve has been added. Furthermore, for $m \neq m'$, the output of the protocol won't in general be a 0-secure key either, so now that $\Pr(m : m \neq m') \neq 0$, one is opening one's self up to further security risks. It follows that if there is an input state to the protocol which roughly saturates the ε-security bound in the fixed length security, then using the strategy above to construct the adaptive key length protocol will only be ε'-secure on that input state where ε' is (currently) unknown but is likely to satisfy $\varepsilon' > \varepsilon$.

For this reason it is not obvious that using the strategy presented will 'lift' ε-secure fixed key length protocols to an adaptive key length protocol with no cost to the security. Therefore we maintain, for the time being, we must consider and implement fixed length QKD protocols if we want a method which holds with generality.

Chapter 4

Theory of Numerical Finite Key Analysis

Having explained the theory of finite key analysis, we now would like to be able to apply it to general protocols. This requires numerical methods as in general there aren't enough symmetries and other simplifications to determine the key length analytically. In this chapter we present the numerical method of calculating key rates for device-dependent QKD protocols. The material in this chapter is largely the same as the presentation in the paper resulting from this research, [26]. First we derive the asymptotic key rate from Theorem 6.5.1 of [51]. We then explain how previous works [17, 68] developed a reliable numerical method for determining the infinite key rate. The rest of the chapter is then spent showing how this method may be extended to calculate secure key lengths for collective attacks determined in Theorem 11 for finite key rates. In Section 4.2 we present an SDP which may be used in the numerical method. Because of the technicality of this Chapter, we present it in a simplified form of SDP which will make the subsequent proofs simpler. This allows people to read Sections 4.1 and 4.2 alone to see how the extension works. In Section 4.3, we prove for this simplified case that, just as in the asymptotic case, our numerical method always provides a lower bound on the key rate (reliability) and that, with an ideal computer/solver, we would achieve the true theoretical lower bound (tightness). Finally in Section 4.4 we show that none of the proofs would have gone awry for the most general formulation of the SDP. This gives us a practical numerical tool we can implement for general device-dependent QKD protocols that can be represented in finite dimensional Hilbert spaces.

Furthermore, while it turned out to not be necessary to prove directly, one can prove a certain property needed for the proof of tightness directly via the introduction of semi-

infinite programming to quantum information theory, which has not previously been done in this manner.[1] As this turned out to be not necessary, but is a tool the author hopes may benefit someone else at some point, it has been relegated to Appendix A.

4.1 Background: Asymptotic Numerical Framework

Consider the generic QKD protocol as was depicted in Section 2.2. Under the assumption Alice and Bob are able to send an infinite number of signals, the security proof greatly simplifies as we now only care about the key rate rather than the key length. Define the key rate R as a function of the number of signals used, N, as $R(N) \equiv \ell/N$. Then the key rate of a protocol is defined as $R_\infty = \lim_{N\to\infty} R(N)$. Using Theorem 6.5.1 of [51], we see that as N goes to infinity, one has an infinite number of test signals so that $\mu \to 0$, $n/N \to p_{\text{pass}}$, and all correction terms go to zero, leaving

$$R_\infty = p_{\text{pass}}[\min_{\rho\in\mathbf{S}} H(X|E) - f_{\text{EC}} H(X|Y)] \tag{4.1}$$

where

$$\mathbf{S} = \{\rho \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B) | \operatorname{Tr}(\rho\Gamma_i) = \gamma_i,\ \forall i \in \Lambda\} \tag{4.2}$$

and $\{\Gamma_i\}_{i\in\Lambda} \subset \mathrm{Herm}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is a set of observables. One can see that $H(X|E)$ is the privacy amplification term, $f_{\text{EC}}H(X|Y)$ is the error correction term,[2] and p_{pass} is the probability each signal can be used to generate key. One may then expect, as this is just the rate when Alice and Bob have infinite resources, that this should be the fundamental rate of generating secret keys. This can be verified as it can be viewed as a re-writing and rescaling of the Devetak-Winter bound [19] which is the fundamental bound of establishing secret key. Note that Eqn. 4.1 also tells us that in the asymptotic limit coherent attacks are in effect i.i.d. attacks and so asymptotically one expects information processes to behave in an i.i.d. manner which is an important insight of finite information processing.

Given that R_∞ gives the fundamental bound for QKD protocols, it followed that people wanted a means for calculating the key rate for arbitrary device-dependent QKD protocols

[1] There is one other work we know of which makes use of semi-infinite programming for quantum information theory [57], but it is more specific than the proofs we provide which are necessary for the framework to be applied for general problems. The author of this thesis owes thanks to Jamie Sikora, one of the authors of [57], for bringing semi-infinite programming to the attention of the author of this thesis.

[2] In principle, f_{EC} is also a function of N and by the Slepian-Wolf Coding theorem we know for a good choice of error correction code $f_{\text{EC}} \to 1$ as $N \to \infty$. However, we don't get rid of this term as we are in reality still modeling QKD protocols done with finite resources.

that always provides a lower bound (i.e. reliable) and in principle could produce the 'true theoretical bound' (i.e. tight). A method for doing this was worked out in [17, 68]. In it the idea is to minimize $H(X|E)$ over the set of matrices in Eqn. 4.2. To do this, the idea was to simulate a round of the QKD protocol in a coherent fashion so that $H(X|E)_\rho$ can be calculated using the quantum relative entropy [16]:

$$f(\rho) = D(\mathcal{G}(\rho)||\mathcal{Z}(\mathcal{G}(\rho))) \tag{4.3}$$

where $\mathcal{G}$ is the CPTP map which coherently represents the QKD protocol and $\mathcal{Z}$ is a completely dephasing map on the register representing the key map. This method of determining the conditional entropy between Eve and the key is known as the post-processing framework and we refer the reader to Appendix A of [39] for an in-depth account of it.

Using the post-processing framework, by the joint convexity of quantum relative entropy, the function $f(\rho)$ is a convex function in ρ and thus can be used as the objective function for a semidefinite program. Furthermore, the set in Eqn. 4.2 is clearly a convex set over the semidefinite cone. It follows that one can view $\underset{\rho\in\mathbf{S}}{H}(X|E)$ as a semidefinite program. Therefore the authors of [17, 68] define

$$\alpha \equiv \min_{\rho\in\mathbf{S}} f(\rho) \tag{4.4}$$

However, as they wanted a lower bound that also holds if the numerical optimization routines returns before reaching the true mathematical minimum, they needed to acquire the dual problem of the SDP so that they have a maximization problem. This would guarantee the computer always returns an answer approaching from below the true minimum of the conditional entropy so that they can always guarantee that the answer provides a reliable lower bound on the key rate. Unfortunately, the quantum relative entropy is a highly non-linear function and so determining the dual of this problem is difficult in general. For this reason, they linearized the function about a given density matrix. They were then able to acquire the dual of the *linearization* of the original problem SDP, $\max_{\vec{y}\in\mathbf{S}^*(\sigma)} \vec{\gamma}\cdot\vec{y}$ where

$$\mathbf{S}^*(\sigma) \equiv \{\vec{y}\in\mathbb{R}^{|\Lambda|} | \sum_i y_i\Gamma_i \leq \nabla f(\sigma)\} \tag{4.5}$$

where $\vec{\gamma}$ is just the vector of the set of expectation values $\{\gamma_i\}_{i\in\Lambda}$.

Then the lower bound for any optimal or suboptimal attack σ can be calculated as

$$\beta(\sigma) \equiv f(\sigma) - \mathrm{Tr}(\sigma\nabla f(\sigma)) + \max_{\vec{y}\in\mathbf{S}^*} \vec{\gamma}\cdot\vec{y} \tag{4.6}$$

because it can be shown that for all $\rho \in \mathbf{S}, \alpha \geq \beta(\rho)$ so long as $\nabla f(\rho)$ exists (Theorem 1 of [68]). Here we have defined the gradient of f at point ρ represented in the standard basis $\{|k\rangle\}$ as:[3]

$$\nabla f(\rho) \equiv \sum_{j,k} d_{jk} |k\rangle \langle j|, \text{ with } d_{jk} \equiv \left. \frac{\partial f(\sigma)}{\partial \sigma_{jk}} \right|_{\sigma=\rho}$$

and $\sigma_{jk} \equiv \langle j| \sigma |k\rangle$. Moreover, we can write the gradient of $f(\rho)$ as:

$$\nabla f(\rho) \equiv \mathcal{G}^\dagger(\log_2 \mathcal{G}(\rho)) - \mathcal{G}^\dagger(\log_2 \mathcal{Z}(\mathcal{G}(\rho))) \tag{4.7}$$

Lastly, one can guarantee $\nabla f(\rho)$ exists via perturbing the state sufficiently by mixing the output of $\mathcal{G}(\rho)$ with the maximally mixed state such that all eigenvalues are non-zero.

The expression of $\beta(\sigma)$ in Eqn. 4.6 gives a valid lower bound for the key rate for any σ, but the bound will be tighter the closer σ is to the true optimum. We thus use a near-optimal evaluation of the primal problem (Eqn. 4.4) . This is referred to as Step 1 (see Algorithm 1). For further information on the specifics of this method, we refer to [68].

[3]Note that we have defined the derivative differently than in [68] by absorbing the occuring transposition into the definition of the gradient. This removes transpositions in many equations. Every statement is kept consistent with this definition throughout this thesis.

Algorithm 1: Asymptotic Key Rate lower bound

Result: lower bound on $\min_{\rho \in \mathbf{S}} H(X|E)$ [68]

1. Let $\epsilon > 0$, $\rho_0 \in \mathbf{S}$, maxIter $\in \mathbb{N}$, and $i = 0$.

Step 1

2. Compute $\Delta\rho := \arg\min_{\delta\rho} \mathrm{Tr}[(\delta\rho)\nabla f(\rho_i)]$ subject to $\Delta\rho + \rho_i \in \mathbf{S}$.
3. If $\mathrm{Tr}[(\Delta\rho)\nabla f(\rho_i)] < \epsilon$, then proceed to Step 2
4. Find $\lambda \in (0, 1)$ that minimizes $f(\rho_i + \lambda\Delta\rho)$
5. Set $\rho_{i+1} = \rho_i + \lambda\Delta\rho$, $i \to i + 1$.
6. If $i >$ maxIter, proceed to Step 2

Step 2

7. Let ρ be the result of Step 1. Let $\zeta \geq 0$ be the maximum constraint violation of ρ from the original set $\mathbf{S}$ constraints which satisfy this.
8. Calculate $\nabla f(\rho)$ to use for constructing $\mathbf{S}^*$
9. Expand $\mathbf{S}^*$ such that states which violated the original constraints by ζ are included.
10. Calculate β using the SDP defined above Equation 4.5

4.2 Extension to Finite Key Analysis

Having seen the numerical method for the infinite key analysis, the idea is to be able to use the method for finite key analysis. It is perhaps worth noting the primary issue in determining finite key rate is the optimization problem $H_\mu(X|E) \equiv \min_{\rho \in \mathbf{S}^{\mathcal{Q}}_{\varepsilon_{\mathrm{PE}}}} H(X|E)_\rho$ (Eqn. 3.12). Given the previous section, clearly the goal is to be able to get a good lower bound on the solution to this optimization problem in an efficient manner. To do this using the method of the previous section, we will need to be able to guarantee $\mathbf{S}^{\mathcal{Q}}_{\varepsilon_{\mathrm{PE}}}$ is a convex set. In general, this may not be the case as $\mathcal{Q} \subseteq \mathcal{P}(\Sigma)$ which allows for $\mathbf{S}^{\mathcal{Q}}_{\varepsilon_{\mathrm{PE}}}$ to not be convex.

The non-convexity of this set of density matrices $\mathbf{S}^{\mathcal{Q}}_{\varepsilon_{\text{PE}}}$ in general can be seen from the following simple example. Let Alice and Bob have a joint measurement for two qubits which can detect the entanglement of certain states. This is not an unreasonable demand as in effect this is what parameter estimation is trying to do. Let no coarse-graining be applied (i.e. $\mathcal{N}$ is the identity channel). Let $\mathcal{Q} = \{F_{\text{ent}}, F_{\text{mixed}}\}$ where F_{ent} is the probability distribution if Alice and Bob were to have measured the maximally entangled state and F_{mixed} is the probability distribution if they were to measure the maximally mixed state. By Born's rule and the linearity of quantum mechanics, it follows that one can get a linear combination of these two statistics, $F_\lambda = \lambda F_{\text{ent}} + (1-\lambda)F_{\text{mixed}}$ for any $\lambda \in (0,1)$ simply by measuring the corresponding mixture of the maximally entangled state and maximally mixed state. Then if $\mu < 0.5$, the equal mixture of the maximally mixed and maximally entangled state cannot be included in $\mathbf{S}^{\mathcal{Q}}_{\varepsilon_{\text{PE}}}$ yet the maximally mixed and maximally entangled state are included. Thus the set is not convex as a convex combination of two elements is not also included.

This example tells us we need to impose some further constraint on $\mathcal{Q}$ if we wish to use semidefinite programming. The obvious requirement is that we should force $\mathcal{Q}$ to be a convex set (this will be sufficient as all frequency distributions when written as matrices are a subset of the positive semidefinite cone). However it is important to note we are not losing much in this requirement. Generally the key rate is a function of the observed frequency distribution in a smooth manner. This tells us that over a closed convex set of frequency distributions the worst case will be an extreme point of the set of frequency distributions. For this reason, it should suffice to define $\mathcal{Q}$ as a closed convex set.[4] We therefore define the set of accepted frequency distributions from now on as:

$$\mathcal{Q} = \{F \in \mathcal{P}(\Sigma) : \|\overline{\mathcal{N}}(F) - \overline{\mathcal{N}}(\overline{F})\|_1 \leq t\} \tag{4.8}$$

where Σ is the alphabet of the fine-grained data of the protocol, $\overline{\mathcal{N}}$ is a coarse-graining so that one may abort on data-processed data, $\overline{F}$, and consequently $\overline{\mathcal{N}}(\overline{F})$, is a preferred frequency distribution, and t is the accepted variation threshold from $\overline{\mathcal{N}}(\overline{F})$.

Using this specific choice of form for the set of accepted frequency distributions, we can write the most general set to optimize over for i.i.d. collective attacks which allows for considering multiple coarse-grainings and considering constraints on Alice's marginal state due to source-replacement. In other words, we can rewrite Eqn. 3.5 so that it can

[4] We also note this decision is not specific to device-dependent QKD. In device-independent QKD the entropy term is a function of (generally) a CHSH inequality and also seems to behave in a smooth manner. Furthermore the set of accepted frequency distributions is demanded to be convex (see Section 9.2.3 of [2]).

be handled numerically for any device-dependent QKD protocol that can be represented in finite dimensional Hilbert spaces:

$$\begin{aligned}\mathbf{S}_{\varepsilon_{\text{PE}}} \equiv\{\sigma \in \text{Pos}(\mathcal{H}_A \otimes \mathcal{H}_B) \,:\, &\forall k \in \Xi, \exists F_k \in \mathcal{P}(\Sigma) : \\ &\|\Phi_{\mathcal{P}_k}(\sigma) - \mathcal{N}_k(F_k)\|_1 \leq \mu_k \;\;\&\;\; \|\overline{\mathcal{N}}(F_k) - \overline{\mathcal{N}}(\overline{F})\|_1 \leq t \\ &\&\;\; \text{Tr}(\sigma\Gamma_i) = \gamma_i \;\; \forall i \in \Lambda\}\end{aligned} \tag{4.9}$$

where $\{\Gamma_i\}_{i\in\Lambda} \subset \text{Herm}(\mathcal{H}_A \otimes \mathcal{H}_B)$ are a set of observables which pertain to properties of the state of which we are certain in the protocol. For all protocols one may assume $\Gamma_1 = \mathbb{1}_{\mathcal{H}_{AB}}$ which imposes σ is a quantum state. The set of observables also contains constraints to guarantee σ_A for source-replacement scheme versions of prepare and measure protocols (see Section 2.2.2). We again note that $\Phi_{\mathcal{P}_k}$ is the measurement channel under the action of the k^{th} coarse-graining which simplifies the analysis of the numerics.

While it may not be immediately obvious that Eqn. 4.9 can be written as a semidefinite program, note that it is just a set of trace norms and inner products (traces) of the state σ with observables. As in Section 2.3 we showed the trace norm is a semidefinite program and inner products are trivially Hermitian-preserving maps, the combination of these form an SDP.

For the rest of the chapter the consideration of multiple coarse-grainings will make the notation in the proofs more difficult to follow without adding any nuance. Therefore for the rest of the chapter we will provide proofs for the SDP which considers a *single* coarse-graining. Once these proofs are understood, it will be easy to see how they trivially extend for the multiple coarse-grainings case as is explained in Section 4.4. Therefore we now define the set to optimize over for a single coarse-graining:

$$\begin{aligned}\mathbf{S}_\mu \equiv \{\sigma \in &\text{Pos}(\mathcal{H}_A \otimes \mathcal{H}_B) \,:\, \exists F \in \mathcal{P}(\Sigma) : \;\; \|\Phi_{\mathcal{P}}(\sigma) - \mathcal{N}_k(F)\|_1 \leq \mu \\ &\&\;\; \|\overline{\mathcal{N}}(F) - \overline{\mathcal{N}}(\overline{F})\|_1 \leq t \;\&\; \text{Tr}(\sigma\Gamma_i) = \gamma_i \;\; \forall i \in \Lambda\}\end{aligned} \tag{4.10}$$

Using Eqn. 4.10, we can now define the linearized version of the semidefinite program:

$$\begin{aligned}\text{minimize} \quad & \langle \nabla f(\rho), \sigma \rangle \\ \text{subject to} \quad & \text{Tr}(\Gamma_i\sigma) = \gamma_i \qquad\qquad \forall i \in \Lambda \\ & \|\Phi_{\mathcal{P}}(\sigma) - \mathcal{N}(F)\|_1 \leq \mu \\ & \|\overline{\mathcal{N}}(F) - \overline{\mathcal{N}}(\overline{F})\|_1 \leq t \\ & \text{Tr}(F) = 1 \\ & \sigma, F \succeq 0 \;.\end{aligned} \tag{4.11}$$

When reformatted properly (see Section 4.3) the semidefinite program has the following dual problem:

$$\begin{aligned} \text{maximize} \quad & \vec{\gamma}\cdot\vec{y} + \overline{f}\cdot\vec{\overline{z}} - a\mu - \overline{a}t - b \\ \text{subject to} \quad & \sum_i y_i\Gamma_i + \sum_j z_j\widetilde{\Gamma}_j \leq \nabla f(\rho) \\ & \overrightarrow{\overline{N}^\dagger}(\vec{\overline{z}}) - \overrightarrow{N^\dagger}(\vec{z}) \leq b\vec{1} \\ & -a\vec{1} \leq \vec{z} \leq a\vec{1} \\ & -\overline{a}\vec{1} \leq \vec{\overline{z}} \leq \overline{a}\vec{1} \\ & a, \overline{a} \geq 0, \vec{y} \in \mathbb{R}^{|\Lambda|} \end{aligned} \tag{4.12}$$

where $\overline{f}$ is the vector version of $\overline{\mathcal{N}}(\overline{F})$ and $\overrightarrow{\mathcal{N}^\dagger}$ is the action as the adjoint of the map $\mathcal{N}$, $\mathcal{N}^\dagger$, on the diagonal entries of a matrix. It is sufficient to consider $\overrightarrow{\mathcal{N}^\dagger}$ on the diagonal entries of a matrix because $\mathcal{N}^\dagger$ only acts on the diagonal entries of a matrix, and so it is easy to see that the $\overrightarrow{\mathcal{N}^\dagger}$ map applied to the vector formed by the diagonal entries of a matrix gives the equivalent action as $N^\dagger$ on the matrix.

These semidefinite programs are important as the first is the optimization problem used in Step 1 of Algorithm 1 (i.e. one replaces $\mathbf{S}$ with $\mathbf{S}_\mu$) and the dual problem is what is used in Step 2 (by replacing $\mathbf{S}^*$ by the constraints in Eqn. 4.12). In effect this completes the explanation of the extension of the method to finite key analysis. The rest of the chapter are technical details. In the following subsection we consider a specific simplification of the above problem used in many previous works which we will use in Chapter 5. Then in Section 4.3 we prove that our method is reliable and tight (as well as deriving the dual problem stated in Eqn. 4.12). In Section 4.4 we show that nothing is lost when we consider multiple coarse-grainings.

4.2.1 SDP for Unique Acceptance

Historically, many works have been interested in the case where Alice and Bob accept only a single frequency distribution as it is easier to analyze and perhaps due to a misunderstanding of the relationship between adaptive and fixed key length protocols (See Section 3.5.2). We refer to protocols in which Alice and Bob accept a single frequency distribution as a QKD protocol with *unique acceptance*. As we will consider it in our numerics, we derive the SDP for a QKD protocol with unique acceptance from the general version above as it is a simpler SDP. Most generally, a protocol with unique acceptance may be viewed

as picking $\overline{\mathcal{N}}(\overline{F})$ to be the only distribution Alice and Bob accept on. Then the constraint pertaining to $\mathcal{Q}$ in Eqn. 4.10 vanishes as it must be the case $\overline{\mathcal{N}}(F) = \overline{\mathcal{N}}(\overline{F})$. It follows F could be allowed to vary over all F such that $\overline{\mathcal{N}}(F) = \overline{\mathcal{N}}(\overline{F})$. However, in previous works [10, 11, 53, 54], this nuance is lost as only one coarse-graining is considered, and so it is assumed $F = \overline{\mathcal{N}}(\overline{F})$. For consistency, we also make this assumption in defining a protocol with unique acceptance. We denote $\overline{\mathcal{N}}(\overline{F})$ by $\overline{F}_{\overline{\mathcal{N}}}$ to make it clear it is fixed rather than a variable. Using this notation, we can define the following set:

$$\begin{aligned}\mathbf{S}^{\mathrm{UA}}_{\varepsilon_{\mathrm{PE}}} \equiv \{\rho \in \mathrm{Pos}(\mathcal{H}_A \otimes \mathcal{H}_B) \,|\, \\ \|\Phi_{\mathcal{P}}(\rho) - \mathcal{N}(\overline{F}_{\overline{\mathcal{N}}})\|_1 \leq \mu,\ \mathrm{Tr}(\rho\Gamma_i) = \gamma_i,\ \forall i \in \Lambda\}\end{aligned}$$

where it must be the case that $\mathcal{N}$ coarse-grains data from the alphabet of $\overline{F}_{\overline{\mathcal{N}}}$ as otherwise it would not be well-defined. From this definition we get the following primal problem:

$$\begin{aligned}
\text{minimize} \quad & \langle \nabla f(\rho), \sigma \rangle \\
\text{subject to} \quad & \mathrm{Tr}(\Gamma_i \sigma) = \gamma_i \qquad\qquad \forall i \in \Lambda \\
& \mathrm{Tr}\big(\Delta^+\big) + \mathrm{Tr}\big(\Delta^-\big) \leq \mu \\
& \Delta^+ \succeq \Phi_{\mathcal{P}}(\sigma) - \mathcal{N}(\overline{F}_{\overline{\mathcal{N}}}) \\
& \Delta^- \succeq -(\Phi_{\mathcal{P}}(\sigma) - \mathcal{N}(\overline{F}_{\overline{\mathcal{N}}})) \\
& \sigma, \Delta^+, \Delta^- \succeq 0
\end{aligned} \tag{4.13}$$

The dual of this problem is:

$$\begin{aligned}
\text{maximize} \quad & \vec{\gamma} \cdot \vec{y} + \overline{f} \cdot \vec{z} - a\mu \\
\text{subject to} \quad & \sum_i y_i \Gamma_i + \sum_j z_j \widetilde{\Gamma}_j \leq \nabla f(\rho) \\
& -a\vec{1} \leq \vec{z} \leq a\vec{1} \\
& a \geq 0, \vec{y} \in \mathbb{R}^{|\Lambda|}
\end{aligned} \tag{4.14}$$

where $\overline{f}$ is the vector version of $\mathcal{N}(\overline{F}_{\overline{\mathcal{N}}})$. As this presents a simpler SDP, the numerical method has less free parameters and so, when applicable, this simpler SDP gives less chance to the computer to suffer from numerical imprecisions which can undermine Step 2 of Algorithm 1 as we will discuss next in the following section.

4.3 Tightness and Reliability of Finite Key Method

In this section we prove the tightness (Eqn. 4.28) and reliability (Eqn. 4.23) of our solver, which are stated together in Theorem 13. The tightness tells us that if numerical impre-

cisions were not to occur, the computer would find us the 'true' solution. The reliability property tells us that even when numerical imprecisions do occur, the computer gives us a lower bound on the conditional entropy, and so it is safe to use in determining the secure key length in actual experiment. As both of these properties are in terms of numerical imprecisions, we begin with an account of those. We then present the SDP when it considers numerical imprecisions. Finally, we prove tightness and reliability.

4.3.1 Numerical Imprecision

We recall two sets of constraints defined in the previous section: The set of constraints that are not subject to statistical fluctuation, denoted by $\{\Gamma_i\}_{i\in\Lambda}$, referred to as certainty constraints, and the constraints $\{\widetilde{\Gamma}_j\}_{j\in\Sigma}$ that are subject to statistical fluctuation, which are referred to as uncertainty constraints.

As noted in Sec. 4.1, when one acquires a solution ρ_f after the first step in Algorithm 1, the answer may not truly be feasible; that is, ρ_f is not in the correct set $\mathbf{S}_\mu$, but rather in an enlarged set $\widetilde{\mathbf{S}}_\mu$. This issue arises from the imprecise numerical representation of the POVMs as well as the imprecision of the numerical optimization solver which lead to violation of constraints in the optimization problem. To resolve this issue, one needs to consider the larger set $\widetilde{\mathbf{S}}_\mu$ to guarantee that ρ_f is included. Reference [68] presents a method for the asymptotic case. In Ref. [68], one has to consider only violations pertaining to certainty constraints $\{\Gamma_i\}$. In the finite key scenario, we also need to consider the uncertainty constraints $\{\widetilde{\Gamma}_j\}$. To rigorously account for numerical imprecision, we now adapt the method in [68] to finite key analysis.

An imprecise solver may lead to a solution ρ_f which is not positive semidefinite or that does not satisfy these constraints. To handle the first issue, if the state ρ_f has negative eigenvalues, one first perturbs the state to be $\rho_f' \equiv \rho_f + |\lambda_{\min}(\rho_f)|\mathbb{1}$ so that ρ_f' does not have negative eigenvalues. Then one checks the maximum violation of the certainty constraints of ρ_f', and define $\epsilon_{\text{sol}} \equiv \max_{i\in\Lambda} |\operatorname{Tr}(\rho_f'\Gamma_i) - \gamma_i|$.

Imprecise representations can be seen as deviations from the true POVM and probability representations. One can therefore denote the imprecise representations as follows:

$$\overline{\Gamma}_i = \Gamma_i + \delta\Gamma_i \text{ and } \overline{\gamma}_i = \gamma_i + \delta\gamma_i,$$

where $\|\delta\Gamma_i\|_{\text{HS}} \leq \epsilon_1$ and $|\delta\gamma_i| \leq \epsilon_2$ for all $i \in \Lambda$. By defining $\epsilon_{\text{rep}} \equiv \epsilon_1 + \epsilon_2$, it is shown in Lemma 10 of Ref. [68] that $|\operatorname{Tr}(\overline{\Gamma}_i) - \overline{\gamma}_i| \leq \epsilon_{\text{rep}}, \forall i \in \Lambda$. One then defines $\epsilon' = \max(\epsilon_{\text{sol}},\ \epsilon_{\text{rep}})$ and considers ρ subject to the constraints $\{|\operatorname{Tr}(\rho\overline{\Gamma}_i) - \overline{\gamma}_i| \leq \epsilon'\}$.

These imprecisions may also lead to violation of the variational distance constraint. Therefore, one should redefine μ for the second step to guarantee the ρ_f is considered in the second step. Since the uncertainty constraints pertain to the variational distance which takes the imprecisions as a whole, to properly enlarge μ to take constraint violations into account, one can use the Cauchy-Schwarz inequality along with Lemma 10 of [68] to expand μ as $\mu' = \max(\mu + n\epsilon', \|\Phi_{\mathcal{P}}(\rho_f) - F\|_1 + n\epsilon')$ where $n = |\Lambda|$.

Lastly, there is the possibility that the solver finds an optimal solution (σ, F) such that $\|\overline{\mathcal{N}}(F) - \overline{\mathcal{N}}(\overline{F})\|_1 > t$. In this case, one should expand t. Thus define $t' \equiv \max(t, \|\overline{\mathcal{N}}(F) - \overline{\mathcal{N}}(\overline{F})\|_1)$. Then one defines $\mathbf{S}_{\mu'\epsilon't'}$ to play the role of $\mathbf{S}_\mu$ by the following:

$$\begin{aligned}\mathbf{S}_{\mu'\epsilon't'} =& \{\rho \in \mathrm{Pos}(\mathcal{H}_A \otimes \mathcal{H}_B) \,|\, |\operatorname{Tr}\big(\overline{\Gamma}_i \rho\big) - \overline{\gamma}_i| \leq \epsilon' \;\forall i \in \Lambda, \\ & \|\Phi_{\mathcal{P}}(\rho) - \mathcal{N}(F)\|_1 \leq \mu', \|\overline{\mathcal{N}}(F) - \overline{\mathcal{N}}(\overline{F})\|_1 \leq t'\} \supseteq \mathbf{S}_\mu \ .\end{aligned} \tag{4.15}$$

Clearly, if $\epsilon' = 0$, $t' = t$, and $\mu' = \mu$, one reconstructs the original set $\mathbf{S}_\mu$. This alternative set is used for deriving the dual problem in the second step in the following section. By optimizing over this set $\mathbf{S}_{\mu'\epsilon't'}$, we handle the numerical imprecision related to certainty and uncertainty constraints.

It is important to note that when $\mathcal{G}(\rho)$ is singular, the derivative in Eqn. 4.7 may not exist. To tackle this issue, Ref. [68] introduces a small perturbation as

$$\begin{aligned}\mathcal{G}_\epsilon(\rho) &\equiv (1-\epsilon)\mathcal{G}(\rho) + \epsilon \mathbb{1}/d', \\ f_\epsilon(\rho) &\equiv D\big(\mathcal{G}_\epsilon(\rho) || \mathcal{Z}[\mathcal{G}_\epsilon(\rho)]\big),\end{aligned} \tag{4.16}$$

where d' is the dimension of $\mathcal{G}(\rho)$, and $\epsilon \geq 0$ is chosen in a way such that $\mathcal{G}_\epsilon(\rho)$ is not singular. The derivative of $f_\epsilon(\rho)$ is obtained by replacing $\mathcal{G}$ with $\mathcal{G}_\epsilon$ in Eqn. 4.7.

4.3.2 Finite Key SDP with Numerical Imprecisions

We present the SDP that also takes into account the numerical imprecision discussed above. (However, for ease of writing, we still use $\{\Gamma_i\}$ to denote certainty constraints and $\{\widetilde{\Gamma}_j\}$ to denote uncertainty constraints.) For simplicity, we present here derivations in the case of one variation bound and state the result related to multiple coarse-grainings in Sec. 4.4.

The primal problem of our SDP at $\rho \in \mathbf{S}_{\mu'\epsilon't'}$ is

$$\begin{aligned}
\text{minimize} \quad & \langle \nabla f_\epsilon(\rho), \sigma \rangle \\
\text{subject to} \quad & \mathrm{Tr}(G) + \mathrm{Tr}(H) \leq \mu' \\
& G \succeq \Phi_{\mathcal{P}}(\sigma) - \mathcal{N}(F) \\
& H \succeq \mathcal{N}(F) - \Phi_{\mathcal{P}}(\sigma) \\
& \mathrm{Tr}\big(\overline{G}\big) + \mathrm{Tr}\big(\overline{H}\big) \leq t' \\
& \overline{G} \succeq \overline{\mathcal{N}}(F) - \overline{F}_{\overline{\mathcal{N}}} \\
& \overline{H} \succeq \overline{F}_{\overline{\mathcal{N}}} - \overline{\mathcal{N}}(F) \\
& \mathrm{Tr}(F) = 1 \\
& |\mathrm{Tr}(\Gamma_i \sigma) - \gamma_i| \leq \epsilon' \ \forall i \in \Lambda \\
& \sigma, F, G, H, \overline{G}, \overline{H} \succeq 0.
\end{aligned} \tag{4.17}$$

where $\overline{F}_{\overline{\mathcal{N}}} \equiv \overline{\mathcal{N}}(\overline{F})$. We use this notation to emphasize $\overline{\mathcal{N}}(\overline{F})$ is fixed and is not an optimization variable because $\overline{F}$ and $\overline{\mathcal{N}}$ are both fixed. Let $\alpha_0(\rho)$ denote the optimal value of this primal problem. To derive its dual problem, Eqn. 4.17 can be reformatted to fit the definition of the primal problem of an SDP, Eqn. 2.11, as follows:

$$\begin{aligned}
A &= \mathrm{diag}(\nabla f_\epsilon(\rho), \overline{\mathbf{0}}) \\
B &= \mathrm{diag}(\mu', 0, 0, t', \overline{F}_{\overline{\mathcal{N}}}, -\overline{F}_{\overline{\mathcal{N}}}, 1, \sum_i (\epsilon + \gamma_i) |i\rangle\langle i|, \sum_i (\epsilon - \gamma_i) |i\rangle\langle i|) \\
\Psi(X) &= \mathrm{diag}(\mathrm{Tr}(G) + \mathrm{Tr}(H) + z, -G - \mathcal{N}(F) + \Phi_{\mathcal{P}}(\sigma) + I, \\
&\qquad - H + \mathcal{N}(F) - \Phi_{\mathcal{P}}(\sigma) + J, \mathrm{Tr}\big(\overline{G}\big) + \mathrm{Tr}\big(\overline{H}\big) + \overline{z}, \\
&\qquad - \overline{G} + \overline{\mathcal{N}}(F) + \overline{I}, -\overline{H} - \overline{\mathcal{N}}(F) + \overline{J}, \mathrm{Tr}(F), \\
&\qquad \Phi_0(\sigma) + M_1, -\Phi_0(\sigma) + M_2) \\
X &= \widetilde{\mathrm{diag}}(\sigma, F, G, H, z, I, J, \overline{G}, \overline{H}, \overline{z}, \overline{I}, \overline{J}, M_1, M_2)
\end{aligned} \tag{4.18}$$

where $\overline{\mathbf{0}}$ is a shorthand notation to mean that all other blocks are zero matrices of appropriate size, $\widetilde{\mathrm{diag}}$ means it is a matrix whose diagonal blocks I have named, $\Phi_0(X) \equiv \sum_{i\in\Lambda} \mathrm{Tr}(X\Gamma_i) |i\rangle\langle i|$, $\mathcal{N}(X) = \sum_{x,y} p(y|x) |y\rangle\langle x| X |x\rangle\langle y|$, $\overline{\mathcal{N}}(X) = \sum_{x,y} \overline{p}(y|x) |y\rangle\langle x| X |x\rangle\langle y|$, and $z, \overline{z} \in \mathbb{C}, I \in \mathrm{L}(\mathbb{C}^{|\Sigma|}), J \in \mathrm{L}(\mathbb{C}^{|\Sigma|}), \overline{I} \in \mathrm{L}(\mathbb{C}^{|\Sigma_C|}), \overline{J} \in \mathrm{L}(\mathbb{C}^{|\Sigma_C|}), M_1 \in \mathrm{L}(\mathbb{C}^{|\Lambda|})$ and $M_2 \in \mathrm{L}(\mathbb{C}^{|\Lambda|})$ are slack variables. Furthermore Σ_C represents the alphabet for the coarse-graining. It is easy to verify using the definition of adjoint map, $\langle Y, \Psi(X)\rangle = \langle \Psi^\dagger(Y), X\rangle$,

that the adjoint of Ψ is:

$$\begin{aligned}\Psi^\dagger(Y) = \operatorname{diag}(&\Phi_0^\dagger(W_1 - W_2) + \Phi_{\mathcal{P}}^\dagger(K - L), \mathcal{N}^\dagger(L - K) + \overline{\mathcal{N}}^\dagger(\overline{K} - \overline{L}) + b\mathbb{1}_{\mathcal{W}}, \\ &a\mathbb{1}_{\mathcal{W}} - K, a\mathbb{1}_{\mathcal{W}} - L, a, K, L, \overline{a}\mathbb{1}_{\mathcal{W}} - \overline{K}, \overline{a}\mathbb{1}_{\mathcal{W}} - \overline{L}, \overline{a}, \overline{K}, \overline{L}, W_1, W_2)\end{aligned} \tag{4.19}$$

where $Y = \widetilde{\operatorname{diag}}(a, K, L, \overline{a}, \overline{K}, \overline{L}, b, W_1, W_2)$,

$$\Phi_0^\dagger(W) = \sum_{i\in\Lambda} W(i,i)\Gamma_i, \quad \Phi_{\mathcal{P}}^\dagger(V) = \sum_{j\in\Sigma} V(j,j)\widetilde{\Gamma}_j \tag{4.20}$$

If we substitute these definitions in the standard form of SDP from Eqns. 2.11 and 2.12 and flip signs of $a, \overline{a}, b, K, L, \overline{K}$, and $\overline{L}$, we then get the following dual problem:

$$\begin{aligned}
\text{maximize} \quad & \langle \sum_{i\in\Lambda} (\epsilon' + \gamma_i) \left|i\right\rangle\!\left\langle i\right|, W_1 \rangle + \langle \sum_{i\in\Lambda} (\epsilon' - \gamma_i) \left|i\right\rangle\!\left\langle i\right|, W_2 \rangle \\
& \qquad + \langle \overline{F}_{\overline{\mathcal{N}}}, \overline{L} - \overline{K} \rangle - \mu' a - t' \overline{a} - b \\
\text{subject to} \quad & \sum_{i\in\Lambda} [W_1(i,i) - W_2(i,i)]\Gamma_i + \sum_{j\in\Sigma} [L(j,j) - K(j,j)]\widetilde{\Gamma}_j \preceq \nabla f_\epsilon(\rho) \\
& \overline{\mathcal{N}}^\dagger(\overline{L} - \overline{K}) - \mathcal{N}^\dagger(L - K) \preceq b\mathbb{1}_{\mathcal{W}} \\
& 0 \preceq K \preceq a\mathbb{1}_{\mathcal{W}} \qquad\qquad 0 \preceq \overline{K} \preceq \overline{a}\mathbb{1}_{\mathcal{W}} \\
& 0 \preceq L \preceq a\mathbb{1}_{\mathcal{W}} \qquad\qquad 0 \preceq \overline{L} \preceq \overline{a}\mathbb{1}_{\mathcal{W}} \\
& a, \overline{a} \geq 0, \qquad W_1, W_2 \preceq 0,
\end{aligned} \tag{4.21}$$

where $\mathcal{W} \equiv \mathbb{C}^{|\Sigma|}$. Let $\beta_0(\rho)$ denote the optimal value of this dual problem.

From Eqn. 4.21, we observe that off-diagonal entries of K, L, $\overline{K}$, $\overline{L}$, W_1 and W_2, are not important for this optimization problem since for any optimal solution $Y^* = \widetilde{\operatorname{diag}}(a^*, K^*, L^*, \overline{a}^*, \overline{K}^*, \overline{L}^*, b^*, W_1^*, W_2^*)$ of this problem, if K', L',$\overline{K}'$,$\overline{L}'$, W_1' and W_2' are matrices obtained by taking only the diagonal parts of $K^*, L^*, \overline{K}^*, \overline{L}^*, W_1^*$ and W_2^*, respectively, then the matrix $Y' = \operatorname{diag}(a^*, K', L', \overline{K}', \overline{L}', W_1', W_2')$ is also optimal as it is feasible and achieves the same optimal value. Moreover, we may optimize over the difference $L - K$ $(\overline{L} - \overline{K})$ subject to the constraint $-a\mathbb{1}_{\mathcal{W}} \preceq L - K \preceq a\mathbb{1}_{\mathcal{W}}$ $(-\overline{a}\mathbb{1}_{\mathcal{W}} \preceq \overline{L} - \overline{K} \preceq \overline{a}\mathbb{1}_{\mathcal{W}})$ as only the difference $L - K$ $(\overline{L} - \overline{K})$ matters in the optimization and its range is $-a\mathbb{1} \preceq L - K \preceq a\mathbb{1}$ $(-\overline{a}\mathbb{1} \preceq \overline{L} - \overline{K} \preceq \overline{a}\mathbb{1})$ which is determined by the two constraints $0 \preceq K \preceq a\mathbb{1}$ and $0 \preceq L \preceq a\mathbb{1}$ $(0 \preceq \overline{K} \preceq \overline{a}\mathbb{1}$ and $0 \preceq \overline{L} \preceq \overline{a}\mathbb{1})$. If we write $\vec{\gamma}$ as the vector whose i-th entry

is γ_i and $\overline{f} = \mathrm{diag}(\overline{F}_{\overline{\mathcal{N}}})$, the dual problem in Eqn. 4.21 is simplified as

$$\begin{aligned}
\text{maximize} \quad & (\epsilon' + \vec{\gamma}) \cdot \vec{y_1} + (\epsilon' - \vec{\gamma}) \cdot \vec{y_2} + \overline{f} \cdot \vec{\overline{z}} - \mu' a - t' \overline{a} - b \\
\text{subject to} \quad & \sum_{i \in \Lambda} [y_1(i) - y_2(i)] \Gamma_i + \sum_{j \in \Sigma} z(j) \widetilde{\Gamma}_j \preceq \nabla f_\epsilon(\rho) \\
& \overrightarrow{\overline{N}^\dagger}(\vec{\overline{z}}) - \overrightarrow{N^\dagger}(\vec{z}) \preceq b \vec{1} \\
& -a\vec{1} \leq \vec{z} \leq a\vec{1} \\
& -\overline{a}\vec{1} \leq \vec{\overline{z}} \leq \overline{a}\vec{1} \\
& a, \overline{a} \geq 0 \qquad \vec{y_1}, \vec{y_2} \leq \vec{0}.
\end{aligned} \tag{4.22}$$

where $\overrightarrow{\mathcal{N}^\dagger}$ is defined such that $\mathrm{diag}(\mathcal{N}^\dagger(Z)) = \overrightarrow{\mathcal{N}^\dagger}(\mathrm{diag}(Z))$ for arbitrary $Z \in \mathrm{L}(\mathbb{C}^{|\Sigma_C|})$. We remark that when $\epsilon' = 0$, we can replace $\vec{y_1}$ and $\vec{y_2}$ by $\vec{y} \equiv \vec{y_1} - \vec{y_2}$ subject to the constraint $\vec{y} \in \mathbb{R}^{|\Lambda|}$. When $\mu' = \mu$, $t' = t$, and $\epsilon' = 0$, Eqn. 4.22 reduces to the dual presented in Section 4.2, Eqn. 4.12 .

4.3.3 Reliability and Tightness

Now that we have the dual problem needed for Step 2 of Algorithm 1 which takes into account numerical imprecisions from Step 1, we prove that the lower bound obtained using this algorithm is reliable and tight. That is, in the limit where the numerical imprecisions go away, the program will obtain the true answer (tightness) and when the numerical imprecisions don't go away, we always have a lowerbound (reliability). In this section we present the precise mathematical statement of tightness for the SDP in Eqn. 4.17 in Theorem 13 which considers the issues of numerical imprecision discussed in Sec. 4.3.1. We then prove it. The extension to multiple coarse-grainings is then straightforward as we show in Section 4.4 This theorem is a finite-size version of Theorem 3 in Ref. [68]. In proving this theorem, we will adapt the proofs in Appendixes D and E of [68] as well as technical lemmas in Appendixes A-C of [68].

As our optimization problem comes from a physical scenario and we are only interested in the situation where the set $\mathbf{S}_{\mu'\epsilon't'}$ is not empty (otherwise we may trivially set the key rate to be zero), we restrict our attention to this situation.

Theorem 13. *(General Proof of Tightness of Numerical Method) Let $\mathbf{S}_{\mu'\epsilon't'}$ be defined in Eqn. (4.15) and assume $\mathbf{S}_{\mu'\epsilon't'} \neq \emptyset$. Let $\rho \in \mathbf{S}_{\mu'\epsilon't'}$ where $\mathcal{G}(\rho)$ is of size $d' \times d'$ and $\epsilon' > 0$. For $0 < \epsilon \leq 1/[e(d'-1)]$, then*

$$\alpha \geq \beta_{\mu'\epsilon't'\epsilon}(\rho) - \zeta_\epsilon \tag{4.23}$$

where

$$\alpha = \min_{\sigma \in \mathbf{S}_\mu} f(\sigma), \tag{4.24}$$

$$\beta_{\mu'\epsilon't't\epsilon}(\sigma) \equiv f_\epsilon(\sigma) - \mathrm{Tr}[\sigma \nabla f_\epsilon(\sigma)] + \max_{(a,\overline{a},\vec{y_1},\vec{y_2},\vec{z},\vec{\overline{z}},b) \in \mathbf{S}^*_{\mu'\epsilon't'}(\sigma)} [(\epsilon' + \vec{\gamma}) \cdot \vec{y_1} + (\epsilon' - \vec{\gamma}) \cdot \vec{y_2} + \overline{f} \cdot \vec{\overline{z}} - \mu' a - t' \overline{a} - b], \tag{4.25}$$

and

$$\zeta_\epsilon \equiv 2\epsilon(d' - 1) \log_2 \frac{d'}{\epsilon(d' - 1)}. \tag{4.26}$$

The set $\mathbf{S}^*_{\mu'\epsilon't'}(\sigma)$ *is defined by*

$$\begin{aligned} \mathbf{S}^*_{\mu'\epsilon't'}(\sigma) \equiv \{ & (a, \overline{a}, \vec{y_1}, \vec{y_2}, \vec{z}, \vec{\overline{z}}, b) \in (\mathbb{R}, \mathbb{R}, \mathbb{R}^{|\Lambda|}, \mathbb{R}^{|\Lambda|}, \mathbb{R}^{|\Sigma|}, \mathbb{R}^{|\Sigma|}, \mathbb{R}) | \\ & a, \overline{a} \geq 0, -a\vec{1} \leq \vec{z} \leq a\vec{1}, -\overline{a}\vec{1} \leq \vec{\overline{z}} \leq \overline{a}\vec{1}, \vec{y_1} \leq 0, \vec{y_2} \leq 0, \\ & \sum_{i \in \Lambda} [y_1(i) - y_2(i)] \Gamma_i + \sum_{j \in \Sigma} z(j) \widetilde{\Gamma}_j \preceq \nabla f_\epsilon(\sigma), \overrightarrow{N^\dagger}(\vec{\overline{z}}) - \overrightarrow{N^\dagger}(\vec{z}) \preceq b\vec{1} \} \end{aligned} \tag{4.27}$$

Moreover, if ρ^* *is an optimal solution to the primal problem,*

$$\lim_{\epsilon \to 0+} \lim_{\substack{\epsilon' \to 0+ \\ \mu' \to \mu \\ t' \to t}} [\beta_{\mu' t' \epsilon \epsilon'}(\rho^*) - \zeta_\epsilon] = \alpha. \tag{4.28}$$

To prove Theorem 13, we first show that for any $\rho \in \mathbf{S}_{\mu' t' \epsilon' \epsilon}$, the primal optimal value $\alpha_0(\rho)$ is equal to the dual optimal value $\beta_0(\rho)$ as Lemma 14. Then, we break down the proof of theorem into two parts: reliability in Eqn. (4.23) and tightness in Eqn. (4.28).

Lemma 14. *If* $\mathbf{S}_{\mu'\epsilon't'} \neq \emptyset$*, then* $\alpha_0(\rho) = \beta_0(\rho)$ *for any* $\rho \in \mathbf{S}_{\mu'\epsilon't'}$*.*

Proof. As $\mathbf{S}_{\mu'\epsilon't'} \neq \emptyset$, to apply Slater's condition, we just find a strictly feasible solution to the dual problem. We consider the dual problem in the form of Eqn. (4.21). Let $a = \overline{a} = 3$. Let $W_1 = \mathrm{diag}(x - 3, -1, -1, ..., -1)$ where $x = -|\lambda_{\min}(\nabla f_\epsilon(\rho))|$. Thus $W_1 \leq 0$. Let $W_2 = -\mathbb{1} \leq 0$. Without loss of generality, let $\Gamma_1 = \mathbb{1}$ as we always have the constraint $\mathrm{Tr}(\sigma) = 1$ in the primal problem. Let $L = 2\mathbb{1}$ and $K = \mathbb{1}$.Thus $-a\mathbb{1}_{\mathcal{W}} \prec L - K \prec a\mathbb{1}_{\mathcal{W}}$. Furthermore, $\sum_j [K(j,j) - L(j,j)] \widetilde{\Gamma}_j = \mathbb{1}$ as $\{\widetilde{\Gamma}_j\}$ is a POVM. Thus, $\sum_i [W_1(i,i) - W_2(i,i)] \Gamma_i + \sum_j [K(j,j) - L(j,j)] \widetilde{\Gamma}_j = (x-1)\mathbb{1} \prec \nabla f_\epsilon(\rho)$ by construction of x. Let $\overline{L} = 2\mathbb{1}$, $\overline{K} = \mathbb{1}$ and $b = 2$. Then $-\overline{a}\mathbb{1}_{\mathcal{W}} \prec \overline{L} - \overline{K} \prec \overline{a}\mathbb{1}_{\mathcal{W}}$ and $\overline{\mathcal{N}}^\dagger(\overline{L} - \overline{K}) - \mathcal{N}^\dagger(L - K) = 0 \prec b\mathbb{1}_{\mathcal{W}}$. The last equality followed from the fact $\mathcal{N}$ is a quantum channel and so its adjoint is unital. Thus all inequalities are strictly satisfied. □

We now adapt the proof in Appendix D.3 of [68] to finite-key scenario.

Lemma 15. *In the context of Theorem 13, $\alpha \geq \beta_{\mu'\epsilon't'\epsilon}(\rho) - \zeta_\epsilon$ for any $\rho \in \mathbf{S}_{\mu'\epsilon't'}$, which is Eqn. (4.23).*

Proof. Let $\alpha_{\mu'\epsilon't'\epsilon} \equiv \min_{\sigma \in \mathbf{S}_{\mu'\epsilon't'}} f_\epsilon(\sigma)$. Suppose that $\rho^*_{\mu'\epsilon't'\epsilon} \in \mathbf{S}_{\mu'\epsilon't'}$ is an optimal solution of this optimization. For any $\rho \in \mathbf{S}_{\mu'\epsilon't'}$, since f_ϵ is convex,

$$\begin{aligned} \alpha_{\mu'\epsilon't'\epsilon} = f_\epsilon(\rho^*_{\mu'\epsilon't'\epsilon}) &\geq f_\epsilon(\rho) + \langle (\rho^*_{\mu'\epsilon't'\epsilon} - \rho), \nabla f_\epsilon(\rho) \rangle \\ &\geq f_\epsilon(\rho) - \langle \rho, \nabla f_\epsilon(\rho) \rangle + \min_{\sigma \in \mathbf{S}_{\mu'\epsilon't'}} \langle \sigma, \nabla f_\epsilon(\rho) \rangle \\ &= f_\epsilon(\rho) - \langle \rho, \nabla f_\epsilon(\rho) \rangle + \alpha_0(\rho) \\ &= f_\epsilon(\rho) - \langle \rho, \nabla f_\epsilon(\rho) \rangle + \beta_0(\rho) = \beta_{\mu't'\epsilon'\epsilon}(\rho), \end{aligned} \tag{4.29}$$

where first two inequalities follow from the same argument about this linearization of our convex objective function as it is used in Eqns. (77)-(79) of Ref. [68] and the last line follows from Lemma 14 and the definition of $\beta_{\mu't'\epsilon'\epsilon}(\rho)$. Since $\mathbf{S}_\mu \subseteq \mathbf{S}_{\mu'\epsilon't'}$,

$$\alpha = \min_{\sigma \in \mathbf{S}_\mu} f(\sigma) \geq \min_{\sigma \in \mathbf{S}_{\mu'\epsilon't'}} f(\sigma) \geq \min_{\sigma \in \mathbf{S}_{\mu'\epsilon't'}} f_\epsilon(\sigma) - \zeta_\epsilon = \alpha_{\mu'\epsilon't'\epsilon} - \zeta_\epsilon, \tag{4.30}$$

where the last inequality follows from a continuity argument (which is Lemma 8 and Lemma 9 in Ref. [68]). Combining this result with Eqn. (4.29) leads to Eqn. (4.23). □

As we have shown the reliability of our numerical method, we now proceed with the tightness in Eqn. (4.28). If ρ^* is an optimal solution, an immediate consequence of Lemma 15 is that for any $\rho \in \mathbf{S}_{\mu'\epsilon't'}$, the following equation holds:

$$\min_{\sigma \in \mathbf{S}_{\mu'\epsilon't'}} \mathrm{Tr}[(\sigma - \rho^*)\nabla f(\rho^*)] \leq 0. \tag{4.31}$$

As Eqn. (4.31) holds for any feasible density operator in the set $\mathbf{S}_{\mu'\epsilon't'} \supseteq \mathbf{S}_\mu$, we want to show that if ρ^* optimizes the objective function f, then $\min_{\sigma \in \mathbf{S}_\mu} \mathrm{Tr}[(\sigma - \rho^*)\nabla f(\rho^*)] = 0$ where the optimization is over $\mathbf{S}_\mu$ as Eqn. 4.28 pertains to the limit where that is the set we are interested in. Therefore we just need to prove

$$\min_{\sigma \in \mathbf{S}_{\mu'\epsilon't'}} \mathrm{Tr}[(\sigma - \rho^*)\nabla f(\rho^*)] \geq 0 \tag{4.32}$$

when $\mathbf{S}_{\mu'\epsilon't'} \neq \emptyset$.

Lemma 16. *When* $\mathbf{S}_{\mu'\epsilon't'} \neq \emptyset$,

$$\min_{\sigma \in \mathbf{S}_{\mu'\epsilon't'}} \mathrm{Tr}[(\sigma - \rho^*)\nabla f(\rho^*)] \geq 0 \tag{4.33}$$

Proof. Let $\mathbf{S}_{\mu'\epsilon't'} \neq \emptyset$. By Lemma 14, we know that Eqn. 4.17 obtains its optimal value. Let ρ^* optimize f over $\mathbf{S}_{\mu'\epsilon't'} \neq \emptyset$. As f is a differentiable, convex function (one may consider f_ϵ to guarantee differentiability), it is the case that for all $\sigma \in \mathbf{S}_{\mu'\epsilon't'}$, $\mathrm{Tr}[\nabla f_{\epsilon'}(\rho^*)(\sigma - \rho^*)] \geq 0$ (Eqn. 4.21 of [9]). It follows $\min_{\sigma \in \mathbf{S}_{\mu'\epsilon't'}} \mathrm{Tr}[(\sigma - \rho^*)\nabla f(\rho^*)] \geq 0$ □

Eqn. (4.31) and Lemma 16 imply that, given ρ^* that optimizes f over $\mathbf{S}_{\mu'\epsilon't'}$,

$$\min_{\sigma \in \mathbf{S}_{\mu'\epsilon't'}} \mathrm{Tr}((\sigma - \rho^*)\nabla f(\rho^*)) = 0$$

We can therefore conclude the following:

$$\begin{aligned} f(\rho^*) &= f(\rho^*) + \min_{\sigma \in \mathbf{S}_{\mu'\epsilon't'}} \mathrm{Tr}[(\sigma - \rho^*)\nabla f(\rho^*)] \\ &= f(\rho^*) - \mathrm{Tr}(\rho^* \nabla f(\rho^*)) \\ &\qquad + \max_{(a,\overline{a},\vec{y_1},\vec{y_2},\vec{z},\vec{\overline{z}},b) \in \mathbf{S}^*_{\mu'\epsilon't'}(\rho^*)} [(\epsilon' + \vec{\gamma}) \cdot \vec{y_1} + (\epsilon' - \vec{\gamma}) \cdot \vec{y_2} + \overline{f} \cdot \vec{\overline{z}} - \mu' a - t' \overline{a} - b] \\ &= \beta(\rho^*) \end{aligned}$$

this completes the proof of Eqn. 4.28 and Theorem 13.

Note: While we used Eqn. 4.21 of [9] in proving Lemma 16, one can prove Lemma 16 directly using semi-infinite programming. While this isn't necessary as we can just use this cited result, we have presented this approach in Appendix A because it shows how one can introduce semi-infinite programming for quantum information theory and one application. If this interests the reader, we refer you to Appendix A.

4.4 Multiple Coarse-Graining SDP

In Chapter 3, we showed that we could consider multiple coarse-grainings in hopes of reducing the number of density matrices we need to consider in proving our security. In Eqn. 4.9 we presented this set for our numerics, but then in Section 4.3 we proved reliability and tightness for a single coarse-graining. We now show that it is easy to extend

to the case where one considers multiple coarse-grainings so that nothing is lost. First, we define Σ_f as the alphabet indexing the fine-grained statistics of the experiment. Let k index the set of conditional probability distributions pertaining to coarse-grained data, $\{p_{\Sigma_k|\Sigma_f}\}_k$. Each conditional probability distribution induces a channel $\mathcal{N}_k$ which applies the coarse-graining to the statistics. Define the POVM which pertains to the k^{th} conditional probability distribution as $\{\widetilde{\Gamma}_j^k\}_{j\in\Sigma_k}$ which induces a measurement channel $\Phi_{\mathcal{P}_k}$. In this case j is implicitly dependent on k as different coarse-grainings will construct probability distributions of different sizes. Then, the primal problem may be written as:

$$\begin{aligned}
\text{minimize} \quad & \langle \nabla f_\epsilon(\rho), \sigma \rangle & \\
\text{subject to} \quad & \operatorname{Tr}(\Gamma_i \sigma) = \gamma_i & \forall i \in \Lambda \\
& \|\Phi_{\mathcal{P}_k}(\sigma) - \mathcal{N}_k(F_k)\|_1 \leq \mu_k & \forall k \\
& \|\overline{\mathcal{N}}(F_k) - \overline{\mathcal{N}}(\overline{F})\|_1 \leq t & \forall k \\
& F_k \succeq 0 & \forall k \\
& \sigma \succeq 0 &
\end{aligned} \tag{4.34}$$

where $\Phi_{\mathcal{P}_k}(X) \equiv \sum_{j\in\Sigma_k} \operatorname{Tr}\left(X\widetilde{\Gamma}_j^k\right) |j\rangle\langle j|$, $\mathcal{N}_k(X) = \sum_{x\in\Sigma_f, y\in\Sigma_k} p_{\Sigma_k|\Sigma_f}(y|x) \langle x| X |x\rangle |y\rangle\langle y|$. We stress that F_k is indexed by k given the set considered in Corollary 6.

To convert this linearized primal problem into a semidefinite program, we effectively are just optimizing k copies of Eqn. 4.17 at the same time. This means we can write the equivalent form of Eqn. (4.17):

$$\begin{aligned}
\text{minimize} \quad & \langle \nabla f_\epsilon(\rho), \sigma \rangle \\
\text{subject to} \quad & \operatorname{Tr}(G_k) + \operatorname{Tr}(H_k) \leq \mu_k' \ \ \forall k \\
& G_k \succeq \Phi_{\mathcal{P}_k}(\sigma) - F_k \ \ \forall k \\
& H_k \succeq F_k - \Phi_{\mathcal{P}_k}(\sigma) \ \ \forall k \\
& \operatorname{Tr}\big(\overline{G}_k\big) + \operatorname{Tr}\big(\overline{H}_k\big) \leq t_k' \ \ \forall k \\
& \overline{G}_k \succeq \overline{\mathcal{N}}(F_k) - \overline{F}_{\overline{\mathcal{N}}} \ \ \forall k \\
& \overline{H}_k \succeq \overline{F}_{\overline{\mathcal{N}}} - \overline{\mathcal{N}}(F_k) \ \ \forall k \\
& |\operatorname{Tr}(\Gamma_i \sigma) - \gamma_i| \leq \epsilon' \\
& F_k, G_k, H_k, \overline{G}_k, \overline{H}_k \succeq 0 \ \ \forall k \\
& \sigma \succeq 0
\end{aligned} \tag{4.35}$$

where we have let t_k' be indexed by k in case different coarse-grainings violate the $\mathcal{Q}$ set by different amounts.

To reformat Eqn. (4.35) into the definition in Eqn. (2.11) we can extend the definitions in Eqn. (4.18) in a block diagonal fashion using the matrix direct sum, $\oplus$, over k.

$$
\begin{aligned}
A &= \operatorname{diag}(\nabla f_\epsilon(\rho), \overline{0}) \\
B &= \operatorname{diag}(\oplus_k \mu_k, \oplus_k 0, \oplus_k 0, \oplus_k t, \oplus_k \overline{F}_{\overline{\mathcal{N}}}, \oplus_k -\overline{F}_{\overline{\mathcal{N}}}, \oplus_k 1, \sum_i (\epsilon+\gamma_i)\left|i\right\rangle\left\langle i\right|, \sum_i (\epsilon-\gamma_i)\left|i\right\rangle\left\langle i\right|) \\
\Psi(X) &= \operatorname{diag}(\oplus_k[\operatorname{Tr}(G_k)+\operatorname{Tr}(H_k)+z_k], \oplus_k[-G_k-\mathcal{N}_k(F_k)+\Phi_{\mathcal{P}_k}(\sigma)+I_k], \\
&\qquad \oplus_k[-H_k+\mathcal{N}_k(F_k)-\Phi_{\mathcal{P}_k}(\sigma)+J_k], \oplus_k[\operatorname{Tr}(\overline{G}_k)+\operatorname{Tr}(\overline{H}_k)+\overline{z}_k], \\
&\qquad \oplus_k[-\overline{G}_k+\overline{\mathcal{N}}(F_k)+\overline{I}_k], \oplus_k[-\overline{H}_k-\overline{\mathcal{N}}(F_k)+\overline{J}_k], \oplus_k \operatorname{Tr}(F_k), \\
&\qquad\qquad \Phi_0(\sigma)+M_1, -\Phi_0(\sigma)+M_2) \\
X &= \widetilde{\operatorname{diag}}(\sigma, \oplus_k F_k, \oplus_k G_k, \oplus_k H_k, \oplus_k z_k, \oplus_k I_k, \oplus_k J_k, \oplus_k \overline{G}_k, \oplus_k \overline{H}_k, \oplus_k \overline{z}_k, \oplus_k \overline{I}_k, \oplus_k \overline{J}_k, M_1, M_2)
\end{aligned}
$$

It is straightforward to see the adjoint map of Ψ in this case is

$$
\begin{aligned}
\Psi^\dagger(Y) = \operatorname{diag}(&\Phi_0^\dagger(W_1-W_2)+\sum_k \Phi_{\mathcal{P}_k}^\dagger(K_k-L_k), \oplus_k[\mathcal{N}_k^\dagger(L_k-K_k)+\overline{\mathcal{N}}^\dagger(\overline{K}_k-\overline{L}_k)+b_k\mathbb{1}_{\mathcal{W}}], \\
&\oplus_k [a_k\mathbb{1}_{\mathcal{W}}-K_k], \oplus_k[a_k\mathbb{1}_{\mathcal{W}}-L_k], \oplus_k a_k, \oplus_k K_k, \oplus_k L_k, \oplus_k[\overline{a}_k\mathbb{1}_{\mathcal{W}}-\overline{K}_k], \oplus_k[\overline{a}_k\mathbb{1}_{\mathcal{W}}-\overline{L}_k], \\
&\oplus_k \overline{a}_k, \oplus_k \overline{K}_k, \oplus_k \overline{L}_k, W_1, W_2)
\end{aligned}
$$

where

$$
Y = \widetilde{\operatorname{diag}}(\oplus_k a_k, \oplus_k K_k, \oplus_k L_k, \oplus_k \overline{a}_k, \oplus_k \overline{K}_k, \oplus_k \overline{L}_k, \oplus_k \overline{b}_k, W_1, W_2)
$$

Finally, again because all of the ks are independent, this dual problem is ultimately simplified to:

$$
\begin{aligned}
\text{maximize} \quad & \sum_k \overline{f}\cdot\vec{\overline{z}}_k + (\epsilon'+\vec{\gamma})\cdot\vec{y}_1 + (\epsilon'-\vec{\gamma})\cdot\vec{y}_2 - \vec{\mu}\cdot\vec{a} - \vec{t}\cdot\vec{\overline{a}} - \|\vec{b}\|_1 \\
\text{subject to} \quad & \sum_i [y_1(i)-y_2(i)]\Gamma_i + \sum_k (\sum_j z_k(j)\widetilde{\Gamma}_j^k) \preceq \nabla f_\epsilon(\rho) \\
& \overrightarrow{\overline{\mathcal{N}}^\dagger}(\vec{\overline{z}}_k) - \overrightarrow{\mathcal{N}_k^\dagger}(\vec{z}) \leq b_k\vec{1}_{\mathcal{W}} \qquad \forall k \\
& -a_k\vec{1}_{\mathcal{W}} \leq \vec{z}_k \leq a_k\vec{1}_{\mathcal{W}} \qquad \forall k \\
& -\overline{a}_k\vec{1}_{\mathcal{W}} \leq \vec{\overline{z}}_k \leq \overline{a}_k\vec{1}_{\mathcal{W}} \qquad \forall k \\
& \vec{a}, \vec{\overline{a}} \geq 0 \qquad \vec{y}_1, \vec{y}_2 \leq 0
\end{aligned}
\tag{4.36}
$$

where $\vec{\mu'}$ is just the vector whose k-th entry is given by μ'_k. From these forms, it is clear that strong duality and tightness proofs follow from the single POVM case by indexing over the variable k and scaling things properly in the proof of strong duality. Therefore in this chapter we have shown that we have constructed a numerical method for determining reliable, tight key rates which can consider multiple coarse-grainings.

Chapter 5

Examples of Numerical Finite Key Analysis

So far we have improved the theory of finite key analysis and developed a numerical method for determining the key length for general device-dependent QKD protocols. We would now like to prove that our improved understanding and our numerical method are useful. To do this we consider variations of the Bennett-Brassard 1984 (BB84) protocol [7] under collective attack. Specifically, we consider the efficient single-photon BB84 protocol [41] under simple conditions in which an analytic security proof [53] is known as well as when the channel induces a reference frame misalignment. We also consider measurement-device-independent BB84 [42] and discrete-phase-randomized BB84 [12] which is an optical implementation of BB84. Lastly, as we present all of these results in the case of unique acceptance, we consider the reference frame misaligned BB84 and how it behaves when the protocol accepts a set of observations.

As all of these examples are based on BB84, we begin the chapter with a brief overview of efficient BB84 and its security proof to provide intuition for the discussions of the following protocols. We note these examples are the ones we presented in [26] and the presentation is largely taken from that work, but has been expanded. Lastly, as all of the numerics require the construction of Kraus operators for the $\mathcal{G}$ map in the post-processing framework (See Section 4.1), we have put this information in Appendix B for completeness. Furthermore, we note the code is not provided in this thesis as it would take up too much space. It is planned to be part of upcoming open-source software [40], at which point it will be available.

5.1 Background: BB84 and its Asymptotic Security

Protocol 2: Efficient BB84 [41]

N Number of signals Alice sends to Bob
p_z Probability of sending in the Z-basis
$\mathcal{Q}$ Set of frequency distributions Alice and Bob will accept

Protocol:

1. *Transmission*: For $i \in [N]$, Alice draws $b \in \{0,1\}$ according to $p(b=0) = p_z$. Alice draws $s \in \{0,1\}$ with uniform probability. She then sends the following state depending on her results:

(b,s)	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
State Sent	$\|0\rangle$	$\|1\rangle$	$\|+\rangle$	$\|-\rangle$

   where $|+\rangle \equiv \frac{1}{2}(|0\rangle + |1\rangle)$ and $|-\rangle \equiv \frac{1}{2}(|0\rangle - |1\rangle)$. Alice lets $\overline{A}_i = s$ and $\widetilde{A}_i = b$.

2. *Measurement*: For $i \in [N]$, Bob draws $r \in \{0,1\}$ according to $p(r=0) = p_z$. If $r=0$, Bob uses the POVM $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ If $r=1$, Bob uses the POVM $\{|+\rangle\langle +|, |-\rangle\langle -|\}$. Bob stores the outcome of his measurement q as 0 if the outcome is $|0\rangle$ or $|+\rangle$ and 1 otherwise. He stores $\overline{A}_i = q$ and $\widetilde{A}_i = r$.

3. *Parameter Estimation*: Alice picks a random subset of the N signals sent of size $m = (1-p_z)^2$. For each of the m signals, Alice and Bob announce (b,s) and (q,r) respectively to construct their frequency distribution F. If $F \in \mathcal{Q}$, they proceed. Otherwise, they abort.

4. *Announcements & General Sifting*: Alice and Bob announce $\widetilde{A}_i$ and $\widetilde{B}_i$ (i.e. r and b). If $m \neq b$, Alice and Bob throw out that signal.

5. *Key Map*: For any signal that has not been thrown out where $b=0$, Alice sets $x_i = s_i$. This results in $x \in \{0,1\}^n$ where $n \approx p_z^2(N-m)$.

6. *Error Correction & Privacy Amplification*: Performed as is standard.

Note: We skipped the Data Partitioning step as the protocol is simple enough to embed it into transmission and measurement steps.

BB84 is the first QKD protocol and will probably forever be the most well-behaved protocol. As it is in some sense the foundation of QKD and all of the examples in this thesis,

we present the efficient BB84 prepare-and-measure variation of the protocol (Protocol 2) including the steps that will be used in the finite case. We then present a summary of the asymptotic security proof using the source-replacement.

Protocl 2 is an 'efficient' BB84 protocol because one can see that as the total number of signals sent, N, goes to infinity, one can use an increasingly small fraction of states for parameter estimation, and send almost all states in the Z-basis, which results in effectively all signals sent contributing to the secret key. Furthermore, the observations one expects to see in parameter estimation asymptotically are straightforward:

$$p(i,j) = \begin{bmatrix} \frac{p_z^2}{2}(1-e_x) & \frac{p_z^2}{2}e_x & \frac{p_z(1-p_x)}{4} & \frac{p_z(1-p_z)}{4} \\ \frac{p_z^2}{2}e_x & \frac{p_z^2}{2}(1-e_x) & \frac{p_z(1-p_z)}{4} & \frac{p_z(1-p_z)}{4} \\ \frac{p_z(1-p_z)}{4} & \frac{p_z(1-p_z)}{4} & \frac{(1-p_z)^2}{2}(1-e_z) & \frac{(1-p_z)^2}{2}e_z \\ \frac{p_z(1-p_z)}{4} & \frac{p_z(1-p_z)}{4} & \frac{(1-p_z)^2}{2}e_z & \frac{(1-p_z)^2}{2}(1-e_z) \end{bmatrix} \tag{5.1}$$

where $(i,j) \in \{(0,0),(0,1),(1,0),(1,1)\}$ and e_x and e_z are known as the phase error and the bit error respectively.

With all of this in mind, all that is left is a brief discussion of the security. The security of this protocol is given in Appendix A of [52], and so we just note the points we will need in the rest of this chapter from this security proof. First, as is always the case in this thesis, one does the source-replacement. In the case of the BB84 protocol, this means that Alice will always sends the Bell state $|\Phi^+\rangle \equiv \frac{1}{2}(|00\rangle + |11\rangle)$. As we previously saw, since we are considering the asymptotic security, the attack may be assumed to be collective. It turns out that by the symmetries of the BB84 protocol, one may also the output of the collective attack will be diagonal in the Bell-basis:

$$\rho_{AB} = \lambda_1 \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right| + \lambda_2 \left|\Phi^-\right\rangle\!\left\langle\Phi^-\right| + \lambda_3 \left|\Psi^+\right\rangle\!\left\langle\Psi^+\right| + \lambda_4 \left|\Psi^-\right\rangle\!\left\langle\Psi^-\right| \tag{5.2}$$

where $|\Phi^-\rangle \equiv \frac{1}{2}(|00\rangle + |11\rangle)$ and $|\Psi^\pm\rangle \equiv \frac{1}{2}(|01\rangle \pm |10\rangle)$. Using these observations, [52] is able to show that, under the assumption, the asymptotic key rate for the efficient BB84 protocol is:

$$R^\infty_{BB84} = p_z^2[1 - h(e_x) - f_{EC}h(e_z)] \tag{5.3}$$

where $h(p) \equiv -p\log(p) - (1-p)\log(1-p)$ is the binary entropy function and the Shannon limit of error correction is $h(q)$ as can be verified given the conditional probability table (Eqn. 5.1). Furthermore, this would be expected as the number of bits that need to be corrected when doing the key map in the Z-basis is just asymptoitcally the expectation of the error rate in the Z-basis.

With these points established, we can continue to our numerical results.

5.2 Efficient BB84 with Phase Error Estimation

In [53], they use the asymptotic efficient BB84 key rate presented in the previous section to determine the key length, and thus rate, in the finite regime. They do this in the following manner. First, note that by the asymptotic key rate (Eqn. 5.3) we know only the fluctuations in the phase error will effect the privacy amplification term, $1 - h(e_x)$. It follows that if one accepts only one observed frequency distribution, the worst-case scenario privacy amplification term in the finite regime will be, $h(e_x + \frac{\mu}{2})$. We note that the factor of two is because the largest amount of the total variational bound, μ, that can be added to the outcome e_x is $\mu/2$ as one must then remove $\mu/2$ from the rest of the observations to preserve it being a probability distribution. This differs from [53] in which they add the entirety of μ to e_x. Given this analysis, by the key length under i.i.d. Collective attacks (Eqn. 3.16) one can conclude that analytically we know the finite key length under i.i.d. collective attacks for the efficient BB84 protocol to be:

$$\ell_{\text{BB84}} \leq p_z^2(N-m)[1 - h(e_x + \frac{\mu}{2}) - f_{\text{EC}}h(e_z) - \delta(\bar{\varepsilon})] - \log_2(\frac{2}{\varepsilon_{\text{EC}}}) - 2\log_2(\frac{1}{\varepsilon_{\text{PA}}}) \tag{5.4}$$

where $H_\mu(X|E) = 1 - h(e_x + \frac{\mu}{2})$ and $H(X|Y) = h(e_z)$.

To calculate the key length and rate, one would want to minimize the variational bound μ so as to minimize the increase in the cost of privacy amplification. As the variation bound of μ is partially determined by the size of the parameter estimation alphabet, one would like to reduce the alphabet of parameter estimation. As the key rate only cares about the phase error e_x, it makes sense to assume Alice and Bob only check the phase error, e_x, using the coarse-grained POVM, $\{\Pi_{e_x}, \mathbb{1} - \Pi_{e_x}\}$ where $\Pi_{e_x} \equiv (\mathbb{1}_A \otimes \mathbb{1}_B - \sigma_X \otimes \sigma_X)/2$ and σ_X is the Pauli-X operator.

To show that our approach works, we consider our numerical key rate against the analytical key rate calculated using Eqn. 5.4 in Fig. 5.1. Following [53] and the description of Protocol 2, we assume Alice and Bob sacrifice $(1 - p_z)^2 N$ of the signals to parameter estimation. This is a good choice since as N goes to infinity, one can test an increasingly smaller fraction, so p_z can approach unity, which this a priori decision takes into account. Furthermore, in the simulation we assume that our observations yield that the error rates satisfy $e_z = e_x$ to let $H(X|Y) = h(e_x)$. As can be seen in Fig. 5.1, for this protocol our solver produces a lower bound that effectively matches the analytical result perfectly.[1]

[1] The theory curve is always slightly higher than the numerical curve by the construction of the numerical method. However, by slightly differ, for this protocol, we mean that the *rate* differs by roughly less than half a percent.

Furthermore, in this example, we let $f_{\text{EC}} = 1.2$ as this is a realistic model of the inefficiency of error correction in current experiments [44, 53, 54].

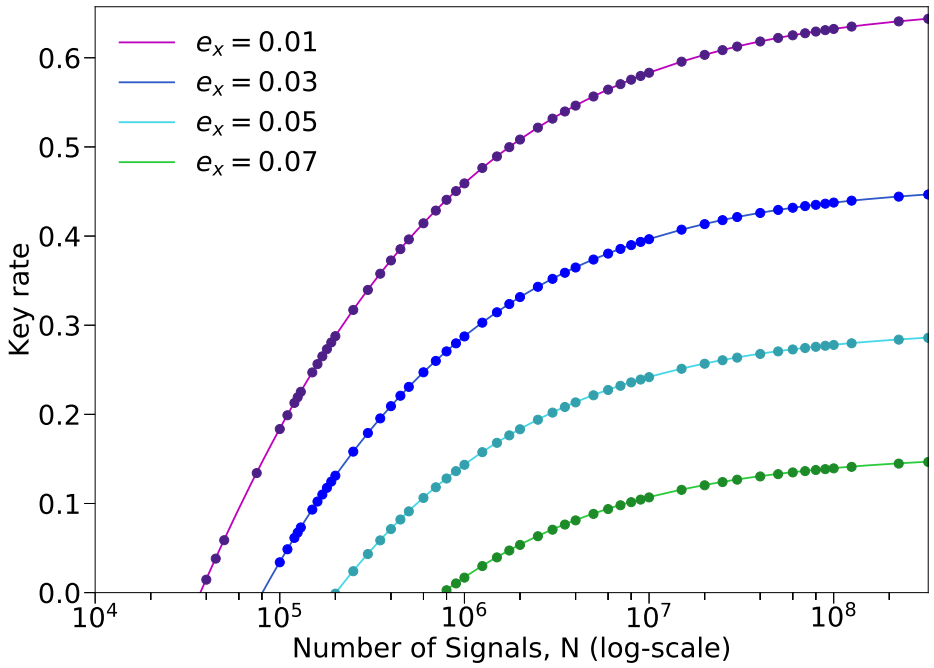


Figure 5.1: Numerical key rate versus analytic key rate for BB84 for four error rates with $\varepsilon_{\text{PE}} = \bar{\varepsilon} = \varepsilon_{\text{EC}} = \varepsilon_{\text{PA}} = \frac{1}{4} \times 10^{-8}$ so that $\varepsilon = 10^{-8}$. The lines are the theory curves, and the dots are the corresponding solutions by our numerical method. We let $e_z = e_x$ and $p_z = 0.9$. We assume here that the sample size is still larger than the block length of error correction, which then gives f_{EC} =1.20. Results generated using SDPT3.

5.2.1 Coherent Attack Rates

The current belief of the field is there should exist information-theoretic theorems which will show that the secure key length of QKD protocols under coherent attacks should behave roughly, i.e. up to an $O(1)$ correction term, the same as the security of collective i.i.d attacks. This belief is primarily based on intuition and the fact this is true for the set of QKD protocols which satisfy the requirements of the Entropy Accumulation Theorem [21]. As such, this chapter focuses on the i.i.d. collective attack key length. However, it

is worthwhile to present a comparison of the i.i.d. collective attack to the coherent attack analysis using the post-selection technique and the Finite Quantum de Finetti theorem for this simple protocol so that readers might have some intuition for how things work currently.

In Fig. 5.2, we consider unique acceptance about the phase error statistics with 1% observed phase error for efficient BB84. As this example is simple enough to not need numerics, the results are analytic and have been optimized over choice of p_z at each point. For all three curves we let the protocol be $\varepsilon = 10^{-8}$-secure. For the collective attack that means letting all the ε-terms be a fourth of the entire term. For the post-selection technique (see Eqn. 3.21), we just take the collective attack ε-terms and divide them all by $(N+1)^{d_{AB}^2-1}$.[2] We note there may be better ratios for the ε-terms for this specific choice (See [53]), but as a simple random sampling over both ε-ratios and observed frequencies at the same time numerically did not show a specific choice, I use the flat distribution for ε-terms since the actual protocol will fix these terms anyways.

For the Finite Quantum de Finetti coherent attack analysis (see Eqn. 3.17), there is more to be said. First we let each of the five ε-terms be a fifth of the entire term. Second, we must decide on the number k of the total signals N Alice and Bob should throw out to, in effect, force Eve's attack to be i.i.d.-esque. As one can see from Eqn. 3.17, one needs the fraction of signals k/N to approach 0 as the number of total signals used, N, approaches to infinity as a correction term is proportional to k. However, as the parameter r (Eqn. 3.20) is proportional to $\ln(k)/k$, and the variational distance μ (Eqn. 3.18) and the correction term δ (Eqn. 3.19) depend on $h(r/m)$ and $h(r/n)$ respectively, one must make sure that k is small enough that r does not grow too big. The optimal choice of k is therefore not obvious. As such, for each number of signals we optimize over both p_z and the fraction of N which we take for k. This is the correct choice as it turns out the optimal value of these two parameters does not seem well behaved at low key rates, as can be seen in Fig. 5.2. We do note however that it becomes well behaved once the number of signals is large enough to achieve at least $\approx 1/2$ of the asymptotic key rate. This also tells us how much more complicated using the Finite Quantum de Finetti coherent attack analysis is than applying the post-selection technique.

As one can see in Fig. 5.2, the post-selection technique does quite well for this protocol compared to the collective attack. It follows that as long as the dimension of the shared quantum state of Alice and Bob remains relatively small, the coherent attack curve will

[2]This isn't totally rigorous as really there should be an extra ε-term due to the fact error correction is not normally a permutation invariant map (See Sec. 3.4.3, and specifically Eqn. 3.92, of [4]), but for our purposes this is sufficient.

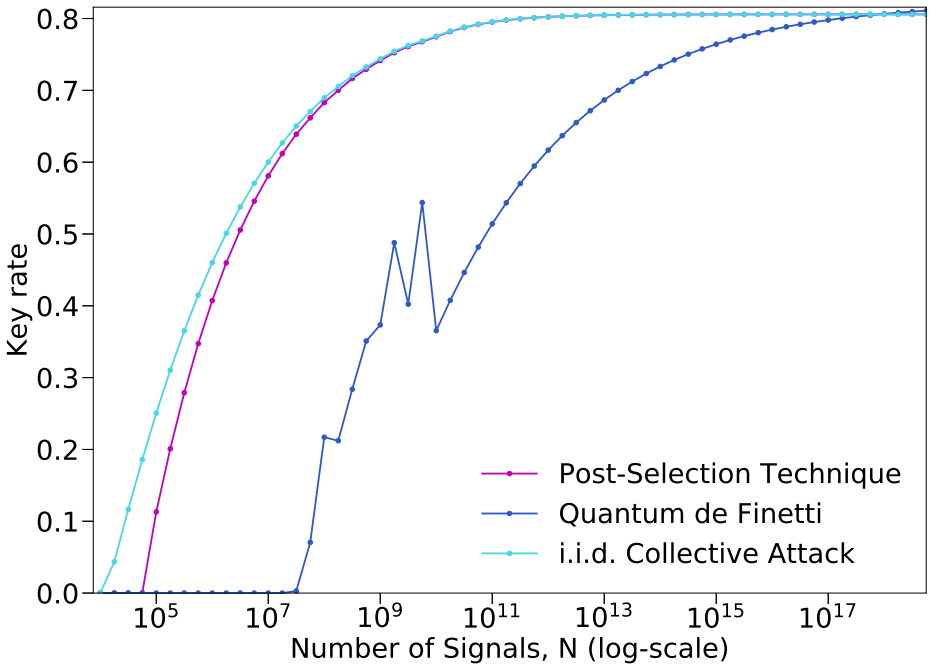


Figure 5.2: Here we consider the two types of coherent attack analysis from Chapter 3. At every point p_z has been optimized for each respective curve. We assume here that the sample size is still larger than the block length of error correction, which then gives f_{EC} =1.20. The Quantum de Finetti theorem does not converge to the same asymptotic result due to how it handles correlations between Eve's purification of the states sifted.

not cost too many signals (depending on if 'too much' gets to be decided by the theorist or someone who actually has to make this work efficiently). We do note though that the range of the plot was chosen because for more signals you have to use a high-precision float to avoid having Matlab round the ε terms to zero for the post-selection technique. In contrast, beyond the aforementioned annoyance that the Quantum de Finetti coherent attack analysis requires one to carefully optimize over multiple parameters, one can see it also requires many signals to achieve a positive key rate or approach the asymptotic limit. It in fact requires $\approx 5.6 \times 10^{19}$ signals to achieve 99% of the asymptotic limit for this very simple protocol. It therefore follows that while the Quantum de Finetti Theorem proves that coherent attack analysis reduces to i.i.d. attack analysis in the limit, it is not particularly practical for applications.

5.3 Reference Frame Misaligned Efficient BB84

One may notice in the previous section one was able to improve the key rate by using the coarse-grained POVM to reduce the value of the variational bound μ. This would imply that somehow Alice and Bob throwing out information allowed them to be more secure against Eve. In this section we explore the effect of fine-grained data versus coarse-grained data on the key rate and the increased importance of the difference in the finite regime. Furthermore we show the advantage of considering multiple coarse-grainings (Theorem 5) rather than only one, which resolves the idea that Alice and Bob throwing out information may allow them to be more secure against Eve.

In the case of constraining the set of density matrices using a single frequency distribution F, there are two competing effects— the rate at which the variation bound μ goes to 0 and the value of the asymptotic key rate. As one can see from Eqn. 3.14, the number of POVM outcomes effects the size of the variation bound μ. This means that more coarse-grained data F^{C_k} has a variation bound μ_k that converges to 0 faster than that of the fine-grained data. It follows that for a case such as in the first example where an element of a coarse-grained probability distribution (e_x) determines the key rate (Eqn. 5.4), the coarse-grained data will lead to a better or equal key rate to the fine-grained data for any amount of signals.

However, we know that if one applies a unitary rotation about the Y-axis on the Bloch sphere to each signal sent to Bob, then the fine-grained statistics will detect the rotation, thereby leaving the key rate unchanged. In contrast, the phase error coarse-grained statistics cannot determine the rotation, thereby decreasing the coarse-grained key rate. This can be seen in the following manner. Consider the efficient BB84 protocol under the source-replacement scheme where Alice sends Bob's half of the Bell state $|\Phi^+\rangle$ through a channel that is the composition of two channels. The first channel is the depolarizing channel with noise value q defined as:

$$\Phi^q_{dp}(X) = \sum_{k=0}^{3} p_k \sigma_k(X)\sigma_k$$

where $p_0 = 1 - \frac{3q}{4}, p_1 = p_2 = p_3 = \frac{q}{4}$ and $\sigma_0 = \mathbb{1}_2, \sigma_1 = \sigma_X, \sigma_2 = \sigma_\mathrm{Y}, \sigma_3 = \sigma_Z$ where $\sigma_X, \sigma_\mathrm{Y}, \sigma_Z$ are the Pauli operators. The depolarizing channel induces a qubit error rate of q in the output state. The second channel is a unitary channel that rotates the state about the Y-axis on the Bloch sphere by an angle θ, $\Phi_U(X) = e^{i\theta\sigma_\mathrm{Y}} X e^{-i\theta\sigma_\mathrm{Y}}$. It follows the

output state will be of the form

$$\begin{aligned}\rho_{out} &= \left(\mathrm{id}_A \otimes (\Phi_U \circ \Phi_{dp}^q)\right)\left(\left|\Phi^+\right\rangle\!\left\langle\Phi^+\right|\right) \\ &= (\mathrm{id}_A \otimes \Phi_U)\left(1 - \frac{3q}{4}\left|\Phi^+\right\rangle\!\left\langle\Phi^+\right| + \frac{q}{4}(\left|\Phi^-\right\rangle\!\left\langle\Phi^-\right| + \left|\Psi^+\right\rangle\!\left\langle\Psi^+\right| + \left|\Psi^-\right\rangle\!\left\langle\Psi^-\right|)\right) \\ &= \begin{bmatrix} a\cos^2\theta + b\sin^2\theta & (b-a)\cos\theta\sin\theta & c\cos\theta\sin\theta & c\cos^2\theta \\ (b-a)\cos\theta\sin\theta & b\cos^2\theta + a\sin^2\theta & d\sin^2\theta & -c\cos\theta\sin\theta \\ c\cos\theta\sin\theta & d\sin^2\theta & b\cos^2\theta + a\sin^2\theta & (c-b)\cos\theta\sin\theta \\ c\cos^2\theta & -c\cos\theta\sin\theta & (c-b)\cos\theta\sin\theta & a\cos^2\theta + b\sin^2\theta \end{bmatrix}\end{aligned}$$

where $a \equiv \left(\frac{1}{2}(1-\frac{3q}{4}) + \frac{q}{8}\right)$, $b \equiv \frac{q}{4}$, $c \equiv \left(\frac{1}{2}(1-\frac{3q}{4}) - \frac{q}{8}\right)$, $d \equiv \left(\frac{1}{2}(-1+\frac{3q}{4}) + \frac{q}{8}\right)$, and the matrix is written in the computational basis. Using this output state from the channel model, we can construct the joint probability distribution table of Alice and Bob's observations just as we did for the simpler case (Eqn. 5.1) using that both Alice and Bob's local POVM is $\{p_z\left|0\right\rangle\!\left\langle 0\right|, p_z\left|1\right\rangle\!\left\langle 1\right|, (1-p_z)\left|+\right\rangle\!\left\langle +\right|, (1-p_z)\left|-\right\rangle\!\left\langle -\right|\}$:

$$p(i,j) = \begin{bmatrix} \frac{p_z^2}{4}(1+x) & \frac{p_z^2}{4}(1-x) & \frac{p_z(1-p_z)}{4}(1-y) & \frac{p_z(1-p_z)}{4}(1+y) \\ \frac{p_z^2}{4}(1-x) & \frac{p_z^2}{4}(1+x) & \frac{p_z(1-p_z)}{4}(1+y) & \frac{p_z(1-p_z)}{4}(1-y) \\ \frac{p_z(1-p_z)}{4}(1+y) & \frac{p_z(1-p_z)}{4}(1-y) & \frac{(1-p_z)^2}{4}(1+x) & \frac{(1-p_z)^2}{4}(1-x) \\ \frac{p_z(1-p_z)}{4}(1-y) & \frac{p_z(1-p_z)}{4}(1+y) & \frac{(1-p_z)^2}{4}(1-x) & \frac{(1-p_z)^2}{4}(1+x) \end{bmatrix} \tag{5.5}$$

where $x \equiv (1-q)\cos 2\theta$ and $y \equiv (1-q)\sin 2\theta$.

It follows from this calculation that in the asymptotic limit where we recover this probability distribution, the asymmetries induced by the rotation can be detected. As a unitary rotation leaks no information to Eve, one could use this to ignore the error due to rotation when determining how much privacy amplification is needed to decouple from Eve. However, if one coarse-grained the results to just check the phase error, this information would be lost and so one would have to perform privacy amplification as if all of the error were due to Eve, therefore reducing the key rate. This means asymptotically we understand that the fine-grained key rate is better than the coarse-grained key rate in the event of such a rotation. Therefore it follows that in the finite regime, even though the coarse-grained statistic variation bound converges to zero faster, the fine-grained key rate must be better than the coarse-grained key rate at some threshold of number of signals used.

We note that this point has been addressed independently of finite size effects in the literature where the fact that certain POVMs are robust to rotations has been utilized in the invention of the 'reference frame independent' and '6-state 4-state' protocols [37,

59]. The idea is that the information extracted by the POVM determines how robust the protocol is to differences in Alice and Bob's reference frames. This is because the signals sacrificed for the parameter estimation step allow them to in effect align their relevant reference frame [3]. For example, if we had rotated the states about the X-axis of the Bloch sphere, not even in the asymptotic limit would the fine-grained data of the BB84 protocol be able to determine the rotation. However, the six-state protocol, which is tomographically complete, would be robust to such a rotation about the Bloch sphere. In this section we present an example of this misalignment in reference frames in BB84 to explore its relation to finite size effects and the advantage of doing parameter estimation with multiple-coarse grainings.

We consider BB84 where we constrain with one or more of the following three conditional probability distributions where for intelligibility we write the corresponding POVM rather than the conditional probability distribution:

1. The fine-grained joint POVM constructed by both Alice and Bob having the local POVM:

$$\{p_z |0\rangle\langle 0|, p_z |1\rangle\langle 1|, (1-p_z) |+\rangle\langle +|, (1-p_z) |-\rangle\langle -|\} \tag{5.6}$$

   This corresponds to applying the identity conditional probability distribution to the fine-grained statistics.

2. The phase error POVM $\{\Pi_{e_x}, \mathbb{1} - \Pi_{e_x}\}$. This corresponds to mapping the frequencies corresponding to Alice and Bob both using the X-basis POVM and getting different results to a single outcome and all other fine-grained outcomes to a second.

3. The *agreement* POVM which simply checks how often Alice and Bob agree:

$$\{p_z^2\Pi_0, p_z^2\Pi_1, (1-p_z)^2\Pi_+, (1-p_z)^2\Pi_-, \Pi_{else}\}$$

   where $\Pi_a = |a\rangle\langle a| \otimes |a\rangle\langle a|$ and Π_{else} is the POVM element that completes the POVM. This corresponds to a conditional probability distribution that retains the statistics pertaining to Alice and Bob getting the same outcome and mapping all other fine-grained outcomes to a single outcome.

To evaluate the resulting key rates, we need to work with simulated observations, to construct the unique the frequency distribution F that Alice and Bob accept on. To do this, we consider the channel previously discussed in which Alice prepares $|\Phi^+\rangle$ and then sends half of it to Bob through the channel $\Phi_U \circ \Phi_{dp}^q$ for the case where the rotation is by 12°.

We then get the statistics by having Alice and Bob perform measurements on the state $(\mathrm{id}_A \otimes (\Phi_U \circ \Phi_{dp}^q))(|\Phi^+\rangle\langle\Phi^+|)$ using one of the POVMs previously described to generate the probabilities.

In Fig. 5.3, we plot the key rate for all three coarse-grainings individually as well as the key rate when we consider both the phase-error statistics and the fine-grained statistics. To look at this, in Fig. 5.3, whenever $m \leq 10^8$ we construct a frequency distribution by randomly sampling the simulated probability distribution using a pseudo-random function and then calculate the key rate for the protocol with unique acceptance which accepts on that obtained frequency distribution. To see how much the key rate fluctuates when sampling m times depending on the frequency distribution Alice and Bob accept, we chose to repeat the simulation 20 times to determine the average key rate and standard deviation of the protocol with unique acceptance with all other parameters fixed. The standard deviation is represented by the error bars in Fig. 5.3. Furthermore, to make the comparison between the different POVMs fair, we optimize the choice of p_z at each point by maximizing the average key rate over p_z for each value of N we plot. As in the previous example, we let $m = (1-p_z)^2 N$ and assume they do the key map only in the Z-basis. Lastly, the (observed) error correction cost for all four key rates is $f_{\mathrm{EC}} H(\mathrm{X}|\mathrm{Y}) = f_{\mathrm{EC}} h(\bar{e}_z)$ where $f_{\mathrm{EC}} = 1.2$ and $\bar{e}_z$ is the bit error frequency determined by the fine-grained statistics in the key-generation basis Z.

Given Fig. 5.3, we now see how in some regime coarse-graining does better than fine-grained data due to the coarse-grained variational bound μ_k converging to zero faster than the fine-grained variational bound. However, we also see that as N increases, the fine-grained data begins to outperform the coarse-grained data because asymptotically the fine-grained data provides a better key rate. We also see that considering both the coarse-grained and fine-grained data together improves the key rate for all N. This is because whatever density matrix satisfies both sets of constraints has the phase error lower than just the fine-grained data and the unitary can be 'undone' to a greater degree than by just using the phase error coarse-grained data. For this reason in the finite regime it will only be beneficial to always optimize over the fine-grained data as well as relevant coarse-grainings. The ability for our solver to do this regardless of the number of outcomes is one property which makes our solver truly general and practical.

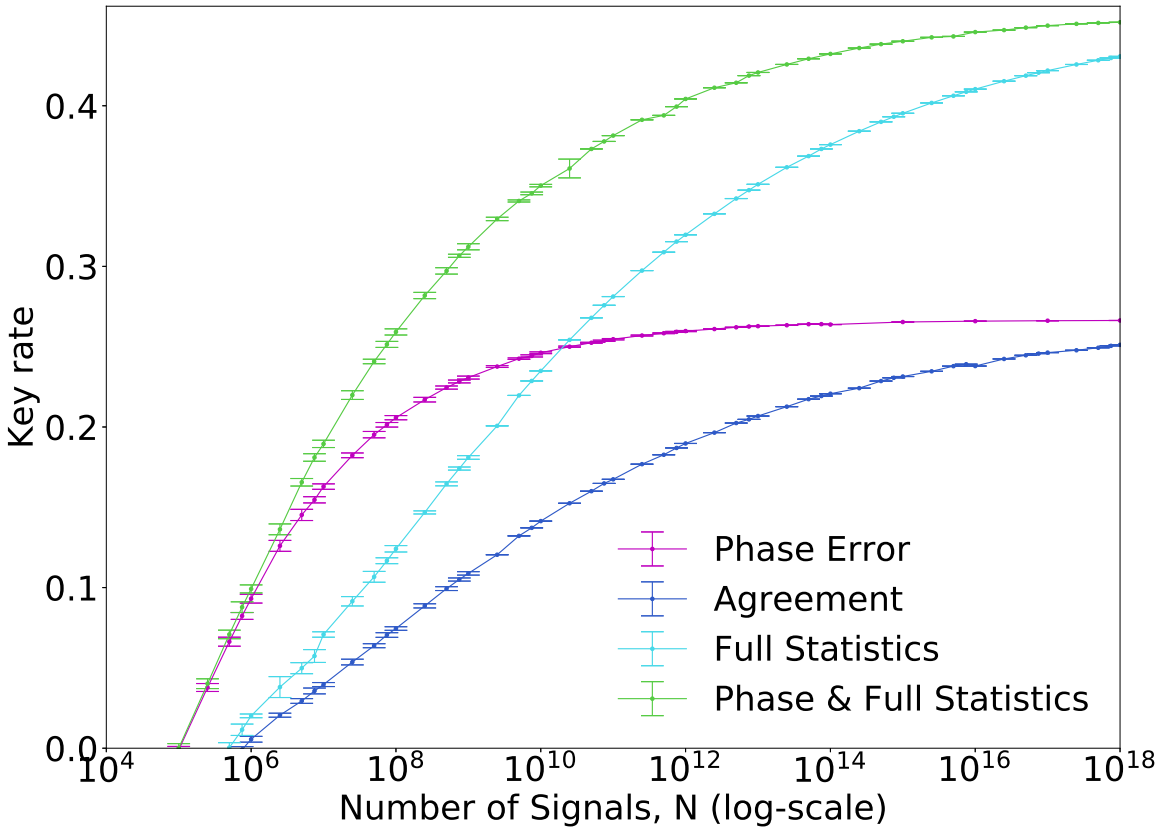


Figure 5.3: We consider four different parameter estimation constraints for $BB84$ transmitted through a depolarizing channel with $q = 0.02$ when the signal states have been rotated by 12° about the Y-axis on the Bloch sphere. Each point has p_z numerically optimized for maximum key rate. The error bars are from checking the key rate for 20 trials of sampling the distribution whenever the number of signals used for parameter estimation was less than 10^8 and calculating the standard deviation. For all curves we let $\varepsilon_{\text{PE}} = \bar{\varepsilon} = \varepsilon_{\text{EC}} = \varepsilon_{\text{PA}} = \frac{1}{4} \times 10^{-8}$. Results generated using SDPT3.

5.4 Measurement-Device-Independent BB84

In this section we consider the simple extension to MDI-QKD protocols which are designed to be immune to side-channel attacks on measurement devices [42]. Specifically we consider MDI-BB84 with perfect single photon sources in which Alice and Bob both send BB84 states to an untrusted third party Charlie who performs Bell state measurements on the two signals. Charlie then announces on which signals his measurement was successful as well as the outcome. Alice and Bob then do sifting on this subset and finally construct the key. The primary extension for finite key is that in MDI-QKD there is a third party. This means that there is a joint probability distribution over three alphabets and a joint POVM over three parties. This however is an immediate extension to our previous discussion on parameter estimation as parameter estimation can be defined for tripartite states and the third party in MDI QKD is a classical announcement and so does not effect Alice and Bob's fine-grained data.

To simulate data for the protocol, we apply source-replacement to both Alice and Bob's signal states resulting in a state $\rho_{ABA'B'}$. In our calculation, we assume the setup is using linear optics, so Charlie can only discriminate unambiguously two of the Bell state measurements, Ψ_+ and Ψ_- where $\Psi_\pm \equiv \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle)$. For simulating the statistics, we consider that the signal portions of the states, A' and B', each go through a separate depolarizing channel Φ_{dp}^q as they are sent to Charlie. Lastly, we assume Alice and Bob only do the key map in the Z-basis for simplicity. In Fig. 5.4 we consider MDI-BB84 with $p_z = 0.5$ for two depolarizing parameter values to see the rate of converging to the asymptotic key rate as a simple example.

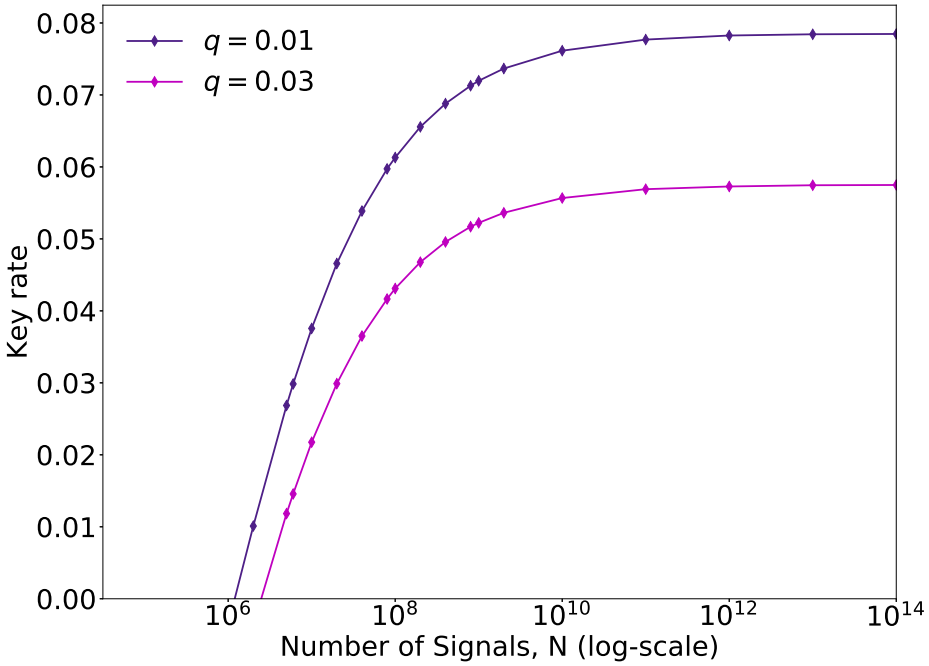


Figure 5.4: Here we see the MDI-BB84 protocol with unique acceptance converging to its asymptotic value as the number of signals is increased for depolarizing channels with depolarizing parameter values $q = 0.01$ and $q = 0.03$. For all curves, the security is defined by $\varepsilon_{\text{PE}} = \bar{\varepsilon} = \varepsilon_{\text{EC}} = \varepsilon_{\text{PA}} = \frac{1}{4} \times 10^{-8}$. Results generated using SDPT3.

5.5 Discrete-Phase-Randomized BB84

We next apply our method to a QKD protocol optically implemented with weak coherent pulses. Since each state that Bob receives is an optical mode and is in principle manipulated by Eve, a full description of the POVM usually involves an infinite-dimensional Hilbert space (e.g. Fock space). This also means that the density operator ρ_{AB} in our optimization problem is infinite-dimensional such that no numerical optimization algorithm can solve the problem directly. Fortunately, for many discrete-variable QKD protocols, there exists a squashing model [5, 48, 69] that reduces the apparent infinite-dimensional representation to an effective finite-dimensional subspace representation. Here, we present our finite key analysis for the discrete-phase-randomized BB84 protocol [12], which is based on phase-encoding and has a squashing model [5].

We consider the following simple model for determining the statistics.

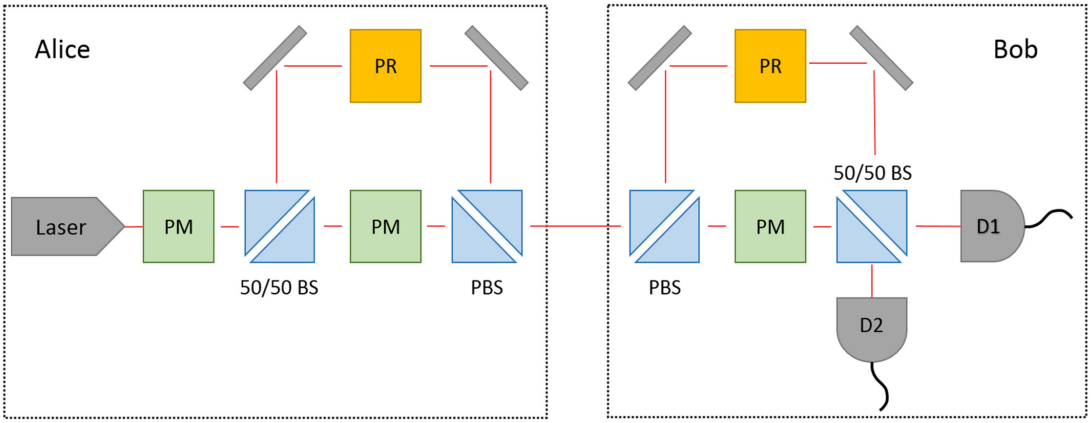


Figure 5.5: Schematic for discrete-phase-randomized BB84. PM stands for phase modulator, PBS stands for polarizing beam splitter, BS stands for beam splitter, PR stands for polarization rotator, and D1 and D2 are two threshold detectors.

As depicted in Fig. 5.5, the quantum part of the protocol is

1. Alice sends two-mode coherent states $\left|\sqrt{\nu}e^{i\theta}\right\rangle_r \left|\sqrt{\nu}e^{i(\theta+\phi_A)}\right\rangle_s$, to Bob where the first mode is the reference pulse and the second mode is the signal pulse. The global phase θ is chosen at random from the set $\{\frac{2\pi k}{c} : k = 0, \ldots, c-1\}$ where c is the number of different global phases. The key information is encoded in the relative phase ϕ_A chosen from the Z basis $\{0, \pi\}$ or X basis $\{\frac{\pi}{2}, \frac{3\pi}{2}\}$.

2. After receiving states from Alice, Bob may choose to measure in one of the two basis by applying a relative phase $\phi_B \in \{0, \frac{\pi}{2}\}$ to the reference pulse, where $\phi_B = 0$ corresponds to Z basis and $\phi_B = \frac{\pi}{2}$ to X basis. This results in either one, none, or both of Bob's detectors clicking. In the case where both detectors click, Bob assigns the result to either just detector 1 clicking or just detector 2 clicking.

We remark that the protocol with c=1, in which case Alice does not randomize the global phase, is also studied in [38, 43].

For our simulation, unlike previous examples, we consider a lossy channel parameterized by the single-photon transmittance $\eta = 10^{-\alpha_{att}L/10}$ for a distance L (in kilometers) between Alice and Bob. We also introduce a channel noise parameterized by ζ, which describes

the relative phase drift between the signal pulse and the reference pulse. In addition, imperfection of Bob's detectors is taken into account by the dark count probability p_d and the dectector efficiency η_d. To obtain simulated statistics, we choose $\eta_d = 0.045$, $p_d = 8.5 \times 10^{-7}$, and let the attenuation coefficient be $\alpha_{att} = 0.2$ dB/km, from the experimental parameters reported in [27]. We also set $\zeta = 11°$, which produces a misalignment error of 1% at 0 km distance and let $f_{EC} = 1.16$ as was done in [38]. Due to the fact in most optical QKD analyses there are approximations used such that the probability of the observations considered do not sum to unity, which is an assumption for our solver to give correct lower bounds, the derivation of the observed statistics for this protocol was done for this work specifically. For completeness, this derivation has been included as Appendix C.

Under the squashing model and source-replacement scheme, the fine-grained statistics for this protocol are generated by a $20c$-outcome joint POVM constructed by Alice and Bob's local POVMs where Alice has $4c$ POVM elements which are projectors on to her $4c$ possible signal states and Bob has a 5-outcome POVM defined as:

$$\{1/2\,|0\rangle\langle 0| \oplus 0, 1/2\,|1\rangle\langle 1| \oplus 0,$$
$$1/2\,|+\rangle\langle +| \oplus 0, 1/2\,|-\rangle\langle -| \oplus 0, |\text{vac}\rangle\langle \text{vac}|\}$$

In other words, Bob's local POVM is the standard fine-grained local BB84 POVM (Eqn. 5.6 with $p_z = 1/2$) embedded in a three-dimensional space plus a projector onto the third dimension where the third dimension is the vacuum state and $|\text{vac}\rangle$ denotes the basis of the third dimension.

We take $L = 100$ km and $L = 20$ km and consider both $c = 1$ and $c = 2$ scenarios as an example to show the method works for multiple discrete phases and loss regimes. In this model the dark counts are the primary source of error. In generating this plot, to improve the key rate when less signals are sent, we optimize the fraction of signals that would be used for parameter estimation, which we denote $g_{\text{PE}} \equiv m/N$, heuristically. The fraction is determined as follows:

$$g_{\text{PE}}^{L=20km} = \begin{cases} 0.99 & N < 1.31 \times 10^{11} \\ \frac{1.1\times 10^{11}}{N} + (0.5)^{\log_{10}(N)/4} & \text{else} \end{cases}$$

$$g_{\text{PE}}^{L=100km} = \begin{cases} 0.99 & N < 2.75 \times 10^{14} \\ \frac{2.35\times 10^{14}}{N} + (0.5)^{\log_{10}(N)/5} & \text{else} \end{cases}$$

The first term of line 2 of each g_{PE} was determined by numerically determining for how many signals the key rate could be made positive for $c = 1$. The extra term was decided

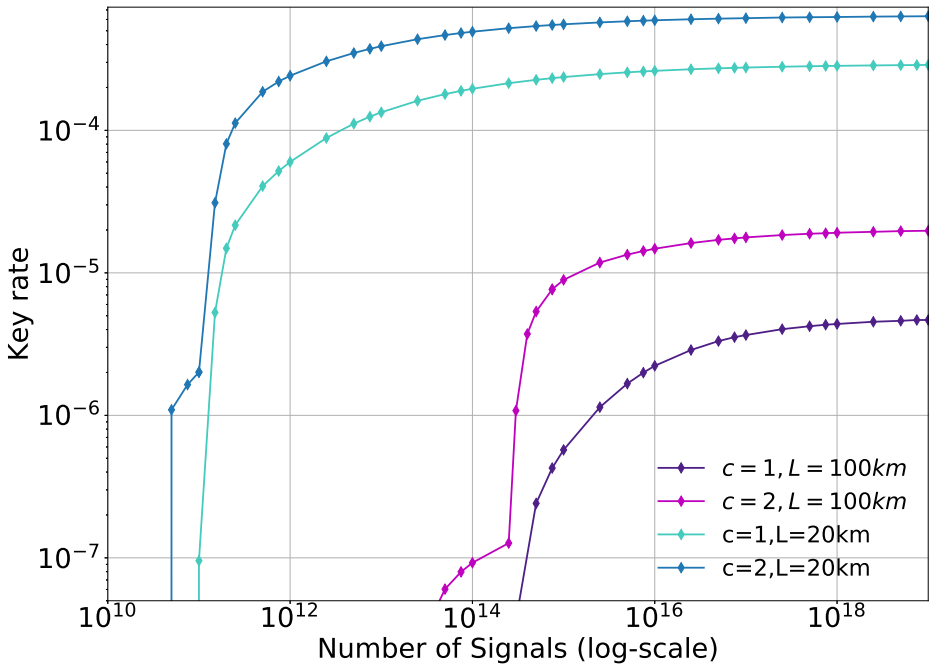


Figure 5.6: Key rate of discrete-phase-randomized BB84 with unique acceptance when not randomizing the global phase ($c = 1$) and randomizing it over 2 choices ($c = 2$). Every point is for optimized coherent state intensity ν. For all curves, the security is defined by $\varepsilon_{\text{PE}} = \bar{\varepsilon} = \varepsilon_{\text{EC}} = \varepsilon_{\text{PA}} = \frac{1}{4} \times 10^{-8}$. For this protocol we let $f_{\text{EC}} = 1.16$. Results generated used Mosek.

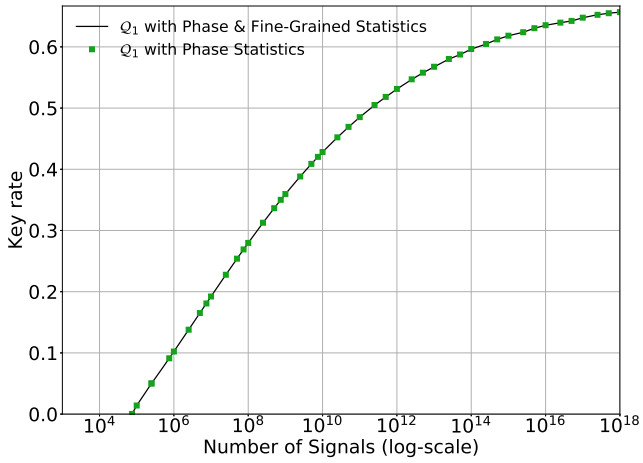
so as to sacrifice a smaller fraction to parameter estimation as N grows so that the key rate is improved.

We notice that with our simulation parameters, at $L = 100$ km considered in Fig. 5.6, a significant amount of signals needs to be sent before the key rate becomes nonzero. The reason is that at $L = 100$ km, the probability of the outcomes that will lead to key generation is quite low, at the order 10^{-6} in the $c = 1$ case. It follows that if the variation bound μ is of an order greater than 10^{-6}, there exists a probability distribution P such that $\|P - F\|_1 \leq \mu$ and P corresponds to a density matrix that lacks sufficient correlation for any key to be distilled. Therefore one needs to sacrifice enough signals to parameter estimation such that the variation bound μ is sufficiently small with respect to the portion of the frequency distribution relevant to key distillation.

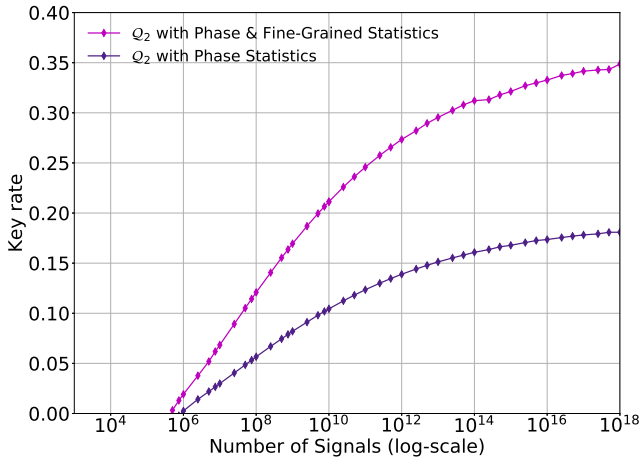
Lastly, we note that in Fig. 5.6, we see a jump in the $c = 2$ key rates at a given number of signals. This is not a poor behaviour of our numerical solver as one can check that the second step of our algorithm improves upon the Frank-Wolfe Algorithm benchmark we use to see if our better lowerbound is behaving properly. Rather we believe this to because at this point the variational bound μ is tight enough to force Eve's optimal attack to change. Such observations are one major advantage of using numerics.

5.6 BB84 with Practical Acceptance Set

So far we have only presented protocols with unique acceptance. However, protocols with unique acceptance are impractical as the probability that an experiment yields the exact frequency distribution of outcomes that match the acceptance criteria is usually very low. Thus one introduces a range of accepted statistics, where the key rate is now to be taken over the worst case scenario of the accepted statistics. Therefore, there is a trade-off between how often one aborts, and the length of the secret key generated when the protocol does not abort. In some cases, especially where the accepted statistics is only based on one observable, such as an error rate, and the key rate has some monotonic behaviour, it is easy to identify the worst-case acceptable statistics. In these cases one can relate the case of a set of accepted statistics back to the case of a single accepted statistics, namely the identified worst case statistics. However, in cases where the observed statistics needed for determining the key rate of the protocol are more complex, it is often not as simple to identify the worst case statistics. In these scenarios, our numerical method is a powerful



(a)



(b)

Figure 5.7: (a) Key rate of the BB84 protocol for accepting statistics in $\mathcal{Q}_1$ where either just the phase statistics or both the phase statistics and the fine-grained statistics are used to determine the key rate. We see in this case the fine-grained data does not help for this protocol. (b) Key rate of the BB84 protocol for accepting observed statistics in $\mathcal{Q}_2$ where either just the phase statistics or both the phase statistics and the fine-grained statistics are used to determine the key rate. We see in this case the fine-grained data help for this protocol, and so an analytical key rate calculation is difficult. For both (a) and (b), each point p_z is optimized and the security is defined by $\varepsilon_{\text{PE}} = \bar{\varepsilon} = \varepsilon_{\text{EC}} = \varepsilon_{\text{PA}} = \frac{1}{4} \times 10^{-8}$. Results generated using SDPT3.

tool for determining a tight lower bound of the secret key rate. Here we present an example of determining the secure key rate for single-photon BB84 in the practical setting where multiple frequency distributions are accepted by Alice and Bob to show how our numerical approach may help.

We again return to the efficient BB84 protocol. We consider two sets of frequency distributions to accept corresponding to whether their protocol has ideal behaviour or is suffering from misalignment due to the quantum channel. Following the notation in Eqn. 4.8, the first set, $\mathcal{Q}_1$, is defined by letting $\overline{\mathcal{N}}(\overline{F})$ be the two-outcome frequency distribution of 'phase error' and 'no phase error' with no observed phase errors ($e_x = 0$). We refer to $\mathcal{Q}_1$ as the phase set. The second set, $\mathcal{Q}_2$, is defined by letting $\overline{\mathcal{N}}(\overline{F}) = \overline{F}$ be the asymptotic results of the fine-grained statistics given the model from Section 5.3. We refer to $\mathcal{Q}_2$ as the rotated set. In both cases, the variation threshold, t, is $2(1-p_z)^2\overline{e}_x$ where $\overline{e}_x$ is the maximum tolerated observed error from $\overline{F}$. For this example we let $\overline{e}_x = 0.02$. The factor of $(1-p_z)^2$ is so that the variation threshold stays the same as p_z is varied to optimize the key rate.

Given the definition of the phase set, $\mathcal{Q}_1$, the key rate can be determined analytically as one can replace e_x in Eqn. 5.4 by $\overline{e}_x$. Furthermore, as no data more fine-grained than the phase error is needed in this case, it is clear that the multiple coarse-grainings will not further improve the key rate. These observations are verified numerically in Fig. 5.7a. However, in the case where the observed statistics would be contained in $\mathcal{Q}_2$ rather than in $\mathcal{Q}_1$, an analytical tight lower bound of the key rate is not a reasonable task as the structure of the worst case scenario is no longer simple. This is seen in Fig. 5.7b, where our numerical result shows that multiple coarse-grainings helps to obtain a tighter key rate when $\mathcal{Q}_2$ is used. It follows that obtaining a tight key rate analytically would be difficult as one needs to utilize both fine-grained statistics and multiple coarse-grainings.

More generally, this tells us the optimal choice of $\mathcal{Q}$ in certain implementations may be difficult due to issues such as misalignment errors. In such cases, even in the honest implementation, the statistics one ought to accept are fine-grained data that, because of complications, lack certain symmetries in Alice and Bob's results. This in turn limits one's a priori knowledge of what form the worst-case scenario observed statistics will take. This is further aggravated by the trade-off between how often the protocol will be aborted and the length of the secret key when the protocol does not abort. For these reasons, constructing a good choice of $\mathcal{Q}$ is a non-trivial task due to common issues in implementing QKD protocols. As it is designed for generic protocols, our numerical method allows for further exploration of these difficulties which cannot be explored analytically.

Chapter 6

Conclusions and Open Problems

It is good to now summarize what we have done and learned. In Chapter 3 we delved into the theory of finite key analysis. In particular we focused on the parameter estimation subprotocol, the definition of ε-securely filtered states, and the description of completeness and robustness of (sub)protocols. In doing this, we showed how one can consider multiple ways of processing one's data to decrease the set of states of which we don't know whether or not they are ε-securely filtered. This new result we saw to improve the key rates for asymmetric observations in QKD protocols in Chapter 5. In Chapter 4, we derived a general numerical method for determining the finite key rate of device-dependent QKD protocols that can be represented in finite-dimensional Hilbert spaces. Finally, in Chapter 5, beyond showing our ability to improve the finite key rate for asymmetric observations, we showed the generality of our method in its ability to determine the secure key length for MDI QKD and optically implemented QKD protocols. These considerations tell us that we need to use all of the information available to us to get the best key rates, particularly when we consider asymmetric observations, and we now have a numerical method that allows us to do this when we cannot analytically.

With all of these points handled, we can therefore conclude that finite key analysis has been solved and we all may move on to new and exciting research problems. Just kidding. We all know that is not how this works. In fact, for a field that presented a clean framework more than a decade ago, the field of finite key analysis seems to continuously get more complicated. We therefore briefly summarize what avenues should be taken next, though the ordering of the paths to be taken is unclear.

Extension of Method for More Protocols

This work has spoken nothing of continuous-variable QKD which is a current popular QKD implementation. The issue in implementing our numerics for this protocol is that they do not (currently) admit a squashing model and thus can't be handled in finite-dimensions directly. This has been handled under some assumptions in the asymptotic regime [39], but the unification of these approximations with the finite-size is not immediate.

There is also the notion of decoy state protocols which are another popular method of implementing QKD. In principle, one could handle it the same way as was done for the discrete-phase-randomized BB84 in this thesis (Section 5.5). However, this would lead to large demands on the memory of the computer. Therefore, more tools must be developed for practical use of the numerics for decoy state protocols.

Adaptive Security

First we should note that in terms of security definitions, things are finished. As noted in the thesis, there are multiple frameworks [6, 47, 50] for establishing the formal security definition of a QKD protocol. However, as we saw in Section 3.5.2, finite key security proofs are effectively done exclusively for fixed length protocols in which the QKD protocol either aborts or produces an ε-secure key of fixed length. This is a problem as intuitively this seems inefficient in terms of the ratio of total bits of key generated over all runs to total number of signals used over all runs.

There is, to the author of this thesis's knowledge, a single publication [30] in which adaptive key length security is proven for a protocol. However, this is proven under the assumption that for all inputs Alice's raw key is a uniform distribution. This does not seem to be a reasonable assumption in coherent attack analysis nor would one expect it to hold for general protocols– even for noisy honest implementations. Furthermore, the proof of adaptive key security in [30] relies on bounding the trace norm between the ideal output key and the implementation's output key using the phase error correction finite key framework rather than the Renner finite key framework.[1] Therefore, it remains an open problem how one can prove adaptive key length security for general protocols. Specifically, it would be ideal if one could construct a near optimal method for constructing a ε-secure adaptive key length security protocol using ε'-secure fixed length security.[2]

[1] A recent publication [65] argues the phase error correction and Renner framework are identical. An interesting way to understand this equivalence better might be to see how to convert the adaptive security argument from the phase error framework to the Renner framework using the conversions of [65].

[2] An intuitive starting point would be to fix an $\varepsilon' > 0$ and then carefully construct a finite family of fixed

Coherent Attack Analysis- Methods of Bounding $H_{\min}^{\varepsilon}$

In this thesis we discussed two methods for performing coherent attack analysis- the post-selection technique and the Quantum de Finetti theorem. Fundamentally, as is understood from the Leftover Hashing Lemma, the secure key that Alice and Bob can generate depends on the smooth min-entropy of the entire raw key with respect to Eve. Therefore, a more recent form of coherent attack analysis has been the Entropy Accumulation Theorem (EAT) [20, 21] which is able to bound the smooth min-entropy of the entire raw key by the i.i.d. collective attack rate (up to an $O(\sqrt{n})$ correction term) [2] at the demand that, roughly, the output of the entire protocol is a Markov chain between the first j rounds of public announcements and the $j+1, ...n$ bits of the raw key for all $j \in [N-1]$. In effect, this means the QKD protocol's side information needs to be seeded exclusively by randomness. However, the Finite Quantum de Finetti coherent attack analysis tells us this Markov condition shouldn't be necessary at least asymptotically, though it should come at some cost to how quickly the finite key rate converges to the asymptotic key rate. We therefore need a result which can interpolate between this ideal Markov chain property and the Finite Quantum de Finetti result. One such proposal [21] would be to find a variation of the EAT for approximate quantum Markov chains [58]. Such a result would allow us to extend the numerical method in this thesis to get key rates for coherent attacks which we would expect to be relatively close to those presented in Chapter 5.

length protocols, $\{\text{QKD}^j\}_{j\in\Sigma}$, for the same implementation as follows. First construct a set of disjoint sets of accepted frequency distributions $\{\mathcal{Q}_j\}_{j\in\Sigma}$ and fix a number of signals used for parameter estimation m such that the set of states considered which will be accepted by each fixed length protocol is disjoint from the others, i.e. $\mathbf{S}_\mu^j \cap \mathbf{S}_\mu^{j'} = \emptyset$ for all $j, j' \in \Sigma$ such that $j \neq j'$. It will certainly follow $\ell^j \neq \ell^{j'}$ for $j \neq j'$. Then if one hashes to ℓ^j whenever $F \in \mathcal{Q}_j$ is observed, the protocol should be $\approx \varepsilon'$-secure (though a correction term is necessary due to the ε-filtered states having probability of leading to each observation). This however is presumably not the best we can do.

Bibliography

[1] Antonio Acín, Serge Massar, and Stefano Pironio. "Efficient quantum key distribution secure against no-signalling eavesdroppers". In: *New Journal of Physics* 8.8 (Aug. 2006), pp. 126–126. DOI: 10.1088/1367-2630/8/8/126.

[2] Rotem Arnon-Friedman. "Reductions to IID in Device-independent Quantum Information Processing". In: (Dec. 28, 2018). arXiv: http://arxiv.org/abs/1812.10922v1 [quant-ph].

[3] Stephen D. Bartlett, Terry Rudolph, and Robert W. Spekkens. "Reference frames, superselection rules, and quantum information". In: *Reviews of Modern Physics* 79.2 (Apr. 2007), pp. 555–609. DOI: 10.1103/revmodphys.79.555.

[4] Normand J. Beaudry. "Assumptions in Quantum Cryptography". In: (May 11, 2015). arXiv: http://arxiv.org/abs/1505.02792v1 [quant-ph].

[5] Normand J. Beaudry, Tobias Moroder, and Norbert Lütkenhaus. "Squashing Models for Optical Measurements in Quantum Communication". In: *Phys. Rev. Lett.* 101 (2008), p. 093601.

[6] M. Ben-Or et al. "The Universal Composable Security of Quantum Key Distribution". In: *Second Theory of Cryptography Conference, TCC 2005, Cambridge, MA, USA, February 10-12, 2005.* Ed. by Joe Kilian. Vol. 3378. Lecture Notes in Computer Science. Berlin: Springer, 2005, pp. 386–406.

[7] C. H. Bennett and G. Brassard. "Quantum Cryptography: Public key distribution and coin tossing." In: *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India.* New York: IEEE, Dec. 1984, pp. 175–179.

[8] C. H. Bennett, G. Brassard, and N. D. Mermin. "Quantum Cryptography without Bell's Theorem". In: *PRL* 68.5 (1992), pp. 557–559.

[9] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge: Cambridge University Press, 2004.

[10] Sylvia Bratzik et al. “Min-entropy and quantum key distribution: Nonzero key rates for “small” numbers of signals”. In: *Physical Review A* 83.2 (Feb. 2011), p. 022330. DOI: 10.1103/physreva.83.022330.

[11] Darius Bunandar et al. “Numerical finite-key analysis of quantum key distribution”. In: (Nov. 18, 2019). arXiv: http://arxiv.org/abs/1911.07860v1 [quant-ph].

[12] Zhu Cao et al. “Discrete-phase-randomized coherent state source and its application in quantum key distribution”. In: *New Journal of Physics* 17.5 (May 2015), p. 053014. DOI: 10.1088/1367-2630/17/5/053014.

[13] Carlton M. Caves, Christopher A. Fuchs, and Rüdiger Schack. “Unknown quantum states: The quantum de Finetti representation”. In: *Journal of Mathematical Physics* 43.9 (Sept. 2002), pp. 4537–4559. DOI: 10.1063/1.1494475.

[14] Matthias Christandl, Robert König, and Renato Renner. “Postselection Technique for Quantum Channels with Applications to Quantum Cryptography”. In: *Physical Review Letters* 102.2 (Jan. 2009), p. 020504. DOI: 10.1103/physrevlett.102.020504.

[15] Roger Colbeck and Renato Renner. “No extension of quantum theory can have improved predictive power”. In: *Nature Communications* 2.1 (Aug. 2011). DOI: 10.1038/ncomms1416.

[16] Patrick J. Coles. “Unification of different views of decoherence and discord”. In: *Physical Review A* 85.4 (Apr. 2012), p. 042103. DOI: 10.1103/physreva.85.042103.

[17] Patrick J. Coles, Eric M. Metodiev, and Norbert Lütkenhaus. “Numerical approach for unstructured quantum key distribution”. In: *Nature Communications* 7.1 (May 2016). DOI: 10.1038/ncomms11712.

[18] M. Curty, M. Lewenstein, and N. Lütkenhaus. “Entanglement as precondition for secure quantum key distribution”. In: *PRL* 92 (2004), p. 217903.

[19] I. Devetak and A. Winter. “Distillation of secret key entanglement from quantum states”. In: *Proc. of the Roy. Soc. of London Series A* 461.2053 (2005), pp. 207–235.

[20] Frederic Dupuis and Omar Fawzi. “Entropy Accumulation With Improved Second-Order Term”. In: *IEEE Transactions on Information Theory* 65.11 (Nov. 2019), pp. 7596–7612. DOI: 10.1109/tit.2019.2929564.

[21] Frederic Dupuis, Omar Fawzi, and Renato Renner. “Entropy accumulation”. In: (July 6, 2016). arXiv: http://arxiv.org/abs/1607.01796v1 [quant-ph].

[22] A. Ekert. “Quantum cryptography based on Bell’s Theorem”. In: 67.6 (1991), pp. 661–663.

[23] Philippe Faist. "Quantum Coarse-Graining: An Information-Theoretic Approach to Thermodynamics". In: (July 11, 2016). arXiv: http://arxiv.org/abs/1607.03104v1 [quant-ph].

[24] Agnes Ferenczi and Norbert Lütkenhaus. "Symmetries in quantum key distribution and the connection between optimal attacks and optimal cloning". In: *Phys. Rev. A* 85.5 (), p. 052310. DOI: 10.1103/PhysRevA.85.052310. URL: http://link.aps.org/doi/10.1103/PhysRevA.85.052310.

[25] K. Ferentios. "On Tcebycheff's type inequalities". In: *Trabajos de Estadistica y de Investigacion Operativa* 33.1 (Feb. 1982), pp. 125–132. DOI: 10.1007/bf02888707.

[26] Ian George, Jie Lin, and Norbert Lütkenhaus. "Numerical Calculations of Finite Key Rate for General Quantum Key Distribution Protocols". In: (Apr. 24, 2020). arXiv: http://arxiv.org/abs/2004.11865v1 [quant-ph].

[27] C. Gobby, Z.L. Yuan, and A.J. Shields. "Quantum key distribution over 122km of standard telecom fiber". In: *Appl. Phys. Lett.* 84 (2004), pp. 3762–3764.

[28] F. Grosshans et al. "Virtual Entanglement and Reconciliation Protocols for Quantum Cryptography with Continuous Variables". In: (June 20, 2003). arXiv: http://arxiv.org/abs/quant-ph/0306141v1 [quant-ph].

[29] Masahito Hayashi. "Upper bounds of eavesdropper's performances in finite-length code with the decoy method". In: *Physical Review A* 76.1 (July 2007). DOI: 10.1103/physreva.76.012329.

[30] Masahito Hayashi and Toyohiro Tsurumaru. "Concise and tight security analysis of the Bennett–Brassard 1984 protocol with finite key lengths". In: *New Journal of Physics* 14.9 (Sept. 2012), p. 093014. DOI: 10.1088/1367-2630/14/9/093014.

[31] Matthias Heid and Norbert Lütkenhaus. "Efficiency of coherent-state quantum cryptography in the presence of loss: Influence of realistic error correction". In: *Physical Review A* 73.5 (May 2006), p. 1. DOI: 10.1103/physreva.73.052316.

[32] C. W. Helstrom. *Quantum detection and estimation theory*. New York: Academic Press, 1976.

[33] R. L. Hudson and G. R. Moody. "Locally normal symmetric states and an analogue of de Finetti's theorem". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 33.4 (Dec. 1976), pp. 343–351. DOI: 10.1007/bf00534784.

[34] A Yu Kitaev. "Quantum computations: algorithms and error correction". In: *Russian Mathematical Surveys* 52.6 (Dec. 1997), pp. 1191–1249. DOI: 10.1070/rm1997v052n06abeh002155.

[35] Robert Konig, Renato Renner, and Christian Schaffner. "The Operational Meaning of Min- and Max-Entropy". In: *IEEE Transactions on Information Theory* 55.9 (Sept. 2009), pp. 4337–4347. DOI: 10.1109/tit.2009.2025545.

[36] Robert König et al. "Small Accessible Quantum Information Does Not Imply Security". In: *Physical Review Letters* 98.14 (Apr. 2007). DOI: 10.1103/physrevlett.98.140502.

[37] Anthony Laing et al. "Reference-frame-independent quantum key distribution". In: *Physical Review A* 82.1 (July 2010), p. 012304. DOI: 10.1103/physreva.82.012304.

[38] Jie Lin. "Security Proofs for Quantum Key Distribution Protocols by Numerical Approaches". MA thesis. University of Waterloo, 2017.

[39] Jie Lin, Twesh Upadhyaya, and Norbert Lütkenhaus. "Asymptotic security analysis of discrete-modulated continuous-variable quantum key distribution". In: (May 26, 2019). arXiv: http://arxiv.org/abs/1905.10896v1 [quant-ph].

[40] Jie Lin et al. "Towards an Open-source Software Platform for Numerical Key Rate Calculation of General Quantum Key Distribution Protocols". In: Qcrypt 2020 (Aug. 11, 2020). URL: https://2020.qcrypt.net/posters/QCrypt2020Poster139Wang.pdf.

[41] H. K. Lo, F. Chau, and M. Ardehali. "Efficient quantum key distribution scheme and proof of its unconditional security". In: *J. Cryptology* 18 (2005), pp. 133–165. DOI: 10.1007/s00145-004-0142-y.

[42] Hoi-Kwong Lo, Marcos Curty, and Bing Qi. "Measurement-Device-Independent Quantum Key Distribution". In: *Physical Review Letters* 108.13 (Mar. 2012), p. 130503. DOI: 10.1103/physrevlett.108.130503.

[43] Hoi-Kwong Lo and John Preskill. "Security of quantum key distribution using weak coherent states with nonrandom phases". In: *Quant. Inf. Comput. 8 (2007) 431-458* (Oct. 23, 2006). arXiv: http://arxiv.org/abs/quant-ph/0610203v2 [quant-ph].

[44] M. Lucamarini et al. "Practical Security Bounds Against the Trojan-Horse Attack in Quantum Key Distribution". In: *Physical Review X* 5.3 (Sept. 2015). DOI: 10.1103/physrevx.5.031030.

[45] Kento Maeda, Toshihiko Sasaki, and Masato Koashi. "Repeaterless quantum key distribution with efficient finite-key analysis overcoming the rate-distance limit". In: *Nature Communications* 10.1 (July 2019). DOI: 10.1038/s41467-019-11008-z.

[46] Lluis Masanes et al. "Full Security of Quantum Key Distribution From No-Signaling Constraints". In: *IEEE Transactions on Information Theory* 60.8 (Aug. 2014), pp. 4973–4986. DOI: 10.1109/tit.2014.2329417.

[47] Ueli Maurer and Renato Renner. "Abstract Cryptography". In: *Innovations in Computer Science - ICS 2011, Tsinghua University, Beijing, China, January 7-9, 2011. Proceedings*. Ed. by Bernard Chazelle. Tsinghua University Press, 2011, pp. 1–21. URL: http://conference.iiis.tsinghua.edu.cn/ICS2011/content/papers/14.html.

[48] Tobias Moroder et al. "Entanglement verification with realistic measurement devices via squashing operations". In: *Physical Review A* 81 (2010), p. 052342.

[49] John von Neumann. *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, 1955.

[50] Christopher Portmann and Renato Renner. "Cryptographic security of quantum key distribution". In: (Sept. 11, 2014). arXiv: http://arxiv.org/abs/1409.3525v1 [quant-ph].

[51] Renato Renner. "Security of Quantum Key Distribution". In: *Int. J. Quantum Inf.* 06.01 (Feb. 2008), pp. 1–127. DOI: 10.1142/s0219749908003256. eprint: arXiv:quant-ph/0512258v2.

[52] V. Scarani et al. "The security of practical quantum key distribution". In: *Review of Modern Physics* 81 (2009), pp. 1301–1350.

[53] Valerio Scarani and Renato Renner. "Quantum Cryptography with Finite Resources: Unconditional Security Bound for Discrete-Variable Protocols with One-Way Postprocessing". In: *Physical Review Letters* 100.20 (May 2008), p. 200501. DOI: 10.1103/physrevlett.100.200501.

[54] Valerio Scarani and Renato Renner. "Security Bounds for Quantum Cryptography with Finite Resources". In: *Proceedings of TQC2008, Lecture Notes in Computer Science 5106 (Springer Verlag, Berlin), pp. 83-95 (2008)* (June 1, 2008). arXiv: http://arxiv.org/abs/0806.0120v1 [quant-ph].

[55] Alexander Shapiro. "Semi-infinite programming, duality, discretization and optimality conditions†". In: *Optimization* 58.2 (Feb. 2009), pp. 133–161. DOI: 10.1080/02331930902730070.

[56] Peter W. Shor and John Preskill. "Simple Proof of Security of the BB84 Quantum Key Distribution Protocol". In: *Physical Review Letters* 85.2 (July 2000), pp. 441–444. DOI: 10.1103/physrevlett.85.441.

[57] Jamie Sikora and John H. Selby. "On the impossibility of coin-flipping in generalized probabilistic theories via discretizations of semi-infinite programs". In: (Jan. 15, 2019). arXiv: http://arxiv.org/abs/1901.04876v1 [quant-ph].

[58] David Sutter. "Approximate Quantum Markov Chains". In: *Approximate Quantum Markov Chains*. Springer International Publishing, 2018, pp. 75–100. DOI: 10.1007/978-3-319-78732-9_5.

[59] Ramy Tannous et al. "Demonstration of a 6 state-4 state reference frame independent channel for quantum key distribution". In: *Applied Physics Letters* 115.21 (Nov. 2019), p. 211103. DOI: 10.1063/1.5125700.

[60] Joy A. Thomas Thomas M. Cover. *Elements of Information Theory*. Wiley John + Sons, Sept. 8, 2006. 776 pp. ISBN: 0471241954.

[61] Marco Tomamichel. "A Framework for Non-Asymptotic Quantum Information Theory". In: (Mar. 9, 2012). arXiv: http://arxiv.org/abs/1203.2142v2 [quant-ph].

[62] Marco Tomamichel. *Quantum Information Processing with Finite Resources*. Springer International Publishing, 2016. DOI: 10.1007/978-3-319-21891-5.

[63] Marco Tomamichel, Roger Colbeck, and Renato Renner. "A Fully Quantum Asymptotic Equipartition Property". In: *IEEE Transactions on Information Theory* 55.12 (Dec. 2009), pp. 5840–5847. DOI: 10.1109/tit.2009.2032797.

[64] Marco Tomamichel, Roger Colbeck, and Renato Renner. "Duality Between Smooth Min- and Max-Entropies". In: *IEEE Transactions on Information Theory* 56.9 (Sept. 2010), pp. 4674–4681. DOI: 10.1109/tit.2010.2054130.

[65] Toyohiro Tsurumaru. "Leftover Hashing From Quantum Error Correction: Unifying the Two Approaches to the Security Proof of Quantum Key Distribution". In: *IEEE Transactions on Information Theory* 66.6 (June 2020), pp. 3465–3484. DOI: 10.1109/tit.2020.2969656.

[66] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. DOI: 10.1017/9781316848142.

[67] Mark M. Wilde. *From Classical to Quantum Shannon Theory*. June 7, 2011. DOI: 10.1017/9781316809976.001. arXiv: http://arxiv.org/abs/1106.1445v8 [quant-ph].

[68] Adam Winick, Norbert Lütkenhaus, and Patrick J. Coles. "Reliable numerical key rates for quantum key distribution". In: *Quantum* 2 (July 2018), p. 77. DOI: 10.22331/q-2018-07-26-77.

[69] Yanbao Zhang et al. "Security proof of practical quantum key distribution with detection-efficiency mismatch". In: (Apr. 9, 2020). arXiv: http://arxiv.org/abs/2004.04383v1 [quant-ph].

APPENDICES

Appendix A

Direct Proof of Tightness

In this appendix we are officially interested in proving Lemma 16 for multiple coarse-grainings in a direct manner. More honestly we are interested in the connection between semi-infinite programming (SIP) and semidefinite programming (SDP) for finite-dimensional quantum information theory. As such, we are interested in proving that when $\mathbf{S}_{\mu'\epsilon't'} \neq \emptyset$,

$$\min_{\sigma \in \mathbf{S}_{\mu'\epsilon't'}} \mathrm{Tr}[(\sigma - \rho^*)\nabla f(\rho^*)] \geq 0 \tag{A.1}$$

where ρ^* is the optimal solution to the original problem, $\min_{\rho \in \mathbf{S}_\mu} f(\rho)$. We split the proof of this into a series of lemmas. I would like to acknowledge Jamie Sikora and Jie Lin who read and gave feedback on various forms of this section and acknowledge Jie Lin helped with the proof of Lemma 24.

First we recall that $\mathbf{S}_{\mu'\epsilon't'}$ is defined as the set of density matrices ρ which satisfy the constraints in the following SDP:

$$\begin{aligned} \text{minimize} \quad & f(\rho) \\ \text{subject to} \quad & \|\Phi_{\mathcal{P}}(\rho) - \mathcal{N}(F)\|_1 \leq \mu' \\ & \|\overline{\mathcal{N}}(F) - \overline{F}\|_1 \leq t' \\ & |\,\mathrm{Tr}(\rho\Gamma_i) - \gamma_i| \leq \epsilon' \qquad \forall i \in \Lambda \\ & F \in \mathcal{P}(\Sigma_f) \\ & \rho \succeq 0 \end{aligned} \tag{A.2}$$

where $\rho \in \mathrm{Pos}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and Σ_f is the alphabet for the fine-grained statistics of the QKD protocol. $\Phi_{\mathcal{P}}$ maps density matrices to probability distributions in $\mathcal{P}(\Sigma)$, $\mathcal{N}$ maps a frequency distribution F to a frequency distribution in $\mathcal{P}(\Sigma)$.

Motivation for Semi-Infinite Programming

First we note the motivation for using semi-infinite programming to prove Eqn. A.1. In [68] (Lemma 2) the authors were able to directly prove this property for the set of density matrices considered in asymptotic QKD, $\mathbf{S}$, using the Karush-Kuhn-Tucker (KKT) conditions for optimality (Section 5.5.3 of [9]). This was because the constraints were all of the form $\mathrm{Tr}(\rho\Gamma_i) = \gamma_i$ which made the proof straightforward. The issue with using the same approach for finite key is that one has to consider the frequency distributions, $\{F_k\}$, which makes converting the set optimized over into one expressed in just terms of $\mathrm{Tr}\big(\rho\overline{\Gamma}_i\big) \leq \overline{\gamma}_i$ non-trivial. This is not only in the case of considering multiple frequency distributions, but because the coarse-graining may make comparing F to $\Phi_{\mathcal{P}}(\rho)$ in such a manner to reduce it to a restriction on ρ impossible. However we note that the type of constraints we want are just by definition descriptions of half-spaces. It is well-known [9] that a closed convex set may be written as the intersection of (a possibly infinite set of) halfspaces. Therefore if our set is is a closed convex set of density matrices, then we could convert it into this type of constraint and possible prove tightness that way. This requires using semi-infinite programming as the number halfspaces may be infinite. Then (using an argument where we take a limit) we can use the semi-infinite program to prove this property directly.[1]

To prove what we want, we first show how to write a semi-infinite program over the semidefinite cone. This uses definitions relevant to semi-infinite programs as defined in [55]. We then move on to a series of theorems to prove we can write our semi-definite program in Eqn. A.2 as a semi-infinite program as we can write it as the intersection of a set of half spaces. From there we prove that the problem is discretizable which effectively means there exists a sequence of finite subsets of the constraints such that the optimal value of the finite problem will approach the optimal value of the semi-infinite program. As the tightness property holds for any of these problems defined by a finite intersection of half spaces (as then KKT is clearly well defined and the approximations satisfy Slater's condition since the original problem did), one can conclude we get what we want in the limit (or up to arbitrary precision, however you'd like to view it).

[1] Depending on what your view of 'directly' proving something is, I suppose.

A.1 Semi-Infinite Programming for Quantum Information

In this section we convert semi-infinite programming [55] from optimizing over $\mathbb{R}^n$ to optimizing over $L(\mathcal{X})$.

Definition 4. *(Variation of Definition in [55]) A semi-infinite program over the linear operators $L(\mathcal{X})$ is of the form*

$$\begin{aligned} &\textit{minimize} && f(\sigma) \\ &\textit{subject to} && g(\sigma,\omega) \leq 0 \quad \forall \omega \in \Omega \\ & && \sigma \in L(\mathcal{X}) \end{aligned} \tag{A.3}$$

where Ω is a (possibly infinite) index set, $f : L(\mathcal{X}) \to \overline{\mathbb{R}}$ where $\overline{\mathbb{R}} \equiv \mathbb{R} \bigcup \{-\infty, \infty\}$, and $g : L(\mathcal{X}) \to \mathbb{R}$.

Definition 5. *Let $\mathcal{X} = \mathbb{C}^{|\Sigma|}$, $\Phi : L(\mathcal{X}) \to \overline{\mathbb{R}}$. Define the following linear bijections:*

- $\Psi : \mathbb{C}^{|\Sigma|} \to \mathbb{R}^{2|\Sigma|}$ *which in matrix representation can be seen as*

$$\begin{pmatrix} a_1 + ib_1 \\ \vdots \\ a_{|\Sigma|} + ib_{|\Sigma|} \end{pmatrix} \to \left(a_1, \ldots, a_{|\Sigma|}, b_1, \ldots, b_{|\Sigma|}\right)^T$$

- $\xi \equiv \Psi \otimes \Psi : \mathbb{C}^{|\Sigma|} \otimes \mathbb{C}^{|\Sigma|} \to \mathbb{R}^{2|\Sigma|} \otimes \mathbb{R}^{2|\Sigma|}$
- $\phi : \mathbb{R}^{2|\Sigma|} \otimes \mathbb{R}^{2|\Sigma|} \to \mathbb{R}^{4|\Sigma|^2}$ *is an isomorphism map between the two spaces.*
- *The vec mapping [66] $vec : L(\mathcal{X}, \mathcal{Y}) \to \mathcal{X} \otimes \mathcal{Y}$ where in this case $\mathcal{Y} = \mathcal{X}$*

Thus we define the corresponding vector map of Φ by $\vec{\Phi} \equiv \Phi \circ vec^{-1} \circ \xi^{-1} \circ \phi^{-1} : \mathbb{R}^{4|\Sigma|^2} \to \overline{\mathbb{R}}$. An identical definition of corresponding vector map may be used in the case that Φ maps to $\mathbb{R}$ instead of $\overline{R}$.

The isomorphisms used in the definition of the corresponding vector map also tell us that Definition 4 is equivalent to the one over $\mathbb{R}^n$. Furthermore, from Definition 4, one can define a semi-infinite program that is over the positive semidefinite cone by making the objective function f such that it has a finite value for some $\rho \in \text{Pos}(\mathcal{X})$ and such that for any $\sigma \not\geq 0$, $f = +\infty$.[2]

[2]Alternatively, one can guarantee that the set of functions $\{g_u(\sigma) \equiv \langle u, \sigma u \rangle : u \in \mathcal{X}\}$ is a subset of $\{g(\sigma,\omega)\}_{\omega \in \Omega}$ since such constraints are sufficient for defining the positive semidefinite cone.

Proposition 17. *If $\Phi : L(\mathcal{X}) \to \mathbb{R}$ is convex, the corresponding vector map $\vec{\Phi}$ is also convex.*

Proof. Let Φ be convex. Let $\vec{\Phi} \equiv \Phi \circ \text{vec}^{-1} \circ \xi^{-1} \circ \phi^{-1}$ and $\Psi \equiv \text{vec}^{-1} \circ \xi^{-1} \circ \phi^{-1}$. As the composition of linear maps are linear, Ψ is linear. Let $v, w \in \mathbb{R}^{4|\Sigma|^2}$, $t \in (0, 1)$.

$$\begin{aligned}
\vec{\Phi}(tv + (1-t)w) &= \Phi\Psi(tv + (1-t)w) \\
&= \Phi(t\Psi(v) + (1-t)\Psi(w)) \\
&\leq t\Phi\Psi(v) + (1-t)\Phi\Psi(w) \\
&= t\vec{\Phi}(v) + (1-t)\vec{\Phi}(w)
\end{aligned}$$

Thus, $\vec{\Phi}$ is convex. □

Definition 6. *A function $f : X \to \mathbb{R} \bigcup \{-\infty, \infty\}$ is lower semi-continuous at a point x_o if for every $y < f(x_o)$ there exists a neighborhood V of x_o such that $y < f(x)$ for all $x \in V$. Furthermore, a function $f : X \to \overline{\mathbb{R}}$ is lower semi-continuous if f is lower semi-continuous at x_o for all $x_o \in dom(f)$.*

Definition 7. *A function $f : X \to \overline{\mathbb{R}}$ is proper convex if f is a convex function taking values on $\overline{\mathbb{R}}$ and $\exists x \in X$ such that $f(x) < +\infty$ and $\forall x \in X$, $f(x) > -\infty$.*

Proposition 18. *If $\Phi : L(\mathcal{X}) \to \overline{\mathbb{R}}$ is proper convex, the corresponding vector map $\vec{\Phi}$ is proper convex.*

Proof. Follows from Definitions 5 and 7 as well as Proposition 17. □

Definition 8. *(Definition 1.1 of [55]) A semi-infinite program is convex if the objective function is lower semi-continuous and proper convex and that for all $\omega \in \Omega$, $g(\sigma, \omega)$ is convex in the first argument.*

Definition 9. *(Definition 3.1 of [55]) Given a semi-infinite program P, a finite approximation of the program P, which is denoted by P_m, is called a discretization of P if P_m is constructed by taking $m \in \mathbb{N}$ of the constraints in P. The semi-infinte program P is discretizable if for all $\delta > 0$, there exists a discretization P_m such that $\alpha - \alpha_m \leq \delta$ where α denotes the optimal value for P and α_m denotes the optimal value for P_m.*

Proposition 19. *(Corollary 3.1 of [55]) If the semi-infinite program P is convex, and its set of optimal solutions $Sol(P)$ is non-empty and bounded, then P is discretizable.*

A.2 Direct Proof of Lemma 16

We will now use the results of the previous section to prove Lemma 16 from which one can prove tightness Eqn. 4.28.

Lemma 20. *(Page 36 of [9]) A closed convex set can be written as the intersection of halfspaces.*

Proposition 21. *The set of density matrices optimized over in the SDP in Eqn. A.2, $\mathbf{S}_{\mu'\epsilon't'}$, is a closed convex set.*

Proof. First re-write Eqn. A.2 as:

$$\begin{aligned} \text{minimize} \quad & f(\rho) \\ \text{subject to} \quad & g(\rho, F) \leq \mu' \\ & h(\rho, F) \leq t' \\ & |\operatorname{Tr}(\rho\Gamma_i) - \gamma_i| \leq \epsilon' \quad \forall i \in \Lambda \\ & \operatorname{Tr}(F) = 1 \\ & \rho, F \succeq 0 \end{aligned} \tag{A.4}$$

where $g(\rho, F) = \|\Phi_{\mathcal{P}}(\rho) - \mathcal{N}(F)\|_1$, and $h(\rho, F) = \|\overline{\mathcal{N}}(F) - \overline{F}\|_1$. As norms are continuous functions, and $[0, \mu']$, $[0, t']$, and $[0, \epsilon']$ are closed intervals, the pre-images of $g^{-1}([0, \mu'])$, $h^{-1}([0, t'])$, $|\operatorname{Tr}(\rho\Gamma_i) - \gamma_i|^{-1}([0, \epsilon'])$ are closed sets. Thus the set

$$\{(\rho, F) \in \operatorname{Pos}(\mathcal{X}) \times \operatorname{Pos}(\mathcal{Y}) : g(\rho, F) \leq \mu', h(\rho, F) \leq t', |\operatorname{Tr}(\rho\Gamma_i) - \gamma_i| \leq \epsilon' \; \forall i \in \Lambda \text{ and } \operatorname{Tr}(F) = 1\}$$

where $\mathcal{Y} \cong \mathbb{C}^{|\Sigma|}$ is a closed convex set. We can rewrite this set as:

$$\{(\rho, F) \in D^{\epsilon'}(\mathcal{X}) \times D(\mathcal{Y}) : g(\rho, F) \leq \mu', f(\rho, F) \leq t' \; \forall k \in \Xi \text{ and } |\operatorname{Tr}(\rho\Gamma_i) - \gamma_i| \leq \epsilon' \; \forall i \in \Lambda\} \tag{A.5}$$

where $D^{\epsilon'}(\mathcal{X}) = \{X \in \operatorname{Pos}(\mathcal{X}) : \operatorname{Tr}(X) \leq 1 + \epsilon'\}$. This can be done as without loss of generality $\Gamma_i = \mathbb{1}$ and so $\rho \in D^{\epsilon'}(\mathcal{X})$ and F is a probability distribution and thus in the set of density matrices. As the set of density matrices is compact [66], the projection map $\pi : D^{\epsilon'}(\mathcal{X}) \times D(\mathcal{Y}) \to D^{\epsilon'}(\mathcal{X})$ is a closed map and so applying this projection onto the set in Eqn. A.5 recovers $\mathbf{S}_{\mu'\epsilon't'}$ which tells us that $\mathbf{S}_{\mu'\epsilon't'}$ is a closed convex set. □

Note: In principle F may numerically fail to be a probability distribution. In that case, if $\|F\|_1 > 1$, just let $F \in D^{\epsilon''}(\mathcal{Y})$ where $\epsilon'' \equiv \|F\|_1 - 1$. This set is also compact as norms are continuous and $[0, 1 + \epsilon'']$ is compact by the Heine-Borel theorem.

Lemma 22. $\mathbf{S}_{\mu'\epsilon't'}$ *can be written as the intersection of halfspaces.*

Proof. Follows from Lemma 20 and Proposition 21. □

Proposition 23. *The SDP in A.2 can be written as the following semi-infinite program:*

$$\begin{aligned} &\textit{minimize} \quad f'(\rho) \\ &\textit{subject to} \quad \langle \rho, \widehat{\Gamma}_i \rangle \leq \widehat{\gamma}_i \quad \forall i \in \Omega \end{aligned} \tag{A.6}$$

where

$$f'(\rho) = \left\{ \begin{array}{ll} f(\rho) \equiv D(\mathcal{G}(\rho)||\mathcal{Z}(\mathcal{G}(\rho)) & \rho \succeq 0 \\ +\infty & \rho \not\succeq 0 \end{array} \right\} \tag{A.7}$$

and Ω *is an (infinite) index set.*

Proof. This follows directly from Lemma 22, Definition 4, and the definition of $f'(\rho)$. □

We now wish to prove the semi-infinite program in Eqn. A.6 is convex. By Definition 8, as we know all of the constraints are convex, we just need to show f' is proper convex and lower semi-continuous.

Lemma 24. *The function* $f' : L(\mathcal{X}) \to \overline{\mathbb{R}}$ *is proper convex and lower semi-continuous.*

Proof. That f' is proper convex follows from the positive definiteness of the quantum relative entropy. We therefore focus on proving it's a lower semi-continuous function (see Definition 6). As the positive semidefinite cone, $\mathrm{Pos}(\mathcal{X})$, is closed, the issue is the boundary points. We will construct the neighbourhood of an arbitrary boundary point necessary to satisfy Definition 6. Let x_o be a boundary point of the positive semidefinite cone. As x_o is positive semidefinite, $f'(x_o)$ is finite, and so $y < f'(x_o)$ is finite. For any point $\overline{x} \in \mathrm{Pos}(\mathcal{X})$, it follows from the continuity of quantum relative entropy that for every $y < f'(\overline{x})$ one can construct a neighbourhood of $\overline{x}$, $\overline{V}$, such that $y < f'(\overline{x}')$ for all $\overline{x}' \in \overline{V}$. Therefore, for all $y < f'(x_o)$ there exists a neighbourhood of x_o, V, such that the set $\widetilde{V} \equiv V \bigcap \mathrm{Pos}(\mathcal{X})$ is such that $y < f'(\widetilde{x})$ for all $\widetilde{x} \in \widetilde{V}$. Thus by Definition 6, we may conclude f' is lower semi-continuous. □

Therefore we know that Eqn. A.6 is a convex semi-infinite program.

Lemma 25. *The semi-infinite program in Eqn. A.6 is discretizable when the feasible set is nonempty.*

Proof. Let P denote the semi-infinite program in Eqn. A.6. Let $\mathcal{A} \neq \emptyset$ as that is the only case we are interested in. By Lemma 14, we know that Eqn. A.2 attains its optimal solution. It follows Eqn. A.6 also attains its optimal solution as they optimize over the same set, and so its set of optimal solutions, $\text{Sol}(P)$, is non-empty. The feasible set is bounded as $D^{\varepsilon}(\mathcal{X})$ is bounded and the subset of a bounded set is bounded. As $\text{Sol}(P)$ is a subset of the feasible set, it must also be bounded. Thus $\text{Sol}(P)$ is non-empty and bounded. Furthermore, we know P is convex. Therefore, by Proposition 19, the problem is discretizable. □

Proposition 26. *For all discretizations of the semi-infinite program in Eqn. A.6 it holds that*

$$\min_{\sigma \in \mathbf{S}_{\mu}^{d} \bigcap Pos(\mathcal{X})} \text{Tr}[(\sigma - \rho^*)\nabla f(\rho^*)] \geq 0 \tag{A.8}$$

where ρ^* optimizes f over the intersection of the feasible set of the discretization of Eqn. A.6, denoted $\mathbf{S}_{\mu}^{d}$, with the positive semidefinite cone, $\mathbf{S}_{\mu}^{d} \bigcap Pos(\mathcal{X})$.

Proof. Consider an arbitrary discretization of the semi-infinite program in Eqn A.6, which can be written in the following form:

$$\begin{aligned} \text{minimize} \quad & f'(\rho) \\ \text{subject to} \quad & \langle \rho, \widehat{\Gamma}_i \rangle \leq \widehat{\gamma}_i \quad \forall i \in \overline{\Omega} \end{aligned} \tag{A.9}$$

where $\overline{\Omega} \subset \Omega$ such that $|\overline{\Omega}| \in \mathbb{N}$. This is well defined for *any* discretization of Eqn. A.6 as $\text{dom}(f') = L(\mathcal{X})$.

One can then define the following SDP using the discretization:

$$\begin{aligned} \text{minimize} \quad & f(\rho) \\ \text{subject to} \quad & \langle \rho, \widehat{\Gamma}_i \rangle \leq \widehat{\gamma}_i \quad \forall i \in \overline{\Omega} \\ & \rho \succeq 0 \end{aligned} \tag{A.10}$$

By the definition of $f'(\rho)$ and $f(\rho)$, we see the optimal values of Eqn. A.9 and A.10 must be the same as every discretization contains some positive-semidefinite matrices if the original problem, Eqn. A.2, is feasible.

As the original problem satisfies Slater's condition and is a subset of the positive-semidefinite cone, not only must the discretization also satisfy Slater's condition (as $\mathbf{S}_{\mu} \subseteq \mathbf{S}_{\mu}^{d}$), but the intersection of the discretization and the positive-semidefinite cone must also satisfy Slater's condition (as $\mathbf{S}_{\mu} \subseteq \mathbf{S}_{\mu}^{d} \bigcap \text{Pos}(\mathcal{X})$). As the set $\mathbf{S}_{\mu}^{d} \bigcap \text{Pos}(\mathcal{X})$ is convex and

satisfies Slater's condition, the Karush-Kuhn-Tucker (KKT) conditions (Page 243 of [9]) hold. Therefore we have the following KKT conditions:

$$\begin{aligned}
\nabla f(\rho^*) + \sum_i \nu_i \widehat{\Gamma}_i - Z &= 0 \\
\nu_i(\langle \rho^*, \widehat{\Gamma}_i \rangle - \widehat{\gamma}_i) &= 0 \ \ \forall i \in \overline{\Omega} \\
\mathrm{Tr}(\rho^* Z) &= 0 \\
\nu_i &\geq 0 \ \ \forall i \in \overline{\Omega} \\
Z &\succeq 0
\end{aligned}$$

It follows: Let $\sigma \in \mathbf{S}_\mu^d \bigcap \mathrm{Pos}(\mathcal{X})$. It follows,

$$\begin{aligned}
&\mathrm{Tr}[(\sigma - \rho^*)\nabla f(\rho^*)] \\
=&\mathrm{Tr}\left[(\sigma - \rho^*)(-\sum_i \nu_i \widehat{\Gamma}_i + Z)\right] \\
=&\sum_i \nu_i (\mathrm{Tr}\left(\rho^* \widehat{\Gamma}_i\right) - \mathrm{Tr}\left(\sigma \widehat{\Gamma}_i\right)) + \mathrm{Tr}(\sigma Z) - \mathrm{Tr}(\rho^* Z) \\
=&\sum_i \nu_i (\mathrm{Tr}\left(\rho^* \widehat{\Gamma}_i\right) - \widehat{\gamma}_i + \widehat{\gamma}_i - \mathrm{Tr}\left(\sigma \widehat{\Gamma}_i\right)) + \mathrm{Tr}(\sigma Z) \\
=&\sum_i \nu_i (\widehat{\gamma}_i - \mathrm{Tr}\left(\sigma \widehat{\Gamma}_i\right)) + \mathrm{Tr}(\sigma Z) \\
\geq& 0
\end{aligned}$$

Where the first and fourth equality follows from the KKT conditions and the inequality follows from the definition of the feasible set. The inequality follows from $\overline{\gamma}_i - \mathrm{Tr}\left(\sigma \widehat{\Gamma}_i\right) \geq 0$ for all $i \in \overline{\Omega}$ for all feasible σ by Eqn. A.10 along with the KKT conditions. □

Proposition 27. *Let $\mathcal{A}_m$ be the feasible set of the discretization P_m of Eqn. A.6 such that $\alpha - \alpha_m \leq \delta$. Let $\mathbf{S}_\mu^\delta \equiv \mathcal{A}_m \bigcap Pos(\mathcal{X})$. Then*

$$\lim_{\delta \to 0^+} \min_{\sigma \in \mathbf{S}_\mu^\delta} \mathrm{Tr}[(\sigma - \rho^*)\nabla f(\rho^*)] \geq 0 \tag{A.11}$$

Proof. Follows from Propositions 19 and 26. □

One could then use Proposition 27 in proving tightness in Chapter 4.

Appendix B

Post-processing Maps for Examples

In this section we provide the post-processing maps, $\mathcal{G}$, for each example for completeness.

Single-Photon BB84

The examples in Sections 5.2 and 5.3 use the same map $\mathcal{G}$. In singe-photon BB84, Alice and Bob perform von Neumann measurements in the Z and X bases with probabilities p_z and $1-p_z$ respectively, the public information Alice and Bob announce are what bases they measure in, the private information is what outcome they got (represented by a 0 or 1) in both bases, and the sifting throws out any measurement where Alice and Bob did not use the same basis. Lastly we note that in Sections 5.2 and 5.3, we assumed Alice only performs the key map in the Z-basis. Therefore we have the following definitions for constructing the $\mathcal{G}$ map:

$$
\begin{aligned}
K_Z^A &= \sqrt{p_z}\,|0\rangle\langle 0|_A \otimes |0\rangle_{\widetilde{A}} \otimes |0\rangle_{\overline{A}} + \sqrt{p_z}\,|1\rangle\langle 1|_A \otimes |0\rangle_{\widetilde{A}} \otimes |1\rangle_{\overline{A}} \\
K_X^A &= \sqrt{1-p_z}\,|+\rangle\langle +|_A \otimes |1\rangle_{\widetilde{A}} \otimes |0\rangle_{\overline{A}} + \sqrt{1-p_z}\,|-\rangle\langle -|_A \otimes |1\rangle_{\widetilde{A}} \otimes |1\rangle_{\overline{A}} \\
K_Z^B &= \sqrt{p_z}\,|0\rangle\langle 0|_B \otimes |0\rangle_{\widetilde{B}} \otimes |0\rangle_{\overline{B}} + \sqrt{p_z}\,|1\rangle\langle 1|_B \otimes |0\rangle_{\widetilde{B}} \otimes |1\rangle_{\overline{B}} \\
K_X^B &= \sqrt{1-p_z}\,|+\rangle\langle +|_B \otimes |1\rangle_{\widetilde{B}} \otimes |0\rangle_{\overline{B}} + \sqrt{1-p_z}\,|-\rangle\langle -|_B \otimes |0\rangle_{\widetilde{B}} \otimes |1\rangle_{\overline{B}} \\
\Pi &= |0\rangle\langle 0|_{\widetilde{A}} \otimes |0\rangle\langle 0|_{\widetilde{B}} + |1\rangle\langle 1|_{\widetilde{A}} \otimes |1\rangle\langle 1|_{\widetilde{B}} \\
V &= |0\rangle_R \otimes |0\rangle\langle 0|_{\widetilde{A}} \otimes |0\rangle\langle 0|_{\overline{A}} \otimes |0\rangle\langle 0|_{\widetilde{B}} + |1\rangle_R \otimes |0\rangle\langle 0|_{\widetilde{A}} \otimes |1\rangle\langle 1|_{\overline{A}} \otimes |0\rangle\langle 0|_{\widetilde{B}}
\end{aligned}
$$

We note that while we used the source-replacement method, we used the Gram-Schmidt process to return Alice's space to the original size as explained in [24], which in this case reconstructs Alice's original POVM.

MDI BB84

For MDI BB84, as we consider the case where Alice and Bob only distill key from the Z basis, using the source-replacement scheme on both Alice and Bob's sources and the simplification rules explained in Appendix A of [39], there is only one Kraus operator for the entire map $\mathcal{G}$:

$$K_Z = (|0\rangle_R \otimes |0\rangle\langle 0|_A + |1\rangle_R \otimes |1\rangle\langle 1|_A) \otimes (|0\rangle\langle 0|_B + |1\rangle\langle 1|_B) \otimes (|0\rangle\langle 0|_C + |1\rangle\langle 1|_C)$$

Discrete-phase-randomized BB84

In the discrete-phase-randomized BB84, we begin from the use of the squashing model which results in Alice preparing 4 states for each global phase, and Bob having the 5-outcome POVM described in Section 5.5. Then by the source-replacement scheme on Alice, Alice's portion of the signal is a $4c$-dimensional Hilbert space $\mathcal{H}_A$ where c is the number of global phases Alice uses. In other words, $\mathcal{H}_A \cong \oplus_c \mathcal{H}_4$ where $\mathcal{H}_4$ is a 4-dimensional Hilbert space and $\oplus$ is the direct sum. To make the expression of the Kraus operators concise, define the projector $\Pi_n = |n\rangle\langle n|$ where $n \in \{0,1,2,3\}$. Then, using that Alice performs the key map along with the simplifications from Appendix A of [39], we have two Kraus operators which describe the action of $\mathcal{G}$:

$$\begin{aligned}
K_Z = |0\rangle_R \otimes \left(\bigoplus_c (\Pi_0)\right) \otimes \sqrt{p_z}(|0\rangle\langle 0|_B + |1\rangle\langle 1|_B) \otimes |0\rangle_{\tilde{A}} \\
+ |1\rangle_R \otimes \left(\bigoplus_c (\Pi_1)\right) \otimes \sqrt{p_z}(|0\rangle\langle 0|_B + |1\rangle\langle 1|_B) \otimes |0\rangle_{\tilde{A}} \\
K_X = |0\rangle_R \otimes \left(\bigoplus_c (\Pi_2)\right) \otimes \sqrt{1-p_z}(|+\rangle\langle +|_B + |-\rangle\langle -|_B) \otimes |1\rangle_{\tilde{A}} \\
+ |1\rangle_R \otimes \left(\bigoplus_c (\Pi_3)\right) \otimes \sqrt{1-p_z}(|+\rangle\langle +|_B + |-\rangle\langle -|_B) \otimes |1\rangle_{\tilde{A}}
\end{aligned}$$

where $\oplus_c \Pi_n$ is well defined for all n as $\Pi_n \in \mathcal{H}_4$ and $\mathcal{H}_A \cong \oplus_c \mathcal{H}_4$.

Appendix C

Derivation of Expected Observations of DPR BB84

In this appendix we derive the expected observed statistics for the discrete-phase-randomized BB84 which are used for the numerics in Section 5.5. We refer the reader to Fig. 5.5 for visualizing the protocol.

First recall that for input coherent states $|\alpha\rangle_a |\beta\rangle_b$, the output of a beamsplitter is

$$\left|\sqrt{t}\alpha + e^{i\varphi}\sqrt{r}\beta\right\rangle_c + \left|-\sqrt{r}e^{-i\varphi}\alpha + \sqrt{t}\beta\right\rangle_d$$

where t is the transmittance, r is the reflectance, and φ is a phase where $r + t = 1$. Furthermore, when the phase is taken to be trivial, you can just parameterize it by a single parameter, η, as $t \equiv \eta$ implies $r \equiv 1 - \eta$.

Denote θ as the discrete phase, ϕ as the phase encoding of the signal, ζ as the relative phase drift in the channel, and φ as Bob's choice of phase for basis choice. Then we can determine the output state as follows:

$$\begin{aligned}
&\left|\sqrt{2\mu}\right\rangle && \text{Alice's laser output}\\
\rightarrow &\left|\sqrt{2\mu}e^{i\theta}\right\rangle && \text{Discrete phase added (PM1)}\\
\rightarrow &\left|\sqrt{\mu}e^{i\theta}\right\rangle_S \left|\sqrt{\mu}e^{i\theta}\right\rangle_R && \text{50:50 BS}\\
\rightarrow &\left|\sqrt{\mu}e^{i(\theta+\phi)}\right\rangle_S \left|\sqrt{\mu}e^{i\theta}\right\rangle_R && \text{Phase Mode Encoding (PM2)}
\end{aligned}$$

$$
\begin{aligned}
&\to \left|\sqrt{\eta\mu}e^{i(\theta+\phi)}\right\rangle_S \left|\sqrt{\eta\mu}e^{i\theta}\right\rangle_R && \text{Channel Loss } \eta \text{ BS}\\
&\to \left|\sqrt{\eta\mu}e^{i(\theta+\phi+\zeta)}\right\rangle_S \left|\sqrt{\eta\mu}e^{i\theta}\right\rangle_R && \text{Relative phase drift } \zeta\\
&\to \left|\sqrt{\eta\mu}e^{i\theta}e^{i(\phi+\zeta)}\right\rangle \left|\sqrt{\eta\mu}e^{i\theta}e^{i\varphi}\right\rangle && \text{Bob makes basis choice (PM3)}\\
&\to \left|\sqrt{\frac{\eta\mu}{2}}e^{i\theta}\left(e^{i(\phi+\zeta)}+e^{i\varphi}\right)\right\rangle \left|\sqrt{\frac{\eta\mu}{2}}e^{i\theta}\left(e^{i\varphi}-e^{i(\phi+\zeta)}\right)\right\rangle && \text{Output to detectors (50:50 BS)}
\end{aligned}
$$

where we will refer to the first mode as the coherent state $|\Delta\rangle$ and the second mode as the coherent state $|\gamma\rangle$.

As we are interested in threshold detectors, both detectors have the following POVM:

$$F_{\text{vac}} = (1-d_b)\,|0\rangle\,0 \qquad\qquad F_{\text{click}} = \mathbb{1} - F_{\text{vac}}$$

where d_B is the probability of dark counts.[1]

This means the joint POVM is of the form:

$$
\begin{aligned}
F_{\text{no click}} &= (1-d_B)^2\,|0\rangle\langle 0|\otimes|0\rangle\langle 0|\\
F_{\text{click 1}} &= (1-d_B)\mathbb{1}\otimes|0\rangle\langle 0| - F_{\text{no click}}\\
F_{\text{click 2}} &= (1-d_B)\,|0\rangle\langle 0|\otimes\mathbb{1} - F_{\text{no click}}\\
F_{\text{dbl click}} &= \mathbb{1}\otimes\mathbb{1} - (1-d_B)\left[\mathbb{1}\otimes|0\rangle\langle 0| + |0\rangle\langle 0|\otimes\mathbb{1}\right] + F_{\text{no click}}\ .
\end{aligned}
$$

From this we need to calculate the basic probabilities of these detectors. Given the way I have expressed the joint POVM, we will just need the overlap of these coherent states the vacuum state:

$$
\begin{aligned}
\langle 0|\Delta\rangle &= \exp\left(-\frac{1}{2}(\frac{\eta\mu}{2})(e^{i(\phi+\zeta)}+e^{i\varphi})(e^{-i(\phi+\zeta)}+e^{-i\varphi})\right)\sum_{n=0}^{\infty}\frac{\Delta^n}{n!}\langle 0|\,(a^\dagger)^n\,|0\rangle\\
&= \exp\left(-\frac{\eta\mu}{4}(2+e^{i(\phi+\zeta-\varphi)}+e^{-i(\phi+\zeta-\varphi)})\right)\\
&= \exp\left(-\frac{\eta\mu}{2}(1+\cos g)\right)
\end{aligned}
$$

where $g \equiv \varphi-\phi-\zeta$ and the second equality follows from noting $\langle 0|\,(a^\dagger)^n\,|0\rangle = \delta_{n,0}$. By the same sort of calculation, we find

$$\langle 0|\gamma\rangle = \exp\left(-\frac{\eta\mu}{2}(1-\cos g)\right)\ .$$

[1] We note this is a simple model for the detectors. A more realistic model that has been handled in the asymptotic case using the numerical framework can be found in [69].

Using these, the calculations are straightforward:

$$P_{\text{no click}} = (1-d_B)^2 |\langle 0|\Delta\rangle|^2 |\langle 0|\gamma\rangle|^2 = (1-d_B)^2 \exp(-2\eta\mu)$$

$$\begin{aligned} P_{\text{click 1}} &= \text{Tr}((1-d_B)(\mathbb{1}\otimes|0\rangle\langle 0|)\,|\Delta\rangle\langle\Delta|\otimes|\gamma\rangle\langle\gamma|) - P_{\text{no click}} \\ &= (1-d_B)|\langle 0|\Delta\rangle|^2 - P_{\text{no click}} \\ &= (1-d_B)\exp(-\eta\mu(1-\cos g)) - (1-d_B)^2\exp(-2\eta\mu) \end{aligned}$$

$$\begin{aligned} P_{\text{click 2}} &= \text{Tr}((1-d_B)(|0\rangle\langle 0|\otimes\mathbb{1})\,|\Delta\rangle\langle\Delta|\otimes|\gamma\rangle\langle\gamma|) - P_{\text{no click}} \\ &= (1-d_B)\exp(-\eta\mu(1+\cos g)) - (1-d_B)^2\exp(-2\eta\mu) \end{aligned}$$

$$\begin{aligned} P_{\text{dbl click}} = 1 &- (1-d_B)[\exp(-\eta\mu(1+\cos g)) + \exp(-\eta\mu(1-\cos g))] \\ &+ (1-d_B)^2\exp(-2\eta\mu) \end{aligned}$$

Using these, as we need Bob to map double clicks to single clicks with a fair coin to satisfy the squashing model [5], we have our actual probabilities:

$$P_{\text{vac}} = P_{\text{no click}} \qquad P_{D1} = P_{\text{click 1}} + 1/2 P_{\text{dbl click}} \qquad P_{D2} = P_{\text{click 2}} + 1/2 P_{\text{dbl click}}\ .$$

Then if the optical states are mapped to the BB84 states in the same manner proposed in [38],

$$|0\rangle \leftrightarrow \phi = 0 \qquad |1\rangle \leftrightarrow \phi = \pi \qquad |+\rangle \leftrightarrow \phi = \frac{\pi}{2} \qquad |-\rangle \leftrightarrow \phi = \frac{3\pi}{2}\ ,$$

then the Z- basis and $X0$-basis measurements are when Bob choose $\varphi = 0$ and $\varphi = \Pi$ respectively. Lastly, if $d_B, \zeta = 0$, then D_1 clicks if both ϕ and φ equal zero or $\frac{pi}{2}$, and D_2 clicks if $(\phi,\varphi) \in \{(\Pi,0),(\frac{3\pi}{2},\frac{\pi}{2})\}$. Given the squashing model for optical BB84 with threshold detectors in [5], this completes the explanation of simulating the statistics for discrete-phase-randomized BB84.