Tools for the Security Analysis of Quantum Key Distribution in Infinite Dimensions

by

Twesh Upadhyaya

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, 2021

© Twesh Upadhyaya 2021

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

Portions of this thesis are based on the paper *Dimension Reduction in Quantum Key Distribution for Continuous- and Discrete-Variable Protocols* [1], written by the author, and co-authored with Thomas van Himbeeck, Jie Lin, and Norbert Lütkenhaus.

Abstract

We develop a method to connect the infinite-dimensional description of optical continuous-variable quantum key distribution (QKD) protocols to a finite-dimensional formulation. The secure key rates of the optical QKD protocols can then be evaluated using recently-developed reliable numerical methods for key rate calculations. We apply this method to obtain asymptotic key rates for discrete-modulated continuous-variable QKD protocols, which are of practical significance due to their experimental simplicity and potential for large-scale deployment in quantum-secured networks. Importantly, our security proof does not require the photon-number cutoff assumption relied upon in previous works. We also demonstrate that our method can provide practical advantages over the flag-state squasher when applied to discrete-variable protocols.

Acknowledgements

First and foremost I would like to thank my supervisor Prof. Norbert Lütkenhaus for his mentorship and guidance. I greatly appreciate his always making time to discuss research, and I benefitted greatly from his experience and insight. His principled approach to exploring new research questions helped me always keep the big picture in mind.

I would also like to thank Jie Lin and Thomas van Himbeeck. I learned a lot from them and it was a pleasure working together.

Prof. Jon Yard and Prof. Thomas Jennewein served on both my MSc advisory and defence committees, and I thank them for their time and valuable advice.

I am grateful to Prof. John Watrous and Prof. Vern Paulsen for enthusiastically answering my many questions regarding quantum information theory.

Thanks also to Ian George for clarifying many details regarding finite-key analysis, and to Nicky Kai Hong Li for providing the code for the flag-state squasher comparison.

A special thank you to my mom and dad, Vandana and Akhilesh, for their incredible love and support throughout this journey.

Dedication

To my dear grandmother, Mithlesh Upadhyaya.

Table of Contents

List of Figures xi

List of Tables xiv

List of Abbreviations xv

1 Introduction 1

2 Background 4

2.1 Quantum Information Theory 4

2.1.1 Hilbert Space 4

2.1.2 Operators 5

2.1.3 States 7

2.1.4 Multipartite Systems 10

2.1.5 Channels 11

2.1.6 Measurements 12

2.1.7 Distance Measures 14

2.2 Quantum Entropies 15

2.2.1 Shannon Entropy and von Neumann Entropy 15

2.2.2 Quantum Relative Entropy 16

2.2.3 Conditional Entropy 16

2.2.4 Mutual Information . . . 17
2.3 Semidefinite Programming and Convex Optimization . . . 17
2.4 Quantum Optics . . . 18
2.4.1 Simple Harmonic Oscillator . . . 18
2.4.2 Important Classes of States . . . 19
2.4.3 Phase Space and Displacement . . . 20
2.4.4 Beam Splitters . . . 21

3 Quantum Key Distribution **22**
3.1 QKD Protocols . . . 23
3.1.1 Protocol Steps . . . 24
3.1.2 Source-Replacement Scheme . . . 25
3.2 Key Rate and Security Definition . . . 26
3.2.1 QKD Security Definition . . . 26
3.2.2 Key Rate . . . 28
3.2.3 Security Promise and Assumptions . . . 32
3.3 Numerical Optimization Formulation . . . 33
3.3.1 Different Form of Objective Function . . . 34
3.3.2 Primal Reformulation . . . 35
3.3.3 Numerical Imprecision . . . 36
3.3.4 Dual SDP . . . 36
3.3.5 Simulation . . . 36

4 Dimension Reduction Method **38**
4.1 General Result . . . 39
4.1.1 Preliminary Definitions . . . 39
4.1.2 Theorem Statement . . . 41
4.1.3 Procedural Viewpoint . . . 43

4.2 Application to QKD . . . 43
4.2.1 Finite Subspace Π . . . 45
4.2.2 Weight W . . . 45
4.2.3 Correction Term Δ . . . 46
4.2.4 Finite Set $\mathbf{S}_N$ . . . 55

5 Discrete-Modulated Continuous-Variable QKD 61
5.1 Protocol Description . . . 62
5.1.1 Bob's Measurement . . . 63
5.2 Generalized Beam Splitting Attack . . . 66
5.3 Observables . . . 68
5.3.1 Ideal Detector . . . 69
5.3.2 Trusted Noise . . . 70
5.4 Infinite-Dimensional Optimization . . . 74
5.5 Applying Dimension Reduction Method . . . 76
5.5.1 Finite Subspace Π . . . 77
5.5.2 Weight W . . . 78
5.5.3 Correction Term Δ . . . 80
5.5.4 Finite Set $\mathbf{S}_N$ . . . 80
5.6 Simulation Results . . . 82
5.6.1 Understanding the Dimension Reduction Method . . . 83
5.6.2 Key Rates and Optimal Protocol Parameters . . . 85
5.6.3 Trusted Noise . . . 87
5.6.4 Comparison to Gaussian Modulation . . . 87
5.7 Improving Protocol Performance . . . 89
5.7.1 Optimal Postselection . . . 90
5.7.2 Larger Constellations . . . 98

6 Comparison to Flag-State Squasher **105**
6.1 Analytical Comparison 105
6.2 Numerical Comparison: Unbalanced Phase-Encoded BB84 107

7 Conclusion **109**

References **111**

APPENDICES **119**

A Uniform Continuity Bound on Conditional Entropy **120**

B Numerical Framework Formulation **124**
B.1 Explicit SDPs 124
B.2 Modification to Dual Expansion 125
B.3 DMCVQKD Numerical Details 127

C Matrix Operations in Displaced Basis **128**
C.1 Constraints 128
C.2 Ideal and Trusted Noise Region Operators 128
C.3 Channels 131

D Simulated Expectations and Error-Correction Cost **133**

E Experimental Implementation **135**
E.1 Choosing Finite Subspace 135
E.2 Determining Expectations 136
E.3 Overall Formulation 137

List of Figures

2.1 Graphical depiction of beam splitter transforming input modes into output modes. 21

4.1 Pictorial representation of Theorem (1). The set $\mathbf{S}_N$ is chosen to contain the projection $\Pi\tilde{\rho}^{\infty}\Pi$. This auxiliary variable is used to relate $f(\tilde{\rho}^{\infty})$ and $f(\tilde{\rho}^{N})$ 42

4.2 Flowchart indicating the key steps in applying the dimension reduction method. 44

5.1 Phase space regions for the key map, with amplitude and phase postselection, for reverse reconciliation in quadrature phase-shift keying ($M = 4$) DMCVQKD. Bob obtains the measurement result ζ from the heterodyne detector. He maps the outcome to the symbol of the region containing ζ. Δ_a and Δ_p are amplitude and phase postselection parameters, respectively. The $\perp$ region corresponds to signals that are discarded. 63

5.2 Schematic of beam splitter implementing the loss-only Gaussian channel. . 66

5.3 Comparison of key rates and uncorrected values from our dimension reduction method with key rates under the photon number cutoff assumption from Ref. [23]. Results are plotted versus distance with excess noise $\xi = 0.01$, and are in the ideal detector scenario. Postselection parameters and signal state intensities from Ref. [23] are used: $\alpha = 0.6$, $\Delta_p = 0$, and Δ_a is optimized with a coarse-grained search over $[0.5, 0.65]$. The subspace dimension parameter is $N = 20$ and $M = 4$. 84

5.4 The fractional correction term versus the subspace dimension parameter N for different values of excess noise, in the ideal detector scenario. The distance is 15 km, $M = 4$, $\alpha = 0.8$, $\Delta_a = 0.5$, and $\Delta_p = 0$. 85

5.5 Ideal detector secure key rates versus transmission distance, for different values of excess noise ξ, with optimized postselection parameters and signal state intensity. For each point, the better of the two results from $N = 30$ and $N = 40$ is used, with the majority of points from $N = 40$. Postselection can improve the noise tolerance and range of the protocol. The protocol is $M = 4$ phase-shift keying. 86

5.6 Optimal signal state amplitude α_{opt} versus transmission distance, for different values of excess noise ξ and in the ideal detector scenario. The amplitude is optimized in the range $[0.5, 2]$, with $\Delta_a = \Delta_p = 0$ and $N = 10$. 87

5.7 Optimal amplitude postselection parameter $\Delta_{a_{opt}}$ versus transmission distance, for different values of excess noise ξ and in the ideal detector scenario. The amplitude is optimized in the range $[0, 1]$, with $\alpha = \alpha_{opt}$, $\Delta_p = 0$ and $N = 10$. 88

5.8 Key rates versus distance for different trusted detector imperfections, with excess noise $\xi = 0.01$. Protocols are evaluated with the same optimized protocol parameters, including postselection, as in Fig. 5.5. The subspace dimension parameter is $N = 5$ and $M = 4$. 88

5.9 Comparison of Gaussian [82] and discrete-modulation key rates for different values of excess noise ξ, in the trusted noise scenario with $\eta_d = 0.6$, $\nu_{el} = 0.05$. Parameters for both protocols are optimized. The subspace dimension parameter is $N = 5$. The discrete-modulated protocol uses $M = 4$ phase-shift keying. 89

5.10 Heatmap of the difference $[H(Z|E)_{\rho_l} - f_l H(Z|A)_{\rho_l}]$ over phase space. The contour where this quantity is 0 is indicated by the dashed blue line. The optimal postselection region (i.e. the set of discarded outcomes) extends from the origin to this contour. For 4-PSK with coset announcements and realistic error correction, in the ideal detector and loss-only scenarios, at distance 20 km and $\alpha = 0.7$. 96

5.11 Ideal detector uncorrected values for phase-shift keying with $M = 4$ and $M = 8$. Protocols are evaluated with optimized protocol parameters, including postselection. The subspace dimension parameter is $N = 10$. 100

5.12 Coherent signal state constellation for the (5,11) protocol. It consists of 16 signal states, 5 spaced equally on the inner circle with radius α^{in}, and 11 on the outer circle with radius α^{out}. Each inner state is chosen with probability $\frac{p^{in}}{5}$, and each outer one with probability $\frac{1-p^{in}}{11}$. 101

5.13 Phase space regions for the reverse reconciliation key map for the (5,11) protocol. Bob obtains the measurement result ζ from the heterodyne detector. He maps the outcome to the symbol of the region containing ζ. Δ_a^{in} and Δ_a^{out} are amplitude postselection parameters corresponding to the radii of the innermost and outermost circles. r_m is the radius of the middle circle. The two $\perp$ regions correspond to signals that are discarded. 102

5.14 Comparison of Gaussian-modulated [82] CVQKD with the (5,11) DMCVQKD protocol. Key rates are plotted against distance, at excess noise $\xi = 0.01$, in the ideal detector scenario. For Gaussian modulation, the variance is optimized at each distance. For discrete modulation, the protocol parameters are approximately optimized at a distance of 100 km, and the same parameters are then used at all distances. The subspace dimension parameter is $N = 5$. 104

6.1 Key rates for unbalanced phase-encoded BB84 versus transmission efficiency η, for different values of asymmetric interferometer loss $1-\kappa$. It is clear that the key rates from our dimension reduction method are nearly identical to those from the flag-state squasher, indicating the tightness of our method in practice. In generating the data for this graph, our dimension reduction method was approximately five times faster than the flag-state squasher as implemented in Ref. [10]. 108

List of Tables

5.1 Optimal protocol parameters for $M = 4$ phase-shift keying, at distance 150 km and 2% excess noise, with error-correction efficiency $\beta_{EC} = 95\%$. 97

5.2 Optimal protocol parameters for $M = 4$ phase-shift keying, at distance 10 km and 5% excess noise, with error-correction efficiency $\beta_{EC} = 95\%$. 98

5.3 Protocol parameters used in Fig. 5.14 for the (5,11) protocol. Values are approximately optimal at distance 100 km and 1% excess noise. Refer to Fig. 5.12 and Fig. 5.13 for descriptions of all the parameters. 103

List of Abbreviations

APSK Amplitude and Phase-Shift Keying

CV Continuous-Variable

DM Discrete Modulation

DR Dimension Reduction

DV Discrete-Variable

EC Error Correction

FSS Flag-State Squasher

GM Gaussian Modulation

LDPC Low-Density Parity-Check

PA Privacy Amplification

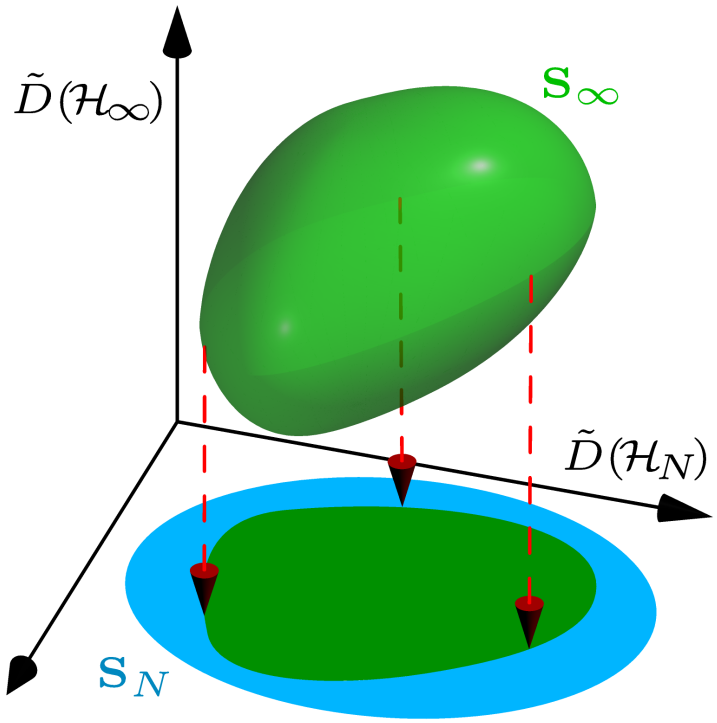
POVM Projective Operator-Valued Measure

PSK Phase-Shift Keying

QBER Quantum Bit Error Rate

QKD Quantum Key Distribution

UCDUP Uniformly Close to Decreasing Under Projection


$\tilde{D}(\mathcal{H}_\infty)$
$\mathbf{S}_\infty$
$\tilde{D}(\mathcal{H}_N)$
$\mathbf{S}_N$

Chapter 1

Introduction

Quantum key distribution (QKD) has emerged as an exciting and practical quantum technology. It enables two remote parties, Alice and Bob, to establish information-theoretically secure keys even in the presence of an eavesdropper, Eve. These keys can then be used in many other cryptographic applications, such as the one-time pad. To prove QKD to be secure with a specified key rate, we need to assume a quantum-mechanical model for Alice and Bob's devices, but do not have to assume anything about the processing power available to the eavesdropper [2].

For a given QKD protocol, the goal of a security proof is to find a lower bound on the secure key rate. Analytical methods for this task can be very involved, tend to be restricted to symmetric protocols, and can introduce looseness in the lower bounds. These issues are ameliorated by the recent development of tight, reliable numerical methods for finding secure key rates [3, 4]. At a high level, these methods determine the key rate by solving a particular convex optimization over the set of quantum states that could be held by Alice and Bob. When the bipartite Hilbert space is infinite dimensional, the numerical methods cannot be used directly.

This is a highly relevant problem because QKD protocols are almost always implemented using quantum optics, which means the quantum state can have any number of photons and is thus infinite dimensional. Fortunately, for many discrete-variable (DV) protocols, the numerical methods can be applied by using the squashing map [5–8] or the more general flag-state squasher [9] to reduce the problem to finite dimensions. However, these squashing approaches do not seem applicable to continuous-variable (CV) protocols. Additionally, even for DV protocols, the flag-state squasher can have challenging runtimes [10].

Discrete-modulated continuous-variable QKD (DMCVQKD) is a family of protocols that utilize existing telecommunication infrastructure, including homodyne or conjugate homodyne detection [11–13]. They are thus promising candidates for deployment in large scale quantum-secured networks. In comparison to Gaussian-modulated CVQKD (GMCVQKD) [14–16], discrete modulation with a small number of signal states is less demanding on the source modulator and on the error-correction protocols, yet is expected to achieve similar key rates. Whille DMCVQKD has clear experimental advantages, its security proofs are challenging.

It is thus of interest to establish tight lower bounds on the secure key rates for DMCVQKD. Of particular interest is DMCVQKD with a reasonably small number of modulated states, say in the range $4 - 16$, which is expected to outperform constellations with just two or three states and essentially achieve the key rates of GMCVQKD, while still maintaining ease of experimental implementation.

There are analytic asymptotic security proofs of DMCVQKD with two [17] or three [18] modulated states. A full finite-key analysis of binary-modulated DMCVQKD has also been recently completed in Ref. [19]. However, these proofs are difficult to generalize to larger numbers of modulated states.

Another class of analytic DMCVQKD security proofs works well for very large numbers of modulated states. Roughly speaking, these proofs work by showing the discrete modulation is close to a Gaussian one. Such an approach has led to analytic security proofs in Ref. [20] and more recently in Ref. [21]. The analytic lower bounds are tight as the number of modulated signal states goes to infinity and the discrete modulation approximates a Gaussian one. Unfortunately, the lower bounds are quite pessimistic for reasonably small numbers of modulated states.

Recent works have numerically studied asymptotic security proofs for DMCVQKD with, in principle, any number of modulated states [22, 23]. However, these approaches assume the state is finite-dimensional, known as the photon-number cutoff assumption. Thus, while these results seem numerically plausible, they do not constitute a rigorous asymptotic security proof, as the cutoff assumption cannot be justified.

In this thesis, we present the *dimension reduction method* [1] to tightly lower bound the key rate of an infinite-dimensional QKD protocol in terms of a finite-dimensional convex optimization. In combination with existing numerical tools for solving finite-dimensional convex optimizations, this enables us to find tight, reliable key rates for general device-dependent QKD protocols in infinite-dimensional Hilbert spaces. Our dimension reduction method can also be applied to study other quantum information tasks, such as entanglement verification [24].

As a result, our method can provide a complete asymptotic security proof for discrete-modulated continuous-variable protocols with any number of modulated states, with tight key rates and without relying on the photon-number cutoff assumption. While our focus in this thesis is on calculating asymptotic key rates, we expect key elements of our method to lift to a finite-key analysis.

Our dimension reduction method also provides an alternative approach to study protocols admitting a flag-state squasher. We consider unbalanced phase-encoded BB84 as an example, and show that our method can have an improved runtime compared to the flag-state squasher, while providing similar results.

The remainder of this thesis is structured as follows. In Chapter 2, we provide a brief theory primer and background on relevant concepts. In Chapter 3, we review the basic principles of quantum key distribution and how the key rate can be formulated as a convex optimization. In Chapter 4 we develop our framework for dimension reduction, in more generality than is needed for QKD. We then specialize our general method to asymptotic key rate calculations. In Chapter 5, we apply our method to calculate key rates for DMCVQKD, and use this to perform a thorough security analysis of the protocol, including modelling postselection and trusted noise. We also explore and evaluate different strategies to improve the key rate. In Chapter 6 we compare our method with the flag-state squasher. Finally, we provide concluding remarks and avenues for future work in Chapter 7. Certain technical details are relegated to the Appendices.

Chapter 2

Background

Quantum mechanics describes phenomena at very small scales. Quantum information combines ideas from quantum mechanics, such as superposition and entanglement, with concepts from information theory, like entropy and algorithms. We first review the basic mathematical framework of quantum information in Sec. 2.1. Our review is based on Refs. [25] and [26]. The former is an excellent reference for finite-dimensional Hilbert spaces, and provides a logical progression to build up the relevant concepts. The latter does the same for infinite-dimensional Hilbert spaces, providing a rigorous and self-contained treatment of functional analysis geared towards quantum information theory. We then touch on relevant aspects of quantum entropies in Sec. 2.2, for which Ref. [27] is a good reference. In Sec. 2.3, we discuss semidefinite programming and convex optimization, which are covered in Refs. [25] and [28]. Finally, we introduce some essentials of quantum optics in Sec. 2.4; a thorough treament is found in Ref. [29].

2.1 Quantum Information Theory

We summarize the building blocks of quantum information, including Hilbert spaces, states, and channels. Our discussion is based on Refs. [26] and [25].

2.1.1 Hilbert Space

A *Hilbert space* is the backdrop for all our quantum information concepts. A Hilbert space is a vector space over the field of complex numbers endowed with some additional structure.

Namely, it has an inner product between any two vectors u, v, denoted by $\langle u|v\rangle$. The inner product is linear, symmetric, and positive-definite.

The inner product induces a norm on the Hilbert space, via $\|v\| = \sqrt{\langle v|v\rangle}$. This is all we need to define a finite-dimensional Hilbert space. For infinite-dimensional Hilbert spaces, there is an additional caveat. This is that the vector space must be complete in the norm induced by the inner product. That is, it must be a Banach space.[1]

An orthonormal set in a Hilbert space is a collection of vectors $\{v_a\}$ satisfying $\langle v_a|v_b\rangle = \delta_{ab}$, where the latter symbol is the Kronecker delta. All Hilbert spaces permit orthonormal bases, conventionally denoted by $\{e_a\}$, which are orthonormal sets that are not the proper subset of any orthonormal set (i.e. they are *maximal*).

2.1.2 Operators

Given two normed vector spaces V and W, consider a linear function, or *operator* T from V to W. The *operator norm* of T is defined as

$$\|T\|_\infty = \sup_{\{x\in V:\|x\|_V\leq T\}} \{\|Tx\|_W\}. \tag{2.1}$$

A function is *bounded* if it has a finite operator norm.

We can now define some important spaces associated to normed linear spaces V, W.

The first is the space of all bounded linear functionals from V to $\mathbb{C}$. This is itself a normed linear space, known as the *dual space* of V, and is denoted by V^d. The dual space is always a Banach space.

For each element h of a Hilbert space $\mathcal{H}$, we can define a bounded linear functional f_h via the inner product

$$f_h(k) = \langle h|k\rangle . \tag{2.2}$$

In fact, it can be shown that every linear functional in $\mathcal{H}^d$ is of this form. The map $h \to f_h$ is a conjugate linear bijection, as for all $k \in \mathcal{H}$,

$$\begin{aligned} f_{ch_1+h_2}(k) &= \langle ch_1 + h_2|k\rangle & (2.3)\\ &= \overline{c}\,\langle h_1|k\rangle + \langle h_2|k\rangle & (2.4)\\ &= \overline{c} f_{h_1}(k) + f_{h_2}(k) & (2.5) \end{aligned}$$

[1]Note this is still a requirement in the finite-dimensional case, but it can be shown that all finite-dimensional normed spaces are necessarily complete.

By the Cauchy-Schwarz inequality, this map is also norm-preserving, i.e. an isometry. Thus, as normed linear spaces, $\mathcal{H}$ and $\mathcal{H}^d$ are conjugate isomorphic. This fact motivates Dirac's bra-ket notation: $|\psi\rangle$ denotes a vector in $\mathcal{H}$ labelled by ψ, while $\langle\phi|$ denotes the linear functional in $\mathcal{H}^d$ which maps an input vector $|\psi\rangle$ to the inner product $\langle\phi|\psi\rangle$. In this notation, $|\psi\rangle\langle\phi|$ is an operator in $B(\mathcal{H})$ sending $|\theta\rangle$ to $\langle\phi|\theta\rangle\,|\psi\rangle$.

We can also consider the space of all bounded operators from V to W. This is denoted by $B(V,W)$, while $B(V,V)$ is denoted simply by $B(V)$. When W is a Banach space, so too is $B(V,W)$. The space of all operators, unbounded or bounded, from V to W is denoted by $\mathcal{L}(V,W)$.

To every operator $T \in B(\mathcal{H})$ we can associate a partner known as its adjoint.[2] The adjoint of T, denoted by $T^\dagger$, is the unique operator satisfying

$$\langle h|Tk\rangle = \left\langle T^\dagger h\middle|k\right\rangle \quad \forall h,k \in \mathcal{H}. \tag{2.6}$$

A bounded operator is defined to be *self-adjoint* or *Hermitian* if it is equal to its adjoint. Defining adjoints for unbounded operators is more involved, and we refer the interested reader to Definition 3.78. in Ref. [26]. We also caution that for unbounded operators, the notions of self-adjointness and Hermiticity are not the same (see Definition 3.82 in Ref. [26]).

$\mathbb{1}_{\mathcal{H}}$ is used to denote the identity operator on $\mathcal{H}$, which maps vectors to themselves.

An operator $U \in B(\mathcal{H},\mathcal{K})$ is an *isometry* if $U^\dagger U = \mathbb{1}_{\mathcal{H}}$. It is a *unitary* if additionally $UU^\dagger = \mathbb{1}_{\mathcal{K}}$.

An important set of bounded operators are the *projections*. Suppose $\mathcal{V}$ is a closed subspace of a Hilbert space $\mathcal{H}$. The perpendicular subspace is defined as

$$\mathcal{V}^\perp = \{h \in \mathcal{H} : \langle h|v\rangle = 0, \forall v \in \mathcal{V}\}. \tag{2.7}$$

Every $h \in \mathcal{H}$ can be uniquely written as $h = v + w$, where $v \in \mathcal{V}$ and $w \in \mathcal{V}^\perp$. The (orthogonal) projection of $\mathcal{H}$ onto $\mathcal{V}$ is the map $\Pi_{\mathcal{V}} : h \to v$. In this thesis, projections will be denoted by Π, and the associated projection onto the perpendicular subspace by $\overline{\Pi}$ (i.e. $\Pi + \overline{\Pi} = \mathbb{1}_{\mathcal{H}}$).

Given an invertible operator $T \in B(\mathcal{H})$, its inverse is denoted by T^{-1}, so that $T^{-1}T = TT^{-1} = \mathbb{1}_{\mathcal{H}}$. Even if an operator is not invertible, one can define the generalized inverse (also known as the Moore-Penrose pseudo-inverse) which inherits some of the properties

[2] To be clear, this is the Hilbert space adjoint. There is also a procedure to take adjoints in all Banach spaces, and the two notions do not agree.

of an inverse. It is denoted by T^g, and satisfies $T^gT = \Pi_{\text{supp}(T)}$ and $TT^g = \Pi_{\overline{\text{im}(T)}}$ (see Theorem 9 and the subsequent corollary of Ref. [30] for proofs). The preceding notation indicates the projectors onto two closed subspaces: the support and the closure of the image of T, respectively. The intuition behind this idea is that when T is restricted to its support, then it is an invertible operator.

2.1.3 States

Given a physical system, its quantum-mechanical state is a description of the system, or of our knowledge of the system, at a given time. We will sometimes use the term register to refer to such a physical system. Mathematically, states are a special subset of bounded operators acting on the Hilbert space of the system. To introduce this class of operators, we need some prerequisites.

Spectrum

The spectrum of an operator can be thought of as a generalization of the concept of eigenvalues. The spectrum of a bounded operator $T \in B(\mathcal{H})$ is

$$\sigma(T) = \{z \in \mathbb{C} : (z\mathbb{1} - T) \text{ does not have an inverse}\}. \tag{2.8}$$

Functional Calculus

We now cover the notion of applying continuous functions to self-adjoint operators. Let $T \in B(\mathcal{H})$ and $T = T^\dagger$. If we have a polynomial $p(x) = a_n x^n + ... + a_0$, then we can define $p(T) = a_n T^n + ... + a_0 \mathbb{1}$, and $p(T) \in B(\mathcal{H})$.

For $f : \sigma(T) \to \mathbb{R}$ a continuous function, there is a sequence of polynomials $\{p_n\}$ converging uniformly to f. The sequence $p_n(T)$ is itself Cauchy, in the operator norm, so we define its limit as $f(T)$. (Recall that a sequence $\{a_n\}$ is Cauchy if, for all $\epsilon > 0$, there exists N such that $\|a_n - a_m\| < \epsilon$ for all $n, m > N$.)

Positive Operators

$T \in B(\mathcal{H})$ is *positive semidefinite* if

$$\langle h|Th\rangle \geq 0 \qquad \forall h \in \mathcal{H}, \tag{2.9}$$

and positive-definite if the above inequality is strict. The notation $T > S$ ($T \geq S$) indicates $T - S$ is positive (semi)definite. In this thesis, the term positive is taken to mean positive semidefinite. We will sometimes use the notation $\mathrm{Pos}\,(\mathcal{H})$ to denote the set of positive operators on a Hilbert space $\mathcal{H}$. Positive operators are necessarily self-adjoint, and have a real and nonnegative spectrum.

Compact Operators

An operator $T \in B(\mathcal{H})$ is defined to be *compact* if the closure of the set

$$\{Th : \|h\| \leq 1\} \tag{2.10}$$

is compact in $\mathcal{H}$. (Recall that a compact set is one where every sequence has a convergent subsequence.) In words, this states that the closure of the image of the unit ball is compact. Defining the rank of an operator as the dimension of its range, the above definition is equivalent to requiring that T can be approximated arbitrarily closely in operator norm by finite-rank operators. By the spectral theorem, compact self-adjoint operators admit an eigendecomposition, with at most countably many eigenvalues. The set of compact operators will be denoted by $\mathbb{K}(\mathcal{H})$.

Trace-Class

For any $T \in B(\mathcal{H})$, $T^\dagger T$ is a manifestly self-adjoint operator with a real and nonnegative spectrum. By the functional calculus, $\sqrt{T^\dagger T}$ is then well-defined and we denote this as the absolute value $|T|$.

Further suppose T is compact, so that $|T|$ is a positive compact operator. It then has a countable set of eigenvalues (counted with multiplicities). The eigenvalues of $|T|$, denoted by $s_i(T)$, are referred to as the *singular values* of T.

We can use this to define an important set of norms. The *Schatten p-norm* of T is

$$\|T\|_p = \left(\sum_n |s_n(T)|^p \right)^{\frac{1}{p}} . \tag{2.11}$$

for $1 \leq p < \infty$. The set of operators with finite p-norm is called the Schatten p-class, denoted by SC_p. As a shorthand, SC_∞ is defined to just be $B(\mathcal{H})$.

SC_1 is also known as the set of *trace-class* operators $TC(\mathcal{H})$. For every trace-class operator T, we can define the *trace* functional via

$$\mathrm{Tr}(T) = \sum \langle e_a | T e_a \rangle \,, \tag{2.12}$$

where the choice of orthonormal basis $\{e_a\}$ does not matter. The trace function is cyclic.

A very useful bound on the trace is given by Hölder's inequality. Let $1 \leq p, q \leq \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$. Such pairs of numbers are called Hölder conjugates. Let $T \in SC_p$ and $S \in SC_q$. Then, TS is a trace-class operator, and

$$|\mathrm{Tr}(TS)| \leq \|T\|_p \|S\|_q. \tag{2.13}$$

Definition

We can pull this all together to mathematically define what a quantum state is. The quantum states, or density operators, on a Hilbert space $\mathcal{H}$ are the positive, trace-class operators with unit trace

$$D(\mathcal{H}) = \{\rho : \rho \in TC(\mathcal{H}), \mathrm{Tr}(\rho) = 1, \rho \geq 0\}. \tag{2.14}$$

We will sometimes deal with the subnormalized states $\tilde{D}(\mathcal{H})$, which are defined the same as above, but with $\mathrm{Tr}(\rho) \leq 1$. It is useful to note that $TC(\mathcal{H})$ is spanned by elements of $D(\mathcal{H})$.

The simplest states are *pure states*. A pure state is a density operator that can be written as $\rho = |\psi\rangle\langle\psi|$, for some unit vector $|\psi\rangle$. We will sometimes just refer to $|\psi\rangle$ as the state. The most general class of states are *mixed states*. Intuitively, these can be thought of as a classical mixture or convex combination of pure states.

Once we have chosen a preferred basis, say $\{|i\rangle\}$, the state of a classical system can be captured in the density operator formulation as a diagonal operator

$$\rho_{classical} = \sum_i p(i) \, |i\rangle\langle i| \,. \tag{2.15}$$

A classical-quantum (CQ) state is a bipartite state with the following form (see Sec 2.1.4 for an explanation of the tensor product)

$$\rho_{CQ} = \sum_i p(i) \, |i\rangle\langle i| \otimes \sigma_i. \tag{2.16}$$

We refer to the σ_i as *conditional states*.

2.1.4 Multipartite Systems

Our discussion covers bipartite systems, and inductively extends to multipartite systems $A_1...A_n$ formed from a finite set of registers. We first clarify the notation used in the rest of this thesis. Given a register R, the associated Hilbert space is denoted by $\mathcal{H}_R$. For classical registers, the alphabet, which is the set of values the register can take, is denoted by S_R.

In order to understand bipartite systems, we must introduce the tensor product. The tensor product space $V \otimes W$ is the unique vector space of dimension $|V||W|$ spanned by *elementary tensors* of the form $v \otimes w$, for $v \in V$ and $w \in W$. The tensor map $\otimes$ is a bilinear mapping.

Suppose we have two registers, A and B. The Hilbert space describing the composite register AB, is given by the tensor product[3] $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. If $\{e_a\}$ and $\{f_b\}$ are orthonormal bases for $\mathcal{H}_A$ and $\mathcal{H}_B$, then the elementary tensors $\{e_a \otimes f_b\}$ form an orthonormal basis for $\mathcal{H}_{AB}$.

Given operators $T \in B(\mathcal{H}_A)$ and $S \in B(\mathcal{H}_B)$, the operator $T \otimes S$ is defined on elementary tensors by $(T \otimes S)(v \otimes w) = Tv \otimes Sw$, and extends to all vectors by linearity.

Given a bipartite state ρ_{AB}, ρ_A represents the *reduced state* on the first register, and is obtained by tracing out the second register $\rho_A = \mathrm{Tr}_B(\rho_{AB})$ (see Sec. 2.1.5). This operation is called the partial trace, and the subscripts indicate the systems being traced out.

Given a quantum state $\rho \in D(\mathcal{H}_1)$, it is always possible to find a *purification* of ρ in a larger Hilbert space. There exists $|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$, where $\mathcal{H}_2 = \mathcal{H}_1$, such that $\mathrm{Tr}_2(|\psi\rangle\langle\psi|) = \rho$. This purification is unique, up to a unitary on the second system.

Entanglement

An important property of bipartite states is *entanglement*. A state ρ_{AB} is *separable* if it can be expressed as

$$\rho_{AB} = \sum_i X_i \otimes Y_i \qquad X_i, Y_i \geq 0 \ \ \forall i \tag{2.17}$$

A state with exactly one term in the above expansion is called a *product state*. A state is entangled if it is not separable. Entanglement is an important nonclassical correlation.

[3] More precisely, the metric space completion of the algebraic tensor product.

2.1.5 Channels

While density operators describe the state of a physical system, *quantum channels* describe processes that change the state of a system. Channels are linear maps which satisfy some additional postulates.

Definition

Mathematically, a channel is a completely positive and trace-preserving (CPTP) linear map Φ from $TC(\mathcal{H}_X)$ to $TC(\mathcal{H}_Y)$. As the name suggests, Φ is trace-preserving if

$$\mathrm{Tr}(\Phi(X)) = \mathrm{Tr}(X) \qquad \forall X \in TC(\mathcal{H}_X). \tag{2.18}$$

Φ is positive if

$$\Phi(P) \geq 0 \qquad \forall P \in TC(\mathcal{H}_X), P \geq 0. \tag{2.19}$$

This ensures the channel maps positive operators to positive operators. Complete positivity is a stronger condition, ensuring that when the channel acts on a subsystem of a composite system, the composite output state is positive. To define this notion precisely, we need the tensor product of two linear maps. Analogously to tensor products of operators, the tensor product of Φ and Ψ is defined on product states by $(\Phi \otimes \Psi)(\rho \otimes \sigma) = \Phi(\rho) \otimes \Psi(\sigma)$, and extends to all density operators by linearity.

Stated precisely, Φ is completely positive if $\Phi \otimes \mathbb{1}_n$ is positive for all n, where $\mathbb{1}_n$ is the identity channel on $\mathcal{H} = \mathbb{C}^n$. Note we will use the notation $\mathbb{1}$ for the identity quantum channel. It should be clear from context whether $\mathbb{1}$ denotes the identity operator or the identity channel.

Representations and Properties

There are different but equivalent representations of quantum channels; two that we will use are the Kraus and Stinespring representations.

The Kraus representation of a channel is

$$\phi(X) = \sum_i K_i X {K_i}^\dagger, \tag{2.20}$$

where $K_i \in B(\mathcal{H}_X, \mathcal{H}_Y)$ are known as *Kraus operators*. The sum $\sum {K_i}^\dagger K_i$ converges to $\mathbb{1}_{\mathcal{H}_X}$ in the strong topology.

Given channels $\phi_A : TC(\mathcal{H}_X^A) \to TC(\mathcal{H}_Y^A)$ and $\phi_B : TC(\mathcal{H}_X^B) \to TC(\mathcal{H}_Y^B)$, with Kraus operators $\{K_i^A\}$ and $\{K_j^B\}$, the composite channel $\phi_A \otimes \phi_B : TC(\mathcal{H}_X^A \otimes \mathcal{H}_X^B) \to TC(\mathcal{H}_Y^A \otimes \mathcal{H}_Y^B)$ is given by the Kraus operators $\{K_i^A \otimes K_j^B\}$.

The trace is an important example of a quantum channel, $\mathrm{Tr} : \mathcal{H} \to \mathbb{C}$, as is the partial trace, $\mathbb{1}_A \otimes \mathrm{Tr}_B : \mathcal{H}_A \otimes \mathcal{H}_B \to \mathcal{H}_A$.[4]

In the Stinespring representation, any channel can be realized isometrically in a larger Hilbert space. That is, for any channel ϕ there exists a Hilbert space $\mathcal{H}_E$ and an isometry $V \in B(\mathcal{H}_X, \mathcal{H}_Y \otimes \mathcal{H}_E)$ such that

$$\phi(X) = \mathrm{Tr}_E(VXV^\dagger). \tag{2.21}$$

The system E is customarily thought of as the environment (or Eve, as we will see later).

If the output of a quantum channel is classical on any input state, then the channel is called *quantum-to-classical*. We will see that such channels can always be realized by a measurement (see Eq. (2.24)).

Every linear map $\phi : TC(\mathcal{H}_X) \to TC(\mathcal{H}_Y)$ has an adjoint $\phi^\dagger : B(\mathcal{H}_Y) \to B(\mathcal{H}_X)$, which is the unique linear map such that

$$\mathrm{Tr}(M\phi(\rho)) = \mathrm{Tr}\big(\phi^\dagger(M)\rho\big) \qquad \forall \rho \in D(\mathcal{H}_X),\ M \in B(\mathcal{H}_Y). \tag{2.22}$$

If the channel ϕ is given by the Kraus operators K_i, then $\phi^\dagger$ is given by the Kraus operators ${K_i}^\dagger$, and is thus a completely positive and unital map (unital means $\phi^\dagger(\mathbb{1}_{\mathcal{H}_Y}) = \mathbb{1}_{\mathcal{H}_X}$).

2.1.6 Measurements

A measurement device in quantum mechanics can be thought of as a instrument which acts on a quantum state, and returns classical outcomes with some probability distribution. For the purposes of this thesis, we will only be concerned with so-called destructive measurements, where the system is assumed to be discarded after measurement.

Projective Operator-Valued Measures

The mathematical description of measurements is in terms of *projective operator-valued measures*, or POVMs for short. In the following discussion S is defined to be the set of all possible measurement outcomes.

[4] We implicitly use the fact that $\mathcal{H} \otimes \mathbb{C} = \mathcal{H}$.

To define a POVM in full generality, we first need the notion of a σ-algebra. A σ-algebra $\mathcal{A}$ is a collection of subsets of S, satisfying:

1. $S \in \mathcal{A}$.
2. $X \in \mathcal{A} \implies S \backslash X \in \mathcal{A}$.
3. $X_i \in \mathcal{A} \implies \bigcup_i X_i \in \mathcal{A}$.

A POVM is a map $P : \mathcal{A} \to B(\mathcal{H})$ satisfying:

1. $P(X) \geq 0 \quad \forall X \in \mathcal{A}$.
2. $P(S) = \mathbb{1}_{\mathcal{H}}$.
3. $\sum_i P(X_i) \xrightarrow{w} P\left(\bigcup_i X_i\right)$ for all sequences of disjoint sets $\{X_i\}$, where $\xrightarrow{w}$ denotes weak convergence.

The probability of getting an outcome in X is

$$p(X) = \mathrm{Tr}(\rho P(X)). \tag{2.23}$$

The probabilities are always well-defined by Hölder's inequality (Eq. (2.13)). We will generally write $P_X \equiv P(X)$. We refer to the P_X as POVM elements, or POVMs for short.

Often, we will consider a finite set S of measurement outcomes. In this case, there are finitely many POVM elements, indexed discretely, and one does not need the framework of σ-algebras. However, for heterodyne measurements (see Sec. 2.4.2), $S = \mathbb{C}$, so we do need to consider the more formal definition of a POVM.

One important fact that we will make use of (see Sec. 4.2.3) is that every quantum-to-classical channel can be realized by a POVM,

$$\phi_{Q\to C}(X) = \sum_i \mathrm{Tr}(X P_i) \left|i\right\rangle\!\left\langle i\right| . \tag{2.24}$$

Observables

Another way to describe a measurement is in terms of observables. An observable is a self-adjoint operator $\hat{O} \in \mathcal{L}(\mathcal{H})$. The expected value of a measurement of the observable

is then $\mathrm{Tr}\left(\rho\hat{O}\right)$. The shorthand $\left\langle\hat{O}\right\rangle$ is often used to denote the expectation value of the observable, where the underlying state is implicit.

If $\hat{O}$ is unbounded, $\left\langle\hat{O}\right\rangle$ may not be defined. This is not paradoxical since discussing observables really only makes sense in terms of expectation values. To gain a better understanding of this situation, consider an observable expressed as

$$\hat{O} = \sum_i w_i P_i. \tag{2.25}$$

for a POVM $\{P_i\}$. Physically, the POVM is the complete description of a measurement apparatus in quantum mechanics. Now, suppose that on outcome i, the measurement apparatus, instead of outputting the outcome i, instead outputs the real-valued weight w_i. On any particular measurement, some value w_i is obtained, while the average weight is $\mathrm{Tr}\left(\rho\hat{O}\right)$. That $\left\langle\hat{O}\right\rangle$ is undefined only means that over a large number of repetitions, the average will not converge.

2.1.7 Distance Measures

Distance measures quantify the notion that two quantum states are more or less different. We have already covered the trace norm above, and the trace distance is simply defined as one half of the trace norm

$$\frac{1}{2}\|\tilde{\rho} - \tilde{\sigma}\|_1. \tag{2.26}$$

The fidelity is another measure of the closeness of states, defined as

$$F(\tilde{\rho}, \tilde{\sigma}) = \mathrm{Tr}\left(\sqrt{\sqrt{\tilde{\sigma}}\tilde{\rho}\sqrt{\tilde{\sigma}}}\right). \tag{2.27}$$

The Fuchs-van de Graaf inequalities [31] relate fidelity and trace distance as follows,

$$1 - F(\tilde{\rho}, \tilde{\sigma}) \leq \frac{1}{2}\|\tilde{\rho} - \tilde{\sigma}\|_1 \leq \sqrt{1 - F(\tilde{\rho}, \tilde{\sigma})^2}. \tag{2.28}$$

The trace distance between a state and its projection can be bounded in the following manner, which is tighter than the result one would get just from applying the Fuchs-van de Graaf inequalities,

$$\frac{1}{2}\|\rho - \Pi\rho\Pi\|_1 \leq \sqrt{\mathrm{Tr}\left(\rho\overline{\Pi}\right)}. \tag{2.29}$$

The proof of this fact can be found in Lemma 5 of Ref. [32], which is a slight tightening of the gentle measurement lemma in Ref. [33].

It is impossible to perfectly distinguish two non-orthogonal quantum states. More generally, any physical process can only bring two states closer together. This is captured by the fact that trace distance is monotonically decreasing under any CPTNI map Φ,

$$\frac{1}{2}\|\tilde{\rho} - \tilde{\sigma}\|_1 \geq \frac{1}{2}\|\Phi(\tilde{\rho}) - \Phi(\tilde{\sigma})\|_1. \tag{2.30}$$

2.2 Quantum Entropies

There are many entropic quantities which are of theoretical interest or quantify the resources required for different tasks in information theory. All such quantities can be viewed as generalizations of a corresponding classical one. The classical one can be recovered by considering classical input states only, so in the following we generally only give the definitions for the quantum ones.

2.2.1 Shannon Entropy and von Neumann Entropy

Entropy is a central concept in information theory. It provides a way to quantify randomness or uncertainty. For a classical discrete probability distribution $p(i)$, the Shannon entropy is defined as

$$H(p) = \sum_i p(i) \log_2 \frac{1}{p(i)}. \tag{2.31}$$

The base of the logarithm in the above expression determines the units. We will exclusively use base-2 logarithms in this thesis, so that the entropy is measured in bits. We will denote the binary entropy by $h(x) \equiv -x \log x - (1-x)\log(1-x)$.

The von Neumann entropy of a quantum state is defined in an analogous manner,

$$H(\rho_X) = H(X)_{\rho_X} = -\operatorname{Tr}(\rho_X \log \rho_X) = \sum_i \lambda(i) \log \frac{1}{\lambda(i)}, \tag{2.32}$$

where $\lambda(i)$ are the eigenvalues, with multiplicities, of ρ_X. Written in this form, it is clear that the von Neumann entropy is the Shannon entropy evaluated on the probability distribution $\lambda(i)$.

A brief remark regarding notation: given a multipartite state $\rho_{X_1X_2\ldots X_n}$, we will either write $H(X_1X_2..X_k)_\rho$ or $H(\rho_{X_1X_2\ldots X_k})$ to denote the entropy of the reduced state $\rho_{X_1X_2\ldots X_k}$.[5]

The von Neumann entropy is strongly subadditive, meaning that for any state ρ_{XYZ},

$$H(XYZ) + H(Z) \leq H(XZ) + H(YZ). \tag{2.33}$$

2.2.2 Quantum Relative Entropy

The quantum relative entropy, known in the classical case as the Kullback-Leibler divergence, is a useful intermediate quantity for defining other entropies. It captures the notion of how different two quantum states are, and is defined as

$$D(\rho||\sigma) = \begin{cases} \mathrm{Tr}(\rho\log\rho - \rho\log\sigma) & \text{if } \mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma) \\ \infty & \text{otherwise.} \end{cases} \tag{2.34}$$

where supp denotes the support.

2.2.3 Conditional Entropy

The conditional entropy is the entropy of a system X, given that one has some side information Y,

$$H(X|Y)_{\rho_{XY}} = H(XY)_{\rho_{XY}} - H(Y)_{\rho_Y}. \tag{2.35}$$

The conditional entropy is scalar-linear, so it can be defined on subnormalized states as

$$H(X|Y)_{\tilde{\rho}_{XY}} = \mathrm{Tr}(\tilde{\rho}_{XY})H(X|Y)_{\rho_{XY}}. \tag{2.36}$$

A useful characterization of the conditional entropy is in terms of the relative entropy

$$H(X|Y)_{\rho_{XY}} = -\min_{\tau_Y} D(\rho_{XY}||\mathbb{1}_X \otimes \tau_Y), \tag{2.37}$$

where the minimum is attained at $\tau_Y = \rho_Y$.

[5] In the first notation, if the state is understood from context, it may not be written explicitly as a subscript. Conversely, in the second notation, the registers may not be written explicitly.

2.2.4 Mutual Information

The mutual information measures the correlation between two registers X and Y,

$$I(X:Y)_{\rho_{XY}} = H(X)_{\rho_X} + H(Y)_{\rho_Y} - H(XY)_{\rho_{XY}} = H(X)_{\rho_X} - H(X|Y)_{\rho_{XY}}. \tag{2.38}$$

The mutual information evaluated on a classical-quantum state is also known as the Holevo quantity, and can be written as

$$I(X:Y)_{\rho_{XY}} = \chi(X|Y)_{\rho_{XY}} = H\left(\sum p(i)\sigma^i\right) - \sum p(i)H(\sigma^i) \tag{2.39}$$

where $\rho_{XY} = \sum_i p(i) \left|i\right\rangle\!\left\langle i\right|_X \otimes \sigma_Y^i$.

2.3 Semidefinite Programming and Convex Optimization

Semidefinite programs (SDPs) are useful tools in a variety of fields, and have been increasingly applied to quantum information in recent years. Convex optimization is a more general notion, which encompasses many classes of interesting problems. Both types of optimizations are particularly useful because they can generally be efficiently solved. This section is based on Ref. [25], which covers SDPs in finite dimensions, and Ref. [34], which has additional details regarding the formulation of SDPs in infinite dimensions.

Let $A \in B(\mathcal{H}_X)$ be a bounded self-adjoint operator, on a potentially infinite-dimensional Hilbert space $\mathcal{H}_X$. Let $B \in TC(\mathcal{H}_Y)$ be a trace-class self-adjoint operator on a finite-dimensional Hilbert space $\mathcal{H}_Y$. Let $\Phi : TC(\mathcal{H}_X) \to TC(\mathcal{H}_Y)$ be a linear map which preserves self-adjointness. Recall that Φ has an adjoint $\Phi^\dagger : B(\mathcal{H}_Y) \to B(\mathcal{H}_X)$, which is the unique linear map satisfying Eq. (2.22).

A semidefinite program is then the pair of convex optimization problems:

Primal Form

$$\begin{aligned} \underset{X}{\text{minimize:}} \quad & \operatorname{Tr}(AX) \\ \text{subject to:} \quad & \Phi(X) \leq B \\ & X \in \operatorname{Pos}(\mathcal{H}_X) \\ & X \in TC(\mathcal{H}_X) \end{aligned} \tag{2.40}$$

Dual Form

$$\begin{aligned} \underset{Y}{\text{maximize:}} \quad & -\operatorname{Tr}(BY) \\ \text{subject to:} \quad & \Phi^\dagger(Y) \geq -A \\ & Y \in \operatorname{Pos}(\mathcal{H}_Y) \\ & Y \in B(\mathcal{H}_Y) \end{aligned} \tag{2.41}$$

Let α and β denote the solutions of the primal and dual respectively. Semidefinite programs obey *weak duality*, which means that $\alpha \geq \beta$. *Strong duality* means $\alpha = \beta$. Slater's theorem provides sufficient conditions for strong duality to hold. Namely, if the dual problem is feasible, and there is a strictly feasible primal solution, i.e. $X > 0$ and $\Phi(X) \leq B$, then strong duality holds [25, 34].

More generally, one can consider convex optimizations, where the objective function and feasible set are both convex. Such optimizations arise in many cases of interest, and can often be efficiently solved numerically.

2.4 Quantum Optics

Quantum key distribution is implemented overwhelmingly using optics. To model QKD systems, it is therefore important to understand some quantum optics. For the purposes of this thesis, it suffices to understand the quantization of single modes of light, the phase space representation, and some basic optical elements. We provide a brief overview of these concepts in this section; our discussion is based on Ref. [29].

2.4.1 Simple Harmonic Oscillator

Given a physical setup, such as a cavity or an optical fiber, consider a particular solution to Maxwell's equations at a given frequency ω. This is known as a *mode*. Following the procedure of canonical quantization, the electric and magnetic fields can be promoted to operators. It turns out that, mathematically, such a single mode of light is exactly equivalent to a quantum simple harmonic oscillator (QSHO). We briefly review this formalism, drawing parallels between the terminology for quantized light and a QSHO. A thorough description of the QSHO can be found in Ref. [35].

A basis for the Hilbert space is formed by the *Fock states* $\{|n\rangle\}_{n=0}^{\infty}$, where n corresponds to the number of photons in the state. The *ladder operators* $\hat{a}^\dagger$ and $\hat{a}$ create or annihilate a photon of energy $\hbar\omega$. That is,

$$\hat{a}\,|n\rangle = \sqrt{n}\,|n-1\rangle\,, \tag{2.42}$$

$$\hat{a}^\dagger\,|n\rangle = \sqrt{n+1}\,|n+1\rangle\,, \tag{2.43}$$

with the exception that $\hat{a}\,|0\rangle$ is the zero vector. $|0\rangle$ is referred to as the *vacuum state*. From these relations, it can be seen that the Fock states are eigenstates of the *photon number operator* $\hat{n} \equiv \hat{a}^\dagger\hat{a}$. The ladder operators obey the commutation relation $[\hat{a}, \hat{a}^\dagger] = 1$.

The *quadrature* operators $\hat{X}$ and $\hat{P}$, analogous to position and momentum, are dimensionless operators proportional to the electric and magnetic field operators. (Note that units can always be restored using dimensional analysis.) There are two different conventions to define the quadratures, *natural units* (NU) and *shot noise units* (SNU). In this thesis, we use SNU unless otherwise mentioned. For SNU, the quadratures are defined by the ladder operators via

$$\hat{X} = \frac{1}{\sqrt{2}}(\hat{a}^\dagger + \hat{a}), \tag{2.44}$$

$$\hat{P} = \frac{i}{\sqrt{2}}(\hat{a}^\dagger - \hat{a}). \tag{2.45}$$

Shot noise units correspond to setting $\hbar = 1$. In natural units, the definitions of the quadratures are the same except there is no prefactor of $\frac{1}{\sqrt{2}}$. This corresponds to setting $\hbar = 2$.

2.4.2 Important Classes of States

Coherent States

A particularly important class of states are the *coherent states* $\{|\alpha\rangle\}_{\alpha\in\mathbb{C}}$. They are the eigenstates of the lowering operator,

$$\hat{a}\,|\alpha\rangle = \alpha\,|\alpha\rangle\,. \tag{2.46}$$

The coherent state $|0\rangle$ is just the vacuum state.

These states are easy to create experimentally, as the output of a laser is a coherent state. Coherent states form an overcomplete basis for the Hilbert space. A measurement with respect to this basis is called a *heterodyne measurement*, and is described by the POVM $\{\frac{1}{\pi}\,|\alpha\rangle\langle\alpha|\}_{\alpha\in\mathbb{C}}$. Note that $\frac{1}{\pi}\int_{\alpha\in\mathbb{C}}|\alpha\rangle\langle\alpha|\,d^2\alpha = \mathbb{1}$.

A coherent state can be represented as a Poisson distribution of Fock states

$$|\alpha\rangle = e^{\frac{-|\alpha|^2}{2}}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}\,|n\rangle\,. \tag{2.47}$$

From this representation follows the well-known formula for the overlap of two coherent states,

$$\langle\alpha_j|\alpha_i\rangle = \exp\left(i\,\mathrm{Im}(\alpha_i\alpha_j^*) - \frac{1}{2}|\alpha_i - \alpha_j|^2\right) \tag{2.48}$$

(see Eq. 3.61 of Ref. [29] for a full derivation).

Gaussian States

Gaussian states are those states whose Husimi Q-function (see Sec. 2.4.3) is a two-dimensional Gaussian distribution over the complex plane. As Gaussian distributions are completely characterized by their first and second moments, Gaussian states have a finite-dimensional representation in terms of their covariance matrix. This makes them relatively easy to handle theoretically. A thorough reference on Gaussian states and Gaussian operations is given in Ref. [36].

Thermal States

A thermal state represents a state in thermal equilibrium, and is defined as

$$\rho_{th}(\bar{n}) = \frac{1}{1+\bar{n}} \sum_{n=0}^{\infty} \left(\frac{\bar{n}}{1+\bar{n}} \right)^n |n\rangle\langle n| \,, \tag{2.49}$$

where $\bar{n}$ is the mean photon number. Note $\mathrm{Tr}(\rho_{th}\hat{n}) = \bar{n}$ and $\mathrm{Tr}(\rho_{th}\hat{n}^2) = \bar{n}(1+2\bar{n})$.

2.4.3 Phase Space and Displacement

Single-mode states can be visualized in a two-dimensional *phase space*. Unlike classical mechanics, where the states are points in phase space, in quantum mechanics the states are quasiprobability distributions over phase space. One particular quasiprobability distribution we will make use of is the *Husimi Q-function* [37], defined as

$$Q(\alpha) = \frac{1}{\pi} \langle \alpha | \rho | \alpha \rangle \,. \tag{2.50}$$

An important unitary operator is the *displacement operator*, defined as

$$\hat{D}(\gamma) = \exp\left(\gamma \hat{a}^\dagger - \overline{\gamma}\hat{a}\right). \tag{2.51}$$

As the name suggests, this operator displaces states in phase space. In particular, its action on coherent states is $\hat{D}(\gamma)|\alpha\rangle = e^{i\,\mathrm{Im}(\gamma\overline{\alpha})}|\alpha+\gamma\rangle$. Given a complex number β, we will use the shorthand $\hat{O}_\beta \equiv \hat{D}(\beta)\hat{O}\hat{D}^\dagger(\beta)$ and $|n_\beta\rangle \equiv \hat{D}(\beta)|n\rangle$.

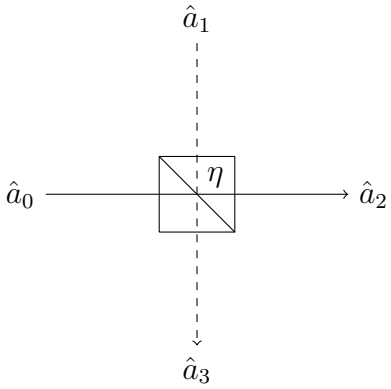


Figure 2.1: Graphical depiction of beam splitter transforming input modes into output modes.

2.4.4 Beam Splitters

There are a variety of optical elements which can change the state of light. When modelling these components, it is convenient to work in the Heisenberg picture. In particular, we assume the state is always the vacuum state, and view the ladder operators as being modified. We briefly summarize a symmetric beam splitter here, but a fuller discussion can be found in Ref. [29]. A beam splitter is a two-input, two-output device. The two input modes are labelled $0, 1$, while the two output modes are labelled $2, 3$ (Fig. 2.1).

A symmetric beam splitter is characterized by its transmittance η, and transforms the ladder operators as

$$\begin{pmatrix} \hat{a}_2 \\ \hat{a}_3 \end{pmatrix} = \begin{pmatrix} \sqrt{\eta} & \sqrt{1-\eta} \\ -\sqrt{1-\eta} & \sqrt{\eta} \end{pmatrix} \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \end{pmatrix}. \tag{2.52}$$

To use this formula, we first express the input state in terms of ladder operators acting on the vacuum. Then, replacing the input ladder operators with output ladder operators according to Eq. (2.52) gives the output state.

Chapter 3

Quantum Key Distribution

Consider the following scenario. Alice and Bob are two honest parties who are remotely separated. They have access to an untrusted quantum channel and an authenticated public classical channel.[1] The quantum channel is entirely under the control of an eavesdropper, Eve. Eve is also able to view (but not tamper with) all messages sent on the classical channel.

Remarkably, in this scenario, it is possible for Alice and Bob to establish a shared, secret key using quantum key distribution (QKD). Broadly speaking, QKD is possible due to two key principles of quantum mechanics: the *no-cloning theorem* [38] and *monogamy of entanglement* [39]. The first implies that when Eve tries to distinguish non-orthogonal states, she will necessarily introduce errors that Alice and Bob can catch and use to determine how much Eve tampers with the channel. The second guarantees that, once Alice and Bob establish they have a highly entangled state, Eve cannot be strongly correlated to their data.

If limited to classical systems only, it is not possible to generate a key in this scenario. If, however, one puts restrictions on the eavesdropper, such as limited computational power, then it is possible to generate secret keys in the classical scenario as well; using techniques such as Diffie-Hellmann key exchange [40]. Thus, one key advantage of QKD over similar classical protocols is that the eavesdropper is limited only by the laws of quantum mechanics [2]. This means keys generated via QKD are information-theoretically secure,

[1] To authenticate their messages over an untrusted, public classical channel, Alice and Bob can use some shared secret bits along with a classical message authentication procedure. In this sense, QKD is more properly referred to as quantum key *growing*, as the parties need to start out with some key bits for authentication on the first run of QKD. In future runs, they can use a portion of the key already generated.

instead of just computationally secure. The second key advantage of QKD over its classical counterparts is *perfect forward secrecy*, meaning that keys generated via QKD are as secure in the future as when they were generated. That is, the discovery of new attacks on or weaknesses in the protocol will not compromise the security of already-generated keys. See Ref. [41] for a nuanced discussion of the security guarantees of QKD, and its potential role in a larger quantum network.

The first ideas on using quantum mechanics for cryptography were introduced by Wiesner in the 1960s [42]. The first complete QKD protocol was then developed by Bennett and Brassard in 1984, and is referred to as BB84 [43]. This protocol relies on single-photon threshold detectors, and is hence classified as a *discrete-variable* (DV) protocol. In 1999, Ralph introduced the first *continuous-variable* (CV) protocol [11]. CV protocols are classed as such because the measurements involve homodyne or heterodyne detection. Due to their simpler detection setup, CV protocols are easier to implement experimentally. However, their security analysis has lagged behind that of DV protocols. Reviews of QKD protocols and security analyses can be found in Refs. [44–46].

Among technologies harnessing quantum information, QKD enjoys relatively widespread experimental implementation, including over long distances [47], with space-based satellites [48], and in commercial plug-and-play products [49].

In this chapter, we first review the steps of a QKD protocol and the entanglement-based and prepare-and-measure formulations in Sec. 3.1. We then briefly outline the security definition of QKD, and the corresponding formula for the secure key rate in Sec. 3.2. Finally, we summarize the existing numerics framework for key rate calculations in finite-dimensional Hilbert spaces in Sec. 3.3.

3.1 QKD Protocols

A *protocol* is a specific procedure that Alice and Bob follow to generate their secret key. It consists of a quantum phase, where quantum states are sent and measured, followed by a classical phase, where the measurement data is postprocessed to correct errors and remove Eve's correlations. There can be different variations of protocols, such as where Alice and Bob both send signals to a third party, Charlie, who then does all the measurements. Here, we consider protocols where only Alice and Bob perform measurements. We consider device-dependent protocols, where the devices of Alice and Bob are assumed to be trusted and fully characterized (see Sec. 3.2.3).

3.1.1 Protocol Steps

An *entanglement-based* (EB) protocol is one where Alice and Bob each receive one part of a bipartite state and perform measurements on it.[2] The steps of a generic entanglement-based protocol are as follows.

1. Alice and Bob start with a bipartite quantum state $\rho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$. This state may be established with the help of an untrusted third party.

2. Alice and Bob measure their subsystems with POVMs $\{P_A^i\}$ and $\{P_B^j\}$.[3] The alphabets for i and j are S_A^{msmt} and S_B^{msmt}. To each outcome i, j, they associate two pieces of classical data: a public announcement a_i, b_j and a private measurement result α_i, β_j. The respective alphabets from which the values are drawn are $S_A^{pub}, S_B^{pub}, S_A^{priv}, S_B^{priv}$.

After repeating the previous two steps for a large number of rounds N, Alice and Bob proceed to the classical phase.

3. Alice and Bob choose a random subset of m rounds to use for *parameter estimation*. For these testing rounds, they announce their measurement outcomes. This allows them to determine the frequency f of each possible outcome (i, j). (The frequency is simply the count of the number of times the joint outcome (i, j) is obtained, divided by m.)

4. Based on the frequencies of each joint outcome $f(i, j)$ from parameter estimation, Alice and Bob decide whether or not to abort the protocol. This is represented by a binary function $\mathcal{A} : \mathcal{F} \to \{0, 1\}$; where $\mathcal{A} = 0$ if the protocol is continued and $\mathcal{A} = 1$ if aborted. Here, $\mathcal{F}$ is the space of all possible frequency distributions on $S_A^{msmt} \times S_B^{msmt}$.

We now proceed with the steps if the protocol is not aborted. All following discussion concerns the data from the remaining $N - m$ rounds only.

5. Alice and Bob announce their public data. Based on the joint announcements, they may decide to discard some rounds. This is referred to as *postselection*. It is represented by a binary function $d : S_A^{pub} \times S_B^{pub} \to \{0, 1\}$; where $d = 0$ if the signal is kept and $d = 1$ if discarded.

[2] As we will see in the next section, this encapsulates prepare-and-measure protocols as well.

[3] We use discrete indices for simplicity, but the POVM can more generally be indexed by a σ-algebra (see Sec. 2.1.6).

6. Based on the public announcements and their own private data, one party performs the *key map*.[4] When Alice (Bob) performs the key map, it is conventionally referred to as direct (reverse) reconciliation. In the following discussion, we consider the case where Bob performs the key map. For a M-ary key, the key map is a function[5] $g : S_A^{pub} \times S_B^{pub} \times S_B^{priv} \to \{0, 1, ..., M-1, \perp\} \equiv S_Z$. The $\perp$ symbol is only used to flag the discarded or *sifted* signals, so $g(a, b, \beta) = \perp \iff d(a, b) = 1$. This models postselection (see above step) in a unified manner as part of the key map.

7. Alice and Bob then perform error correction (EC) using the classical channel to get Alice's data to agree with the sifted key established by Bob. In the process, they will reveal some information over the public classical channel.

8. Alice and Bob perform privacy amplification (PA). To do this, they first determine how long their final secret key can be while still being secure.[6] Then, they choose a two-universal hash function from the appropriate family, communicate this choice publicly, and apply the function to obtain the final shared secret key, of length l.[7]

In practice, the discarded signals are simply removed before performing error correction and privacy amplification. We include them with the discard flag $\perp$ only to formulate the protocol in a trace-preserving manner.

3.1.2 Source-Replacement Scheme

A *prepare-and-measure* (P&M) protocol is one where Alice prepares signal states and sends them to Bob, who then performs a measurement. It is now known that any such protocol can be realized by an entanglement-based one. The first work towards this was in Ref. [50], where it was noted that Einstein-Podolsky-Rosen (EPR) pairs could be used to implement BB84. A more general setting was considered in Ref. [51], and it was noted that prepare-and-measure continuous-variable protocols could be thought of as virtual entanglement-

[4] The nomenclature "key map" can be confusing. Strictly speaking, this map generates the raw key, which only becomes the true final key after error correction and privacy amplification.

[5] While this may seem limited to deterministic key maps, note that one can define the measurement POVMs to introduce randomness into the key.

[6] This is exactly what a security proof for the protocol provides, a lower bound on the length of the key, as a function of the parameter estimation results and the error correction leakage.

[7] The analysis of the classical error correction and privacy amplification steps, including modifications to the order of the two steps and different strategies for PA and EC, is a deep and interesting field in its own right. It is however outside the scope of this thesis.

based protocols. Finally, the ideas from the preceding papers were generalized to arbitrary prepare-and-measure protocols in Ref. [52].

This principle, which allows us to rewrite any P&M protocol as an EB one, is known as the *source-replacement scheme*. Thus, the description of EB protocols above encompasses P&M protocols as well. It should be emphasized that source-replacement is only a conceptual step; the physical implementation of the protocol does not change.

In this thesis, we focus on the case where the signal states are chosen from a finite set of pure states. Concretely, suppose Alice prepares signal states $|\psi_i\rangle$ with probability $p(i)$. In the source-replacement scheme, this is modelled as Alice preparing the state $\tau_{AA'} = \sum_{ij} \sqrt{p(i)p(j)}\, |i\rangle\langle j|_A \otimes |\psi_i\rangle\langle\psi_j|_{A'}$. System A' is sent through the unknown quantum channel to Bob, and becomes system B. Alice's virtual measurement in this EB description is given by the POVM $\{|i\rangle\langle i|_A\}$.

In addition to the usual constraints from parameter estimation, for P&M protocols there is an extra constraint on ρ_{AB}. As Eve cannot access Alice's system, the reduced density matrix of τ_A must be the same as ρ_A. Thus, for P&M protocols the additional constraint is $\rho_A = \tau_A = \sum_{ij} \sqrt{p(i)p(j)}\, \langle\psi_j|\psi_i\rangle\, |i\rangle\langle j|$. Note that τ_A is closely related to the Gram matrix of signal states.

3.2 Key Rate and Security Definition

In order to evaluate and compare QKD protocols, we must first define what it means for a QKD protocol to be secure. Once this is done, we can calculate the *key rate*, which is a measure of the efficiency of a QKD protocol. It quantifies how many bits of secret key are generated per round. For the purposes of this thesis, it is not important to understand the details of the QKD security definition; it suffices to take the Devetak-Winter formula [53] as a starting point. Nevertheless, it is good to have an overview of the underlying concepts.

3.2.1 QKD Security Definition

So far, we have taken an operational view of QKD protocols. To discuss the security definition, we now consider QKD protocols in a slightly more abstract light. At a high-level, the QKD protocol takes quantum states as input, and outputs a pair of keys to Alice and Bob. We desire that Alice and Bob receive the same key (correctness), and that an

eavesdropper does not know the key (secrecy) [54]. We now formalize these two notions. The reader is directed to Sec. 6.1 of Ref. [54] for further details.

Recall that a QKD protocol lasts for N rounds, and each round Alice and Bob receive a (possibly different) state in $\mathcal{H}_A \otimes \mathcal{H}_B$. The cumulative input Hilbert space is then $(\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes N}$, and the input state denoted by $\rho_{A^{\otimes N}B^{\otimes N}}$. We also want to consider the eavesdropper Eve, who is only limited by the laws of quantum mechanics. Without loss of generality, we can assume Eve, with her register denoted by E_N, holds the purification of $\rho_{A^{\otimes N}B^{\otimes N}}$.

We can always assume Eve at least holds the purifying register, as if she does not, giving this information to her could only reduce the key rate. On the other hand, if Eve held the purifying register and any additional register E', the purity of $\rho_{A^{\otimes N}B^{\otimes N}E_N}$ implies the total state would be $\rho_{A^{\otimes N}B^{\otimes N}E_N} \otimes \sigma_{E'}$. The product form of this state means there are no correlations between E' and the remaining registers, so it does not help Eve in any way. For the purification, recall that it suffices to choose $\mathcal{H}_{E_N}$ to have the same dimension as $(\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes N}$ (see Sec. 2.1.4).

Let $\mathcal{S}$ be the set of all possible keys of length l. Then, the output keys to Alice and Bob are held in respective classical registers K_A and K_B, with associated Hilbert spaces $\mathcal{H}_{K_A} = \mathcal{H}_{K_B} = \mathbb{C}^{|\mathcal{S}|}$.

We also need to account for the public information revealed during the course of the protocol. We let $\mathcal{H}_{C_N}$ be the Hilbert space, with dimension as large as needed, containing all classical public communication exchanged by Alice and Bob during the protocol. This includes their parameter estimation information, any announcements regarding sifting, error-correction information, and a choice of hash function for privacy amplification (see Fig. 6.4 in Ref. [54]).

We can now formally define the QKD protocol as being a completely positive and trace non-increasing map. This map is from $D((\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes N} \otimes \mathcal{H}_{E_N})$ to $\tilde{D}(\mathcal{H}_{K_A} \otimes \mathcal{H}_{K_B} \otimes \mathcal{H}_{C_N} \otimes \mathcal{H}_{E_N})$, acting trivially on $\mathcal{H}_{E_N}$, with the requirement that the final state $\tilde{\sigma}$ is classical-classical-classical-quantum. The protocol is not trace-preserving because it may abort, in which case it outputs the zero operator (see Sec. 6.1.3 of Ref. [54]).

With the formal definition of a protocol complete, we can now mathematically define what we mean by security. A protocol is said to be *ϵ-secure* if, for all input states, the output state satisfies

$$\frac{1}{2}\|\tilde{\sigma}_{K_A K_B C_N E_N} - \tilde{\sigma}_{ideal}\|_1 \leq \epsilon. \tag{3.1}$$

Here, $\tilde{\sigma}_{ideal} = \sum_{s\in\mathcal{S}} \frac{1}{|\mathcal{S}|} |s\rangle\langle s|_{K_A} \otimes |s\rangle\langle s|_{K_B} \otimes \tilde{\sigma}_{C_N E_N}$, where $\tilde{\sigma}_{C_N E_N} = \mathrm{Tr}_{K_A K_B}(\tilde{\sigma}_{K_A K_B C_N E_N})$. Up to normalization, $\tilde{\sigma}_{ideal}$ is the output of the ideal QKD protocol, in that the key is the

same between Alice and Bob (correct), and is uniformly random and completely uncorrelated to Eve (secret) (see also Remark 6.1.3 in Ref. [54]). Intuitively then, this security definition captures the notion that a secure QKD protocol produces outputs that are close to ideal, regardless of the input state. ϵ can be interpreted as the failure probability of the protocol. That is, the probability that the protocol does not abort and fails to provide a correct and secret key. For further motivation of this security definition, see the discussion leading up to Eq. 11 in Ref. [55].

Note that a protocol which always aborts is secure under this definition, though it is certainly not useful.

To understand why the security definition in Eq. (3.1) is a *composable* definition, consider applying any subsequent ideal cryptographic protocol, represented by some CPTNI map. By the monotonicity of trace distance under CPTNI maps, it follows that the outputs of this new protocol on the real and ideal state are no further apart in trace distance. Hence, the overall process comprising the initial QKD protocol and the subsequent cryptographic protocol is still ϵ-secure.[8]

3.2.2 Key Rate

Given the complete description of a QKD protocol, except for the size l of the output string, and a particular value of the security parameter ϵ, we wish to determine a value of l such that the protocol is ϵ-secure. In other words, we want to know how much to shrink the key in the final step of privacy amplification. This is known as a *security proof*. By the leftover hash lemma, this can be determined by solving a minimization of an entropic quantity known as the smooth min-entropy [54].

In full generality, computing this quantity is very challenging. There can be $N > 10^6$ rounds so the Hilbert spaces are very large, one has to contend with the fact that the input state could be entangled across multiple rounds, there are statistical fluctuations in the parameter estimation step, etc. Indeed, many basic questions about quantum key distribution remain open. For example, while it is known that entanglement is necessary for QKD [52], it is not known if it is sufficient. In the most general scenarios, one is typically content with any reasonable lower bound on the key rate, though it may not be clear how tight the answer is. In this context, tight means that the value of l is close to the largest value of l such that the protocol is still ϵ-secure.

[8] By the triangle inequality, if the subsequent protocol is instead ϵ'-secure, then the composite protocol is $(\epsilon + \epsilon')$-secure.

To simplify the problem, we can instead consider more restricted scenarios, in which case relatively simple and tight formulas for the key rate do exist. At the risk of overgeneralizing, the development behind most QKD security proofs is to first find the key rate under some restrictive assumptions, and then over time lift these assumptions at the cost of some decrease in the key rate.

One often considers restrictions on the form of the input states, leading to various classes of security proofs. Restrictions on the input state can equivalently be thought of as restrictions on the actions Eve is allowed to take. The most general attack, with no restrictions, is referred to as a *coherent attack*.

Asymptotic IID Collective Scenario

In this thesis, we focus on the independent and identically distributed (IID) collective scenario. It corresponds to a restricted attack where Eve interacts with the quantum state in the same manner and with a fresh ancilla each round. After all the rounds, she is however able to make a joint measurement on all the ancillas. This should not be confused with an *individual* attack, where Eve must measure each ancilla separately for each round.

Mathematically, this corresponds to the input state having the form $\rho_{A^{\otimes N}B^{\otimes N}E_N} = \rho_{ABE}^{\otimes N}$. As has been discussed above, without loss of generality, E is taken to be the purifying register of ρ_{ABE}.

The main advantage of studying the IID collective scenario is that we can determine the key rate just by considering one round. We do not have to consider the huge input state across all of the (potentially millions of!) rounds.

We further specialize to the asymptotic scenario, which is when the number of rounds $N \to \infty$. The main advantage of this is that we do not have to worry about statistical fluctuations, as all the frequencies from parameter estimation become probabilities. We model this by saying the expectations $\{\gamma_k\}$ of certain *coarse-grained* observables $\{\Gamma_{k_{AB}}\}$ are known. Alice and Bob could always benefit from using their full *fine-grained* data, so that $\{\Gamma_k\} = \{P_A^i \otimes P_B^j\}$. However, it can be challenging to deal with this set of observables when the alphabet is infinite (e.g. see Sec. 5.1.1), so we allow the flexibility to use a different set of observables, which are linear combinations of the fine-grained POVMs. The number of rounds m used for parameter estimation is chosen to be a vanishingly small fraction of N, so that there is no reduction in the key rate due to rounds consumed in parameter estimation.

Ultimately, one hopes to lift asymptotic collective security proofs to coherent attacks and to the regime of finite N. For protocols where $\mathcal{H}_A$ and $\mathcal{H}_B$ are finite-dimensional, and

where the protocol is invariant under permuting the different rounds, asymptotic coherent attacks are no stronger than collective ones, due to the *quantum de Finetti theorem* [56]. In brief, this theorem states that, after discarding some subsystems of an approximately symmetric (permutation-invariant) state, the state on the remaining subsystems can be approximated by a convex combination of IID states.

For protocols in infinite-dimensional spaces, the situation is a bit more subtle. While there are quantum de Finetti theorems that apply in infinite dimensions, they are slightly more involved than their finite-dimensional counterparts, requiring additional assumptions to make similar claims [57]. It is thus expected that the asymptotic collective key rate for infinite-dimensional protocols is the same as the asymptotic coherent key rate, but rigorously proving this requires some small additional steps.

Postprocessing Channel

Recall that by the IID collective assumption, we only need to consider a single round of the QKD protocol. To model this, we now define the channel corresponding to the measurement and key generation (collectively called postprocessing) that Alice and Bob perform.

We introduce registers $\tilde{A}$ and $\tilde{B}$ to hold the public announcements, $\overline{A}$ and $\overline{B}$ to hold the private measurement data, and Z to hold the result of the key map. As noted above, without loss of generality, Eve has access to the register E purifying ρ_{AB} and the public information.

It is worth noting that the public information in this context only includes the public announcements a_i, b_j. It does not include parameter estimation announcements, because the IID structure means Eve does not learn anything about the non-discarded signals from this information. It also does not include the information leaked during error correction, as that will be modelled separately. Finally it does not include the announcement of which hash function is used, as the information this gives Eve is already accounted for in the key rate formula.

Alice and Bob's measurement can be described by a channel Φ_M^{AB} that is simply given by $\mathrm{Tr}_{AB}(\Phi_T^{AB})$, where

$$\Phi_T^{AB}(\rho_{ABE}) = \sum_{i,j} |a_i\rangle\langle a_i|_{\tilde{A}} \otimes |\alpha_i\rangle\langle\alpha_i|_{\overline{A}} \otimes |b_j\rangle\langle b_j|_{\tilde{B}} \otimes |\beta_j\rangle\langle\beta_j|_{\overline{B}} \otimes \left[\left(\sqrt{P_A^i} \otimes \sqrt{P_B^j}\right) \rho_{ABE} \left(\sqrt{P_A^i} \otimes \sqrt{P_B^j}\right)\right]. \quad (3.2)$$

The action of the key map can be represented by the isometry

$$V = \sum_{S_A^{pub}, S_B^{pub}, S_B^{priv}} |g(a,b,\beta)\rangle_Z \otimes |a\rangle\langle a|_{\tilde{A}} \otimes |b\rangle\langle b|_{\tilde{B}} \otimes |\beta\rangle\langle\beta|_{\overline{B}}\,. \tag{3.3}$$

The final state between all parties is then $V\Phi_M^{AB}(\rho_{ABE})\,V^\dagger$. We can pull the partial trace in the measurement channel Φ_M^{AB} through the isometry V. Then, the final classical-quantum state between the register holding the result of the key map and Eve is

$$\sigma_{Z[E]} = \mathrm{Tr}_{A\overline{A}B\overline{B}}(V\Phi_T^{AB}(\rho_{ABE})V^\dagger), \tag{3.4}$$
$$\equiv \Phi(\rho_{ABE}). \tag{3.5}$$

The channel Φ characterizes the postprocessing steps and $[E]$ denotes the composite register $E\tilde{A}\tilde{B}$. Note that discarded signals will not contribute to the key rate, as $\sigma_{Z[E]}$ is block-diagonal in the classical announcements. To compute Φ, it is thus simpler and equivalent to not apply the POVM elements leading to discarded outcomes, rather than use a discard symbol [23]. In this case, the postprocessing map is completely positive and trace non-increasing.

Asymptotic Key Rate Formula

Given ρ_{AB}, the asymptotic secure key rate in the IID collective scenario is given by the Devetak-Winter formula [53]:

$$R^\infty = \frac{l}{N} = I(Z:X) - \chi(Z|[E]), \tag{3.6}$$

where the entropic quantities are evaluated on the postprocessed state after a single round, $\Phi(\rho_{ABE})$, and X refers to the party who does not perform the key map. Note that there is no dependence on the security parameter ϵ, due to the asymptotic limit.

In general, the state ρ_{AB} is unknown. However, it is constrained by Alice and Bob through testing. To determine a lower bound on the key rate, the Devetak-Winter formula should be evaluated on the worst-case state compatible with these constraints. This can be formalized as a convex optimization problem [3].

The Devetak-Winter formula is first rearranged as,

$$R^\infty = H(Z|[E]) - H(Z) + I(Z:X), \tag{3.7}$$
$$= H(Z|[E]) - H(Z|X). \tag{3.8}$$

Only the first term needs to be optimized over. The second term is replaced by the actual error-correction cost δ_{EC}^{leak} which is the number of bits leaked to perform error correction, normalized per round. For realistic error correction, δ_{EC}^{leak} will be larger than the Shannon limit $H(Z|X)$ (see Sec. 3.3.5).

As a shorthand notation, we write the convex objective function as

$$f^{QKD}(\rho_{AB}) = H(Z|[E])_{\Phi(\rho_{ABE})}. \tag{3.9}$$

To define the feasible set of the optimization, we use the information from parameter estimation. Additionally, for P&M protocols we have a constraint on the reduced density matrix ρ_A. This defines the convex feasible set as

$$\begin{aligned} \mathbf{S}^{QKD} = \{\rho \in \mathrm{Pos}\,(\mathcal{H}_{AB}) : &\mathrm{Tr}(\rho) = 1, \\ &\mathrm{Tr}_B(\rho) = \tau_A, \\ &\mathrm{Tr}(\rho\Gamma_i) = \gamma_i\}. \end{aligned} \tag{3.10}$$

Then, the convex optimization for the key rate is [4]

$$R^{\infty} = \min_{\rho_{AB} \in \mathbf{S}^{QKD}} [f^{QKD}(\rho_{AB})] - \delta_{EC}^{leak}. \tag{3.11}$$

Note that both the objective function and feasible set are convex so that Eq. (3.11) is a convex optimization.

When the Hilbert space $\mathcal{H}_{AB}$ is finite-dimensional, this problem can be reliably solved numerically [3, 4]. However, when the Hilbert space is infinite-dimensional, it is clearly not possible to directly solve this optimization numerically. We develop the dimension reduction method (see Sec. 4) which, for all QKD protocols, allows us to compute tight lower bounds on the asymptotic key rate by relating the infinite-dimensional optimization to a finite-dimensional one. We can then numerically solve the finite-dimensional problem using methods similar to Ref. [4] to get tight lower bounds on the key rate for protocols where the state lives in an infinite-dimensional Hilbert space.

3.2.3 Security Promise and Assumptions

In this thesis, we are interested in *device-dependent* QKD. Under this assumption, $\mathcal{H}_A$ and $\mathcal{H}_B$ are known, as is the exact CPTNI map the protocol implements.

In reality, the behaviour of a physical apparatus may deviate from the idealized model. Eavesdropping attacks enabled in this manner are referred to as *side-channel* attacks. Discovering new side channels and improving the modelling of realistic devices is an important virtuous cycle in QKD security analysis.

We note that alternatives exist. In measurement-device-independent QKD [58], Alice and Bob both have sources and the measurement is done by an untrusted third party. In this case, one only needs to model Alice and Bob's sources, and not the detectors. Device-independent QKD [59] enables secure key generation without any detailed modelling of Alice and Bob's devices.[9] Presently however, it suffers from poor performance compared to device-dependent QKD.

3.3 Numerical Optimization Formulation

From the previous section, we see that the asymptotic key rate is given by a minimization over the possible states held by Alice and Bob. For finite-dimensional Hilbert spaces, this optimization can be solved numerically, in principle. In practice, however, no numerical solver is perfect and computers have finite floating-point precision. Thus, directly solving this minimization will return at best an approximate minimum. This does not suffice for a rigorous QKD security proof, where it is essential that we obtain true lower bounds on the key rate. The numerical framework presented in Ref. [4] addresses the issue of determining reliable lower bounds on the secure key rate despite numerical imprecision, when the Hilbert space is finite-dimensional.

This section summarizes the numerical framework presented in Ref. [4], with modifications to the calculation of the loosened dual problem, due to the fact that we consider a different feasible set, and due to the fact that we have a slightly different method to calculate the expansion.

Beginning with the original primal problem from Eq. (3.11), the optimization is modified in successive steps to account for numerical precision. Ultimately, we wish to consider some form of dual problem, which will provide reliable lower bounds on the key rate. However, directly working with the dual of Eq. (3.11) is challenging because f is a nonlinear function. Hence, we first construct a linearized primal problem, and then take its dual.

[9] There are always some minimal assumptions on the devices. For example, they cannot just stealthily leak the key to Eve.

3.3.1 Different Form of Objective Function

In Section 4.0.2 of Ref. [4], the objective function is rewritten as a quantum relative entropy. The definition is equivalent to Eq. (3.9), and involves similar postprocessing maps. However, it is slightly different in directly discarding the postselected signals, and hence involving a completely positive and trace non-increasing map. We briefly summarize the definition here.

First, for each announcement $b \in S_B^{pub}$, we define an operator K_b as

$$K_b = \sum_{\{j \,:\, b_j = b\}} \sqrt{P_B^j} \otimes |b_j\rangle_{\tilde{B}} \otimes |\beta_j\rangle_{\overline{B}}\,, \tag{3.12}$$

and similarly for each $a \in S_A^{pub}$

$$K_a = \sum_{\{i \,:\, a_i = a\}} \sqrt{P_A^i} \otimes |a_i\rangle_{\tilde{A}} \otimes |\alpha_i\rangle_{\overline{A}}\,. \tag{3.13}$$

Then, the projector which implements the sifting step is defined as

$$\Pi_{sift} = \sum_{\{a,b \,:\, d(a,b)=0\}} |a\rangle\langle a|_{\tilde{A}} \otimes |b\rangle\langle b|_{\tilde{B}}\,. \tag{3.14}$$

Finally, the key map isometry V is the same as defined above (Eq. (3.3)). Then, the Kraus operators of the map $\mathcal{G}$ are defined as

$$\{V\ \Pi_{sift}\ (K_a \otimes K_b)\}_{a \in S_A^{pub},\ b \in S_B^{pub}}. \tag{3.15}$$

The map $\mathcal{Z}$ simply dephases the Z register, so has Kraus operators

$$\{|z\rangle\langle z|\}_{z \in S_Z}. \tag{3.16}$$

With these maps defined, the alternative form of the objective function is

$$f^{QKD}(\rho_{AB}) = D(\mathcal{G}(\rho_{AB}) || \mathcal{Z}(\mathcal{G}(\rho_{AB}))). \tag{3.17}$$

3.3.2 Primal Reformulation

The first step of the numerical framework is to determine an approximate solution to Eq. (3.11). There are different algorithms to do this; the one used in Ref. [4] and in this thesis is the Frank-Wolfe method [60].[10]

If the solution returned by the Frank-Wolfe method is not positive semidefinite, then it is perturbed by taking a mixture with a small multiple of the identity. We refer to this final solution as ρ_{guess}. It will not necessarily be an optimal point, nor satisfy all the constraints. However, it is guaranteed to be positive semidefinite and have unit trace.

A convex function is lower-bounded by its tangent hyperplane at any point. Using the same convention as Ref. [62], the gradient of f^{QKD} is defined as

$$\nabla f^{QKD}(\rho) = \sum_{ij} \frac{\partial f^{QKD}(\sigma)}{\partial \sigma_{ij}}\Big|_{\sigma=\rho} |j\rangle\langle i| \,. \tag{3.18}$$

We then have that

$$f^{QKD}(\sigma) \geq f^{QKD}(\rho_{guess}) + \mathrm{Tr}\big((\sigma - \rho_{guess})\nabla f^{QKD}(\rho_{guess})\big). \tag{3.19}$$

Re-introducing the minimization,

$$\begin{aligned}\min_{\sigma\in\mathbf{S}^{QKD}} f^{QKD}(\sigma) \geq\ & f^{QKD}(\rho_{guess}) \\ & + \mathrm{Tr}\big(-\rho_{guess}\nabla f^{QKD}(\rho_{guess})\big) + \min_{\sigma\in\mathbf{S}^{QKD}} \mathrm{Tr}\big(\sigma\nabla f^{QKD}(\rho_{guess})\big).\end{aligned} \tag{3.20}$$

The last term on the right-hand side is the reformulated primal problem. With regards to tightness, note that Eq. (3.20) holds with equality when ρ_{guess} is the true minimum.

Before proceeding with the dual problem, we note a technical caveat. The gradient of the function is not defined if $\mathcal{G}(\rho_{opt})$ is not full-rank. Hence, a perturbed function $f^{QKD}_{\epsilon_{pert}}$ is introduced, defined by

$$f^{QKD}_{\epsilon_{pert}}(\rho) = D(\mathcal{G}_{\epsilon_{pert}}(\rho)||\mathcal{Z}(\mathcal{G}_{\epsilon_{pert}}(\rho))), \tag{3.21}$$

where $\mathcal{G}_{\epsilon_{pert}}(\rho) = (1-\epsilon)\mathcal{G}(\rho) + \epsilon\frac{\mathbb{1}}{\mathrm{Tr}(\mathbb{1})}$. This perturbed function is then used in Eq. (3.20). In practice, $\epsilon_{pert} = 10^{-14}$ is used.

[10] A promising new approach has been recently presented in Ref. [61], though it currently applies to a restricted class of optimizations only.

3.3.3 Numerical Imprecision

It is not always possible to perfectly represent the set $\mathbf{S}^{QKD}$ numerically. To ensure a reliable lower bound, one should ensure a potentially larger set $\mathbf{S}^{QKD}_{expanded}$ is used. The details of how large this expansion should be depend on how $\mathbf{S}^{QKD}$ is defined, so we defer a detailed discussion on this matter to Appendix B. More details can be found in Appendix D of Ref. [4].

3.3.4 Dual SDP

After expanding the feasible set, and perturbing the objective function, the primal problem of interest is

$$\min_{\sigma \in \mathbf{S}^{QKD}_{expanded}} \mathrm{Tr}\Big(\sigma \nabla f^{QKD}_{\epsilon_{pert}}(\rho_{guess})\Big). \tag{3.22}$$

Finding the approximate minimum ρ_{guess} has been the objective of the first step of the numerical framework. The second step of the framework is to numerically solve the dual problem to Eq. (3.22). Of course, any numerical solution will only be approximately the optimal solution. This is why we solve the dual, since it ensures that any feasible answer is a lower bound on the primal problem. The specific form of the dual depends on how $\mathbf{S}^{QKD}_{expanded}$ is defined, so we defer the detailed discussion to Appendix B.

3.3.5 Simulation

In a real experiment, the expectations are obtained from measurement, and the number of bits leaked during error correction is directly observed. As theorists, we simulate data as expected from a typical implementation, in order to evaluate the expected performance of a protocol. It should be emphasized that this *simulation* step is not a security proof assumption. The security proof does *not* assume that Eve's attack is known. Simulation is merely determining representative inputs on which to run the security proof.

Expectations

In order to calculate the expectation values γ_i for the observables Γ_i (Eq. (3.10)), one first chooses a model for the channel. Then, Bob's conditional states can be calculated, and the conditional expectations determined.

Error-Correction Cost

The Shannon limit $H(Z|X)$ gives the minimum cost for asymptotic error correction. However, real error-correcting codes do not always reach this Shannon limit. Therefore, to calculate error-correction cost, one first uses the channel model to calculate $H(Z|X)$. Then, a further factor is introduced to model the realistic error-correction cost. There are two different parameterizations that are typically used.

The first is

$$\delta_{EC}^{leak} = fH(Z|X). \tag{3.23}$$

The range of the parameter is $1 \leq f \leq \frac{H(Z)}{H(Z|X)}$. The lower bound is because the EC cost is never less than the Shannon limit, while the upper bound arises because the EC cost is never more than the entropy of the key. Reporting error-correction costs in terms of f is common in DVQKD.

The second parametrization is

$$\delta_{EC}^{leak} = (1 - \beta_{EC})H(Z) + \beta_{EC}H(Z|X). \tag{3.24}$$

The range of the parameter is $0 \leq \beta_{EC} \leq 1$. Here the upper bound is to enforce $\delta_{EC}^{leak} \geq H(Z|X)$, while the lower bound enforces $\delta_{EC}^{leak} \leq H(Z)$. Reporting error-correction costs in terms of β_{EC} is common in CVQKD. (Just like $H(Z|X)$, $H(Z)$ is calculated using the channel model.)

As many theoretical QKD analyses focus on bounding $H(Z|[E])$, the modelling of error-correction costs can sometimes be an afterthought. However, especially when it comes to designing protocols and comparing their performance, it is important to accurately consider the effect of realistic error correction. See for example the discussion of the error-correction cost for DMCVQKD in Secs. 5.6 and 5.7.1.

Typically, the choice of modelling via f or β_{EC} is made so that setting the parameter to a constant value, independent of distance or other channel parameters, accurately reflects the EC cost. Constant f and constant β_{EC} however, are rather different heuristics, and can lead to substantial differences in the protocol performance. This is not to be confused with the fact that, at any fixed value of all the channel parameters, one could always express the EC cost equivalently in terms of f or β_{EC} just by reparametrizing.

Chapter 4

Dimension Reduction Method

The motivation for the dimension reduction method is as follows. On one hand, we have an expression for the key rate as a convex minimization (Eq. (3.11)). When the Hilbert space is finite-dimensional, we can use the existing numerical framework to efficiently compute tight and reliable lower bounds on the key rate. On the other hand, for all continuous-variable protocols, and for almost all experimental implementations of discrete-variable protocols, the relevant Hilbert space is one or more optical modes, and is hence infinite-dimensional. We cannot directly compute the key rate for these protocols numerically. Ostensibly, we should be able to work in some finite-dimensional subspace and still compute a good approximation to the key rate. Proving that we can indeed do so, and rigorously finding a lower bound from the approximate answer, is the core of the dimension reduction method.

Note that the squashing model [5–8] and flag-state squasher [9] solve a similar problem for the class of discrete-variable protocols. In fact, as is discussed in Sec. 6, the dimension reduction method can serve as an alternative to the flag-state squasher for DV protocols, where it can offer better numerical performance while achieving the same key rates.

While our primary application is to QKD, we present the key ideas of the dimension reduction method in greater generality; as a procedure to lower bound infinite-dimensional convex minimizations. Such an optimization problem arises in many different contexts. Thus, not only is our method relevant for QKD asymptotic [4] and finite key rate calculations [62], but it applies to evaluating other tasks in quantum communication, such as entanglement verification [24].

We first establish the general method in Sec. 4.1, and then show how to apply it to QKD in Sec. 4.2.

4.1 General Result

We start by defining the infinite-dimensional optimization of interest. Let $\mathcal{H}_\infty$ be a separable Hilbert space. As the notation suggests, $\mathcal{H}_\infty$ may be an infinite-dimensional Hilbert space. (It may also just be a very large finite-dimensional space.) Let $\mathbf{S}_\infty$ be a convex subset of $\tilde{D}(\mathcal{H}_\infty)$. Finally, let f be a convex function from $\tilde{D}(\mathcal{H}_\infty)$ to $\mathbb{R}$. Consider the convex optimization problem

$$\inf_{\tilde{\rho}\in\mathbf{S}_\infty} f(\tilde{\rho}). \tag{4.1}$$

Our goal is to find tight lower bounds on Eq. (4.1), without being able to directly solve the optimization. To do this, we will relate it to a suitably chosen finite-dimensional convex minimization (Eq. (4.2)), which can then be solved numerically[1]; along with an analytic correction term.

The intuition which guides our solution is that solving Eq. (4.1) over some finite-dimensional subspace should give an approximation to the actual optimum. However, in the context of QKD key rate calculations, an approximate lower bound does not suffice, and hence we need to rigorously relate the minima from the finite- and infinite-dimensional problems.

4.1.1 Preliminary Definitions

This section covers the relevant concepts and notation for proving our main result.

Finite-Dimensional Optimization

We first define the notation for the finite-dimensional optimization. This requires choosing the finite-dimensional subspace of $\mathcal{H}_\infty$ in which we will work. Let $\mathcal{H}_N$ be this subspace, and Π the projection onto this subspace. Let the new feasible set be $\mathbf{S}_N$, which is a convex subset of $\tilde{D}(\mathcal{H}_N)$. Then, the finite-dimensional optimization we will consider is

$$\inf_{\tilde{\rho}\in\mathbf{S}_N} f(\tilde{\rho}). \tag{4.2}$$

[1]Note that in order for Eq. (4.2) to be computationally tractable, there is an implicit assumption that computing f on finite-dimensional operators is numerically possible.

Uniform Closeness to Decreasing Under Projection

To help prove our main theorem, we next introduce a property of the function f that we will later need.

Definition 4.1.1. Given a projection Π, a function $f : \tilde{D}(\mathcal{H}_\infty) \to \mathbb{R}$ is *uniformly close to decreasing under projection (UCDUP) on* $\mathbf{S} \subseteq \tilde{D}(\mathcal{H}_\infty)$ *with correction term* Δ if, for every state $\tilde{\sigma} \in \mathbf{S}$, it satisfies

$$F(\tilde{\sigma}, \Pi\tilde{\sigma}\Pi) \geq \mathrm{Tr}(\tilde{\sigma}) - W \implies f(\Pi\tilde{\sigma}\Pi) - f(\tilde{\sigma}) \leq \Delta(W), \tag{4.3}$$

where Δ is a nonnegative, increasing function satisfying $\Delta(0) = 0$.

We make some remarks regarding this definition. Being UCDUP involves four quantities: a specific projection Π, the function f, a set $\mathbf{S}$ and a correction term Δ. The choice of correction term is not unique, there may be smaller (better) and larger (worse) ones. Given the same function, there may be different choices of correction term depending on the set.

The UCDUP property has some similarities to uniform-continuity with respect to trace distance. In fact, it is implied by the latter. Given this, one might wonder why we have introduced the UCDUP property. The answer is, because being UCDUP is a weaker condition, we may be able to find smaller Δ than implied by a uniform-continuity bound. A good example of this is in Appendix A.

In some cases, it may be possible to set $\Delta = 0$. This is the case for bounding the QKD objective function when the key map POVM elements commute with Π (see Sec. 4.2.3). It is also the case for some entanglement measures [24].

The following way to re-express the condition for being UCDUP will be very useful to us later on.

Lemma 1.

$$F(\tilde{\sigma}, \Pi\tilde{\sigma}\Pi) \geq \mathrm{Tr}(\tilde{\sigma}) - W \iff \mathrm{Tr}\big(\tilde{\sigma}\overline{\Pi}\big) \leq W. \tag{4.4}$$

Proof. Expanding the definition of fidelity,

$$F(\tilde{\sigma}, \Pi\tilde{\sigma}\Pi) = \mathrm{Tr}\left(\sqrt{\sqrt{\tilde{\sigma}}\Pi\tilde{\sigma}\Pi\sqrt{\tilde{\sigma}}}\right) \tag{4.5}$$

$$= \mathrm{Tr}\left(\sqrt{\sqrt{\tilde{\sigma}}\Pi\sqrt{\tilde{\sigma}}\sqrt{\tilde{\sigma}}\Pi\sqrt{\tilde{\sigma}}}\right) \tag{4.6}$$

$$= \mathrm{Tr}\left(\sqrt{\tilde{\sigma}}\Pi\sqrt{\tilde{\sigma}}\right) \tag{4.7}$$

$$= \mathrm{Tr}(\tilde{\sigma}\Pi). \tag{4.8}$$

Then,

$$F(\tilde{\sigma}, \Pi\tilde{\sigma}\Pi) \geq \mathrm{Tr}(\tilde{\sigma}) - W \iff \mathrm{Tr}(\tilde{\sigma}\Pi) \geq \mathrm{Tr}(\tilde{\sigma}) - W \tag{4.9}$$

$$\iff \mathrm{Tr}\left(\tilde{\sigma}\overline{\Pi}\right) \leq W \tag{4.10}$$

□

4.1.2 Theorem Statement

Theorem 1 (Relating Finite- and Infinite-Dimensional Optimizations)**.** *Assume* $\mathrm{Tr}\left(\tilde{\rho}\overline{\Pi}\right) \leq W$ *for all* $\tilde{\rho} \in \mathbf{S}_\infty$*. Further assume* $\Pi\mathbf{S}_\infty\Pi \subseteq \mathbf{S}_N$*. Finally, suppose* f *is UCDUP on* $\mathbf{S}_\infty$ *with correction term* Δ*. Then,*

$$\left[\inf_{\tilde{\rho}\in\mathbf{S}_N} f(\tilde{\rho})\right] - \Delta(W) \leq \left[\inf_{\tilde{\rho}\in\mathbf{S}_\infty} f(\tilde{\rho})\right]. \tag{4.11}$$

Proof. For clarity, we first prove the theorem under the assumption that there exist feasible operators $\tilde{\rho}^\infty$ and $\tilde{\rho}^N$ achieving the respective infima. We then extend the proof to hold without this assumption.

At a high level, our proof method is illustrated in Fig. 4.1. In order to relate the objective function f evaluated at the two optimal operators, $\tilde{\rho}^\infty$ and $\tilde{\rho}^N$, we introduce an auxiliary variable, $\tilde{\rho}^\Pi \equiv \Pi\tilde{\rho}^\infty\Pi$. This variable is the projection of the infinite-dimensional optimum $\tilde{\rho}^\infty$ onto the chosen finite subspace. We relate the optima to this auxiliary variable separately, and then to each other.

Inequality 1, Finite Set: We first relate $f(\tilde{\rho}^N)$ and $f(\tilde{\rho}^\Pi)$. By definition, $\tilde{\rho}^\Pi \in \Pi\mathbf{S}_\infty\Pi$. By the assumption that $\Pi\mathbf{S}_\infty\Pi \subseteq \mathbf{S}_N$, it follows that $\tilde{\rho}^\Pi \in \mathbf{S}_N$. Thus, $\tilde{\rho}^\Pi$ is feasible for the minimization over $\mathbf{S}_N$. Since $\tilde{\rho}^N$ achieves the infimum, it follows that

$$f(\tilde{\rho}^N) \leq f(\tilde{\rho}^\Pi). \tag{4.12}$$

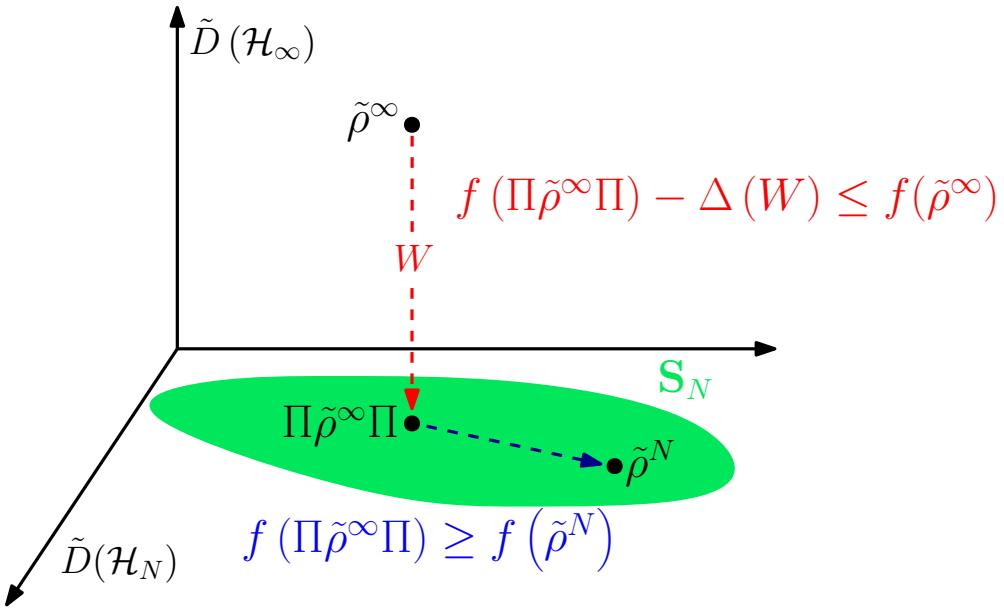


Figure 4.1: Pictorial representation of Theorem (1). The set $\mathbf{S}_N$ is chosen to contain the projection $\Pi\tilde{\rho}^\infty\Pi$. This auxiliary variable is used to relate $f(\tilde{\rho}^\infty)$ and $f(\tilde{\rho}^N)$

Inequality 2, Projection: By assumption, $\mathrm{Tr}\left(\tilde{\rho}^\infty\overline{\Pi}\right) \leq W$, and f is UCDUP on $\mathbf{S}_\infty$ with correction term Δ. By definition (see Eq. (4.3)),

$$f(\tilde{\rho}^\Pi) - \Delta(W) \leq f(\tilde{\rho}^\infty). \tag{4.13}$$

Thus, under the assumption that the infima are reached, the theorem follows from chaining the inequalities in Eq. (4.12) and Eq. (4.13).

We now extend the proof to the case where the infima are not necessarily reached. Let $\tilde{\rho}_n^\infty$ be a sequence of operators approaching the infimum of the infinite-dimensional problem Eq. (4.1). Let $\tilde{\rho}_n^\Pi \equiv \Pi\tilde{\rho}_n^\infty\Pi$. Each $\tilde{\rho}_n^\Pi$ is feasible for the finite-dimensional minimization, so $\inf_{\tilde{\rho}\in\mathbf{S}_N} f(\tilde{\rho}) \leq f(\tilde{\rho}_n^\Pi)$ for all n. For each n, it is also still the case that $f(\tilde{\rho}_n^\Pi) - \Delta(W) \leq f(\tilde{\rho}_n^\infty)$. We thus have

$$\left[\inf_{\tilde{\rho}\in\mathbf{S}_N} f(\tilde{\rho})\right] - \Delta(W) \leq f(\tilde{\rho}_n^\infty) \quad \forall n. \tag{4.14}$$

Taking the infimum over all n gives the desired theorem statement. □

4.1.3 Procedural Viewpoint

In order to apply this theorem, in addition to choosing $\mathcal{H}_N$ and $\mathbf{S}_N$, we need to determine an expression for the correction term Δ and a suitable value for W (see Sec. 4.2). For the latter quantity, we want to choose the smallest value of W satisfying

$$W \geq \sup_{\tilde{\rho} \in \mathbf{S}_\infty} \mathrm{Tr}\big(\tilde{\rho}\overline{\Pi}\big). \tag{4.15}$$

We thus see that the determination of W itself involves an infinite-dimensional optimization. In practice, this optimization tends to be considerably easier to solve than the original one (Eq. (4.1)). In particular, when $\mathbf{S}_\infty$ is the feasible set of a semidefinite program (SDP), known as a spectrahedron, then any feasible solution to the dual problem is a feasible value for W.[2] Note that W will not only be used to determine the correction term $\Delta(W)$, but also to parametrize the set $\mathbf{S}_N$ we choose (see Sec. 4.2.4).

Another remark is that, assuming one wants to obtain the largest lower bound on the original minimization, the best choice of the finite set is $\mathbf{S}_N = \Pi\mathbf{S}_\infty\Pi$. However, in general it seems difficult to find a closed form expression for this set: so in practice we often resort to slightly looser definitions of $\mathbf{S}_N$.

To clarify the roles of the different quantities, and to better enable its application to different problems of interest, we summarize the dimension reduction method in a procedural manner in Fig. 4.2.

4.2 Application to QKD

Having developed a general method to lower-bound convex minimizations, we now show how it can be applied to asymptotic QKD key rate calculations. The initial infinite-dimensional optimization is Eq. (3.11) (see Sec. 3.2.2), so that $\mathcal{H}_\infty = \mathcal{H}_{AB}$, $f = f^{QKD}$, and $\mathbf{S}_\infty = \mathbf{S}^{QKD}$.

As shown in the flowchart, four inputs are needed to apply Theorem (1): the subspace $\mathcal{H}_N$ onto which Π projects, the bound on weight outside the subspace W, the correction term Δ, and the finite set $\mathbf{S}_N$. Choosing or determining the first two of these quantities is very specific to the QKD protocol being considered. Thus, in this section we only give some general remarks on their derivation. On the other hand, the forms of the second two quantities are protocol-agnostic, so we can derive them in detail here.

[2] For the QKD protocols we study, we can obtain tight values of W by analytically solving the dual or a relaxed version of the primal problem (see Sec. 4.2.2).

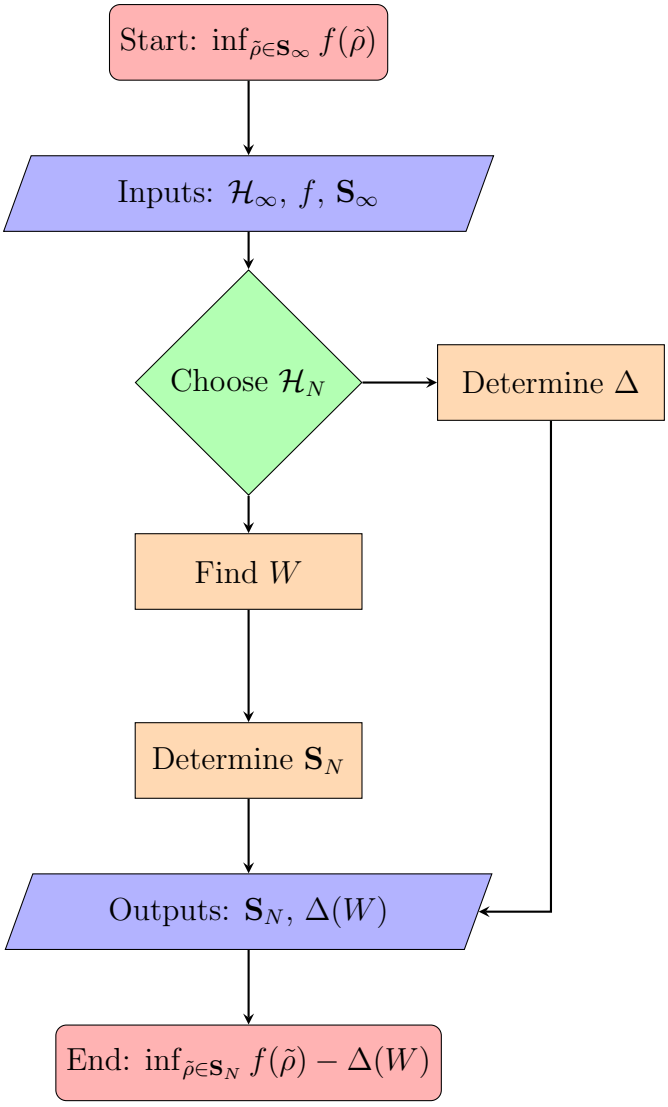


Figure 4.2: Flowchart indicating the key steps in applying the dimension reduction method.

4.2.1 Finite Subspace Π

This step is intimately related to the protocol being studied. There are no restrictions on the choice of finite subspace $\mathcal{H}_N$. However, there are two considerations to keep in mind to get the best key rates.

The first consideration is that certain choices will lead to a smaller W. The dimension of the finite subspace will generally be upper-bounded by considerations of runtime or numerical solver limitations. For a fixed finite dimension N then, the choice of subspace minimizing W would be the one containing the most weight of $\tilde{\rho}^{\infty}$. That is, choosing the first N eigenvectors of $\tilde{\rho}^{\infty}$ minimizes the weight outside the subspace. Here, the eigenvectors are ordered so that the corresponding eigenvalues are listed largest to smallest. However, the state $\tilde{\rho}^{\infty}$ is unknown. A good heuristic is to instead choose the subspace containing the most weight of the expected state under a representative channel model (see Sec. 5.5.1). That is, one derives ρ_{AB} for an honest implementation of the protocol, and chooses the subspace to be the first N eigenvectors. This is conceptually similar to assuming a particular channel for the purpose of designing a QKD protocol or error-correcting code.

The second consideration is choosing the projection to commute, or almost commute, with the objective function POVM elements (see Sec. 4.2.3) or the constraint operators (see Sec. 4.2.4). In the former case, this leads to a smaller correction term Δ. In the latter case, it enables tighter definitions of the finite set $\mathbf{S}_N$.

Note that although the original Hilbert space $\mathcal{H}_{\infty}$ is bipartite, if Π is not of the form $\Pi_A \otimes \Pi_B$, the projected space $\mathcal{H}_N$ will not be bipartite. Nevertheless, it is still possible to derive constraints on ρ_A, though one has to be cautious in defining the operations of partial trace and its adjoint (see Appendix C.3).

4.2.2 Weight W

This step is also protocol-specific. For $\mathbf{S}_{\infty}$ as given in Eq. (3.10), the bound on W shown in Eq. (4.15) is a semidefinite program. After writing down the dual of this SDP, finding any feasible solution suffices for W. Indeed, this is the approach taken for DMCVQKD (see Sec. 5.5.2). A different approach is used for unbalanced phase-encoded BB84. The monotonicity of cross-clicks or double-clicks with increasing photon number, along with Markov's inequality, is used to bound W [10].

4.2.3 Correction Term Δ

Recall our objective function $f = f^{QKD}$ is given in Eq. (3.9). We will show that it is UCDUP and determine the correction term Δ as a function of W. We first present a general analytic correction term which does not depend on the details of the postprocessing map (see Eq. (3.5)) and is thus applicable to all protocols. We then show that when the postprocessing map satisfies a certain property, which holds for some protocols, we can omit the correction term entirely, i.e. $\Delta = 0$. Finally, we consider a third form of the correction term, which again applies to all protocols, and interpolates between the two previous cases. While we present all three cases for completeness, the reader who is not interested in the development of the concepts should directly skip to the third form of the correction term, as it completely subsumes the previous two cases.

Recall our goal is to compute an upper bound on

$$f(\Pi\tilde{\rho}^{\infty}\Pi) - f(\tilde{\rho}^{\infty}), \tag{4.16}$$

where $f(\rho_{AB}) = H(Z|[E])_{\Phi(\rho_{ABE})}$.

Define Ξ_{AB} to be a dephasing channel associated with the projector Π as

$$\Xi_{AB}(\rho) \equiv \Pi\rho\Pi + \overline{\Pi}\rho\overline{\Pi}. \tag{4.17}$$

While the correction terms differ in how they bound the trace distance, all use the same continuity bound for conditional entropy. We thus present it in a lemma here.

Lemma 2. *Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be two Hilbert spaces, where the dimension of $\mathcal{H}_A$ is $|A|$ while $\mathcal{H}_B$ can be infinite-dimensional. Let $\tilde{\rho}_{AB}, \tilde{\sigma}_{AB} \in \tilde{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be two subnormalized, classical-quantum states with $\mathrm{Tr}(\tilde{\rho}_{AB}) \geq \mathrm{Tr}(\tilde{\sigma}_{AB})$. If $\frac{1}{2}\|\tilde{\rho}_{AB} - \tilde{\sigma}_{AB}\|_1 \leq \epsilon$, then*

$$H(A|B)_{\tilde{\sigma}_{AB}} - H(A|B)_{\tilde{\rho}_{AB}} \leq \epsilon \log_2 |A| + (1+\epsilon) h\left(\frac{\epsilon}{1+\epsilon}\right), \tag{4.18}$$

where $h(x)$ is the binary entropy function.

Proof. This result is a generalization of Lemma 2 in Ref. [63], which was derived for normalized states. The proof of the extension to subnormalized states is given in Appendix A. □

General Case

With Lemma 2 in hand, we will prove that f is UCDUP on $D(\mathcal{H}_\infty)$ in the following theorem. Note this is sufficient as for any QKD protocol, $\mathbf{S}_\infty \subseteq D(\mathcal{H}_\infty)$.

Theorem 2. *For any projection Π, the QKD objective function f is UCDUP on $D(\mathcal{H}_\infty)$ with correction term*

$$\Delta(W) = \sqrt{W}\log_2|Z| + \left(1+\sqrt{W}\right) h\left(\frac{\sqrt{W}}{1+\sqrt{W}}\right), \tag{4.19}$$

where $|Z|$ is the dimension of the key map register.

Proof. As per the definition of UCDUP, let $\sigma_{AB} \in D(\mathcal{H}_\infty)$ be a state satisfying $F(\sigma_{AB}, \Pi\sigma_{AB}\Pi) = \mathrm{Tr}(\sigma_{AB}\Pi) \geq 1 - W$. Let σ_{ABE} be a purification of σ_{AB} and note that $\tilde{\sigma}^{\Pi}_{ABE} \equiv \Pi\sigma_{ABE}\Pi$ is a purification of $\Pi\sigma_{AB}\Pi$. Since Π only acts on the AB subsystem, $\mathrm{Tr}\left(\tilde{\sigma}^{\Pi}_{ABE}\right) = \mathrm{Tr}(\Pi\sigma_{AB}\Pi)$. By Eq. (2.29), it follows that

$$\frac{1}{2}\left\|\sigma_{ABE} - \tilde{\sigma}^{\Pi}_{ABE}\right\|_1 \leq \sqrt{W}. \tag{4.20}$$

By the monotonicity of trace distance under channels, we have

$$\frac{1}{2}\left\|\tau_{Z[E]} - \tilde{\tau}^{\Pi}_{Z[E]}\right\|_1 \leq \sqrt{W} \tag{4.21}$$

where $\tau_{Z[E]} = \Phi(\sigma_{ABE})$ and $\tilde{\tau}^{\Pi}_{Z[E]} = \Phi(\tilde{\sigma}^{\Pi}_{ABE})$ (recall Φ is the postprocessing map defined in Eq. (3.5)).

Since Φ is trace-preserving, $\mathrm{Tr}\left(\tau_{Z[E]}\right) \geq \mathrm{Tr}\left(\tilde{\tau}^{\Pi}_{Z[E]}\right)$. Thus, we can apply Lemma 2 to obtain

$$f(\Pi\sigma_{AB}\Pi) - f(\sigma_{AB}) = H(Z|[E])_{\tilde{\tau}^{\Pi}_{Z[E]}} - H(Z|[E])_{\tau_{Z[E]}} \tag{4.22}$$

$$\leq \sqrt{W}\log_2|Z| + (1+\sqrt{W})h\left(\frac{\sqrt{W}}{1+\sqrt{W}}\right). \tag{4.23}$$

This is precisely the condition we require for f to be UCDUP, so we identify the right-hand side as $\Delta(W)$. □

For a key map with M outcomes, the dimension of the key map register is $|Z| = M$. For the purpose of this argument, the discard symbol $\perp$ does not count towards a key outcome. The reason for this is that instead of $\perp$, one could use any pre-existing key symbol to flag discarded signals. Since the classical-quantum state $\sigma_{Z[E]}$ between the key map register and Eve is block-diagonal in the classical announcements (Eq. (3.5)), Eve could identify the discarded signals from those public announcements alone. This would leave the value of the objective function unchanged. Indeed, the $\perp$ symbol is only used for clarity in our presentation.

Block-Diagonal (Commuting) Case

If the key map POVM elements are block-diagonal with respect to Π and $\overline{\Pi}$, the correction term is zero.

Theorem 3. *Let Φ be defined by the POVMs $\{P_A^i\}$ and $\{P_B^j\}$ and a key map isometry V. If all POVM elements are block-diagonal, that is $[P_A^i \otimes P_B^j, \Pi] = 0\ \forall i, j$, then*

$$\Delta(W) = 0. \tag{4.24}$$

Proof. Recall that Π only acts on the AB subsystem. Then, $\Pi\rho_{ABE}\Pi$ is a purification of $\Pi\rho_{AB}\Pi$. Π commutes with V as they act on different subsystems, and commutes with all elements of the POVMs $\{P_A^i\}$ and $\{P_B^j\}$ by assumption. By the definition of Φ_T^{AB} (Eq. (3.2)), it then follows that

$$\Pi\ V\ \Phi_T^{AB}\left(\rho_{ABE}\right)\ V^\dagger\ \Pi = V\ \Phi_T^{AB}\left(\Pi\rho_{ABE}\Pi\right)\ V^\dagger \tag{4.25}$$

and analogously for $\overline{\Pi}$.

We have

$$\begin{aligned}
f(\rho_{AB}) &= H(Z|[E])_{\Phi(\rho_{ABE})} && (4.26)\\
&= H(Z|[E])_{\mathrm{Tr}_{A\overline{A}B\overline{B}}\left(V\Phi_T^{AB}(\rho_{ABE})V^\dagger\right)} && (4.27)\\
&= H(Z|[E])_{\mathrm{Tr}_{A\overline{A}B\overline{B}}\left(\Xi_{AB}\left(V\Phi_T^{AB}(\rho_{ABE})V^\dagger\right)\right)} && (4.28)\\
&= H(Z|[E])_{\mathrm{Tr}_{A\overline{A}B\overline{B}}\left(V\Phi_T^{AB}(\Xi_{AB}(\rho_{ABE}))V^\dagger\right)} && (4.29)\\
&\geq H(Z|[E])_{\Phi(\Pi\rho_{ABE}\Pi)} + H(Z|[E])_{\Phi(\overline{\Pi}\rho_{ABE}\overline{\Pi})} && (4.30)\\
&\geq H(Z|[E])_{\Phi(\Pi\rho_{ABE}\Pi)} && (4.31)\\
&= f(\Pi\rho_{AB}\Pi), && (4.32)
\end{aligned}$$

where we have in Eq. (4.28) freely introduced the dephasing channel (defined in Eq. (4.17)) as it will be traced out, in Eq. (4.30) used the concavity of conditional entropy, and in Eq. (4.31) used the nonnegativity of conditional entropy for classical-quantum states (see Eq. (3.5)). □

Almost Block-Diagonal (Partially-Commuting) Case

In principle, we expect there to be a correction term which interpolates between the two previous cases, with a dependence on 'how much' the projection commutes with the post-processing channel. This would allow a smaller correction term than Theorem 2 for protocols where the POVM is almost, but not fully, block-diagonal. We derive such a correction term in this section.

We begin by expressing the postprocessing map in a specific manner. First, let us combine the public announcement registers $\tilde{A}$ and $\tilde{B}$ (see Sec. 3.2.2) into a single register C, over an alphabet $S_C = S_A^{pub} \times S_B^{pub}$. Observe that the map Φ (Eq. (3.5)) is a quantum-to-classical channel from AB to ZC. It follows that Φ can be realized by a measurement (Eq. (2.24)). Then, Φ has the form

$$\Phi(\rho_{ABE}) = \sum_{\substack{z \in S_Z \\ c \in S_C}} |z\rangle\langle z|_Z \otimes |c\rangle\langle c|_C \otimes \mathrm{Tr}_{AB}\left[(P_{AB}^{z,c} \otimes \mathbb{1}_E)\rho_{ABE}\right], \tag{4.33}$$

where

$$\{P_{AB}^{z,c}\}_{\substack{z \in S_Z \\ c \in S_C}} \tag{4.34}$$

is some POVM.[3] For simplicity, we re-index the POVM by $k \in S_K \equiv S_Z \times S_C$.

We now state the theorem for the improved correction term. Recall that X^g denotes the generalized inverse of X. Also recall that the postselection symbol $\perp$ does not count toward the dimension of the key map register $|Z|$.

Theorem 4. *Consider the QKD objective function $f(\rho_{AB}) = H(Z|[E])_{\Phi(\rho_{ABE})}$, with the map Φ defined by a POVM $\{P_k\}_{k \in S_K}$. With respect to any projection Π, write each P_k as a block matrix*

$$P_k = \begin{pmatrix} \Pi P_k \Pi & \Pi P_k \overline{\Pi} \\ \overline{\Pi} P_k \Pi & \overline{\Pi} P_k \overline{\Pi} \end{pmatrix} \equiv \begin{pmatrix} A_k & B_k \\ {B_k}^\dagger & D_k \end{pmatrix}. \tag{4.35}$$

[3] Concretely, one can construct this POVM by following through the action of the key map and the corresponding coarse-graining of the original POVM it induces.

For this projection Π*, the QKD objective function* f *is UCDUP on* $D(\mathcal{H}_\infty)$ *with correction term*

$$\Delta(W) = c \log_2 |Z| + (1+c)\, h\left(\frac{c}{1+c}\right), \tag{4.36}$$

where $|Z|$ *is the dimension of the key map register and*

$$c = \sqrt{W} \max_{k \in S_K} \left\| \sqrt{A_k}^g B_k \sqrt{D_k}^g \right\|_\infty. \tag{4.37}$$

The quantity $\max_{k\in S_K} \left\|\sqrt{A_k}^g B_k \sqrt{D_k}^g\right\|_\infty$ has a nice interpretation as quantifying how close to block-diagonal the POVM elements are. It is 0 when all the POVMs are block-diagonal, since then $B_k = 0$ for all k. By the positivity of each P_k, this quantity is also upper-bounded by 1 (this is made clear in the proof of Lemma 4). Hence, this correction term interpolates between the two cases we considered earlier, as c ranges from 0 to $\sqrt{W}$.

In order to prove this theorem, we introduce two lemmas. The following lemma lets us work with the dephased state instead of the projected one.

Lemma 3. *For any state* ρ_{AB}*,* $f(\Pi\rho_{AB}\Pi) \leq H(Z|[E])_{\Phi(\Xi(\rho_{ABE}))}$

Proof. Expanding definitions, we have that

$$\begin{aligned}
f(\Pi\rho_{AB}\Pi) &= H(Z|[E])_{\Phi(\Pi\rho_{ABE}\Pi)} && (4.38)\\
&\leq H(Z|[E])_{\Phi(\Pi\rho_{ABE}\Pi)} + H(Z|[E])_{\Phi(\overline{\Pi}\rho_{ABE}\overline{\Pi})} && (4.39)\\
&\leq H(Z|[E])_{\Phi(\Pi\rho_{ABE}\Pi)+\Phi(\overline{\Pi}\rho_{ABE}\overline{\Pi})} && (4.40)\\
&= H(Z|[E])_{\Phi(\Pi\rho_{ABE}\Pi+\overline{\Pi}\rho_{ABE}\overline{\Pi})} && (4.41)\\
&= H(Z|[E])_{\Phi(\Xi(\rho_{ABE}))}. && (4.42)
\end{aligned}$$

Line (4.39) follows because the second term is the conditional entropy of a CQ state and thus nonnegative, (4.40) follows because conditional entropy is concave, and (4.41) follows simply because the map Φ is linear. □

The next lemma bounds the trace norm for a specific form of operator.

Lemma 4. *Let* P *be a POVM element. With respect to a projection* Π*, write* P *as a block matrix*

$$P = \begin{pmatrix} \Pi P \Pi & \Pi P \overline{\Pi} \\ \overline{\Pi} P \Pi & \overline{\Pi} P \overline{\Pi} \end{pmatrix} \equiv \begin{pmatrix} A & B \\ B^\dagger & D \end{pmatrix}. \tag{4.43}$$

Define $H = \begin{pmatrix} 0 & B \\ B^\dagger & 0 \end{pmatrix}$ *as the off-diagonal portion of* P.

Let ρ *be a state, and define two new states corresponding to the normalized on-diagonal blocks of* ρ: $\rho^\Pi = \frac{\Pi\rho\Pi}{\mathrm{Tr}(\rho\Pi)}$ *and* $\rho^{\overline{\Pi}} = \frac{\overline{\Pi}\rho\overline{\Pi}}{\mathrm{Tr}\left(\rho\overline{\Pi}\right)}$. *Define the measurement probabilities* $r = \mathrm{Tr}\left(\rho^\Pi P\right)$ *and* $s = \mathrm{Tr}\left(\rho^{\overline{\Pi}} P\right)$. *As usual, let* $W \geq \mathrm{Tr}\left(\rho\overline{\Pi}\right)$.

It holds that

$$\|\sqrt{\rho}H\sqrt{\rho}\|_1 \leq (r+s)\sqrt{W}\left\|\sqrt{A}^g B \sqrt{D}^g\right\|_\infty, \tag{4.44}$$

where $(\cdot)^g$ denotes the generalized inverse.

Proof. Using the SDP formulation for trace norm [25],

$$\begin{aligned} \left\|\sqrt{\rho}H\sqrt{\rho}\right\|_1 \quad = \quad & \underset{X}{\text{minimize:}} \quad \mathrm{Tr}\, X \\ & \text{subject to:} \quad X \geq \sqrt{\rho}H\sqrt{\rho} \\ & \qquad\qquad\quad X \geq -\sqrt{\rho}H\sqrt{\rho} \\ & \qquad\qquad\quad X \geq 0. \end{aligned} \tag{4.45}$$

A moment's reflection convinces us that X can always be chosen to satisfy $X = \Pi_{\mathrm{im}(\rho)} X \Pi_{\mathrm{im}(\rho)}$ ($\Pi_{\mathrm{im}(\rho)}$ is the projection onto the image of ρ, which is a closed subspace). WLOG, we can then reparametrize X as $\sqrt{\rho}R\sqrt{\rho}$. This lets us rewrite the SDP (4.45) as

$$\begin{aligned} \left\|\sqrt{\rho}H\sqrt{\rho}\right\|_1 \quad = \quad & \underset{R}{\text{minimize:}} \quad \mathrm{Tr}(\rho R) \\ & \text{subject to:} \quad \sqrt{\rho}R\sqrt{\rho} \geq \sqrt{\rho}H\sqrt{\rho} \\ & \qquad\qquad\quad \sqrt{\rho}R\sqrt{\rho} \geq -\sqrt{\rho}H\sqrt{\rho} \\ & \qquad\qquad\quad R \geq 0. \end{aligned} \tag{4.46}$$

As this is a minimization, any feasible value provides an upper bound. It suffices to choose a positive R satisfying $R \geq \pm H$.

The remainder of the proof consists of three steps. We first choose the feasible guess for R. We then show that it is indeed feasible. Finally, we calculate the corresponding value of the objective function.

To specify the guess for R, note the following facts. P is positive since it is a POVM element. In terms of the block matrix characterization,

$$\begin{pmatrix} A & B \\ B^\dagger & D \end{pmatrix} \geq 0. \tag{4.47}$$

By Theorem IX.5.9 of Ref. [64], this positivity implies $\sqrt{A}^g B \sqrt{D}^g \equiv K$ is a contraction, i.e. $\|K\|_\infty \leq 1$.[4] Let $\|K\|_\infty = \lambda$. Note λ ranges from 0 to 1 and quantifies how close to block-diagonal the POVM element is. In particular, $\lambda = 0$ when the POVM is exactly block-diagonal, since then $B = K = 0$.

We now specify our guess to be

$$R = a\Pi P \Pi + b\overline{\Pi} P \overline{\Pi}, \tag{4.48}$$

with the constants $a = \lambda\sqrt{W}$ and $b = \frac{\lambda}{\sqrt{W}}$ (assuming $W \neq 0$, as the lemma follows immediately otherwise).[5]

Let us verify that this guess is feasible. Since R is manifestly positive, all that remains is to show $R \pm H \geq 0$. Written in terms of block matrices, this condition is equivalent to

$$R \pm H = \begin{pmatrix} aA & \pm B \\ \pm B^\dagger & bD \end{pmatrix} \geq 0. \tag{4.49}$$

Again by Theorem IX.5.9 of Ref. [64], this condition is satisfied if $\sqrt{aA}^g (\pm B) \sqrt{bD}^g$ is a contraction. Noting that $\sqrt{ab} = \lambda$, this simplifies as

$$\begin{aligned} \sqrt{aA}^g (\pm B) \sqrt{bD}^g &= \frac{1}{\sqrt{ab}} \sqrt{A}^g (\pm B) \sqrt{D}^g && (4.50) \\ &= \pm \frac{1}{\lambda} K && (4.51) \\ &= \frac{\pm K}{\|K\|_\infty}. && (4.52) \end{aligned}$$

The operator on the last line clearly has unit norm so is a contraction. Thus, it follows that $R \pm H \geq 0$ and R is a feasible guess.

The objective function value is

$$\begin{aligned} \operatorname{Tr}(\rho R) &= a \operatorname{Tr}(\rho A) + b \operatorname{Tr}(\rho D) && (4.53) \\ &= a \operatorname{Tr}(\rho \Pi P \Pi) + b \operatorname{Tr}\big(\rho \overline{\Pi} P \overline{\Pi}\big). && (4.54) \end{aligned}$$

[4] In Ref. [64], this theorem is proven for finite-dimensional matrices where B is a square block. However, nothing precludes the proof from applying in infinite dimensions and with rectangular blocks.

[5] We thank Thomas van Himbeeck for key ideas leading to this feasible guess.

The first term can be upper bounded as

$$
\begin{aligned}
a \operatorname{Tr}(\rho \Pi P \Pi) &= a \operatorname{Tr}(\Pi \rho \Pi \ \Pi P \Pi) && (4.55)\\
&= a \operatorname{Tr}(\Pi \rho \Pi) \operatorname{Tr}\left(\rho^{\Pi} P\right) && (4.56)\\
&\leq a \operatorname{Tr}\left(\rho^{\Pi} P\right) && (4.57)\\
&= \lambda \sqrt{W} r. && (4.58)
\end{aligned}
$$

Similarly for the second term,

$$
\begin{aligned}
b \operatorname{Tr}\left(\rho \overline{\Pi} P \overline{\Pi}\right) &= b \operatorname{Tr}\left(\overline{\Pi} \rho \overline{\Pi} \ \overline{\Pi} P \overline{\Pi}\right) && (4.59)\\
&= b \operatorname{Tr}\left(\overline{\Pi} \rho \overline{\Pi}\right) \operatorname{Tr}\left(\rho^{\overline{\Pi}} P\right) && (4.60)\\
&\leq bW \operatorname{Tr}\left(\rho^{\overline{\Pi}} P\right) && (4.61)\\
&= \lambda \sqrt{W} s. && (4.62)
\end{aligned}
$$

Thus, the feasible choice of R in Eq. (4.48) leads to the following upper bound on Eq. (4.46),

$$
\left\| \sqrt{\rho} H \sqrt{\rho} \right\|_1 \leq (r+s) \lambda \sqrt{W}, \tag{4.63}
$$

and the proof is complete. □

Proof of Theorem 4. As per the definition of UCDUP, let $\rho_{AB} \in D(\mathcal{H}_\infty)$ be a state satisfying $\operatorname{Tr}\left(\rho_{AB} \overline{\Pi}\right) \leq W$. We need to bound the trace distance between $\Phi(\Xi(\rho_{ABE}))$ and $\Phi(\rho_{ABE})$. We have

$$
\begin{aligned}
\Phi(\Xi(\rho_{ABE})) &= \sum_k |k\rangle\langle k|_K \otimes \operatorname{Tr}_{AB}\left[(P^k_{AB} \otimes \mathbb{1}_E) \Xi(\rho_{ABE})\right] && (4.64)\\
&= \sum_k |k\rangle\langle k|_K \otimes \operatorname{Tr}_{AB}\left[(\Xi(P^k_{AB}) \otimes \mathbb{1}_E) \rho_{ABE}\right] && (4.65)
\end{aligned}
$$

since the channel Ξ is self-adjoint. $\{\Xi(P^k)\}$ is also a POVM. In writing $\Phi(\Xi(\rho_{ABE}))$ in this manner, it now looks like we are comparing the effect of two different channels on the same input, instead of the same channel on two different inputs.

Since the trace norm is additive over blocks, we have

$$\left\|\Phi(\rho_{ABE}) - \Phi(\Xi(\rho_{ABE}))\right\|_1 = \left\|\sum_k |k\rangle\langle k|_K \otimes \mathrm{Tr}_{AB}\left[\left(\left[P^k_{AB} - \Xi(P^k_{AB})\right] \otimes \mathbb{1}_E\right)\rho_{ABE}\right]\right\|_1 \tag{4.66}$$

$$= \sum_k \left\|\mathrm{Tr}_{AB}\left[\left(\left[P^k_{AB} - \Xi(P^k_{AB})\right] \otimes \mathbb{1}_E\right)\rho_{ABE}\right]\right\|_1. \tag{4.67}$$

To proceed, we find a more useful form for Eve's conditional states. Recall that Eve's register E purifies ρ_{AB}. We can thus assume that Eve's register has the same dimension as Alice and Bob's (see Sec. 2.1.4). That is, $\mathcal{H}_E = \mathcal{H}_{AB}$. There then exists a bijective isometry $V : \mathcal{H}_{AB} \to \mathcal{H}_E$. (To construct such an isometry, simply choose a basis $|i\rangle_{AB}$ for $\mathcal{H}_{AB}$ and a basis $|i\rangle_E$ for $\mathcal{H}_E$, and define $V|i\rangle_{AB} = |i\rangle_E$.) Via the vectorization mapping, it can easily be shown that

$$\mathrm{Tr}_{AB}\left(\left(P^k_{AB} \otimes \mathbb{1}_E\right)\rho_{ABE}\right) = V\left(\sqrt{\rho_{AB}}P^k_{AB}\sqrt{\rho_{AB}}\right)V^\dagger, \tag{4.68}$$

and similarly for $\Xi(P^k_{AB})$.

Applying this identity to Eq. (4.67), we have

$$\left\|\Phi(\rho_{ABE}) - \Phi(\Xi(\rho_{ABE}))\right\|_1 = \sum_k \left\|V\left(\sqrt{\rho_{AB}}\left[P^k_{AB} - \Xi(P^k_{AB})\right]\sqrt{\rho_{AB}}\right)V^\dagger\right\|_1 \tag{4.69}$$

$$= \sum_k \left\|\left(\sqrt{\rho_{AB}}\left[P^k_{AB} - \Xi(P^k_{AB})\right]\sqrt{\rho_{AB}}\right)\right\|_1 \tag{4.70}$$

$$= \sum_k \left\|\left(\sqrt{\rho_{AB}}\left[\Pi P^k_{AB}\overline{\Pi} + \overline{\Pi}P^k_{AB}\Pi\right]\sqrt{\rho_{AB}}\right)\right\|_1. \tag{4.71}$$

In keeping with our previous notation, we define $\rho^\Pi = \frac{\Pi\rho\Pi}{\mathrm{Tr}(\rho\Pi)}$ and $\rho^{\overline{\Pi}} = \frac{\overline{\Pi}\rho\overline{\Pi}}{\mathrm{Tr}\left(\rho\overline{\Pi}\right)}$, as well as the probability distributions $r(k) = \mathrm{Tr}\left(\rho^\Pi P_k\right)$ and $s(k) = \mathrm{Tr}\left(\rho^{\overline{\Pi}} P_k\right)$. By Lemma 4 we

have

$$\sum_k \left\| \left(\sqrt{\rho_{AB}} \left[\Pi P_{AB}^k \bar{\Pi} + \bar{\Pi} P_{AB}^k \Pi \right] \sqrt{\rho_{AB}} \right) \right\|_1 \leq \sum_k (r(k) + s(k)) \sqrt{W} \left\| \sqrt{A_k}^g B_k \sqrt{D_k}^g \right\|_\infty \tag{4.72}$$

$$= \sqrt{W} \left(\sum_k r(k) \left\| \sqrt{A_k}^g B_k \sqrt{D_k}^g \right\|_\infty + \sum_k s(k) \left\| \sqrt{A_k}^g B_k \sqrt{D_k}^g \right\|_\infty \right) \tag{4.73}$$

$$\leq \sqrt{W} \left(\max_k \left\| \sqrt{A_k}^g B_k \sqrt{D_k}^g \right\|_\infty + \max_k \left\| \sqrt{A_k}^g B_k \sqrt{D_k}^g \right\|_\infty \right) \tag{4.74}$$

$$= 2\sqrt{W} \left(\max_k \left\| \sqrt{A_k}^g B_k \sqrt{D_k}^g \right\|_\infty \right), \tag{4.75}$$

where in Eq. (4.73) we used the fact that the sums over $r(k)$ and over $s(k)$ are both convex combinations of $\left\| \sqrt{A_k}^g B_k \sqrt{D_k}^g \right\|_\infty$, and hence upper-bounded by the largest of these terms. Re-inserting this bound in Eq. (4.71), we have

$$\| \Phi(\rho_{ABE}) - \Phi(\Xi(\rho_{ABE})) \|_1 \leq 2\sqrt{W} \left(\max_k \left\| \sqrt{A_k}^g B_k \sqrt{D_k}^g \right\|_\infty \right) \tag{4.76}$$

or

$$\frac{1}{2} \| \Phi(\rho_{ABE}) - \Phi(\Xi(\rho_{ABE})) \|_1 \leq \sqrt{W} \left(\max_k \left\| \sqrt{A_k}^g B_k \sqrt{D_k}^g \right\|_\infty \right). \tag{4.77}$$

Letting $c = \sqrt{W} \left(\max_k \left\| \sqrt{A_k}^g B_k \sqrt{D_k}^g \right\|_\infty \right)$, by the continuity bound in Lemma 2, we have

$$H(Z|[E])_{\Phi(\Xi(\rho_{ABE}))} - f(\rho_{AB}) \leq c \log_2 |Z| + (1+c)\, h\left(\frac{c}{1+c} \right). \tag{4.78}$$

By Lemma 3, and the definition of UCDUP (Eq. (4.3)), the theorem statement follows. □

4.2.4 Finite Set $\mathbf{S}_N$

Recall that the feasible set $\mathbf{S}_\infty$ for the infinite-dimensional optimization is given in Eq. (3.10). The form of this set is common to all protocols. To define the expanded set, the approach we take is to individually expand each constraint of $\mathbf{S}_\infty$, using W as a parameter. Note that in order to have the highest key rates, we want $\mathbf{S}_N$ to be as small as possible, while still satisfying the containment $\mathbf{S}_N \supseteq \Pi \mathbf{S}_\infty \Pi$.

It is an interesting question whether one can do better than this. Indeed, one might expect that given $\mathbf{S}_\infty$ and Π, an analytic formula for $\Pi\mathbf{S}_\infty\Pi$ exists. However, we are not aware of any such result, and hence resort to the heuristics in this section.[6]

Assuming $\sigma \in \mathbf{S}_\infty$, we will now list constraints that $\tilde{\sigma}^\Pi \equiv \Pi\sigma\Pi$ must satisfy. Generally speaking, there are multiple options. We try to present them in order of the tightest ones requiring the most assumptions first and the looser but more general ones last.

Trace Constraint

We begin with the trace constraint, $\mathrm{Tr}(\sigma) = 1$, which can be easily expanded. By the definition of W in Eq. (4.15), $1 - W \leq \mathrm{Tr}\big(\tilde{\sigma}^\Pi\big) \leq 1$.

Expectation Constraints

We next consider the expectation constraints, $\mathrm{Tr}(\sigma\Gamma_i) = \gamma_i$. We choose to define the loosened constraints as $\gamma_i^{min} \leq \mathrm{Tr}\big(\tilde{\sigma}^\Pi\Gamma_i\big) \leq \gamma_i^{max}$, where

$$\gamma_i^{max} = \sup_{\rho\in\mathbf{S}_\infty} \mathrm{Tr}(\Gamma_i\ \Pi\rho\Pi), \tag{4.79}$$

$$\gamma_i^{min} = \inf_{\rho\in\mathbf{S}_\infty} \mathrm{Tr}(\Gamma_i\ \Pi\rho\Pi). \tag{4.80}$$

That is, we do not change the constraint operators Γ_i. We now find bounds on γ_i^{max} and γ_i^{min} in two different cases.

We first consider the case where $[\Pi, \Gamma_i] = 0$ and $\Gamma_i \geq 0\ \forall i$. This condition is satisfied for the protocols we study in this thesis. We emphasize that this is not a particularly strong assumption. With a judicious choice of Π, this condition can be achieved for many protocols. In the following theorem, we derive the desired bounds on expectations in the finite subspace.

Theorem 5. *Let* $\Gamma_i \geq 0$ *and* $[\Pi, \Gamma_i] = 0$. *If* $\mathrm{Tr}\big(\rho\overline{\Pi}\big) \leq W$ *and* $\mathrm{Tr}(\rho\Gamma_i) = \gamma_i$, *then*

$$\gamma_i - W\|\Gamma_i\|_\infty \leq \mathrm{Tr}(\Pi\rho\Pi\ \Gamma_i) \leq \gamma_i. \tag{4.81}$$

[6] The problem is, given the constraints specifying a spectrahedron and a projection, find the constraints defining the projection of this spectrahedron, i.e. its *shadow*. Solving this problem for polyhedra is easy. One first enumerates the extreme points of the polyhedron; noting that the polyhedron is the convex hull of its extreme points. Then, the shadow of the polyhedron is just the convex hull of the projections of the extreme points. In contrast, a literature search suggests this problem is difficult to solve for an arbitrary spectrahedron.

Proof. By the commutation relation, it follows that

$$\operatorname{Tr}(\Pi\rho\Pi\Gamma_i) = \operatorname{Tr}\Big(\sqrt{\Gamma_i}\Pi\rho\Pi\sqrt{\Gamma_i}\Big) \tag{4.82}$$

$$= \operatorname{Tr}\Big(\Pi\sqrt{\Gamma_i}\rho\sqrt{\Gamma_i}\Pi\Big). \tag{4.83}$$

We now find the upper and lower bounds separately.

For the upper bound, we simply note that the trace of a positive operator can only decrease under projection. Then,

$$\operatorname{Tr}\Big(\Pi\sqrt{\Gamma_i}\rho\sqrt{\Gamma_i}\Pi\Big) \leq \operatorname{Tr}\Big(\sqrt{\Gamma_i}\rho\sqrt{\Gamma_i}\Big) \tag{4.84}$$

$$= \gamma_i. \tag{4.85}$$

For the lower bound,

$$\operatorname{Tr}\Big(\Pi\sqrt{\Gamma_i}\rho\sqrt{\Gamma_i}\Pi\Big) = \operatorname{Tr}\Big(\sqrt{\Gamma_i}\rho\sqrt{\Gamma_i}\Big) - \operatorname{Tr}\Big(\overline{\Pi}\sqrt{\Gamma_i}\rho\sqrt{\Gamma_i}\overline{\Pi}\Big) \tag{4.86}$$

$$= \gamma_i - \operatorname{Tr}\big(\overline{\Pi}\rho\overline{\Pi}\ \Gamma_i\big) \tag{4.87}$$

$$\geq \gamma_i - \big\|\overline{\Pi}\rho\overline{\Pi}\big\|_1 \|\Gamma_i\|_\infty \tag{4.88}$$

$$\geq \gamma_i - W\|\Gamma_i\|_\infty, \tag{4.89}$$

where in Eq. (4.88) we used Hölder's inequality (Eq. (2.13)). Note that this lower bound is trivial for unbounded observables. □

We next consider the general case without any assumptions on Γ_i. While the following theorem is more general, it is not as tight as the preceding one in the cases where both can be applied.

Theorem 6. *Let Γ_i be self-adjoint. If $\operatorname{Tr}\big(\rho\overline{\Pi}\big) \leq W$ and $\operatorname{Tr}(\rho\Gamma_i) = \gamma_i$, then*

$$\gamma_i - 2\sqrt{W}\|\Gamma_i\|_\infty \leq \operatorname{Tr}(\Pi\rho\Pi\ \Gamma_i) \leq \gamma_i + 2\sqrt{W}\|\Gamma_i\|_\infty. \tag{4.90}$$

Proof. We have

$$|\gamma_i - \operatorname{Tr}(\Pi\rho\Pi\ \Gamma_i)| = |\operatorname{Tr}(\rho\Gamma_i) - \operatorname{Tr}(\Pi\rho\Pi\ \Gamma_i)| \tag{4.91}$$

$$= |\operatorname{Tr}[(\rho - \Pi\rho\Pi)\Gamma_i]| \tag{4.92}$$

$$\leq \|\rho - \Pi\rho\Pi\|_1 \|\Gamma_i\|_\infty \tag{4.93}$$

$$\leq 2\sqrt{W}\|\Gamma_i\|_\infty, \tag{4.94}$$

where in Eq. (4.93) we used Hölder's inequality (Eq. (2.13)) and in Eq. (4.94) we used Eq. (2.29). □

The previous two theorems seem to suggest that it is not possible to derive loosened constraints when Γ_i are unbounded observables. This is not the case. It is indeed possible to bound the change in expectation under projection for unbounded, non-commuting, and non-positive observables Γ_i. However, doing so seems to be very specific to the set of observables considered, and hence it is difficult to give an overarching result. See Ref. [24] for an example of loosened bounds derived for the specific unbounded observables $\hat{X}$ and $\hat{P}$ under a projection in the Fock basis.

Reduced State Constraint

Finally we consider the reduced state constraint $\mathrm{Tr}_B(\sigma) = \tau_A$. There are (at least) three different ways to expand this constraint. It is not clear which constraint is the tightest, especially when considered together with the other constraints in the optimization. From a practical perspective, one can simply use all three constraints simultaneously, ensuring the smallest $\mathbf{S}_N$ and highest key rate.

The first loosened constraint uses a complete Hermitian basis $\{H_k\}$ on system A. This basis can always be chosen so that each H_k is positive and $\|H_k\|_\infty = 1$. The original constraint can then be expressed as $\mathrm{Tr}((H_k \otimes \mathbb{1}_B)\sigma) = \mathrm{Tr}(H_k\tau_A)$, and expanded as an expectation constraint in the previous section. If the projection is of the form $\Pi = \mathbb{1}_A \otimes \Pi_B$, then each $H_k \otimes \mathbb{1}_B$ commutes with Π, so the tighter expansion from Theorem 5 can be used. If some $H_k \otimes \mathbb{1}_B$ do not commute with Π, then Theorem 6 must be used which means this method of expansion is looser than the third method discussed below.[7] For this reason, we can think of this first method as really only being useful when $\Pi = \mathbb{1}_A \otimes \Pi_B$.

The second method applies when the projection has the form $\Pi = \Pi_A \otimes \Pi_B$. We have

$$\begin{aligned}
\Pi_A \,\mathrm{Tr}_B(\sigma)\Pi_A &= \mathrm{Tr}_B\left[(\Pi_A \otimes \mathbb{1}_B)\sigma(\Pi_A \otimes \mathbb{1}_B)\right] && (4.95)\\
&= \mathrm{Tr}_B\left[(\Pi_A \otimes [\Pi_B + \overline{\Pi}_B])\sigma(\Pi_A \otimes [\Pi_B + \overline{\Pi}_B])\right] && (4.96)\\
&= \mathrm{Tr}_B\left[(\Pi_A \otimes \Pi_B)\sigma(\Pi_A \otimes \Pi_B)\right] + \mathrm{Tr}_B\left[(\Pi_A \otimes \overline{\Pi}_B)\sigma(\Pi_A \otimes \overline{\Pi}_B)\right] && (4.97)\\
&\geq \mathrm{Tr}_B\left[(\Pi_A \otimes \Pi_B)\sigma(\Pi_A \otimes \Pi_B)\right] && (4.98)\\
&= \mathrm{Tr}_B(\tilde{\sigma}^{\Pi}), && (4.99)
\end{aligned}$$

where in Eq. (4.97) we used the fact that partial trace is cyclic over the subsystem being traced out. The constraint is thus $\mathrm{Tr}_B(\tilde{\sigma}^{\Pi}) \leq \Pi_A\tau_A\Pi_A$. Note that this constraint can be

[7] This is because a bound of $2\sqrt{W}$ on each element of the matrix is worse than a bound of $2\sqrt{W}$ on the trace distance of the matrix.

used even when τ_A is infinite-dimensional. When $\Pi = \mathbb{1}_A \otimes \Pi_B$, this loosened constraint is implied by the first one. At first glance, this constraint seems useless since it is always satisfied by the zero matrix. However, there are other constraints in the optimization, namely one which lower bounds the trace. Taking this into account, it is difficult to determine a priori which reduced state constraint expansion is best.

The third method is to use a trace distance argument. By Eq. (2.29), $\frac{1}{2}\left\|\sigma - \tilde{\sigma}^{\Pi}\right\|_1 \leq \sqrt{W}$. By the monotonicity of trace distance under channels, and by the fact that taking the partial trace is a channel, this implies the constraint $\frac{1}{2}\left\|\tau_A - \mathrm{Tr}_B(\tilde{\sigma}^{\Pi})\right\|_1 \leq \sqrt{W}$.

Summary

For a generic protocol, one may use combinations of all the expanded constraints listed above. By construction, $\mathbf{S}_N$ will be a convex subset of $\tilde{D}(\mathcal{H}_N)$ containing $\Pi\mathbf{S}_\infty\Pi$. The tightness of the feasible set is guaranteed by the fact that for $W = 0$, $\mathbf{S}_N = \mathbf{S}_\infty$.

We remark that there may be even better methods to define $\mathbf{S}_N$ by using specific details of the protocol being considered. In practice however, the results we obtain, especially in comparison to the flag-state squasher (see Sec. 6.2), provide strong evidence that our choice of $\mathbf{S}_N$ is close to the optimal set $\Pi\mathbf{S}_\infty\Pi$.

In this thesis, we will focus on the case where the observables are positive operators and commute with the projection. For ease of reference, we summarize the definitions of the original infinite-dimensional optimization:

$$\begin{aligned} \underset{\rho}{\text{minimize:}} \quad & f(\rho) \\ \text{subject to:} \quad & \mathrm{Tr}(\rho) = 1 \\ & \mathrm{Tr}_B(\rho) = \tau_A \\ & \mathrm{Tr}(\rho\Gamma_i) = \gamma_i \\ & \rho \in \mathrm{Pos}\left(\mathcal{H}_\infty\right), \end{aligned} \tag{4.100}$$

and the expanded finite-dimensional optimization:

$$\begin{aligned} \underset{\tilde{\rho}}{\text{minimize:}} \quad & f(\tilde{\rho}) \\ \text{subject to:} \quad & 1 - W \leq \mathrm{Tr}(\tilde{\rho}) \leq 1 \\ & \frac{1}{2}\|\mathrm{Tr}_B(\tilde{\rho}) - \tau_A\|_1 \leq \sqrt{W} \\ & \gamma_i - W\|\Gamma_i\|_\infty \leq \mathrm{Tr}(\tilde{\rho}\Gamma_i) \leq \gamma_i \\ & \tilde{\rho} \in \mathrm{Pos}\left(\mathcal{H}_N\right). \end{aligned} \tag{4.101}$$

When $\Pi = \Pi_A \otimes \Pi_B$, the constraint on the reduced density matrix is instead $\mathrm{Tr}_B(\tilde{\rho}^{\Pi}) \leq \Pi_A \tau_A \Pi_A$.

Note that the constraint operators Γ_i and the POVM elements P_k defining f are themselves infinite-dimensional. However, as $\mathrm{Tr}(\Pi\rho\Pi\ X) = \mathrm{Tr}[(\Pi\rho\Pi)(\Pi X \Pi)]$, we can equivalently set $\Gamma_i \to \Pi\Gamma_i\Pi$ and $P_k \to \Pi P_k \Pi$. We tacitly assume this substitution is made in order to represent the optimizations numerically.

Chapter 5

Discrete-Modulated Continuous-Variable QKD

Continuous-variable QKD emerged as an alternative to discrete-variable protocols [11]. Instead of relying on single-photon detectors, which can be expensive, require low operating temperatures, or have issues with dark counts and low repetition rates, CVQKD utilizes homodyne or heterodyne detectors.[1] As these detectors can be integrated into existing telecommunication infrastructure, CVQKD protocols are promising candidates for deployment in large-scale quantum-secured networks.

Our interest is in a particular class of prepare-and-measure CVQKD protocols. These protocols are characterized by Alice sending coherent states to Bob, who then performs a homodyne or heterodyne measurement. Information is encoded in the location of the states in phase space. If Alice samples the coherent state amplitudes according to a Gaussian probability distribution over phase space, it is referred to as *Gaussian modulation* (GM) [14–16]. If she instead chooses the amplitudes from a finite set, it is referred to as *discrete modulation* (DM) [11–13].

In comparison to GM, DM is less demanding on the source modulator and on the error-correction protocols, yet is expected to achieve similar key rates. This makes it an enticing protocol to study.

The security analysis for GM is significantly more mature than for DM. This is because the optimal attack for GM, under typical observed statistics, is also Gaussian [65, 66]. Gaussian states have a finite-dimensional representation in terms of covariance matrices,

[1] In this context, heterodyne refers to conjugate homodyne detection.

and this makes it easier to handle the security proofs. Key rates for GMCVQKD have even been calculated in the finite-key regime [67, 68].

This chapter is organized as follows. In Secs. 5.1 and 5.2, we review the DMCVQKD protocol. After setting up the infinite-dimensional optimization in Secs. 5.3 and 5.4, we apply the dimension reduction method in Sec. 5.5. In Sec. 5.6 we consider some representative numerical results, and in Sec. 5.7 we explore different strategies to improve the protocol's performance.

5.1 Protocol Description

Strictly speaking, DMCVQKD is a blanket term for a family of protocols with variations in the signal states, measurement devices, and classical postprocessing steps.

The steps of a generic protocol are as follows. In each round of the quantum phase, Alice chooses a coherent signal state from the *constellation* $\{|\alpha_i\rangle\}_{i=0}^{d-1}$ with probability $p(i)$. She sends the state to Bob, who then performs a homodyne or heterodyne measurement. Alice or Bob then performs the key map, which may include postselection. Finally, they perform the usual steps of error correction and privacy amplification. As discussed in Sec. 3.1.2, we can recast this prepare-and-measure protocol as an entanglement-based one. Then, we think of Alice preparing the state

$$\sum_{ij} \sqrt{p(i)p(j)}\, |i\rangle\langle j|_A \otimes |\alpha_i\rangle\langle\alpha_j|_{A'}\,, \tag{5.1}$$

sending A' to Bob, and measuring with the POVM $\{|i\rangle\langle i|_A\}_{i=0}^{d-1}$.

In this thesis, we will focus on the case where Bob performs a heterodyne measurement. We will also focus on Bob performing the key map (see Sec. 3.1.1), also known as reverse reconciliation [69], as it outperforms direct reconciliation at long distances. The security proof we will present can easily be applied to direct reconciliation as well. However, the generalization to homodyne detection would be more involved. The result of a heterodyne measurement is two real numbers, which we represent as a complex number ζ. To perform the key map (see Sec. 3.1.1), Bob partitions the complex plane into regions A_z, where $z \in \{0, 1, ..., M-1, \perp\} = S_Z$. Then, the outcome ζ is assigned to the key symbol z when ζ lies in the region A_z. Recall that $\perp$ corresponds to postselection, i.e. signals that are discarded.

Our security proof works for any set of coherent signal states, postselection region, and key map. Indeed, most of our following results will be presented in this generality. Later,

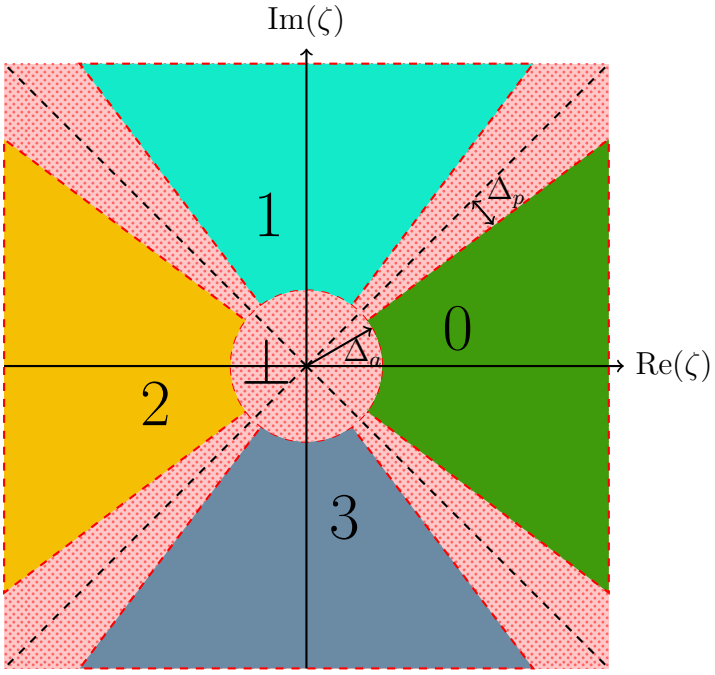


Figure 5.1: Phase space regions for the key map, with amplitude and phase postselection, for reverse reconciliation in quadrature phase-shift keying ($M = 4$) DMCVQKD. Bob obtains the measurement result ζ from the heterodyne detector. He maps the outcome to the symbol of the region containing ζ. Δ_a and Δ_p are amplitude and phase postselection parameters, respectively. The $\perp$ region corresponds to signals that are discarded.

we will primarily focus on the specific case of *phase-shift keying*. In this case, the signal states are chosen uniformly at random from $\{|\alpha e^{2\pi ik/M}\rangle\}_{k=0}^{M-1}$ for a positive signal state amplitude α; and each A_z is a slice of pie in the complex plane from $\theta = \frac{2\pi(z-1/2)}{M}, \frac{2\pi(z+1/2)}{M}]$ (not accounting for postselection). The key map is visualized in Fig. 5.1, for the specific case $M = 4$, which is known as *quadrature phase-shift keying*, and a specific postselection pattern.

5.1.1 Bob's Measurement

Up till now, we have been deliberately vague about Bob's measurement procedure, as there are a number of details to consider.

Ideal Detectors and Trusted Noise

We will actually consider two different ways of modelling the heterodyne detection.

The first model treats the entire heterodyne apparatus as ideal. Bob's POVM is then a (scaled) projection onto the coherent states $\{\frac{1}{\pi}|\zeta\rangle\langle\zeta|\}_{\zeta\in\mathbb{C}}$.

The second model, which encompasses the first, is the *trusted noise* scenario [70, 71]. It is a particular way to model imperfections in the heterodyne apparatus which is commonly used in CVQKD. Namely, the two homodyne detectors comprising the heterodyne measurement have an associated efficiency and electronic noise. It is an important aspect of our security proof that it extends to this scenario where Bob has imperfect but characterized detectors. The security of DMCVQKD under trusted noise has been investigated in Ref. [72] under the photon-number cutoff assumption.

To illustrate our approach, we focus on the case where the two detectors have the same efficiency η_d and electronic noise ν_{el}. The POVM is then a (scaled) projection on displaced thermal states $\left\{\frac{1}{\eta_d\pi}\hat{D}\left(\frac{\zeta}{\sqrt{\eta_d}}\right)\rho_{th}(\bar{n})\hat{D}^\dagger\left(\frac{\zeta}{\sqrt{\eta_d}}\right)\right\}_{\zeta\in\mathbb{C}}$ where the mean photon number of the thermal state is $\bar{n}=\frac{1-\eta_d+\nu_{el}}{\eta_d}$ [72]. We refer the reader to Ref. [72] for a derivation of this POVM from the optical model for the apparatus. Setting $\eta_d=1$, $\nu_{el}=0$ recovers the ideal detector scenario, where the projection is onto displaced vacuum states.

Generically, we will use $\{G_\zeta\}_{\zeta\in\mathbb{C}}$ to denote Bob's POVM. Unless otherwise specified, it should be assumed that all discussion is in the ideal detector case.

Coarse-Graining of Measurement Result

The result of each heterodyne measurement can be represented as a complex number. However, the numerics framework cannot directly handle a register holding a continuous outcome in the definition of the postprocessing map (Eq. (3.5)) for the objective function (Eq. (3.9)).[2] Thus, we instead work with a coarse-graining of Bob's measurement for the key map.

Recall that for the key map, Bob assigns the measurement result ζ to the key symbol $z\in S_Z$ when ζ lies in the region A_z of the complex plane. This notion is effectively captured

[2]It might be possible to consider a discretization or binning of the heterodyne outcome. One would have to find a middle ground between the dimension the numerics framework can handle, and the fineness of the binning.

by defining the coarse-grained POVM

$$R^z = \int_{A_z} G_\zeta \, d^2\zeta. \tag{5.2}$$

The POVM elements R^z are referred to as *region operators*. It suffices to only use this coarse-grained POVM in the key rate objective function.

Similarly for parameter estimation, it is not clear how to represent the constraints arising from all the fine-grained measurement results. One approach might be to note that in the ideal detector case, and in the asymptotic limit, Bob's measurement determines the Husimi Q-function. Then, he could perform complete tomography to uniquely determine the received conditional states. To determine the key rate, it would then only remain to optimize over the off-diagonal blocks of ρ_{AB}. However, such a security proof would almost certainly not be suitable for a finite-key analysis. This is because this type of security proof would seemingly only work in the finite regime if a large fraction of rounds were used for parameter estimation, in order to perform approximate tomography. Since only a small fraction of rounds would remain for key generation, the key rate would likely be poor; compared to if fewer rounds were used for parameter estimation along with a different security proof technique.

So for the constraints, we again assume Bob performs a coarse-graining of his data, which translates to him determining the expectations of some finite set of observables. Note the coarse-graining is just a classical step after the physical measurement, so we have the freedom to perform different coarse-grainings after the fact, as long as the fine-grained data is stored in the interim.

In particular, we will consider coarse-grained observables which depend on the signal state that was sent. We re-emphasize that this is easy to physically realize. Bob simply stores all the fine-grained data from his measurements during the quantum phase of the protocol. During parameter estimation, Alice announces which signal state was sent on the testing rounds. For each signal state, Bob considers the data from the corresponding rounds only, and coarse grains it accordingly. In Appendix E, we show in detail how to calculate the coarse-grained expectations from the fine-grained measurement data.

By no means is the choice of observables unique. In Sec. 5.3, we discuss how we settled on the particular choice of observables used.

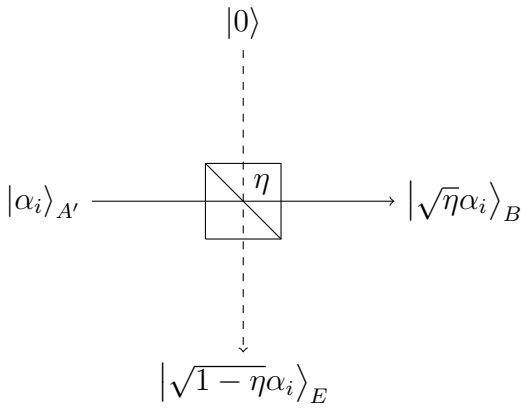


Figure 5.2: Schematic of beam splitter implementing the loss-only Gaussian channel.

5.2 Generalized Beam Splitting Attack

The idea behind the *generalized beam splitting attack*, discovered in Ref. [73], is that under certain circumstances Bob and Eve's conditional states decouple and it is possible to determine Eve's conditional states. In particular, suppose Alice and Bob share a Gaussian channel with loss η. This is equivalent to modelling the signal as passing through a beam splitter with a vacuum state at the second input port (Fig. 5.2). They can determine that the channel is indeed (unitarily equivalent to) a pure-loss channel, only given that they have *statistics* compatible with said channel. In other words, ρ_{ABE} is known (up to an irrelevant unitary on Eve's system). In still other words, the key rate can be determined analytically by simply evaluating the objective function on this state. This fact is very important in guiding our intuition in later sections. For this reason, before further discussion of the security proof for DMCVQKD, it is instructive to elucidate some key details of this generalized beam splitting scenario. Bob is assumed to have ideal detectors, but we will show in Sec. 5.3.2 that the reasoning of this section extends to the trusted noise scenario.

Under this channel, the joint state is $\rho_{ABE} = |\Psi\rangle\langle\Psi|$ with

$$|\Psi\rangle_{ABE} = \sum_i \sqrt{p(i)}\,|i\rangle_A\,|\sqrt{\eta}\alpha_i\rangle_B\left|\sqrt{1-\eta}\alpha_i\right\rangle_E. \tag{5.3}$$

Pretending that the state is unknown, let's see how Alice and Bob can determine that this is indeed the tripartite state.

Recall that in the source-replacement picture, Alice prepares the state

$$|\Psi\rangle_{AA'} = \sum_i \sqrt{p(i)}\,|i\rangle_A\,|\alpha_i\rangle_{A'} \tag{5.4}$$

and sends A' to Bob. Bob determines the expectations of $\hat{X}$, $\hat{P}$, $\hat{X}^2$ and $\hat{P}^2$ for each conditional state. He will find that these states saturate Heisenberg's uncertainty principle [74],

$$(\Delta X)(\Delta P) \geq \frac{1}{2}, \tag{5.5}$$

where

$$\Delta X \equiv \sqrt{\left\langle \hat{X}^2 \right\rangle - \left\langle \hat{X} \right\rangle^2}, \tag{5.6}$$

$$\Delta P \equiv \sqrt{\left\langle \hat{P}^2 \right\rangle - \left\langle \hat{P} \right\rangle^2}. \tag{5.7}$$

These so-called minimum uncertainty states can only be squeezed or coherent states [75]. These states are uniquely determined by their first and second moments. In particular, Bob will observe that ΔX and ΔP are the same, so he knows his conditional state is a coherent state. Because he knows $\left\langle \hat{X} \right\rangle$ and $\left\langle \hat{P} \right\rangle$, he also knows where the coherent state is centered. In summary, Bob will know that his conditional states are $\left|\sqrt{\eta}\alpha_i\right\rangle$.

In the Stinespring representation, the channel action is some isometry $V : A' \to BE$, taking the signal states $|\alpha_i\rangle_{A'}$ to $|\phi_i\rangle_{BE} = V\,|\alpha_i\rangle_{A'}$. Then,

$$|\Psi\rangle_{ABE} = (\mathbb{1}_A \otimes V)\,|\Psi\rangle_{AA'} \tag{5.8}$$

$$= \sum_i \sqrt{p(i)}\,|i\rangle_A\,|\phi_i\rangle_{BE}\,. \tag{5.9}$$

By the Schmidt decomposition, write

$$|\phi_i\rangle_{BE} = \sum_{r=0}^{R_i} \lambda_i^r\,|\theta_i^r\rangle_B\,|\omega_i^r\rangle_E\,. \tag{5.10}$$

But we know that $\mathrm{Tr}_E(|\phi_i\rangle\langle\phi_i|_{BE}) = \left|\sqrt{\eta}\alpha_i\right\rangle\!\left\langle\sqrt{\eta}\alpha_i\right|_B$. It follows immediately that $R_i = 1\ \forall i$. In words, the only way the partial trace of a bipartite pure state is pure, is if the bipartite state is a product state. We can thus write

$$|\phi_i\rangle_{BE} = |\sqrt{\eta}\alpha_i\rangle_B\,|\omega_i\rangle_E\,. \tag{5.11}$$

This allows us to express the overlap of the signal states as

$$\langle \phi_j | \phi_i \rangle = \langle \sqrt{\eta}\alpha_j | \sqrt{\eta}\alpha_i \rangle \langle \omega_j | \omega_i \rangle \tag{5.12}$$

$$= \exp\left[\eta \left(i \operatorname{Im}(\alpha_i \alpha_j^*) - \frac{1}{2}|\alpha_i - \alpha_j|^2 \right)\right] \langle \omega_j | \omega_i \rangle , \tag{5.13}$$

(see Eq. (2.48) for the formula for the overlap of two coherent states). But on the other hand, we also know the inner product of the signal states as prepared by Alice

$$\langle \phi_j | \phi_i \rangle = \langle \alpha_j | V^\dagger V | \alpha_i \rangle \tag{5.14}$$

$$= \langle \alpha_j | \alpha_i \rangle \tag{5.15}$$

$$= \exp\left(i \operatorname{Im}(\alpha_i \alpha_j^*) - \frac{1}{2}|\alpha_i - \alpha_j|^2 \right). \tag{5.16}$$

Equating Eq. (5.13) and Eq. (5.16), we have

$$\exp\left[\eta \left(i \operatorname{Im}(\alpha_i \alpha_j^*) - \frac{1}{2}|\alpha_i - \alpha_j|^2 \right)\right] \langle \omega_j | \omega_i \rangle = \exp\left(i \operatorname{Im}(\alpha_i \alpha_j^*) - \frac{1}{2}|\alpha_i - \alpha_j|^2 \right) \tag{5.17}$$

$$\langle \omega_j | \omega_i \rangle = \exp\left[(1-\eta) \left(i \operatorname{Im}(\alpha_i \alpha_j^*) - \frac{1}{2}|\alpha_i - \alpha_j|^2 \right)\right]. \tag{5.18}$$

Thus, up to an irrelevant unitary on her system, Eve's conditional states are $|\omega_i\rangle_E = \left|\sqrt{1-\eta}\alpha_i\right\rangle_E$. We now know ρ_{ABE}, so the key rate can be determined analytically (for any key map, postselection choice, and other classical postprocessing steps).

There are two important takeaways from this example. The first is that ρ_{ABE}, and hence the key rate, is completely determined. The second is that Bob did not have to perform a complete tomographic measurement to determine his conditional states; they were known from the expectations of a small number of observables.

5.3 Observables

As discussed, we need to choose a set of coarse-grained observables for Bob's side of parameter estimation. Typically, these observables are taken to be the quadratures, $\hat{X}$ and $\hat{P}$, and their higher moments [23]. We introduce a new set of observables. These observables are built to capture the intuition behind the generalized beam splitting attack.

5.3.1 Ideal Detector

Let's first discuss the ideal detector scenario. For any operator $\hat{O}$, recall we use the shorthand notation $\hat{O}_\gamma \equiv \hat{D}(\gamma)\hat{O}\hat{D}^\dagger(\gamma)$ where $\hat{D}(\gamma)$ is the displacement operator with complex parameter γ.

Consider the displacements[3]

$$\beta_i = \sqrt{\eta}\alpha_i \qquad i = 0, ..., d-1. \tag{5.19}$$

Denoting the photon number operator by $\hat{n} \equiv \hat{a}^\dagger\hat{a}$, consider the observable $\hat{n}_{\beta_i}$. It is easy to see that the vacuum state is the only state for which $\langle\hat{n}\rangle = 0$. Since coherent states are just displaced vacuum states, it is immediate that $|\beta_i\rangle$ is the only state for which the displaced expectation value $\langle\hat{n}_{\beta_i}\rangle = 0$. Thus, for the i^{th} conditional state, if Bob only determines $\langle\hat{n}_{\beta_i}\rangle$, he can still confirm if the channel is loss-only, and apply the reasoning of the generalized beam splitting attack.

Another way to put it is that originally, we considered the expectations of four different observables, $\hat{X}$, $\hat{P}$, $\hat{X}^2$ and $\hat{P}^2$, in order to uniquely determine the conditional state. Now, we can do it with just one. Note also that $\hat{n}_{\beta_i}$ is a linear combination of the quadratures and their second moments, and in some sense we see that it is the right linear combination.

Intuitively, this observable measures how spread out a state is compared to the coherent state $|\beta_i\rangle$. We can think of this as characterizing the deviation from a generalized beam splitting attack.

The discussion so far has applied to the protocol itself and to its original infinite-dimensional formulation. This choice of coarse-grained observables is of independent interest as it elucidates some of the essential working principles of the DMCVQKD protocol.

But this choice of observables also has important implications for the finite-dimensional formulation via the dimension reduction method. Recall that in the dimension reduction method, one step is to bound the weight outside the subspace. The observables we choose here will have an important role to play in that capacity as well. Unfortunately, it turns out that just having $\hat{n}_{\beta_i}$ is insufficient to get W and the correction term $\Delta(W)$ small enough. By small enough, we mean that the correction term is a small fraction of the key rate. Hence, we add in a second-order constraint $\hat{n}^2_{\beta_i}$.

[3]It may seem that η and α_i have to be known exactly. This is not the case. This particular choice of $\{\beta_i\}$ is based on the expected channel behavior in an honest implementation of the protocol. We emphasize however that our security proof method works for any choice of $\{\beta_i\}$. For example, in an experiment where the parameters are not known exactly, a good guess for the displacements will suffice. See Appendix E for a detailed discussion of how this security proof for DMCVQKD is applied with data from an experiment.

Finally, recall that one consideration in choosing the projection is to have it commute with the observables. The observables we have chosen here will commute with the projection we later define (Eq. (5.65)), dovetailing with a natural choice for the finite subspace.

Adding in Alice's POVM from the source-replacement scheme, the overall observables for parameter estimation are

$$\{\Gamma_i\} = \{|i\rangle\langle i| \otimes \hat{n}_{\beta_i}, |i\rangle\langle i| \otimes \hat{n}^2_{\beta_i}\}_{i=0}^{d-1}. \tag{5.20}$$

5.3.2 Trusted Noise

We now turn our attention to the trusted noise scenario. It turns out that, for our specific choice of observables in Eq. (5.20), there is a simple relationship between the ideal observables we've defined and their noisy counterparts. We begin with some general background, and then proceed to analyze the specific observables for DMCVQKD.

Definition of Observables

Recall that an observable is simply a weighted linear combination of some POVM elements, where the weights represent a value which is associated with each POVM element (see Sec. 2.1.6).

Then, an observable $\hat{O}$ in the ideal detector scenario is just a weighted linear combination of the ideal POVM elements,

$$\hat{O} = \int_{\zeta \in \mathbb{C}} w_{\hat{O}}(\zeta) \, \frac{1}{\pi} \, |\zeta\rangle\langle\zeta| \; d^2\zeta. \tag{5.21}$$

Here $w_{\hat{O}}$ is a function representing the weight associated to the outcome ζ.

Now consider the trusted noise scenario. If Bob performs the same coarse-graining, in the sense of mapping his raw measurement outcomes ζ to weights via the same function $w_{\hat{O}}(\zeta)$, then he is really measuring the observable

$$\left[\hat{O}\right]' = \int_{\zeta \in \mathbb{C}} w_{\hat{O}}(\zeta) \, G_\zeta \; d^2\zeta. \tag{5.22}$$

Here, we've introduced the notation $[\cdot]'$ for the noisy version of an observable.

Trusted Noise Observables

In the trusted noise scenario, we use the following coarse-grained observables

$$\{\Gamma_i\} = \{|i\rangle\langle i| \otimes \left[\hat{n}_{\sqrt{\eta_d}\beta_i}\right]', |i\rangle\langle i| \otimes \left[\hat{n}^2_{\sqrt{\eta_d}\beta_i}\right]'\}_{i=0}^{d-1}. \tag{5.23}$$

We now motivate why we consider this choice. Aside from the obvious analogy with the ideal detector scenario, another reason is that these observables turn out to be closely related to the ones defined in the ideal detector case, as we now prove. It is difficult to directly compare the expressions for the observables themselves. Instead, we consider the Husimi Q-function of the observables (see Sec. 2.4.3 for a definition of this quasiprobability). As the observables are uniquely specified by their Husimi Q-functions, it suffices to show that their Q-functions are equal.

Observable Weights

The first step is to determine the weights $w_{\hat{n}_{\beta_i}}$ and $w_{\hat{n}^2_{\beta_i}}$, so that we can compute the requisite integral. We begin by finding $w_{\hat{n}}$ and $w_{\hat{n}^2}$. A method to do this is found in Eq. (7.20) of Ref. [76]. We first write the operator in antinormal ordering and then replace the ladder operators $\hat{a}, \hat{a}^\dagger$ with ζ, ζ^*.

$$\hat{n} = \hat{a}\hat{a}^\dagger - 1 \implies w_{\hat{n}}(\zeta) = |\zeta|^2 - 1, \tag{5.24}$$

$$\hat{n}^2 = \hat{a}^2(\hat{a}^\dagger)^2 - 3\hat{n} - 2 \implies w_{\hat{n}^2}(\zeta) = |\zeta|^4 - 3|\zeta|^2 + 1. \tag{5.25}$$

To find w for the displaced observables, we simply perform a change of variables,

$$\hat{n}_\beta = \hat{D}(\beta)\,\hat{n}\hat{D}^\dagger(\beta) \tag{5.26}$$

$$= \hat{D}(\beta)\left(\frac{1}{\pi}\int (|\zeta|^2 - 1)\,|\zeta\rangle\langle\zeta|\,d^2\zeta\right)\hat{D}^\dagger(\beta) \tag{5.27}$$

$$= \frac{1}{\pi}\int (|\zeta|^2 - 1)\,|\zeta+\beta\rangle\langle\zeta+\beta|\,d^2\zeta \tag{5.28}$$

$$= \frac{1}{\pi}\int (|\zeta-\beta|^2 - 1)\,|\zeta\rangle\langle\zeta|\,d^2\zeta, \tag{5.29}$$

and similarly for $\hat{n}^2_\beta$. Thus,

$$w_{\hat{n}_\beta}(\zeta) = |\zeta-\beta|^2 - 1, \tag{5.30}$$

and

$$w_{\hat{n}^2_\beta}(\zeta) = |\zeta-\beta|^4 - 3|\zeta-\beta|^2 + 1. \tag{5.31}$$

Q-function of Noisy Observables

We can now calculate the noisy observables using Eq. (5.22). We simplify our notation by writing $\rho_{th} \equiv \rho_{th}(\bar{n})$. We make use of the following identity,

$$\langle\gamma|\rho_{th}(\bar{n})|\gamma\rangle = \frac{e^{-|\gamma|^2/(1+\bar{n})}}{1+\bar{n}}, \tag{5.32}$$

which can be easily verified by a calculation in the Fock basis.

By definition,

$$[\hat{n}_\beta]' = \int (|\zeta-\beta|^2 - 1) G_\zeta \; d^2\zeta \tag{5.33}$$

$$= \frac{\eta_d}{\pi} \int \left(\left| \zeta - \frac{\beta}{\sqrt{\eta_d}} \right|^2 - \frac{1}{\eta_d} \right) D(\zeta)\rho_{th}D^\dagger(\zeta) \; d^2\zeta. \tag{5.34}$$

Then,

$$\langle\alpha|\,[\hat{n}_\beta]'\,|\alpha\rangle = \frac{\eta_d}{\pi} \int \left(|\zeta-\beta'|^2 - \frac{1}{\eta_d} \right) \langle\alpha-\zeta|\,\rho_{th}\,|\alpha-\zeta\rangle \; d^2\zeta \tag{5.35}$$

$$= \frac{\eta_d}{\pi(1+\bar{n})} \int \left(|\zeta+\alpha-\beta'|^2 - \frac{1}{\eta_d} \right) e^{\frac{-|\zeta|^2}{1+\bar{n}}} d^2\zeta. \tag{5.36}$$

where $\beta' = \beta/\sqrt{\eta_d}$. Converting to polar coordinates, the integral is

$$\langle\alpha|\,[\hat{n}_\beta]'\,|\alpha\rangle = \frac{\eta_d}{\pi(1+\bar{n})} \int \left(r^2 + \gamma^* r e^{i\theta} + \gamma r e^{-i\theta} + |\gamma|^2 - \frac{1}{\eta_d} \right) \; e^{\frac{-r^2}{1+\bar{n}}} \; r \; dr \; d\theta \tag{5.37}$$

$$= \eta_d|\gamma|^2 + \nu_{el}, \tag{5.38}$$

where $\gamma = \alpha - \beta'$.

Similarly,

$$\left[\hat{n}_\beta^2\right]' = \frac{1}{\pi\eta_d} \int \left(|\zeta-\beta|^4 - 3|\zeta-\beta|^2 + 1 \right) \hat{D}\left(\frac{\zeta}{\sqrt{\eta_d}}\right) \rho_{th} \hat{D}^\dagger \left(\frac{\zeta}{\sqrt{\eta_d}}\right) d^2\zeta \tag{5.39}$$

$$= \frac{1}{\pi} \int \left(\eta_d^2|\zeta-\beta'|^4 - 3\eta_d|\zeta-\beta'|^2 + 1 \right) \hat{D}\left(\zeta\right) \rho_{th} \hat{D}^\dagger\left(\zeta\right) d^2\zeta. \tag{5.40}$$

The Q-function is then

$$\langle\alpha|\left[\hat{n}_\beta^2\right]'|\alpha\rangle = \frac{1}{\pi(1+\bar{n})}\int\left(\eta_d^2|\zeta+\alpha-\beta'|^4 - 3\eta_d|\zeta+\alpha-\beta'|^2+1\right)e^{\frac{-|\zeta|^2}{1+\bar{n}}}d^2\zeta \tag{5.41}$$

$$= \frac{1}{\pi(1+\bar{n})}\int\left(\eta_d^2(r^4+4r^2|\gamma|^2+|\gamma|^4) - 3\eta_d(r^2+|\gamma|^2)+1\right)e^{\frac{-r^2}{1+\bar{n}}}\ r\ dr\ d\theta \tag{5.42}$$

$$= 2\left(1+2\nu_{el}+\nu_{el}^2+2\eta_d(1+\nu_{el})|\gamma|^2 - 3\frac{1+\nu_{el}}{2}+\eta_d^2|\gamma|^4\frac{1}{2}-\frac{3}{2}\eta_d|\gamma|^2+\frac{1}{2}\right) \tag{5.43}$$

$$= \eta_d^2|\gamma|^4+\eta_d(4\nu_{el}+1)|\gamma|^2+2\nu_{el}^2+\nu_{el}. \tag{5.44}$$

Relationship between Ideal and Noisy Observables

We have now determined the Q-function of the noisy observables. The Q-functions of the ideal observables are easy to calculate. Namely,

$$\langle\alpha|\,\hat{n}_{\beta'}\,|\alpha\rangle = \langle\alpha-\beta'|\,\hat{n}\,|\alpha-\beta'\rangle \tag{5.45}$$

$$= |\alpha-\beta'|^2 \tag{5.46}$$

$$= |\gamma|^2, \tag{5.47}$$

and

$$\langle\alpha|\,\hat{n}_{\beta'}^2\,|\alpha\rangle = \langle\alpha-\beta'|\,\hat{n}^2\,|\alpha-\beta'\rangle \tag{5.48}$$

$$= |\alpha-\beta'|^4 \tag{5.49}$$

$$= |\gamma|^4. \tag{5.50}$$

Comparing the Q-functions of the noisy observables in Eqs. (5.38) and (5.44) to those of the ideal observables in Eqs. (5.47) and (5.50), and using the uniqueness of the Husimi Q-function, we have

$$[\hat{n}_\beta]' = \eta_d\hat{n}_{\frac{\beta}{\sqrt{\eta_d}}}+\nu_{el}\mathbb{1}, \tag{5.51}$$

$$\left[\hat{n}_\beta^2\right]' = \eta_d^2\hat{n}^2_{\frac{\beta}{\sqrt{\eta_d}}}+\eta_d(4\nu_{el}+1-\eta_d)\hat{n}_{\frac{\beta}{\sqrt{\eta_d}}}+(2\nu_{el}^2+\nu_{el})\mathbb{1}. \tag{5.52}$$

Written slightly differently, the ideal and noisy observables are related by linear combinations,

$$\left[\hat{n}_{\sqrt{\eta_d}\beta_i}\right]' = \eta_d \hat{n}_{\beta_i} + \nu_{el}\mathbb{1}, \tag{5.53}$$

$$\left[\hat{n}^2_{\sqrt{\eta_d}\beta_i}\right]' = \eta_d^2 \hat{n}^2_{\beta_i} + \eta_d(4\nu_{el} + 1 - \eta_d)\hat{n}_{\beta_i} + (2\nu_{el}^2 + \nu_{el})\mathbb{1}. \tag{5.54}$$

So in the trusted noise scenario, Bob determines the expectations of the observables displaced by $\sqrt{\eta_d}\beta_i$. With these noisy expectations, he can then recreate the expectations of the ideal observables by inverting the relationships in Eqs. (5.53), (5.54). Explicitly,

$$\langle \hat{n}_{\beta_i} \rangle^{\text{eff}} = \frac{\left\langle \left[\hat{n}_{\sqrt{\eta_d}\beta_i}\right]' \right\rangle - \nu_{el}}{\eta_d}, \tag{5.55}$$

$$\langle \hat{n}^2_{\beta_i} \rangle^{\text{eff}} = \frac{1}{\eta_d^2}\left(\left\langle \left[\hat{n}^2_{\sqrt{\eta_d}\beta_i}\right]' \right\rangle - 2\nu_{el}^2 - \nu_{el} - (4\nu_{el} + 1 - \eta_d)\left(\left\langle \left[\hat{n}_{\sqrt{\eta_d}\beta_i}\right]' \right\rangle - \nu_{el} \right) \right). \tag{5.56}$$

This provides one interesting piece of evidence about why trusted noise does not significantly impact key rates. Namely, given a fixed channel model, the set of states in the key rate optimization is the same whether Bob's detection is ideal or noisy. The only change in the key rate optimization is in the objective function, through the POVM defining Bob's key map.

As an interesting corollary, this shows that even in the trusted noise scenario, Alice and Bob can verify the statistics of the generalized beam splitting attack, as long as $\eta_d \neq 0$.

5.4 Infinite-Dimensional Optimization

In order to specify the protocol optimization, all that remains is to explicitly define the objective function. The only public announcements in this protocol (not including parameter estimation, error correction nor privacy amplification) are for sifting. Since Bob defines the key map and makes the sifting decision, this means that Alice's data is irrelevant for the postprocessing map.[4] Thus, for the purpose of the postprocessing map, we can take Alice's POVM to trivially be $\{\mathbb{1}_A\}$. Recall Bob's POVM is comprised of the region operators (Eq. (5.2)).

[4] It is, of course, relevant in calculating the error-correction cost.

We are considering reverse reconciliation. Recall Bob maps his fine-grained result ζ to the key symbol z, where A_z is the region of the complex plane containing ζ (see Sec. 5.1). As discussed in Sec. 5.1.1, for the purposes of the key map, Bob's measurement is described by the coarse-grained POVM formed by the region operators, defined as $R^z = \int_{A_z} \frac{1}{\pi} |\zeta\rangle\langle\zeta| \ d^2\zeta$, with $z \in \{0, ..., M-1, \perp\}$.

So, the key map g simply maps Bob's private measurement result to the key symbol. The key map isometry is then trivial, and we can think of it as a relabelling of the register $\overline{B}$ to Z. The simplified completely positive and trace non-increasing form of the postprocessing map Φ in Eq. (3.5) is then

$$\Phi^{ideal}(\rho_{ABE}) = \sum_{z=0}^{M-1} |z\rangle\langle z|_Z \otimes \mathrm{Tr}_{AB}[\rho_{ABE} \ (\mathbb{1}_A \otimes R_B^z)], \tag{5.57}$$

For the trusted noise scenario, the only difference is that the region operators are noisy

$$\Phi^{noisy}(\rho_{ABE}) = \sum_{z=0}^{M-1} |z\rangle\langle z|_Z \otimes \mathrm{Tr}_{AB}[\rho_{ABE} \ (\mathbb{1}_A \otimes [R_B^z]')] \tag{5.58}$$

where $[R_B^z]' = \int_{A_z} \frac{1}{\eta_d \pi} \hat{D}\left(\frac{\zeta}{\sqrt{\eta_d}}\right) \rho_{th}(\bar{n}) \hat{D}^\dagger\left(\frac{\zeta}{\sqrt{\eta_d}}\right) \ d^2\zeta$.

Accordingly, the two objective functions are

$$f^{ideal}(\rho_{AB}) = H(Z|E)_{\Phi^{ideal}(\rho_{ABE})}, \tag{5.59}$$

and

$$f^{noisy}(\rho_{AB}) = H(Z|E)_{\Phi^{noisy}(\rho_{ABE})}. \tag{5.60}$$

Having formalized the description of the protocol, we are now able to write down the infinite-dimensional optimization for DMCVQKD, in both the ideal detector and trusted noise scenarios. $\mathcal{H}_A$ is the Hilbert space with dimension equal to the number of signal states d; $\mathcal{H}_B$ is the Hilbert space of a single optical mode. Recall that in addition to the constraints from parameter estimation, from the source-replacement scheme (see Sec. 3.1.2) we have a constraint that the reduced density matrix is

$$\tau_A \equiv \sum_{i,j} \sqrt{p(i)p(j)} \, \langle \alpha_j | \alpha_i \rangle \, |i\rangle\langle j| \, . \tag{5.61}$$

The optimizations are defined in Eq. (5.62) and Eq. (5.63).

Infinite-dimensional optimization for DMCVQKD (Ideal Detector)

$$\begin{aligned}
\underset{\rho}{\text{minimize:}} \quad & f^{ideal}(\rho) \\
\text{subject to:} \quad & \mathrm{Tr}(\rho) = 1 \\
& \mathrm{Tr}_B(\rho) = \tau_A \\
& \mathrm{Tr}\left[\rho \left(\tfrac{1}{p(i)} |i\rangle\langle i| \otimes \hat{n}_{\beta_i}\right)\right] = \langle \hat{n}_{\beta_i} \rangle \\
& \mathrm{Tr}\left[\rho \left(\tfrac{1}{p(i)} |i\rangle\langle i| \otimes \hat{n}^2_{\beta_i}\right)\right] = \langle \hat{n}^2_{\beta_i} \rangle \\
& \rho \in \mathrm{Pos}\left(\mathcal{H}_A \otimes \mathcal{H}_B\right).
\end{aligned} \tag{5.62}$$

Infinite-dimensional optimization for DMCVQKD (Trusted Noise)

$$\begin{aligned}
\underset{\rho}{\text{minimize:}} \quad & f^{noisy}(\rho) \\
\text{subject to:} \quad & \mathrm{Tr}(\rho) = 1 \\
& \mathrm{Tr}_B(\rho) = \tau_A \\
& \mathrm{Tr}\left[\rho \left(\tfrac{1}{p(i)} |i\rangle\langle i| \otimes \hat{n}_{\beta_i}\right)\right] = \langle \hat{n}_{\beta_i} \rangle^{\text{eff}} \\
& \mathrm{Tr}\left[\rho \left(\tfrac{1}{p(i)} |i\rangle\langle i| \otimes \hat{n}^2_{\beta_i}\right)\right] = \langle \hat{n}^2_{\beta_i} \rangle^{\text{eff}} \\
& \rho \in \mathrm{Pos}\left(\mathcal{H}_A \otimes \mathcal{H}_B\right).
\end{aligned} \tag{5.63}$$

where $\langle \hat{n}_{\beta_i} \rangle^{\text{eff}}$ and $\langle \hat{n}^2_{\beta_i} \rangle^{\text{eff}}$ are as defined in Eqs. (5.55), (5.56).

5.5 Applying Dimension Reduction Method

Having completed the formulation of the infinite-dimensional key rate optimization for the DMCVQKD protocol (Eq. (5.62)), we can now apply the dimension reduction method to get a tractable finite-dimensional optimization. Recall that the nontrivial steps of this method are choosing the finite subspace Π and determining W, as the other steps are protocol-agnostic. We present some of these sections in more of a chronological fashion, covering some of the ideas that *didn't* work. The purpose of this is to illustrate how these

challenges were overcome, thereby providing greater intuition for future applications of this method.

5.5.1 Finite Subspace Π

The first idea for the finite subspace was to simply use a truncated Fock basis. This is a common choice for approximating quantum optics calculations [22, 72]. However, the weight outside this subspace was too large, leading to a correction term larger than the key rate.

Instead, consider a Gaussian channel with loss and excess noise. As discussed in Sec. 5.6, this is a typical channel model for a fiber-based implementation of the DMCVQKD protocol. If the excess noise is zero, the conditional states are coherent states, so a truncation in the Fock basis necessarily throws away some information. This suggests that using a basis with the coherent states themselves would be better. In particular, for the i^{th} signal state, we would want the basis to include the received coherent state $|0\rangle_{\beta_i}$ (Recall our notation that complex subscripts indicate displacement, so that $|n_\gamma\rangle \equiv \hat{D}(\gamma)|n\rangle$.).

When there is nonzero excess noise, the conditional states are instead displaced thermal states. This means some weight leaks into displaced Fock states with $n > 0$. For the i^{th} signal state, the best projection on $\mathcal{H}_B$ is thus

$$\Pi^N_{B_{\beta_i}} = \sum_{n=0}^{N} |n_{\beta_i}\rangle\langle n_{\beta_i}| . \tag{5.64}$$

We will refer to N as the *subspace dimension parameter*.[5] Combining these projectors for each conditional state, the projection operator on the total Hilbert space is

$$\Pi^N \equiv \sum_{i=0}^{d-1} |i\rangle\langle i|_A \otimes \Pi^N_{B_{\beta_i}} . \tag{5.65}$$

Note this is an application of our general principle: choose the subspace containing the most weight of the state under a typical channel model. Also note that this projection commutes with the observables (Eq. (5.20)) because $\Pi^N_{B_{\beta_i}}$ commutes with $\hat{n}_{\beta_i}$. We use this fact to define the finite set (see Sec. 5.5.4).

[5] This nomenclature is admittedly clunky, but we want to be careful not to confuse N with the dimension of each conditional subspace $(N+1)$ nor with the total finite dimension $(d(N+1))$.

We re-emphasize that unlike the truncated Fock basis considered in previous work [22, 23], our finite subspace contains the full weight of the state when the channel is purely lossy. This is important as it ensures our numerics exactly reproduces the analytically solvable loss-only case.

5.5.2 Weight W

To bound the weight outside the subspace, we analytically solve the dual of the SDP in Eq. (4.15). Our result is stated in the following theorem. It applies to both the ideal detector and trusted noise cases, since the same expectations are determined in both cases.

Theorem 7 (Bound on W for DMCVQKD)**.** *For the DMCVQKD protocol, with* Π^N *as defined in Eq.* (5.65)*, the weight outside the subspace* W *is bounded by*

$$W = \sum_{i=0}^{d-1} p(i) \frac{\langle \hat{n}_{\beta_i}^2 \rangle - \langle \hat{n}_{\beta_i} \rangle}{N(N+1)}, \tag{5.66}$$

for $N > 0$.

Proof. To prove this theorem, we will consider Bob's conditional states $\rho_B^i = \frac{1}{p(i)} \mathrm{Tr}_A \left[\rho_{AB} \left(|i\rangle\langle i|_A \otimes \mathbb{1}_B \right) \right]$. Let $\overline{\Pi}_{B_{\beta_i}}^N \equiv \mathbb{1}_B - \Pi_{B_{\beta_i}}^N$ and let $W_i \equiv \mathrm{Tr}\left(\rho_B^i \overline{\Pi}_{B_{\beta_i}}^N \right)$ be the weight of the i^{th} conditional state. We first show that $W = \sum_i p(i) W_i$.

$$\mathrm{Tr}\left(\rho \overline{\Pi}^N\right) = \mathrm{Tr}\left[\rho \left(\sum_{i=0}^{d-1} |i\rangle\langle i|_A \otimes \overline{\Pi}_{B_{\beta_i}}^N \right) \right] \tag{5.67}$$

$$= \sum_{i=0}^{d-1} \mathrm{Tr}\left(\tilde{\rho}_B^i \overline{\Pi}_{B_{\beta_i}}^N \right) \tag{5.68}$$

$$= \sum_{i=0}^{d-1} p(i) W_i. \tag{5.69}$$

Now, we only need to bound the weight of each conditional state. Using the constraints

from Eq. (5.62), each of these bounds can be expressed as a primal SDP.

$$\begin{aligned} \underset{\rho}{\text{maximize:}} \quad & \operatorname{Tr}\left(\overline{\Pi}^N_{B_{\beta_i}} \rho\right) \\ \text{subject to:} \quad & \operatorname{Tr}(\rho) = 1 \\ & \operatorname{Tr}(\hat{n}_{\beta_i} \rho) = \langle \hat{n}_{\beta_i} \rangle \\ & \operatorname{Tr}\left(\hat{n}^2_{\beta_i} \rho\right) = \langle \hat{n}^2_{\beta_i} \rangle \\ & \rho \in \operatorname{Pos}\left(\mathcal{H}_B\right). \end{aligned} \tag{5.70}$$

In order to find an upper bound on this primal SDP we consider its dual. By weak duality, it holds that a feasible solution to the dual SDP upper bounds the primal. In fact, strong duality holds for this SDP, so this upper bound can be made tight.

$$\begin{aligned} \underset{\vec{y}}{\text{minimize:}} \quad & y_1 + \langle \hat{n}_{\beta_i} \rangle y_2 + \langle \hat{n}^2_{\beta_i} \rangle y_3 \\ \text{subject to:} \quad & y_1 \mathbb{1}_B + y_2 \hat{n}_{\beta_i} + y_3 \hat{n}^2_{\beta_i} - \overline{\Pi}^N_{B_{\beta_i}} \geq 0 \\ & \vec{y} \in \mathbb{R}^3. \end{aligned} \tag{5.71}$$

For $N > 0$, a feasible solution for the dual is

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{-1}{N(N+1)} \\ \frac{1}{N(N+1)} \end{pmatrix}. \tag{5.72}$$

The feasibility of this solution can be easily verified as all operators in the constraint are diagonal in the $|n_{\beta_i}\rangle$ basis, so positivity is implied if and only if the diagonal entries are nonnegative. In fact this dual solution is optimal, as can be seen by exhibiting the primal feasible solution

$$\begin{aligned} \rho = \hat{D}\left(\beta_i\right) \Bigg[& \left(1 - \langle \hat{n}_{\beta_i} \rangle \left(1 + \frac{1}{N+1}\right) + \langle \hat{n}^2_{\beta_i} \rangle \frac{1}{N+1}\right) |0\rangle\langle 0| \\ & + \left(\langle \hat{n}_{\beta_i} \rangle \left(1 + \frac{1}{N}\right) - \langle \hat{n}^2_{\beta_i} \rangle \frac{1}{N}\right) |1\rangle\langle 1| + \frac{\langle \hat{n}^2_{\beta_i} \rangle - \langle \hat{n}_{\beta_i} \rangle}{N(N+1)} |N+1\rangle\langle N+1| \Bigg] \hat{D}^\dagger\left(\beta_i\right). \end{aligned} \tag{5.73}$$

The dual solution leads to the objective value $W_i = \frac{\langle \hat{n}^2_{\beta_i} \rangle - \langle \hat{n}_{\beta_i} \rangle}{N(N+1)}$. Substituting into Eq. (5.69), the proof is complete. □

5.5.3 Correction Term Δ

The projection does not commute with the POVMs forming the key map, and hence we use the general correction term from Theorem 2. The correction term depends on the weight W and key map register dimension $|Z|$. We use the value of W determined in Theorem 7. Regardless of whether postselection is performed, $|Z|$ is equal to the number of key outcomes only, not including $\perp$ (see Sec. 4.2.3).

We could do no worse by using the correction term in the partially-commuting case from Theorem 4. However, numerical calculations suggest that the POVMs for DMCVQKD typically have large off-diagonal blocks. That is, the quantity $\frac{c}{\sqrt{W}}$ in Theorem 4, which is a normalized measure of the "block-diagonalness" of the POVM, is around 1. Hence for this protocol we may not see much qualitative improvement over the general correction term by using the partially-commuting correction term. This seems plausible, as information in this protocol is encoded into the amplitude and/or phase of states in phase space, But the Fock states are phase-invariant, so a projection in the Fock basis, or even a displaced Fock basis, is very non-commutative with the POVMs. It is an interesting avenue for future work to explore this further.

5.5.4 Finite Set $\mathbf{S}_N$

By design, the projection Π^N (Eq. (5.65)) commutes with the positive observables Γ_i (Eq. (5.20)). We can thus use the form of $\mathbf{S}_N$ in Eq. (4.101). Note that the lower bounds in Eq. (4.101) are not useful as the observables Γ_i are unbounded. Recall $\mathcal{H}_N$ is the subspace of $\mathcal{H}_{AB}$ onto which Π^N projects. The finite-dimensional optimizations are presented in Eq. (5.74) and Eq. (5.75).

Finite-dimensional optimization for DMCVQKD (Ideal Detector)

$$
\begin{aligned}
\underset{\tilde{\rho}}{\text{minimize:}} \quad & f^{ideal}(\tilde{\rho}) \\
\text{subject to:} \quad & 1 - W \leq \mathrm{Tr}(\tilde{\rho}) \leq 1 \\
& \frac{1}{2}\|\mathrm{Tr}_B(\tilde{\rho}) - \tau_A\|_1 \leq \sqrt{W} \\
& \mathrm{Tr}\left[\tilde{\rho}\left(\tfrac{1}{p(i)}\,|i\rangle\langle i| \otimes \hat{n}_{\beta_i}\right)\right] \leq \langle \hat{n}_{\beta_i} \rangle \\
& \mathrm{Tr}\left[\tilde{\rho}\left(\tfrac{1}{p(i)}\,|i\rangle\langle i| \otimes \hat{n}^2_{\beta_i}\right)\right] \leq \langle \hat{n}^2_{\beta_i} \rangle \\
& \tilde{\rho} \in \mathrm{Pos}\left(\mathcal{H}_N\right)
\end{aligned}
\tag{5.74}
$$

where $W = \sum_i p(i) \frac{\langle \hat{n}^2_{\beta_i} \rangle - \langle \hat{n}_{\beta_i} \rangle}{N(N+1)}$.

Finite-dimensional optimization for DMCVQKD (Trusted Noise)

$$
\begin{aligned}
\underset{\tilde{\rho}}{\text{minimize:}} \quad & f^{noisy}(\tilde{\rho}) \\
\text{subject to:} \quad & 1 - W \leq \mathrm{Tr}(\tilde{\rho}) \leq 1 \\
& \frac{1}{2}\|\mathrm{Tr}_B(\tilde{\rho}) - \tau_A\|_1 \leq \sqrt{W} \\
& \mathrm{Tr}\left[\tilde{\rho}\left(\tfrac{1}{p(i)}\,|i\rangle\langle i| \otimes \hat{n}_{\beta_i}\right)\right] \leq \langle \hat{n}_{\beta_i} \rangle^{\text{eff}} \\
& \mathrm{Tr}\left[\tilde{\rho}\left(\tfrac{1}{p(i)}\,|i\rangle\langle i| \otimes \hat{n}^2_{\beta_i}\right)\right] \leq \langle \hat{n}^2_{\beta_i} \rangle^{\text{eff}} \\
& \tilde{\rho} \in \mathrm{Pos}\left(\mathcal{H}_N\right)
\end{aligned}
\tag{5.75}
$$

where $W = \sum_i p(i) \frac{\langle \hat{n}^2_{\beta_i} \rangle^{\text{eff}} - \langle \hat{n}_{\beta_i} \rangle^{\text{eff}}}{N(N+1)}$ and $\langle \hat{n}_{\beta_i} \rangle^{\text{eff}}$ and $\langle \hat{n}^2_{\beta_i} \rangle^{\text{eff}}$ are as defined in Eqs. (5.55), (5.56).

5.6 Simulation Results

To understand the performance of this protocol and demonstrate our security proof approach, we simulate data obtained from a typical experiment. Note that this simulation discussion pertains not only to this section, but the following sections as well.

We model the signal states as passing through a noisy and lossy Gaussian channel. The transmittance η is modelled as a function of distance d according to $\eta = 10^{-k \cdot d/10}$, where k is the attenuation factor of the channel. We use a typical value for commercial-grade fiber, $k = 0.2$ dB/km. The excess noise ξ is taken to be fixed at the channel input, for example as preparation noise, so that Bob sees the effective noise $\delta = \eta\xi$. The expectation values for this simulation are $\langle \hat{n}_{\beta_i} \rangle = \delta/2$ and $\langle \hat{n}^2_{\beta_i} \rangle = \delta(1+\delta)/2$, as derived in Appendix D. This implies $W = \delta^2/[2N(N+1)]$. We emphasize that our security proof does not depend on these parameter choices and simulation model, which are only used to illustrate the performance of the protocol in a typical implementation.

The error-correction cost is calculated by simulating the joint probability distribution obtained by Alice and Bob (see Appendix D).To account for realistic error-correction costs, we model the δ^{leak}_{EC} term in the key rate (see Eq. (3.11)), using an efficiency parameter β_{EC} (see Eq. (3.24)). We use a representative value of $\beta_{EC} = 0.95$ for all modulation schemes and at all distances. This efficiency can likely be achieved by using low-density parity-check (LDPC) codes [77] adapted to discrete modulations. Very recent work [78] has achieved similar efficiencies using a variation on LDPC codes known as the multiple decoding attempts scheme. In principle, these codes can be applied to arbitrary modulation schemes [78], which will be relevant when we consider simulations for different signal state constellations in Sec. 5.7.2. Nevertheless, we caution that setting $\beta_{EC} = 0.95$ is ultimately only a heuristic for modelling the EC cost, and more study of the details of these EC codes is warranted to ascertain the accuracy of this heuristic, especially at long distances and correspondingly low signal-to-noise ratios.

Our algorithms are implemented in MATLAB (versions R2019B or R2021A), using the convex optimization package CVX (versions 2.1 or 2.2) [79, 80] with the solver MOSEK (versions 8.0.0.60 or 9.1.9) [81]. The Frank-Wolfe algorithm, with a maximum of 30 iterations, is used to solve the first step of the numerical method.

Some general remarks regarding parameter optimization: unless otherwise stated, parameter optimizations use the fminbnd algorithm included in the MATLAB distribution, which uses a combination of parabolic interpolation and golden-section search. In some cases, we instead use a coarse-grained search over parameters, either individually or jointly. Due to the inherent unpredictability of numerical solvers, and the fact that search algo-

rithms can get stuck in local minima, the parameter values returned by these numerical methods are not always exactly optimal. Generally speaking, this has a rather small impact on the results, because the key rate does not seem to vary rapidly with respect to any of the protocol parameters we consider in this or later sections. In some cases, a large outlier in the returned optimal parameters is evident. In this case, in order to produce relatively smooth plots, we slightly tweak the numerically-returned optimal parameters at these points by hand, roughly fitting them to the overall trend from the other points.

We emphasize that all key rate plots include the correction term unless stated otherwise. That is, $R^{\infty} = C_{num} - \delta_{EC}^{leak} - \Delta(W)$, where C_{num} denotes the reliable numerical lower bound on $f(\tilde{\rho}^N)$. In order to evaluate the effect of the correction term, and to compare with previous work using the photon number cutoff assumption, we will find it useful to consider the uncorrected values, defined as $C_{num} - \delta_{EC}^{leak}$.

All results in this section are for phase-shift keying with $M = 4$, using amplitude and phase postselection (see Fig. 5.1). We present key rate plots for different choices of channel parameters η and ξ, protocol parameters α, Δ_a, and Δ_p, and the subspace dimension parameter N. For the trusted noise scenario, we also consider η_d and ν_{el}.

5.6.1 Understanding the Dimension Reduction Method

We first examine some results relating to the dimension reduction method itself.

In Fig. 5.3, we compare the key rates and uncorrected values from our dimension reduction method to the key rates under the photon number cutoff assumption obtained in Ref. [23]. In order to enable a meaningful comparison, we use the protocol parameters from Ref. [23]. The uncorrected values, which are equal to the key rate before subtracting the correction term Δ, are essentially identical to those in Ref. [23]. As the results from Ref. [23] are an upper bound on the key rate, this indicates our choice of $\mathbf{S}_N$ is tight. Further, our corrected key rates are very close to the uncorrected values. This illustrates our correction term is small for reasonable values of the subspace dimension parameter N, at low channel excess noise (see Fig. 5.4).

As our key rates are similar to the ones under the cutoff assumption, the qualitative conclusions of previous work [23, 72] are confirmed to hold under our precise treatment, without the previous working assumptions. That is, we are able to lift the assumption that the state is finite-dimensional, with minimal impact on the key rate.

To illustrate the relative size of the correction term, we plot it as a fraction of the uncorrected value in Fig. 5.4, at a fixed distance of 15 km. More precisely, for each value

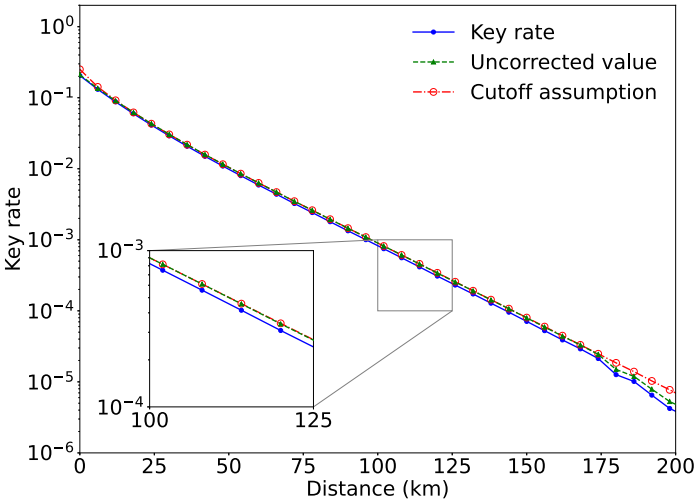


Figure 5.3: Comparison of key rates and uncorrected values from our dimension reduction method with key rates under the photon number cutoff assumption from Ref. [23]. Results are plotted versus distance with excess noise $\xi = 0.01$, and are in the ideal detector scenario. Postselection parameters and signal state intensities from Ref. [23] are used: $\alpha = 0.6$, $\Delta_p = 0$, and Δ_a is optimized with a coarse-grained search over $[0.5, 0.65]$. The subspace dimension parameter is $N = 20$ and $M = 4$.

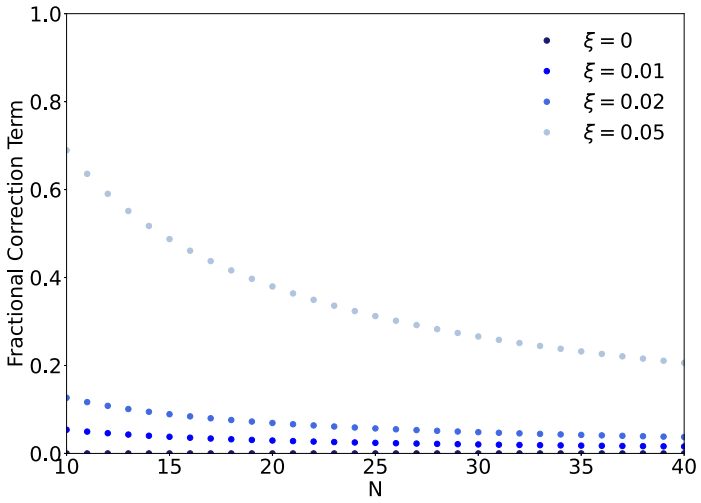


Figure 5.4: The fractional correction term versus the subspace dimension parameter N for different values of excess noise, in the ideal detector scenario. The distance is 15 km, $M = 4$, $\alpha = 0.8$, $\Delta_a = 0.5$, and $\Delta_p = 0$.

of excess noise ξ and subspace dimension parameter N, we plot the fractional correction term $\Delta(W)/C_{num}$, where $W = (\eta\xi)^2/[2N(N+1)]$, while C_{num} is computed at a fixed value of $N = 40$ for each ξ. (In the range considered here, C_{num} has a negligible dependence on N.) The protocol parameters are $M = 4$, $\alpha = 0.8$, $\Delta_a = 0.5$, and $\Delta_p = 0$ for each value of excess noise.

As expected, for a pure-loss channel the correction term is zero. This is because the finite subspace fully contains the loss-only simulation state, as discussed in Sec. 5.5.1. For small values of excess noise, the correction term is negligible even for small N. For larger values of excess noise, and especially in the high loss regime, N must be increased to obtain reasonable results. This is because the correction term scales like the loss η for all values of excess noise ξ, while, for nonzero ξ, the key rate scales worse than η.

5.6.2 Key Rates and Optimal Protocol Parameters

In Fig. 5.5, we plot the ideal detector key rates for different channel parameters, using amplitude and phase postselection. The signal state intensity and postselection parameters are numerically optimized for each distance and value of excess noise, using $N = 10$. The key rates are calculated using $N = 40$, except for a small number of points where we use

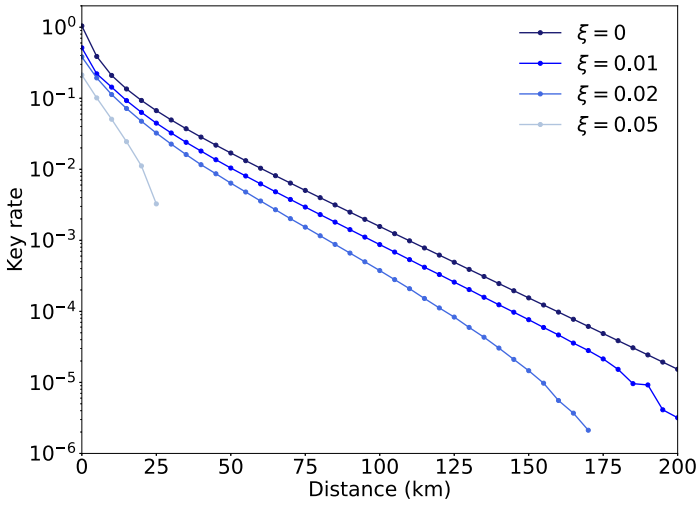


Figure 5.5: Ideal detector secure key rates versus transmission distance, for different values of excess noise ξ, with optimized postselection parameters and signal state intensity. For each point, the better of the two results from $N = 30$ and $N = 40$ is used, with the majority of points from $N = 40$. Postselection can improve the noise tolerance and range of the protocol. The protocol is $M = 4$ phase-shift keying.

$N = 30$ to ameliorate numeric issues, as discussed in Appendix B.2. Using postselection extends the range of the protocol for high excess noise while also reducing the amount of data processing for error correction, which can be a bottleneck in actual implementations. For example, for 2% excess noise, postselection increases the maximum distance by around 50 km, while discarding 40% of the signals. The small number of outlying points that deviate from the trend are due to numerical issues inherent to convex solvers. We emphasize that these key rates are still rigorous, and can be improved by using higher-precision numerical methods.

We plot the optimal protocol parameters to gain a better understanding of the protocol behaviour. The optimal signal state amplitudes α_{opt} are shown in Fig. 5.6. The optimization range $[0.5, 2]$ is sufficient for almost all parameter choices, though α_{opt} tends to infinity as distance and excess noise tend to zero. The general trend is that α_{opt} decreases as distance and excess noise increase. With α_{opt} fixed, Δ_a is optimized over $[0, 1]$. (One could jointly optimize all protocol parameters, but we do not expect this to noticeably improve the key rates.) The optimal postselection amplitudes $\Delta_{a_{opt}}$ are shown in Fig. 5.7. The general trend is that $\Delta_{a_{opt}}$ increases as distance and excess noise increase. We find that the

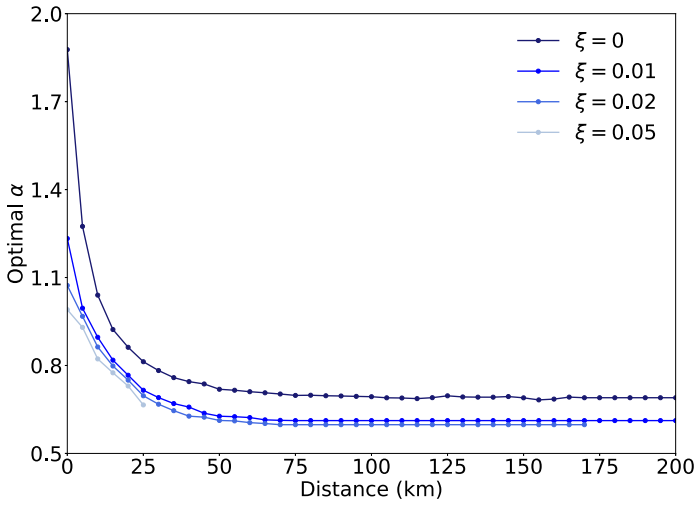


Figure 5.6: Optimal signal state amplitude α_{opt} versus transmission distance, for different values of excess noise ξ and in the ideal detector scenario. The amplitude is optimized in the range $[0.5, 2]$, with $\Delta_a = \Delta_p = 0$ and $N = 10$.

optimal value for Δ_p seems to always be zero, so phase postselection is omitted altogether. We will consider other postselection patterns in Sec. 5.7.1.

5.6.3 Trusted Noise

In Fig. 5.8, we consider the key rates in the trusted detector noise scenario. We take the channel to have 1% excess noise, and consider different values of detector efficiency and electronic noise. The protocol parameters are optimized for the ideal detector scenario, and the same parameters are used for each of the different trusted noise cases. We observe that even with large detector imperfections and 1% channel excess noise, it is possible to generate secure key at long distances. As expected, trusted detector noise does not significantly alter the scaling of the key rates. This is markedly different from the effect of channel excess noise (see Fig. 5.5).

5.6.4 Comparison to Gaussian Modulation

Discrete-modulated (DM) CVQKD is intended to be a more experimentally feasible alternative to Gaussian-modulated (GM) CVQKD. It is thus of interest to compare the

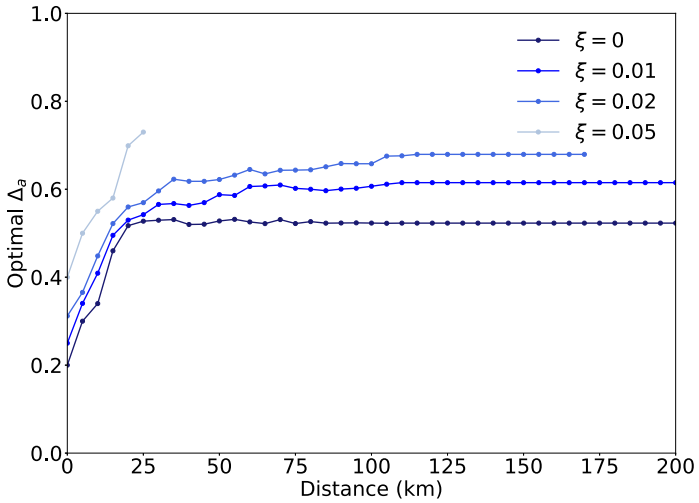


Figure 5.7: Optimal amplitude postselection parameter $\Delta_{a_{opt}}$ versus transmission distance, for different values of excess noise ξ and in the ideal detector scenario. The amplitude is optimized in the range $[0, 1]$, with $\alpha = \alpha_{opt}$, $\Delta_p = 0$ and $N = 10$.

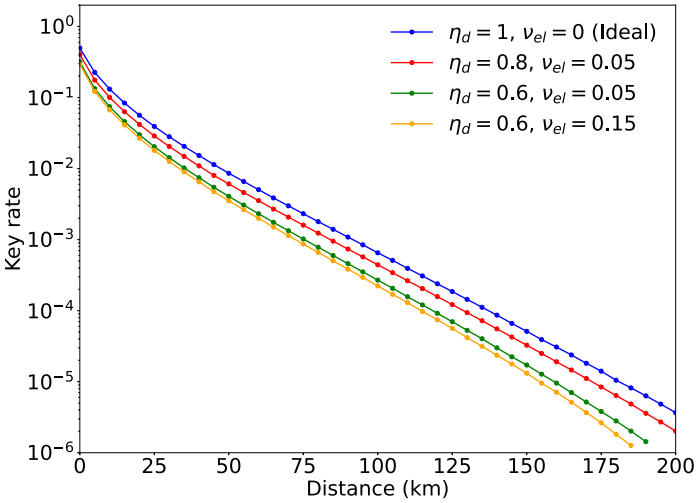


Figure 5.8: Key rates versus distance for different trusted detector imperfections, with excess noise $\xi = 0.01$. Protocols are evaluated with the same optimized protocol parameters, including postselection, as in Fig. 5.5. The subspace dimension parameter is $N = 5$ and $M = 4$.

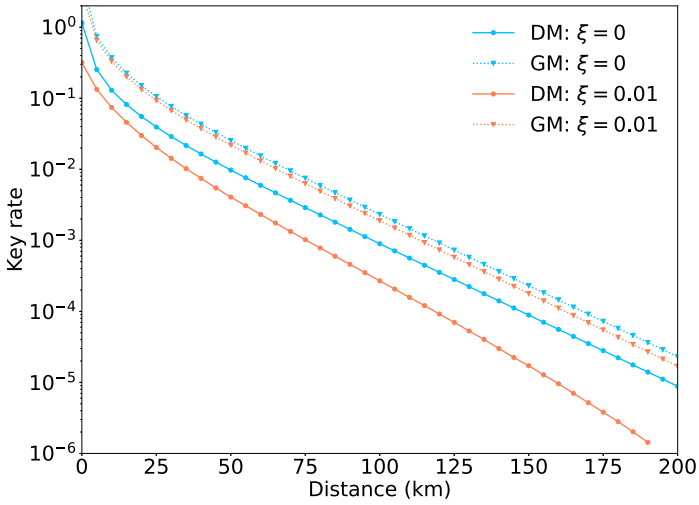


Figure 5.9: Comparison of Gaussian [82] and discrete-modulation key rates for different values of excess noise ξ, in the trusted noise scenario with $\eta_d = 0.6$, $\nu_{el} = 0.05$. Parameters for both protocols are optimized. The subspace dimension parameter is $N = 5$. The discrete-modulated protocol uses $M = 4$ phase-shift keying.

performance of the two protocols. We perform a very basic comparison in Fig. 5.9, using the same error-correction efficiency $\beta_{EC} = 0.95$, detector loss $\eta_d = 0.6$, and electronic noise $\nu_{el} = 0.05$. The modulation variance is optimized for GM, using the key rate formula in Ref. [82]. The optimized protocol parameters for DM are the same as in Fig. 5.8. We note that a complete and in-depth comparison of the two protocols would have to account for many more implementation details. For a pure-loss channel, the GM key rates are around an order of magnitude higher. At higher excess noise, the gap is larger as GM is more robust to channel noise. However, both protocols are largely unaffected by trusted detector imperfections. We will see in Sec. 5.7.2 that the tolerance of DM to channel noise can be improved by using a larger constellation of signal states.

5.7 Improving Protocol Performance

With the numerical key rate toolbox in hand, we can explore variations on the DMCVQKD protocol to enhance its performance. In particular, we consider different postselection regions and different signal state constellations. All simulation results in this section follow

the same setup as outlined at the beginning of Sec. 5.6.

5.7.1 Optimal Postselection

The idea of postselection in CVQKD was first introduced in Ref. [13] in the context of a Gaussian-modulated protocol, as a generalization of the classical idea of advantage distillation [83]. It was shown that postselection enables the protocol with direct reconciliation to reach losses greater than 3 dB. It was not considered relevant for reverse reconciliation, as then Eve's information is always less than Bob's.

Postselection was later studied in Ref. [73] for binary-modulated DMCVQKD, and the optimal postselection region determined in the loss-only scenario. Importantly, realistic error correction (i.e not at the Shannon limit) was considered, which made postselection relevant for reverse reconciliation. Postselection was also studied in Ref. [84] for phase-shift keying with larger M, but only for direct reconciliation as realistic error correction was not considered.

Our contribution is firstly to clarify the role of the announcement structure in designing optimal postselection. In doing so, we arrive at a generalization of the concept of effective binary channels which was introduced in Ref. [73]. Secondly, we use this result to extend the study of optimal postselection to reverse reconciliation with realistic error-correction costs for phase-shift keying with larger M. We also demonstrate the sensitivity of the optimal postselection region to different error-correction costs. Finally, we propose heuristics for the postselection region for the DMCVQKD protocol without announcements.

Intuitively, the idea behind postselection is as follows. Unlike the situation with Wyner's wire-tap channel [85], Alice and Bob share a "resource" in the form of authenticated communication, and they can harness this to gain an advantage over the eavesdropper. When Bob gets an ambiguous measurement result, he simply lets Alice know and they discard that round. But when Eve gets an ambiguous result, she cannot do so. By discarding some data in this manner, Alice and Bob can improve the correlations of their remaining information and *then* by Wyner's wire-tap bound can generate a secret key.

General Problem Formulation

We first precisely define the scenario we are considering for designing optimal postselection, and then see how the announcement structure aids in solving the resultant optimization.

Without loss of generality, we suppose Bob performs the key map and makes the postselection decision. Assume the input state ρ_{ABE} is known, and that the protocol steps,

excluding postselection, are fixed. Consider the state after measurement, announcements, and key generation. Recalling our definition of the postprocessing map in Sec. 3.2.2, this is the state $V\Phi_M^{AB}(\rho_{ABE})V^\dagger$. We model each party as having a copy of all classical public announcements, in registers C_A, C_B, and C_E. To simplify the notation, we use composite registers $[A] = \overline{A}C_A$, $[B] = \overline{B}C_B$, and $[E] = EC_E$. The postprocessed state is a CCQ state between Alice, Bob, and Eve. As usual, $S_{[A]}$ and $S_{[B]}$ denote the alphabets for the corresponding classical registers.

Now, we introduce a new classical register L. This holds the conditioning variable $l \in S_L$, on which we condition whether or not to postselect. l is a function of Bob's information, represented by $l = g_p(j)$, with $j \in S_{[B]}$. In other words, the information in L is a coarse-graining of that in $[B]$.[6] Writing the new register explicitly, the postprocessed CCQC state is

$$\rho_{post} = \sum_{\substack{i\in S_{[A]}\\ j\in S_{[B]}}} p(i,j)\,|g(j)\rangle\langle g(j)|_Z \otimes |i\rangle\langle i|_{[A]} \otimes |j\rangle\langle j|_{[B]} \otimes \rho_{[E]}^{i,j} \otimes |g_p(j)\rangle\langle g_p(j)|_L\,. \tag{5.76}$$

Recall that the key map g is a function of j with output alphabet $S_Z = \{0, ..., M-1, \perp\}$.

We model postselection in the following manner. A subset S_0 of the alphabet S_L is chosen. If $l \in S_0$, then "keep" is announced. If $l \notin S_0$, then "discard" is announced, and the key symbol is changed to $\perp$. The optimal postselection choice is then the subset S_0 maximizing the key rate. Finding this subset can be formulated as an optimization problem.

First, the state after postselection is

$$\begin{aligned}\rho(S_0) = &\sum_{\substack{i\\ j:l(j)\in S_0}} p(i,j)\,|g(j)\rangle\langle g(j)|_Z \otimes |i\rangle\langle i|_{[A]} \otimes |j\rangle\langle j|_{[B]} \otimes \rho_{[E]}^{i,j} \otimes |g_p(j)\rangle\langle g_p(j)|_L \otimes |keep\rangle\langle keep|_D \\ + &\sum_{\substack{i\\ j:l(j)\notin S_0}} p(i,j)\,|\perp\rangle\langle\perp|_Z \otimes |i\rangle\langle i|_{[A]} \otimes |j\rangle\langle j|_{[B]} \otimes \rho_{[E]}^{i,j} \otimes |g_p(j)\rangle\langle g_p(j)|_L \otimes |discard\rangle\langle discard|_D\,.\end{aligned} \tag{5.77}$$

The key rate is then[7]

$$R(\rho(S_0)) = H(Z|[E]D)_{\rho(S_0)} - H(Z|[A]D)_{\rho(S_0)} \tag{5.78}$$

[6] The astute reader may wonder why we would ever limit L to not contain all possible decision-making information. As we will see shortly, this will be because we announce L publicly.

[7] Assuming error correction is performed at the Shannon limit. We will later consider realistic error-correction costs.

As we have noted earlier, the conditional entropy is additive over blocks so the discard block does not contribute to the key rate (since the conditional entropy of a CQ state is bounded by the logarithm of the number of outcomes, and there is only one key outcome, $\perp$, when "discard" is announced). Thus, we can equivalently think of the key rate formula as being

$$R(\rho(S_0)) = H(Z|[E])_{\tilde{\rho}(S_0)} - H(Z|[A])_{\tilde{\rho}(S_0)} \tag{5.79}$$

where

$$\tilde{\rho}(S_0) = \sum_{\substack{i \\ j:l(j)\in S_0}} p(i,j) \left|g(j)\right\rangle\!\left\langle g(j)\right|_Z \otimes \left|i\right\rangle\!\left\langle i\right|_{[A]} \otimes \left|j\right\rangle\!\left\langle j\right|_{[B]} \otimes \rho_{[E]}^{i,j} \otimes \left|g_p(j)\right\rangle\!\left\langle g_p(j)\right|_L . \tag{5.80}$$

Then, the optimization we want to solve is

$$\sup_{S_0} R(\tilde{\rho}(S_0)). \tag{5.81}$$

This optimization is over all possible subsets S_0, or in other words, over the power set of S_L. This can be a challenging computation. In particular, when S_L is continuous, it is not even clear how to formulate such an optimization. We thus need a better way to frame the problem.

The key idea is to have the register L become public. Then, the key rate formula becomes

$$R(\tilde{\rho}(S_0)) = H(Z|[E]L)_{\tilde{\rho}(S_0)} - H(Z|[A]L)_{\tilde{\rho}(S_0)} \tag{5.82}$$

The conditional entropy is additive over blocks, *when the classical register is being conditioned on.* Thus, the key rate can be decomposed into contributions from each block. With $p(l) \equiv \sum_{j:g_p(j)=l} p(j)$, the conditional states ρ_l are

$$\rho_l = \frac{1}{p(l)} \left\langle l\right|_L \rho_{post} \left|l\right\rangle_L \tag{5.83}$$

and the key rate is then

$$R(\tilde{\rho}(S_0)) = \sum_{l\in S_0} p(l)[H(Z|[E])_{\rho_l} - H(Z|[A])_{\rho_l}]. \tag{5.84}$$

From this expression, we can read off that the optimal postselection set S_0 simply consists of those l for which $H(Z|EL)_{\rho_l} - H(Z|XL)_{\rho_l} > 0$; where the strict inequality is so that we reduce our data processing burden by discarding signals that will not help us. There

is no longer any need to search over the power set. Thus, by announcing the conditioning variable, we greatly simplify the determination of the optimal postselection set.

As mentioned above, this can be understood as a generalization of the concept of effective binary channels [73]. For a key map with M outcomes (not including $\perp$), each l corresponds to an effective M-ary channel, and we only keep those channels which contribute positively to the key rate.

Application to DMCVQKD

We now consider finding the optimal postselection region for DMCVQKD with phase-shift keying. We first need to make a choice for the conditioning variable. We consider postselecting on Bob's outcomes only. The most fine-grained choice we could make is to let l be Bob's heterodyne measurement result ζ. But since we are announcing l, this would reveal all information to Eve.

We make the simplifying (and likely optimal) assumption that the discard regions respect the $\frac{2\pi}{M}$ rotational symmetry of the protocol. In that case, ζ is discarded if and only if all elements in the coset $C(\zeta) = \{\zeta, e^{2\pi i/M}\zeta, ..., e^{2\pi i(M-1)/M}\zeta\}$ are discarded. Intuitively, the key information is encoded in the phase, so revealing the coset should not help Eve much. We thus set the conditioning variable to be the coset of Bob's outcome $g_p(\zeta) = C(\zeta)$. We will conventionally label each coset by its representative in the first region, i.e. the slice of the complex plane from $\theta : [\frac{-\pi}{M}, \frac{\pi}{M}]$. This region will be denoted by A_{coset}. Since we are considering symmetric postselection regions, we only need to determine which ζ in the first region are kept, and can then just rotate this region to obtain the others.

Unfortunately, as has been discussed, the numerics framework cannot directly model continuous registers, though it might be possible to consider some coarse-graining or binning of the measurement outcome. We thus study the optimal postselection region in the loss-only case, where the key rate can be determined analytically. We then use this result, which is for the protocol with coset announcements, to motivate heuristics for the protocol without announcements, which is the one we can implement in the numerical key rate framework.

Loss-Only Case

In the loss-only scenario, ρ_{ABE} is known (Eq. (5.3)) by the generalized beam splitting attack (see Sec. 5.2). Any measurement outcome ζ for Bob is uniquely identified by its coset $l \in A_{coset}$ and its index within the coset z, and we notate this via $\zeta = \zeta_{zl} = le^{2\pi iz/M}$.

Visually, the index z identifies which $\frac{2\pi}{M}$ rotation of A_{coset} contains ζ, and l identifies the location of the complex number within this rotated A_{coset}. Instead of integrating over all ζ, we can instead sum over z and integrate over l. Recall that the coset label l is taken to lie in the region A_{coset}. The postprocessed state ρ_{ZABEL} is then

$$\rho = \sum_{i=0}^{M-1} \sum_{z=0}^{M-1} \int_{A_{coset}} d^2 l \; p(i, \zeta_{zl}) \, |z\rangle\langle z|_Z \otimes |i\rangle\langle i|_A \otimes |\zeta_{zl}\rangle\langle\zeta_{zl}|_B \otimes \rho_E^i \otimes |l\rangle\langle l|_L \, . \tag{5.85}$$

Note that Eve's conditional states $\rho_E^i = \left|\sqrt{1-\eta}\alpha_i\right\rangle\left\langle\sqrt{1-\eta}\alpha_i\right|$ have no dependence on ζ, due to the product form of Bob and Eve's combined conditional states (Eq. (5.3)). $p(i)$ is the a priori signal state preparation probability, which we have taken to be $\frac{1}{M}$. Conditioned on Alice sending $|\alpha_i\rangle$, Bob's probability distribution follows directly from the (ideal) heterodyne POVM as $p(\zeta|i) = \frac{1}{\pi}\left|\left\langle\zeta\right|\sqrt{\eta}\alpha_i\right\rangle\right|^2 = \frac{1}{\pi}e^{-\left|\zeta-\sqrt{\eta}\alpha_i\right|^2}$.

We thus have

$$\rho_l = \frac{1}{p(l)} \sum_{z,i} \; p(i, \zeta_{zl}) \, |z\rangle\langle z|_Z \otimes |i\rangle\langle i|_A \otimes |\zeta_{zl}\rangle\langle\zeta_{zl}|_B \otimes \rho_E^i \tag{5.86}$$

$$= \frac{1}{p(l)} \sum_{z,i} \; \frac{1}{M\pi} e^{-\left|\zeta_{zl}-\sqrt{\eta}\alpha_i\right|^2} \, |z\rangle\langle z|_Z \otimes |i\rangle\langle i|_A \otimes |\zeta_{zl}\rangle\langle\zeta_{zl}|_B \otimes \left|\sqrt{1-\eta}\alpha_i\right\rangle\left\langle\sqrt{1-\eta}\alpha_i\right|_E . \tag{5.87}$$

The normalization factor is $p(l) = \sum_{zi} \; \frac{1}{M\pi} e^{-\left|\zeta_{zl}-\sqrt{\eta}\alpha_i\right|^2}$. By symmetry, this is equal to $p(l) = \sum_i \frac{1}{\pi} e^{-\left|\zeta_{0l}-\sqrt{\eta}\alpha_i\right|^2}$.

With the state written out explicitly, we now compute $H(Z|E)_{\rho_l}$ and $H(Z|A)_{\rho_l}$ (we are considering reverse reconciliation).

$H(Z|A)$ is simply calculated on the classical probability distribution $q(z,i) = \frac{p(i,\zeta_{zl})}{p(l)}$ as $H(q_{ZA}) - H(q_A)$.

$H(Z|E)$ is calculated on the CQ state

$$\sum_z \; |z\rangle\langle z|_Z \otimes \sum_i q(z,i) \left|\sqrt{1-\eta}\alpha_i\right\rangle\left\langle\sqrt{1-\eta}\alpha_i\right|_E , \tag{5.88}$$

which expands as

$$H(Z|E)_{\rho_l} = \sum_z H\left(\sum_i q(z,i)\left|\sqrt{1-\eta}\alpha_i\right\rangle\!\!\left\langle\sqrt{1-\eta}\alpha_i\right|_E\right) - H\left(\sum_{zi} q(z,i)\left|\sqrt{1-\eta}\alpha_i\right\rangle\!\!\left\langle\sqrt{1-\eta}\alpha_i\right|_E\right). \quad (5.89)$$

Eve's states are infinite-dimensional, so to evaluate the entropy we use the fact that $H(\sum_i p(i)\left|\psi_i\right\rangle\!\left\langle\psi_i\right|) = H(\sum_{ij}\sqrt{p(i)p(j)}\left\langle\psi_j|\psi_i\right\rangle\left|i\right\rangle\!\left\langle j\right|)$. (This fact follows by considering the pure state $\sum_i \sqrt{p(i)}\left|\psi_i\right\rangle\left|i\right\rangle$, and noting that both reduced density matrices of a bipartite pure state have the same entropy.) Finally, the total key rate is $\int_l p(l)[H(Z|E)_{\rho_l} - f_l H(Z|A)_{\rho_l}]$, where the integral is only taken over l giving a positive contribution to the key rate.

Influence of Error Correction

As mentioned above, when error correction is performed at the Shannon limit, reverse reconciliation does not benefit from postselection. Hence, it is important to account for the realistic cost of error correction. We find that the optimal postselection region is very sensitive to the cost of error correction. Thus, we emphasize that the qualitative conclusions we make in the following section are intimately related to how we have assigned the error correction cost. For different error-correcting codes, the analysis will need to be redone. Our general method however, can always be run again with the new cost modelled, and indeed for experimental runs can directly use the known leakage.

Recall we can think of each l as corresponding to an effective channel, with associated key rate $H(Z|E)_{\rho_l} - H(Z|A)_{\rho_l}$. Importantly, in each effective channel, Alice and Bob share a discrete probability distribution $q(z,i)$. We can thus apply error-correction techniques for discrete-variable protocols. For simplicity, we consider $M = 4$ and assume that Bob and Alice assign their 4 outcomes to 2 bits, via the mapping $0 \to 00$, $1 \to 01$, $2 \to 11$ and $3 \to 10$. This creates an effective binary symmetric channel, with an associated quantum bit error rate (QBER). (Note that the specific choice of mapping minimizes the effective QBER, as the coherent states diametrically opposite each other are the least likely to be confused with each other.) For error correction, we consider the Cascade protocol, which is the prototypical choice for discrete-variable QKD. One caveat is that most error-correction codes for DV protocols are designed to work at low QBER. In our case, effective channels with l close to the origin will have QBER close to $\frac{1}{2}$. It is part of our assumptions on

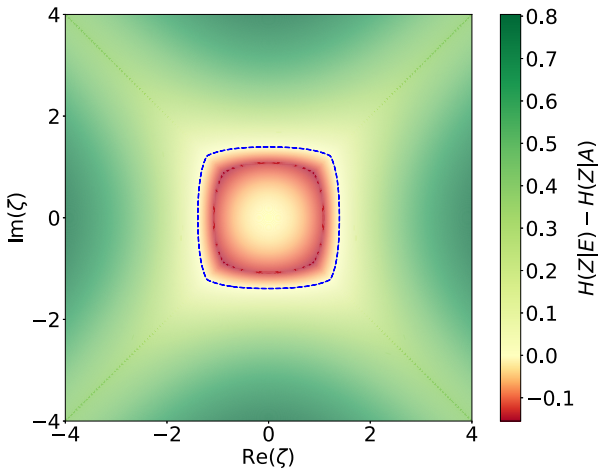


Figure 5.10: Heatmap of the difference $[H(Z|E)_{\rho_l} - f_l H(Z|A)_{\rho_l}]$ over phase space. The contour where this quantity is 0 is indicated by the dashed blue line. The optimal postselection region (i.e. the set of discarded outcomes) extends from the origin to this contour. For 4-PSK with coset announcements and realistic error correction, in the ideal detector and loss-only scenarios, at distance 20 km and $\alpha = 0.7$.

the modelling of error correction that one can find codes extrapolating the performance of Cascade up to maximal QBER.

Denote the QBER for each effective channel l by $QBER_l$. Linearly fitting to the efficiency of Cascade [86], we set $f_l = 1.05 + \frac{5}{11} QBER_l$.[8] The postselected region then consists of those l for which $[H(Z|E)_{\rho_l} - f_l H(Z|A)_{\rho_l}] \leq 0$. In Figure 5.10, we show the optimal postselection region at a distance of 20 km, with $\alpha = 0.7$. The shape is akin to a rounded square, and, interestingly, includes the centers of the received coherent signal states. As an indication of the sensitivity of the optimal postselection region to the error-correction cost, if we set $\beta_l = 0.95$, performing no postselection is optimal.

Postselection Heuristics

We now consider heuristics for better postselection for DMCVQKD without announcements. We will still consider symmetric postselection. This means the first region A_0 is

[8] We also put an upper bound of $\delta_{EC}^{leak} \leq \log_2(M) = 2$.

Amplitude and Phase			Hyperbolic		
Parameter	Initialized Value	Optimal Value	Parameter	Initialized Value	Optimal Value
α	0.6	0.67	α	0.6	0.67
Δ_a	0.5	0.69	Δ_h	0.5	0.40
Δ_p	0.05	0.0047	Δ_s	0.95	0.04

Table 5.1: Optimal protocol parameters for $M = 4$ phase-shift keying, at distance 150 km and 2% excess noise, with error-correction efficiency $\beta_{EC} = 95\%$.

some non-discarded subset of the slice $\theta : [-\frac{\pi}{M}, \frac{\pi}{M}]$, and the remaining regions are obtained by rotations of $\frac{2\pi}{M}$.

We consider a hyperbolic postselection region. For the first region operator, we keep points within the first slice which additionally satisfy $x^2 - \Delta_s y^2 \geq \Delta_h$. The ranges of the parameters are $\Delta_h : [0, \infty)$ and $\Delta_s : [-\infty, \infty]$.

We now compare this hyperbolic scheme to the amplitude and phase scheme. To evaluate the optimal parameters, we use the fminsearch function in MATLAB. We consider a distance of 150 km, with $\xi = 0.02$. We jointly optimize α, Δ_a and Δ_p, for the amplitude and phase postselection pattern. We jointly optimize α, Δ_h and Δ_s for the hyperbolic postselection region. The function fminsearch does not accept ranges on the parameters to be optimized, but instead takes an initialization point. We set fminsearch to target an accuracy of 10^{-6} in the objective function value, and 10^{-2} in the norm of the parameter vector. The error-correction efficiency β_{EC} is set to 0.95.

The parameter results are summarized in Table 5.1. The uncorrected key rate for amplitude and phase is 2.79×10^{-5}, while for hyperbolic it is 2.80×10^{-5}. Both postselection schemes seem to perform essentially the same. Though we recognize that the parameter optimization is not exact, it seems that the optimal values are $\Delta_p = 0$ and $\Delta_s = 0$ (this is consistent with our observations in Sec. 5.6 regarding phase postselection).

We consider another point, this time at distance 10 km and 5% excess noise. We find that the MOSEK solver is not as consistent at this point, so we instead use SDPT3. Moreover, fminsearch gets stuck in local minima, so we instead perform a coarse-grained search jointly over all three parameters. For the amplitude and phase, we optimize α from 0.6 to 1.1 in increments of 0.1, Δ_a from 0.4 to 0.9 in increments of 0.05, and Δ_p from 0 to 0.05 in increments of 0.01. For hyperbolic, we optimize α the same, Δ_h from 0.4 to 0.7 in increments of 0.05, and Δ_s from 0 to 1 in increments of 0.1.

The parameter results are summarized in Table 5.2. The uncorrected key rate for

Amplitude and Phase		Hyperbolic	
Parameter	Optimal Value	Parameter	Optimal Value
α	0.8	α	0.8
Δ_a	0.5	Δ_h	0.4
Δ_p	0	Δ_s	-1

Table 5.2: Optimal protocol parameters for $M = 4$ phase-shift keying, at distance 10 km and 5% excess noise, with error-correction efficiency $\beta_{EC} = 95\%$.

amplitude and phase is 0.028, while for hyperbolic it is 0.027. Again, the two postselection schemes perform similarly. In fact, note that the choice $\Delta_s = -1$ is just a circular arc.

The conclusion we draw is that hyperbolic postselection does not perform much better than amplitude and phase postselection, which in turn works just as well with only amplitude postselection; which is what we will thus consider in the following section.

We remark however that the numerical solver sometimes has large gaps between the approximate primal result and the dual lower bound. This indicates that the dual value is not always tight. Moreover, the approximate primal result can also be inconsistent. This is especially apparent when considering small changes to the input parameters, which sometimes lead to large fluctuations in the key rate, purely due to the unpredictability of numerical solvers. Unlike for key rate plots, we do not simply adjust the returned values by hand (see the discussion at the beginning of Sec. 5.6), as this would defeat the purpose of the comparison. Given a more robust numerical method, one might find somewhat different qualitative conclusions regarding postselection regions; especially with regards to small perturbations of input parameters. We defer a more extensive exploration of postselection regions, especially in concert with different signal state constellations, to future work: especially once the more robust numerical approach from Ref. [61] can be extended to apply to DMCVQKD with the dimension reduction method.

5.7.2 Larger Constellations

Using larger constellations offers a path to improving the performance of DMCVQKD, for two reasons. The first is that Alice has a larger alphabet, allowing her to encode more information per signal state. The second is that the parameter estimation from each signal state further constrains Eve's channel action. These two notions can be decoupled, and this is precisely the idea behind decoy states [87, 88], which are only used to constrain

the channel action, not to generate the raw key. Recall that, asymptotically, parameter estimation is performed using only a vanishing fraction of the rounds of the QKD protocol. Thus, using decoy states can only improve the asymptotic key rate.

In order to understand the impact of decoy states on the performance of DMCVQKD, we consider the limiting case of infinite decoy states. Suppose all coherent states are used as decoy states. For each coherent state, Bob can determine the received conditional state by performing state tomography. Since the coherent states form a basis for density operators on Fock space, the channel action on any state is known, and hence the channel itself is fully determined.

Thus, in the limit of infinite decoy states, Alice and Bob know their shared state ρ_{AB}. The key rate objective function is then evaluated directly on the known state. Computing the key rates for $M = 4$ phase-shift keying in this manner, we find they are very similar to the key rates without using decoy states. We are led to conclude that, given an extra state prepared by Alice, it is more beneficial to put it towards key generation than decoy states, and hence we focus on this idea in our constellations.

Again, we refer the reader to Sec. 5.6 for all details regarding the simulation for the following plots. Based on our findings in previous sections, we only use amplitude postselection for phase-shift keying.

In Fig. 5.11, we compare the performance of $M = 4$ and $M = 8$ phase-shift keying. As our interest is in the relative performance of these two schemes, we plot the uncorrected values rather than the corrected key rates. The protocol parameters for $M = 4$ are the same as in Fig. 5.5. For $M = 8$, first the signal state amplitude α is coarsely optimized over the range 0.6 to 2 in increments of 0.1, with no postselection. Then, with α_{opt} fixed, the postselection parameter Δ_a is optimized over the range 0.5 to 1.4, also in increments of 0.1 ($\Delta_p = 0$).

At 0% and 1 % excess noise, the larger constellation has a relatively modest improvement over the smaller one. The improvement is approximately a constant factor, though it is slightly more pronounced at longer distances. At 2% excess noise, the difference is much more significant. For 8-PSK, the key rate scaling is approximately linear, with a slope only slightly decreased from the loss-only case, while for 4-PSK, the scaling is initially linear but drops off at around 120 km. At 5% excess noise, 8-PSK almost doubles the maximum distance, and also achieves a large constant factor improvement in the key rate.

It thus seems that using larger constellations is most impactful at somewhat higher values of excess noise. In practice, the choice to use a larger constellation would have to be balanced with an associated increase in preparation noise. We remark that once the preparation noise is characterized (say as a function of M), it can be easily modelled in our

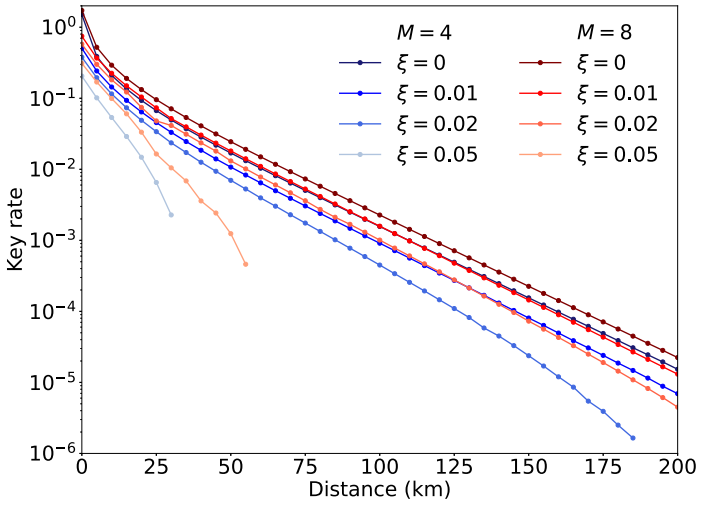


Figure 5.11: Ideal detector uncorrected values for phase-shift keying with $M = 4$ and $M = 8$. Protocols are evaluated with optimized protocol parameters, including postselection. The subspace dimension parameter is $N = 10$.

existing numerics. One simply adds an additional term expanding the constraint on the trace distance between the simulated and actual reduced density matrices (see the third line of Eq. (5.75)).[9] With this, one could optimize the value of M taking into account realistic modulator performance.

At least in classical telecommunications, it is known that even larger values of M in phase-shift keying offer only marginal improvements. We find a similar saturation of performance when testing $M = 16$ PSK. As an alternative, one typically considers modulation schemes which encode information in both phase *and* amplitude. This is known as amplitude and phase-shift keying (APSK).

Classically, there is an extensive body of literature on the design of discrete-modulated constellations which approach the performance of Gaussian modulation (e.g. see Ref. [89]). These designs include variations on the location of the signal states, as well as the a priori probabilities with which they are prepared. They provide good heuristics for designing quantum constellations in the context of DMCVQKD.

A fairly extensive exploration of different classical constellations is completed in Ref. [90]. We consider the so-called *(5,11)* constellation, which offers some of the best perfor-

[9] This observation regarding modelling source imperfections is due to Shlok Nahar.

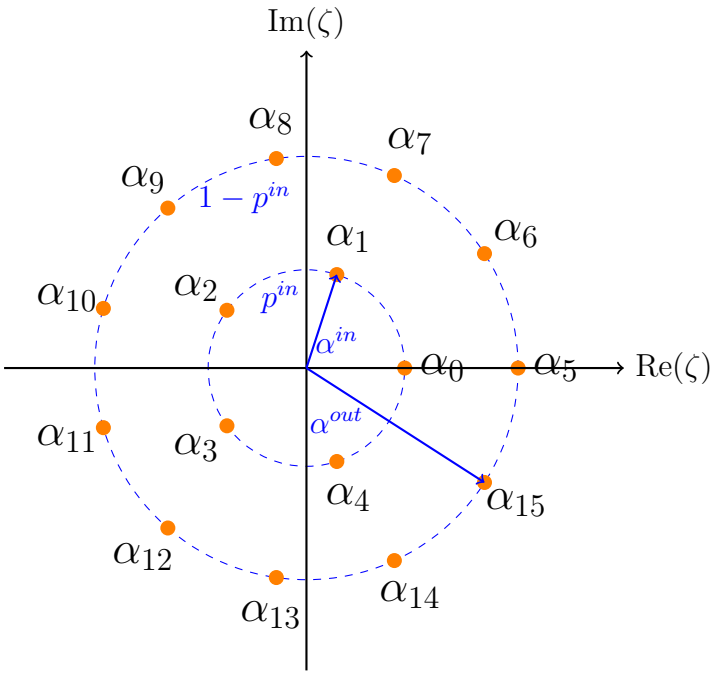


Figure 5.12: Coherent signal state constellation for the (5,11) protocol. It consists of 16 signal states, 5 spaced equally on the inner circle with radius α^{in}, and 11 on the outer circle with radius α^{out}. Each inner state is chosen with probability $\frac{p^{in}}{5}$, and each outer one with probability $\frac{1-p^{in}}{11}$.

mance of all 16 state modulations that are considered in Ref. [90]. As the name suggests, the (5,11) constellation consists of two concentric rings with 5 and 11 states respectively. As a natural decoding scheme, we construct the region operators in a similar pattern. A visualization of this constellation and its parameters is provided in Fig. 5.12, while the region operators and their parameters are depicted in Fig. 5.13. We considered a relative phase shift between the two rings, but found that this had a negligible impact on the key rate.

In Fig. 5.14, we compare the performance of the (5,11) discrete-modulated constellation with Gaussian modulation. We use $\xi = 0.01$ which is a typical value for channel excess noise. For simplicity, the DM protocol parameters are coarsely optimized at a distance of 100 km, and the same parameters are used at all distances (see Table 5.3). This is evident at very short distances, where the key rate could be pushed higher by using larger signal state amplitudes (see Fig. 5.6 for a representative example of how the signal state

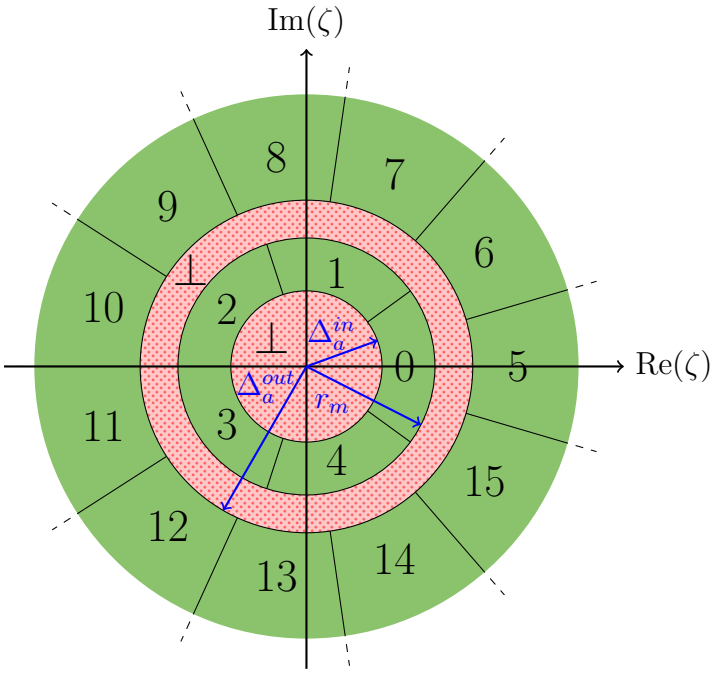


Figure 5.13: Phase space regions for the reverse reconciliation key map for the (5,11) protocol. Bob obtains the measurement result ζ from the heterodyne detector. He maps the outcome to the symbol of the region containing ζ. Δ_a^{in} and Δ_a^{out} are amplitude postselection parameters corresponding to the radii of the innermost and outermost circles. r_m is the radius of the middle circle. The two $\perp$ regions correspond to signals that are discarded.

Parameter	Optimal Value
α^{in}	0.743
α^{out}	1.708
p^{in}	0.623
Δ_a^{in}	0.319
Δ_a^{out}	0.977
r_m	0.977

Table 5.3: Protocol parameters used in Fig. 5.14 for the (5,11) protocol. Values are approximately optimal at distance 100 km and 1% excess noise. Refer to Fig. 5.12 and Fig. 5.13 for descriptions of all the parameters.

amplitude is large at zero distance but quickly decreases, plateauing at around 40 km). We use a subspace dimension parameter of $N = 5$, which at the considered value of excess noise, is sufficient to ensure the correction term is small.

The DMCVQKD protocol with just 16 modulated states achieves essentially the same key rate as Gaussian modulation, even at very long distances and with excess noise. This demonstrates that discrete-modulated CVQKD is a promising alternative to Gaussian modulation. As we have noted before, a more thorough comparison of the protocols, delving into implementation details like error-correction cost and modulation precision, would be in order.

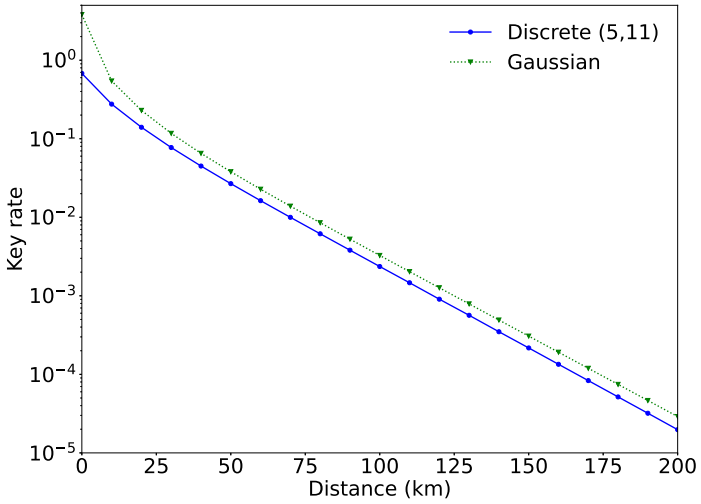


Figure 5.14: Comparison of Gaussian-modulated [82] CVQKD with the (5,11) DMCVQKD protocol. Key rates are plotted against distance, at excess noise $\xi = 0.01$, in the ideal detector scenario. For Gaussian modulation, the variance is optimized at each distance. For discrete modulation, the protocol parameters are approximately optimized at a distance of 100 km, and the same parameters are then used at all distances. The subspace dimension parameter is $N = 5$.

Chapter 6

Comparison to Flag-State Squasher

Our dimension reduction (DR) approach encompasses another method known as the flag-state squasher (FSS) [9]. The FSS also obtains lower bounds on the key rate by solving a finite-dimensional optimization. However, the FSS is restricted to protocols where both the key map POVM elements and constraint observables commute with the projection. Notably, this is not the case for DMCVQKD. In this section, we compare our method to the FSS both analytically and numerically. This demonstrates the advantages of our method and offers further insight into the FSS approach.

6.1 Analytical Comparison

We briefly summarize the FSS, deferring a complete description to Ref. [9]. As usual, Alice's POVM is given by $\{|j\rangle\langle j|_A\}$, while Bob's POVM is $\{\Gamma_B^i\}$. The corresponding probabilities are $\{\gamma_{ij}\}$. The FSS also requires choosing a projection $\Pi = \mathbb{1}_A \otimes \Pi_B$ onto a finite subspace and upper-bounding the weight W outside that subspace. It is assumed that $[\Gamma_B^i, \Pi_B] = 0$.

Define a flag Hilbert space $\mathcal{H}_F$, with dimension equal to the number of elements in Bob's POVM. The finite-dimensional optimization is over density matrices in $\mathcal{H}_N \oplus \mathcal{H}_F$. Bob's new POVM is $\tilde{\Gamma}^i = \Pi\ \Gamma^i\ \Pi \oplus |i\rangle\langle i|_F$. Alice's POVM and the expectation values are unchanged. The objective function is $f_{FSS}(\tilde{\rho}_N \oplus \tilde{\sigma}_F) = f(\tilde{\rho}_N)$, i.e. it simply discards the flag portion and evaluates the usual key rate function on the remaining portion. This completes the formulation of the finite-dimensional optimization for the FSS.

The squashing map Λ_B is a channel from $\mathcal{H}_\infty$ to $\mathcal{H}_N \oplus \mathcal{H}_F$ defined as $\Lambda_B(\rho_\infty) = \Pi\rho_\infty\Pi \oplus \sum_i \mathrm{Tr}_B[\rho_\infty(\Gamma^i - \Pi\Gamma^i\Pi)]\,|i\rangle\langle i|_F$. Note that ρ_∞ is feasible for the original infinite-dimensional optimization if and only if $\Lambda(\rho_\infty)$ is feasible for the finite-dimensional one. In this sense, we can think of the flag-state squasher as implicitly solving over the tightest possible choice of $\mathbf{S}_N$, namely $\Pi\mathbf{S}_\infty\Pi$.

As a special case, our method can be applied to any protocol admitting a flag-state squasher. The FSS requires choosing a projection Π and getting a bound on weight W, which establishes the first two steps of our method. Since the key map POVM elements commute with the projection, we can set $\Delta = 0$ (Theorem 3). All observables are POVM elements, so we can use the explicit form of $\mathbf{S}_N$ in Eq. (4.101). This establishes the last two steps of our method. We can now compare both approaches in the following theorem.

Theorem 8. *For a fixed projection Π and weight W, our dimension reduction (DR) method gives the same key rate as the flag-state squasher (FSS) when $\mathbf{S}_N = \Pi\mathbf{S}_\infty\Pi$.*

Proof. Let $\Lambda(\rho_\infty) = \tilde{\rho}_N \oplus \tilde{\sigma}_F$ be a state reaching the minimum in the FSS optimization. By definition, $f_{FSS}(\tilde{\rho}_N \oplus \tilde{\sigma}_F) = f_{DR}(\tilde{\rho}_N)$. By the definition of Λ, $\tilde{\rho}_N = \Pi\rho_\infty\Pi$. Thus, $\tilde{\rho}_N \in \mathbf{S}_N$, so is feasible for the dimension reduction optimization. Since this optimization is a minimization, $R^\infty_{FSS} \geq R^\infty_{DR}$.

Conversely, let $\tilde{\rho}_N$ be a subnormalized state reaching the minimum in the dimension reduction optimization. By the definition of $\mathbf{S}_N$, $\tilde{\rho}_N = \Pi\rho_\infty\Pi$ for some state $\rho_\infty \in \mathbf{S}_\infty$. By definition, $f_{DR}(\tilde{\rho}_N) = f_{FSS}(\Lambda(\rho_\infty))$. By the property of the squashing map, $\Lambda(\rho_\infty)$ is feasible for the FSS optimization. Since the FSS optimization is a minimization, $R^\infty_{DR} \geq R^\infty_{FSS}$. □

If $\mathbf{S}_N$ is not chosen optimally, then our dimension reduction method gives a lower key rate. In practice, our explicit prescription for choosing $\mathbf{S}_N$ in Eq. (4.101) gives very similar key rates to the flag state squasher (see Sec. 6.2), suggesting this choice is essentially optimal.

In addition to being more general, our method has an important advantage compared to the flag-state squasher. Our finite-dimensional optimization is over a smaller Hilbert space, since we do not require flag-state dimensions. Therefore, if we compare at a fixed *total* dimension, which roughly determines the runtime, our method can give higher key rates than the flag-state squasher. Some protocols can have a very large number of POVM constraints. For the flag-state squasher, using all the constraints would make the runtime prohibitive, as the dimension of the problem depends on the number of constraints. Thus, a smaller set of coarse-grained POVM elements is typically used. Our dimension reduction

method is not limited by the number of constraints and can thus handle the fine-grained POVM directly, potentially giving better key rates.

6.2 Numerical Comparison: Unbalanced Phase-Encoded BB84

Having provided an analytical comparison between our method and the flag-state squasher, we now perform a sample numerical comparison of the two methods. We will consider the unbalanced phase-encoded BB84 protocol, to which the flag state squasher has recently been applied in Ref. [10]. We defer a complete description of the protocol to Ref. [10]. Briefly, this is a phase-encoded BB84 protocol where Alice and Bob's interferometers each have a loss $1-\kappa$ only in the arm with the phase modulator; hence the term unbalanced. The projection is $\Pi_B = \sum_{\substack{0\leq n_1,n_2 \\ n_1+n_2\leq N}} |n_1,n_2\rangle\langle n_1,n_2|$, where $|n_1,n_2\rangle$ are two-mode Fock basis states. The weight W is bounded using the fact that the frequency of cross-clicks increases with photon number [10].

In Fig. 6.1 we compare the key rates from our method and the flag-state squasher, for a channel with transmittance η and for different interferometer asymmetric transmittance κ. All parameters are the same as in Figure 3(a) of Ref. [10], and the signal state intensity is optimized separately for each method and parameter choice. We see that our method gives essentially identical key rates. In conjunction with Theorem 8, this provides strong numerical evidence that our heuristic choice of $\mathbf{S}_N$ in Eq. (4.101) is tight. While a more thorough benchmarking would be in order, we remark that for generating the data in Fig. 6.1, our method was approximately five times faster than the flag-state squasher as implemented in Ref. [10] (using the same SDPT3 solver [91, 92]).

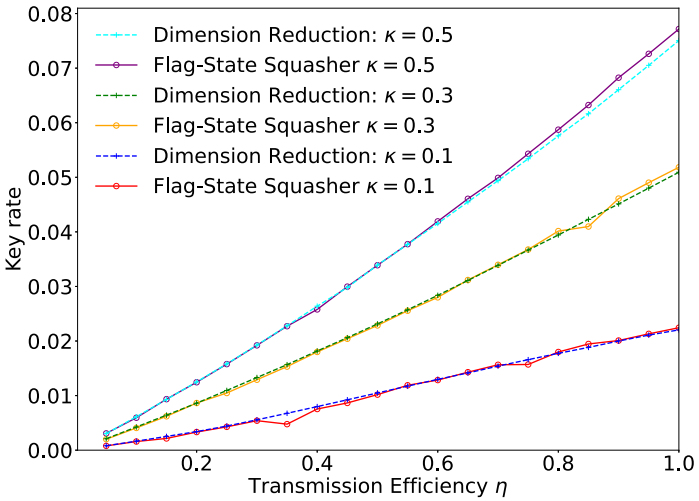


Figure 6.1: Key rates for unbalanced phase-encoded BB84 versus transmission efficiency η, for different values of asymmetric interferometer loss $1 - \kappa$. It is clear that the key rates from our dimension reduction method are nearly identical to those from the flag-state squasher, indicating the tightness of our method in practice. In generating the data for this graph, our dimension reduction method was approximately five times faster than the flag-state squasher as implemented in Ref. [10].

Chapter 7

Conclusion

In summary, we establish a framework to lower bound a large dimensional convex optimization using a judiciously chosen smaller dimensional one. We show how this framework can be used to reduce the dimension of QKD key rate calculations. This allows existing numerical tools for finite-dimensional key rate calculations to be applied to protocols in infinite-dimensional Hilbert spaces. This extends the advantages of numerical key rate calculations, including modelling imperfections in devices and exploring classical postprocessing strategies, to such protocols. An important application of our method is to prove the asymptotic security of DMCVQKD with an arbitrary constellation and postselection parameters.

As concrete examples, we apply this method to the quadrature phase-shift keying scheme in both ideal and trusted detector noise scenarios. We show that discrete modulation key rates can scale similarly to Gaussian modulation. Moreover, we rigorously demonstrate that postselection of data can improve the key rates for DMCVQKD. We harness the numerical framework to explore different strategies to improve the key rate. We demonstrate that discrete-modulated CVQKD, with as few as 16 modulated states, can achieve key rates extremely similar to Gaussian-modulated CVQKD.

We also show how the flag-state squasher can be understood in the language of the dimension reduction method, and compare the methods analytically. Using unbalanced phase-encoded BB84 as an example, we show that our approach can achieve key rates nearly identical to those from the flag-state squasher, while having an improved runtime.

Three directions for future work are as follows.

The first direction is using our flexible numerical key rate framework to explore further variations on the DMCVQKD protocol. As has been discussed, there is a wealth of liter-

ature on optimizing the performance of classical discrete-modulated communication. It is both interesting and practically relevant to see how these strategies can be adapted to the quantum case to improve the key rate.

Another avenue of research is applying the dimension reduction method to obtain security proofs for other unsolved protocols. One interesting example is Gaussian modulation with postselection. Due to the postselection step, it can no longer be assumed that the optimal attack is Gaussian. By projecting on both Alice and Bob's Fock spaces, we expect one could establish the security of this protocol using the dimension reduction method. Beyond QKD, the dimension reduction method can be applied to study other quantum communication protocols, such as entanglement verification.

Finally, given the recent development of a finite-key numerical framework [62], we hope to see the dimension reduction method extended to finite-key analysis of protocols in infinite-dimensional Hilbert spaces. We expect that key elements of the method, including bounding the weight outside the subspace and expanding the feasible set, will lift to the finite-key analysis.

References

[1] T. Upadhyaya, T. van Himbeeck, J. Lin, and N. Lütkenhaus, *Dimension Reduction in Quantum Key Distribution for Continuous- and Discrete-Variable Protocols*, PRX Quantum **2**, 020325 (2021).

[2] R. Colbeck and R. Renner, *No extension of quantum theory can have improved predictive power*, Nature Communications **2**, 411 (2011).

[3] P. J. Coles, E. M. Metodiev, and N. Lütkenhaus, *Numerical approach for unstructured quantum key distribution*, Nature Communications **7**, 11712 (2016).

[4] A. Winick, N. Lütkenhaus, and P. J. Coles, *Reliable numerical key rates for quantum key distribution*, Quantum **2**, 77 (2018).

[5] N. J. Beaudry, T. Moroder, and N. Lütkenhaus, *Squashing Models for Optical Measurements in Quantum Communication*, Phys. Rev. Lett. **101**, 093601 (2008).

[6] T. Tsurumaru and K. Tamaki, *Security proof for quantum-key-distribution systems with threshold detectors*, Phys. Rev. A **78**, 032302 (2008).

[7] T. Tsurumaru, *Squash operator and symmetry*, Phys. Rev. A **81**, 012328 (2010).

[8] O. Gittsovich, N. J. Beaudry, V. Narasimhachar, R. R. Alvarez, T. Moroder, and N. Lütkenhaus, *Squashing model for detectors and applications to quantum-key-distribution protocols*, Phys. Rev. A **89**, 012325 (2014).

[9] Y. Zhang, P. J. Coles, A. Winick, J. Lin, and N. Lütkenhaus, *Security proof of practical quantum key distribution with detection-efficiency mismatch*, Phys. Rev. Research **3**, 013076 (2021).

[10] N. K. H. Li and N. Lütkenhaus, *Improving key rates of the unbalanced phase-encoded BB84 protocol using the flag-state squashing model*, Phys. Rev. Research **2**, 043172 (2020).

[11] T. C. Ralph, *Continuous variable quantum cryptography*, Phys. Rev. A **61**, 010303 (1999).

[12] M. Hillery, *Quantum cryptography with squeezed states*, Phys. Rev. A **61**, 022309 (2000).

[13] C. Silberhorn, T. C. Ralph, N. Lütkenhaus, and G. Leuchs, *Continuous Variable Quantum Cryptography: Beating the 3 dB Loss Limit*, Phys. Rev. Lett. **89**, 167901 (2002).

[14] F. Grosshans and P. Grangier, *Continuous Variable Quantum Cryptography Using Coherent States*, Phys. Rev. Lett. **88**, 057902 (2002).

[15] F. Grosshans, G. Van Assche, J. Wenger, R. Brouri, N. J. Cerf, and P. Grangier, *Quantum key distribution using gaussian-modulated coherent states*, Nature **421**, 238 (2003).

[16] C. Weedbrook, A. M. Lance, W. P. Bowen, T. Symul, T. C. Ralph, and P. K. Lam, *Quantum Cryptography Without Switching*, Phys. Rev. Lett. **93**, 170504 (2004).

[17] Y.-B. Zhao, M. Heid, J. Rigas, and N. Lütkenhaus, *Asymptotic security of binary modulated continuous-variable quantum key distribution under collective attacks*, Phys. Rev. A **79**, 012307 (2009).

[18] K. Brádler and C. Weedbrook, *Security proof of continuous-variable quantum key distribution using three coherent states*, Phys. Rev. A **97**, 022310 (2018).

[19] T. Matsuura, K. Maeda, T. Sasaki, and M. Koashi, *Finite-size security of continuous-variable quantum key distribution with digital signal processing*, Nature Communications **12**, 252 (2021).

[20] E. Kaur, S. Guha, and M. M. Wilde, *Asymptotic security of discrete-modulation protocols for continuous-variable quantum key distribution*, Phys. Rev. A **103**, 012412 (2021).

[21] A. Denys, P. Brown, and A. Leverrier, *Explicit asymptotic secret key rate of continuous-variable quantum key distribution with an arbitrary modulation*, arXiv:2103.13945 [quant-ph] (2021).

[22] S. Ghorai, P. Grangier, E. Diamanti, and A. Leverrier, *Asymptotic Security of Continuous-Variable Quantum Key Distribution with a Discrete Modulation*, Phys. Rev. X **9**, 021059 (2019).

[23] J. Lin, T. Upadhyaya, and N. Lütkenhaus, *Asymptotic Security Analysis of Discrete-Modulated Continuous-Variable Quantum Key Distribution*, Phys. Rev. X **9**, 041064 (2019).

[24] N. Killoran and N. Lütkenhaus, *Strong quantitative benchmarking of quantum optical devices*, Phys. Rev. A **83**, 052320 (2011).

[25] J. Watrous, *The Theory of Quantum Information*, 1st ed. (Cambridge University Press, Cambridge, UK, 2018).

[26] J. Crann, D. W. Kribs, and V. I. Paulsen, Functional Analysis for Quantum Information (2021).

[27] M. Ohya and D. Petz, *Quantum Entropy and Its Use* (Springer-Verlag Berlin Heidelberg, New York, USA, 2004).

[28] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, USA, 2004).

[29] C. Gerry and P. Knight, *Introductory Quantum Optics* (Cambridge University Press, Cambridge, UK, 2004).

[30] F. J. Beutler, *The operator theory of the pseudo-inverse II. Unbounded operators with arbitrary range*, Journal of Mathematical Analysis and Applications **10**, 471 (1965).

[31] C. A. Fuchs and J. van de Graaf, *Cryptographic distinguishability measures for quantum-mechanical states*, IEEE Transactions on Information Theory **45**, 1216 (1999).

[32] T. Ogawa and H. Nagaoka, *A new proof of the channel coding theorem via hypothesis testing in quantum information theory*, in *Proceedings IEEE International Symposium on Information Theory,* (2002) p. 73, *Full text available at arXiv:quant-ph/0208139*.

[33] A. Winter, *Coding theorem and strong converse for quantum channels*, IEEE Transactions on Information Theory **45**, 2481 (1999).

[34] J. M. Borwein and A. S. Lewis, *Partially finite convex programming, Part II: Explicit lattice models*, Mathematical Programming **57**, 49 (1992).

[35] R. Shankar, *Principles of Quantum Mechanics, 2nd Ed.* (Springer Science+Business Media New York, New York, USA, 2014).

[36] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, *Gaussian quantum information*, Rev. Mod. Phys. **84**, 621 (2012).

[37] K. Husimi, *Some Formal Properties of the Density Matrix*, Proceedings of the Physico-Mathematical Society of Japan. 3rd Series **22**, 264 (1940).

[38] W. K. Wootters and W. H. Zurek, *A single quantum cannot be cloned*, Nature **299**, 802 (1982).

[39] V. Coffman, J. Kundu, and W. K. Wootters, *Distributed entanglement*, Phys. Rev. A **61**, 052306 (2000).

[40] W. Diffie and M. Hellman, *New directions in cryptography*, IEEE Transactions on Information Theory **22**, 644 (1976).

[41] D. Stebila, M. Mosca, and N. Lütkenhaus, *The Case for Quantum Key Distribution*, in *Quantum Communication and Quantum Networking*, edited by A. Sergienko, S. Pascazio, and P. Villoresi (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010) pp. 283–296.

[42] S. Wiesner, *Conjugate Coding*, SIGACT News **15**, 78–88 (1983), *Work originally completed around 1969.*

[43] 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* (IEEE, New York, USA, 1984) pp. 175–179.

[44] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, *The security of practical quantum key distribution*, Rev. Mod. Phys. **81**, 1301 (2009).

[45] F. Xu, X. Ma, Q. Zhang, H.-K. Lo, and J.-W. Pan, *Secure quantum key distribution with realistic devices*, Rev. Mod. Phys. **92**, 025002 (2020).

[46] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani, J. L. Pereira, M. Razavi, J. S. Shaari, M. Tomamichel, V. C. Usenko, G. Vallone, P. Villoresi, and P. Wallden, *Advances in quantum cryptography*, Adv. Opt. Photon. **12**, 1012 (2020).

[47] R. Ursin, F. Tiefenbacher, T. Schmitt-Manderbach, H. Weier, T. Scheidl, M. Lindenthal, B. Blauensteiner, T. Jennewein, J. Perdigues, P. Trojek, B. Ömer, M. Fürst,

M. Meyenburg, J. Rarity, Z. Sodnik, C. Barbieri, H. Weinfurter, and A. Zeilinger, *Entanglement-based quantum communication over 144 km*, Nature Physics **3**, 481 (2007).

[48] S.-K. Liao, W.-Q. Cai, W.-Y. Liu, L. Zhang, Y. Li, J.-G. Ren, J. Yin, Q. Shen, Y. Cao, Z.-P. Li, F.-Z. Li, X.-W. Chen, L.-H. Sun, J.-J. Jia, J.-C. Wu, X.-J. Jiang, J.-F. Wang, Y.-M. Huang, Q. Wang, Y.-L. Zhou, L. Deng, T. Xi, L. Ma, T. Hu, Q. Zhang, Y.-A. Chen, N.-L. Liu, X.-B. Wang, Z.-C. Zhu, C.-Y. Lu, R. Shu, C.-Z. Peng, J.-Y. Wang, and J.-W. Pan, *Satellite-to-ground quantum key distribution*, Nature **549**, 43 (2017).

[49] D. Stucki, N. Gisin, O. Guinnard, G. Ribordy, and H. Zbinden, *Quantum key distribution over 67 km with a plug&play system*, New Journal of Physics **4**, 41 (2002).

[50] C. H. Bennett, G. Brassard, and N. D. Mermin, *Quantum cryptography without Bell's theorem*, Phys. Rev. Lett. **68**, 557 (1992).

[51] F. Grosshans, N. J. Cerf, J. Wenger, R. Tualle-Brouri, and P. Grangier, *Virtual Entanglement and Reconciliation Protocols for Quantum Cryptography with Continuous Variables*, Quantum Info. Comput. **3**, 535–552 (2003).

[52] M. Curty, M. Lewenstein, and N. Lütkenhaus, *Entanglement as a Precondition for Secure Quantum Key Distribution*, Phys. Rev. Lett. **92**, 217903 (2004).

[53] I. Devetak and A. Winter, *Distillation of secret key and entanglement from quantum states*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **461**, 207 (2005).

[54] R. Renner, *Security of Quantum Key Distribution*, Ph.D. thesis, ETH Zurich (2005), available at http://arxiv.org/abs/quant-ph/0512258.

[55] C. Portmann and R. Renner, *Cryptographic security of quantum key distribution*, arXiv:1409.3525 [quant-ph] (2014).

[56] R. Renner, *Symmetry of large physical systems implies independence of subsystems*, Nature Physics **3**, 645 (2007).

[57] R. Renner and J. I. Cirac, *de Finetti Representation Theorem for Infinite-Dimensional Quantum Systems and Applications to Quantum Cryptography*, Phys. Rev. Lett. **102**, 110504 (2009).

[58] H.-K. Lo, M. Curty, and B. Qi, *Measurement-Device-Independent Quantum Key Distribution*, Phys. Rev. Lett. **108**, 130503 (2012).

[59] J. Barrett, L. Hardy, and A. Kent, *No Signaling and Quantum Key Distribution*, Phys. Rev. Lett. **95**, 010503 (2005).

[60] M. Frank and P. Wolfe, *An algorithm for quadratic programming*, Naval Research Logistics Quarterly **3**, 95 (1956), https://onlinelibrary.wiley.com/doi/pdf/10.1002/nav.3800030109 .

[61] H. Hu, J. Im, J. Lin, N. Lütkenhaus, and H. Wolkowicz, *Robust Interior Point Method for Quantum Key Distribution Rate Computation*, arXiv:2104.03847 [quant-ph] (2021).

[62] I. George, J. Lin, and N. Lütkenhaus, *Numerical calculations of the finite key rate for general quantum key distribution protocols*, Phys. Rev. Research **3**, 013274 (2021).

[63] A. Winter, *Tight Uniform Continuity Bounds for Quantum Entropies: Conditional Entropy, Relative Entropy Distance and Energy Constraints*, Communications in Mathematical Physics **347**, 291 (2016).

[64] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics (Springer Science+Business Media New York, New York, USA, 1996).

[65] R. García-Patrón and N. J. Cerf, *Unconditional Optimality of Gaussian Attacks against Continuous-Variable Quantum Key Distribution*, Phys. Rev. Lett. **97**, 190503 (2006).

[66] M. Navascués, F. Grosshans, and A. Acín, *Optimality of Gaussian Attacks in Continuous-Variable Quantum Cryptography*, Phys. Rev. Lett. **97**, 190502 (2006).

[67] A. Leverrier, F. Grosshans, and P. Grangier, *Finite-size analysis of a continuous-variable quantum key distribution*, Phys. Rev. A **81**, 062343 (2010).

[68] A. Leverrier, *Security of Continuous-Variable Quantum Key Distribution via a Gaussian de Finetti Reduction*, Phys. Rev. Lett. **118**, 200501 (2017).

[69] F. Grosshans and P. Grangier, *Reverse reconciliation protocols for quantum cryptography with continuous variables*, arXiv:quant-ph/0204127 [quant-ph] (2021).

[70] P. Jouguet, S. Kunz-Jacques, A. Leverrier, P. Grangier, and E. Diamanti, *Experimental demonstration of long-distance continuous-variable quantum key distribution*, Nature Photonics **7**, 378 (2013).

[71] V. C. Usenko and R. Filip, *Trusted Noise in Continuous-Variable Quantum Key Distribution: A Threat and a Defense*, Entropy **18**, 10.3390/e18010020 (2016).

[72] J. Lin and N. Lütkenhaus, *Trusted Detector Noise Analysis for Discrete Modulation Schemes of Continuous-Variable Quantum Key Distribution*, Phys. Rev. Applied **14**, 064030 (2020).

[73] M. Heid and N. Lütkenhaus, *Efficiency of coherent-state quantum cryptography in the presence of loss: Influence of realistic error correction*, Phys. Rev. A **73**, 052316 (2006).

[74] W. Heisenberg, *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*, Zeitschrift für Physik **43**, 172 (1927).

[75] D. Stoler, *Equivalence Classes of Minimum Uncertainty Packets*, Phys. Rev. D **1**, 3217 (1970).

[76] K. E. Cahill and R. J. Glauber, *Ordered Expansions in Boson Amplitude Operators*, Phys. Rev. **177**, 1857 (1969).

[77] C. Pacher, J. Martinez-Mateo, J. Duhme, T. Gehring, and F. Furrer, *Information Reconciliation for Continuous-Variable Quantum Key Distribution using Non-Binary Low-Density Parity-Check Codes*, arXiv:1602.09140 [quant-ph] (2016).

[78] K. Gümüş, T. A. Eriksson, M. Takeoka, M. Fujiwara, M. Sasaki, L. Schmalen, and A. Alvarado, *A novel error correction protocol for continuous variable quantum key distribution*, Scientific Reports **11**, 10465 (2021).

[79] M. Grant and S. Boyd, CVX: Matlab Software for Disciplined Convex Programming, version 2.1, `http://cvxr.com/cvx` (2014).

[80] M. C. Grant and S. P. Boyd, *Graph Implementations for Nonsmooth Convex Programs*, in *Recent Advances in Learning and Control*, edited by V. D. Blondel, S. P. Boyd, and H. Kimura (Springer London, London, UK, 2008) pp. 95–110.

[81] M. ApS, The MOSEK optimization toolbox for MATLAB Manual. Version 8.0.0.60. (2016).

[82] S. Fossier, E. Diamanti, T. Debuisschert, R. Tualle-Brouri, and P. Grangier, *Improvement of continuous-variable quantum key distribution systems by using optical preamplifiers*, Journal of Physics B: Atomic, Molecular and Optical Physics **42**, 114014 (2009).

[83] U. Maurer, *Secret key agreement by public discussion from common information*, IEEE Transactions on Information Theory **39**, 733 (1993).

[84] D. Sych and G. Leuchs, *Coherent state quantum key distribution with multi letter phase-shift keying*, New Journal of Physics **12**, 053019 (2010).

[85] A. D. Wyner, *The wire-tap channel*, Bell System Technical Journal **54**, 1355 (1975).

[86] J. Martinez-Mateo, C. Pacher, M. Peev, A. Ciurana, and V. Martin, *Demystifying the information reconciliation protocol Cascade*, Quantum Information and Computation , 453 (2015).

[87] W.-Y. Hwang, *Quantum Key Distribution with High Loss: Toward Global Secure Communication*, Phys. Rev. Lett. **91**, 057901 (2003).

[88] H.-K. Lo, X. Ma, and K. Chen, *Decoy State Quantum Key Distribution*, Phys. Rev. Lett. **94**, 230504 (2005).

[89] H. Méric, *Approaching the Gaussian Channel Capacity With APSK Constellations*, IEEE Communications Letters **19**, 1125 (2015).

[90] C. Thomas, M. Weidner, and S. Durrani, *Digital Amplitude-Phase Keying with M-Ary Alphabets*, IEEE Transactions on Communications **22**, 168 (1974).

[91] K. C. Toh, M. J. Todd, and R. H. Tütüncü, *SDPT3 — A Matlab software package for semidefinite programming, Version 1.3*, Optimization Methods and Software **11**, 545 (1999).

[92] R. H. Tütüncü, K. C. Toh, and M. J. Todd, *Solving semidefinite-quadratic-linear programs using SDPT3*, Mathematical Programming **95**, 189 (2003).

APPENDICES

Appendix A

Uniform Continuity Bound on Conditional Entropy

Here we prove an extension of Lemma 2 in Ref. [63] to subnormalized states. Our development closely parallels that result. Although we are only interested in showing the conditional entropy is uniformly close to decreasing under projection with correction Δ, we will effectively have to derive uniform continuity to determine Δ; so for completeness we give the overall uniform continuity bound as well. Note that the correction term in Eq. (A.3) is smaller than Eq. (A.2).

Theorem 9 (Uniform Continuity and UCDUP of Conditional Entropy)**.** *Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be two Hilbert spaces where the dimension of $\mathcal{H}_A$ is $|A|$ while $\mathcal{H}_B$ can be infinite-dimensional. Let $\tilde{\rho}_{AB}, \tilde{\sigma}_{AB} \in \tilde{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be two subnormalized states; we will omit the system subscripts for readability. WLOG, suppose $\operatorname{Tr}(\tilde{\rho}) \geq \operatorname{Tr}(\tilde{\sigma})$. Let $\frac{1}{2}\|\tilde{\rho} - \tilde{\sigma}\|_1 \leq \epsilon \leq 1$, $\frac{1}{2}\operatorname{Tr}(\tilde{\rho} - \tilde{\sigma}) = \delta$ and $\frac{1}{2}\operatorname{Tr}(\tilde{\rho} + \tilde{\sigma}) = a$. Let $\epsilon' = \epsilon + \delta$ and $\epsilon'' = \epsilon - \delta$. Then, it holds that*

$$|H(A|B)_{\tilde{\rho}} - H(A|B)_{\tilde{\sigma}}| \leq 2\epsilon \log_2 |A| + (a+\epsilon) \max\left\{ h\left(\frac{\epsilon'}{a+\epsilon}\right), h\left(\frac{\epsilon''}{a+\epsilon}\right)\right\}. \tag{A.1}$$

If $\tilde{\rho}$ and $\tilde{\sigma}$ are classical-quantum states, that is $\tilde{\rho} = \sum_{i=1}^{|A|} |i\rangle\langle i|_A \otimes \tilde{\rho}_B^i$ and $\tilde{\sigma} = \sum_{i=1}^{|A|} |i\rangle\langle i|_A \otimes \tilde{\sigma}_B^i$, then

$$|H(A|B)_{\tilde{\rho}} - H(A|B)_{\tilde{\sigma}}| \leq \epsilon' \log_2 |A| + (a+\epsilon) \max\left\{ h\left(\frac{\epsilon'}{a+\epsilon}\right), h\left(\frac{\epsilon''}{a+\epsilon}\right)\right\}, \tag{A.2}$$

and

$$H(A|B)_{\tilde{\sigma}} - H(A|B)_{\tilde{\rho}} \leq \epsilon'' \log_2 |A| + (a+\epsilon) h\left(\frac{\epsilon''}{a+\epsilon}\right). \tag{A.3}$$

Proof. We can assume $\frac{1}{2}\|\tilde{\rho}-\tilde{\sigma}\|_1 = \epsilon$ since our bound will be increasing in ϵ. Note that $\delta \leq \epsilon$. As usual, ρ and σ denote the normalized $\tilde{\rho}$ and $\tilde{\sigma}$. Let $\cdot_+$ denote the positive part of a Hermitian operator. The proof consists of a series of operator inequalities and applications of strong subadditivity.

We first determine the trace of the positive and negative parts of $\tilde{\rho}-\tilde{\sigma}$. To do this, consider the eigenvalues λ_i of $\tilde{\rho}-\tilde{\sigma}$. By assumption, $\sum|\lambda_i| = \|\tilde{\rho}-\tilde{\sigma}\|_1 = 2\epsilon$ and $\sum\lambda_i = \mathrm{Tr}(\tilde{\rho}-\tilde{\sigma}) = 2\delta$. Thus, $\mathrm{Tr}[(\tilde{\rho}-\tilde{\sigma})_+] = \sum_{\lambda_i\geq 0}\lambda_i = \epsilon+\delta = \epsilon'$. Similarly, $\mathrm{Tr}[(\tilde{\rho}-\tilde{\sigma})_-] = -\sum_{\lambda_i<0}\lambda_i = \epsilon-\delta = \epsilon''$.

Thus, $\frac{1}{\epsilon'}(\tilde{\rho}-\tilde{\sigma})_+$ and $\frac{1}{\epsilon''}(\tilde{\rho}-\tilde{\sigma})_-$ are normalized states. Denote them by τ and τ' respectively. After some rearrangement, we can define a third state ω satisfying

$$\omega = \frac{\mathrm{Tr}\,\tilde{\sigma}}{\mathrm{Tr}\,\tilde{\sigma}+\epsilon'}\sigma + \frac{\epsilon'}{\mathrm{Tr}\,\tilde{\sigma}+\epsilon'}\tau = \frac{\mathrm{Tr}\,\tilde{\rho}}{\mathrm{Tr}\,\tilde{\rho}+\epsilon''}\rho + \frac{\epsilon''}{\mathrm{Tr}\,\tilde{\rho}+\epsilon''}\tau'. \tag{A.4}$$

Note that $\mathrm{Tr}\,\tilde{\sigma}+\epsilon' = \mathrm{Tr}\,\tilde{\rho}+\epsilon'' = a+\epsilon$. We will find an upper and lower bound on $H(A|B)_\omega$, and combine them to get our final result.

The lower bound simply follows from the concavity of conditional entropy and the definition of ω in Eq. (A.4),

$$H(A|B)_\omega \geq \frac{\mathrm{Tr}\,\tilde{\sigma}}{a+\epsilon}H(A|B)_\sigma + \frac{\epsilon'}{a+\epsilon}H(A|B)_\tau. \tag{A.5}$$

For the upper bound, we first rewrite the conditional entropy in terms of the relative entropy as follows [27],

$$-H(A|B)_{\omega_{AB}} = \min_{\xi_B} D(\omega_{AB}||\mathbb{1}_A\otimes\xi_B). \tag{A.6}$$

Note that the minimum is achieved at $\xi_B = \omega_B = \mathrm{Tr}_A(\omega_{AB})$. Expanding the definition of the relative entropy, we have

$$H(A|B)_\omega = -D(\omega_{AB}||\mathbb{1}_A\otimes\omega_B) \tag{A.7}$$

$$= H(\omega) + \mathrm{Tr}[\omega(\mathbb{1}_A\otimes\log_2\omega_B)]. \tag{A.8}$$

We upper bound the first term using strong subadditivity,

$$H(\omega) = H\left(\frac{\mathrm{Tr}\,\tilde{\rho}}{a+\epsilon}\rho + \frac{\epsilon''}{a+\epsilon}\tau'\right) \tag{A.9}$$

$$\leq \frac{\mathrm{Tr}\,\tilde{\rho}}{a+\epsilon}H(\rho) + \frac{\epsilon''}{a+\epsilon}H(\tau') + h\left(\frac{\epsilon''}{a+\epsilon}\right). \tag{A.10}$$

In the second term, we simply insert the definition of ω and expand. Thus, we have

$$H(A|B)_\omega \leq \frac{\operatorname{Tr}\tilde{\rho}}{a+\epsilon}H(\rho) + \frac{\epsilon''}{a+\epsilon}H(\tau') + \frac{\operatorname{Tr}\tilde{\rho}}{a+\epsilon}\operatorname{Tr}[\rho(\mathbb{1}_A \otimes \log_2 \omega_B)] \tag{A.11}$$

$$+ \frac{\epsilon''}{a+\epsilon}\operatorname{Tr}[\tau'(\mathbb{1}_A \otimes \log_2 \omega_B)] + h\left(\frac{\epsilon''}{a+\epsilon}\right) \tag{A.12}$$

$$= -\frac{\operatorname{Tr}\tilde{\rho}}{a+\epsilon}D(\rho||\mathbb{1}_A \otimes \omega_B) - \frac{\epsilon''}{a+\epsilon}D(\tau'||\mathbb{1}_A \otimes \omega_B) + h\left(\frac{\epsilon''}{a+\epsilon}\right), \tag{A.13}$$

where we have recombined the terms into relative entropies. We now use the relation in Eq. (A.6) again, to obtain

$$H(A|B)_\omega \leq \frac{\operatorname{Tr}\tilde{\rho}}{a+\epsilon}H(A|B)_\rho + \frac{\epsilon''}{a+\epsilon}H(A|B)_{\tau'} + h\left(\frac{\epsilon''}{a+\epsilon}\right). \tag{A.14}$$

The upper and lower bounds on $H(A|B)_\omega$, in Eq. (A.14) and Eq. (A.5) respectively, can be combined to obtain

$$\frac{\operatorname{Tr}\tilde{\sigma}}{a+\epsilon}H(A|B)_\sigma + \frac{\epsilon'}{a+\epsilon}H(A|B)_\tau \leq \frac{\operatorname{Tr}\tilde{\rho}}{a+\epsilon}H(A|B)_\rho + \frac{\epsilon''}{a+\epsilon}H(A|B)_{\tau'} + h\left(\frac{\epsilon''}{a+\epsilon}\right), \tag{A.15}$$

$$H(A|B)_{\tilde{\sigma}} - H(A|B)_{\tilde{\rho}} \leq \epsilon'' H(A|B)_{\tau'} - \epsilon' H(A|B)_\tau + (a+\epsilon)h\left(\frac{\epsilon''}{a+\epsilon}\right). \tag{A.16}$$

By repeating the proof but interchanging the two expressions for ω, we similarly obtain

$$H(A|B)_{\tilde{\rho}} - H(A|B)_{\tilde{\sigma}} \leq \epsilon' H(A|B)_\tau - \epsilon'' H(A|B)_{\tau'} + (a+\epsilon)h\left(\frac{\epsilon'}{a+\epsilon}\right). \tag{A.17}$$

Conditional entropies of normalized states are bounded between $\pm\log_2|A|$. Thus, we have

$$|H(A|B)_{\tilde{\rho}} - H(A|B)_{\tilde{\sigma}}| \leq 2\epsilon\log_2|A| + (a+\epsilon)\max\left\{h\left(\frac{\epsilon'}{a+\epsilon}\right), h\left(\frac{\epsilon''}{a+\epsilon}\right)\right\}. \tag{A.18}$$

When $\tilde{\rho}$ and $\tilde{\sigma}$ are both classical-quantum states, τ and τ' are also both classical-quantum states. Then, their conditional entropy is between 0 and $\log_2|A|$. This gives the tighter bound of

$$|H(A|B)_{\tilde{\rho}} - H(A|B)_{\tilde{\sigma}}| \leq \epsilon'\log_2|A| + (a+\epsilon)\max\left\{h\left(\frac{\epsilon'}{a+\epsilon}\right), h\left(\frac{\epsilon''}{a+\epsilon}\right)\right\}. \tag{A.19}$$

Similarly,

$$H(A|B)_{\tilde{\sigma}} - H(A|B)_{\tilde{\rho}} \leq \epsilon'' \log_2 |A| + (a+\epsilon) h\left(\frac{\epsilon''}{a+\epsilon}\right). \tag{A.20}$$

□

Corollary 1. *Let $\tilde{\rho}_{AB}$ and $\tilde{\sigma}_{AB}$ be two bipartite subnormalized classical-quantum states with $\mathrm{Tr}(\tilde{\rho}) \geq \mathrm{Tr}(\tilde{\sigma})$; the dimension of system B can be infinite. Let $\frac{1}{2}\|\tilde{\rho} - \tilde{\sigma}\|_1 \leq \epsilon \leq 1$. Then,*

$$H(A|B)_{\tilde{\sigma}} - H(A|B)_{\tilde{\rho}} \leq \epsilon \log_2 |A| + (1+\epsilon) h\left(\frac{\epsilon}{1+\epsilon}\right). \tag{A.21}$$

Proof. Begin with the third statement of Theorem 9. We can upper bound ϵ'' in the first term on the right-hand side by ϵ. Then, since the function $g(a) = (a+\epsilon) h\left(\frac{c}{a+\epsilon}\right)$ is increasing on $a \in [0,1]$, we can upper bound the second term on the right-hand side by evaluating it at $a = 1$. We have

$$\frac{\epsilon''}{1+\epsilon} \leq \frac{\epsilon}{1+\epsilon} \leq \frac{1}{2}. \tag{A.22}$$

Since the binary entropy is increasing on $[0, \frac{1}{2}]$,

$$h\left(\frac{\epsilon''}{1+\epsilon}\right) \leq h\left(\frac{\epsilon}{1+\epsilon}\right). \tag{A.23}$$

Thus we can replace ϵ'' with ϵ in the second term as well. This leaves us with the desired expression. □

Appendix B

Numerical Framework Formulation

In this Appendix, we provide additional details regarding the numerics framework and its formulation for the DMCVQKD protocol with the dimension reduction method.

B.1 Explicit SDPs

To apply the numerical framework, we need the feasible set of the convex minimization to be that of an SDP. To show that $\mathbf{S}_N$ (Eq. (4.101)) is such a set, we rewrite the trace distance constraint for the reduced density matrix. The trace norm can be expressed as an SDP [25], which allows us to rewrite the constraint using slack variables. This equivalent reformulation of the finite-dimensional optimization is given by:

$$\begin{aligned}
\underset{\tilde{\rho},R,S}{\text{minimize:}} \quad & f(\tilde{\rho}) \\
\text{subject to:} \quad & 1 - W \leq \mathrm{Tr}(\tilde{\rho}) \leq 1 \\
& \mathrm{Tr}(R) + \mathrm{Tr}(S) \leq 2\sqrt{W} \\
& \mathrm{Tr}_B(\tilde{\rho}) - R \leq \tau_A \\
& -S - \mathrm{Tr}_B(\tilde{\rho}) \leq -\tau_A \\
& \gamma_i - W\|\Gamma_i\|_\infty \leq \mathrm{Tr}(\tilde{\rho}\Gamma_i) \leq \gamma_i \\
& \tilde{\rho} \in \mathrm{Pos}\left(\mathcal{H}_N\right) \\
& R, S \in \mathrm{Pos}\left(\mathcal{H}_A\right).
\end{aligned} \tag{B.1}$$

Let ξ denote the adjoint of the partial trace map Tr_B restricted to operators on $\mathcal{H}_N$.

Following the numerical framework, the corresponding dual linearized SDP is:

$$\begin{aligned}
\underset{\vec{y},y_s,Y_1,Y_2}{\text{maximize:}} \quad & -\vec{y}\cdot(\vec{\gamma}+\vec{\epsilon}_{rep}) - y_s(2\sqrt{W}+\epsilon'_{rep}) \\
& - \mathrm{Tr}(\tau_A Y_1) + \mathrm{Tr}(\tau_A Y_2) \\
\text{subject to:} \quad & \sum_{i=1}^{2m} y_i\Gamma_i + \xi(Y_1) - \xi(Y_2) \geq -\nabla f(\rho_{opt}) \\
& y_s \mathbb{1}_A \geq Y_1 \\
& y_s \mathbb{1}_A \geq Y_2 \\
& \vec{y} \in \mathbb{R}^{2m}_{\geq 0} \\
& y_s \in \mathbb{R}_{\geq 0} \\
& Y_1, Y_2 \in \mathrm{Pos}\left(\mathcal{H}_A\right),
\end{aligned} \tag{B.2}$$

where $\vec{\gamma} = (\{\gamma_i\}_{i=1}^m, \{W\|\Gamma_i\| - \gamma_i\}_{i=1}^m)$, $\vec{\Gamma} = (\{\Gamma_i\}_{i=1}^m, \{-\Gamma_i\}_{i=1}^m)$, and $\vec{\epsilon}_{rep}$ and ϵ'_{rep} are the expansion parameters which account for finite numerical precision (see Appendix B.2 for further discussion).

As noted in Sec. 4.2.4, the finite-dimensional optimization has a slightly different form when the projection only acts on Bob's system, that is $\Pi = \mathbb{1}_A \otimes \Pi_B$. In this case, the dual SDP is the same as above, but the term in the objective function and both constraints involving y_s are removed, Y_2 is set to zero, and the term $-\epsilon'_{rep}\,\mathrm{Tr}(Y_1)$ is added to the objective function.

B.2 Modification to Dual Expansion

As has been summarized in Sec. 3.3, the numerical framework presented in Ref. [4] is tight in principle. We observe the following issue in practice. The near-optimal ρ_{guess} computed in the first step often has constraint violations, due to the inherent imprecision of convex solvers. For example, for the quadrature phase-shift keying protocol, at all distances and values of excess noise, and for our particular implementation using MATLAB and CVX with the MOSEK solver, these violations are typically $10^{-7} - 10^{-6}$. At distances approaching 200 km, the simulated expectation values $\langle \hat{n}_{\beta_i} \rangle$ and $\langle \hat{n}^2_{\beta_i} \rangle$ are both small; approximately 10^{-6} for the nonzero values of excess noise we consider. Since the constraint violation is the same order of magnitude as the expectation, the first step solution is effectively an optimal state for double the excess noise. Given the poor scaling of the protocol with excess noise, this implies the approximate key rate from this first step will be much lower

than its theoretical value. Since the first step upper bounds the second step, this implies a poor second step result. One way to ameliorate this is to solve with a smaller N so that the solver returns a better first step solution. Thus, purely due to numerical precision issues, solving with $N = 30$ instead of $N = 40$ can improve the key rate at long distances. Even though the correction term $\Delta(W)$ is slightly larger, this is more than offset by the improved quality of the first step solution. This suggests that due to numerical issues, one should choose the finite dimension carefully, even though analytically a larger dimension is always better.

The reason the first step upper bounds the second step is due to the expansion of the feasible set. Referring to the notation in Appendix D of Ref. [4], the large constraint violations lead to a large value of ϵ', which controls how much the set is expanded for the second step. However, the choice $\epsilon' = \max(\epsilon_{rep}, \epsilon_{sol})$ (Equation (165) of Ref. [4]) is pessimistic. One only needs to choose $\epsilon' = \epsilon_{rep}$. As noted in Equation (162) of Ref. [4], this is sufficient to provide a reliable lower bound when accounting for finite numerical precision. Further, note that $f(\rho)$ is lower bounded by a tangent hyperplane at any point in its domain. Thus, it is not necessary to expand the feasible set further to include the point returned by the first step. This change gives improved results in practice, while still being reliable and tight.

In previous work using the numerical framework of Ref. [4], it has been assumed that $\epsilon_{rep} \leq \epsilon_{sol}$. Hence the issue of how to suitably choose ϵ_{rep} has not been considered. As noted in Ref. [4], rigorously determining ϵ_{rep} for a particular implementation can be an involved process. For our MATLAB implementation, which has precision better than 10^{-15}, we conservatively use 10^{-10} for ϵ'_{rep} and all elements of $\vec{\epsilon}_{rep}$. Finally note that in our numerical evaluation of the unbalanced phase-encoded BB84 protocol, we continue to use the original, larger set expansion as in Ref. [4]. This is to ensure a fair comparison to the flag-state squasher numerical results.

As is remarked in Ref. [4], the numerical framework does not account for errors in function evaluation, as these are implementation-dependent; but in principle, the effect of these errors can be accounted for for a given implementation. One expects that this would not have a noticeable effect on the key rates. For the specific case of DMCVQKD, one area of caution is that we use numerical integration to generate the matrix elements for the POVM which defines the key map, which necessarily introduces some small errors due to the limits of floating-point arithmetic. While comparison to the analytic key rates in simple cases makes it clear our results are accurate and reliable, it would be nice to have a self-contained bound on the objective function evaluation error, though it is likely negligible. This could then be accounted for in the dual by adding $\epsilon_{func}\mathbb{1}$ to $\nabla f(\rho_{opt})$ (see Eq. (B.2)).

B.3 DMCVQKD Numerical Details

The objective function can be equivalently formulated as a conditional entropy (Eq. (3.9)) and as a quantum relative entropy (Eq. (3.17)). The former is used in this thesis for analytic results, while the latter is used in the numerical framework. We thus give the quantum relative entropy formulation here.

The Kraus operator for the $\mathcal{G}$ map in the quantum relative entropy formulation is

$$K = \sum_{z=0}^{M-1} \mathbb{1}_A \otimes \sqrt{R_B^z} \otimes |z\rangle_Z \,. \tag{B.3}$$

Note that in keeping with the trace non-increasing definition of $\mathcal{G}$, the Kraus operator does not include the region operator for discarded signals. We have omitted trivial or redundant registers [23].

In Ref. [4], the initial point for the Frank-Wolfe iteration is itself found using a SDP. For DMCVQKD, we simply use the simulation state for the loss-only scenario as the initialization point. Note that this point always satisfies the constraints, as nonzero excess noise only increases the conditional photon number expectations.

Appendix C

Matrix Operations in Displaced Basis

Recall our basis is $\{|i\rangle_A \otimes |n_{\beta_i}\rangle_B\}$. We calculate the matrix elements of certain operators in this basis and evaluate the action of relevant channels.

C.1 Constraints

Our constraint operators take a particularly simple form in the displaced basis. The matrix elements are

$$\langle i| \langle m_{\beta_i}| (|k\rangle\langle k| \otimes \hat{n}_{\beta_k}) |j\rangle \left|n_{\beta_j}\right\rangle = \delta_{ik}\delta_{jk} \langle m_{\beta_k}| \hat{n}_{\beta_k} |n_{\beta_k}\rangle \tag{C.1}$$

$$= \delta_{ik}\delta_{jk}\delta_{mn} n. \tag{C.2}$$

Similarly, for $\hat{n}^2_{\beta_k}$, they are $\delta_{ik}\delta_{jk}\delta_{mn}n^2$.

C.2 Ideal and Trusted Noise Region Operators

Recall the ideal detector POVM elements (Eq. (5.57)) are

$$P^k = \mathbb{1}_A \otimes R^k_B, \tag{C.3}$$

where R_B^k are the region operators. The matrix elements are

$$P^k_{ijmn} = \langle i| \langle m_{\beta_i}| \left(\mathbb{1}_A \otimes R_B^k\right) |j\rangle \left|n_{\beta_j}\right\rangle \tag{C.4}$$

$$= \delta_{ij} \langle m_{\beta_i}| R_B^k \left|n_{\beta_j}\right\rangle \tag{C.5}$$

$$= \langle m_{\beta_i}| R_B^k |n_{\beta_i}\rangle \tag{C.6}$$

$$= \frac{1}{\pi} \int_{A^k} r e^{-|\kappa|^2} \frac{\kappa^m \kappa^{*n}}{\sqrt{m!n!}}\, d\theta\, dr, \tag{C.7}$$

where $\kappa = re^{i\theta} - \beta_i$.

Analogously, recall the trusted noise POVM elements (Eq. (5.58)) are

$$P^k = \mathbb{1}_A \otimes \left[R_B^k\right]', \tag{C.8}$$

The matrix elements are similarly

$$\left[P^k\right]'_{ijmn} = \langle m_{\beta_i}| \left[R_B^k\right]' |n_{\beta_i}\rangle\,, \tag{C.9}$$

$$= \int_{A^k} \langle m_{\beta_i}| G_\zeta |n_{\beta_i}\rangle\, d^2\zeta, \tag{C.10}$$

$$= \frac{1}{\eta_d \pi} \int_{A^k} \langle m| D\left(\frac{\zeta}{\sqrt{\eta_d}} - \beta_i\right) \rho_{th}(\bar{n}) D^\dagger \left(\frac{\zeta}{\sqrt{\eta_d}} - \beta_i\right) |n\rangle\, d^2\zeta, \tag{C.11}$$

$$= \int_{A^k} \langle m| G_{\zeta - \sqrt{\eta_d}\beta_i} |n\rangle\, d^2\zeta. \tag{C.12}$$

We use the expression in Equation (B1) in Ref. [72] for the Fock basis matrix elements of the noisy fine-grained POVM G_ζ. Again with $\kappa = re^{i\theta} - \sqrt{\eta_d}\beta_i$ and $\kappa' = \kappa/\sqrt{\eta_d} = re^{i\theta}/\sqrt{\eta_d} - \beta_i$ we have

$$\langle m| G_\kappa |n\rangle = \frac{1}{\eta_d \pi} \exp\left(\frac{-|\kappa'|^2}{(1+\bar{n})}\right) \frac{\bar{n}^m}{(1+\bar{n})^{n+1}} \left(\overline{\kappa'}\right)^{n-m} \sqrt{\frac{m!}{n!}} L_m^{(n-m)}\left(\frac{-|\kappa'|^2}{\bar{n}(1+\bar{n})}\right). \tag{C.13}$$

We have now specified the integrand in polar coordinates for both the ideal and trusted noise cases. The next step is to specify the regions over which we integrate. As we will see shortly, we only need to directly compute the integrals for the first region operator.

For amplitude and phase postselection, the integration limits for R_0 are

$$\int_{\frac{-\pi}{M}+\Delta_p}^{\frac{\pi}{M}-\Delta_p} \int_{\Delta_a}^{\infty} [...]\, dr\, d\theta. \tag{C.14}$$

For hyperbolic postselection, the integration limits for R_0 are

$$\int_{\frac{-\pi}{M}}^{\frac{\pi}{M}} \int_{\sqrt{\frac{\Delta_h}{\cos^2\theta - \Delta_s \sin^2\theta}}}^{\infty} [...]\ dr\ d\theta. \tag{C.15}$$

These integrals are computed in MATLAB.

Harnessing Rotational Symmetry

Computing these integrals is numerically intensive, so we make use of a symmetry property to simplify the calculations, at least for the phase-shift keying case. We give the argument in the trusted noise case only, and the ideal detector scenario is recovered as a special case. All indexing in this section is modulo M. $U_k = e^{\frac{2\pi i k}{M}\hat{n}}$ is the unitary effecting a counter-clockwise rotation of $\frac{2\pi k}{M}$ around the origin in phase space.

For phase-shift keying we have that $\left[R_B^k\right]' = U_k \left[R_B^0\right]' {U_k}^\dagger$. Moreover, for the particular channel model we consider, $|\beta_k\rangle = U_k |\beta_0\rangle$. Given this rotational symmetry, we can simplify the calculation of matrix elements as follows.

We first note that the matrix elements of the region operators in the displaced basis are the same as the Fock basis elements of the displaced region operators. That is,

$$\langle m_{\beta_i}| \left[R_B^k\right]' |n_{\beta_i}\rangle = \langle m| \hat{D}^\dagger(\beta_i) \left[R_B^k\right]' \hat{D}(\beta_i) |n\rangle . \tag{C.16}$$

By symmetry, and the commutation relation between the rotation and displacement unitaries, we have

$$\begin{aligned} U_k \hat{D}^\dagger(\beta_i) \left[R_B^0\right]' \hat{D}(\beta_i) {U_k}^\dagger &= \hat{D}^\dagger(\beta_{i+k}) U_k \left[R_B^0\right]' {U_k}^\dagger \hat{D}(\beta_{i+k}) && \text{(C.17)} \\ &= \hat{D}^\dagger(\beta_{i+k}) \left[R_B^k\right]' \hat{D}(\beta_{i+k}) && \text{(C.18)} \end{aligned}$$

Each U_k is diagonal in the Fock basis, with $\langle n| U_k |n\rangle = \exp\left(n\frac{2\pi i k}{M}\right)$, so commutes with the projection $\Pi_{Fock} = \sum_{n=0}^{N} |n\rangle\langle n|$. Thus, it suffices to determine the M matrices $\langle m_{\beta_i}| \left[R_B^0\right]' |n_{\beta_i}\rangle$, and then use matrix multiplication (in the Fock basis) with U_k to generate the remaining $\langle m_{\beta_i}| \left[R_B^k\right]' |n_{\beta_i}\rangle$. This is more efficient than directly computing the integrals for all M^2 matrices.

C.3 Channels

Our basis for the bipartite Hilbert space is not of the form $|i\rangle_A \otimes |j\rangle_B$, where $|i\rangle_A$ and $|j\rangle_B$ are bases for $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively. Matrix multiplication proceeds as normal, since we simply have some orthonormal basis. However, operations that care about subsystems, namely the partial trace and its adjoint, have a different matrix representation than the typical presentation. We have

$$\rho_{AB} = \sum_{i,j,m,n} c_{ijmn} \, |i\rangle\langle j| \otimes \left|m_{\beta_i}\right\rangle\left\langle n_{\beta_j}\right| , \tag{C.19}$$

where the coefficients c are the matrix elements of ρ. We denote this matrix by M_ρ, and its matrix elements are

$$M_\rho = \sum_{i,j,m,n} c_{ijmn} \, |i\rangle\langle j| \otimes |m\rangle\langle n| . \tag{C.20}$$

The reduced density matrix is

$$\rho_A = \sum_{i,j,m,n} c_{ijmn} \, |i\rangle\langle j| \left\langle n_{\beta_j}\middle|m_{\beta_i}\right\rangle . \tag{C.21}$$

Defining

$$G = \sum_{i,j,m,n} \left\langle n_{\beta_j}\middle|m_{\beta_i}\right\rangle |i\rangle\langle j| \otimes \left|m_{\beta_i}\right\rangle\left\langle n_{\beta_j}\right| , \tag{C.22}$$

we have that

$$\langle i|\rho_A|j\rangle = \rho_{ij} \odot G_{ij}, \tag{C.23}$$

where the subscripts on the bipartite operators indicate the respective block matrix, and $\odot$ is the element-wise dot product. Note that each G_{ij} can be thought of as a basis change unitary in $\mathcal{H}_B$. An explicit formula for the elements of G is

$$\begin{aligned}
\left\langle n_{\beta_j}\middle|m_{\beta_i}\right\rangle &= \langle n| \, D^\dagger(\beta_j) D(\beta_i) \, |m\rangle && \text{(C.24)} \\
&= \exp(i \operatorname{Im}(-\beta_j \beta_i^*)) \, \langle n| \, D(\beta_i - \beta_j) \, |m\rangle && \text{(C.25)} \\
&= \exp\left(i \operatorname{Im}(-\beta_j \beta_i^*) - \frac{|\beta_i - \beta_j|^2}{2} \right) \sqrt{m! n!} \\
&\quad \sum_{k=0}^{\min(m,n)} \frac{1}{k!(m-k)!(n-k)!} (\beta_i - \beta_j)^{n-k} (\beta_j^* - \beta_i^*)^{m-k}. && \text{(C.26)}
\end{aligned}$$

We compute and store this matrix once at the beginning of the optimization algorithm, and use it each time to calculate the partial trace.

The adjoint of the partial trace also has a matrix representation involving G. The adjoint of the partial trace is $\xi(\sigma_A) = \sigma_A \otimes \mathbb{1}_B$. Letting

$$\sigma_A = \sum_{ij} c_{ij} \left|i\right\rangle\!\left\langle j\right|, \tag{C.27}$$

we seek d_{ijmn} such that

$$\sigma_A \otimes \mathbb{1}_B = \sum_{ijmn} d_{ijmn} \left|i\right\rangle\!\left\langle j\right| \otimes \left|m_{\beta_i}\right\rangle\!\left\langle n_{\beta_j}\right|. \tag{C.28}$$

This implies

$$\sum_{mn} d_{ijmn} \left|m_{\beta_i}\right\rangle\!\left\langle n_{\beta_j}\right| = c_{ij}\mathbb{1}_B \quad \forall i, j. \tag{C.29}$$

Taking the bra-ket on both sides, we obtain $d_{ijmn} = c_{ij} \left\langle m_{\beta_i} \middle| n_{\beta_j} \right\rangle$. We recognize the factor on the right-hand side as G^*. Thus, we have that

$$\xi(\sigma_A) = \sum_{ijmn} c_{ij} \left|i\right\rangle\!\left\langle j\right| \otimes G^*_{ij}. \tag{C.30}$$

Appendix D

Simulated Expectations and Error-Correction Cost

We discuss how the coarse-grained expectations can be determined from a heterodyne measurement, and what the expectation values are for the simulation. We focus on the trusted detector noise scenario, as the ideal detector results can be recovered as a special case.

Bob's measurement results determine a probability density $p(\zeta) = \mathrm{Tr}(\rho G_\zeta)$ over the complex plane. In general, given an observable $\Gamma = \int_{\zeta\in\mathbb{C}} w_\Gamma(\zeta)\ G_\zeta\ d^2\zeta$, the expectation value is then

$$\mathrm{Tr}(\rho\Gamma) = \int_{\zeta\in\mathbb{C}} w_\Gamma(\zeta)\ p(\zeta)\ d^2\zeta. \tag{D.1}$$

For the i^{th} conditional state, Bob's coarse-grained observables are $\hat{n}_{\beta_i}$ and $\hat{n}^2_{\beta_i}$. The corresponding functions $w_{\hat{n}_{\beta_i}}$ and $w_{\hat{n}^2_{\beta_i}}$ are given in Eqs. (5.30) and (5.31). Using his measurement result $p(\zeta)$, Bob can thus compute the integral (D.1) to determine the desired expectations for each conditional state (see Appendix E for more details).

Note that for the typical quadratures $\hat{X}$ and $\hat{P}$, we have that

$$\hat{X} = \frac{1}{\sqrt{2}}(\hat{a}^\dagger + \hat{a}) \implies w_{\hat{X}}(\zeta) = \sqrt{2}\,\mathrm{Re}(\zeta), \tag{D.2}$$

$$\hat{P} = \frac{i}{\sqrt{2}}(\hat{a}^\dagger - \hat{a}) \implies w_{\hat{P}}(\zeta) = \sqrt{2}\,\mathrm{Im}(\zeta). \tag{D.3}$$

Thus, by expanding $w_{\hat{n}_{\beta_i}}(\zeta)$ and $w_{\hat{n}^2_{\beta_i}}(\zeta)$ as polynomials in $\mathrm{Re}(\zeta)$ and $\mathrm{Im}(\zeta)$, we can also relate the expectations of $\hat{n}_{\beta_i}$ and $\hat{n}^2_{\beta_i}$ to the moments and cross-terms of the measurement

data of quadratures $\hat{X}$ and $\hat{P}$. Again, for a detailed description of this calculation, see Appendix E.

We now consider the expectations under the simulated channel model. After passing through a Gaussian channel with loss η and excess noise ξ, a coherent signal state becomes a displaced thermal state

$$|\alpha_i\rangle\langle\alpha_i| \to \hat{D}\left(\beta_i\right)\rho_{th}(\tfrac{\delta}{2})\hat{D}^{\dagger}\left(\beta_i\right) \tag{D.4}$$

where $\delta = \eta\xi$ and $\beta_i = \sqrt{\eta}\alpha_i$. The expectation values for each conditional state are straightforward to calculate,

$$\mathrm{Tr}\Big(\hat{n}_{\beta_i}\hat{D}\left(\beta_i\right)\rho_{th}(\tfrac{\delta}{2})\hat{D}^{\dagger}\left(\beta_i\right)\Big) = \mathrm{Tr}\big(\hat{n}\ \rho_{th}(\tfrac{\delta}{2})\big) = \frac{\delta}{2} \tag{D.5}$$

and

$$\mathrm{Tr}\Big(\hat{n}^2_{\beta_i}\hat{D}\left(\beta_i\right)\rho_{th}(\tfrac{\delta}{2})\hat{D}^{\dagger}\left(\beta_i\right)\Big) = \frac{\delta(1+\delta)}{2}. \tag{D.6}$$

For the reduced state constraint, we simply use the formula for the overlap of two coherent states (Eq. (2.48)), reproduced below for convenience,

$$\langle\alpha_j|\alpha_i\rangle = \exp\left(i\,\mathrm{Im}(\alpha_i\alpha_j^*) - \frac{1}{2}|\alpha_i - \alpha_j|^2\right). \tag{D.7}$$

The error-correction cost is determined by the simulated joint probability distribution. Given Alice prepares $|\alpha_i\rangle$, the probability Bob gets the key map outcome j, with $i \in \{0, 1, ..., M-1\}$ and $j \in \{0, 1, ..., M-1, \perp\}$, is given by the following integral

$$p(j|i) = \mathrm{Tr}\Big(R_j\hat{D}\left(\beta_i\right)\rho_{th}(\tfrac{\delta}{2})\hat{D}^{\dagger}\left(\beta_i\right)\Big) \tag{D.8}$$

$$= \int_{A_j}\mathrm{Tr}\Big(G_{\zeta}\hat{D}\left(\beta_i\right)\rho_{th}(\tfrac{\delta}{2})\hat{D}^{\dagger}\left(\beta_i\right)\Big)d^2\zeta. \tag{D.9}$$

The integrand, which is the overlap of two displaced thermal states, is given by [72]

$$\mathrm{Tr}\Big(G_{\zeta}\hat{D}\left(\beta_i\right)\rho_{th}(\tfrac{\delta}{2})\hat{D}^{\dagger}\left(\beta_i\right)\Big) = \frac{1}{\pi(1+\frac{1}{2}\eta_d\delta+\nu_{el})}\exp\left(\frac{-\left|\zeta-\sqrt{\eta_d}\beta_i\right|^2}{1+\frac{1}{2}\eta_d\delta+\nu_{el}}\right). \tag{D.10}$$

The integral in Eq. (D.9) is converted to polar coordinates and computed in MATLAB. As the signal states are distributed uniformly, $p_A(i) = \frac{1}{M}$. Then, $p_{AB}(i,j) = \frac{1}{M}p(j|i)$. Since discarded signals do not incur an error-correction cost, we remove the outcome $j = \perp$ and renormalize p accordingly. Denoting this sifted probability distribution by q, the error-correction cost is

$$\delta_{EC}^{leak} = (1-p(\perp))(\log_2(M) - \beta_{EC}[H(q_A) + H(q_B) - H(q_{AB})]). \tag{D.11}$$

Appendix E

Experimental Implementation

In this section, we outline how raw heterodyne measurement data from an experimental implementation is processed to be used in our numerical framework for DMCVQKD key rate calculations. Again we give the results in the trusted noise scenario, and the ideal detector scenario is recovered as a special case.

From the experiment, the known parameters are the signal state amplitudes α_i, the a priori probabilities $p(i)$, the detector efficiency η_d, the detector electronic noise ν_{el} (in shot noise units), and Bob's heterodyne measurement results from the testing rounds.

In order to apply the numerical framework, the main steps are to specify the parameters for the finite subspace and determine the constraints for the optimization. The formula for W and the correction term $\Delta(W)$ can be used directly.

Each heterodyne measurement, which is comprised of two homodyne detections, returns two real numbers, x and p. These correspond to the quadratures $\hat{X}$ and $\hat{P}$, and we assume that the values are reported in shot-noise units (see Sec. 2.4.1). If they are instead reported in natural units, one simply divides both numbers by $\sqrt{2}$ to convert to shot noise units.

E.1 Choosing Finite Subspace

The subspace dimension N can be chosen freely by the experimentalist. Larger N reduce the correction term, but also increase the numerical run time. A typical choice would be $N = 10$.

To specify the subspace, the only other step is to choose the displacements β_k (see Sec. 5.5.1). To do this, we use the information from the testing rounds. Once Alice announces

which state was sent on each testing round, Bob can group together the measurement results by the corresponding conditional state. For the k^{th} conditional state, let $\{(x_j^k, p_j^k)\}_{j=1}^{C_k}$ be the list of all corresponding measurement results.

Recall the weight functions $w_{\hat{n}}(\zeta) = |\zeta|^2 - 1$ and $w_{\hat{n}^2}(\zeta) = |\zeta|^4 - 3|\zeta|^2 + 1$ (Eqs. (5.24), (5.25)), as well as $w_{\hat{X}}(\zeta) = \sqrt{2}\,\mathrm{Re}(\zeta)$ and $w_{\hat{P}}(\zeta) = \sqrt{2}\,\mathrm{Im}(\zeta)$ (Eqs. (D.2), (D.3)).

We simply set each β_k to be the average complex displacement, with a normalization due to the detector efficiency and to the factor of $\sqrt{2}$ in the weight function,

$$\beta_k \equiv \frac{1}{\sqrt{2\eta_d}} \frac{1}{C_k} \sum_{j=1}^{C_k} x_j^k + i p_j^k. \tag{E.1}$$

E.2 Determining Expectations

Next, we need to calculate the expectations $\left\langle \left[\hat{n}_{\sqrt{\eta_d}\beta_k}\right]' \right\rangle$ and $\left\langle \left[\hat{n}^2_{\sqrt{\eta_d}\beta_k}\right]' \right\rangle$. To do this, it is simplest to displace the raw data first. The displaced values are

$$\tilde{x}_j^k = x_j^k - \sqrt{2\eta_d}\,\mathrm{Re}(\beta_k), \tag{E.2}$$

$$\tilde{p}_j^k = p_j^k - \sqrt{2\eta_d}\,\mathrm{Im}(\beta_k). \tag{E.3}$$

Calculating the expectations of the displaced observables on the original undisplaced data is the same as calculating the expectations of the undisplaced observables on the displaced data. We have that

$$w_{\hat{n}}(\zeta) = |\zeta|^2 - 1 \tag{E.4}$$

$$= \mathrm{Re}(\zeta)^2 + \mathrm{Im}(\zeta)^2 - 1 \tag{E.5}$$

$$= \frac{\left[\sqrt{2}\,\mathrm{Re}(\zeta)\right]^2 + \left[\sqrt{2}\,\mathrm{Im}(\zeta)\right]^2}{2} - 1 \tag{E.6}$$

$$= \frac{w_{\hat{X}}^2 + w_{\hat{P}}^2}{2} - 1. \tag{E.7}$$

Similarly,

$$w_{\hat{n}^2}(\zeta) = |\zeta|^4 - 3|\zeta|^2 + 1 \tag{E.8}$$

$$= \left[\mathrm{Re}(\zeta)^2 + \mathrm{Im}(\zeta)^2\right]^2 - 3\left[\mathrm{Re}(\zeta)^2 + \mathrm{Im}(\zeta)^2\right] + 1 \tag{E.9}$$

$$= \mathrm{Re}(\zeta)^4 + 2\,\mathrm{Re}(\zeta)^2\,\mathrm{Im}(\zeta)^2 + \mathrm{Im}(\zeta)^4 - 3\left[\mathrm{Re}(\zeta)^2 + \mathrm{Im}(\zeta)^2\right] + 1 \tag{E.10}$$

$$= \frac{\left[\sqrt{2}\,\mathrm{Re}(\zeta)\right]^4 + 2\left[\sqrt{2}\,\mathrm{Re}(\zeta)\right]^2\left[\sqrt{2}\,\mathrm{Im}(\zeta)\right]^2 + \left[\sqrt{2}\,\mathrm{Im}(\zeta)\right]^4}{4} - \frac{3\left(\left[\sqrt{2}\,\mathrm{Re}(\zeta)\right]^2 + \left[\sqrt{2}\,\mathrm{Im}(\zeta)\right]^2\right)}{2} + 1 \tag{E.11}$$

$$= \frac{w_{\hat{X}}^4 + 2w_{\hat{X}}^2 w_{\hat{P}}^2 + w_{\hat{P}}^4}{4} - \frac{3\left(w_{\hat{X}}^2 + w_{\hat{P}}^2\right)}{2} + 1. \tag{E.12}$$

Now, referring to Eq. (D.1), we can simply replace the integral over the probability distribution $p(\zeta)$, with a summation over the discrete probability distribution obtained from measurement. Directly from Eqs. (E.7), (E.12), we then have

$$\left\langle \left[\hat{n}_{\sqrt{\eta_d}\beta_k}\right]' \right\rangle = \frac{1}{C_k}\sum_{j=1}^{C_k}\left[\frac{1}{2}(\tilde{x}_j^k)^2 + \frac{1}{2}(\tilde{p}_j^k)^2 - 1\right], \tag{E.13}$$

and

$$\left\langle \left[\hat{n}^2_{\sqrt{\eta_d}\beta_k}\right]' \right\rangle = \frac{1}{C_k}\sum_{j=1}^{C_k}\left[\frac{1}{4}(\tilde{x}_j^k)^4 + \frac{1}{2}(\tilde{x}_j^k)^2(\tilde{p}_j^k)^2 + \frac{1}{4}(\tilde{p}_j^k)^4 - \frac{3}{2}(\tilde{x}_j^k)^2 - \frac{3}{2}(\tilde{p}_j^k)^2 + 1\right] \tag{E.14}$$

From this, one reconstructs the effective ideal expectations, per Eqs. (5.55), (5.56). This is also an appropriate place to remark that the challenge in generalizing our security proof to the DMCVQKD protocol with homodyne detection is in estimating the cross terms between the two quadratures (the second term in Eq. (E.14)).

E.3 Overall Formulation

Recall the form of the finite-dimensional optimization in Eq. (5.75). For each key generation round, Bob maps his measurement result (x, p) to the key symbol k when the complex number $x + ip$ lies in the region A_k of the complex plane. The choices of A_k define the

region operators, which in turn define the objective function f^{noisy} for the optimization (see Eq. (5.60)). The expectation constraints have been derived above, and the reduced state constraint τ_A is known directly from the experimental values of α_i and $p(i)$. This completes the processing of experimental data into the numerics framework.