Information-Theoretic Tools for Analyzing Non-IID Structures in Quantum Key Distribution

by

Amir Arqand

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

Waterloo, Ontario, Canada, 2025

© Amir Arqand 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Mark M. Wilde Associate Professor, School of Electrical and Computer Engineering Cornell University
Supervisor:	Norbert Lütkenhaus Professor, Department of Physics and Astronomy University of Waterloo
Internal Member:	Alan Jamison Assistant Professor, Department of Physics and Astronomy University of Waterloo
Internal-External Member:	Graeme S. B. Smith Associate Professor, Department of Applied Mathematics University of Waterloo
Committee Member:	Michele Mosca Professor, Department of Combinatorics and Optimization University of Waterloo

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

This thesis is based on the following articles, of which I am the first author.

> [7] Arqand, A., Hahn, T. A., & Tan, E. Y.-Z. (2024). Generalized Rényi entropy accumulation theorem and generalized quantum probability estimation.
>
> I did not contribute to the calculations in Sec. 4.6.2; those were done by my coauthors.
>
> [8] Arqand, A., Metger, T., & Tan, E. Y.-Z. (2024). Mutual information chain rules for security proofs robust against device imperfections.
>
> I did not contribute to the calculations in Theorem 6.2.2 and Corollary 6.2.2; those were done by my coauthors.
>
> [9] Arqand, A., & Tan, E. Y.-Z. (2025). Marginal-constrained entropy accumulation theorem.

The thesis is also based on the following article, of which I am a co-first author.

> [65] Kamin, L., Arqand, A., George, I., Lütkenhaus, N., & Tan, E. Y.-Z. (2024). Finite-size analysis of prepare-and-measure and decoy-state QKD via entropy accumulation.
>
> I contributed equally with Lars Kamin. My contribution was primarily in the theoretical foundations of the work, as well as in parts of the coding for numerical analysis. In particular, the contributions to the completeness parameter and most of the contributions to the decoy-state analysis were made by my coauthors.

Abstract

Quantum key distribution (QKD) is the task of establishing an information-theoretically secure key between two parties (Alice and Bob) by exploiting the rules of quantum mechanics. A central goal in QKD is to provide finite-size security proofs against the most general class of attacks—namely, coherent attacks. One approach for obtaining such proofs is through the original entropy accumulation theorem (EAT) and the generalized entropy accumulation theorem (GEAT). In this thesis, we improve and extend these EAT-style techniques in several aspects. We begin by presenting techniques for applying the GEAT in the finite-size analysis of generic prepare-and-measure protocols. We then improve the statistical analysis in these EAT-style techniques, yielding a significant improvement in the key rates obtained by these methods. In addition, these improvements can be applied directly at the level of Rényi entropies if desired, yielding fully-Rényi security proofs. Beyond this, we develop an EAT-style framework that can accommodate specific forms of marginal-state constraints "compatible" with the source-replacement scheme. This allows us to overcome the repetition-rate restriction previously imposed on the GEAT when analyzing prepare-and-measure protocols. Furthermore, it yields "fully adaptive" protocols that can, in principle, update the entropy estimation procedure during the protocol itself. Finally, we investigate the effect of imperfections in QKD by deriving a number of chain rules for mutual information suited for analyzing protocols involving information leakage from imperfect devices. In particular, we derive chain rules between smooth min-entropy and smooth max-information for characterizing one-shot information leakage caused by an additional conditioning register. Furthermore, we introduce an EAT-style theorem for mutual information to simplify the evaluation of leakage measured by smooth max-information. In addition, we derive suitable chain rules to incorporate the leakage process into the GEAT framework.

Acknowledgements

First of all, I would like to thank my supervisor, Professor Norbert Lütkenhaus, who gave me the opportunity to pursue my studies at the Institute for Quantum Computing.

I would like to especially thank Prof. Ernest Y.-Z. Tan, without whom I would not have been able to carry out the research that led to this thesis.

I am grateful to the members of the OQCT research group, especially Lars Kamin, Devashish Tupkary, John Burniston, and Shlok Nahar, for fruitful discussions over the course of my studies. My sincere thanks go to my co-authors: Prof. Ernest Y.-Z. Tan, Prof. Norbert Lütkenhaus, Thomas Hahn, Dr. Ian George, Lars Kamin, and Tony Metger. I extend my sincere appreciation to Prof. Renato Renner, Prof. Peter Brown, Dr. Thomas van Himbeeck, Dr. Ashutosh Marwah, and Dr. Srijita Kundu. I would also like to thank my past and present advisory committee: Prof. Jon Yard, Prof. Graeme Smith, Prof. Michele Mosca, Prof. Thomas Jennewein, Prof. Alan Jamison, and Prof. Norbert Lütkenhaus.

I would also like to thank all my examiners: Prof. Graeme Smith, Prof Michele Mosca, Prof. Alan Jamison, Prof. Norbert Lütkenhaus, and Prof. Mark Wilde. I would like to express my sincere gratitude to Prof. Mark Wilde for providing detailed feedback on this thesis, especially for a discussion which led to Remark 6.3.1.

Finally, I would like to thank my parents, my sister, and especially my wife, Saba, for her love, encouragement, and support.

Table of Contents

Examining Committee ii

Author's Declaration iii

Statement of Contributions iv

Abstract v

Acknowledgements vi

List of Figures xi

List of Tables xii

1 Introduction 1

1.1 Outline of the thesis 3

2 Background 7

2.1 Notations and definitions 7

2.2 Entropic quantities 11

2.3 Basic convex optimization concepts 19

2.3.1 Conjugate and duality 20

2.3.2 Semi-definite program 20

2.4 Introduction to Quantum Key Distribution (QKD) . . . 22
2.4.1 Generic device-dependent QKD protocols . . . 22
2.4.2 Source-replacement technique . . . 26
2.4.3 Security definition . . . 27
2.5 Security against coherent attack . . . 28
2.5.1 Privacy amplification and Rényi Leftover-Hashing Lemma . . . 28
2.5.2 Original entropy accumulation theorem . . . 29
2.5.3 Generalized entropy accumulation theorem . . . 33

3 Finite-size analysis of prepare-and-measure and decoy-state QKD via generalized entropy accumulation 37

3.1 Introduction . . . 37
3.2 Protocol description . . . 39
3.3 Improved entropy accumulation analysis for PM protocols . . . 41
3.3.1 Security . . . 41
3.3.2 Completeness . . . 44
3.3.3 The rate function . . . 45
3.4 Single-round analysis . . . 49
3.4.1 Secure lower bound . . . 49
3.4.2 Optimizing parameter choices . . . 51
3.5 Qubit BB84 with loss . . . 59
3.5.1 Numerical results . . . 61
3.6 Decoy-state analysis . . . 61
3.6.1 Numerical results . . . 64
3.7 Conclusion . . . 65

4 Improved Statistical analysis via generalized Rényi entropy accumulation 68
4.1 Introduction . . . 68
4.2 Preliminary definitions . . . 69
4.3 QEF-type bounds from the GEAT . . . 71
4.3.1 QES-entropies . . . 71
4.3.2 QES-entropy accumulation or chaining . . . 80
4.3.3 Applications . . . 89
4.4 Simplified versions . . . 92
4.4.1 First-order analysis . . . 95
4.4.2 Infrequent-sampling channels . . . 98
4.4.3 Convexity of relaxed lower bounds . . . 100
4.4.4 Proof of Theorem 4.4.1 . . . 101
4.5 Original EAT with improved statistical analysis . . . 110
4.6 Tightness of bounds and numerical example . . . 112
4.6.1 Remaining steps for improvement . . . 113
4.6.2 Entanglement-based BB84 . . . 113
4.7 Conclusion and open questions . . . 120

5 Marginal-constrained entropy accumulation 121
5.1 Introduction . . . 121
5.2 Rényi channel conditional entropy . . . 124
5.3 Chain rule and additivity . . . 125
5.3.1 Weak additivity under tensor products . . . 126
5.3.2 Duality . . . 127
5.3.3 Superadditivity under channel composition . . . 128
5.3.4 Strong additivity under tensor products . . . 134
5.4 Marginal-constrained entropy accumulation theorem (MEAT) . . . 137
5.4.1 f-weighted $H_\alpha^{\uparrow,f}$ entropies . . . 137

5.4.2 MEAT (simplified) . . . 144
5.4.3 MEAT (general statement) . . . 147
5.5 Security of PM-QKD protocols via the MEAT . . . 157
5.6 Conclusion and open questions . . . 161

6 Mutual information chain rules for security proofs robust against device imperfections 163
6.1 Introduction . . . 163
6.2 One-shot chain rule . . . 165
6.3 Information bounding theorem . . . 170
6.4 Entropy accumulation with leakage . . . 176
6.4.1 Relating Rényi conditional entropies to Rényi mutual information . 176
6.4.2 Incorporating the GEAT . . . 181
6.4.3 Incorporating generalized entropy accumulation with testing . . . 184
6.4.4 GEAT without non-signaling . . . 188
6.5 Example applications . . . 189
6.5.1 Physical leakage . . . 190
6.5.2 Source correlation . . . 190
6.5.3 Bounding the mutual information . . . 191
6.6 Conclusion and open questions . . . 193

7 Concluding remarks 195

References 197

APPENDICES 209

A Missing proofs in Chapter 4 210
A.1 Proof of Lemma 4.4.1 . . . 210
A.2 Proof of Lemma 4.4.2 . . . 211

B Generalized Rényi entropy accumulation for $\alpha < 1$ 214

List of Figures

2.1 The setting of the entropy accumulation theorem 31

2.2 The setting of generalized entropy accumulation theorem 33

3.1 Key rate comparison for BB84 with loss 62

3.2 Key rates for the qubit BB84 protocol with loss 63

3.3 Key rates for the decoy-state protocol under the GEAT analysis and the IID setting 65

3.4 Key rates for the decoy-state protocol 66

4.1 Graphical depiction of a non-signaling channel 70

4.2 Key rates for EB-BB84 from our approach, for an example with QBER threshold $Q_{\text{thresh}} = 0.025$, error-correction efficiency $\xi_{\text{EC}} = 1.1$, and security parameter $\varepsilon^{\text{secure}} = 10^{-10}$ 118

4.3 Key rates for EB-BB84 from our approach, for an example with QBER threshold $Q_{\text{thresh}} = 0.0451$, error-correction efficiency $\xi_{\text{EC}} = 1.19$, and security parameter $\varepsilon^{\text{secure}} = 10^{-10}$ 119

5.1 The setting for the marginal-constrained entropy accumulation theorem . . 147

6.1 Schematic setup of a sequence of GEATTL channels 185

6.2 The upper bound on Rényi mutual information in Eq (6.82) versus leakage probability δ_{leak} 193

List of Tables

2.1 List of notation 8

Chapter 1

Introduction

Quantum key distribution (QKD) is the task of establishing an information-theoretically secure key between two parties (Alice and Bob) by exploiting the laws of quantum mechanics. In a QKD task, Alice and Bob are connected by an insecure quantum channel controlled by the adversary and an authenticated classical channel. The adversary is allowed to intercept any states sent over the quantum channel and perform any valid attack within the realm of quantum mechanics, but she cannot alter or impersonate messages sent over the authenticated classical channel. The goal of a security proof for a QKD protocol is to show that Alice and Bob can generate a secure key under these conditions, regardless of the attack Eve employs.

What makes QKD special is that its security is based on the fundamental physical principles as opposed to classical cryptographic schemes, which rely on computationally difficult problems and can potentially be broken if the adversary gains enough computational power. In fact, QKD relies on unique quantum phenomena. For instance, the no-cloning theorem ensures that an eavesdropper cannot make perfect copies of states being transmitted, and acting on the transmitted states would inevitably disturb the system and reveal the presence of interception. Moreover, by exploiting entanglement between parties, Alice and Bob can verify their correlations through violations of Bell's inequality [12, 43], which rule out classical explanations and certify the presence of genuine quantum correlations.[1]

The first QKD protocol was introduced by Bennett and Brassard in their seminal 1984 paper [13], now widely referred to as the "BB84 protocol". Several years later, in 1991,

[1] Equivalently, in the ideal case of QKD security, Alice and Bob can verify that the correlations between them are completely decoupled from Eve. In this case, the entropic bound in the Leftover-Hashing Lemma is maximal (see Sec. 2.5.1).

Ekert introduced another protocol [43] based on the notion of entanglement, which is now known as the "E91 protocol". One of the most important features of QKD is a property known as "forward secrecy", a property which, in the QKD context, states that once a secure key is established via a QKD protocol, it remains secure forever, irrespective of the adversary's present and future computational power.

Most of the early security proofs for QKD were based on the assumption that the trusted parties (Alice and Bob) can send an unlimited number of signals to establish a secret key. In practice, though, Alice and Bob do not have access to an unlimited number of resources. Therefore, it is vital to construct a framework capable of analyzing the length of the secure key given a finite number of resources. Once we enter this regime, the analysis becomes more involved and complicated. For this reason, a large amount of the analysis in the literature has been done either for simple and symmetric protocols or under some restrictive assumptions on the adversary's attack [67, 104, 132, 52] (see also Sec. 2.5). Following the framework introduced in [104], these techniques are mostly designed to obtain lower bounds on some suitable form of smoothed min-entropy of the global state prior to the privacy amplification step (see Sec. 2.5.1).

Among techniques that do not restrict the adversary's capability, the *entropy accumulation theorem* (EAT) and its variations [37, 41, 89] are shown to provide a flexible framework for finite-size analysis while imposing minimal restrictions on the set of applicable protocols. The EAT provides a recipe to reduce the problem of evaluating smooth min-entropy over the global state to an analysis that depends on the information obtained in a single round of the protocol, which, informally speaking, can be described as follows. Suppose the state ρ in the protocol can be produced by n channels acting "sequentially" in some sense, and denote the resulting "secret" raw data as S_1^n and some final side-information register as E_n. Then, letting Ω denote the event that the protocol accepts (based on some "test data" computed over the n rounds), the EAT states that as long as ρ satisfies some technical conditions, its smooth min-entropy conditioned on Ω satisfies a bound of the following form:

$$\text{(informal summary)} \quad H_\infty^{\uparrow,\varepsilon}(S_1^n|E_n)_{\rho_{|\Omega}} \geq nh_{\text{vN}} - O(\sqrt{n}), \tag{1.1}$$

where, loosely speaking, h_{vN} is a value slightly smaller than the minimum von Neumann entropy over all single-round states that are "compatible with" the accept condition. This suffices to yield a security proof, since the $O(\sqrt{n})$ "finite-size correction" becomes negligible in comparison to the leading-order nh_{vN} term at large n, so $(1/n)H_\infty^{\uparrow,\varepsilon}(S_1^n|E_n)_{\rho_{|\Omega}}$ converges asymptotically to h_{vN}.

The original EAT was mostly suitable for analyzing a certain class of QKD protocols—namely, device-independent (DI) QKD and device-dependent entanglement-based

(EB) QKD (see Sec. 2.4). Subsequently, a generalized version known as the *generalized entropy accumulation theorem* (GEAT) was developed [89], which was designed to adapt to another important class of QKD protocols: device-dependent prepare-and-measure (PM) QKD. However, there are some limitations and drawbacks in using EAT/GEAT for security proofs, which we will briefly explain here.

- The finite-size corrections in both the original EAT and the GEAT are suboptimal. In fact, they are currently outperformed by other security proof techniques [119, 80].
- Although the GEAT generalizes the original EAT framework to accommodate PM-QKD protocols, it does so at the cost of imposing another form of restriction, namely, the repetition rate of the protocol must be limited such that there is only one signal at a time in the channel between Alice and Bob. We will refer to this as a "repetition-rate restriction".
- Both the original EAT and the GEAT rely on the assumption that the source used to prepare states in Alice's lab, as well as the detectors in both Alice's and Bob's labs, do not have imperfections involving memory across rounds.

In this thesis, we address all of the above limitations by providing an enhanced statistical analysis to significantly improve the finite-size corrections, introducing a variant of EAT to overcome the repetition-rate restriction, and presenting a framework to incorporate memory-dependent imperfections into the GEAT framework.

1.1 Outline of the thesis

In this section, we provide an overview of the results in this thesis.

Chapter 2: Background

In this chapter, we lay out the required concepts and notions that will be used throughout the thesis. We briefly introduce QKD and discuss the frameworks of EAT and GEAT.

Chapter 3: Finite-size analysis of prepare-and-measure and decoy-state QKD via GEAT

In this chapter, we provide a flexible framework for analyzing the finite-size effects of prepare-and-measure QKD protocols via the GEAT. We present a scheme to optimize various parameters in the GEAT, in particular a vital quantity termed the "min-tradeoff function", which significantly impacts both first- and second-order terms in the finite-size analysis. We apply this to an example of a decoy-state protocol and qubit BB84 with loss, demonstrating that reasonably robust key rates can be achieved in the finite-size regime.

Chapter 4: Improved Statistical analysis via generalized Rényi EAT

In this chapter, we extend the connection between the framework known as quantum probability estimation (QPE) [136] and the GEAT [89]. By doing so, we address some of the challenges we faced in Chapter 3; namely, we significantly improve the finite-size corrections in the GEAT model. In particular, we obtain a bound that does not involve a min-tradeoff function, thus completely avoiding the difficulties associated with optimizing its choice. Furthermore, the resulting bound appears to be extremely tight. Informally, the bound we obtain has the following form:

$$\text{(informal summary)} \quad H_\alpha^\uparrow(S_1^n|E_n)_{\rho_{|\Omega}} \geq nh_{\widehat{\alpha}} - \frac{\alpha}{\alpha-1}\log\frac{1}{\Pr[\Omega]}, \tag{1.2}$$

where $H_\alpha^\uparrow$ is a (sandwiched) Rényi entropy with $\alpha > 1$, and $h_{\widehat{\alpha}}$ is a Rényi version of h_{vN} in Eq. (1.1), i.e. loosely speaking, a value somewhat smaller than the minimum Rényi entropy over all single-round states that are "compatible with" the accept condition. Importantly, $h_{\widehat{\alpha}}$ can be computed as a convex optimization that does not involve any specification of an affine min-tradeoff function. However, we note that our analysis in this chapter is based on the GEAT model, meaning its application to the prepare-and-measure QKD protocol is still restricted by the limitations of GEAT for prepare-and-measure QKD protocols—i.e., the so-called repetition rate restriction. In the next chapter, we provide a framework that resolves this limitation.

Chapter 5: Marginal-constrained entropy accumulation

In this chapter, we address the repetition-rate restriction in the GEAT model. We begin by noting that, via the "source-replacement technique", any prepare-and-measure QKD protocol can be converted into an equivalent entanglement-based QKD protocol,

with an additional constraint: the marginal state in Alice's lab must remain untouched by the adversary. Unfortunately, such a constraint cannot be enforced within the GEAT framework due to certain technical limitations (see Chapter 5). To overcome this, we introduce another variant of EAT that accommodates this constraint, thus allowing us to convert a prepare-and-measure QKD protocol into an equivalent entanglement-based protocol where the repetition-rate restriction no longer applies (this is because, in that entanglement-based QKD protocol, the adversary is assumed to have access to the full state at the beginning of the protocol).

We obtain this EAT variant by introducing a notion of "channel conditional entropy", which is defined as

$$H_\alpha^\uparrow(\mathcal{M}, B, [\psi_A]) \coloneqq \inf_{\substack{\rho_{A\widetilde{A}\widetilde{R}} \\ \rho_A = \psi_A}} H_\alpha^\uparrow(B|C\widetilde{R})_{\mathcal{M}[\rho]}, \tag{1.3}$$

where $\mathcal{M} \in \mathrm{CPTP}(A\widetilde{A}, BC)$, and the optimization is over all density operators $\rho_{A\widetilde{A}\widetilde{R}}$ subject to a marginal constraint on A, with $\widetilde{R}$ being a stabilizing register. We then show that this quantity is superadditive under the composition of channels. Finally, we show that a statistical analysis similar to that of Chapter 4 can also be carried out within this model, yielding extremely tight bounds on key rates under this framework.

Chapter 6: Mutual information chain rules for security proofs robust against device imperfections

In this chapter, we go one step further and turn our attention to the scenarios where there are memory-based imperfections, either in the source that Alice uses to prepare states or the measurement devices that Alice and Bob use. To address these scenarios in a security proof, we propose various chain rules. In particular, we propose a technique to accommodate such imperfections in the GEAT framework. For instance, we show that for states satisfying a suitable channel structure, we can write

$$\text{(informal summary)} \quad H_\alpha^\uparrow(A|BC)_\rho \geq H_\alpha^\uparrow(A|B)_\rho - \xi_\alpha, \tag{1.4}$$

where ξ_α is essentially a Rényi mutual information involving C, after taking the "worst-case" value over possible channel outputs. We use this chain rule iteratively to accommodate imperfections in the GEAT framework. Alternatively, we prove a chain rule of the form

$$\text{(informal summary)} \quad H_\infty^{\uparrow,\epsilon}(S_1^n|L_1^n E_n)_\rho \geq H_\infty^{\uparrow,\epsilon'}(S_1^n|E_n)_\rho - I_{\max}^{\epsilon''}(S_1^n E_n; L_1^n)_\rho - [\text{small corrections}], \tag{1.5}$$

where the registers can be interpreted similarly to those in Eq. (1.1) with L_1^n being the registers containing leaked information due to imperfections in the devices. The chain rule in (1.5) has the appealing property of being applicable in any EAT-based framework, with the drawback that the adversary cannot interact with the leakage until the very end of the protocol. In contrast, the chain rule in (1.4), while allowing the adversary to interact actively with the leaked information, is only valid within the GEAT framework. To ease the calculation of smoothed max-information in Eq. (1.5), we also derive a bound on the Rényi mutual information of a state generated by a suitable sequence of channels as

$$I_\alpha^\downarrow(Z;L_1^n)_{\mathcal{M}_n\circ\cdots\circ\mathcal{M}_1(\rho^0)} \leq \sum_{i=1}^{n} \sup_{\omega_{\widetilde{Z}R_{i-1}}} I_\alpha^\downarrow(\widetilde{Z};L_i)_{\mathcal{M}_i(\omega)}, \tag{1.6}$$

where $\mathcal{M}_i \in \mathrm{CPTP}(R_{i-1}, R_iL_i)$ is a sequence of quantum channels, ρ^0 is some initial state, and Z is an overall "reference" register not acted on by the channels, while $\widetilde{Z}$ is a purifying register for each individual channel. We note that the above bound can be related to the smoothed max-information term in Eq. (1.5) via the conversions in [22].

Chapter 7: Concluding remarks

Finally, we conclude this thesis by sketching open questions that remain unresolved, as well as possible future directions for extending the frameworks introduced here and exploring their potential applications.

Chapter 2

Background

2.1 Notations and definitions

In this section, we present the basic notations and definitions used throughout this thesis. Table 2.1 summarizes some fundamental notations. In addition to these, several other key concepts are introduced below. In this thesis, we will assume that all systems are finite-dimensional, but we will not impose any bounds on the system dimensions unless otherwise specified. A "register" will refer to a labeled quantum system, that is, a finite-dimensional Hilbert space used to store quantum or classical information. A quantum channel (or completely positive trace-preserving map) will denote a physical transformation acting on registers, representing the most general allowed evolution of quantum systems.

Definition 2.1.1. (Frequency distributions) For a string $z_1^n \in \mathcal{Z}^n$ on some alphabet $\mathcal{Z}$, $\text{freq}_{z_1^n}$ denotes the following probability distribution on $\mathcal{Z}$:

$$\text{freq}_{z_1^n}(z) \coloneqq \frac{\text{number of occurrences of } z \text{ in } z_1^n}{n}. \tag{2.1}$$

Definition 2.1.2 (2-universal hash function)**.** A family $\mathcal{F}_{\mathcal{A}\to\mathcal{Z}}$ of functions from some set $\mathcal{A}$ to $\mathcal{Z} = \{0,1\}^l$ with a probability distribution p_f is called 2-universal if $\forall a, a' \in \mathcal{A}$ with $a \neq a'$, we have:

$$\Pr_{\mathcal{F}_{\mathcal{A}\to\mathcal{Z}}}[f(a) = f(a')] \leq \frac{1}{2^l}, \tag{2.2}$$

where the probability is computed with respect to the distribution p_f.

Table 2.1: List of notation

Symbol	*Definition*
$\log$	Base-2 logarithm
H	Base-2 von Neumann entropy
h_{bin}	Binary entropy function; $h_{\text{bin}}(x) \coloneqq -x \log x - (1-x)\log(1-x)$
$\mathbb{R} \cup \{-\infty, +\infty\}$	Extended real line
$\lfloor \cdot \rfloor$ (resp. $\lceil \cdot \rceil$)	Floor (resp. ceiling) function
$\|\cdot\|_p$	Schatten p-norm
$\langle A, B \rangle$	Inner product between two operators A, B
$\|\cdot\|$	Absolute value of operator; $\|M\| \coloneqq \sqrt{M^\dagger M}$
$A \perp B$	A and B are orthogonal; $AB = BA = 0$
$X \geq Y$ (resp. $X > Y$)	$X - Y$ is positive semi-definite (resp. positive definite)
$\text{CPTP}(A, B)$	Set of completely positive and trace preserving maps from A to B.
$\text{CP}(A, B)$	Set of completely positive maps from A to B.
$\text{End}(A)$	Set of linear operators on register A
$\text{Hom}(A, B)$	Set of linear operators from register A to B
$\text{Herm}(A)$	Set of hermitian operators on register A
$\text{Pos}(A)$	Set of positive semi-definite operators on register A
$\text{S}_=(A)$ (resp. $\text{S}_\leq(A)$)	Set of normalized (resp. subnormalized) states on register A
$\mathbb{U}_A$	Maximally mixed state on register A
A_j^k	Registers $A_j \dots A_k$

Definition 2.1.3. A state $\rho \in \mathrm{S}_{\leq}(CQ)$ is said to be **classical on** C (with respect to a specified basis on C) if it is in the form

$$\rho_{CQ} = \sum_c \lambda_c \left|c\right\rangle\!\left\langle c\right| \otimes \sigma_c, \tag{2.3}$$

for some normalized states $\sigma_c \in \mathrm{S}_{=}(Q)$ and weights $\lambda_c \geq 0$, with $|c\rangle$ being the specified basis states on C. In most circumstances, we will not explicitly specify this "classical basis" of C, leaving it to be implicitly defined by context. It may be convenient to absorb the weights λ_c into the states σ_c, writing them as subnormalized states $\omega_c = \lambda_c \sigma_c \in \mathrm{S}_{\leq}(Q)$ instead.

Definition 2.1.4. (Conditioning on classical events) For a state $\rho \in \mathrm{S}_{\leq}(CQ)$ classical on C, written in the form $\rho_{CQ} = \sum_c \left|c\right\rangle\!\left\langle c\right| \otimes \omega_c$ for some $\omega_c \in \mathrm{S}_{\leq}(Q)$, and an event Ω defined on the register C, we will define a corresponding **partial state** and **conditional state** as, respectively,

$$\rho_{\wedge\Omega} \coloneqq \sum_{c \in \Omega} \left|c\right\rangle\!\left\langle c\right| \otimes \omega_c, \qquad\qquad \rho_{|\Omega} \coloneqq \frac{\mathrm{Tr}\left[\rho\right]}{\mathrm{Tr}\left[\rho_{\wedge\Omega}\right]} \rho_{\wedge\Omega} = \frac{\sum_c \mathrm{Tr}\left[\omega_c\right]}{\sum_{c\in\Omega} \mathrm{Tr}\left[\omega_c\right]} \rho_{\wedge\Omega}. \tag{2.4}$$

The process of taking partial states is commutative and "associative", in the sense that for any events Ω, Ω' we have $(\rho_{\wedge\Omega})_{\wedge\Omega'} = (\rho_{\wedge\Omega'})_{\wedge\Omega} = \rho_{\wedge(\Omega\wedge\Omega')}$; hence for brevity we will denote all of these expressions as

$$\rho_{\wedge\Omega\wedge\Omega'} \coloneqq (\rho_{\wedge\Omega})_{\wedge\Omega'} = (\rho_{\wedge\Omega'})_{\wedge\Omega} = \rho_{\wedge(\Omega\wedge\Omega')}. \tag{2.5}$$

On the other hand, some disambiguating parentheses are needed when combined with taking conditional states (due to the normalization factors).

In light of the preceding two definitions, for a normalized state $\rho \in \mathrm{S}_{=}(CQ)$ that is classical on C, it is reasonable to write it in the form

$$\rho_{CQ} = \sum_c \rho(c) \left|c\right\rangle\!\left\langle c\right| \otimes \rho_{Q|c}, \tag{2.6}$$

where $\rho(c)$ denotes the probability of $C = c$ according to ρ, and $\rho_{Q|c}$ can indeed be interpreted as the conditional state on Q corresponding to $C = c$, i.e. $\rho_{Q|c} = \mathrm{Tr}_C\left[\rho_{|\Omega}\right]$ where Ω is the event $C = c$. We may sometimes denote the distribution on C induced by ρ as the tuple

$$\boldsymbol{\rho}_C \coloneqq (\rho(1), \rho(2), \dots). \tag{2.7}$$

Definition 2.1.5. For any $\rho \in \mathrm{S}_=(AB)$ we write $\rho_{A|B}$ to denote **the state of A conditioned on B**, which is defined as:

$$\rho_{A|B} \coloneqq \rho_B^{-\frac{1}{2}} \rho_{AB} \rho_B^{-\frac{1}{2}}. \tag{2.8}$$

We highlight that Definition 2.1.4 involves states conditioned on an *event* (or a specific value of some classical register), while Definition 2.1.5 involves states conditioned on a *register*. Qualitatively, to draw an analogue to classical probability distributions P_{AB} on random variables AB, the former is for instance about the notion of the conditional distributions $P_{A|b}$ conditioned on a specific event $B = b$, while the latter is about the notion of the conditional distribution $P_{A|B}$, which is a random variable that is a function of the random variable B.

Definition 2.1.6. (Measure-and-prepare or read-and-prepare channels) A (projective) **measure-and-prepare channel** is a channel $\mathcal{E} \in \mathrm{CPTP}(Q, QQ')$ of the form

$$\mathcal{E}[\rho_Q] = \sum_j (P_j \rho_Q P_j) \otimes \sigma_{Q'|j}, \tag{2.9}$$

for some projective measurement $\{P_j\}$ on Q and some normalized states $\sigma_{Q'|j}$. If Q is classical and the measurement is a projective measurement in its classical basis, we shall refer to it as a **read-and-prepare channel**. Note that a read-and-prepare channel always simply extends the state "without disturbing it", i.e., tracing out Q' results in the original state again.

We now state several measures used to distinguish between states.

Definition 2.1.7. For $\rho, \sigma \in \mathrm{S}_=(A)$, the **trace distance** is

$$T(\rho, \sigma) \coloneqq \frac{1}{2} \left\| \rho - \sigma \right\|_1. \tag{2.10}$$

Definition 2.1.8. For $\rho, \sigma \in \mathrm{S}_=(A)$, the **fidelity** is

$$F(\rho, \sigma) \coloneqq \left\| \sqrt{\rho}\sqrt{\sigma} \right\|_1 \tag{2.11}$$

Definition 2.1.9. For $\rho, \sigma \in \mathrm{S}_\leq(A)$, the **generalized fidelity** is

$$F(\rho, \sigma) \coloneqq \left\| \sqrt{\rho}\sqrt{\sigma} \right\|_1 + \sqrt{(1 - \mathrm{Tr}\left[\rho\right])(1 - \mathrm{Tr}\left[\sigma\right])}, \tag{2.12}$$

and the **purified distance** is $P(\rho, \sigma) \coloneqq \sqrt{1 - F(\rho, \sigma)^2}$. If either ρ or σ is normalized, this reduces to the standard definition of fidelity between normalized states.

2.2 Entropic quantities

We begin this section with an intuitive, informal description of classical entropy and divergence, and refer the reader to [117] for a formal and precise treatment of the material discussed here. Claude Shannon, in his seminal paper [106], provides a natural definition for a function that is supposed to characterize the information content of an event. Essentially, he showed that given an event Ω, and a probability of its occurrence P_Ω, a function $I(P_\Omega)$ characterizes the information content (sometimes referred to as the surprisal function) in that event if it satisfies the following conditions:

- $I(P_\Omega)$ is a non-negative and continuous function on $(0, 1]$.
- $I(P_\Omega)$ is monotone decreasing.
- for two independent events Ω and Ω', with probabilities P_Ω and $P_{\Omega'}$, respectively, we have $I(P_\Omega P_{\Omega'}) = I(P_\Omega) + I(P_{\Omega'})$.

As it turns out [106], the only function satisfying these requirements, up to a multiplicative constant, is the following:[1]

$$I(P_\Omega) = \log\left(\frac{1}{P_\Omega}\right). \tag{2.13}$$

With this in mind, consider a random variable X, and a probability distribution $\mathbf{p}$ associated with X, i.e., $\{p_x\}_{x\in X}$. The Shannon entropy of a probability distribution $\mathbf{p}$ associated with the random variable X is defined as the weighted arithmetic mean of the information content over $\mathbf{p}$, i.e.,

$$H(\mathbf{p}) \coloneqq \sum_{x\in X} p_x I(p_x). \tag{2.14}$$

Shannon entropy plays a crucial role in the foundations of information theory. For example, it characterizes the optimal asymptotic rate for the task of source compression [106].

Another quantity that can be obtained from the information content function, which is widely used, is "Kullback-Leibler divergence" (also referred to as "relative entropy" or "Shannon's information gain") [105, 74]. Let X be a random variable, and $\mathbf{q}$ be a probability

[1] We highlight that the requirements for the information content function are somewhat intuitive. For example, the second requirement imposes the fact that an event with less probability of occurrence contains more information.

distribution associated with X, i.e., $\{q_x\}_{x\in X}$. Furthermore, let Ω be an event; we are interested in how much information is gained upon observing the event Ω. To this end, let $\mathbf{p}$ be the probability distribution associated with X *conditioned* on observing the event Ω, i.e., $\{p_x\}_{x\in X}$. For any $x \in X$, the information gain, upon observing Ω is given by:

$$I_{\text{gain}}(p_x, q_x) = \log\left(\frac{1}{q_x}\right) - \log\left(\frac{1}{p_x}\right). \tag{2.15}$$

Then, the Kullback-Leibler divergence between $\mathbf{p}$ and $\mathbf{q}$ is denoted by $D(\mathbf{p}\|\mathbf{q})$, and is given by the weighted arithmetic mean of the information gain over $\mathbf{q}$, i.e.,[2]

$$D(\mathbf{p}\|\mathbf{q}) \coloneqq \sum_{x\in X} p_x I_{\text{gain}}(p_x, q_x). \tag{2.16}$$

We highlight that both Shannon entropy in Eq. (2.14) and Kullback-Leibler divergence in Eq. (2.16) were defined such that they preserve the third condition of the information content function. In other words, the following relation holds:

$$H(\mathbf{p}\otimes\mathbf{q}) = H(\mathbf{p}) + H(\mathbf{q}), \tag{2.17}$$
$$D(\mathbf{p}\otimes\mathbf{r}\|\mathbf{q}\otimes\mathbf{s}) = D(\mathbf{p}\|\mathbf{q}) + D(\mathbf{r}\|\mathbf{s}). \tag{2.18}$$

A natural question at this point is what alternative quantities would be obtained if we instead define the entropy and divergence through different notions of means. Alfred Rényi, in his seminal paper [105], addressed this question by focusing on a particular class of generalized means while requiring that the relations in Eq. (2.17) be preserved. The class of generalized means he studied is referred to as "Quasi-arithmetic mean" and defined as follows. Consider a random variable X, and a probability distribution $\mathbf{p}$ associated with X, i.e., $\{p_x\}_{x\in X}$, furthermore, let, g be a monotonic function with an inverse g^{-1}. Then, the general mean associated with g is defined as

$$g^{-1}\left(\sum_{x\in X} p_x g(x)\right). \tag{2.19}$$

Informally, we can think of the function g as a transformation that maps into the space where the arithmetic mean is "natural", then taking the arithmetic mean and transforming the result back into the original space.

[2]We note that, in general, these quantities can be defined for generalized probability distributions. However, for simplicity and to obtain a better intuition, we only defined them for probability distributions.

Rényi observed that aside from the arithmetic mean, the only class of g for which, up to a multiplicative constant, the resulting entropy and divergence satisfy the relation in (2.17), is $g(x) = 2^{(1-\alpha)x}$, for $\alpha \geq 0$.[3] Using this particular choice of mean function, the corresponding entropy and divergence are defined as:

$$H_\alpha(\mathbf{p}) \coloneqq \frac{1}{1-\alpha} \log \left(\sum_{x \in X} p_x^\alpha \right), \tag{2.20}$$

$$D_\alpha(\mathbf{p} \| \mathbf{q}) \coloneqq \frac{1}{\alpha - 1} \log \left(\sum_{x \in X} p_x^\alpha q_x^{1-\alpha} \right). \tag{2.21}$$

Rényi divergences have found many applications in classical information theory [105, 135, 58, 125, 91, 84]. However, unlike the classical case, once we enter the quantum world, there are multiple ways to generalize these divergences. In what follows, we state the definition of some of the quantum generalizations of Rényi divergence, along with related entropic quantities derived from these divergences that will be used throughout this thesis. For a more comprehensive discussion of these quantities, we refer the reader to [117, 68].

Definition 2.2.1. (Sandwiched Rényi divergence [93, 131]) For any $\rho, \sigma \in \mathrm{Pos}(A)$ with $\mathrm{Tr}[\rho] \neq 0$, and $\alpha \in (0,1) \cup (1,\infty)$, the (sandwiched) Rényi divergence between ρ, σ is defined as:

$$D_\alpha(\rho \| \sigma) = \begin{cases} \frac{1}{\alpha - 1} \log \frac{\mathrm{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right]}{\mathrm{Tr}[\rho]} & (\alpha < 1 \ \wedge \ \rho \not\perp \sigma) \vee (\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)) \\ +\infty & \text{otherwise,} \end{cases} \tag{2.22}$$

where for $\alpha > 1$ the $\sigma^{\frac{1-\alpha}{2\alpha}}$ terms are defined via the Moore-Penrose pseudoinverse if σ is not full-support [117]. The above definition is extended to $\alpha \in \{0, 1, \infty\}$ by taking the respective limits. For the $\alpha = 1$ case, it reduces to the Umegaki divergence [124]:

$$D(\rho \| \sigma) = \begin{cases} \frac{\mathrm{Tr}[\rho \log \rho - \rho \log \sigma]}{\mathrm{Tr}[\rho]} & \mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma) \\ +\infty & \text{otherwise.} \end{cases} \tag{2.23}$$

For any two classical probability distributions $\mathbf{p}, \mathbf{q}$ on a common alphabet, we also define the Rényi divergence $D_\alpha(\mathbf{p} \| \mathbf{q})$ analogously, e.g. by viewing the distributions as diagonal density matrices in the above formulas; in the $\alpha = 1$ case this gives the Kullback–Leibler (KL) divergence.

[3] More precisely, Rényi axiomatized the notion of divergence, and then he proved that the only two glass of g that satisfy those axioms are linear functions $g(x) = cx$ (which leads to Shannon's entropy and information gain function), and exponential functions of the form $g(x) = 2^{(1-\alpha)x}$, for $\alpha \geq 0$. [105]

Another generalization of the KL divergence to the quantum setting is the Petz divergence [98, 99, 116], which is defined as the following:

Definition 2.2.2. (Petz Rényi divergence [116]) For any $\rho, \sigma \in \mathrm{Pos}(A)$ with $\mathrm{Tr}[\rho] \neq 0$, and $\alpha \in (0,1) \cup (1,\infty)$, the Petz divergence between ρ, σ is defined as:

$$\bar{D}_\alpha(\rho\|\sigma) = \begin{cases} \frac{1}{\alpha-1} \log \frac{\mathrm{Tr}\left[\rho^\alpha \sigma^{1-\alpha}\right]}{\mathrm{Tr}[\rho]} & (\alpha < 1 \ \wedge \ \rho \not\perp \sigma) \vee (\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)) \\ +\infty & \text{otherwise,} \end{cases} \tag{2.24}$$

where for $\alpha > 1$ the $\sigma^{1-\alpha}$ term is defined via the Moore-Penrose pseudoinverse if σ is not full-support [117]. The above definition is extended to $\alpha \in \{0, 1, \infty\}$ by taking the respective limits. Similar to the sandwiched Rényi divergences, for the $\alpha = 1$ case it also reduces to the Umegaki divergence, and for classical states it also reduces to the classical Rényi divergences.

We will also be using the measured Rényi divergence, which is defined as follows.

Definition 2.2.3. (Measured Rényi divergence [51, Eqs. 3.116-3.117]) For any state $\rho, \sigma \in \mathrm{Pos}(A)$, with $\mathrm{Tr}[\rho] \neq 0$, and $\alpha \in [0,\infty]$, the measured Rényi divergence between ρ, σ is defined as:

$$D_{\mathbb{M},\alpha}(\rho\|\sigma) = \sup_{\mathcal{X},M} D_\alpha(P_{\rho,M}\|P_{\sigma,M}), \tag{2.25}$$

where the optimization is over finite sets $\mathcal{X}$, and POVMs M on it, with $P_{\rho,M}(x) = \mathrm{Tr}[M(x)\rho]$, and $D_\alpha(P_{\rho,M}\|P_{\sigma,M})$ is the classical Rényi divergence between the two distributions.

Rényi divergences can also be defined for a pair of CPTP maps, by taking the worst case over the input states [78].

Definition 2.2.4. (Stabilized Rényi channel divergence) For $\mathcal{M} \in \mathrm{CPTP}(A,B)$, $\mathcal{N} \in \mathrm{CPTP}(A,B)$, and $\alpha \in [\frac{1}{2}, 1) \cup (1,\infty)$, the stabilized Rényi channel divergence is defined as

$$\mathbb{D}_\alpha(\mathcal{M}||\mathcal{N}) \coloneqq \sup_{\omega \in \mathrm{S}_=(AR)} \mathbb{D}_\alpha\left(\mathcal{M}(\omega_{AR})||\mathcal{N}(\omega_{AR})\right), \tag{2.26}$$

where $\mathbb{D}_\alpha$ can be any one of $\{D_\alpha, D_{\mathbb{M},\alpha}, \bar{D}_\alpha\}$, and R is isomorphic to A.

A fundamental quantity in quantum key distribution is the sandwiched Rényi conditional entropy. This quantity provides a bound on the length of the key that can be securely extracted in the presence of quantum side information [42].

Definition 2.2.5. (Sandwiched Rényi conditional entropies) For any bipartite state $\rho \in \mathrm{S}_{=}(AB)$, and $\alpha \in [0, \infty]$, we define the following two (sandwiched) Rényi conditional entropies:

$$\begin{aligned} H_\alpha(A|B)_\rho &= -D_\alpha(\rho_{AB} \| \mathbb{I}_A \otimes \rho_B) \\ H_\alpha^\uparrow(A|B)_\rho &= \sup_{\sigma_B \in \mathrm{S}_{=}(B)} -D_\alpha(\rho_{AB} \| \mathbb{I}_A \otimes \sigma_B). \end{aligned} \tag{2.27}$$

For $\alpha = 1$, both of the above values coincide and are equal to the von Neumann entropy. Note that the cases $\alpha = \infty$ and $\alpha = 1/2$ are commonly referred to as the "conditional min-entropy" "and conditional max-entropy", respectively.

Definition 2.2.6. (Petz Rényi conditional entropies) For any bipartite state $\rho \in \mathrm{S}_{=}(AB)$, and $\alpha \in [0, \infty]$, we define the following two Petz conditional entropies:

$$\begin{aligned} \bar{H}_\alpha(A|B)_\rho &= -\bar{D}_\alpha(\rho_{AB} \| \mathbb{I}_A \otimes \rho_B) \\ \bar{H}_\alpha^\uparrow(A|B)_\rho &= \sup_{\sigma_B \in \mathrm{S}_{=}(B)} -\bar{D}_\alpha(\rho_{AB} \| \mathbb{I}_A \otimes \sigma_B). \end{aligned} \tag{2.28}$$

For $\alpha = 1$, both of the above values coincide and are equal to the von Neumann entropy.

In this thesis, we will also make use of Rényi mutual information. In particular, informally, we will be using it to characterize the amount of leakage occurring in a quantum key distribution protocol.

Definition 2.2.7. (Rényi mutual information [131, Definition 12]) For any bipartite state $\rho \in \mathrm{S}_{\leq}(AB)$ with $\mathrm{Tr}\,[\rho] \neq 0$, and $\alpha \in [0, \infty]$, we define the following (sandwiched) Rényi mutual information:

$$\begin{aligned} I_\alpha(A:B)_\rho &\coloneqq D_\alpha(\rho_{AB} || \rho_A \otimes \rho_B), \\ I_\alpha^\downarrow(A;B)_\rho &\coloneqq \inf_{\sigma_B \in \mathrm{S}_{\leq}(B)} D_\alpha(\rho_{AB} || \rho_A \otimes \sigma_B), \\ I_\alpha^{\downarrow\downarrow}(A:B)_\rho &\coloneqq \inf_{\omega_A \in \mathrm{S}_{\leq}(A), \sigma_B \in \mathrm{S}_{\leq}(B)} D_\alpha(\rho_{AB} || \omega_A \otimes \sigma_B). \end{aligned} \tag{2.29}$$

All the above optimizations can instead be restricted to normalized states without changing the optimal values. For $\alpha = 1$ and $\mathrm{Tr}\,[\rho] = 1$, we have $I_\alpha(A:B) = I_\alpha^\downarrow(A;B) = I_\alpha^{\downarrow\downarrow}(A:B)$ and their values are equal to the von Neumann mutual information. Note that while I_α and $I_\alpha^{\downarrow\downarrow}$ are symmetric with respect to partial registers, $I_\alpha^\downarrow$ is not (i.e., $I_\alpha^\downarrow(A;B)_\rho \neq I_\alpha^\downarrow(B;A)_\rho$).

Definition 2.2.8. (Rényi conditional mutual information) For any tripartite state $\rho \in \mathrm{S}_{=}(ABC)$, and $\alpha \in [0, \infty]$, we define the following (sandwiched) Rényi conditional mutual information

$$I_{\alpha}^{\downarrow,\mathrm{diff}}(A;B|C)_{\rho} \coloneqq I_{\alpha}^{\downarrow}(A;BC)_{\rho} - I_{\alpha}^{\downarrow}(A;C)_{\rho} \tag{2.30}$$

For $\alpha = 1$, this definition is equal to the von Neumann conditional mutual information. Note that this definition is not symmetric with respect to partial registers(i.e., $I_{\alpha}^{\downarrow,\mathrm{diff}}(A;B|C)_{\rho} \neq I_{\alpha}^{\downarrow,\mathrm{diff}}(B;A|C)_{\rho}$).

We now briefly introduce smooth entropies [116, 31, 117].

Definition 2.2.9. (ε-smoothed max-divergence [28]) For $\rho \in \mathrm{S}_{\leq}(A), \sigma \in \mathrm{Pos}(A)$ with $\mathrm{Tr}\left[\rho\right] \neq 0$, and any $\epsilon \geq 0$, the ε-smoothed max-divergence is defined as

$$D_{\infty}^{\varepsilon}(\rho\|\sigma) \coloneqq \inf_{\substack{\tilde{\rho}\in\mathrm{S}_{\leq}(A) \text{ s.t.} \\ P(\tilde{\rho},\rho)\leq\varepsilon}} D_{\infty}(\tilde{\rho}\|\sigma). \tag{2.31}$$

Note that the infimum is attained in the above definition, but it may not necessarily be attained by a normalized state, even if ρ is normalized [117] (see [102] for discussions of circumstances under which the infimum may fail to be attained by a normalized state).

Similarly, we define various entropic quantities based on smoothed max-divergence.

Definition 2.2.10. (Smoothed entropies) For $\rho \in \mathrm{S}_{\leq}(AB)$ with $\mathrm{Tr}\left[\rho\right] \neq 0$, and any $\varepsilon \in \left[0, \sqrt{\mathrm{Tr}\left[\rho_{AB}\right]}\right)$, let

$$H_{\infty}^{\uparrow,\varepsilon}(A|B)_{\rho} \coloneqq \sup_{\substack{\tilde{\rho}\in\mathrm{S}_{\leq}(AB) \text{ s.t.} \\ P(\tilde{\rho},\rho)\leq\varepsilon}} H_{\infty}^{\uparrow}(A|B)_{\tilde{\rho}} = -\inf_{\sigma_B\in\mathrm{S}_{=}(B)} D_{\infty}^{\varepsilon}(\rho_{AB}\|\mathbb{I}_A \otimes \sigma_B), \tag{2.32}$$

$$H_{1/2}^{\uparrow,\varepsilon}(A|B)_{\rho} \coloneqq \sup_{\substack{\tilde{\rho}\in\mathrm{S}_{\leq}(AB) \text{ s.t.} \\ P(\tilde{\rho},\rho)\leq\varepsilon}} H_{1/2}^{\uparrow}(A|B)_{\tilde{\rho}}, \tag{2.33}$$

$$I_{\infty}^{\varepsilon}(A:B)_{\rho} \coloneqq \inf_{\substack{\tilde{\rho}\in\mathrm{S}_{\leq}(AB) \text{ s.t.} \\ P(\tilde{\rho},\rho)\leq\varepsilon}} I_{\infty}(A:B)_{\tilde{\rho}}, \tag{2.34}$$

$$I_{\infty}^{\downarrow,\varepsilon}(A;B)_{\rho} \coloneqq \inf_{\substack{\tilde{\rho}\in\mathrm{S}_{\leq}(AB) \text{ s.t.} \\ P(\tilde{\rho},\rho)\leq\varepsilon}} I_{\infty}^{\downarrow}(A;B)_{\tilde{\rho}}, \tag{2.35}$$

$$I_{\infty}^{\downarrow\downarrow,\varepsilon}(A:B)_{\rho} \coloneqq \inf_{\substack{\tilde{\rho}\in\mathrm{S}_{\leq}(AB) \text{ s.t.} \\ P(\tilde{\rho},\rho)\leq\varepsilon}} I_{\infty}^{\downarrow\downarrow}(A:B)_{\tilde{\rho}} = \inf_{\omega_A\in\mathrm{S}_{=}(A),\sigma_B\in\mathrm{S}_{=}(B)} D_{\infty}^{\varepsilon}(\rho_{AB}\|\omega_A \otimes \sigma_B). \tag{2.36}$$

Note that the quantities in the first and second lines are commonly referred to as the "smoothed min-entropy" and "smoothed max-entropy", respectively. We highlight that, again, in general, the infima above may not necessarily be attained by a normalized state. However, for $I_\infty^{\downarrow,\varepsilon}(A;B)_\rho$ with normalized ρ, the optimum is always attainable by a normalized $\tilde{\rho}$. To see this, observe that $D_\infty(t\tilde{\rho}_{AB}||t\tilde{\rho}_A \otimes \sigma_B) = D_\infty(\tilde{\rho}_{AB}||\tilde{\rho}_A \otimes \sigma_B)$ for any $t \geq 0$ (and any σ), therefore, we can just normalize any $\tilde{\rho}$ attaining the optimum (also noting that since ρ is normalized, this normalization will only increase the value of $F(\tilde{\rho}, \rho)$ according to Definition 2.1.9).

We also note that we only expressed the first and last quantities in terms of D_∞^ε. This is because, in the definition of the smoothed max-divergence above, the smoothing optimization is only over the first argument of the divergence, while for $I_\infty^{\downarrow,\varepsilon}$ and I_∞^ε, a marginal of the state in the first argument of the divergence also appears in the second argument. For the purposes of this thesis, however, these discrepancies are not a significant concern, as the quantities can be converted into one another with some loss in the smoothing parameter via [22, Theorem 3].

We highlight that prior to [42], smoothed min-entropy was the primary quantity used to obtain bounds on the length of the secure key that can be extracted in the presence of quantum side information. This is because it was shown in [104, 119] that smoothed min-entropy provides an achievable rate for such a task.

With these definitions in place, we now present several properties of these quantities that are useful in this thesis.

Theorem 2.2.1. *(Data-processing [50, Theorem 1]; see also [93, 11, 92, 117]) For any $\alpha \in [1/2, \infty]$, any $\rho, \sigma \in \mathrm{Pos}(Q)$ with $\mathrm{Tr}[\rho] \neq 0$, and any channel $\mathcal{E} \in \mathrm{CPTP}(Q, Q')$, we have:*

$$D_\alpha(\rho\|\sigma) \geq D_\alpha(\mathcal{E}[\rho]\|\mathcal{E}[\sigma]), \tag{2.37}$$

and thus also for any $\rho \in \mathrm{S}_=(Q''Q)$,

$$\begin{aligned} H_\alpha(Q''|Q)_\rho &\leq H_\alpha(Q''|Q')_{\mathcal{E}[\rho]}, \quad H_\alpha^\uparrow(Q''|Q)_\rho \leq H_\alpha^\uparrow(Q''|Q')_{\mathcal{E}[\rho]} \\ I_\alpha(Q'':Q)_\rho &\geq I_\alpha(Q'':Q')_{\mathcal{E}[\rho]}, \quad I_\alpha^{\downarrow\downarrow}(Q'':Q)_\rho \geq I_\alpha^{\downarrow\downarrow}(Q'':Q')_{\mathcal{E}[\rho]}, \\ I_\alpha^\downarrow(Q'';Q)_\rho &\geq I_\alpha^\downarrow(Q'';Q')_{\mathcal{E}[\rho]}, \quad I_\alpha^\downarrow(Q;Q'')_\rho \geq I_\alpha^\downarrow(Q';Q'')_{\mathcal{E}[\rho]}. \end{aligned} \tag{2.38}$$

If $\mathcal{E}$ is an isometry, all the above bounds hold with equality.

Lemma 2.2.1. *(Conditioning on classical registers; see [93, Sec. III.B.2 and Proposition 9] or [117, Eq. (5.32) and Proposition 5.1]) Let $\rho, \sigma \in \mathrm{S}_=(CQ)$ be states classical on C. Then*

$$\forall \alpha \in (0,1) \cup (1,\infty), \qquad \mathbb{D}_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1} \log \left(\sum_c \rho(c)^\alpha \sigma(c)^{1-\alpha} 2^{(\alpha-1)\mathbb{D}_\alpha(\rho_{Q|c}\|\sigma_{Q|c})} \right), \tag{2.39}$$

where $\mathbb{D}_\alpha \in \{D_\alpha, \bar{D}_\alpha\}$. Hence, for a state $\rho \in \mathrm{S}_=(CQQ')$ classical on C,

$$\forall \alpha \in (0,1) \cup (1,\infty), \qquad \mathbb{H}_\alpha(Q|CQ') = \frac{1}{1-\alpha} \log \left(\sum_c \rho(c) 2^{(1-\alpha)\mathbb{H}_\alpha(Q|Q')_{\rho|c}} \right), \tag{2.40}$$

$$\forall \alpha \in [1/2,1) \cup (1,\infty), \qquad \mathbb{H}^\uparrow_\alpha(Q|CQ') = \frac{\alpha}{1-\alpha} \log \left(\sum_c \rho(c) 2^{\frac{1-\alpha}{\alpha}\mathbb{H}^\uparrow_\alpha(Q|Q')_{\rho|c}} \right), \tag{2.41}$$

where $\mathbb{H}_\alpha \in \{H_\alpha, \bar{H}_\alpha\}$.

Proposition 2.2.1. *(Relations between smooth divergence and sandwiched Rényi divergence; Proposition 6.5 in [117]) Let $\rho \in \mathrm{S}_=(A)$, $\sigma \in \mathrm{Pos}(A)$. Then, for any $\epsilon \in (0,1)$, $\alpha \in (1,\infty)$, we have*

$$D^\epsilon_\infty(\rho||\sigma) \leq D_\alpha(\rho\|\sigma) + \frac{g(\epsilon)}{\alpha - 1}, \tag{2.42}$$

where $g(\epsilon) := \log\left(1 - \sqrt{1-\epsilon^2}\right)$. In particular, for any $\rho \in \mathrm{S}_=(AB)$,

$$I^{\downarrow\downarrow,\epsilon}_\infty(A:B)_\rho \leq I^{\downarrow\downarrow}_\alpha(A:B)_\rho + \frac{g(\epsilon)}{\alpha - 1}. \tag{2.43}$$

To obtain the second bound from the first (which was proven in [117, Proposition 6.5]): observe that for any $\rho \in \mathrm{S}_=(AB), \omega_A \in \mathrm{S}_=(A), \sigma_B \in \mathrm{S}_=(B)$, from the first bound we have

$$D^\varepsilon_\infty(\rho_{AB}||\omega_A \otimes \sigma_B) \leq D_\alpha(\rho_{AB}||\omega_A \otimes \sigma_B) + \frac{g(\varepsilon)}{\alpha - 1}. \tag{2.44}$$

Taking infimum over ω, σ in the above inequality, and noting the relation between $I^{\downarrow\downarrow,\epsilon}_\infty$ and D^ϵ_∞ described in Definition 2.2.10, the claim follows. (Generalizing this property to, say, $I^{\downarrow,\epsilon}_\infty$ would require a suitable handling of the reduced state on A in the second argument, as discussed above.)

Proposition 2.2.2. *(Duality relation [117, Proposition 5.3, 5.4]; see also [93, Theorem 10],[118, Theorem 3]) For any pure state $\rho \in \mathrm{S}_{=}(ABC)$, we have*

$$\begin{aligned} &H_{\alpha}^{\uparrow}(A|B)_{\rho} + H_{\beta}^{\uparrow}(A|C)_{\rho} = 0, \quad \frac{1}{\alpha} + \frac{1}{\beta} = 2, \qquad \alpha, \beta \in [\frac{1}{2}, \infty] \\ &\bar{H}_{\alpha}^{\uparrow}(A|B)_{\rho} + H_{\beta}(A|C)_{\rho} = 0, \quad \alpha\beta = 1, \qquad \alpha, \beta \in [0, \infty] \end{aligned} \tag{2.45}$$

2.3 Basic convex optimization concepts

Convex optimization has become an essential tool in quantum information theory. Many optimization problems in this field can be formulated as convex optimization problems. This is especially important, as this class of problems can be solved "efficiently" using numerical algorithms.

In this section, we briefly discuss the tools needed for this thesis and refer the reader to [18] for a comprehensive treatment of the topic.

Definition 2.3.1 (Convex set)**.** A vector space $\mathcal{S}$ is called convex if for any $X, Y \in \mathcal{S}$ and any $\lambda \in [0, 1]$, we have $\lambda X + (1 - \lambda)Y \in \mathcal{S}$.

Definition 2.3.2 (Convex hull)**.** For a vector space $\mathcal{S}$, the convex hull of $\mathcal{S}$ is denoted by $\mathrm{conv}(\mathcal{S})$ and defined as the set of all the convex combinations of points in $\mathcal{S}$:

$$\mathrm{conv}(\mathcal{S}) \coloneqq \left\{ \sum_{i=1}^{k} \lambda_i X_i \mid \ X_1, \cdots, X_k \in \mathcal{S} \ \wedge \ \sum_{i=1}^{k} \lambda_i = 1, \ \text{for some} \ \ k \in \mathbb{N} \right\}. \tag{2.46}$$

Definition 2.3.3. (Convex function) Let $\mathcal{S}$ be a vector space, and let $F : \mathcal{S} \to \mathbb{R}$ be a real-valued function. Then, the function F is convex if $\mathcal{S}$ is a convex set, and for all $X, Y \in \mathcal{S}$, and $\lambda \in [0, 1]$ we have:

$$F\left(\lambda X + (1 - \lambda)Y\right) \leq \lambda F(X) + (1 - \lambda)F(Y). \tag{2.47}$$

Definition 2.3.4. (Affine function) A function $F : \mathbb{R}^n \to \mathbb{R}^m$ is called affine if there exists $A \in \mathbb{R}^{mn}$ and $B \in \mathbb{R}^m$ such that F can be written as:

$$F(X) = AX + B \tag{2.48}$$

2.3.1 Conjugate and duality

Definition 2.3.5. (Convex conjugate) For a function $F : \mathcal{D} \to \mathbb{R} \cup \{-\infty, +\infty\}$ where $\mathcal{D}$ is a convex subset of $\mathbb{R}^k$, its **convex conjugate** is the function $F^* : \mathbb{R}^k \to \mathbb{R} \cup \{-\infty, +\infty\}$ given by

$$F^*(\mathbf{y}) \coloneqq \sup_{\mathbf{x} \in \mathcal{D}} \left(\mathbf{x} \cdot \mathbf{y} - F(\mathbf{x})\right). \tag{2.49}$$

Intuitively, the convex conjugate function provides the best linear approximation from below of the original function f at a given slope y (see [18] for more details).

Proposition 2.3.1. *(Sion's minimax theorem [109]; see also[68, Theorem 2.24]) Let S be a compact, convex subset of a normed vector space, and S' be a convex subset of a normed vector space. Let $F : S \times S' \to \mathbb{R}$ be a real-valued function such that*

- *$F(X, \cdot) : S' \to \mathbb{R}$ is upper semi-continuous and quasi-concave on S' for every $X \in S$.*
- *$F(\cdot, Y) : S \to \mathbb{R}$ is lower semi-continuous and quasi-convex on S for every $Y \in S'$.*

Then, the following holds:

$$\inf_{X \in S} \sup_{Y \in S'} F(X, Y) = \sup_{Y \in S'} \inf_{X \in S} F(X, Y). \tag{2.50}$$

Furthermore, the infimum can be replaced with a minimum.

2.3.2 Semi-definite program

A semi-definite program (SDP) is an important type of convex optimization problem. It plays a significant role in quantum information theory. Many quantities in quantum information theory, such as fidelity and trace norm, can be formulated equivalently as an SDP (see [68, 130]). In this thesis, we will also use SDP properties to obtain some of the results in Chapter 5. Therefore, we state the definition of an SDP and briefly review some of its properties.

Definition 2.3.6. Let $\Psi \in \mathrm{Hom}(A, B)$ be a hermiticity-preserving map. Let $A \in \mathrm{Herm}(A)$, $B \in \mathrm{Herm}(B)$ be hermitian operators. A semi-definite program is a triple (Ψ, A, B), corresponds to the following optimization problems:

$$\sup_{X \in \mathrm{Pos}(A)} \langle A, X \rangle \quad \text{s.t. } \Psi(X) = B \tag{2.51}$$

$$\inf_{Y \in \mathrm{Herm}(B)} \langle B, Y \rangle \quad \text{s.t. } \Psi^\dagger(Y) \geq A. \tag{2.52}$$

We call (2.51) the primal problem and (2.52) the dual problem.

For the primal problem in (2.51), we call a point X feasible if $X \in \mathrm{Pos}(A)$ and $\Psi(X) = B$. Similarly, for the dual problem in (2.52), a point Y is feasible if $Y \in \mathrm{Herm}(B)$ and $\Psi^\dagger(Y) \geq A$.

Definition 2.3.7. (Weak and Strong Duality) Let (Ψ, A, B) be a semi-definite program. Denote by α and β the optimal values of the primal and dual problems in Eqs. (2.51) and (2.52), respectively. Then, the above SDP satisfies the **weak duality** property iff

$$\alpha \leq \beta, \tag{2.53}$$

and it satisfies **strong duality** iff

$$\alpha = \beta. \tag{2.54}$$

The following propositions provide the criteria under which these properties hold.

Proposition 2.3.2. *Let (Ψ, A, B) be a semi-definite program. Denote by α and β the optimal values of the primal and dual problems. Then,*

$$\alpha \leq \beta. \tag{2.55}$$

We now state the following proposition, which provides a sufficient condition for strong duality to hold.

Proposition 2.3.3. *(Slater's condition) Let (Ψ, A, B) be a semi-definite program with α and β denoting the optimal values of the primal and dual problems, respectively. Then*

1.

$$\begin{aligned}&\left(\exists\, X \geq 0 \text{ s.t. } \Psi(X) = B\right) \wedge \left(\exists\, Y > 0 \text{ s.t. } \Psi^\dagger(Y) > A\right) \\ &\qquad \implies (\alpha = \beta) \wedge \left(\exists\, X \in \mathrm{Pos}(A) \text{ s.t. } \Psi(X) = B \wedge \alpha = \mathrm{Tr}[AX]\right)\end{aligned} \tag{2.56}$$

2.

$$\begin{aligned}&\left(\exists\, Y \geq 0 \text{ s.t. } \Psi^\dagger(Y) \geq A\right) \wedge \left(\exists\, X > 0 \text{ s.t. } \Psi(X) = B\right) \\ &\qquad \implies (\alpha = \beta) \wedge \left(\exists\, Y \in \mathrm{Herm}(B) \text{ s.t. } \Psi^\dagger(Y) \geq A \wedge \beta = \mathrm{Tr}[BY]\right)\end{aligned} \tag{2.57}$$

2.4 Introduction to Quantum Key Distribution (QKD)

In this section, we present the general structure of QKD protocols, briefly discuss their steps, and provide a formal definition of security.

There are two major types of QKD protocols that differ based on the assumptions made about Alice's and Bob's devices: device-dependent and device-independent protocols. In device-dependent QKD, it is assumed that Alice's and Bob's devices are (possibly up to some small imperfections) fully characterized and trusted, whereas in the device-independent case, their devices are only minimally characterized. In this thesis, we focus solely on the device-dependent case, while noting that some of the results we obtain are also applicable to device-independent scenarios.

2.4.1 Generic device-dependent QKD protocols

In this part, we sketch the general structure of generic device-dependent protocols. We begin with entanglement-based protocols. We highlight that while the protocol steps in this chapter allow for a quite generic range of protocols, in the upcoming chapters, we restrict our attention to a suitable subset that can be analyzed using the techniques we develop.

Protocol 1. Entanglement-based protocols
Parameters:

$n \in \mathbb{N}_0$	:	Total number of rounds
$l \in \mathbb{N}_0$	:	Length of final key
$\|\psi\rangle_{A_1^n Q_1^n \widetilde{E}}$	:	Initial joint state of Alice-Bob-Eve
$\mathcal{S}$	:	alphabet of raw-key registers, usually $\{0, 1\}$
S_{acc}	:	Acceptance set of accepted frequencies
λ_{EC}	:	Number of bits exchanged during error correction step
ε_{PA}	:	Security parameter contribution from privacy amplification
ε_{EV}	:	Security parameter contribution from error verification
Ω_{AT}	:	Event of passing the acceptance test
Ω_{EV}	:	Event of passing the error verification
$\Omega_{\text{acc}} = \Omega_{\text{AT}} \wedge \Omega_{\text{EV}}$	:	Event of the protocol not aborting

Protocol steps:

1. **State preparation and transmission**: Alice or Eve prepares a pure state $|\psi\rangle_{A_1^n Q_1^n \widetilde{E}}$. The systems A_1^n are given to (or kept by, depending on whether the initial state was *given* to Alice or was prepared by her) Alice, and Q_1^n to Bob.

2. For each round $i \in \{1, 2, \ldots, n\}$, Alice and Bob perform the following steps:

   (a) **Measurements**: Alice and Bob perform some POVM on the register they receive and store their outcomes in registers X_i and Y_i, respectively.

   (b) **Public announcement**: Alice and Bob make some public announcements. They also compute a register $\widehat{C}_i$ based on their announcements. In addition, Alice and Bob decide whether to label the round as a *test-round* (whose data would be used in the acceptance test step), and they set $\widehat{C}_i = \perp$ if a round is a generation-round.[4] Let I_i denote a register that contains all public communications in this step, including a copy of $\widehat{C}_i$.

   (c) **Sifting and key map**: Alice applies a sifting procedure in which, for the sifted round, she sets S_i to a fixed value in the alphabet of S_i[5], and/or uses a key mapping procedure based on her raw data X_i and the public announcements I_i, to produce a classical register S_i.

3. **Acceptance test:** Alice and Bob compute the frequency distribution $\mathbf{F}^{\text{obs}}$ of the string C_1^n for some predetermined acceptance set S_{acc}. They accept if $\mathbf{F}^{\text{obs}} \in S_{\text{acc}}$, and abort otherwise.

4. **Error correction and verification:** Alice and Bob publicly communicate λ_{EC} bits for error correction. Using those communicated bits along with his data stored in $Y_1^n I_1^n$, Bob produces a guess $\widehat{S}_1^n$ for Alice's string S_1^n. Afterwards, for the purpose of error verification, Alice sends a 2-universal hash of S_1^n with length $\lceil \log(1/\varepsilon_{\text{EV}}) \rceil$ to Bob. He then compares it to the hash of his guess and accepts if the hashes match, and aborts otherwise.

5. **Privacy amplification:** Alice and Bob randomly choose a 2-universal hash function from some family (with fixed output length l) and apply it to their strings S_1^n and $\widehat{S}_1^n$, producing final keys K_A and K_B.

[4] For the purpose of this thesis, unless otherwise stated, we consider the test-round decision to be made by Alice.

[5] Generally, in a sifting procedure, Alice and Bob "erase" parts of their data. For instance, when Alice and Bob independently choose their basis, for those rounds where they chose different bases, they "sift out" that round (i.e., set S_i to a fixed value in its alphabet). Another example would be when (for optical protocols) the detectors do not detect an incoming signal.

An example of such a protocol is given in Chapter 4, Protocol 4. We now move on to the structure of a generic prepare-and-measure protocol.

Protocol 2. Prepare-and-measure protocols

Parameters:

$n \in \mathbb{N}_0$	:	Total number of rounds
$l \in \mathbb{N}_0$	:	Length of final key
$\{\sigma\}_{i=1\cdots d_X}$	:	Initial state prepared by Alice
$\mathcal{S}$	:	alphabet of raw-key registers, usually $\{0,1\}$
S_{acc}	:	Acceptance set of accepted frequencies
λ_{EC}	:	Number of bits exchanged during error correction step
ε_{PA}	:	Security parameter contribution from privacy amplification
ε_{EV}	:	Security parameter contribution from error verification
Ω_{AT}	:	Event of passing the acceptance test
Ω_{EV}	:	Event of passing the error verification
$\Omega_{\text{acc}} = \Omega_{\text{AT}} \wedge \Omega_{\text{EV}}$	:	Event of the protocol not aborting

Protocol steps:

1. For each round $i \in \{1, 2, \ldots, n\}$, Alice and Bob perform the following steps:

   (a) **State preparation and transmission**: Alice prepares one of the d_X states $\{\sigma\}_{i=1\cdots d_X}$ according to some distribution. She then stores the label of her choice in X_i and computes a classical register C_i^A for announcement. Then, she sends the state to Bob through a quantum channel.

   (b) **Measurements**: Bob performs some POVM on the register he receives and stores his outcomes in registers Y_i.

   (c) **Public announcement**: Alice and Bob make some public announcements. They also compute a register $\widehat{C_i}$ based on their announcements. In addition, Alice and Bob decide whether to label the round as a *test-round* (whose data would be used in the acceptance test step), and they set $\widehat{C_i} = \perp$ if a round is a generation-round. Let I_i denote a register that contains all public communications in this step, including a copy of $\widehat{C_i}$.

   (d) **Sifting and key map**: Alice applies a sifting procedure in which, for the sifted round, she sets S_i to a fixed value in the alphabet of S_i, and/or uses a key mapping procedure based on her raw data X_i and the public announcements I_i, to produce a classical register S_i.

The rest of the protocol is identical to Protocol 1.

An example of prepare-and-measure protocols is given in the next chapter, Protocol 3. The steps 3–5 are essentially classical post-processing aimed at transforming the raw key into the final secret key. We highlight that, in the acceptance test step, the choice of the acceptance set is arbitrary; however, in the next chapter, we will provide a suitable explicit example. We now explain the "error correction and verification" step in more detail and defer the"privacy amplification" step to Sec. 2.5.

Error correction and verification

The purpose of this step is to ensure that Bob has the same key string S_1^n as Alice. In particular, Alice tries to send as little information as possible to Bob so that he can obtain a correct guess for Alice's string with high probability; we refer to this part as "error correction". This procedure can be seen as the task of source compression with side information [110, 113], since the goal is for Alice to send a minimal amount of information to Bob, who already possesses some side information in his registers $Y_1^n I_1^n$. The optimal rate for this task in the asymptotic limit is known to be $r_{\mathrm{EC}} \approx H(S|YI)$. However, in the finite-size regime, this rate cannot be achieved. In [104], a one-shot result was proved that provides an upper bound on the length of the error correction string, which we restate their claim for completeness.

Theorem 2.4.1. *([103, Theorem 1] Consider a cq state $\rho \in \mathrm{S}_=(CQ)$ such that Alice holds onto the classical register C and Bob holds onto the quantum register Q. Suppose Alice would like to send l bits of information to Bob such that with probability at least $1-\varepsilon$, he can produce a correct guess for C. Then, for any $\bar{\varepsilon} \in (0, \varepsilon]$, there exists a protocol that achieve this task with*

$$H_{1/2}^{\uparrow,\bar{\varepsilon}}(C|Q)_\rho \leq l \leq H_{1/2}^{\uparrow,\bar{\varepsilon}}(C|Q)_\rho + 2\log\frac{1}{\varepsilon-\bar{\varepsilon}} + 4. \tag{2.58}$$

Therefore, denoting λ_{EC} as the length of the error correction string, the above theorem provides an upper-bound on λ_{EC} as follows:

$$\lambda_{\mathrm{EC}} \leq H_{1/2}^{\uparrow,\varepsilon}(S_1^n|Y_1^n I_1^n), \tag{2.59}$$

for some ε satisfying the condition of the above theorem. We highlight that in case the protocol described in [103] is difficult to implement, other error-correction procedures can be

employed. In Chapter 3, we will use a different bound on the length of the error correction string that is shown to be practical to implement.

After error correction, to check whether the error correction procedure was successful, Alice sends a 2-universal hash of S_1^n with length $\lceil \log(1/\varepsilon_{\text{EV}}) \rceil$ to Bob, who compares this to the hash of his guess and accepts if they match; we refer to this part as "error verification". Essentially, this step ensures that if Bob's guess is incorrect, then the probability of their hashes matching (which serves as a "testing" procedure here) is bounded by ε_{EV}.

2.4.2 Source-replacement technique

The source-replacement technique is a procedure that allows us to convert a PM-QKD protocol into an equivalent EB protocol [26, 49]. To build intuition for this approach, we illustrate it with a simple example. Suppose Alice selects one of the states from $\{|\psi_j\rangle_Q\}_{j=1,\cdots,d}$, according to some probability distribution $\{p_j\}_{j=1,\cdots,d}$, and sends the chosen state to Bob through a quantum channel $\mathcal{N} \in \text{CPTP}(Q, B)$. This prepare-and-measure process is equivalent to a scenario where Alice instead prepares the entangled state $|\Psi\rangle_{AQ} = \sum_{j=1}^{d} p_j |j\rangle_A |\psi_j\rangle_Q$, and then sends only the Q register to Bob through the same channel. In this formulation, the pre-measurement state shared between Alice and Bob would be

$$\rho_{AB} = (\mathbb{I}_A \otimes \mathcal{E}) (|\Psi\rangle \langle\Psi|). \tag{2.60}$$

The procedure of Alice performing a measurement on her register A in the computational basis and obtaining the outcome j is now equivalent to selecting the state $|\psi_j\rangle$ with probability p_j, and sending it to Bob. To build an equivalence between PM and EB protocols via the source-replacement technique for security proofs, it is vital to note that this equivalent EB protocol is such that the marginal of Alice's state on A register must remain untouched, as it has never passed through the channel (see the discussion in Sec. III.A of [52] for a detailed elaboration on this requirement).

We highlight that, in the above description, for simplicity, we only considered scenarios where Alice prepares pure states. However, in the cases where she prepares a mixed state, the preparation step can be understood as follows [62]. Alice's mixed-state preparation can equivalently be thought of as her preparing a pure state and then tracing out the purifying register (this register is commonly known as the "shield system"). Given that this trace-out operation commutes with the measurements performed by the parties, one can assume, without loss of generality, that Alice is preparing a pure state, sending part of it to Bob, while keeping her register along with the shield system in her lab, and only after the measurements are done, she traces out the shield system.

2.4.3 Security definition

In a QKD protocol, the trusted parties (Alice and Bob) aim to generate a shared key string that is effectively decoupled from any potential adversary (Eve). To do so, they execute a protocol consisting of multiple local operations, classical communication over an authenticated channel, and a series of testing steps. The goal is that, at the end of the protocol, the key strings K_A and K_B held by Alice and Bob are identical and decoupled from Eve's register. Accordingly, we define the following notion of security [104, 101]. We emphasize that this definition is "composable", meaning that a key extracted from a protocol satisfying this definition can safely be used in other protocols.

Definition 2.4.1. Consider a generic QKD protocol, and let $\rho_{K_A K_B E}$ be the final state at the end of the protocol between Alice and Bob's key register, and Eve's side information register (including all public announcements made during the protocol). We define the following terms:

- The protocol is $\varepsilon^{\text{secret}}$-secret if any possible output state of the protocol satisfies
$$\frac{1}{2}\Pr[\Omega_{\text{acc}}]\left\|\rho_{K_A E|\Omega_{\text{acc}}} - \frac{\mathbb{I}_{K_A}}{|\mathcal{K}_A|}\otimes\rho_{E|\Omega_{\text{acc}}}\right\|_1 \leq \varepsilon^{\text{secret}}. \tag{2.61}$$
where Ω_{acc} is the event that the protocol accepts.
- The protocol is $\varepsilon^{\text{correct}}$-correct if any possible output state of the protocol satisfies
$$\Pr[K_A \neq K_B \wedge \text{accept}] \leq \varepsilon^{\text{correct}}. \tag{2.62}$$
- The protocol is ε^{com}-complete if, for the state produced by the honest protocol implementation (described by some given noise model), we have
$$\Pr[\text{abort}] \leq \varepsilon^{\text{com}}. \tag{2.63}$$

If the protocol is $\varepsilon^{\text{secret}}$-secret and $\varepsilon^{\text{correct}}$-correct, we refer to it as $\varepsilon^{\text{secure}}$-secure, where $\varepsilon^{\text{secure}} = \varepsilon^{\text{secret}} + \varepsilon^{\text{correct}}$.

We highlight that the reason the secrecy definition is expressed in terms of the trace norm is not merely a matter of mathematical convenience, but rather reflects a deep connection between the trace norm and the notion of indistinguishability [101]. In particular, let us, for simplicity, define $\rho_1 := \rho_{K_A E|\Omega_{\text{acc}}}$ and $\rho_2 := \frac{\mathbb{I}_{K_A}}{|\mathcal{K}A|}\otimes\rho_{E|\Omega_{\text{acc}}}$. Suppose these two states were prepared *a priori* with equal probability, and the adversary's goal is to guess which one is which. It has been shown [61] that the maximum probability with which the adversary can correctly guess the state is $\frac{1}{2}(1+\varepsilon^{\text{secret}})$. Therefore, the states are almost indistinguishable in an "operational" sense.

2.5 Security against coherent attack

Based on the adversary's capabilities, we can define two notions of security: security against collective IID attacks and coherent attacks. In a collective attack, the adversary applies an independent operation in each round of the protocol, such that the global attack is merely the tensor product of the single-round operations. In this thesis, we focus on coherent attacks, which refer to the most general form of attack that an adversary can perform.

There are multiple approaches to analyzing such security proofs, such as complementarity-based proofs [71, 121, 79, 119, 80], de Finetti theorems [104], post-selection techniques [21, 94], and entropy accumulation theorem techniques [41, 37, 90, 88].

As mentioned in the introduction, this thesis focuses solely on the entropy accumulation theorem techniques. We note that a full description of techniques other than EAT-style ones is outside the scope of this thesis. However, to motivate our choice of proof technique, we address some disadvantages of the other methods compared to EAT. Informally, complementarity-based techniques—although they often achieve better key rates—are somewhat protocol-specific, making them difficult to adapt for generic protocols. Post-selection and de Finetti techniques require a permutation-symmetry condition, which usually involves explicitly implementing a random permutation at the start of the protocol. Additionally, they have been shown to provide worse key rates in the finite-size regime compared to EAT proofs [53]. Nonetheless, as we will show in Chapter 4, using an EAT-style technique with improved statistical analysis, one can, in principle, obtain results comparable to those from the EUR approach. Given our focus on the entropy accumulation theorem in this thesis, the upcoming sections provide the necessary tools and a brief overview of the current EAT techniques.

2.5.1 Privacy amplification and Rényi Leftover-Hashing Lemma

Privacy amplification is the final step towards constructing the secure and common key. Essentially, the goal of this step is twofold: firstly, it ensures that the adversary is decoupled from the key register, and secondly, it ensures the key is uniformly distributed. One way to meet such conditions is via the Leftover-Hashing Lemma [104, 119]. Basically, it states that by sacrificing a number of bits and reducing the length of the raw key register to some fixed value, the secrecy condition can be met. Abstractly speaking, the Leftover-Hashing Lemma connects the length of the securely extractable key to an appropriate conditional entropy, thereby reducing the problem of determining the length of the securely extractable key to

bounding a suitable conditional entropy. Informally, the following lemma gives the final key length that Alice should hash to in the privacy amplification step so that the resulting key string remains secret. There are many forms of the Leftover-Hashing Lemma that provide bounds on the length of the secret key based on different definitions of conditional entropy, such as smoothed conditional min-entropy. However, in this thesis, we focus on the recent development in [42] that connects the length of the secret key to a suitable sandwiched Rényi conditional entropy. We highlight that the bound obtained via the Rényi Leftover-Hashing Lemma is tighter than those based on smoothed entropies in [104, 119].

Lemma 2.5.1 (Rényi Leftover-Hashing Lemma (Theorem 9, [42]))**.** *Let ρ_{AE} be a classical-quantum state, and let $(\mathcal{F}_{\mathcal{A}\to\mathcal{Z}}, p_f)$ be a family of 2-universal hash functions with $\mathcal{Z} = \{0,1\}^l$. Then if f is a function drawn uniformly from that family, for $\alpha \in (1,2)$ we have*

$$\frac{1}{2}\mathbb{E}_f \left\| \rho^f_{ZE} - \frac{1}{|\mathcal{Z}|}\mathbb{I}_z \otimes \rho_E \right\|_1 = \frac{1}{2} \left\| \rho_{ZEF} - \frac{1}{|\mathcal{Z}|}\mathbb{I}_z \otimes \rho_{EF} \right\|_1 \leq 2^{\frac{1-\alpha}{\alpha}\left(H^{\uparrow}_{\alpha}(A|E)_\rho - l + 2\right)}, \tag{2.64}$$

where F is the register that stores the choice of the hash function.

If we interpret the Z register as the key register and E as the adversary's side information, the above lemma states that if Alice hashes her raw key register A to a string in Z register of length l, then as long as l is somewhat smaller than $H^{\uparrow}_{\alpha}(A|E)_\rho$, ρ_{ZEF} is close in trace distance to a uniform state on Z that is independent from the adversary's side information. Comparing the above bound with the definition of security in Definition 2.4.1, we can see that the bound in Eq. (2.64) is almost the same as the security definition. However, we highlight that bounding the above trace distance via lower bounding the conditional entropy in the exponent with l is not equivalent to the security definition, and one has to take into account nuances regarding conditioning on the acceptance event, and the probability of such an event (see Sec. 3.3.1).

2.5.2 Original entropy accumulation theorem

Here, we briefly discuss the framework and main results of the original entropy accumulation theorem (EAT) [37, 41]. As stated in the introduction, the original EAT provides a powerful tool for bounding the smooth min-entropy of a large system by decomposing it into a sum of contributions from individual parts under certain conditions. To understand the conditions required for this technique, we define the notion of a quantum Markov chain [72, 59, 97].

Definition 2.5.1. (Quantum Markov chain [72]) A state $\rho_{ABC} \in \mathrm{S}_{=}(ABC)$ is a Markov chain $A \leftrightarrow C \leftrightarrow B$ if there exists families of Hilbert spaces $\{a_j\}_{j\in J}$ and $\{b_j\}_{j\in J}$ such that C satisfies the following decomposition

$$C \cong \bigoplus_{j\in J} a_j \otimes b_j, \tag{2.65}$$

with its corresponding isometry denoted by $V : C \to (\oplus_{j\in J} a_j) \otimes (\oplus_{j\in J} b_j)$, such that

$$V\rho_{ABC}V^\dagger = \bigoplus_{j\in J} p_j \rho_{Aa_{|_j}} \otimes \rho_{Bb_{|_j}}, \tag{2.66}$$

where $\{\rho_{Aa_{|_j}}\}_{j\in J}$ and $\{\rho_{Bb_{|_j}}\}_{j\in J}$ are conditional states as defined in (2.4).

There are a number of equivalent conditions for a state to satisfy a Markov chain, which we state in the following lemma [59, 97]:

Lemma 2.5.2. *Let $\rho_{ABC} \in \mathrm{S}_{=}(ABC)$ be a state such that it obeys the Markov chain condition $A \leftrightarrow C \leftrightarrow B$. The following three conditions are equivalent.*

1. $A \leftrightarrow C \leftrightarrow B$
2. $I(A;B|C)_\rho = 0$
3. $\exists\ \mathcal{R} \in \mathrm{CPTP}(C, BC)$ *s.t.* $\rho_{ABC} = (\mathrm{id}_A \otimes \mathcal{R})[\rho_{AC}]$.

Intuitively, the second relation in the above lemma states that the register B does not contain more information about A than what was already obtained from C. With these, we can state the chain rule version of the original entropy accumulation theorem [41].

Theorem 2.5.1. *(Original EAT) Consider a sequence of channels $\{\mathcal{M}_j\}_{j=1}^n$ such that for any $j \in \{1, \cdots, n\}$, we have $\mathcal{M}_j \in \mathrm{CPTP}(R_{j-1}, R_j S_j B_j)$ such that $S_1^{j-1} \leftrightarrow B_1^{j-1} E \leftrightarrow B_j$. Then, for any $\varepsilon \in (0,1)$ and any $\rho_{R_0 E} \in \mathrm{S}_{=}(R_0 E)$*

$$\begin{aligned} &H_\infty^{\uparrow,\varepsilon}(S_1^n|B_1^n E_n)_{\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1[\rho_{R_0E}]} \\ &\qquad \geq \sum_{j=1}^n \inf_{\omega \in \mathrm{S}_{=}(R_{j-1}\widetilde{R})} H(S_j|B_j\widetilde{R})_{\mathcal{M}_j[\omega]} - n(\alpha-1)\log^2(1+2d_S) - \frac{g(\varepsilon)}{\alpha-1}, \end{aligned} \tag{2.67}$$

where $g(\varepsilon) = -\log\left(1-\sqrt{1-\varepsilon^2}\right)$, $d_S = \max_j \dim(S_j)$, and $\widetilde{R}$ is a register of large enough dimension to serve as a purifying register for any of the R_j registers.

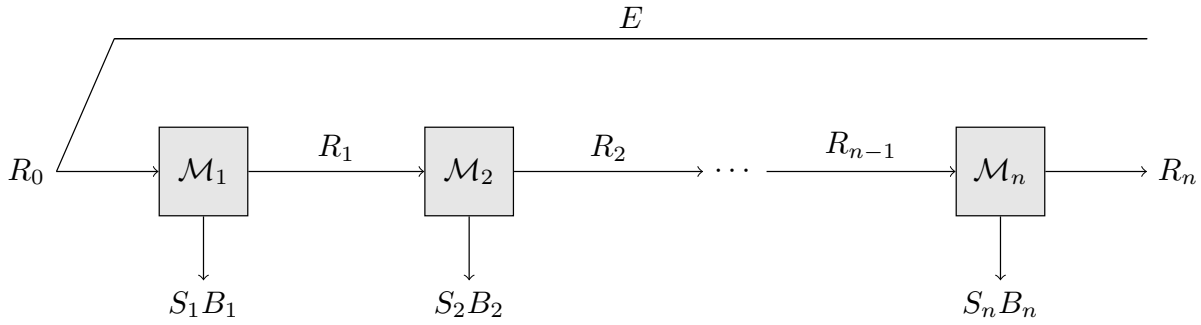


Figure 2.1: The setting of the entropy accumulation theorem.

Informally, the above theorem states that as long as the global state is generated by applying a sequence of channels $\mathcal{M}_j$ to some arbitrary state, such that in each round B_j does not get any more information from previous S_1^{j-1} registers that what is already revealed to it from $B_1^{j-1}E$. Figure 2.1 depicts a schematic representation of the setup.

When applying the original EAT to QKD protocols to obtain a bound on the amount of extractable secret key, we are typically interested in evaluating the conditional entropy of the global state conditioned on the event that the protocol is accepted. However, in such scenarios, notice that the right-hand side of the above bound contains an optimization over *all* input states. Therefore, there exist states for which the result of that optimization is zero. To obtain non-trivial values, we resort to the "EAT-with-testing" procedure. We first state the notion of EAT channels.

Definition 2.5.2. $\{\mathcal{M}_j\}_{j=1}^n$ is called a **sequence of EAT channels** if each $\mathcal{M}_j \in \mathrm{CPTP}(R_{j-1}, R_jS_jB_jC_j)$ is a channel such that the output register C_j is always classical with the same alphabet $\mathcal{C}$, and $\mathcal{M}_j$ satisfies the following conditions:

- For any initial state $\omega_{R_0E} \in \mathrm{S}_=(R_0E)$, setting $\rho \coloneqq \mathcal{M}_j \circ \cdots \circ \mathcal{M}_1[\omega_{R_0E}]$, we have $S_1^{j-1} \leftrightarrow B_1^{j-1}E \leftrightarrow B_j$.
- $\exists\, \mathcal{N}_j \in \mathrm{CPTP}(R_{j-1}, R_jS_jB_j)$ such that $\mathcal{M}_j = \mathcal{P}_j \circ \mathcal{N}_j$ with $\mathcal{P}_j \in \mathrm{CPTP}(S_jB_j, S_jB_jC_j)$ being defined as

$$\mathcal{P}_j(\omega_{S_jB_j}) = \sum_{s,t} \left(\Pi_{S_j}^{(s)} \otimes \Pi_{B_j}^{(t)}\right) \omega_{S_jB_j} \left(\Pi_{S_j}^{(s)} \otimes \Pi_{B_j}^{(t)}\right) \otimes |r(s,t)\rangle\langle r(s,t)|_{C_j}\,, \tag{2.68}$$

  where $\{\Pi_{S_j}^{(s)}\}_{s\in\mathcal{S}}$ and$\{\Pi_{B_j}^{(t)}\}_{t\in\mathcal{T}}$ are families of orthogonal projectors on S_j and B_j respectively, and $r : \mathcal{S} \times \mathcal{T} \to \mathcal{C}$ is a deterministic function. In other words, the statistics can be generated solely from the S_j and B_j systems.

We also define the notions of "rate function" and "min-tradeoff function".

Definition 2.5.3. A real function rate($\mathbf{p}$) on $\mathcal{C}$ is called a **rate function** for a sequence of EAT channels $\{\mathcal{M}_j\}_{j=1}^n$, if it satisfies:

$$\text{rate}(\mathbf{p}) \leq \inf_{\nu \in \Sigma_j(\mathbf{p})} H(S_j|B_j\widetilde{R})_\nu, \qquad \forall j \in \{1, 2, \ldots, n\} \tag{2.69}$$

where[6]

$$\Sigma_j(\mathbf{p}) \coloneqq \left\{\nu_{R_j S_j B_j C_j \widetilde{R}} = \mathcal{M}_j[\omega_{R_{j-1}\widetilde{R}}] \mid \ \omega_{R_{j-1}\widetilde{R}} \in \mathrm{S}_=(R_{j-1}\widetilde{R}) \ \wedge \ \boldsymbol{\nu}_{C_j} = \mathbf{p}\right\}. \tag{2.70}$$

Furthermore, an affine rate function is termed a **min-tradeoff function**, and it can always be expressed in the form

$$f(\mathbf{p}) = \mathbf{f} \cdot \mathbf{p} + k_f, \tag{2.71}$$

for some vector $\mathbf{f}$ (which is just the gradient of f) and scalar k_f.

With these definitions, we now state the original EAT with testing.

Theorem 2.5.2. *Let ρ be a state generated by a sequence of EAT channels $\{\mathcal{M}_j\}_{j=1}^n$ (Definition 2.5.2). Suppose furthermore that $\rho = p_\Omega \rho_{|\Omega} + (1 - p_\Omega)\rho_{|\overline{\Omega}}$ for some $p_\Omega \in (0, 1]$. Let f be a min-tradeoff function, and $h = \min_{c^n \in \Omega} f(\text{freq}(c^n))$. Then, for any $\varepsilon \in (0, 1)$ we have*

$$H_\infty^{\uparrow,\varepsilon}(S_1^n|B_1^n E)_{\rho|\Omega} \geq nh - n(\alpha - 1)\frac{\ln 2}{2}V(\mathbf{p}, f) - \frac{g(\varepsilon) + \alpha \log\left(1/\Pr[\Omega]\right)}{\alpha - 1} - n(\alpha - 1)^2 K(\alpha), \tag{2.72}$$

where

$$\begin{aligned}
g(\varepsilon) &= -\log\left(1 - \sqrt{1 - \varepsilon^2}\right), \\
\text{Var}(\mathbf{p}, f) &= \sum_x p(x) f(\delta_x)^2 - \left(\sum_x p(x) f(\delta_x)\right)^2, \\
V(\mathbf{p}, f) &= \left(\log(1 + 2d_S^2) + \sqrt{2 + \text{Var}(\mathbf{p}, f)}\right)^2 \\
K(\alpha) &= \frac{1}{6(2 - \alpha)^3 \ln 2} 2^{(\alpha-1)(2\log d_S + \max(f) - \min_{\mathcal{Q}}(f))} \ln^3\left(2^{2\log d_S + \max(f) - \min_{\mathcal{Q}}(f)} + e^2\right),
\end{aligned} \tag{2.73}$$

with $d_S = \max_j \dim(S_j)$, and δ_x being the distribution with all weight on element x.

[6] Note that if the set $\Sigma_j(\mathbf{p})$ is empty we define the infimum to be $+\infty$.

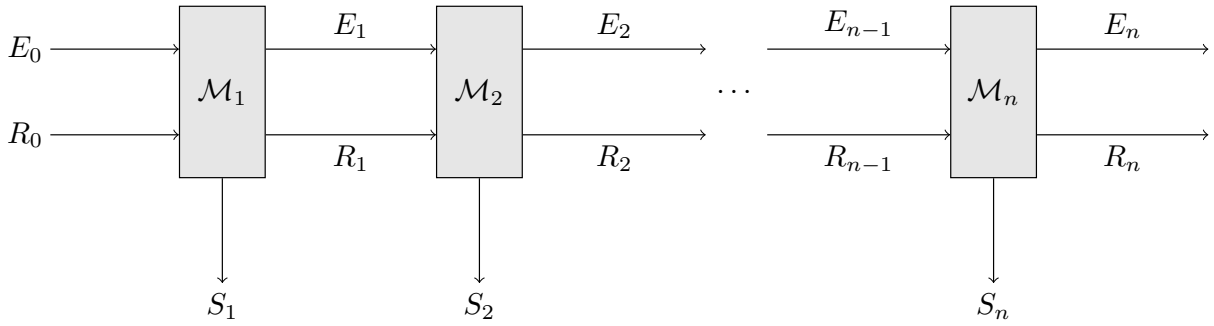


Figure 2.2: The setting of generalized entropy accumulation theorem.

2.5.3 Generalized entropy accumulation theorem

We now turn to the generalized entropy accumulation (GEAT) [89]. Essentially, the GEAT, as the name suggests, provides a generalization of the original EAT by allowing a more general structure for channels. In particular, the original EAT considers side information consisting of two parts: untouched side information E, and other sequentially generated side information B_1^n. However, as shown in Fig. 2.1, the side information was not updated during the process. This is the case because the E part remains untouched, and the B_j values are generated from some internal memory register R_{j-1} but are never updated throughout the protocol. The GEAT, however, considers a more general model for side information that allows it to be updated in each round. In addition, the GEAT also relaxes the Markov chain condition to a particular non-signaling (NS) condition, which we will explain after stating the main theorem.

Theorem 2.5.3. *(GEAT) Consider a sequence of channels $\{\mathcal{M}_j\}_{j=1}^n$ such that for any $j \in \{1, \cdots, n\}$, $\mathcal{M}_j \in \mathrm{CPTP}(R_{j-1}E_{j-1}, R_jE_jS_j)$, and furthermore*

$$\exists \text{ a channel } \mathcal{R}_j \in \mathrm{CPTP}(E_{j-1}, E_j) \text{ such that } \mathrm{Tr}_{S_jR_j} \circ \mathcal{M}_j = \mathcal{R}_j \circ \mathrm{Tr}_{R_{j-1}}. \tag{2.74}$$

Take any $\alpha \in (1,2)$ and any $\rho_{R_0E_0} \in \mathrm{S}_=(R_0E_0)$. Then, letting $\widehat{\alpha} = 1/(2-\alpha)$, we have

$$H_\alpha^\uparrow(S_1^n|E_n)_{\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1[\rho_{R_0E_0}]} \geq \sum_{j=1}^n \inf_{\omega \in \mathrm{S}_=(R_{j-1}E_{j-1}\widetilde{E})} H_{\widehat{\alpha}}(S_j|E_j\widetilde{E})_{\mathcal{M}_j[\omega]} \tag{2.75}$$

where $\widetilde{E}$ is a register of large enough dimension to serve as a purifying register for any of the R_jE_j registers.

The condition in Eq. (2.74) is commonly referred to as "one-way non-signaling". Informally, it states that the channels $\mathcal{M}_j$ should not transmit information from R_{j-1} to the E_j register. Fig. 2.2 presents a schematic representation of the GEAT setup.

Similar to the original EAT, when applied to QKD protocols, one is typically interested in evaluating the conditional entropy over a state conditioned on the event that the protocol is accepted. Thus, to obtain non-trivial bounds on the right-hand side of the above equation, we again introduce the notion of the "GEAT-with-testing" procedure.

Definition 2.5.4. $\{\mathcal{M}_j\}_{j=1}^n$ is called a **sequence of GEAT channels** if each $\mathcal{M}_j \in \mathrm{CPTP}(R_{j-1}E_{j-1}, R_jE_jS_jC_j)$ is a channel such that the output register C_j is always classical with the same alphabet $\mathcal{C}$, and $\mathcal{M}_j$ satisfies the following conditions:

- $\exists\ \mathcal{P} \in \mathrm{CPTP}(S_1^nE_n, C_1^nS_1^nE_n)$ such that $\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1 = \mathcal{P} \circ \mathcal{N}_n \circ \cdots \circ \mathcal{N}_1$ with $\mathcal{N}_j = \mathrm{Tr}_{C_j} \circ \mathcal{M}_j$, and

$$\mathcal{P}(\rho_{S_1^nE_n}) = \sum_{s\in\mathcal{S},t\in\mathcal{T}} \left(\Pi_{S_1^n}^{(s)} \otimes \Pi_{E_n}^{(t)}\right) \rho_{S_1^nE_n} \left(\Pi_{S_1^n}^{(s)} \otimes \Pi_{E_n}^{(t)}\right) \otimes |r(s,t)\rangle\langle r(s,t)|_{C_1^n}\,, \quad (2.76)$$

  where $\{\Pi_{S_1^n}^{(s)}\}_{s\in\mathcal{S}}$ and $\{\Pi_{E_n}^{(t)}\}_{t\in\mathcal{T}}$ are families of orthogonal projectors on S_1^n and E_n respectively, while $r : \mathcal{S}\times\mathcal{T} \to \mathcal{C}$ is a deterministic function.

- $\exists\ \mathcal{R}_j \in \mathrm{CPTP}(E_{j-1}, E_j)$ such that $\mathrm{Tr}_{A_jR_jC_j} \circ \mathcal{M}_j = \mathcal{R}_j \circ \mathrm{Tr}_{R_{j-1}}$. Intuitively, this condition states that the channels are "non-signaling" from the R_{j-1} registers to the E_j registers.

Similar to the original EAT, we require the notion of rate function and min-tradeoff function, which are defined similarly to Definition 2.5.3. However, for completeness, we restate these definitions, explicitly adapted to the GEAT channels.

Definition 2.5.5. A real function $\mathrm{rate}(\mathbf{p})$ on $\mathcal{C}$ is called a **rate function** for a sequence of GEAT channels $\{\mathcal{M}_j\}_{j=1}^n$, if it satisfies:

$$\mathrm{rate}(\mathbf{p}) \leq \inf_{\nu\in\Sigma_j(\mathbf{p})} H(S_j|E_j\widetilde{E})_\nu, \qquad \forall j \in \{1,2,\ldots,n\} \quad (2.77)$$

where[7]

$$\Sigma_j(\mathbf{p}) \coloneqq \left\{\nu_{R_jS_jE_jC_j\widetilde{E}} = \mathcal{M}_j[\omega_{R_{j-1}E_{j-1}\widetilde{E}}] \mid\ \omega_{R_{j-1}E_{j-1}\widetilde{E}} \in \mathrm{S}_=(R_{j-1}E_{j-1}\widetilde{E})\ \wedge\ \boldsymbol{\nu}_{C_j} = \mathbf{p}\right\}. \quad (2.78)$$

[7]Again if the set $\Sigma_j(\mathbf{p})$ is empty we define the infimum to be $+\infty$.

Similarly, an affine rate function is called a min-tradeoff function, and it can always be expressed in the form given in Eq. (2.71).

Before stating the main theorem of the GEAT with testing, for the sake of clarity and in preparation for the next chapter, we define the notions of *crossover rate function* and *crossover min-tradeoff function*. We highlight that while these concepts do not appear in the main statement of the GEAT with testing, they will be predominantly used in the next chapter.

Definition 2.5.6. Let $\{\mathcal{M}_j^{\text{gen}}\}_{j=1}^n$ and $\{\mathcal{M}_j^{\text{test}}\}_{j=1}^n$, be two sets of channels, where $\mathcal{M}_j^{\text{gen}} \in \text{CPTP}(E_{j-1}R_{j-1}, E_jR_jS_j)$ and $\mathcal{M}_j^{\text{test}} \in \text{CPTP}(E_{j-1}R_{j-1}, E_jR_jS_jC_j)$. Furthermore, let $\lambda \in [0,1]$ and $\{\mathcal{M}_j\}_{j=1}^n$ be a sequence of GEAT channels such that

$$\mathcal{M}_j = \lambda \mathcal{M}_j^{\text{test}} + (1-\lambda)\mathcal{M}_j^{\text{gen}} \otimes |\perp\rangle\langle\perp|_{C_j}, \tag{2.79}$$

A real function $r_{\text{cross}}(\mathbf{q})$ on $\mathcal{C}$ is called a **crossover rate function** for a sequence of channels $\{\mathcal{M}_j\}_{j=1}^n$ defined above, if it satisfies:

$$r_{\text{cross}}(\mathbf{q}) \leq \inf_{\nu \in \Sigma_j(\mathbf{q})} H(S_j|E_j\widetilde{E})_\nu, \qquad \forall j \in \{1,2,\ldots,n\} \tag{2.80}$$

where

$$\Sigma_j(\mathbf{q}) \coloneqq \left\{\nu_{R_jS_jE_jC_j\widetilde{E}} = \mathcal{M}_j^{\text{test}}[\omega_{R_{j-1}E_{j-1}\widetilde{E}}] \mid \omega_{R_{j-1}E_{j-1}\widetilde{E}} \in \text{S}_=(R_{j-1}E_{j-1}\widetilde{E}) \ \wedge \ \boldsymbol{\nu}_{C_j} = \mathbf{q}\right\}. \tag{2.81}$$

Furthermore, an affine crossover rate function is termed a **crossover min-tradeoff function**, and it can always be expressed in the form

$$g(\mathbf{q}) = \mathbf{g}\cdot\mathbf{q} + k_g, \tag{2.82}$$

for some vector $\mathbf{g}$ (which is just the gradient of g) and scalar k_g.[8]

Theorem 2.5.4. *Let ρ be a state generated by a sequence of GEAT channels $\{\mathcal{M}_j\}_{j=1}^n$ (Definition 2.5.4). Suppose furthermore that $\rho = p_\Omega \rho_{|\Omega} + (1-p_\Omega)\rho_{|\overline{\Omega}}$ for some $p_\Omega \in (0,1]$. Let f be a min-tradeoff function (Definition 2.5.5), and $h = \min_{c^n\in\Omega} f(\text{freq}(c^n))$. Then, for any $\varepsilon \in (0,1)$, $\alpha \in (1,3/2)$ we have*

$$H_\infty^{\uparrow,\varepsilon}(S_1^n|B_1^nE)_{\rho_{|\Omega}} \geq nh - n\frac{\alpha-1}{2-\alpha}\frac{\ln 2}{2}V(\mathbf{p},f) - \frac{g(\varepsilon)+\alpha\log\left(1/\Pr[\Omega]\right)}{\alpha-1} - n\left(\frac{\alpha-1}{2-\alpha}\right)^2 K(\alpha), \tag{2.83}$$

[8] We note that these notions can also be defined similarly for EAT channels.

where

$$g(\varepsilon) = -\log\left(1-\sqrt{1-\varepsilon^2}\right),$$

$$\mathrm{Var}(\mathbf{p}, f) = \sum_x p(x) f(\delta_x)^2 - \left(\sum_x p(x) f(\delta_x)\right)^2,$$

$$V(\mathbf{p}, f) = \left(\log(1+2d_S^2) + \sqrt{2+\mathrm{Var}(\mathbf{p}, f)}\right)^2$$

$$K(\alpha) = \frac{(2-\alpha)^3}{6(3-2\alpha)^3 \ln 2} 2^{\frac{\alpha-1}{2-\alpha}(2\log d_S + \max(f) - \min_{\mathcal{Q}}(f))} \ln^3\left(2^{2\log d_S + \max(f) - \min_{\mathcal{Q}}(f)} + e^2\right), \quad (2.84)$$

with $d_S = \max_j \dim(S_j)$*, and* δ_x *being the distribution with all weight on element* x*.*

Chapter 3

Finite-size analysis of prepare-and-measure and decoy-state QKD via generalized entropy accumulation

3.1 Introduction

An important question in QKD is how to bound the amount of extractable key in a protocol such that it remains secure against the most general class of attacks—coherent attacks—by an adversary. As pointed out in the previous chapter, a key development in addressing this has been the introduction of the original entropy accumulation theorem (EAT) [41, 37], which provides a framework for deriving such bounds. The original EAT was formulated for device-independent QKD and device-dependent entanglement-based (EB) QKD protocols. However, applying it to prepare-and-measure (PM) QKD protocols remained an open challenge until recently, when a generalized version—GEAT—was introduced [88, 90], enabling security proofs for PM protocols under coherent attacks.

Previous works based on the GEAT mainly provided proof-of-principle demonstrations of such security proofs, without much focus on optimizing the parameter choices to maximize the finite-size key rates in practical applications—in particular, the GEAT relies on the construction of the *min-tradeoff function*, which is challenging to construct optimally for protocols with large numbers of signal states or measurement outcomes. In this chapter,

we provide techniques to overcome this difficulty and incorporate multiple improvements for finite-size effects.

Furthermore, when considering PM protocols of practical interest, decoy-state protocols are an important family of such protocols. More specifically, in practical QKD implementations, the source typically uses a highly attenuated laser that does not necessarily emit single-photon pulses. There is a non-trivial probability that these pulses contain more than one photon. This can introduce a security vulnerability, as Eve could exploit these multiphoton pulses to perform a photon-number splitting (PNS) attack [14, 19]. To mitigate the effects of such attacks, decoy-state protocols were introduced [63, 82, 129]. In this approach, Alice randomly selects the intensity of each pulse from a predefined set before sending it to Bob. Since Eve cannot distinguish the signals based on their intensities, Alice and Bob can design an acceptance test that aims to detect PNS attacks. Thus, the decoy-state method enhances the protocol's resilience to such threats.

We provide a flexible framework to analyze the finite-size effects of PM protocols via the GEAT. We apply techniques similar to those in [53], with improvements and modifications to address challenges specific to PM protocols. Furthermore, we incorporate the refined finite-size analysis from [81] into our framework, resulting in improved key rates. However, the enhancements introduced in that work involve an additional optimization problem, which makes key rate computations more time-consuming. We address this issue by demonstrating that this optimization can be integrated with the numerical methods used for bounding single-round entropies. Moreover, we develop a method to automatically optimize the choice of the min-tradeoff function, using techniques similar to those in [53].[1]

We highlight that there is a technical condition in the GEAT framework which restricts the adversary to interacting with only a single signal at a time,[2] thereby limiting the repetition rate of the protocol. However, this restriction can be lifted using the techniques introduced in Chapter 5. In fact, in that chapter, we prove that the key rates derived in this chapter remain valid under the more general model introduced there, which does not impose any repetition-rate constraint.

This chapter is based on [65].

[1] Note that we do not obtain a closed-form expression for the optimal min-tradeoff function here. However, in the next chapter, we introduce a different approach to study the statistical analysis that allows us to obtain a bound independent of any min-tradeoff function, thus avoiding the problem of optimizing it.

[2] Note that Eve may still act in an arbitrary non-IID manner.

3.2 Protocol description

Here, we first state the generic structure of the protocols we analyze:[3]

Protocol 3. Prepare-and-Measure Protocol.
Parameters:

$n \in \mathbb{N}_0$	:	Total number of rounds
$l \in \mathbb{N}_0$	:	Length of final key
$\{\sigma_i\}_{i=1\ldots d_X}$	:	States sent by Alice
$\mathcal{S}$	:	alphabet of raw-key registers, usually $\{0,1\}$
$\mathcal{C}$	:	alphabet of acceptance test decision register
M_k^B	:	POVM elements acting on Bob's system describing his measurement outcomes k
γ	:	Probability that Alice chooses a round to be a test round
S_{acc}	:	Acceptance set of accepted frequencies
λ_{EC}	:	Number of bits exchanged during error correction step
ε_{PA}	:	Security parameter contribution from privacy amplification
ε_{EV}	:	Security parameter contribution from error verification
Ω_{AT}	:	Event of passing the acceptance test
Ω_{EV}	:	Event of passing the error verification
$\Omega_{\text{acc}} = \Omega_{\text{AT}} \wedge \Omega_{\text{EV}}$	:	Event of the protocol not aborting

Protocol steps:

1. For each round $i \in \{1, 2, \ldots, n\}$, Alice and Bob perform the following steps:

   (a) **State preparation and transmission:** Alice, with probability γ $(1-\gamma)$, in each round chooses to be a **test** (**generation**) round. She then prepares a state from $\{\sigma_k\}_{k=1\ldots d_X}$, according to some distribution [4]. Alice stores the label for her choice of signal state in a classical register X_i, and computes a classical register C_i^A for public announcement. She then sends the signal state to Bob through a quantum channel.

   (b) **Measurements:** Bob performs a measurement on the received states using a POVM with elements $\{M_k^B\}_{k=1\ldots d_B}$, recording the outcomes in a classical register

[3] In this chapter, we only focus on protocols that make an accept/abort decision and produce a key of fixed length when they accept

[4] The distribution could depend on the choice of a test or generation round.

Y_i with alphabet $\mathcal{Y}$. Furthermore, he computes a classical register C_i^B for public announcement.

(c) **Public announcement:** Alice and Bob announce their values $C_i^A C_i^B$, and jointly compute a pair of values (C_i, I_i) as a function of their announcements, i.e., $(C_i, I_i) = f(C_i^A, C_i^B)$, where f is such that the resulting I_i contains all the public communication. Moreover, C_i is a value that will be used later in the acceptance test to decide whether to abort, and we use the convention that it is set to a fixed symbol $\perp$ whenever Alice chooses the round to be a generation round.[5]

(d) **Sifting and key map:** Alice applies a sifting procedure in which, for the sifted round, she sets S_i to a fixed value in the alphabet of S_i, and/or uses a key mapping procedure based on her raw data X_i and the public announcements I_i, to produce a classical register S_i.

2. **Acceptance test (parameter estimation):** Alice and Bob compute the frequency distribution $\mathbf{F}^{\text{obs}}$ of the string C_1^n. They accept if $\mathbf{F}^{\text{obs}} \in S_{\text{acc}}$, and abort otherwise. Note that S_{acc} is the predetermined acceptance set.

3. **Error correction and verification:** Alice and Bob publicly communicate λ_{EC} bits for error correction. Using those communicated bits along with his data stored in $Y_1^n I_1^n$, Bob produces a guess $\widehat{S_1^n}$ for Alice's string S_1^n. Afterwards, for the purpose of error verification, Alice sends a 2-universal hash of S_1^n with length $\lceil \log(1/\varepsilon_{\text{EV}}) \rceil$ to Bob. He then compares it to the hash of his guess and accepts if the hashes match, and aborts otherwise.

4. **Privacy amplification:** Alice and Bob randomly choose a 2-universal hash function from some family (with fixed output length l) and apply it to their strings S_1^n and $\widehat{S_1^n}$, producing final keys K_A and K_B.

It is worth noting that steps 1c–1d do not have to take place in individual rounds. In fact, they can be postponed until all the measurements are completed, as long as the $C_i I_i S_i$ are generated solely based on the public information available in their respective rounds. This is because they commute them with the measurements from earlier rounds.

[5] Note that here we allow Bob to independently decide whether to measure a round on a generation or testing basis, but this decision is independent of Alice's choice, and the test/generation labeling is entirely determined by Alice.

3.3 Improved entropy accumulation analysis for PM protocols

In this section, we apply the generalized entropy accumulation theorem from [89], along with the improvements from [81], to bound the secure key length of a PM QKD protocol. We begin by stating the improved version of the GEAT:[6]

Theorem 3.3.1 ([89] Theorem 4.3 and [81] Supplement, Theorem 2)**.** *For a sequence of GEAT channels $\{\mathcal{M}_i\}_{i\in\{1,2,\dots,n\}}$ defined in Definition 2.5.4, let $\varepsilon \in (0,1)$, $\alpha \in (1,3/2)$, let Ω be an event on registers C_1^n, let $\rho_{R_0E_0} \in \mathcal{D}(R_0E_0)$, let f be a min-tradeoff function (Definition 2.5.5), and let $h = \min_{c^n\in\Omega} f(\mathrm{freq}(c^n))$. Then:*

$$H_\alpha^\uparrow(S^n|E_n)_{\mathcal{M}_n\circ\cdots\circ\mathcal{M}_1(\rho_{R_0E_0})_{|\Omega}} \geq nh + nT_\alpha(f) - \frac{\alpha}{\alpha-1}\log\frac{1}{\Pr[\Omega]} - n\left(\frac{\alpha-1}{2-\alpha}\right)^2 K(\alpha), \quad (3.1)$$

where (letting $\kappa = 1$ if the A_i systems are classical and $\kappa = 2$ if they are quantum):

$$\begin{aligned}
T_\alpha(f) &= \inf_{\mathbf{p}\in\mathcal{Q}}\left(\mathrm{rate}(\mathbf{p}) - f(\mathbf{p}) - \frac{\alpha-1}{2-\alpha}\frac{\ln(2)}{2}V(\mathbf{p},f)\right)\\
\mathrm{Var}(\mathbf{p},f) &= \sum_x p(x)f(\delta_x)^2 - \left(\sum_x p(x)f(\delta_x)\right)^2\\
V(\mathbf{p},f) &= \left(\log(1+2d_A^\kappa) + \sqrt{2+\mathrm{Var}(\mathbf{p},f)}\right)^2\\
K(\alpha) &= \frac{(2-\alpha)^3}{6(3-2\alpha)^3\ln 2}2^{\frac{\alpha-1}{2-\alpha}(\kappa\log d_A+\max(f)-\min_{\mathcal{Q}}(f))}\ln^3\left(2^{\kappa\log d_A+\max(f)-\min_{\mathcal{Q}}(f)}+e^2\right), \quad (3.2)
\end{aligned}$$

with $d_A = \max_i \dim(A_i)$, δ_x being the distribution with all weight on element x, and $\mathcal{Q}$ being the set of all distributions on $\mathcal{C}$ that could be produced by a GEAT channel $\mathcal{M}_i$.

We now proceed to prove the security of Protocol 3, following the definition provided in Definition 2.4.1.

3.3.1 Security

Using the framework of GEAT instead of the original EAT allows us to include the test-round announcements in the conditioning register. Moreover, by exploiting the Rényi

[6]We highlight that the bound here is slightly different from the one in Theorem 2.5.4, as we used some recent improvements from [81].

Leftover-Hashing Lemma in 2.5.1, we obtain a significantly simplified formula for the secrecy parameter that involves only a single parameter, $\varepsilon_{\mathrm{PA}}$. Informally speaking, this parameter represents the cost of privacy amplification in the key-length formula.

Theorem 3.3.2. *For any $\alpha \in (1, 3/2)$, $\varepsilon_{\mathrm{PA}}, \varepsilon_{\mathrm{EV}} \in (0, 1]$, Protocol 3 is $\varepsilon_{\mathrm{PA}}$-secret, $\varepsilon_{\mathrm{EV}}$-correct, and hence $(\varepsilon_{\mathrm{PA}} + \varepsilon_{\mathrm{EV}})$-secure, when the length l of the final key satisfies*

$$l \leq nh + nT_\alpha(f) - n\left(\frac{\alpha-1}{2-\alpha}\right)^2 K(\alpha) - \lambda_{\mathrm{EC}} - \left\lceil \log \frac{1}{\varepsilon_{\mathrm{EV}}} \right\rceil - \frac{\alpha}{\alpha-1} \log \frac{1}{\varepsilon_{\mathrm{PA}}} + 2, \tag{3.3}$$

where $h, T_\alpha(f)$, and $K(\alpha)$ are as defined in Theorem 3.3.1, and λ_{EC} is the length of the error-correction string.

Proof. Proving correctness is straightforward: let us denote the event of Bob correctly guessing Alice's string by Ω_g, then the expression in Eq. (2.62) can be written as:

$$\begin{aligned} \Pr[K_A \neq K_B \wedge \Omega_{\mathrm{acc}}] &\leq \Pr[K_A \neq K_B \wedge \Omega_{\mathrm{EV}}] \\ &\leq \Pr\left[\Omega_g^c \wedge \Omega_{\mathrm{EV}}\right] \leq \Pr\left[\Omega_{\mathrm{EV}} | \Omega_g^c\right] \leq 2^{-\left\lceil \log \frac{1}{\varepsilon_{\mathrm{EV}}} \right\rceil} \leq \varepsilon_{\mathrm{EV}}, \end{aligned} \tag{3.4}$$

where Ω_g^c is the complementary event to Ω_g^c. Thus, the protocol has a correctness parameter $\varepsilon^{\mathrm{correct}} = \varepsilon_{\mathrm{EV}}$.

To prove secrecy, let us denote the final state at the end of the QKD protocol by $\rho_{K_A K_B I^n E_n' L}$, where K_A, K_B are Alice and Bob's key registers respectively, I^n denotes the public communication registers over all rounds, E_n' is Eve's quantum side information register, and L is the register containing the error correction and error verification information. Note that the event in which the protocol does not abort is $\Omega_{\mathrm{acc}} = \Omega_{\mathrm{EV}} \wedge \Omega_{\mathrm{AT}}$, where Ω_{EV} is the event where the error verification succeeds, and Ω_{AT} is the event of the acceptance test passing. Then, the secrecy condition in Eq. (2.61) can be written as

$$\frac{1}{2} \Pr[\Omega_{\mathrm{EV}} \wedge \Omega_{\mathrm{AT}}] \left\| \rho_{K_A I^n E_n' L | \Omega_{\mathrm{acc}}} - \frac{\mathbb{I}_{K_A}}{|\mathcal{K}_A|} \otimes \rho_{I^n E_n' L | \Omega_{\mathrm{acc}}} \right\|_1 \leq \varepsilon^{\mathrm{secret}}. \tag{3.5}$$

To show the above bound holds, we first apply Theorem 2.5.1 to find:

$$\frac{1}{2} \left\| \rho_{K_A I^n E_n' L | \Omega_{\mathrm{acc}}} - \frac{\mathbb{I}_{K_A}}{|\mathcal{K}_A|} \otimes \rho_{I^n E_n' L | \Omega_{\mathrm{acc}}} \right\|_1 \leq 2^{\frac{2(1-\alpha)}{\alpha}} 2^{\frac{1-\alpha}{\alpha}\left(H_\alpha^\uparrow(S^n | I^n E_n' L)_{\rho | \Omega_{\mathrm{acc}}} - l\right)}, \tag{3.6}$$

where S^n is the key register before applying the hash function. On the other hand, the entropic quantity in the exponent of the RHS of the above inequality can be bounded as follows:

$$\begin{aligned} H_\alpha^\uparrow(S^n|I^nE_n'L)_{\rho_{|\Omega_{\text{acc}}}} &\geq H_\alpha^\uparrow(S^n|I^nE_n'L)_{\rho_{|\Omega_{\text{AT}}}} + \frac{\alpha}{1-\alpha}\log\frac{1}{\Pr[\Omega_{\text{EV}}|\Omega_{\text{AT}}]} \\ &\geq H_\alpha^\uparrow(S^n|I^nE_n')_{\rho_{|\Omega_{\text{AT}}}} + \frac{\alpha}{1-\alpha}\log\frac{1}{\Pr[\Omega_{\text{EV}}|\Omega_{\text{AT}}]} - \lambda_{\text{EC}} - \lceil\log(1/\varepsilon_{\text{EV}})\rceil\,, \end{aligned} \tag{3.7}$$

where in the first line we used Lemma B.5 in [41], and in the second line we used Lemma 6.8 in [117]. Note that the λ_{EC} term arises from the length of the error-correction string, while the $\lceil\log(1/\varepsilon_{\text{EV}})\rceil$ term arises from the length of the error-verification hash. It remains to bound the first term in Eq. (3.7). To do so, we first note that by setting $E_i \equiv E_i' I_1 \cdots I_i$ in Theorem 3.3.1, we can apply that theorem to lower bound $H_\alpha^\uparrow(S^n|E_n)_{\rho_{|\Omega_{\text{AT}}}}$ via Eq. (3.1), modeling the protocol with a sequence of GEAT channels as shown in [90, Claim 10]. Combining this with Eq. (3.6)–(3.7), we can write

$$\begin{aligned} &\frac{1}{2}\Pr[\Omega_{\text{EV}}\wedge\Omega_{\text{AT}}]\left\|\rho_{KI^nE_n'L_{|\Omega_{\text{acc}}}} - \frac{\mathbb{I}_{K_A}}{|\mathcal{K}_A|}\otimes\rho_{I^nE_n'L_{|\Omega_{\text{acc}}}}\right\|_1 \\ =&\frac{1}{2}\Pr[\Omega_{\text{EV}}|\Omega_{\text{AT}}]\Pr[\Omega_{\text{AT}}]\left\|\rho_{KI^nE_n'L_{|\Omega_{\text{acc}}}} - \frac{\mathbb{I}_{K_A}}{|\mathcal{K}_A|}\otimes\rho_{I^nE_n'L_{|\Omega_{\text{acc}}}}\right\|_1 \\ \leq&2^{\frac{1-\alpha}{\alpha}\left(2+nh+T_\alpha(f)-n\left(\frac{\alpha-1}{2-\alpha}\right)^2K(\alpha)-\lambda_{\text{EC}}-l-\lceil\log(1/\varepsilon_{\text{EV}})\rceil\right)}, \end{aligned} \tag{3.8}$$

Comparing this to the secrecy definition in Eq. (3.5), we conclude that the protocol is ε_{PA}-secret as long as the key length satisfies the formula (3.3).

Combining Eq. (3.3) and Eq. (3.4), we obtain the statements of the theorem. □

It is worth noting that the formula given in Eq. (3.3) converges to the Devetak–Winter formula [34] in the asymptotic regime by taking $\alpha = 1 + O(1/\sqrt{n})$ [37, Theorem V.2]. However, we found that in the finite-size regime, this may not be the optimal choice of α, and thus it is best to optimize this parameter for each n numerically.

We now turn to the proof of completeness for Protocol 3.

3.3.2 Completeness

The completeness parameter is simply an upper bound on the probability of an accidental abort in the honest case[7]. In the description of Protocol 3, there are only two steps where the protocol may abort: the acceptance test (Step 2) and the error verification (Step 3). Therefore, by the union bound, it suffices to choose:

$$\varepsilon^{\text{com}} = \varepsilon^{\text{com}}_{\text{AT}} + \varepsilon^{\text{com}}_{\text{EV}}, \tag{3.9}$$

where $\varepsilon^{\text{com}}_{\text{AT}}$ and $\varepsilon^{\text{com}}_{\text{EV}}$ are upper bounds on the probability of an honest-behavior abort in the acceptance test and error verification steps, respectively. For the purpose of calculations in this chapter, we take $\varepsilon^{\text{com}} = 10^{-3}$.[8] We begin by analyzing the first term, $\varepsilon^{\text{com}}_{\text{AT}}$. For the purposes of the calculations in this chapter and to compute explicit bounds on $\varepsilon^{\text{com}}_{\text{AT}}$, we consider the following structure of the acceptance set S_{acc}: the acceptance set is defined by entry-wise constraints with respect to the honest behavior. In other words, letting $\mathbb{P}_{\mathcal{C}}$ be the probability distribution on $\mathcal{C}$, we have:

$$S_{\text{acc}} = \left\{ \mathbf{p}^{\text{acc}} \in \mathbb{P}_{\mathcal{C}} \;\middle|\; \forall c \in \mathcal{C},\ p^{\text{hon}}_c - t^{\text{low}}_c \leq p^{\text{acc}}_c \leq p^{\text{hon}}_c + t^{\text{upp}}_c \right\}, \tag{3.10}$$

where $\mathbf{t}^{\text{low}}, \mathbf{t}^{\text{upp}} \in \mathbb{R}^{|\mathcal{C}|}_{\geq 0}$, and $\mathbf{p}^{\text{hon}}$ denotes the distribution produced on a single-round register C_j by the honest behavior. With this choice of acceptance set, let us consider a specific value $c \in \mathcal{C}$. Since the honest behavior is IID, the frequency of outcome c follows a binomial distribution with success probability p^{hon}_c. Therefore, the probability that the frequency of c lies outside the acceptance interval for this specific c is:

$$\Pr_{X \sim \text{Binom}\left(n, p^{\text{hon}}_c\right)} \left[\frac{X}{n} \notin [p^{\text{hon}}_c - t^{\text{low}}_c, p^{\text{hon}}_c + t^{\text{upp}}_c] \right], \tag{3.11}$$

where $X \sim \text{Binom}(n, p)$ denotes a random variable X following a binomial distribution with parameters (n, p). Thus, employing union bound, we have:

$$\varepsilon^{\text{com}}_{\text{AT}} \geq \sum_c \Pr_{X \sim \text{Binom}\left(n, p^{\text{hon}}_c\right)} \left[\frac{X}{n} \notin [p^{\text{hon}}_c - t^{\text{low}}_c, p^{\text{hon}}_c + t^{\text{upp}}_c] \right]. \tag{3.12}$$

In other words, $\varepsilon^{\text{com}}_{\text{AT}}$ can be any value that satisfies the above relation. For the purposes of this chapter, we set $\varepsilon^{\text{com}}_{\text{AT}} = 10^{-3}$ and optimize $\mathbf{t}^{\text{low}}, \mathbf{t}^{\text{upp}}$ while ensuring the above bound is satisfied.

[7] In the most extreme case, a protocol where Alice and Bob always abort is technically secure but never useful. The completeness parameter essentially addresses such scenarios.

[8] Note that the choice of completeness parameter does not affect the secrecy, as this parameter does not appear in the secret key-length calculations

Remark 3.3.1. *Given the freedom in choosing the parameters* $\mathbf{t}^{\text{low}}, \mathbf{t}^{\text{upp}}$*, one can set them both to zero. This would correspond to a particular form of the acceptance set, commonly referred to as unique acceptance, in which case* $S_{\text{acc}} = \{\mathbf{p}^{\text{hon}}\}$*. While this choice has been used in some previous works to simplify the analysis, it is not ideal in practice. When the protocol is executed, if the observed frequency distribution lies outside the acceptance set, Alice and Bob abort the protocol. Therefore, to ensure that the protocol does not abort under honest behavior, the acceptance set must be chosen carefully. Unique acceptance is particularly problematic in the finite-size regime, where statistical fluctuations typically cause the frequency distribution to deviate from its ideal value, resulting in the protocol almost always aborting.*

We now move on to bounding the $\varepsilon_{\text{EV}}^{\text{com}}$ term. Observe that the error-verification step effectively serves as a flag to check whether the error correction procedure has succeeded. Therefore, $\varepsilon_{\text{EV}}^{\text{com}}$ is upper bounded by the probability that the error correction procedure fails (i.e., produces an incorrect guess of Alice's string). This corresponds to the task of source compression with side information, where Bob's side information is contained in the $Y_1^n I_1^n$ registers [110, 113].

As shown in [104, Corollary 6.3.5], in the IID asymptotic regime, an achievable length for this task is given by $\lambda_{\text{EC}} \approx nH(S|YI)_{\text{hon}}$, where $H(S|YI)_{\text{hon}}$ denotes the single-round conditional entropy under honest behavior. However, in the finite-size regime, this characterization does not accurately reflect the correct length. Instead, in [120, Corollary 2](see also [60]), the length of the error-correction string is set to:

$$\lambda_{\text{EC}} = nf_{\text{EC}}H(S|YI)_{\text{hon}}, \tag{3.13}$$

where $f_{\text{EC}} > 1$ is an "error-correction efficiency" parameter that quantifies how much it differs from the asymptotic value. Heuristically, we take $f_{\text{EC}} = 1.16$ as a reasonable choice corresponding to the $\varepsilon_{\text{EV}}^{\text{com}} = 10^{-3}$ used here.

3.3.3 The rate function

In this part, we describe the rate function for Protocol 3. In each round in that protocol, by the source-replacement technique, Alice's preparation process can be described by the following equivalent process: with probability γ, she prepares a test-round state $|\xi^t\rangle_{AA'}$ and sends out the register A' through a quantum channel $\mathcal{E} \in \text{CPTP}(A', B)$ controlled by Eve, and otherwise she prepares a generation-round state $|\xi^g\rangle_{AA'}$, followed by Eve's attack. The

resulting state after Eve's attack in the test and generation rounds then becomes:

$$\begin{aligned}\rho^t_{\mathcal{E}} &\coloneqq (\mathrm{id}_A \otimes \mathcal{E})\left(\left|\xi^t\right\rangle\!\left\langle\xi^t\right|_{AA'}\right), \\ \rho^g_{\mathcal{E}} &\coloneqq (\mathrm{id}_A \otimes \mathcal{E})\left(\left|\xi^g\right\rangle\!\left\langle\xi^g\right|_{AA'}\right).\end{aligned} \tag{3.14}$$

Afterwards, Bob performs measurements on the B register, followed by some public announcements made by Alice and Bob, denoted by I. Finally, Alice uses the public announcements to process her raw measurement outcome into a value S.

Let $W(\rho^g_{\mathcal{E}})$ denote the conditional entropy $H(S|IE)$ of the state at the end of the procedure described above, where E is the purification register held by the adversary. Note that since Alice announces her outcome in test rounds, we have $H(S|IE) = 0$ for test rounds, and thus, it suffices to consider only generation rounds.

Moreover, even though the state $\rho^g_{\mathcal{E}} \in \mathrm{S}_=(AB)$ does not have any component in the E register, any arbitrary purification of it would suffice, since, by the isometric equivalence of purifications and the isometric invariance of conditional entropy, the value of $H(S|IE)$ can be computed using only the state $\rho^g_{\mathcal{E}}$. Additionally, as shown in [132], $W(\rho^g_{\mathcal{E}})$ is a convex function.

Going back to the test round data, let $\Phi_c[\rho^t_{\mathcal{E}}]$ denote the conditional probability of obtaining the value c on the C register when measuring $\rho^t_{\mathcal{E}}$. Then, for protocol 3, we have:

$$\Pr\left[C = c \wedge \text{measuring } \rho^g_{\mathcal{E}}\right] = \begin{cases}\gamma\Phi_c[\rho^t_{\mathcal{E}}] & \text{if } c \neq \perp \\ 1-\gamma & \text{if } c = \perp\end{cases} \tag{3.15}$$

For brevity, we denote $\mathcal{C}_{\setminus\perp}$ as the alphabet of C excluding the element $\perp$. Similarly, for any probability distribution $\mathbf{p}$, we denote $\mathbf{p}_{\setminus\perp}$ as the components of $\mathbf{p}$ with $p_\perp$ omitted. Furthermore, we indicate the values of tuple $\left\{\Phi_c[\rho^t_{\mathcal{E}}] \,\middle|\, c \in \mathcal{C}_{\setminus\perp}\right\}$ as $\mathbf{\Phi}[\rho^t_{\mathcal{E}}]$[9]. With these notations, a valid choice of rate function is given by the following optimization problem:

$$\begin{aligned}\mathrm{rate}(\mathbf{p}) \coloneqq \inf_{\mathcal{E}} \quad & W(\rho^g_{\mathcal{E}}) \\ \text{s.t.} \quad & \gamma\mathbf{\Phi}[\rho^t_{\mathcal{E}}] = \mathbf{p}_{\setminus\perp},\end{aligned} \tag{3.16}$$

where $\rho^t_{\mathcal{E}}, \rho^g_{\mathcal{E}}$ are defined in terms of the channel $\mathcal{E}$ via (3.14). We highlight that for any distribution $\mathbf{p}$ such that $p_\perp \neq 1-\gamma$, the optimization is infeasible by definition, and thus we have $\mathrm{rate}(\mathbf{p}) = +\infty$. Note that any affine lower bound on the rate function defined in Eq. (3.16) is a valid min-tradeoff function. This optimization can alternatively be rewritten

[9]Note that having $\mathcal{C}_{\setminus\perp}$ here causes no issues, since the definition of $\Phi_c[\rho^t_{\mathcal{E}}]$ excludes $c = \perp$.

in terms of the Choi state of the channel $\mathcal{E}$. In particular, the Choi state J is characterized by the following relation:

$$\mathcal{E}(X) = \mathrm{Tr}_{A'}\left[J\left(X^T \otimes \mathbb{I}_B\right)\right]. \tag{3.17}$$

Thus, Eq. (3.16) can be written as

$$\begin{aligned} \mathrm{rate}(\mathbf{p}) = \inf_{J} \quad & W(\rho_J^g) \\ \text{s.t.} \quad & \gamma \mathbf{\Phi}[\rho_J^t] = \mathbf{p}_{\setminus\perp}, \end{aligned} \tag{3.18}$$

where

$$\begin{aligned} \rho_J^g &:= \mathrm{Tr}_{A'}\left[(\mathbb{I}_A \otimes J)\left(\left|\xi^g\right\rangle\!\left\langle\xi^g\right|^{T_{A'}} \otimes \mathbb{I}_B\right)\right], \\ \rho_J^t &:= \mathrm{Tr}_{A'}\left[(\mathbb{I}_A \otimes J)\left(\left|\xi^t\right\rangle\!\left\langle\xi^t\right|^{T_{A'}} \otimes \mathbb{I}_B\right)\right]. \end{aligned} \tag{3.19}$$

Furthermore, the $\Phi_c[\rho_J^t]$ term can also be written in terms of Choi state J as

$$\Phi_c[\rho_J^t] = \mathrm{Tr}\left[\left(M_a^A \otimes J\right)\left(\left|\xi^t\right\rangle\!\left\langle\xi^t\right|_{AA'}^{T_{A'}} \otimes M_b^B\right)\right], \tag{3.20}$$

where $c = (a, b)$. It is worth noting that the above optimization is a convex optimization problem, since the objective function is convex with respect to its input, and the constraints are affine functions with respect to their inputs.

To improve numerical stability in our calculations, we take into account that the channels under consideration are infrequent-sampling channels, as defined in [37]. In other words, with probability γ, they perform a test round and store some value in the C register; otherwise, they set the C register to $\perp$. These channels satisfy the conditions in Definition 2.5.6, and thus one can construct a crossover rate function and a crossover min-tradeoff function for these channels. As stated in Definition 2.5.6, informally, each of these notions bounds the single-round conditional entropy in terms of the distribution conditioned on being a test round and then crosses over the resulting entropy bound into a function of the full distribution (see also [37, 41] for a more detailed discussion). The following function is a crossover rate function for our setting:

$$\begin{aligned} r_{\mathrm{cross}}(\mathbf{q}) := \inf_{J} \quad & W(\rho_J^g) \\ \text{s.t.} \quad & \mathbf{\Phi}[\rho_J^t] = \mathbf{q}, \end{aligned} \tag{3.21}$$

where $\mathbf{q}$ is a distribution on $\mathcal{C}_{\setminus\perp}$. Note that for any feasible solution $\mathbf{p}$ in Eq. (3.18), we can write it as $\mathbf{p} = (\gamma\mathbf{q}, 1-\gamma)$ for some distribution $\mathbf{q}$, and thus we have $\mathrm{rate}(\mathbf{p}) = r_{\mathrm{cross}}(\mathbf{q})$.

Similarly, any affine lower bound on the crossover rate function is a valid crossover min-tradeoff function. Denoting $g(\mathbf{q})$ as the crossover min-tradeoff function, it can be written as:

$$g(\mathbf{q}) = \mathbf{g} \cdot \mathbf{q} + k_g, \tag{3.22}$$

where $\mathbf{g}$ is a gradient vector, and k_g is some scalar. Any crossover min-tradeoff function $g(\mathbf{q})$ can be converted into a min-tradeoff function $f(\mathbf{p})$ using the procedure provided in [37] as follows:

$$f(\delta_c) = \begin{cases} \frac{1}{\gamma} g(\delta_c) + \left(1 - \frac{1}{\gamma}\right) \mathrm{Max}(g) & \text{if } c \neq \perp \\ \mathrm{Max}(g) & \text{if } c = \perp \end{cases}, \tag{3.23}$$

where δ_c denotes the point distribution on the value c. With this formula, we have:

$$\begin{aligned} f(\mathbf{p}) = \sum_c p_c f(\delta_c) &= \left(\sum_{c \neq \perp} p_c \left(\frac{1}{\gamma} g(\delta_c) + \left(1 - \frac{1}{\gamma}\right) \mathrm{Max}(g) \right) \right) + p_\perp \, \mathrm{Max}(g) \\ &= \frac{1}{\gamma} \mathbf{g} \cdot \mathbf{p}_{\setminus \perp} + (1 - p_\perp) \left(\frac{k_g}{\gamma} + \left(1 - \frac{1}{\gamma}\right) \mathrm{Max}(g) \right) + p_\perp \, \mathrm{Max}(g) \\ &= \frac{1}{\gamma} \mathbf{g} \cdot \mathbf{p}_{\setminus \perp} + p_\perp \left(\frac{\mathrm{Max}(g) - k_g}{\gamma} \right) + \left(\left(1 - \frac{1}{\gamma}\right) \mathrm{Max}(g) + \frac{k_g}{\gamma} \right), \end{aligned} \tag{3.24}$$

where in the second line onward we have further rewritten the expressions in terms of the formulation (3.22) for g. Inspecting the last line of the above equation, we see that the following holds:

$$\mathbf{f} = \frac{1}{\gamma} (\mathbf{g}, \mathrm{Max}(g) - k_g) = \frac{1}{\gamma} (\mathbf{g}, \max(\mathbf{g})), \tag{3.25}$$

where $\max(\mathbf{g})$ is the maximum of vector components g_c. Moreover, for any distribution of

the form $\mathbf{p} = (\gamma\mathbf{q}, 1-\gamma)$, this choice of f satisfies

$$\begin{aligned}
\mathrm{Var}(\mathbf{p}, f) &= \sum_c p_c \Big(f(\delta_c) - \sum_{c'} p_{c'} f(\delta_{c'}) \Big)^2 \\
&= (1-\gamma)\Big(\mathrm{Max}(g) - g(\mathbf{q}) \Big)^2 + \sum_{c \neq \perp} \gamma q_c \Big(\mathrm{Max}(g) + \frac{1}{\gamma}\big(g(\delta_c) - \mathrm{Max}(g)\big) - g(\mathbf{q}) \Big)^2 \\
&= (1-\gamma)\Big(\mathrm{Max}(g) - g(\mathbf{q}) \Big)^2 + \sum_{c \neq \perp} \frac{q_c}{\gamma} \big(g(\delta_c) - \mathrm{Max}(g)\big)^2 \\
&\quad + 2(\mathrm{Max}(g) - g(\mathbf{q})) \sum_{c \neq \perp} q_c \big(g(\delta_c) - \mathrm{Max}(g)\big) + \gamma(\mathrm{Max}(g) - g(\mathbf{q}))^2 \sum_{c \neq \perp} q_c \\
&= (1-\gamma)\Big(\mathrm{Max}(g) - g(\mathbf{q}) \Big)^2 + \sum_{c \neq \perp} \frac{q_c}{\gamma} \big(g(\delta_c) - \mathrm{Max}(g)\big)^2 \\
&\quad + 2(\mathrm{Max}(g) - g(\mathbf{q}))\big(g(\mathbf{q}) - \mathrm{Max}(g)\big) + \gamma(\mathrm{Max}(g) - g(\mathbf{q}))^2 \\
&= \sum_{c \neq \perp} \frac{q_c}{\gamma} \Big(\mathrm{Max}(g) - g(\delta_c) \Big)^2 - \Big(\mathrm{Max}(g) - g(\mathbf{q}) \Big)^2 \\
&= \sum_{c \neq \perp} \frac{q_c}{\gamma} \Big(\max(\mathbf{g}) - g_c \Big)^2 - \Big(\max(\mathbf{g}) - \mathbf{g} \cdot \mathbf{q} \Big)^2 . \qquad (3.26)
\end{aligned}$$

Since the variance term only depends on $\mathbf{q}$, the gradient $\mathbf{g}$, and the test probability γ, setting $\mathbf{p} = (\gamma\mathbf{q}, 1-\gamma)$, we have:

$$V(\mathbf{p}, f) = \widetilde{V}(\mathbf{q}, \mathbf{g}) \coloneqq \left(\log(1 + 2d_A^{\kappa}) + \sqrt{2 + \sum_{c \neq \perp} \frac{q_c}{\gamma} \Big(\max(\mathbf{g}) - g_c \Big)^2 - \Big(\max(\mathbf{g}) - \mathbf{g} \cdot \mathbf{q} \Big)^2} \right)^2 . \qquad (3.27)$$

3.4 Single-round analysis

3.4.1 Secure lower bound

In this part, we show how to explicitly obtain a secure lower bound on the key length for Protocol 3. We highlight that the analysis in this section is valid for any given choice of min-tradeoff function. In the upcoming section, we will provide a procedure for choosing an approximately optimal min-tradeoff function.

To simplify the key length analysis, we express the key length formula in Eq. (3.3) in terms of the honest probability distribution $\mathbf{p}^{\text{hon}}$. In particular, the h term in that formula can be expressed as:

$$\begin{aligned} h = \inf_{c^n \in \Omega} f(\text{freq}_{c^n}) &= \inf_{\mathbf{p}^{\text{acc}} \in S_{\text{acc}}} f(\mathbf{p}^{\text{acc}}) \\ &= f(\mathbf{p}^{\text{hon}}) + \inf_{\mathbf{p}^{\text{acc}} \in S_{\text{acc}}} \left(f(\mathbf{p}^{\text{acc}}) - f(\mathbf{p}^{\text{hon}}) \right) \\ &= f(\mathbf{p}^{\text{hon}}) - \Delta_{\text{com}} \quad \text{for} \quad \Delta_{\text{com}} \coloneqq \sup_{\mathbf{p}^{\text{acc}} \in S_{\text{acc}}} -\mathbf{f} \cdot (\mathbf{p}^{\text{acc}} - \mathbf{p}^{\text{hon}}), \end{aligned} \tag{3.28}$$

where S_{acc} is the set of all frequency distributions that are accepted during the acceptance test step. This parameterization highlights the effect of using a non-unique acceptance set in the protocol, as captured by the parameter Δ_{com}.

Another advantage of this parameterization is that, to obtain a secure key length, one merely needs to find lower bounds on $f(\mathbf{p}^{\text{hon}}) + T_\alpha(f)$, regardless of whether the underlying protocol uses a unique-acceptance condition. Accordingly, we begin by explicitly expressing $T_\alpha(f)$ in terms of the rate function obtained in Eq. (3.18):

$$\begin{aligned} & f(\mathbf{p}^{\text{hon}}) + T_\alpha(f) \\ =& f(\mathbf{p}^{\text{hon}}) + \inf_J \left(W(\rho_J^g) - f(\mathbf{p}_J) - \frac{\alpha - 1}{2 - \alpha} \frac{\ln(2)}{2} V(\mathbf{p}_J, f) \right) \text{ where } \mathbf{p}_J \coloneqq \left(\gamma \mathbf{\Phi}[\rho_J^t], 1 - \gamma \right) \\ =& \inf_J \left(W(\rho_J^g) + \mathbf{f} \cdot (\mathbf{p}^{\text{hon}} - \mathbf{p}_J) - \frac{\alpha - 1}{2 - \alpha} \frac{\ln(2)}{2} V(\mathbf{p}_J, f) \right) \\ =& \inf_J \left(W(\rho_J^g) + \mathbf{g} \cdot \left(\mathbf{q}^{\text{hon}} - \mathbf{\Phi}[\rho_J^t] \right) - \frac{\alpha - 1}{2 - \alpha} \frac{\ln(2)}{2} \widetilde{V}\left(\mathbf{\Phi}[\rho_J^t], \mathbf{g} \right) \right), \end{aligned} \tag{3.29}$$

where the last line follows from the application of Eqs. (3.25) and (3.27), along with the fact that we can always write $\mathbf{p}^{\text{hon}} = \left(\gamma \mathbf{q}^{\text{hon}}, 1 - \gamma \right)$ for some distribution $\mathbf{q}^{\text{hon}}$. Substituting this back into Eq. (3.3), we have:

$$\begin{aligned} l \leq n & \left(\inf_J \left(W(\rho_J^g) + \mathbf{g} \cdot \left(\mathbf{q}^{\text{hon}} - \mathbf{\Phi}[\rho_J^t] \right) - \frac{\alpha - 1}{2 - \alpha} \frac{\ln(2)}{2} \widetilde{V}\left(\mathbf{\Phi}[\rho_J^t], \mathbf{g} \right) \right) - \Delta_{\text{com}} \right) \\ & - n \left(\frac{\alpha - 1}{2 - \alpha} \right)^2 K(\alpha) - \lambda_{\text{EC}} - \left\lceil \log \frac{1}{\varepsilon_{\text{EV}}} \right\rceil - \frac{\alpha}{\alpha - 1} \log \frac{1}{\varepsilon_{\text{PA}}} + 2, \end{aligned} \tag{3.30}$$

where $\rho_J^g, \rho_{\mathcal{E}}^g$ are defined via Eq. (3.19). We highlight that, as shown in [81, Supplement, Lemma 2], the term $\widetilde{V}\left(\mathbf{\Phi}[\rho_J^t], \mathbf{g} \right)$ is concave with respect to $\mathbf{\Phi}[\rho_J^t]$, which is itself an affine function of J. Thus, $\widetilde{V}\left(\mathbf{\Phi}[\rho_J^t], \mathbf{g} \right)$ is concave with respect to J. Therefore, the optimization

in the above formula is a convex optimization problem that can be efficiently addressed using the Frank–Wolfe algorithm [76]. It is worth noting that the above bound depends only on the gradient of the crossover min-tradeoff function, not on the offset factor k_g. This simplifies the calculations, as one only needs to determine the gradient rather than the full crossover min-tradeoff function. However, to achieve the best key lengths, it is necessary to choose the optimal gradient, which we will address in the upcoming section.

3.4.2 Optimizing parameter choices

Min-tradeoff function

We begin by optimizing the choice of the min-tradeoff function. The method introduced here provides an approximately optimal choice of this function. It is important to note, however, that the analysis from the previous section yields a secure lower bound on the key length regardless of the specific choice of min-tradeoff function. For the optimization, we fix the values of the testing probability γ and the Rényi parameter α. We then employ a technique similar to those used in [53].

We first note that the only terms that depend on $\mathbf{g}$ are the objective function in the optimization in Eq. (3.29), Δ_{com}, and $K(\alpha)$. However, to simplify the optimization, we consider only the dependence of the min-tradeoff function on the objective function in Eq. (3.29). In fact, we perform an additional round of approximation and consider an approximate bound on the objective function in Eq. (3.29).[10]

To that end, similar to [37], we lower bound the second line in Eq. (3.29) as follows:

$$\begin{aligned}
&f(\mathbf{p}^{\mathrm{hon}}) + \inf_J \left(W(\rho_J^g) - f(\mathbf{p}_J) - \frac{\alpha-1}{2-\alpha}\frac{\ln(2)}{2} V(\mathbf{p}_J, f) \right) \\
\geq & f(\mathbf{p}^{\mathrm{hon}}) + \inf_J \left(-\frac{\alpha-1}{2-\alpha}\frac{\ln(2)}{2} V(\mathbf{p}_J, f) \right) \\
\geq & f(\mathbf{p}^{\mathrm{hon}}) - \frac{\alpha-1}{2-\alpha}\frac{\ln(2)}{2} \left(\log(1+2d_A^{\kappa}) + \sqrt{2 + \frac{1}{\gamma}\left(\max(\mathbf{g}) - \min(\mathbf{g})\right)^2} \right)^2 ,
\end{aligned} \tag{3.31}$$

where the first inequality holds because $W(\rho_J^g) - f(\mathbf{p}_J) \geq 0$, and the second inequality is from [37, Lemma V.5]. Expanding the square in the above expression and recalling that

[10] The reason for applying these approximation steps is that the method we use here relies on explicitly calculating the Legendre–Fenchel conjugate functions, which we were unable to compute without these approximations.

$f(\mathbf{p}^{\text{hon}}) = g(\mathbf{q}^{\text{hon}})$, we see the above is equal to

$$\begin{aligned}
&g(\mathbf{q}^{\text{hon}}) - \frac{\alpha-1}{2-\alpha}\frac{\ln(2)}{2}\Bigg(\left(\log(1+2d_A^\kappa)\right)^2 + 2\log(1+2d_A^\kappa)\sqrt{2+\frac{1}{\gamma}\left(\max(\mathbf{g})-\min(\mathbf{g})\right)^2} + \\
&\qquad + 2 + \frac{1}{\gamma}\left(\max(\mathbf{g})-\min(\mathbf{g})\right)^2 \Bigg) \\
&\approx g(\mathbf{q}^{\text{hon}}) - \varphi_0((\max(\mathbf{g})-\min(\mathbf{g}))^2 - \varphi_1\left(\max(\mathbf{g})-\min(\mathbf{g})\right) - \varphi_2 \\
&= g(\mathbf{q}^{\text{hon}}) - \hat{T}(\mathbf{g}) - \varphi_2 \qquad \text{for } \hat{T}(\mathbf{g}) \coloneqq \varphi_0((\max(\mathbf{g})-\min(\mathbf{g}))^2 + \varphi_1\left(\max(\mathbf{g})-\min(\mathbf{g})\right),
\end{aligned} \tag{3.32}$$

where the approximation in the second line is obtained by dropping the 2 in the square root term[11], and writing

$$\begin{aligned}
\varphi_0 &\coloneqq \frac{\alpha-1}{2-\alpha}\frac{\ln(2)}{2}\frac{1}{\gamma}, \\
\varphi_1 &\coloneqq \frac{\alpha-1}{2-\alpha}\frac{\ln(2)}{2}\frac{2\log(1+2d_A^\kappa)}{\sqrt{\gamma}}, \\
\varphi_2 &\coloneqq \frac{\alpha-1}{2-\alpha}\frac{\ln(2)}{2}\left(2+\left(\log(1+2d_A^\kappa)\right)^2\right).
\end{aligned} \tag{3.33}$$

With the above simplifications, we now turn to finding the optimal choice of $\mathbf{g}$ such that it maximizes $g(\mathbf{q}^{\text{hon}}) - \hat{T}(\mathbf{g})$.[12] We begin with the following lemma:

Lemma 3.4.1. *For any constants $\varphi_0, \varphi_1 \geq 0$, the function $\hat{T}$ defined as*

$$\hat{T}(\mathbf{g}) \coloneqq \varphi_0((\max(\mathbf{g})-\min(\mathbf{g}))^2 + \varphi_1\left(\max(\mathbf{g})-\min(\mathbf{g})\right), \tag{3.34}$$

is a convex function with Legendre-Fenchel conjugate $\hat{T}^*$ given by

$$\hat{T}^*(\boldsymbol{\lambda}) \coloneqq \begin{cases} s(\|\boldsymbol{\lambda}\|_1/2) & \text{if } \sum_c \lambda_c = 0 \\ +\infty & \text{otherwise} \end{cases}, \quad \text{where } s(x) \coloneqq \begin{cases} \frac{(x-\varphi_1)^2}{4\varphi_0} & \text{if } x \geq \varphi_1 \\ 0 & \text{otherwise} \end{cases}, \tag{3.35}$$

and satisfies

$$\hat{T}(\mathbf{g}) = \hat{T}^{**}(\mathbf{g}) = \sup_{\boldsymbol{\lambda}} \left(\mathbf{g}\cdot\boldsymbol{\lambda} - \hat{T}^*(\boldsymbol{\lambda})\right). \tag{3.36}$$

[11] This approximation is reasonably accurate whenever $(\max(\mathbf{g})-\min(\mathbf{g}))^2/\gamma \gg 2$, which is typically the case for the parameter regimes used in this chapter, where γ is small. Since all the other computations here are genuine lower bounds, this at least suggests it is unlikely that we will end up with some choice of $\mathbf{g}$ that is "good" for (3.32) but yields worse results in the actual key length formula (3.30).

[12] Note that the ϕ_2 term is independent of $\mathbf{g}$ and thus is not considered for the purpose of optimizing the choice of $\mathbf{g}$.

Proof. We start the proof by showing that $\hat{T}(\mathbf{g})$ is a convex function. We first restate the definition of $\hat{T}$ from (3.32):

$$\hat{T}(\mathbf{g}) \coloneqq \varphi_0((\max(\mathbf{g}) - \min(\mathbf{g}))^2 + \varphi_1 (\max(\mathbf{g}) - \min(\mathbf{g}))\,.$$

First note that $\max(\mathbf{g})$ is convex and $\min(\mathbf{g})$ is concave. Thus, $\max(\mathbf{g}) - \min(\mathbf{g})$ is convex; moreover, it is non-negative; therefore, its square is also a convex function. By noting that $\varphi_0, \varphi_1 \geq 0$, we conclude $\hat{T}(\mathbf{g})$ is a convex function.

Now to prove its Legendre-Fenchel conjugate is the expression (3.35), we first write out the definition of the conjugate (see e.g. [18, Chapter 3.3]):

$$\hat{T}^*(\boldsymbol{\lambda}) = \sup_{\mathbf{g}} \left(\mathbf{g} \cdot \boldsymbol{\lambda} - \varphi_0((\max(\mathbf{g}) - \min(\mathbf{g}))^2 - \varphi_1 (\max(\mathbf{g}) - \min(\mathbf{g}))\right). \tag{3.37}$$

Let us first consider the case where the sum of the elements of $\boldsymbol{\lambda}$ is zero, i.e., $\sum_c \lambda_c = 0$. Note that for any vector $\mathbf{g}$, we can define another vector $\tilde{\mathbf{g}}$, whose elements are given by:

$$\tilde{g}_c = \begin{cases} \max(g) & \text{if } \lambda_c \geq 0 \\ \min(g) & \text{otherwise} \end{cases}. \tag{3.38}$$

It is not hard to see that for any vector $\mathbf{g}$, the quantity within the supremum in (3.37) is upper bounded by the value with $\tilde{\mathbf{g}}$ in place of $\mathbf{g}$. Furthermore, letting $\mathcal{C}_{\geq 0}$ be the set of c values with $\lambda_c \geq 0$, we have

$$\tilde{\mathbf{g}} \cdot \boldsymbol{\lambda} = \max(g) \sum_{c \in \mathcal{C}_{\geq 0}} \lambda_c + \min(g) \sum_{c \notin \mathcal{C}_{\geq 0}} \lambda_c = (\max(g) - \min(g))\|\boldsymbol{\lambda}\|_1/2, \tag{3.39}$$

using the fact that $\sum_{c \in \mathcal{C}_{\geq 0}} \lambda_c = -\sum_{c \notin \mathcal{C}_{\geq 0}} \lambda_c = \|\boldsymbol{\lambda}\|_1/2$ since $\sum_c \lambda_c = 0$. Therefore, the RHS of (3.37) simplifies to

$$\begin{aligned} \hat{T}^*(\boldsymbol{\lambda}) &= \sup_{\mathbf{g}} \left((\|\boldsymbol{\lambda}\|_1/2 - \varphi_1)\,(\max(\mathbf{g}) - \min(\mathbf{g})) - \varphi_0\,(\max(\mathbf{g}) - \min(\mathbf{g}))^2\right) \\ &= \sup_{\beta \geq 0} \left((\|\boldsymbol{\lambda}\|_1/2 - \varphi_1)\beta - \varphi_0\beta^2\right), \end{aligned} \tag{3.40}$$

where, in the second line, we use the fact that the optimization only depends on the difference $\max(\mathbf{g}) - \min(\mathbf{g})$. The optimization in (3.40) can be solved using simple calculus, which leads to the following:

$$\hat{T}^*(\boldsymbol{\lambda}) = \begin{cases} \frac{(\|\boldsymbol{\lambda}\|_1/2 - \varphi_1)^2}{4\varphi_0} & \text{if } \|\boldsymbol{\lambda}\|_1 \geq 2\varphi_1 \\ 0 & \text{if } \|\boldsymbol{\lambda}\|_1 < 2\varphi_1 \end{cases} \tag{3.41}$$

Let us now look at the situation where $\sum_c \lambda_c \neq 0$. Consider the particular choice $\mathbf{g} = k\mathbf{1}$ for some $k \in \mathbb{R}$. Taking the supremum over such choices of $\mathbf{g}$, from (3.37) we have:

$$\begin{aligned} \hat{T}^*(\boldsymbol{\lambda}) &\geq \sup_k \left(k \sum_c \lambda_c \right) \\ &= +\infty, \end{aligned} \tag{3.42}$$

where the second equality follows by setting $k \to +\infty$ (resp. $k \to -\infty$) if $\sum_c \lambda_c > 0$ (resp. $\sum_c \lambda_c < 0$). Combining (3.41) with (3.42) results the expression in (3.35).

Finally, to show that (3.36) holds, we simply use the fact that $\hat{T}$ is a convex function with its domain being all of $\mathbb{R}^{|\mathcal{C}|-1}$. Therefore its epigraph is a convex closed set, and hence it is equal to the conjugate of its conjugate (see e.g. [18, Chapter 3.3.2]), i.e. we have $\hat{T}(\mathbf{g}) = \hat{T}^{**}(\mathbf{g}) = \sup_{\boldsymbol{\lambda}} \left(\mathbf{g} \cdot \boldsymbol{\lambda} - \hat{T}^*(\boldsymbol{\lambda}) \right)$ as claimed. □

Note that $\hat{T}^*(\boldsymbol{\lambda})$ is also a convex function, as the 1-norm is convex and the prefactor s is a nondecreasing convex function on $\mathbb{R}$. With this lemma, we have the following theorem:

Theorem 3.4.1. *Let r_{best} be defined as the value*

$$r_{\text{best}} \coloneqq \sup_g \left(g(\mathbf{q}^{\text{hon}}) - \hat{T}(\mathbf{g}) \right), \tag{3.43}$$

where the supremum is taken over all valid crossover min-tradeoff functions g. Then we have

$$\begin{aligned} r_{\text{best}} = \inf_{J,\boldsymbol{\lambda}} \quad & W(\rho_J^g) + \hat{T}^*(\boldsymbol{\lambda}) \\ s.t. \quad & \mathbf{q}^{\text{hon}} - \boldsymbol{\Phi}[\rho_J^t] - \boldsymbol{\lambda} = \mathbf{0} \end{aligned} \tag{3.44}$$

$$\begin{aligned} = \inf_{J,\boldsymbol{\lambda},\boldsymbol{\tau}} \quad & W(\rho_J^g) + s\left(\sum_c \tau_c/2 \right) \\ s.t. \quad & \mathbf{q}^{\text{hon}} - \boldsymbol{\Phi}[\rho_J^t] - \boldsymbol{\lambda} = \mathbf{0} \\ & -\boldsymbol{\tau} \leq \boldsymbol{\lambda} \leq \boldsymbol{\tau}, \end{aligned} \tag{3.45}$$

where the domain in each infimum is to be understood as the set of all Choi matrices J and vectors $\boldsymbol{\lambda}, \boldsymbol{\tau} \in \mathbb{R}^{|\mathcal{C}_{\setminus\perp}|}$. Furthermore, suppose $L(\rho_J^g, \boldsymbol{\tau})$ is an affine lower bound on $W(\rho_J^g) + s\left(\sum_c \tau_c/2\right)$, and define

$$\begin{aligned} r_{\text{SDP}} \coloneqq \inf_{J,\boldsymbol{\lambda},\boldsymbol{\tau}} \quad & L(\rho_J^g, \boldsymbol{\tau}) \\ s.t. \quad & \mathbf{q}^{\text{hon}} - \boldsymbol{\Phi}[\rho_J^t] - \boldsymbol{\lambda} = \mathbf{0} \\ & -\boldsymbol{\tau} \leq \boldsymbol{\lambda} \leq \boldsymbol{\tau}. \end{aligned} \tag{3.46}$$

Then the above optimization is an SDP where the optimal dual solution to the constraint $\mathbf{q}^{\text{hon}} - \mathbf{\Phi}[\rho_J^t] - \boldsymbol{\lambda} = \mathbf{0}$ *is a vector* $\mathbf{g}^\star \in \mathbb{R}^{|\mathcal{C}_{\setminus\perp}|}$ *with the following property: there exists a crossover min-tradeoff function* $g^\star$ *with gradient* $\mathbf{g}^\star$ *such that*

$$g^\star(\mathbf{q}^{\text{hon}}) - \hat{T}(\mathbf{g}^\star) \geq r_{\text{SDP}}, \tag{3.47}$$

i.e. it is a feasible point of the supremum in (3.43) *that attains value at least* r_{SDP}.

Proof. We begin by recalling that any affine lower bound on $r_{\text{cross}}(\mathbf{q})$ is a valid choice of crossover min-tradeoff function. That is, any choice of $\mathbf{g}$ and k_g such that $\mathbf{g}\cdot\mathbf{q}+k_g \leq r_{\text{cross}}(\mathbf{q})$ yields a valid crossover min-tradeoff function. Therefore, Eq. (3.43) suggests that for any gradient vector $\mathbf{g}$, the best choice of offset k_g is the maximum value satisfying this lower bound condition. In particular, equivalently, the best choice of this offset is the highest lower bound on the set of probability distributions $\mathbf{q}$. i.e.,

$$\begin{aligned} k_g &= \inf_{\mathbf{q}} \left(r_{\text{cross}}(\mathbf{q}) - \mathbf{g}\cdot\mathbf{q}\right) \\ &= \inf_{J} \left(W(\rho_J^g) - \mathbf{g}\cdot\mathbf{\Phi}[\rho_J^t]\right), \end{aligned} \tag{3.48}$$

where the second line follows from (3.21), with ρ_J^g, ρ_J^t being functions of J via (3.19). Let us define the convex set $\mathcal{D}$ as:

$$\mathcal{D} = \left\{ (J, \boldsymbol{\lambda}) \;\middle|\; J \text{ is Choi matrix } \wedge\; \boldsymbol{\lambda} \in \mathbb{R}^{|\mathcal{C}|-1} \;\wedge\; \sum_c \lambda_c = 0 \right\}. \tag{3.49}$$

Substituting Eq. (3.36) and (3.48) in (3.43), we have:

$$\begin{aligned} r_{\text{best}} &= \sup_{\mathbf{g}} \left(\mathbf{g}\cdot\mathbf{q}^{\text{hon}} + \inf_J \left(W(\rho_J^g) - \mathbf{g}\cdot\mathbf{\Phi}[\rho_J^t]\right) - \sup_{\boldsymbol{\lambda}} \left(\mathbf{g}\cdot\boldsymbol{\lambda} - \hat{T}^*(\boldsymbol{\lambda})\right)\right) \\ &= \sup_{\mathbf{g}} \inf_{(J,\boldsymbol{\lambda})\in\mathcal{D}} \left(W(\rho_J^g) + \hat{T}^*(\boldsymbol{\lambda}) + \mathbf{g}\cdot\left(\mathbf{q}^{\text{hon}} - \mathbf{\Phi}[\rho_J^t] - \boldsymbol{\lambda}\right)\right), \end{aligned} \tag{3.50}$$

where in the second line we use the property that for $\boldsymbol{\lambda}$ such that $\sum_c \lambda_c \neq 0$, we have $\hat{T}(\mathbf{g}) = +\infty$, and thus, by restricting the optimization, no generality is lost. The optimization in the last line of Eq. (3.50) is the Lagrange dual of the following constrained convex optimization:

$$\begin{aligned} \inf_{(J,\boldsymbol{\lambda})\in\mathcal{D}} \quad & W(\rho_J^g) + \hat{T}^*(\boldsymbol{\lambda}) \\ \text{s.t.} \quad & \mathbf{q}^{\text{hon}} - \mathbf{\Phi}[\rho_J^t] - \boldsymbol{\lambda} = \mathbf{0}. \end{aligned} \tag{3.51}$$

To prove Eq. (3.44), we must first show that the domain of optimization in Eq. (3.51) matches that in (3.44), and second, demonstrate that strong duality holds. The first requirement is trivial, as the constraints in (3.51) already, by definition, restrict the domain of $\mathcal{D}$ to $\boldsymbol{\lambda}$ such that $\sum_c \lambda_c = 0$. For the second requirement, we note that by setting $J = \mathbb{I}$ and $\boldsymbol{\lambda} = \mathbf{q}^{\text{hon}} - \boldsymbol{\Phi}[\rho^t_{\mathbb{I}}]$, we obtain a strictly feasible point. Therefore, strong duality follows from the Slater condition in Proposition 2.3.3. Consequently, the optimal solution to Eq. (3.51) is the same as that of Eq. (3.44).

We now turn to prove Eq. (3.45). Observe that for every feasible point $(J, \boldsymbol{\lambda})$ for the optimization in (3.44), taking $J' = J$, $\boldsymbol{\lambda}' = \boldsymbol{\lambda}$, and $\tau'_c = |\lambda_c|$ yields a feasible solution to the optimization in (3.45), and thus we conclude the former is lower bounded by the latter.

For the other direction, for any feasible point $(J, \boldsymbol{\lambda}, \boldsymbol{\tau})$ in (3.45), take $J' = J$ and $\boldsymbol{\lambda}' = \boldsymbol{\lambda}$. Then, we have $\sum_c \tau_c \geq \|\boldsymbol{\lambda}\|_1$, and therefore $s(\sum_c \tau_c) \geq \hat{T}^*(\boldsymbol{\lambda})$. Thus, we see that $(J', \boldsymbol{\lambda}')$ yields a feasible solution with an objective value no higher than that of (3.45). Hence, the optimizations in (3.44) and (3.45) are equal.

We now prove the last piece of the theorem, which is Eq. (3.47). We begin by noting that the SDP in Eq. (3.46) is strictly feasible. This can be seen by taking $J = \mathbb{I}$, $\boldsymbol{\lambda} = \mathbf{q}^{\text{hon}} - \boldsymbol{\Phi}[\rho^t_{\mathbb{I}}]$, and any $\tau_c > |\lambda_c|$. Therefore, by invoking Slater's condition, the dual optimal solution is equal to the primal optimal solution.

As a result, the optimal dual solution to the constraint $\mathbf{q}^{\text{hon}} - \boldsymbol{\Phi}[\rho^t_J] - \boldsymbol{\lambda} = \mathbf{0}$ in the SDP (3.46) is a vector $\mathbf{g}^\star \in \mathbb{R}^{|\mathcal{C}_{\setminus\perp}|}$ such that:

$$r_{\text{SDP}} = \inf_{(J,\boldsymbol{\lambda},\boldsymbol{\tau})\in\mathcal{D}'} \; L(\rho^g_J, \boldsymbol{\tau}) + \mathbf{g}^\star \cdot \left(\mathbf{q}^{\text{hon}} - \boldsymbol{\Phi}[\rho^t_J] - \boldsymbol{\lambda}\right), \tag{3.52}$$

where

$$\mathcal{D}' = \left\{ (J, \boldsymbol{\lambda}, \boldsymbol{\tau}) \;\middle|\; J \text{ is Choi matrix } \wedge \; \boldsymbol{\lambda}, \boldsymbol{\tau} \in \mathbb{R}^{|\mathcal{C}|-1} \; \wedge \; -\boldsymbol{\tau} \leq \boldsymbol{\lambda} \leq \boldsymbol{\tau} \right\} \tag{3.53}$$

For the above choice of gradient vector $\mathbf{g}^\star$, choosing the offset term to be the one in Eq. (3.48), we obtain the following:

$$\begin{aligned}
g^\star(\mathbf{q}^{\text{hon}}) - \hat{T}(\mathbf{g}^\star) &= \mathbf{g}^\star \cdot \mathbf{q}^{\text{hon}} + \inf_J \left(W(\rho^g_J) - \mathbf{g}^\star \cdot \boldsymbol{\Phi}[\rho^t_J]\right) - \sup_{\boldsymbol{\lambda}} \left(\mathbf{g}^\star \cdot \boldsymbol{\lambda} - \hat{T}^*(\boldsymbol{\lambda})\right) \\
&= \inf_{(J,\boldsymbol{\lambda})\in\mathcal{D}} \left(W(\rho^g_J) + \hat{T}^*(\boldsymbol{\lambda}) + \mathbf{g}^\star \cdot \left(\mathbf{q}^{\text{hon}} - \boldsymbol{\Phi}[\rho^t_J] - \boldsymbol{\lambda}\right)\right) \\
&= \inf_{(J,\boldsymbol{\lambda},\boldsymbol{\tau})\in\mathcal{D}'} \left(W(\rho^g_J) + s\left(\sum_c \tau_c/2\right) + \mathbf{g}^\star \cdot \left(\mathbf{q}^{\text{hon}} - \boldsymbol{\Phi}[\rho^t_J] - \boldsymbol{\lambda}\right)\right) \\
&\geq \inf_{(J,\boldsymbol{\lambda},\boldsymbol{\tau})\in\mathcal{D}'} \left(L(\rho^g_J, \boldsymbol{\tau}) + \mathbf{g}^\star \cdot \left(\mathbf{q}^{\text{hon}} - \boldsymbol{\Phi}[\rho^t_J] - \boldsymbol{\lambda}\right)\right) \\
&= r_{\text{SDP}},
\end{aligned} \tag{3.54}$$

where the first line follows from Eq. (3.36), the second line follows by substituting from Eq. (3.48), the third line follows a similar argument as in the derivation of Eq. (3.45), and the penultimate line holds since $L(\rho_J^g, \boldsymbol{\tau})$ is a lower bound on $W(\rho_J^g) + s\,(\sum_c \tau_c/2)$. □

Essentially, the above theorem states that when applying the Frank–Wolfe algorithm in the context of the optimization in (3.44), we obtain a series of affine lower bounds $L(\rho_J^g, \boldsymbol{\tau})$ such that their corresponding r_{SDP} values converge to r_{best}. The bound in (3.47) ensures that the optimization in the final step of the Frank–Wolfe algorithm has the desired property: its dual solution yields a choice of gradient $\mathbf{g}$ that achieves an objective value of at least r_{SDP} in the supremum in (3.43).

Remark 3.4.1. *It is usually more numerically stable to have as few constraints as possible in a convex optimization. Therefore, it might be more stable to use the following optimization instead of the one in Eq.* (3.44)*.*

$$\begin{aligned} \inf_{J,\boldsymbol{\tau}} \quad & W(\rho_J^g) + s\left(\sum_c \tau_c/2\right) \\ s.t. \quad & -\boldsymbol{\tau} \leq \mathbf{q}^{\text{hon}} - \boldsymbol{\Phi}[\rho_J^t] \leq \boldsymbol{\tau}. \end{aligned} \tag{3.55}$$

Completeness penalty term Δ_{com}

We now analyze the completeness term Δ_{com}. We begin with the formula in the last line of Eq. (3.28). As stated by that formula, evaluating the effect of Δ_{com} requires a choice of $\mathbf{f}$, $\mathbf{p}^{\text{hon}}$, and S_{acc}. In the previous section, we discussed how to obtain a relatively optimal choice of $\mathbf{f}$. Here, we focus on the acceptance set of the form in (3.10). Specifically, for a given choice of $\varepsilon_{\text{AT}}^{\text{com}}$, we minimize the Δ_{com} term over the choice of tolerance parameters $\mathbf{t}^{\text{low}}, \mathbf{t}^{\text{upp}} \in \mathbb{R}_{\geq 0}^{|\mathcal{C}|}$, subject to the constraints given in (3.12). Moreover, since Δ_{com} is itself an optimization problem, we are dealing with a nested optimization. To speed up the process,

we begin by simplifying the optimization in (3.28):

$$\begin{aligned} & \sup_{\mathbf{p}^{\text{acc}}} && -\mathbf{f}\cdot(\mathbf{p}^{\text{acc}}-\mathbf{p}^{\text{hon}}) \\ & \text{s.t.} && \mathbf{p}^{\text{hon}}-\mathbf{t}^{\text{low}}\leq\mathbf{p}^{\text{acc}}\leq\mathbf{p}^{\text{hon}}+\mathbf{t}^{\text{upp}} \\ & && \sum_c p_c^{\text{acc}}=1 \end{aligned} \tag{3.56}$$

$$\begin{aligned} = \sup_{\mathbf{v}} && \overline{\mathbf{f}}\cdot\mathbf{v} \\ \text{s.t.} && -\mathbf{t}^{\text{low}}\leq\mathbf{v}\leq\mathbf{t}^{\text{upp}} \\ && \sum_c v_c=0, \end{aligned} \tag{3.57}$$

where in the second line we introduce $\overline{\mathbf{f}}\coloneqq-\mathbf{f}$. This LP is always feasible since $\mathbf{t}^{\text{low}},\mathbf{t}^{\text{upp}}\in\mathbb{R}_{\geq 0}^{|\mathcal{C}|}$, and thus $\sum_c -t_c^{\text{low}}\leq 0\leq\sum_c t_c^{\text{upp}}$. We now explain the optimization procedure. Let $\mathcal{C}_{\geq 0}$ (respectively, $\mathcal{C}_{<0}$) be the set of $c\in\mathcal{C}$ such that $\overline{f}_c\geq 0$ ($\overline{f}_c<0$). We begin by setting v_c as:

$$v_c=\begin{cases} t_c^{\text{upp}} & \text{if } c\in\mathcal{C}_{\geq 0} \\ t_c^{\text{low}} & \text{if } c\in\mathcal{C}_{<0} \end{cases} \tag{3.58}$$

This choice for $\mathbf{v}$ maximizes the objective function and satisfies the first constraint ($\mathbf{p}^{\text{hon}}-\mathbf{t}^{\text{low}}\leq\mathbf{p}^{\text{acc}}\leq\mathbf{p}^{\text{hon}}+\mathbf{t}^{\text{upp}}$). However, it might violate the other constraint. Let us assume the violation is such that $\sum_c v_c>0$, in which case we pick the element v_c corresponding to the smallest $\overline{f}_c$, as it has the least effect on the value of the objective function. We then continuously decrease the value of v_c until the constraint is satisfied (i.e., $\sum_c v_c=0$). If the constraint is not met, we move on to the next smallest $\overline{f}_c$ value over $c\in\mathcal{C}\geq 0$ and repeat the process until we reach $\sum_c v_c=0$, at which point we stop. Once the constraint is met, we take the corresponding $\mathbf{v}$ as the optimal solution. A similar technique is applied in the case where $\sum_c v_c<0$. Finally, we run the optimization in Δ_{com} with the optimal values of $\mathbf{t}^{\text{upp}}$ and $\mathbf{t}^{\text{low}}$ using a standard LP solver to verify the result obtained using the above method.

Other parameters optimizations

The remaining parameters to optimize are the testing probability γ, the Rényi parameter α, and the security parameters $\varepsilon^{\text{secure}}$ and ε_{PA}. We begin by optimizing the choice of security

parameters. Let us define the following function:

$$f_\alpha(\varepsilon^{\text{secure}}, \varepsilon_{\text{PA}}) \coloneqq \frac{\alpha}{\alpha - 1} \log\left(\frac{1}{\varepsilon_{\text{PA}}}\right) + \log\left(\frac{2}{\varepsilon^{\text{secure}} - \varepsilon_{\text{PA}}}\right). \tag{3.59}$$

Then, noting that $\varepsilon^{\text{secure}} = \varepsilon_{\text{PA}} + \varepsilon_{\text{EV}}$, we use simple calculus to determine the following choices for the secrecy and correctness parameters as optimal.

$$\varepsilon_{\text{PA}} = \left(\frac{\alpha}{2\alpha - 1}\right) \varepsilon^{\text{secure}}, \qquad \varepsilon_{\text{EV}} = \left(\frac{\alpha - 1}{2\alpha - 1}\right) \varepsilon^{\text{secure}}. \tag{3.60}$$

Finally, we perform a grid search to determine the optimal values for the testing probability and the Rényi parameter.

3.5 Qubit BB84 with loss

In this section, we examine a variation of the BB84 protocol [13] that employs a perfect single-photon source. We begin by explicitly explaining what happens in a single round of this protocol.

In each round, with probability γ, Alice decides whether that round is a test round. If it is, she sends one of $|+\rangle, |-\rangle$ in the X-basis; otherwise, she sends $|0\rangle, |1\rangle$ in the Z-basis. Upon arrival, with probability γ, Bob measures the received state in the X-basis; otherwise, he measures it in the Z-basis. Alice and Bob record their basis choices and the actual states in classical registers X and Y using the following alphabets:

$$\begin{aligned} \mathcal{X} &= \{(\texttt{Z}, 0), (\texttt{Z}, 1), (\texttt{X}, 0), (\texttt{X}, 1)\} \\ \mathcal{Y} &= \{(\texttt{Z}, 0), (\texttt{Z}, 1), (\texttt{Z}, \texttt{no-det}), (\texttt{X}, 0), (\texttt{X}, 1), (\texttt{X}, \texttt{no-det})\}. \end{aligned} \tag{3.61}$$

Then, Alice and Bob announce classical registers C^A and C^B, respectively, as follows:

$$\begin{aligned} C^A &= \begin{cases} \perp & \text{if} \quad \text{Generation round} \\ X & \text{if} \quad \text{Test round,} \end{cases}, \\ C^B &= \begin{cases} (\texttt{Z}, \texttt{det}) & \text{if} \quad C^A = \perp \ \wedge \ \text{Measured in Z-basis} \ \wedge \ \text{detection} \\ (\texttt{Z}, \texttt{no-det}) & \text{if} \quad C^A = \perp \ \wedge \ \text{Measured in Z-basis} \ \wedge \ \text{no-detection}. \\ Y & \text{if} \quad \text{otherwise} \end{cases} \end{aligned} \tag{3.62}$$

Furthermore, for this protocol, we set the C register in Protocol 3 as follows:

$$C = \begin{cases} \perp & \text{if} \quad C^A = \perp \\ (C^A, C^B) & \text{if} \quad \text{otherwise} \end{cases}. \tag{3.63}$$

We also set $I = (C^A, C^B)$. Afterwards, Alice performs the sifting step to produce the S register. Let $\pi_2 : \mathcal{X} \to \{0, 1\}$ be such that it outputs the second entry of X (e.g., $\pi_2[(\mathtt{Z}, 0)] = 0$). Then, the register S has the following form after the sifting step:

$$S = \begin{cases} \pi_2[X] & \text{if} \quad C^A = \perp \ \wedge \ C^B = (\mathtt{Z}, \mathtt{det}) \\ 0 & \text{if} \quad \text{otherwise} \end{cases}. \tag{3.64}$$

Note that it is clear from the above construction that $\dim(S) = 2$. We now turn to explaining the model for honest behavior.

The honest behavior is modeled as a depolarization channel followed by a probabilistic loss model, as follows: we define a depolarizing parameter $p_{\text{hon}}^{\text{depol}} \in [0, 1]$ and a loss parameter $\zeta_{\text{hon}} \in [0, \infty)$. The states sent by Alice to Bob first undergo depolarizing noise, described as:

$$\mathcal{E}_{\text{depol}}[\rho] \coloneqq (1 - p_{\text{hon}}^{\text{depol}})\rho + p_{\text{hon}}^{\text{depol}} \frac{\mathbb{I}}{2}. \tag{3.65}$$

Then, they undergo a loss mechanism in which, with probability $1 - 10^{-\frac{\zeta_{\text{hon}}}{10,\text{dB}}}$, the qubit is lost. Specifically, they undergo a quantum erasure channel with the erasure parameter $\varepsilon = 1 - 10^{-\frac{\zeta_{\text{hon}}}{10,\text{dB}}}$:

$$\mathcal{E}_{\text{erasure}}[\rho] = \varepsilon \, |\perp\rangle\langle\perp| + (1 - \varepsilon)\rho, \tag{3.66}$$

where $|\perp\rangle\langle\perp|$ is some erasure flag orthogonal to ρ. Bob, on the other hand, performs a projective measurement in some basis on the received states.

Given the above form of announcements, we can also simplify the calculation for the length of the error-correction string λ_{EC} in Eq. (3.13). Specifically, we exploit the fact that for those rounds that are sifted out, the value of the S register is deterministically set to 0. Therefore, we have:

$$H(S|YI)_{\text{hon}} = (1 - \gamma)^2 p_{\text{hon}}^{\mathtt{det}|\mathtt{ZZ}} H(S|Y; C^A = \perp \wedge C^B = (\mathtt{Z}, \mathtt{det}))_{\text{hon}}, \tag{3.67}$$

where $p_{\text{hon}}^{\mathtt{det}|\mathtt{ZZ}}$ is the single-round probability of a detection event, conditioned on both sending and measuring in the Z basis, which in our case is given by $p_{\text{hon}}^{\mathtt{det}|\mathtt{ZZ}} = 10^{-\frac{\zeta_{\text{hon}}}{10,\text{dB}}}$. Moreover, we omit the I register since, for non-sifted rounds, it takes a fixed value.

With these descriptions, one can follow a procedure similar to [132] to evaluate key rates. We highlight that the full details of this step are outside the scope of this thesis, and we refer to [65] for further information.

3.5.1 Numerical results

We are now in a position to present the results obtained by implementing the above method in MATLAB to compute secure key rates for qubit BB84 with loss. For further simplification and a more detailed explanation of how to implement the procedure described in Sec. 3.4 for the protocol outlined in the previous section, we refer to [65].

In Fig. 3.1, we plot the finite-size key rates for the qubit BB84 with loss protocol described in the previous section, as a function of the loss parameter in dB. Motivated by realistic setups, we chose the security parameter $\varepsilon^{\text{secure}} = 10^{-8}$ and the depolarization parameter $p_{\text{hon}}^{\text{depol}} = 10^{-2}$. For simplicity, we used the unique-acceptance set $S_{\text{acc}} = \{\mathbf{p}^{\text{hon}}\}$. We compare our results with the key rates obtained in [67], which were calculated under the assumption of IID attacks.

Fig. 3.2 presents a comparison between unique acceptance and a realistic acceptance set, where the realistic set is chosen such that $\Pr[\text{freq}(c^n) \notin S_{\text{acc}}] \leq \varepsilon^{\text{com}} = 10^{-3}$. We observe that the difference between these two acceptance sets is small. Additionally, we compare the key rates obtained using our method with those reported in [122, 83]. While the EUR-based method yields higher key rates, the difference diminishes as n increases.

3.6 Decoy-state analysis

As another example, we analyze a decoy variant of the BB84 protocol described earlier. We note that we only provide a high-level description of this protocol, as the complete detailed analysis is outside the scope of this thesis. For the full details, we refer to [65].

In this variant, Alice sends one of the fully phase-randomized weak coherent pulses (WCP) with intensities $\{\mu_{\text{sig}}, \mu_2, \mu_3\}$, where μ_{sig} is used mainly for the generation rounds, and the others are used in testing rounds. Moreover, the information is encoded in the polarization degree of freedom $\{\rho_H, \rho_V, \rho_A, \rho_D\}$.

In particular, in each round, with probability γ, Alice chooses whether the round is a testing round. If it is a testing round, then with probability $p(\mu_i|t)$, she selects an intensity μ_i and sends a uniformly random choice of ρ_A, ρ_D in the X-basis with the chosen intensity.

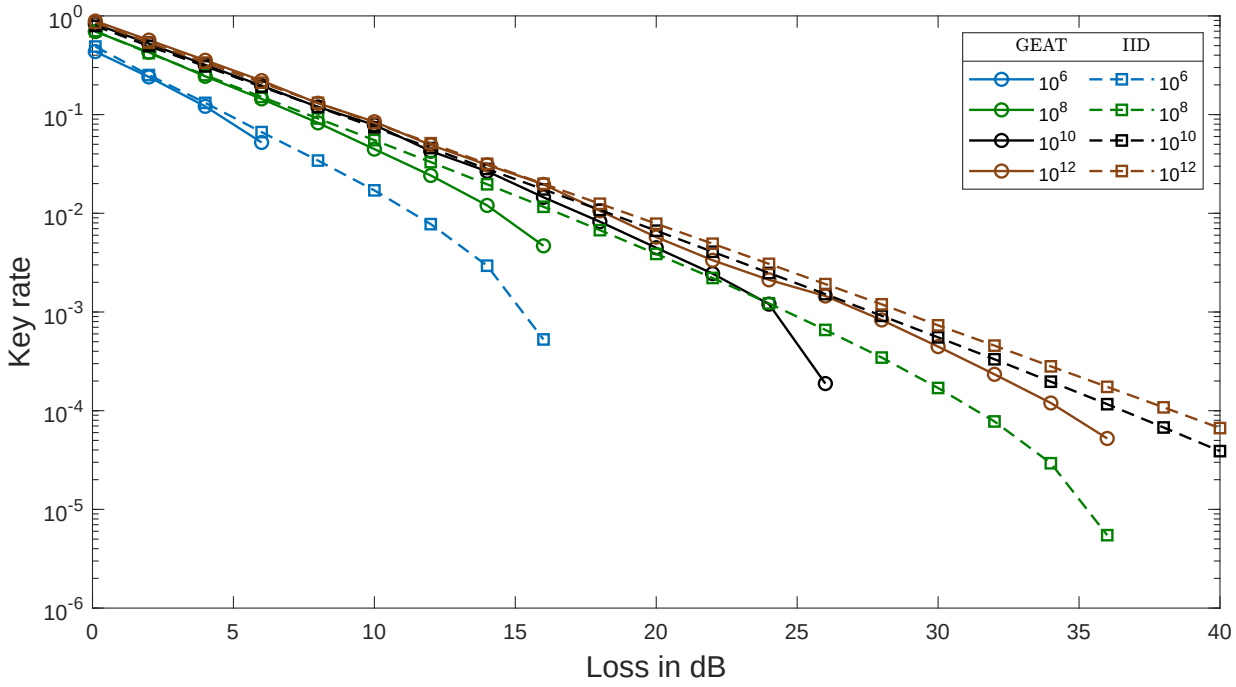


Figure 3.1: Key rate comparison for the qubit BB84 protocol with loss. The GEAT analysis is shown in solid lines, while the IID setting is represented by dashed lines. In all cases, the signals are depolarized with $p_{\text{hon}}^{\text{depol}} = 10^{-2}$, the security parameter is $\varepsilon^{\text{secure}} = 10^{-8}$, and the results correspond to unique acceptance. The testing probability γ and the Rényi parameter α are optimized for the EAT analysis at each data point. The testing probability was also optimized in the IID analysis.

Otherwise, it is a generation round, and she sends a uniformly random choice of states ρ_H, ρ_V in the Z-basis with the μ_{sig} signal.

On Bob's side, we consider an active measurement setup, meaning Bob is actively controlling the polarization rotator that determines the basis for measurement on his side. In particular, similar to the qubit case, with probability γ he measures in the X-basis; otherwise, he measures in the Z-basis. Then, he applies the post-processing procedure from [10, 54] to convert the detection outcome into an equivalent qubit case. Afterwards, he stores the outcome of his measurement in the Y register.

Moving on to the public announcement steps, Alice and Bob announce classical registers C_A and C_B exactly as in the qubit case given in Eq. (3.62). Similarly, the C and I registers are generated exactly as in the qubit case. Finally, the post-sifting register S takes the same form as in Eq. (3.64).

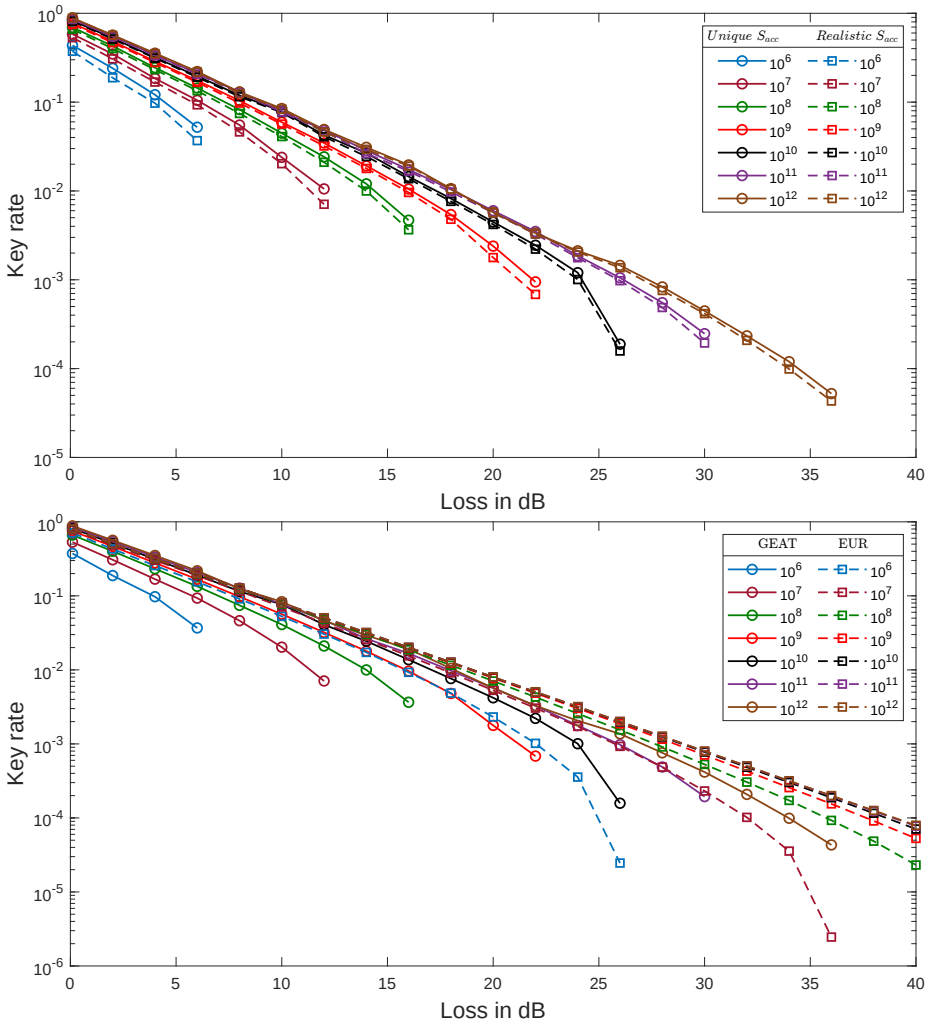


Figure 3.2: Key rates for the qubit BB84 protocol with loss. Upper plot: The solid lines correspond to the unique-acceptance condition, while the dashed lines represent the realistic acceptance condition as described in Sec. 3.3.2. In all cases, the signals are depolarized with $p_{\text{hon}}^{\text{depol}} = 10^{-2}$; the security parameter is $\varepsilon^{\text{secure}} = 10^{-8}$; the completeness parameter for the realistic tolerance intervals is $\varepsilon^{\text{com}} = 10^{-3}$; and both the testing probability γ and the Rényi parameter α are optimized for each data point. Lower plot: The solid lines show the key rates for the realistic acceptance condition using GEAT, whereas the dashed lines indicate the key rates obtained in [122], incorporating the recent improvements developed in [83]. The same values for $p_{\text{hon}}^{\text{depol}}$, $\varepsilon^{\text{secure}}$, and ε^{com} are used, with other parameters similarly optimized.

The honest behavior is modeled as stated in [128], namely a misalignment followed by a loss process. More specifically, we assume the signals sent from Alice undergo a misalignment channel, which is a unitary rotation by an angle $\theta_{\text{hon}}^{\text{misalign}}$ of the incoming polarization mode. Afterwards, they undergo a loss procedure with parameter ζ_{hon}.[13] Given the honest model and the procedure to generate the S register, it can be seen that the same simplification as in Eq. (3.67) holds for the entropic term in the length of the error-correction string.

3.6.1 Numerical results

In this part, we implement the model described above for the decoy-state protocol in MATLAB to obtain key rates for this protocol. For the explicit formulation of the key rate optimization in the decoy-state protocol, we refer to [65].

In Fig. 3.3, we plot key rates for the decoy-state BB84 protocol against loss in dB. Signals are misaligned with the parameter $\theta_{\text{hon}}^{\text{misalign}} = \sin^{-1}(0.1)$. The intensities were chosen as $\{\mu_{\text{sig}} = 0.9,\ \mu_2 = 2 \times 10^{-2},\ \mu_3 = 10^{-3}\}$, and the security parameter was chosen to be $\varepsilon^{\text{secure}} = 10^{-8}$. We compare our results with the IID key rates presented in [67]. Fig. 3.4 shows the comparison between the unique-acceptance and realistic-acceptance cases for the decoy-state BB84 protocol. We observe that, compared to the qubit case, the difference between these two choices of acceptance sets is larger. In particular, the difference becomes more significant in the high-loss regime. Additionally, we compare the key rates for the realistic acceptance set with those reported in [123, 126]. It appears that the EUR approach performs better in the high-loss regime, while the GEAT-based analysis yields higher key rates in the low-loss regime.

[13] It is worth noting that, since we are dealing with WCP signals, the loss model here implements a beamsplitter model for loss from multiphoton states.

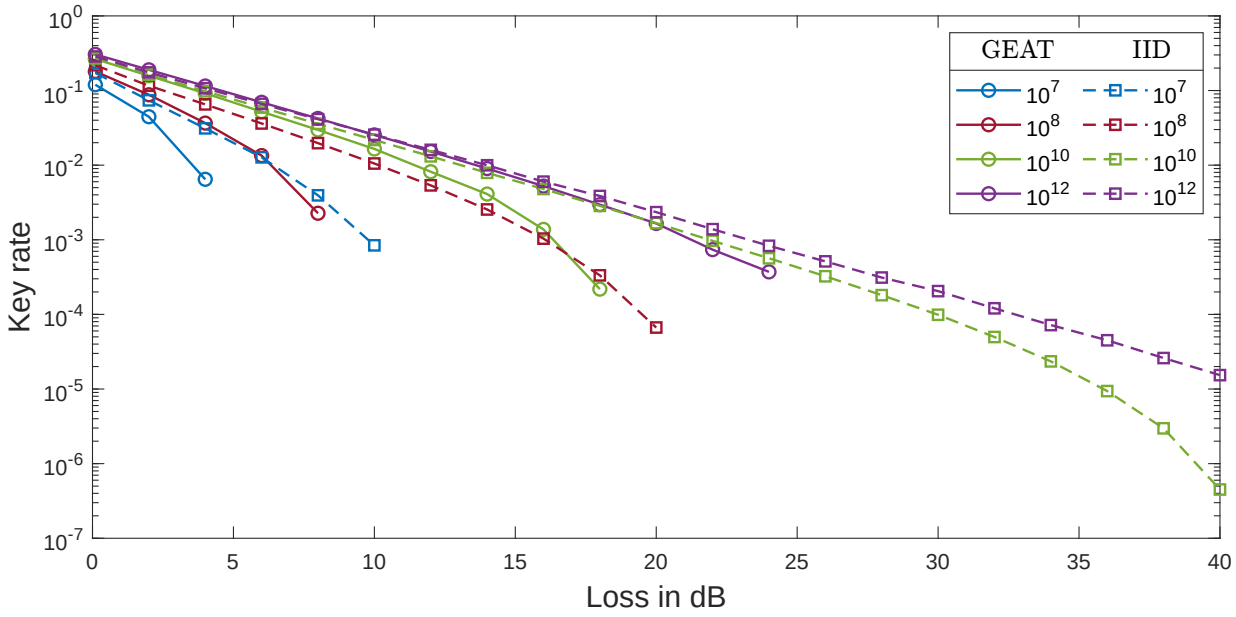


Figure 3.3: Key rates for the decoy-state protocol under the GEAT analysis (solid lines) and the IID setting (dashed lines). In all cases, WCP signals with intensities $\mu_{\text{sig}} = 0.9, \mu_2 = 2 \times 10^{-2}, \mu_3 = 10^{-3}$ were used. Signals are misaligned with $\theta_{\text{hon}}^{\text{misalign}} = \sin^{-1}(0.1)$, and the security parameter is $\varepsilon^{\text{secure}} = 10^{-8}$. The results are obtained for unique acceptance. The testing probability γ and the Rényi parameter α are optimized for the GEAT analysis. Additionally, γ was also optimized for the IID analysis.

3.7 Conclusion

In conclusion, this chapter presents a flexible framework for establishing the security of PM protocols against general attacks. We applied this framework to analyze decoy-state protocols and introduced several methods to refine finite-size corrections. These include an approach for optimizing the min-tradeoff function and integrating additional improvements to the finite-size analysis. Applying this framework to an example of a decoy-state protocol, we demonstrate that reasonably robust key rates can be achieved even in the finite-size regime.

By using the GEAT instead of the original EAT, we achieved key rates with improved loss tolerance and addressed a key limitation in the EAT-based analysis of [53]. Specifically, finite-size key rates derived using the original EAT included a subtractive penalty term proportional to the testing probability. This posed a challenge in loss regimes where the asymptotic key rate was of a similar scale to the testing probability. In contrast, the GEAT

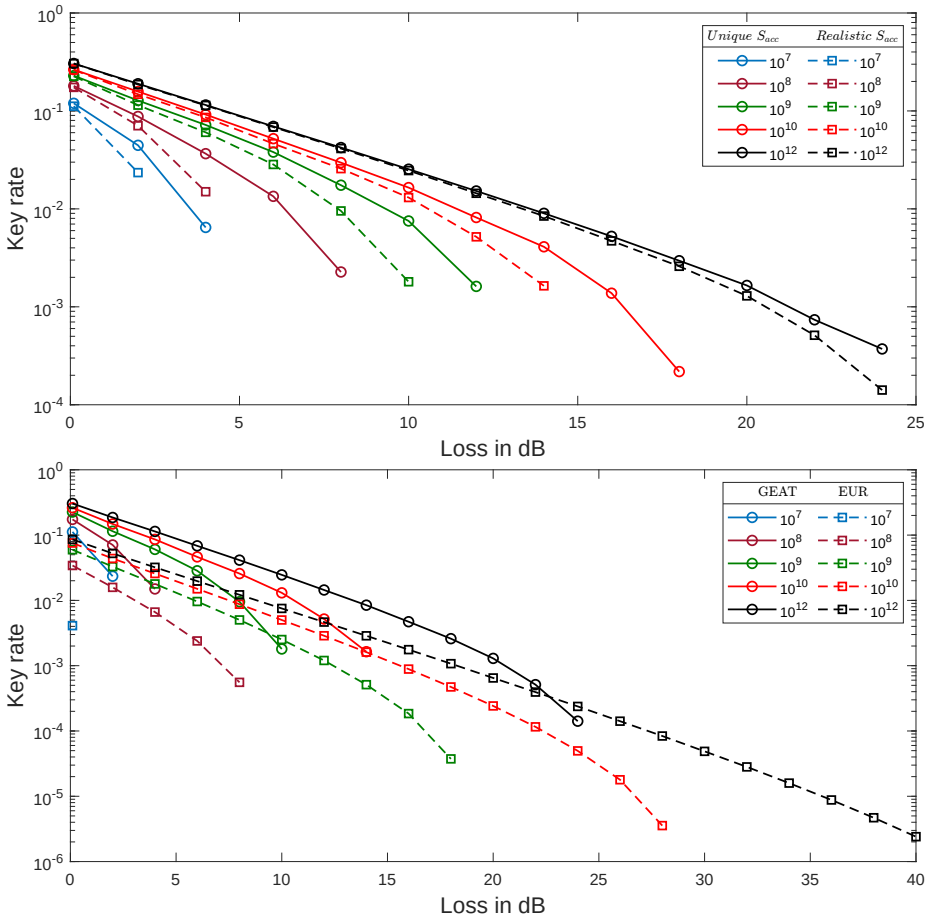


Figure 3.4: Key rates for the decoy-state protocol. Upper plot: The solid lines correspond to the unique-acceptance condition, while the dashed lines represent the realistic acceptance condition. In all cases, two decoy signals are used with intensities $\mu_2 = 2 \times 10^{-2}$ and $\mu_3 = 10^{-3}$, and the actual signal has an intensity of $\mu_{\text{sig}} = 0.9$. The signals are chosen uniformly at random in each round. The misalignment parameter is $\theta_{\text{hon}}^{\text{misalign}} = \sin^{-1}(0.1)$, the security parameter is $\varepsilon^{\text{secure}} = 10^{-8}$, and the completeness parameter is $\varepsilon^{\text{com}} = 10^{-3}$. The testing probability γ and the Rényi parameter α are optimized for each data point. Lower plot: The solid lines show key rates for the realistic acceptance condition, while the dashed lines represent key rates obtained in [122], incorporating recent improvements from [83]. The same values for $\theta_{\text{hon}}^{\text{misalign}}$, $\varepsilon^{\text{secure}}$, and ε^{com} were used, and other parameters were optimized.

transforms the impact of test-round announcements into a multiplicative adjustment of the first-order term, which is significantly less detrimental than the subtractive penalty introduced by the original EAT.

We also note that although the results obtained here are derived within the GEAT framework—and thus subject to its repetition-rate restriction—Chapter 5 demonstrates that these key rates remain valid under the more general framework introduced there, which does not impose any repetition-rate constraint. While these key rates may not be the most optimized, they are nonetheless applicable within this broader and more flexible setting. It is also worth mentioning that a more refined and improved analysis of decoy-state protocols under the framework of Chapter 5 has been done recently in [66].

Chapter 4

Improved Statistical analysis via generalized Rényi entropy accumulation

4.1 Introduction

In the previous chapter, we applied the GEAT to analyze finite-size effects in PM protocols. As we have seen, the GEAT generalizes the original EAT framework by allowing a more flexible treatment of side information, thereby extending its applicability to a broader class of protocols. However, the existing EAT-based techniques have some limitations. In particular, as discussed in the previous chapter, when computing the single-round quantity in those works, one must specify the min-tradeoff function, which significantly affects both the first-order term and finite-size corrections. Therefore, choosing a good min-tradeoff function is of special importance. In addition, current finite-size corrections in EAT-based techniques are outperformed by other approaches, such as the entropic uncertainty principle (EUR) for smooth entropies [119, 80].

In this chapter, we address both of these challenges simultaneously. In particular, we present a bound that has no dependence on the min-tradeoff function, thus avoiding the difficulties associated with optimizing its choice. Moreover, the finite-size bounds in our approach are extremely tight and yield significant improvements in finite-size corrections.

More specifically, the bounds we obtained have $O(1)$ finite-size corrections, as opposed to $O(\sqrt{n})$ in previous EAT approaches. Moreover, they hold for $\alpha > 1$ under the original

EAT model and $\alpha \in (1, 2)$ under the GEAT model. In addition, the single-round quantity is expressed in terms of Rényi entropies; therefore, by exploiting the Rényi privacy amplification in [39], we achieve the goal put forward in [39] of finding a "fully Rényi" approach for security proofs. We also highlight that, with a slight modification, a similar improved statistical analysis compatible with the framework introduced in Chapter 5 can be obtained; however, we defer the details to that chapter.[1]

We highlight that, if desired, the Rényi entropy term in our single-round quantity can be relaxed into von Neumann entropy via a suitable continuity argument [37, 41], which has been the main single-round quantity of interest in previous EAT variations. However, even after this conversion, our bounds still outperform those of previous EAT versions. In particular, we show that for the simple qubit BB84 protocol, our analysis provides improvements over previous results based on the entropic uncertainty relation (EUR) for smooth entropies [119, 80][2]. Our proof techniques also yield upper bounds on smooth max-entropy or for Rényi entropies with $\alpha < 1$, which will be presented in Appendix B.

Our proof techniques extend the connection between entropy accumulation and quantum probability estimation (QPE) developed in [136]. In particular, QPE was originally tailored to the EAT framework, and the connection it provided to QKD was suboptimal for DI-QKD and incompatible with PM-QKD.[3] Here, we build upon their work to provide an analysis that is compatible with the GEAT model and suitable for device-dependent QKD.

This chapter is based on [7].

4.2 Preliminary definitions

In this section, we present some definitions that will be used throughout the chapter. We begin by providing a slightly modified definition of a "sequence of GEAT-with-testing" channels compared to that in Definition 2.5.4.

[1] We emphasize that a framework to analyze the PM-QKD protocol was introduced in [57] prior to [9], which obtained bounds similar to those presented in this chapter. However, the results here offer slight improvements over theirs in the context of fixed-length QKD protocols.

[2] We note that in a recent work, [66] provides a numerical framework to directly tackle the evaluation of Rényi entropies. Therefore, in scenarios where the only reason behind relaxing the single-round Rényi entropy into von Neumann entropy is the difficulty in evaluating the former, such a relaxation may no longer be necessary. Provided that the setting under consideration is compatible with the ones studied in [66], one can directly analyze the single-round Rényi entropy. Thus eliminating the need to convert the single-round entropy to von Neumann entropy when the difficulty lies in evaluating that quantity.

[3] We note that the incompatibility of the QPE framework with EAT mainly stems from the incompatibility of EAT with PM-QKD.

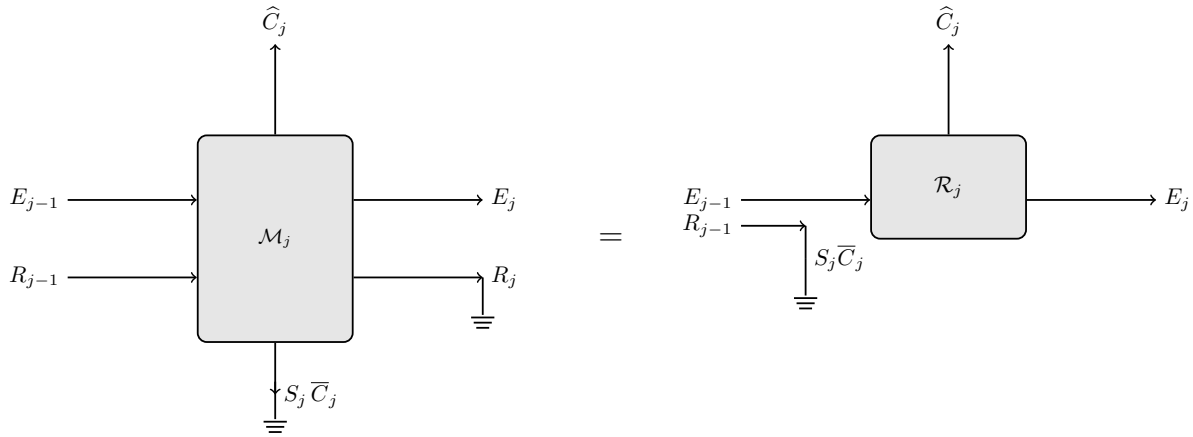


Figure 4.1: Graphical depiction of a non-signaling channel as defined in Definition 4.2.1

Definition 4.2.1. $\{\mathcal{M}_j\}_{j=1}^n$ is called a **sequence of GEAT-with-testing (GEATT) channels** if each $\mathcal{M}_j \in \mathrm{CPTP}(R_{j-1}E_{j-1}, S_jR_jE_j\overline{C}_j\widehat{C}_j)$ is a channel such that the output registers $\overline{C}_j\widehat{C}_j$ are always classical (for any input state), and it satisfies the following **non-signaling (NS) condition** (Fig. 4.1):

$$\exists \text{ a channel } \mathcal{R}_j \in \mathrm{CPTP}(E_{j-1}, E_j\widehat{C}_j) \text{ such that } \mathrm{Tr}_{S_j\overline{C}_jR_j} \circ \mathcal{M}_j = \mathcal{R}_j \circ \mathrm{Tr}_{R_{j-1}}. \tag{4.1}$$

If a state $\rho \in \mathrm{S}_=(S_1^n\overline{C}_1^n\widehat{C}_1^nE_nR_n)$ is of the form $\rho = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1[\omega^0]$ (leaving some identity channels implicit) for some initial state $\omega^0 \in \mathrm{S}_=(R_0E_0)$, we say it is **generated by the sequence of GEATT channels** $\{\mathcal{M}_j\}_{j=1}^n$.

It is worth noting that this definition is slightly different from that of [89] (see also Definition 2.5.4) in that an additional condition was imposed, namely, the classical registers (here $\overline{C}_1^n\widehat{C}_1^n$) should be "projectively reconstructed" from the $S_1^nE_n$ registers. Therefore, the entropic bounds we derive here are of the form $\mathbb{H}(S_1^n\overline{C}_1^n|\widehat{C}_1^nE_n)$, where $\overline{C}_1^n\widehat{C}_1^n$ are explicitly involved in the final bound. However, we highlight that if we were to impose such a condition, then $\mathbb{H}(S_1^n\overline{C}_1^n|\widehat{C}_1^nE_n) = \mathbb{H}(S_1^n|E_n)$, for at least the entropies considered in this chapter, and thus our results would become equivalent in that sense. We note that similar considerations also hold for single-round entropies.[4]

[4]Although it is not explicitly stated in [89] that projective reconstruction is needed for single-round entropies, given their proof methods, it is required to obtain the bound in Eq. (4.10) of [89].

In much of the analysis in this chapter, we deal with purifications of input states to some channel. Therefore, to simplify the setup, we introduce the concept of a purifying function.

Definition 4.2.2. For registers Q, Q' with $\dim(Q) \leq \dim(Q')$, a **purifying function for Q onto Q'** is a function $\texttt{Pur} : \mathrm{S}_{\leq}(Q) \to \mathrm{S}_{\leq}(QQ')$ such that for any state ρ_Q, the state $\texttt{Pur}(\rho_Q)$ is a purification of ρ_Q onto the register Q', i.e. a (possibly subnormalized) rank-1 operator such that $\mathrm{Tr}_{Q'}\left[\texttt{Pur}(\rho_Q)\right] = \rho_Q$.

Note that a purifying function is *not* a channel (i.e. CPTP map), for instance, because it is necessarily nonlinear. As a concrete example, one can choose, for instance, the function

$$\texttt{Pur}(\rho_Q) = \dim(Q) \left(\sqrt{\rho_Q} \otimes \mathbb{I}_{Q'}\right) \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right| \left(\sqrt{\rho_Q} \otimes \mathbb{I}_{Q'}\right), \tag{4.2}$$

where $|\Phi^+\rangle$ is some normalized maximally entangled state across QQ' (up to the support of Q) in some arbitrary basis. However, all results we present in this chapter should be basically independent of any choice of purifying function, by exploiting the isometric equivalence of purifications.

4.3 QEF-type bounds from the GEAT

In this section, we derive similar bounds to the QEF framework as defined in [136] under the GEAT model.

4.3.1 QES-entropies

We begin by defining a variant of QEF in [136], which we term "Quantum Estimation Score-System". We note that this definition is closely related to the notion of QEF in [136] (see also the discussion after Eq. (4.4)), and to the concept of f-weighted entropies introduced in [126], which we extend in Chapter 5. We defer further discussion until after we have introduced and examined some of the properties of these entropies.

Definition 4.3.1. Let $\rho \in \mathrm{S}_{=}(\overline{C}\widehat{C}QQ')$ be a state where $\overline{C}$ and $\widehat{C}$ are classical with alphabets $\overline{\mathcal{C}}$ and $\widehat{\mathcal{C}}$ respectively. A **quantum estimation score-system (QES) on $\overline{C}\widehat{C}$** is simply a function $f : \overline{\mathcal{C}} \times \widehat{\mathcal{C}} \to \mathbb{R}$; equivalently, we may denote it as a real-valued tuple

$\mathbf{f} \in \mathbb{R}^{|\overline{\mathcal{C}} \times \widehat{\mathcal{C}}|}$ where each term in the tuple specifies the value $f(\bar{c}\hat{c})$. Given a QES f and a value[5] $\alpha \in (0,1) \cup (1,\infty)$, the **QES-entropy of order** α for ρ is defined as

$$\begin{aligned} H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho &:= \frac{1}{1-\alpha} \log \left(\sum_{\bar{c}\hat{c}} \rho(\bar{c}\hat{c})^\alpha \rho(\hat{c})^{1-\alpha} \, 2^{(1-\alpha)\left(-f(\bar{c}\hat{c}) - D_\alpha\left(\rho_{QQ'|\bar{c}\hat{c}} \middle\| \mathbb{I}_Q \otimes \rho_{Q'|\hat{c}}\right)\right)} \right) \\ &= \frac{1}{1-\alpha} \log \left(\sum_{\bar{c}\hat{c}} \rho(\bar{c}\hat{c}) 2^{(1-\alpha)\left(-f(\bar{c}\hat{c}) - D_\alpha\left(\rho_{QQ'\wedge\bar{c}\hat{c}} \middle\| \mathbb{I}_Q \otimes \rho_{Q'\wedge\hat{c}}\right)\right)} \right) \\ &= \frac{1}{1-\alpha} \log \left(\sum_{\bar{c}\hat{c}} 2^{-(1-\alpha)f(\bar{c}\hat{c})} \operatorname{Tr}\left[\left(\left(\rho_{Q'\wedge\hat{c}}\right)^{\frac{1-\alpha}{2\alpha}} \rho_{QQ'\wedge\bar{c}\hat{c}} \left(\rho_{Q'\wedge\hat{c}}\right)^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right] \right), \end{aligned} \tag{4.3}$$

where the sum is over all $\bar{c}\hat{c}$ values such that $\rho(\bar{c}\hat{c}) > 0$, and we leave some tensor factors of identity implicit in the last expression.

We highlight that the restriction in the sum is just a convention to avoid complications in defining $D_\alpha\left(\rho_{QQ'\wedge\bar{c}\hat{c}} \| \mathbb{I}_Q \otimes \rho_{Q'\wedge\hat{c}}\right)$ terms when $\rho(\bar{c}\hat{c}) = 0$, especially in Sec. 4.4. However, we note that, in principle, it is possible to assign arbitrary values to the divergence term in such cases, and avoiding that provides cleaner reasoning. With this convention, all terms that appear in the sum are well-defined and finite, because for all such terms we have $\operatorname{Tr}\left[\rho_{QQ'\wedge\bar{c}\hat{c}}\right] > 0$ and

$$\operatorname{supp}\left(\rho_{QQ'\wedge\bar{c}\hat{c}}\right) \subseteq \operatorname{supp}\left(\rho_{QQ'\wedge\hat{c}}\right) \subseteq \operatorname{supp}\left(\mathbb{I}_Q \otimes \rho_{Q'\wedge\hat{c}}\right), \tag{4.4}$$

and similarly for the normalized conditional states, hence the condition for finiteness in (2.22) is satisfied.

Before discussing the intuition behind these entropies, we briefly outline the similarities between QES and the QEF in [136]. We begin by noting that if the Q register were trivial, the QES entropy would be exactly equivalent to a QEF as defined in [136], except we take a logarithmic function of the definition in that work to make the result more apparent in terms of entropies. In other words, they work with a quantity they termed **Rényi powers**, which is essentially the trace term in the last line of Eq. (4.3). Furthermore, a QEF is defined such that it imposes a condition equivalent to $H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho \geq 0$ over some class of states under consideration. However, we will see later that such a condition can be enforced using the normalization property in Lemma 4.3.7. We also highlight that in the case where $\overline{C}$ is trivial, our QES entropies are quite similar to the notion of f-weighted entropy defined

[5] This definition could also be extended to $\alpha \in \{0, 1, \infty\}$ by taking the respective limits.

in [126]. However, as we will extend the definition of f-weighted entropies in Chapter 5, we defer such a comparison to that chapter.

We now turn to building an intuition for QES entropies. Consider the special case where the $\overline{C}$ register is trivial. In that case, the definition becomes:

$$H_\alpha^f(Q|\widehat{C}Q')_\rho = \frac{1}{1-\alpha}\log\left(\sum_{\hat{c}}\rho(\hat{c})\,2^{(1-\alpha)\left(H_\alpha(Q|Q')_{\rho_{|\hat{c}}}-f(\hat{c})\right)}\right), \tag{4.5}$$

Note that if we set $f=0$ (let f be a constant function 0), then, Lemma 2.2.1 shows that the above definition reduces to the usual notion of sandwiched Rényi conditional entropy in Definition 2.2.5 (i.e., $H_\alpha^0(Q|\widehat{C}Q')_\rho = H_\alpha(Q|\widehat{C}Q')_\rho$). Generally, by inspecting the result in Lemma 2.2.1, we can see that the above definition differs from that of Eq. (2.40) by a shift in the exponent that is determined by the value of $f(\hat{c})$. Another way to gain intuition about this quantity, which also clarifies why we are interested in it in this thesis, is as follows. Suppose we want $f(\hat{c})$ to lower bound $H_\alpha(Q|Q')_{\rho|\hat{c}}$ in some average sense. We could impose such a condition by considering the mean value defined as: $\mathbb{E}_{P_X}\left[g(X)\right] \coloneqq \sum_x P_X(x)g(x)$, to satisfy $\mathbb{E}_{\boldsymbol{\rho}_{\widehat{C}}}\left[H_\alpha(Q|Q')_{\rho_{|\widehat{C}}} - f(\widehat{C})\right] \geq 0$. However, for the purposes of this thesis, analyzing such a quantity appears to be challenging. To address this difficulty, we introduce an alternative notion of mean, referred to as the "log-mean-exponential" value. Specifically, we define the log-mean-exponential of a random variable with respect to a base $b \in (0,\infty)$ as:

$$\operatorname*{lme}_{P_X}{}_b\left[g(X)\right] \coloneqq \log_b\left(\sum_x P_X(x)\,b^{g(x)}\right). \tag{4.6}$$

With this notion, Eq. (4.5) can be rewritten as

$$H_\alpha^f(Q|\widehat{C}Q')_\rho = \operatorname*{lme}_{\boldsymbol{\rho}_{\widehat{C}}}{}_b\left[H_\alpha(Q|Q')_{\rho_{|\widehat{C}}} - f(\widehat{C})\right], \quad \text{where } b = 2^{1-\alpha} \in (0,\infty). \tag{4.7}$$

This means the following relation holds:

$$H_\alpha^f(Q|\widehat{C}Q')_\rho \geq 0 \iff \operatorname*{lme}_{\boldsymbol{\rho}_{\widehat{C}}}{}_b\left[H_\alpha(Q|Q')_{\rho_{|\widehat{C}}} - f(\widehat{C})\right] \geq 0. \tag{4.8}$$

In other words, $f(\widehat{C})$ lower bounds $H_\alpha(Q|Q')_{\rho_{|\widehat{C}}}$ in a log-mean-exponential sense.

In addition, in the following lemma, we show that by constructing a new register D that "encodes the value of $f(\overline{C}\widehat{C})$" via its entropy, one can relate these entropies to the usual sandwiched Rényi conditional entropies.

Lemma 4.3.1. *Let f be a QES on some classical registers $\overline{C}\widehat{C}$, let $M \in \mathbb{R}$ be any value such that $M - f(\bar{c}\hat{c}) > 0$ for all $\bar{c}\hat{c}$, and take any $\alpha \in (0,1) \cup (1,\infty)$. Consider any read-and-prepare channel $\mathcal{D} \in \mathrm{CPTP}(\overline{C}\widehat{C}, \overline{C}\widehat{C}D)$ such that the state it prepares on D always satisfies*

$$H_\alpha(D)_{\rho_{|\bar{c}\hat{c}}} = H_\alpha^\uparrow(D)_{\rho_{|\bar{c}\hat{c}}} = M - f(\bar{c}\hat{c}). \tag{4.9}$$

(It is always possible to construct such a channel, and construct it such that D is classical.) Then for any $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$ classical on $\overline{C}\widehat{C}$, extending ρ with this channel yields[6]

$$H_\alpha(DQ\overline{C}|\widehat{C}Q')_\rho = M + H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho. \tag{4.10}$$

Proof. First we briefly verify that such channels can indeed be constructed: note that for any $\alpha > 0$ and $h > 0$, given a register D with dimension at least 2^h, one can straightforwardly construct a state on D with $H_\alpha(D) = H_\alpha^\uparrow(D) = h$ (the first equality holds simply because $H_\alpha = H_\alpha^\uparrow$ when there is no conditioning system), e.g. by taking a mixture between some pure state on D and the maximally mixed state; furthermore this can clearly be achieved using a classical D. Therefore, to obtain a read-and-prepare channel satisfying (4.9), we simply need it to read the classical value $\bar{c}\hat{c}$ and prepare a state on D with the corresponding desired entropy value.

We now show that such a read-and-prepare channel indeed achieves (4.10) (using $\mathbb{U}$ to

[6]There is no danger of ambiguity in having used ρ to denote all states in this lemma, since a read-and-prepare channel always simply extends a state without "disturbing" any registers.

denote maximally mixed states):

$$
\begin{aligned}
&H_\alpha(DQ\overline{C}|\widehat{C}Q')_\rho \\
=&\log\left(d_D d_Q d_{\overline{C}}\right) - D_\alpha\left(\rho_{DQ\overline{C}\widehat{C}Q'}\middle\|\mathbb{U}_{DQ\overline{C}}\otimes\rho_{\widehat{C}Q'}\right) \\
=&\log\left(d_D d_Q d_{\overline{C}}\right) - D_\alpha\left(\sum_{\bar{c}\hat{c}}\rho(\bar{c}\hat{c})\,|\bar{c}\hat{c}\rangle\langle\bar{c}\hat{c}|_{\overline{C}\widehat{C}}\otimes\rho_{DQQ'|\bar{c}\hat{c}}\middle\|\sum_{\bar{c}\hat{c}}\frac{\rho(\hat{c})}{d_{\overline{C}}}\,|\bar{c}\hat{c}\rangle\langle\bar{c}\hat{c}|_{\overline{C}\widehat{C}}\otimes\mathbb{U}_{DQ}\otimes\sigma_{Q'|\hat{c}}\right) \\
=&\log\left(d_D d_Q d_{\overline{C}}\right) - \frac{1}{\alpha-1}\log\left(\sum_{\bar{c}\hat{c}}\rho(\bar{c}\hat{c})^\alpha\left(\frac{\rho(\hat{c})}{d_{\overline{C}}}\right)^{1-\alpha}2^{(\alpha-1)D_\alpha\left(\rho_{DQQ'|\bar{c}\hat{c}}\middle\|\mathbb{U}_{DQ}\otimes\rho_{Q'|\hat{c}}\right)}\right) \\
=&\log\left(d_D d_Q d_{\overline{C}}\right) - \frac{1}{\alpha-1}\log\left(\sum_{\bar{c}\hat{c}}\rho(\bar{c}\hat{c})^\alpha\left(\frac{\rho(\hat{c})}{d_{\overline{C}}}\right)^{1-\alpha}2^{(\alpha-1)\left(D_\alpha\left(\rho_{D|\bar{c}\hat{c}}\middle\|\mathbb{U}_D\right)+D_\alpha\left(\rho_{QQ'|\bar{c}\hat{c}}\middle\|\mathbb{U}_Q\otimes\rho_{Q'|\hat{c}}\right)\right)}\right) \\
=&-\frac{1}{\alpha-1}\log\left(\sum_{\bar{c}\hat{c}}\rho(\bar{c}\hat{c})^\alpha\rho(\hat{c})^{1-\alpha}\,2^{(\alpha-1)\left(-H_\alpha(D)_{\rho_{|\bar{c}\hat{c}}}+D_\alpha\left(\rho_{QQ'|\bar{c}\hat{c}}\middle\|\mathbb{I}_Q\otimes\rho_{Q'|\hat{c}}\right)\right)}\right) \\
=&\frac{1}{1-\alpha}\log\left(\sum_{\bar{c}\hat{c}}\rho(\bar{c}\hat{c})^\alpha\rho(\hat{c})^{1-\alpha}\,2^{(1-\alpha)\left(M-f(\bar{c}\hat{c})-D_\alpha\left(\rho_{QQ'|\bar{c}\hat{c}}\middle\|\mathbb{I}_Q\otimes\rho_{Q'|\hat{c}}\right)\right)}\right) \\
=&M+\frac{1}{1-\alpha}\log\left(\sum_{\bar{c}\hat{c}}\rho(\bar{c}\hat{c})^\alpha\rho(\hat{c})^{1-\alpha}\,2^{(1-\alpha)\left(-f(\bar{c}\hat{c})-D_\alpha\left(\rho_{QQ'|\bar{c}\hat{c}}\middle\|\mathbb{I}_Q\otimes\rho_{Q'|\hat{c}}\right)\right)}\right) \\
=&M+H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho, \qquad (4.11)
\end{aligned}
$$

where the fourth line holds by Lemma 2.2.1, the fifth holds because $\rho_{DQQ'|\bar{c}\hat{c}} = \rho_{D|\bar{c}\hat{c}}\otimes\rho_{QQ'|\bar{c}\hat{c}}$, and the seventh line follows by substitution from Eq. (4.9). □

The above lemma essentially implies that, apart from an additive constant M [7], QES-entropies are just Rényi conditional entropies evaluated over an extension of the original state. Therefore, they satisfy many properties of Rényi conditional entropies. Before stating the properties of these entropies, we first present an approximate version of the above lemma that accounts for all Rényi parameters simultaneously.

Lemma 4.3.2. *Let f be a QES on some classical registers $\overline{C}\widehat{C}$, let $M > 0$ be any value such that $M - f(\bar{c}\hat{c}) > M/2 > 0$ for all $\bar{c}\hat{c}$. Consider any read-and-prepare channel $\mathcal{D} \in \mathrm{CPTP}(\overline{C}\widehat{C}, \overline{C}\widehat{C}D)$ such that the state it prepares on D always satisfies*

$$\forall\alpha\in[0,\infty],\quad H_\alpha(D)_{\rho_{|\bar{c}\hat{c}}} = H_\alpha^\uparrow(D)_{\rho_{|\bar{c}\hat{c}}} \in \left[M - f(\bar{c}\hat{c}), M - f(\bar{c}\hat{c}) + 2^{-\frac{M}{2}}\log e\right]. \qquad (4.12)$$

[7] We note that this additive term, as we will see later, is unimportant in the sense that it does not affect most of the properties we discuss here.

(It is always possible to construct such a channel, and construct it such that D is classical.) Then for any $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$ classical on $\overline{C}\widehat{C}$, extending ρ with this channel yields

$$M + H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho \leq H_\alpha(DQ\overline{C}|\widehat{C}Q')_\rho \leq M + 2^{-\frac{M}{2}} \log e + H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho, \tag{4.13}$$

for any $\alpha \in (0,1) \cup (1,\infty)$.

Proof. To see that such channels indeed exist, simply consider a read-and-prepare channel that reads the classical value $\bar{c}\hat{c}$ and prepares a state on D that is a *uniform* mixture with support size $\left\lceil 2^{M-f(\bar{c}\hat{c})} \right\rceil$; again, this can clearly be achieved with D classical. Since all (unconditioned) Rényi entropies take the same value for a uniform distribution, we see that this means for *every* $\alpha \in [0,\infty]$, this state prepared on D has Rényi entropy

$$H_\alpha(D)_{\rho_{|\bar{c}\hat{c}}} = H_\alpha^\uparrow(D)_{\rho_{|\bar{c}\hat{c}}} = \log\left(\left\lceil 2^{M-f(\bar{c}\hat{c})} \right\rceil\right), \tag{4.14}$$

which lies within the interval described in (4.12) — the upper bound holds from observing that

$$\begin{aligned}
\log\left(\left\lceil 2^{M-f(\bar{c}\hat{c})} \right\rceil\right) &\leq \log\left(2^{M-f(\bar{c}\hat{c})} + 1\right) \\
&= \log\left(2^{M-f(\bar{c}\hat{c})}\right) + \log\left(1 + 2^{f(\bar{c}\hat{c})-M}\right) \\
&\leq M - f(\bar{c}\hat{c}) + 2^{-\frac{M}{2}} \log e,
\end{aligned} \tag{4.15}$$

where in the second line we simply factor out $2^{M-f(\bar{c}\hat{c})}$, and the third line follows from $\ln(1+x) \leq x$ along with the lemma condition $M - f(\bar{c}\hat{c}) > M/2 > 0$.

The remainder of the proof follows the same way as in Lemma 4.3.1, except that we replace the seventh line with inequalities in either direction by bounding $H_\alpha(D)_{\rho_{|\bar{c}\hat{c}}}$ using the interval in (4.12) (together with the observation that the preceding line is monotone increasing with respect to the $H_\alpha(D)_{\rho_{|\bar{c}\hat{c}}}$ terms; note that this is true in both the $\alpha < 1$ and $\alpha > 1$ regimes). □

We now exploit these lemmas to transfer multiple properties of Rényi entropies to QES-entropies.

Lemma 4.3.3. *(Data-processing) Let $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$ be classical on $\overline{C}\widehat{C}$, let f be a QES on $\overline{C}\widehat{C}$, and take any $\alpha \in [\frac{1}{2},1) \cup (1,\infty)$. Then for any channel $\mathcal{E} \in \mathrm{CPTP}(Q',Q'')$,*

$$H_\alpha^f(Q\overline{C}|\widehat{C}Q'')_{\mathcal{E}[\rho]} \geq H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho. \tag{4.16}$$

If $\mathcal{E}$ is an isometry, then we have equality in the above bound.

Proof. Let $\mathcal{D}$ be a read-and-prepare channel as described in Lemma 4.3.1. Then we have (here for clarity we explicitly write $\mathcal{D}$ in the formulas, instead of leaving it implicit in the state extension as in (4.10)):

$$\begin{aligned} H_\alpha^f(Q\overline{C}|\widehat{C}Q'')_{\mathcal{E}[\rho]} &= H_\alpha(DQ\overline{C}|\widehat{C}Q'')_{\mathcal{D}\circ\mathcal{E}[\rho]} - M \\ &= H_\alpha(DQ\overline{C}|\widehat{C}Q'')_{\mathcal{E}\circ\mathcal{D}[\rho]} - M \\ &\geq H_\alpha(DQ\overline{C}|\widehat{C}Q'')_{\mathcal{D}[\rho]} - M \\ &= H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho, \end{aligned} \tag{4.17}$$

where the second line holds since $\mathcal{D}$ commutes with $\mathcal{E}$, and the third line is simply the usual data-processing inequality for Rényi conditional entropies in the range $\alpha \in [\frac{1}{2}, \infty)$ (Theorem 2.2.1). Similarly, the equality condition is simply inherited from that data-processing inequality.

Alternatively, a proof directly from the definition: first considering $\alpha \in [\frac{1}{2}, 1)$, the data-processing inequality for sandwiched Rényi divergence itself gives

$$\begin{aligned} & D_\alpha\left(\mathcal{E}\left[\rho_{QQ'\wedge\bar{c}\hat{c}}\right] \| \mathcal{E}\left[\mathbb{I}_Q \otimes \rho_{Q'\wedge\hat{c}}\right]\right) \leq D_\alpha\left(\rho_{QQ'\wedge\bar{c}\hat{c}} \| \mathbb{I}_Q \otimes \rho_{Q'\wedge\hat{c}}\right) \\ \iff & \rho(\hat{c}\bar{c}) 2^{(1-\alpha)\left(-f(\hat{c}\bar{c}) - D_\alpha\left(\mathcal{E}\left[\rho_{QQ'\wedge\bar{c}\hat{c}}\right] \| \mathcal{E}\left[\mathbb{I}_Q \otimes \rho_{Q'\wedge\hat{c}}\right]\right)\right)} \geq \rho(\hat{c}\bar{c}) 2^{(1-\alpha)\left(-f(\hat{c}\bar{c}) - D_\alpha\left(\rho_{QQ'\wedge\bar{c}\hat{c}} \| \mathbb{I}_Q \otimes \rho_{Q'\wedge\hat{c}}\right)\right)}, \end{aligned} \tag{4.18}$$

summing over all values of $\hat{c}\bar{c}$, taking logarithm, and dividing by $1-\alpha$ would prove the result. The proof for the case where $\alpha \in (1, \infty)$ follows from a similar argument. □

Lemma 4.3.4. *(Conditioning on classical registers) Let $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ'Z)$ be classical on $\overline{C}\widehat{C}Z$, let f be a QES on $\overline{C}\widehat{C}$, and take any $\alpha \in (0,1) \cup (1,\infty)$. Then*

$$H_\alpha^f(Q\overline{C}|\widehat{C}Q'Z)_\rho = \frac{1}{1-\alpha} \log \left(\sum_z \rho(z)\, 2^{(1-\alpha) H_\alpha^f(Q\overline{C}|\widehat{C}Q')_{\rho_{|z}}} \right). \tag{4.19}$$

Proof. Extend ρ with a read-and-prepare channel as described in Lemma 4.3.1. Then

$$\begin{aligned} H_\alpha^f(Q\overline{C}|\widehat{C}Q'Z)_\rho &= H_\alpha(DQ\overline{C}|\widehat{C}Q'Z)_\rho - M \\ &= \frac{1}{1-\alpha}\log\left(\sum_z \rho(z)\, 2^{(1-\alpha)H_\alpha(DQ\overline{C}|\widehat{C}Q')_{\rho_{|z}}}\right) - M \\ &= \frac{1}{1-\alpha}\log\left(\sum_z \rho(z)\, 2^{(1-\alpha)\left(M+H_\alpha^f(Q\overline{C}|\widehat{C}Q')_{\rho_{|z}}\right)}\right) - M \\ &= \frac{1}{1-\alpha}\log\left(\sum_z \rho(z)\, 2^{(1-\alpha)H_\alpha^f(Q\overline{C}|\widehat{C}Q')_{\rho_{|z}}}\right), \end{aligned} \tag{4.20}$$

where in the second line we applied Lemma 2.2.1, and in the third line we implicitly exploited the fact that the read-and-prepare channel in Lemma 4.3.1 achieves the relation (4.10) for all states (in particular, also the conditional states $\rho_{|z}$).

Alternatively, a proof directly from the definition: rewriting the formula for QES-entropy for the state in the lemma, we have:

$$\begin{aligned} &H_\alpha^f(Q\overline{C}|\widehat{C}Q'Z)_\rho \\ &= \frac{1}{1-\alpha}\log\left(\sum_z\sum_{\bar{c}\hat{c}} 2^{-(1-\alpha)f(\bar{c}\hat{c})}\operatorname{Tr}\left[\left(\left(\rho_{Q'\wedge z\hat{c}}\right)^{\frac{1-\alpha}{2\alpha}}\rho_{QQ'\wedge z\bar{c}\hat{c}}\left(\rho_{Q'\wedge z\hat{c}}\right)^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]\right) \\ &= \frac{1}{1-\alpha}\log\left(\sum_z \rho(z)\sum_{\bar{c}\hat{c}} 2^{-(1-\alpha)f(\bar{c}\hat{c})}\operatorname{Tr}\left[\left(\left(\rho_{Q'\wedge\hat{c}|z}\right)^{\frac{1-\alpha}{2\alpha}}\rho_{QQ'\wedge\bar{c}\hat{c}|z}\left(\rho_{Q'\wedge\hat{c}|z}\right)^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]\right) \\ &= \frac{1}{1-\alpha}\log\left(\sum_z \rho(z)\, 2^{(1-\alpha)H_\alpha^f(Q\overline{C}|\widehat{C}Q')_{\rho_{|z}}}\right). \end{aligned} \tag{4.21}$$

□

Lemma 4.3.5. *(Monotonicity in α) Let $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$ be classical on $\overline{C}\widehat{C}$, let f be a tradeoff function on $\overline{C}\widehat{C}$, and take any $\alpha, \alpha' \in [\frac{1}{2}, 1) \cup (1, \infty)$ such that $\alpha \geq \alpha'$. Then,*

$$H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho \leq H_{\alpha'}^f(Q\overline{C}|\widehat{C}Q')_\rho. \tag{4.22}$$

Proof. Let $M > 0$ be any value such that $M - f(\bar{c}\hat{c}) > M/2 > 0$, and extend ρ with a

read-and-prepare channel as described in Lemma 4.3.2, then we have

$$\begin{aligned} H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho &\leq H_\alpha(DQ\overline{C}|\widehat{C}Q')_\rho - M \\ &\leq H_{\alpha'}(DQ\overline{C}|\widehat{C}Q')_\rho - M \\ &\leq 2^{-\frac{M}{2}} \log e + H_{\alpha'}^f(Q\overline{C}|\widehat{C}Q')_\rho, \end{aligned} \tag{4.23}$$

where the first line follows from the first inequality in Eq. (4.13), the second line follows from the monotonicity of sandwiched conditional entropies in α [93], and the last line follows from the second inequality in Eq. (4.13). Since this inequality holds for *arbitrary* (sufficiently large) M, we can take the $M \to \infty$ limit so the $2^{-\frac{M}{2}}$ term vanishes, yielding the desired result. □

Lemma 4.3.6. *(Chain rules) Let f be a QES on classical registers $\overline{C}\widehat{C}$, take any $\alpha, \alpha', \alpha'' \in (\frac{1}{2}, 1) \cup (1, \infty)$ such that $\frac{\alpha}{\alpha-1} = \frac{\alpha'}{\alpha'-1} + \frac{\alpha''}{\alpha''-1}$. Then we have for any $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}RQQ')$:*

$$\begin{aligned} H_\alpha^f(RQ\overline{C}|\widehat{C}Q')_\rho &\leq H_{\alpha'}^f(Q\overline{C}|\widehat{C}Q'R)_\rho + H_{\alpha''}(R|\widehat{C}Q')_\rho \quad \text{if } (\alpha-1)(\alpha'-1)(\alpha''-1) < 0, \\ H_\alpha^f(RQ\overline{C}|\widehat{C}Q')_\rho &\geq H_{\alpha'}^f(Q\overline{C}|\widehat{C}Q'R)_\rho + H_{\alpha''}(R|\widehat{C}Q')_\rho \quad \text{if } (\alpha-1)(\alpha'-1)(\alpha''-1) > 0. \end{aligned} \tag{4.24}$$

Furthermore, we also have for any $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$:

$$\begin{aligned} H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho &\leq H_{\alpha'}^f(Q|\overline{C}\widehat{C}Q')_\rho + H_{\alpha''}(\overline{C}|\widehat{C}Q')_\rho \quad \text{if } (\alpha-1)(\alpha'-1)(\alpha''-1) < 0, \\ H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho &\geq H_{\alpha'}^f(Q|\overline{C}\widehat{C}Q')_\rho + H_{\alpha''}(\overline{C}|\widehat{C}Q')_\rho \quad \text{if } (\alpha-1)(\alpha'-1)(\alpha''-1) > 0, \end{aligned} \tag{4.25}$$

where by $H_\alpha^f(Q|\overline{C}\widehat{C}Q')_\rho$ we mean

$$\begin{aligned} H_\alpha^f(Q|\overline{C}\widehat{C}Q')_\rho &= \frac{1}{1-\alpha} \log \left(\sum_{\bar{c}\hat{c}} \rho(\bar{c}\hat{c}) 2^{(1-\alpha)\left(-f(\bar{c}\hat{c}) - D_\alpha\left(\rho_{QQ'\wedge\bar{c}\hat{c}} \middle\| \mathbb{I}_Q \otimes \rho_{Q'\wedge\bar{c}\hat{c}}\right)\right)} \right) \\ &= \frac{1}{1-\alpha} \log \left(\sum_{\bar{c}\hat{c}} \rho(\bar{c}\hat{c}) 2^{(1-\alpha)\left(-f(\bar{c}\hat{c}) + H_\alpha(Q|Q')_{\rho_{|\bar{c}\hat{c}}}\right)} \right). \end{aligned} \tag{4.26}$$

Proof. We focus only on proving the bounds in (4.25); the bounds in (4.24) hold by precisely analogous arguments, but simpler since we do not need to consider the consistency of (4.26) with our preceding QES-entropy properties.

Take any $M > 0$ such that $M - f(\bar{c}\hat{c}) > \frac{M}{2} > 0$, and extend ρ with a read-and-prepare channel as described in Lemma 4.3.2. Then, as stated in the lemma, we have

$$M + H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho \leq H_\alpha(DQ\overline{C}|\widehat{C}Q')_\rho \leq M + 2^{-\frac{M}{2}} \log e + H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho, \tag{4.27}$$

and also (by viewing $H_\alpha^f(Q|\overline{C}\widehat{C}Q')_\rho$ as an instance of QES-entropy where *both* the $\overline{C}$ and $\widehat{C}$ registers are viewed as the "public" part, as is consistent with the formula (4.26) above)

$$M + H_{\alpha'}^f(Q|\overline{C}\widehat{C}Q')_\rho \leq H_{\alpha'}(DQ|\overline{C}\widehat{C}Q')_\rho \leq M + 2^{-\frac{M}{2}} \log e + H_{\alpha'}^f(Q|\overline{C}\widehat{C}Q')_\rho. \tag{4.28}$$

Then, we have the following chain of inequalities if $(\alpha - 1)(\alpha' - 1)(\alpha'' - 1) < 0$:

$$\begin{aligned} H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho &\leq H_\alpha(DQ\overline{C}|\widehat{C}Q')_\rho - M \\ &\leq H_{\alpha'}(DQ|\overline{C}\widehat{C}Q')_\rho + H_{\alpha''}(\overline{C}|\widehat{C}Q')_\rho - M \\ &\leq H_{\alpha'}^f(Q|\overline{C}\widehat{C}Q')_\rho + H_{\alpha''}(\overline{C}|\widehat{C}Q')_\rho + 2^{-\frac{M}{2}} \log e, \end{aligned} \tag{4.29}$$

where the first line follows from the first inequality in Eq. (4.27), the second line is a special case of [38, Proposition 8], and the third line is a result of applying the second inequality in Eq. (4.28). Taking the $M \to \infty$ limit, we obtain the first bound in (4.25). The second bound (for $(\alpha - 1)(\alpha' - 1)(\alpha'' - 1) > 0$) holds by a similar argument, using [38, Proposition 7] instead. □

Lemma 4.3.7. *(Normalization property) Let $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$ be classical on $\overline{C}\widehat{C}$ and let f be a QES on $\overline{C}\widehat{C}$. Then for any constant $\kappa \in \mathbb{R}$, we have*

$$H_\alpha^{f+\kappa}(Q\overline{C}|\widehat{C}Q')_\rho = H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho - \kappa, \tag{4.30}$$

where $f + \kappa$ simply denotes the QES with values $f(\bar{c}\hat{c}) + \kappa$.

Proof. This follows directly from the definition. □

4.3.2 QES-entropy accumulation or chaining

In this part, we show that QES-entropies can accumulate to provide an estimate of the conditional entropy of the global state. We highlight that this is similar to the QEF chaining property in [136], except that it is compatible with a more general framework of GEAT.

Theorem 4.3.1. *(QES-entropy accumulation) Let ρ be a state generated by a sequence of GEATT channels $\{\mathcal{M}_j\}_{j=1}^n$ (Definition 4.2.1). For each j, suppose that for every value $\bar{c}_1^{j-1}\hat{c}_1^{j-1}$, we have a QES $f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ on registers $\overline{C}_j\widehat{C}_j$.*[8] *Define the following QES on $\overline{C}_1^n\widehat{C}_1^n$:*

$$f_{\text{full}}(\bar{c}_1^n\hat{c}_1^n) \coloneqq \sum_{j=1}^n f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j). \tag{4.31}$$

Take any $\alpha \in (1,2)$ and let $\widehat{\alpha} = 1/(2-\alpha)$. Then we have

$$\begin{aligned} H_\alpha^{f_{\text{full}}}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_\rho &\geq \sum_j \min_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}} \kappa_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}}, \\ \text{where} \quad \kappa_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}} &\coloneqq \inf_{\nu\in\Sigma_j} H_{\widehat{\alpha}}^{f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}}(S_j\overline{C}_j|\widehat{C}_j E_j\widetilde{E})_\nu, \end{aligned} \tag{4.32}$$

where Σ_j denotes the set of all states of the form $\mathcal{M}_j\left[\omega_{R_{j-1}E_{j-1}\widetilde{E}}\right]$ for some initial state $\omega \in \mathrm{S}_=(R_{j-1}E_{j-1}\widetilde{E})$, with $\widetilde{E}$ being a register of large enough dimension to serve as a purifying register for any of the R_jE_j registers.

Consequently, if we instead define the following "normalized" QES on $\overline{C}_1^n\widehat{C}_1^n$:

$$\hat{f}_{\text{full}}(\bar{c}_1^n\hat{c}_1^n) \coloneqq \sum_{j=1}^n \hat{f}_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j), \quad \text{where} \quad \hat{f}_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j) \coloneqq f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j) + \kappa_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}}, \tag{4.33}$$

then

$$H_\alpha^{\hat{f}_{\text{full}}}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_\rho \geq 0. \tag{4.34}$$

Proof. Let $M > 0$ be any value such that $M - f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j) > \frac{M}{2} > 0$ for all the $f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j)$ in the theorem statement (for all j). Now for each j, define a read-and-prepare channel $\mathcal{D}_j \in \mathrm{CPTP}(\overline{C}_1^j\widehat{C}_1^j, \overline{C}_1^j\widehat{C}_1^j D_j)$ of the form described in Lemma 4.3.2, so that

[8] To avoid potential confusion, we highlight that this means for each j we have *multiple* different QES-s, which are indexed by the values $\bar{c}_1^{j-1}\hat{c}_1^{j-1}$. We also highlight that each such QES is a function on the alphabet of $\overline{C}_j\widehat{C}_j$ only, *not* the "preceding" registers $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$. Of course, by the equivalence between k-tuples and functions on a discrete set of size k, in principle we could view the tuple of values $\left\{f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j)\right\}_{\bar{c}_1^j\hat{c}_1^j}$ as being a function on the alphabet of $\overline{C}_1^j\widehat{C}_1^j$, but for this theorem statement, it is more convenient to use the presented form. (In other contexts, it can be convenient to view it as such a function, but we defer that discussion to whenever it should be relevant.)

the state it prepares on D_j satisfies

$$\forall \alpha \in [0,\infty], \quad H_\alpha(D_j)_{\rho_{|\bar{c}_1^j \hat{c}_1^j}} \in \left[M - f_{|\bar{c}_1^{j-1} \hat{c}_1^{j-1}}(\bar{c}_j \hat{c}_j), M - f_{|\bar{c}_1^{j-1} \hat{c}_1^{j-1}}(\bar{c}_j \hat{c}_j) + 2^{-\frac{M}{2}} \log e \right]. \tag{4.35}$$

Now let $\mathcal{N}_j \in \mathrm{CPTP}(R_{j-1}E_{j-1}\overline{C}_1^{j-1}\widehat{C}_1^{j-1}, D_j\overline{\overline{C}}_j S_j R_j E_j \overline{C}_1^j \widehat{C}_1^j)$ denote a channel that does the following (note that its input registers differ from $\mathcal{M}_j$ by also including $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$, and its output registers also include $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$ and two additional registers $\overline{\overline{C}}_j, D_j$):

1. Apply $\mathcal{M}_j \otimes \mathcal{P}_j$, where $\mathcal{P}_j$ is a pinching channel on $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$ (in its classical basis).
2. Copy the classical register $\overline{C}_j$ onto another classical register $\overline{\overline{C}}_j$.
3. Generate a D_j register by applying the above read-and-prepare channel $\mathcal{D}_j$ on $\overline{C}_1^j \widehat{C}_1^j$.

Note that the third step can indeed be implemented as a read-and-prepare channel for *any* input state, even if it is not originally classical on $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$, because the pinching channel in the first step forces the resulting state to be classical on those registers. These channels $\mathcal{N}_j$ satisfy the required condition of Theorem 2.5.3, by identifying the notation in these channels with the notation in that theorem as follows: for the input registers,

- $R_{j-1}\overline{C}_1^{j-1} \leftrightarrow R_{j-1}$,
- $E_{j-1}\widehat{C}_1^{j-1} \leftrightarrow E_{j-1}$,

and for the output registers,

- $D_j\overline{\overline{C}}_j S_j \leftrightarrow S_j$,
- $R_j\overline{C}_1^j \leftrightarrow R_j$,
- $E_j\widehat{C}_1^j \leftrightarrow E_j$.

To verify that the NS condition of Eq. (2.74) is satisfied by this identification, we note that by letting $\mathcal{R}_j$ be the channel in Definition 4.2.1, and defining $\mathcal{R}'_j \coloneqq \mathcal{R}_j \otimes \mathcal{P}'_j$ where $\mathcal{P}'_j$ is a

pinching channel on $\widehat{C}_1^{j-1}$ (in its classical basis), we have

$$\begin{aligned}\mathrm{Tr}_{D_j\overline{\overline{C}}_j S_j R_j \overline{C}_1^j} \circ \mathcal{N}_j &= \mathrm{Tr}_{S_j \overline{C}_1^j R_j} \circ (\mathcal{M}_j \otimes \mathcal{P}_j) \\ &= \left(\mathrm{Tr}_{S_j \overline{C}_j R_j} \circ \mathcal{M}_j\right) \otimes \left(\mathrm{Tr}_{\overline{C}_1^{j-1}} \circ \mathcal{P}_j\right) \\ &= \left(\mathcal{R}_j \circ \mathrm{Tr}_{R_{j-1}}\right) \otimes \left(\mathcal{P}'_j \circ \mathrm{Tr}_{\overline{C}_1^{j-1}}\right) \\ &= \mathcal{R}'_j \circ \mathrm{Tr}_{R_{j-1}\overline{C}_1^{j-1}}, \end{aligned} \tag{4.36}$$

as desired. In the above, the first line holds because the partial trace $\mathrm{Tr}_{D_j\overline{\overline{C}}_j}$ removes all registers generated by $\mathcal{N}_j$ after applying $\mathcal{M}_j \otimes \mathcal{P}_j$, the second line is just a channel regrouping, the third line holds by the NS condition on $\mathcal{M}_j$ in Definition 4.2.1 of GEATT channels (and collapsing part of the pinching channel with the partial trace), and the last line is again a channel regrouping.

These channels $\mathcal{N}_j$ have the critical property that the state $\mathcal{N}_n \circ \cdots \circ \mathcal{N}_1[\omega^0]$ would be identical to $\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1[\omega^0]$ on all registers that are present in the latter.[9] Therefore, we can now write $\rho = \mathcal{N}_n \circ \cdots \circ \mathcal{N}_1[\omega^0]$ without danger of ambiguity with respect to the state ρ in the theorem statement, by viewing the former as just an extension of the latter. Then according to [89, Lemma 3.6] (Theorem 2.5.3), since $\mathcal{N}_j$ are a sequence of GEAT channels, we have

$$H_\alpha(D_1^n S_1^n \overline{\overline{C}}_1^n | \widehat{C}_1^n E_n)_\rho \geq \sum_j \inf_{\nu' \in \Sigma'_j} H_{\widehat{\alpha}}(D_j S_j \overline{\overline{C}}_j | \widehat{C}_1^j E_j \widetilde{E})_{\nu'}, \tag{4.37}$$

where Σ'_j denotes the set of all states that could be produced by $\mathcal{N}_j$ acting on some initial state $\omega' \in \mathrm{S}_=(R_{j-1}E_{j-1}\overline{C}_1^{j-1}\widehat{C}_1^{j-1}\widetilde{E})$.[10] However, recalling that the $\overline{\overline{C}}_j$ register produced by each $\mathcal{N}_j$ channel is always just a copy of the $\overline{C}_j$ register (and also that in the final state, we still have $\overline{\overline{C}}_1^n = \overline{C}_1^n$ because the subsequent channels do not disturb the classical $\overline{C}_j$ registers), this is equivalent to

$$H_\alpha(D_1^n S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_\rho \geq \sum_j \inf_{\nu' \in \Sigma'_j} H_{\widehat{\alpha}}(D_j S_j \overline{C}_j | \widehat{C}_1^j E_j \widetilde{E})_{\nu'}. \tag{4.38}$$

[9] While $\mathcal{N}_j$ performs a pinching channel on $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$ in contrast to $\mathcal{M}_j$ which acts as identity on those registers, this does not make a difference for states produced by applying those channels in sequence, because $\mathcal{N}_{j-1} \circ \cdots \circ \mathcal{N}_1[\omega^0]$ is always already classical on $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$. Similarly, the second and third steps performed by $\mathcal{N}_j$ do not disturb the classical registers they act on.

[10] In this step, let us take $\widetilde{E}$ to be of large enough dimension to be a purifying register for the input registers in the $\mathcal{N}_j$ scenario as well; this can be achieved without loss of generality by expanding its dimension as necessary.

We shall now bound the terms on the right-hand-side in terms of quantities involving only the original channels $\mathcal{M}_j$. Consider any j and take any state $\nu'_{D_j\overline{\overline{C}}_j S_j R_j E_j \overline{C}_1^j \widehat{C}_1^j \widetilde{E}} = \mathcal{N}_j\left[\omega'_{R_{j-1}E_{j-1}\overline{C}_1^{j-1}\widehat{C}_1^{j-1}\widetilde{E}}\right]$ (for some ω') in the corresponding infimum. This state is classical on $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$; furthermore, recalling that $\mathcal{N}_j$ always begins by applying a pinching channel on $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$, we can also take the input state ω' to be classical on $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$ without loss of generality. Letting $\nu'_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ and $\omega'_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ denote the respective states conditioned on[11] $\overline{C}_1^{j-1}\widehat{C}_1^{j-1} = \bar{c}_1^{j-1}\hat{c}_1^{j-1}$, it is straightforward to see that we also have (since $\mathcal{N}_j$ does not disturb $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$ if it is already classical):

$$\nu'_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}} = \mathcal{N}_j\left[\omega'_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}\right], \tag{4.39}$$

a property we will need later in our proof. For now, we observe that since ν' is a mixture of the conditional states $\nu'_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$, we can lower bound it by the "worst-case" term in the mixture, as follows:

$$\begin{aligned} H_{\widehat{\alpha}}(D_j S_j \overline{C}_j | \widehat{C}_1^j E_j \widetilde{E})_{\nu'} &\geq H_{\widehat{\alpha}}(D_j S_j \overline{C}_j | \overline{C}_1^{j-1}\widehat{C}_1^j E_j \widetilde{E})_{\bar{\nu}'} \\ &\geq \min_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}} H_{\widehat{\alpha}}(D_j S_j \overline{C}_j | \widehat{C}_j E_j \widetilde{E})_{\bar{\nu}'_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}}, \end{aligned} \tag{4.40}$$

where the first inequality is by data-processing (Theorem 2.2.1) and the second inequality holds by considering Lemma 2.2.1. (Alternatively, one can directly apply quasi-convexity of Rényi divergence, as described in [68, Proposition 7.35], which implies quasi-concavity of the corresponding conditional entropy.) Now consider any particular value of $\bar{c}_1^{j-1}\hat{c}_1^{j-1}$, and for brevity in the upcoming calculation, let us write $\sigma \coloneqq \nu'_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$. This is classical on

[11] For strict technical accuracy in the following steps, these conditional states should be interpreted to still include the registers $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$, though with those registers simply taking the fixed value $\bar{c}_1^{j-1}\hat{c}_1^{j-1}$.

$\overline{C}_j\widehat{C}_j$, so we can follow a similar chain of calculations as in (4.11) to obtain

$$\begin{aligned}
&H_{\widehat{\alpha}}(D_jS_j\overline{C}_j|\widehat{C}_jE_j\widetilde{E})_\sigma \\
=&\frac{1}{1-\widehat{\alpha}}\log\left(\sum_{\bar{c}_j\hat{c}_j}\sigma(\bar{c}_j\hat{c}_j)^{\widehat{\alpha}}\sigma(\hat{c}_j)^{1-\widehat{\alpha}}\,2^{(1-\widehat{\alpha})\left(H_{\widehat{\alpha}}(D_j)_{\sigma_{|\bar{c}_j\hat{c}_j}}-D_{\widehat{\alpha}}\left(\sigma_{S_jE_j\widetilde{E}|\bar{c}_j\hat{c}_j}\middle\|\mathbb{I}_{S_j}\otimes\sigma_{E_j\widetilde{E}|\hat{c}}\right)\right)}\right) \\
\geq&\frac{1}{1-\widehat{\alpha}}\log\left(\sum_{\bar{c}_j\hat{c}_j}\sigma(\bar{c}_j\hat{c}_j)^{\widehat{\alpha}}\sigma(\hat{c}_j)^{1-\widehat{\alpha}}\,2^{(1-\widehat{\alpha})\left(M-f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j)-D_{\widehat{\alpha}}\left(\sigma_{S_jE_j\widetilde{E}|\bar{c}_j\hat{c}_j}\middle\|\mathbb{I}_{S_j}\otimes\sigma_{E_j\widetilde{E}|\hat{c}}\right)\right)}\right) \\
=&M+H_{\widehat{\alpha}}^{f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}}(S_j\overline{C}_j|\widehat{C}_jE_j\widetilde{E})_\sigma, \qquad (4.41)
\end{aligned}$$

where the inequality in the third line holds because the expression is monotone increasing with respect to the $H_{\widehat{\alpha}}\left(D_j\right)_{\sigma_{|\bar{c}_j\hat{c}_j}}$ terms (for any $\widehat{\alpha}\in(0,1)\cup(1,\infty)$), so we can apply the lower bound in (4.35). (More precisely, in order to apply that lower bound, we are recalling that $\sigma\coloneqq\nu'_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}=\mathcal{N}_j\left[\omega'_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}\right]$ by (4.39), and therefore the D_j register in σ has indeed been produced by a read-and-prepare channel satisfying (4.35), acting on a state in which $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}=\bar{c}_1^{j-1}\hat{c}_1^{j-1}$.)

Critically (for a fixed $\bar{c}_1^{j-1}\hat{c}_1^{j-1}$), the only registers involved in this lower bound are the registers $S_j\overline{C}_j\widehat{C}_jE_j\widetilde{E}$ that can be produced by $\mathcal{M}_j$ (with an extension to an identity channel on $\widetilde{E}$).[12] Furthermore, recall that $\nu'_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ can be produced by $\mathcal{N}_j$ according to (4.39), so we can write

$$\begin{aligned}
H_{\widehat{\alpha}}^{f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}}(S_j\overline{C}_j|\widehat{C}_jE_j\widetilde{E})_{\nu'_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}}&=H_{\widehat{\alpha}}^{f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}}(S_j\overline{C}_j|\widehat{C}_jE_j\widetilde{E})_{\mathcal{N}_j\left[\omega'_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}\right]} \\
&=H_{\widehat{\alpha}}^{f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}}(S_j\overline{C}_j|\widehat{C}_jE_j\widetilde{E})_{\mathcal{M}_j[\omega]}, \qquad (4.42)
\end{aligned}$$

where ω is just $\omega'_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ with $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$ traced out. This is now only a function of $\mathcal{M}_j$ evaluated on some input state $\omega_{R_{j-1}E_{j-1}\widetilde{E}}$, and the specified QES values. Putting together the above, we conclude

$$\inf_{\nu'\in\Sigma'_j}H_{\widehat{\alpha}}(D_jS_j\overline{C}_j|\widehat{C}_1^jE_j\widetilde{E})_{\nu'}\geq M+\min_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}}\inf_{\nu\in\Sigma_j}H_{\widehat{\alpha}}^{f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}}(S_j\overline{C}_j|\widehat{C}_jE_j\widetilde{E})_\nu, \qquad (4.43)$$

[12]While the QES $f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ does depend on $\bar{c}_1^{j-1}\hat{c}_1^{j-1}$, its value is fixed by the initial choices in the theorem, not by the state.

giving us a lower bound on the right-hand-side of (4.38) that only involves the original GEATT channels $\mathcal{M}_j$.

Let us now turn to the left-hand side of (4.38). Again performing calculations analogous to (4.11), we obtain

$$
\begin{aligned}
&H_\alpha(D_1^n S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_\rho \\
&= \frac{1}{1-\alpha} \log \left(\sum_{\bar{c}_1^n \hat{c}_1^n} \rho(\bar{c}_1^n \hat{c}_1^n)^\alpha \rho(\hat{c}_1^n)^{1-\alpha} \, 2^{(1-\alpha)\left(H_\alpha\left(D_1^n\right)_{\rho_{|\bar{c}_1^n \hat{c}_1^n}} - D_\alpha\left(\rho_{S_j E_n | \bar{c}_1^n \hat{c}_1^n} \middle\| \mathbb{I}_{S_j} \otimes \rho_{E_n | \hat{c}_1^n}\right)\right)} \right) \\
&= \frac{1}{1-\alpha} \log \left(\sum_{\bar{c}_1^n \hat{c}_1^n} \rho(\bar{c}_1^n \hat{c}_1^n)^\alpha \rho(\hat{c}_1^n)^{1-\alpha} \, 2^{(1-\alpha)\left(\sum_j H_\alpha(D_j)_{\rho_{|\bar{c}_1^j \hat{c}_1^j}} - D_\alpha\left(\rho_{S_j E_n | \bar{c}_1^n \hat{c}_1^n} \middle\| \mathbb{I}_{S_j} \otimes \rho_{E_n | \hat{c}_1^n}\right)\right)} \right) \\
&\leq \frac{1}{1-\alpha} \log \Bigg(\sum_{\bar{c}_1^n \hat{c}_1^n} \rho(\bar{c}_1^n \hat{c}_1^n)^\alpha \rho(\hat{c}_1^n)^{1-\alpha} \\
&\qquad\qquad 2^{(1-\alpha)\left(\sum_j \left(M - f_{|\bar{c}_1^{j-1} \hat{c}_1^{j-1}}(\bar{c}_j \hat{c}_j) + 2^{-\frac{M}{2}} \log e \right) - D_\alpha\left(\rho_{S_j E_n | \bar{c}_1^n \hat{c}_1^n} \middle\| \mathbb{I}_{S_j} \otimes \rho_{E_n | \hat{c}_1^n}\right)\right)} \Bigg) \\
&= \frac{1}{1-\alpha} \log \left(\sum_{\bar{c}_1^n \hat{c}_1^n} \rho(\bar{c}_1^n \hat{c}_1^n)^\alpha \rho(\hat{c}_1^n)^{1-\alpha} \, 2^{(1-\alpha)\left(nM - f_{\text{full}}(\bar{c}_1^n \hat{c}_1^n) + n 2^{-\frac{M}{2}} \log e - D_\alpha\left(\rho_{S_j E_n | \bar{c}_1^n \hat{c}_1^n} \middle\| \mathbb{I}_{S_j} \otimes \rho_{E_n | \hat{c}_1^n}\right)\right)} \right) \\
&= nM + H_\alpha^{f_{\text{full}}}(S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_\rho + n 2^{-\frac{M}{2}} \log e. \qquad (4.44)
\end{aligned}
$$

In the above, the third line holds because each D_j register is produced by independently applying a read-and-prepare channel on $\overline{C}_1^j \widehat{C}_1^j$, and therefore we have $\rho_{D_1^n | \bar{c}_1^n \hat{c}_1^n} = \bigotimes_{j=1}^n \rho_{D_j | \bar{c}_1^j \hat{c}_1^j}$. The fourth line holds because the expression is again monotone increasing with respect to all the $H_{\widehat{\alpha}}(D_j)$ terms (for any $\alpha \in (0,1) \cup (1,\infty)$), so we can bound it using the upper bound in (4.35). The penultimate line is simply substituting the definition of $f_{\text{full}}(\bar{c}_1^n \hat{c}_1^n)$ in (4.31).

Substituting (4.43) and (4.44) into (4.38), we see that the nM terms can be cancelled off, yielding the conclusion

$$
H_\alpha^{f_{\text{full}}}(S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_\rho \geq \sum_j \min_{\bar{c}_1^{j-1} \hat{c}_1^{j-1}} \inf_{\nu \in \Sigma_j} H_{\widehat{\alpha}}^{f_{|\bar{c}_1^{j-1} \hat{c}_1^{j-1}}}(S_j \overline{C}_j | \widehat{C}_j E_j \widetilde{E})_\nu - n 2^{-\frac{M}{2}} \log e. \qquad (4.45)
$$

However, since this bound holds for arbitrary (sufficiently large) M, we can take the $M \to \infty$ limit so the $2^{-\frac{M}{2}}$ term vanishes, yielding the desired final bound (4.32).

To obtain (4.34), note that by Lemma 4.3.7, each QES $\hat{f}_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ as defined in (4.33) satisfies

$$\inf_{\nu\in\Sigma_j} H_{\widehat{\alpha}}^{\hat{f}_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}}(S_j\overline{C}_j|\widehat{C}_jE_j\widetilde{E})_\nu = \inf_{\nu\in\Sigma_j} H_{\widehat{\alpha}}^{f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}}(S_j\overline{C}_j|\widehat{C}_jE_j\widetilde{E})_\nu - \kappa_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}} = 0, \tag{4.46}$$

and thus by simply applying (4.32) for these QES-s instead, we obtain (4.34). □

We note that the optimization in the calculation of $\kappa_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ is not, as written, a convex optimization. However, the following lemma allows us to express it as a convex optimization problem.

Lemma 4.3.8. *Let $\mathcal{M} \in \mathrm{CPTP}(\widetilde{Q}, SE\overline{C}\widehat{C})$ where the output is always classical on $\overline{C}\widehat{C}$, and let f be a QES on $\overline{C}\widehat{C}$. Let $\widetilde{E}$ be a register such that $\dim(\widetilde{Q}) \leq \dim(\widetilde{E})$, and let* Pur *be a purifying function for $\widetilde{Q}$ onto $\widetilde{E}$ (Definition 4.2.2). Then for any state $\omega_{\widetilde{QE}}$ and $\alpha \in [\frac{1}{2}, 1) \cup (1, \infty)$,*

$$H_\alpha^f(S\overline{C}|\widehat{C}E\widetilde{E})_{\mathcal{M}[\omega_{\widetilde{QE}}]} \geq H_\alpha^f(S\overline{C}|\widehat{C}E\widetilde{E})_{\mathcal{M}[\mathtt{Pur}(\omega_{\widetilde{Q}})]}, \tag{4.47}$$

with equality if $\omega_{\widetilde{QE}}$ is pure, and furthermore for $\alpha \in (1, \infty)$, the right-hand-side is a convex function of $\omega_{\widetilde{Q}}$.

Proof. If $\omega_{\widetilde{QE}}$ is pure, then by using the isometric equivalence of purifications along with the corresponding invariance of QES-entropies (one instance of equality condition in Lemma 4.3.3), we conclude the equality condition in Eq. (4.47). If $\omega_{\widetilde{QE}}$ is not pure, then there exists a purification of it such that $\mathrm{Tr}_{\widetilde{E}'}\left[\omega_{\widetilde{QE}\widetilde{E}'}\right] = \omega_{\widetilde{QE}}$. Thus, we have the following:

$$\begin{aligned} H_\alpha^f(S\overline{C}|\widehat{C}E\widetilde{E})_{\mathcal{M}[\omega_{\widetilde{QE}}]} &= H_\alpha^f(S\overline{C}|\widehat{C}E\widetilde{E})_{\mathcal{M}\circ\mathrm{Tr}_{\widetilde{E}'}[\omega_{\widetilde{QE}\widetilde{E}'}]} \\ &\geq H_\alpha^f(S\overline{C}|\widehat{C}E\widetilde{E}\widetilde{E}')_{\mathcal{M}[\omega_{\widetilde{QE}\widetilde{E}'}]} \\ &= H_\alpha^f(S\overline{C}|\widehat{C}E\widetilde{E})_{\mathcal{M}[\mathtt{Pur}(\omega_{\widetilde{Q}})]}, \end{aligned} \tag{4.48}$$

where the second line holds by applying Lemma 4.3.3, and noting that the partial trace $\mathrm{Tr}_{\widetilde{E}'}[\cdot]$ and $\mathcal{M}$ commute; the last line follows by noting that purifications are isometrically equivalent, and the equality condition in Lemma 4.3.3.

To prove the convexity property, for any two states $\omega_{\widetilde{Q}}^{(1)}, \omega_{\widetilde{Q}}^{(2)}$, and probability distribution $\{p(i)\}_{i=0}^1$, we construct the following state, which is easily verified to be an extension of the

"averaged" state $\omega_{\widetilde{Q}} := \sum_{i=0}^{1} p(i)\omega_{\widetilde{Q}}^{(i)}$:

$$\omega_{\widetilde{Q}\widetilde{E}\widetilde{F}} := \sum_{i=0}^{1} p(i)\, \mathtt{Pur}\left(\omega_{\widetilde{Q}}^{(i)}\right) \otimes |i\rangle\langle i|_{\widetilde{F}}\,. \tag{4.49}$$

Now, consider the state $\mathtt{Pur}\left(\omega_{\widetilde{Q}}\right)$. Since this is a purification of $\omega_{\widetilde{Q}}$ onto $\widetilde{E}$, there exists a channel $\mathcal{E} : \widetilde{E} \to \widetilde{E}\widetilde{F}$ such that

$$\mathcal{E}\left[\mathtt{Pur}\left(\omega_{\widetilde{Q}}\right)\right] = \omega_{\widetilde{Q}\widetilde{E}\widetilde{F}}. \tag{4.50}$$

Therefore, we can bound the function value at the "averaged" state $\omega_{\widetilde{Q}}$ via

$$\begin{aligned} H_\alpha^f(S\overline{C}|\widehat{C}E\widetilde{E})_{\mathcal{M}\left[\mathtt{Pur}\left(\omega_{\widetilde{Q}}\right)\right]} &\leq H_\alpha^f(S\overline{C}|\widehat{C}E\widetilde{E}\widetilde{F})_{\mathcal{E}\circ\mathcal{M}\left[\mathtt{Pur}\left(\omega_{\widetilde{Q}}\right)\right]} \\ &= H_\alpha^f(S\overline{C}|\widehat{C}E\widetilde{E}\widetilde{F})_{\mathcal{M}\circ\mathcal{E}\left[\mathtt{Pur}\left(\omega_{\widetilde{Q}}\right)\right]} \\ &= \frac{1}{1-\alpha}\log\left(\sum_{i=0}^{1} p(i) 2^{(1-\alpha)\left(H_\alpha^f(S\overline{C}|\widehat{C}E\widetilde{E})_{\mathcal{M}\left[\mathtt{Pur}\left(\omega_{\widetilde{Q}}^{(i)}\right)\right]}\right)}\right) \\ &\leq \sum_{i=0}^{1} p(i) H_\alpha^f(S\overline{C}|\widehat{C}E\widetilde{E})_{\mathcal{M}\left[\mathtt{Pur}\left(\omega_{\widetilde{Q}}^{(i)}\right)\right]}, \end{aligned} \tag{4.51}$$

as desired, where the first line is due to the data-processing property in Lemma 4.3.3, the second line holds since $\mathcal{E}$ and $\mathcal{M}$ commute, the third line follows from Eqs. (4.49)–(4.50) and Lemma 4.3.4, and the last line follows by concavity of the log function, and noting $\alpha \in (1, \infty)$. □

Tightness of QES-entropy accumulation theorem

We highlight that for α close to 1, the bound in Theorem 4.31 is almost tight, since it can be saturated by an IID example. More specifically, let $\mathcal{M}_j$ be such that it independently generates a fresh state $\sigma \in \mathrm{S}_=(S_j\overline{C}_j\widehat{C}_j\breve{E}_j)$ and appends $\breve{E}_j$ to E_{j-1}. Then, the global state is an IID state, i.e., $\rho = \sigma^{\otimes n}$. Therefore, if all QES-s are equal to each other, we have:

$$H_\alpha^{f_{\text{full}}}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_\rho = nH_\alpha^f(S_1\overline{C}_1|\widehat{C}_1\breve{E}_1)_\sigma = \sum_j \inf_{\nu\in\Sigma_j} H_\alpha^f(S_j\overline{C}_j|\widehat{C}_j E_j\widetilde{E})_\nu, \tag{4.52}$$

where the only difference from the bound in Eq. (4.32) is the Rényi parameter. However, note that if we let $\alpha = 1 + \mu$ for some $\mu > 0$, we have:

$$\widehat{\alpha} = \frac{1}{1-\mu} = 1 + \frac{\mu}{1-\mu} = 1 + \mu + O(\mu^2). \tag{4.53}$$

In other words, the change in the Rényi parameter is only a second-order correction with respect to $\alpha - 1$. However, we will see later in Section 4.4.1 that this corresponds to a third-order correction with respect to key rates.

It is worth noting that in the case where the QES-s are different in each round, the bound in Eq. (4.40) can also be saturated by constructing a proper initial state. Therefore, any further improvement must be at the level of the "GEAT without testing" bound in (4.37) from [89].

4.3.3 Applications

The variable-length protocols

In this section, we describe how our analysis can be used to prove the security of variable-length protocols. In [126], it was shown that a lower bound on the value of QES entropies leads to a security statement for variable-length protocols. In particular, suppose we have the "normalized QES" bound (4.34) from Theorem 4.3.1. Then, a variable-length secret key is obtained by evaluating the value $\hat{c}_1^n$ on the publicly announced registers $\widehat{C}_1^n$, computing $\hat{f}_{\text{full}}(\hat{c}_1^n)$, and hashing to a key length determined by that value.[13]

One of the main challenges in such an argument is computing $\hat{f}_{\text{full}}(\hat{c}_1^n)$, or more generally $\hat{f}_{\text{full}}(\bar{c}_1^n \hat{c}_1^n)$, since according to Eq. (4.33), evaluating the full estimator requires specifying all the QES-s $f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ across all rounds, along with the values of $\kappa_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$. However, we present a procedure that significantly simplifies these calculations. This simplification was already proposed in [136]. Before stating the procedure, we note that it is based on the critical observation that, in practice, one only needs to compute the full estimator $\hat{f}_{\text{full}}(\bar{c}_1^n \hat{c}_1^n)$ for the single value of $\bar{c}_1^n \hat{c}_1^n$ that occurred during the protocol. We now state the procedure for evaluating the full estimator as follows:

- Without choosing any QES-s, implement the protocol and observe the values of $\bar{c}_1^n \hat{c}_1^n$ on registers $\overline{C}_1^n \widehat{C}_1^n$.

[13] Alternatively, see [66, Theorem 12] for an explicit security proof for variable-length protocols.

- For each j, choose an arbitrary procedure that takes the values of $\bar{c}_1^{j-1}\hat{c}_1^{j-1}$ as input and deterministically outputs a QES on $\overline{C}_j\widehat{C}_j$, denoted by $f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$. Evaluate $\kappa_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ for that choice of QES.
- construct the normalized QES $\hat{f}_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ via the right equation in (4.33).
- Add up the values of normalized single-round QES $\hat{f}_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ to obtain a lower bound on the full estimator defined in the left equation in (4.33).

In fact, the above procedure essentially states that we only need to choose the single QES $f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ actually observed in that round, evaluate the value of the corresponding $\kappa_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$, and then construct the full estimator using Eq. (4.33) to obtain a bound on the global conditional entropy as stated in Eq. (4.34).

The fixed-length protocols

The bounds obtained in Theorem 4.3.1 can be used to bound the secure key rate for fixed-length protocols that make a simple accept/abort decision and produce a fixed-length key upon acceptance. We begin with the following lemma, which is a generalization of [89, Lemma 4.5].

Lemma 4.3.9. *Consider any state $\rho_{SED\overline{C}\widehat{C}}$ that is classical on $\overline{C}\widehat{C}$, and has the form*

$$\rho_{SED\overline{C}\widehat{C}} = \sum_{\bar{c}\hat{c}\in\Omega} \rho(\bar{c}\hat{c})\, |\bar{c}\hat{c}\rangle\langle\bar{c}\hat{c}|_{\overline{C}\widehat{C}} \otimes \rho_{SE|\bar{c}\hat{c}} \otimes \rho_{D|\bar{c}\hat{c}}, \tag{4.54}$$

with Ω being a subset of the alphabet of registers $\overline{C}\widehat{C}$. Then for any $\alpha \in (1,\infty)$, we have the following:

$$H_\alpha^\uparrow(DS\overline{C}|\widehat{C}E)_\rho \leq H_\alpha^\uparrow(S\overline{C}|\widehat{C}E)_\rho + \max_{\bar{c}\hat{c}\in\Omega} H_\alpha(D)_{\rho|\bar{c}\hat{c}}, \tag{4.55}$$

$$H_\alpha(DS\overline{C}|\widehat{C}E)_\rho \leq H_\alpha(S\overline{C}|\widehat{C}E)_\rho + \max_{\bar{c}\hat{c}\in\Omega} H_\alpha(D)_{\rho|\bar{c}\hat{c}}. \tag{4.56}$$

Proof. Let $\sigma_{\widehat{C}E}$ be the optimizer in Definition 2.2.5 for the conditional entropy in the LHS of Eq. (4.55), i.e.,

$$H_\alpha^\uparrow(DS\overline{C}|\widehat{C}E)_\rho = -D_\alpha(\rho_{DSE\overline{C}\widehat{C}}\|\mathbb{I}_{DS\overline{C}} \otimes \sigma_{\widehat{C}E}). \tag{4.57}$$

From Lemma 2.2.1, we have

$$
\begin{aligned}
&H_\alpha^\uparrow(DS\overline{C}|\widehat{C}E)_\rho \\
&= \frac{1}{1-\alpha}\log\left(\sum_{\bar{c}\hat{c}\in\Omega}\rho(\bar{c}\hat{c})^\alpha\sigma(\hat{c})^{1-\alpha}2^{(\alpha-1)D_\alpha(\rho_{DSE|\bar{c}\hat{c}}\|\mathbb{I}_{DS}\otimes\sigma_{E|\hat{c}})}\right) \\
&= \frac{1}{1-\alpha}\log\left(\sum_{\bar{c}\hat{c}\in\Omega}\rho(\bar{c}\hat{c})^\alpha\sigma(\hat{c})^{1-\alpha}2^{(\alpha-1)\left(D_\alpha(\rho_{D|\bar{c}\hat{c}}\|\mathbb{I}_D)+D_\alpha(\rho_{SE|\bar{c}\hat{c}}\|\mathbb{I}_S\otimes\sigma_{E|\hat{c}})\right)}\right) \\
&\leq \frac{1}{1-\alpha}\log\left(\min_{\bar{c}\hat{c}\in\Omega}\left[2^{(\alpha-1)D_\alpha(\rho_{D|\bar{c}\hat{c}}\|\mathbb{I}_D)}\right]\sum_{\bar{c}\hat{c}\in\Omega}\rho(\bar{c}\hat{c})^\alpha\sigma(\hat{c})^{1-\alpha}2^{(\alpha-1)D_\alpha(\rho_{SE|\bar{c}\hat{c}}\|\mathbb{I}_S\otimes\sigma_{E|\hat{c}})}\right) \\
&\leq \max_{\bar{c}\hat{c}\in\Omega} H_\alpha(D)_{\rho|\bar{c}\hat{c}} + H_\alpha^\uparrow(S\overline{C}|\widehat{C}E)_\rho, \qquad (4.58)
\end{aligned}
$$

where the second line follows from the fact that $\rho_{DSE|\bar{c}\hat{c}} = \rho_{D|\bar{c}\hat{c}} \otimes \rho_{SE|\bar{c}\hat{c}}$, and the last line follows from Definition 2.2.5, and concavity of log function. This yields the first bound (4.55). To get the second bound (4.56), simply repeat the above computations with σ replaced by ρ itself. □

With the above lemma in hand, we have the following corollary.

Corollary 4.3.1. *Consider the same conditions and notation as in Theorem 4.3.1. Suppose furthermore that the state ρ is of the form*[14] $p_\Omega\rho_{|\Omega} + (1-p_\Omega)\rho_{|\overline{\Omega}}$ *for some $p_\Omega \in (0,1]$ and normalized states $\rho_{|\Omega}, \rho_{|\overline{\Omega}}$, and let $\widetilde{\Omega}$ denote the set of all values $\bar{c}_1^n\hat{c}_1^n$ that have nonzero probability in $\rho_{|\Omega}$. Then*

$$
\begin{aligned}
H_\alpha^\uparrow(S_1^n\overline{C}_1^n|\widehat{C}_1^nE_n)_{\rho_{|\Omega}} &\geq \min_{\bar{c}_1^n\hat{c}_1^n\in\widetilde{\Omega}} \hat{f}_{\text{full}}(\bar{c}_1^n\hat{c}_1^n) - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega} \\
&\geq \left(\sum_j \min_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}} \kappa_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}}\right) + \min_{\bar{c}_1^n\hat{c}_1^n\in\widetilde{\Omega}} f_{\text{full}}(\bar{c}_1^n\hat{c}_1^n) - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega} \qquad (4.59)
\end{aligned}
$$

Proof. Let $\mathcal{D} \in \text{CPTP}(\overline{C}_1^n\widehat{C}_1^n, \overline{C}_1^n\widehat{C}_1^nD)$ be a read-and-prepare channel in the sense of Lemma 4.3.1 for the QES $\hat{f}_{\text{full}(\bar{c}_1^n\hat{c}_1^n)}$ stated in the theorem conditions, i.e. it reads the registers $\overline{C}_1^n\widehat{C}_1^n$ and generates a corresponding state on register D. By linearity, we can

[14] We write this condition in this form to highlight that Ω does not have to be an event defined entirely on the state ρ itself, which allows more flexibility in applications. This is also why we use a separate symbol $\widetilde{\Omega}$ to denote the set of values $\bar{c}_1^n\hat{c}_1^n$ such that $\rho_{|\Omega}(\bar{c}_1^n\hat{c}_1^n) > 0$.

write $\mathcal{D}[\rho] = p_\Omega \mathcal{D}[\rho_{|\Omega}] + (1-p_\Omega)\mathcal{D}[\rho_{|\overline{\Omega}}]$ (in this proof, for clarity we explicitly write out the $\mathcal{D}$ channel when extending the states with it). Note that by construction of the $\mathcal{D}$ channel, the $\mathcal{D}[\rho_{|\Omega}]$ term is of the form (4.54), with the sum over $\bar{c}\hat{c}$ being restricted to $\widetilde{\Omega}$ since that was the case in the original state $\rho_{|\Omega}$. Hence we are allowed to apply Lemma 4.3.9, obtaining

$$\begin{aligned}
H_\alpha^\uparrow(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_{\mathcal{D}[\rho_{|\Omega}]} &\geq H_\alpha^\uparrow(DS_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_{\mathcal{D}[\rho_{|\Omega}]} - \max_{\bar{c}_1^n\hat{c}_1^n\in\widetilde{\Omega}} H_\alpha(D)_{\mathcal{D}[\rho]_{|\bar{c}_1^n\hat{c}_1^n}} \\
&\geq H_\alpha(DS_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_{\mathcal{D}[\rho]} - \max_{\bar{c}_1^n\hat{c}_1^n\in\widetilde{\Omega}} H_\alpha(D)_{\mathcal{D}[\rho]_{|\bar{c}_1^n\hat{c}_1^n}} - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega} \\
&= M + H_\alpha^{\hat{f}_{\text{full}}}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_\rho - \max_{\bar{c}_1^n\hat{c}_1^n\in\widetilde{\Omega}} H_\alpha(D)_{\mathcal{D}[\rho]_{|\bar{c}_1^n\hat{c}_1^n}} - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega} \\
&= H_\alpha^{\hat{f}_{\text{full}}}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_\rho + \min_{\bar{c}_1^n\hat{c}_1^n\in\widetilde{\Omega}} \hat{f}_{\text{full}}(\bar{c}_1^n\hat{c}_1^n) - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega} \\
&\geq \min_{\bar{c}_1^n\hat{c}_1^n\in\widetilde{\Omega}} \hat{f}_{\text{full}}(\bar{c}_1^n\hat{c}_1^n) - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega} \\
&= \min_{\bar{c}_1^n\hat{c}_1^n\in\widetilde{\Omega}} \sum_{j=1}^n \left(f_{|\bar{c}_1^{j-1}\hat{c}_j^{j-1}}(\bar{c}_j\hat{c}_j) + \kappa_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}} \right) - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega} \\
&\geq \left(\sum_{j=1}^n \min_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}} \kappa_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}} \right) + \min_{\bar{c}_1^n\hat{c}_1^n\in\widetilde{\Omega}} f_{\text{full}}(\bar{c}_1^n\hat{c}_1^n) - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega}, \quad (4.60)
\end{aligned}$$

where the first line follows from Lemma 4.3.9, the second line is by applying [41, Lemma B.5] along with the fact that $H_\alpha^\uparrow \geq H_\alpha$, the third line is due to Eq. (4.10), the fourth line is the result of Eq. (4.9), the fifth follows from Eq. (4.34), and the sixth line is just the substitution of $\hat{f}_{\text{full}}(\bar{c}_1^n\hat{c}_1^n)$ from Eq. (4.33). □

The above corollary still depends on the choice of QES. However, in the next section, we discuss potential approaches for obtaining the optimal choice of QES-s.

4.4 Simplified versions

In this section, we analyze the result of Corollary 4.3.1 in more detail when applied to fixed-length protocols. In particular, we show that in situations where the sequence of channels can be relaxed to a common channel, the result of Corollary 4.3.1 can be significantly simplified.

We begin by introducing the notion of the "convex range" of a set of channels, which is heavily used in the subsequent analysis.

Definition 4.4.1. (Convex range) Let $\{\mathcal{E}_j\}_{j=1}^n$ be any set of channels $\mathcal{E}_j \in \mathrm{CPTP}(\widetilde{Q}_j, Z_j' Z_j'' \dots)$ (for some finite set of output registers $Z_j', Z_j'', \dots$), such that the registers Z_j' (resp. $Z_j'', \dots$) can all be embedded in a common register Z' (resp. $Z'', \dots$). The **convex range** of these channels simply refers to the set of all mixtures of states of the form $\mathcal{E}_j[\omega]$ for some j and some $\omega \in \mathrm{S}_=(\widetilde{Q}_j)$, interpreting the mixtures as taking place in the embedding space.

We note that we did not specify the choice of embedding, as the subsequent analysis relying on this concept remains valid for any such choice of embedding.

In scenarios such as the set of channels studied in [41] for analyzing device-dependent EB-QKD, where all the channels $\mathcal{E}_j$ are isomorphic to each other, the convex range of the channels can be taken to be the range of any one of the channels. More generally, in many applications of EAT-style theorems, the single-round analysis is "indifferent" with respect to the specific choice of channel in the sequence. In many such cases, the single-round analysis already implicitly analyzes the convex range of all the channels, since in each round, an adversary could anyway be applying some mixture of attack strategies. Hence, most existing methods for analyzing single rounds in a protocol should also immediately generalize to the convex range; for instance, this is indeed the case for the analysis of device-dependent PM-QKD and EB-QKD using the GEAT in [90], as well as the analysis of device-independent QKD in [6].

With that in mind, we have the following result:

Theorem 4.4.1. *Let ρ be a state generated by a sequence of GEATT channels $\{\mathcal{M}_j\}_{j=1}^n$ (Definition 4.2.1). Take any $\alpha \in (1,2)$ and let $\widehat{\alpha} = 1/(2-\alpha)$. Suppose furthermore that $\rho = p_\Omega \rho_{|\Omega} + (1-p_\Omega)\rho_{|\overline{\Omega}}$ for some $p_\Omega \in (0,1]$ and normalized states $\rho_{|\Omega}, \rho_{|\overline{\Omega}}$. Let S_Ω be a convex set of probability distributions on the alphabet $\overline{\mathcal{C}} \times \widehat{\mathcal{C}}$, such that for all $\bar{c}_1^n \hat{c}_1^n$ with nonzero probability in $\rho_{|\Omega}$, the frequency distribution $\mathrm{freq}_{\bar{c}_1^n \hat{c}_1^n}$ lies in S_Ω. Then letting $\boldsymbol{\sigma}_{\overline{C}\widehat{C}}$ denote the distribution on $\overline{C}\widehat{C}$ induced by any state $\sigma_{\overline{C}\widehat{C}}$, we have*

$$H_\alpha^\uparrow(S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\rho_{|\Omega}} \geq n h_{\widehat{\alpha}} - \frac{\alpha}{\alpha - 1} \log \frac{1}{p_\Omega},$$

$$\text{where} \quad h_{\widehat{\alpha}} = \inf_{\mathbf{q} \in S_\Omega} \inf_{\nu \in \Sigma} \left(\frac{1}{\widehat{\alpha} - 1} D\left(\mathbf{q} \middle\| \boldsymbol{\nu}_{\overline{C}\widehat{C}}\right) - \sum_{\bar{c}\hat{c} \in \mathrm{supp}(\boldsymbol{\nu}_{\overline{C}\widehat{C}})} q(\bar{c}\hat{c}) D_{\widehat{\alpha}}\left(\nu_{SE\widetilde{E} \wedge \bar{c}\hat{c}} \middle\| \mathbb{I}_S \otimes \nu_{E\widetilde{E} \wedge \hat{c}}\right) \right), \tag{4.61}$$

where Σ denotes the convex range (Definition 4.4.1) of the channels $\{\mathcal{M}_j \otimes \mathrm{id}_{\widetilde{E}}\}_{j=1}^n$, with $\widetilde{E}$ being a register of large enough dimension to serve as a purifying register for any of the R_jE_j registers.

Furthermore, if there exists some channel $\mathcal{M} \in \mathrm{CPTP}(\widetilde{Q}, SE\overline{C}\widehat{C})$ such that Σ is equal to the set of output states of $\mathcal{M} \otimes \mathrm{id}_{\widetilde{E}}$ (i.e. $\left\{\mathcal{M}\left[\omega_{\widetilde{Q}\widetilde{E}}\right] \,\middle|\, \omega \in \mathrm{S}_=(\widetilde{Q}\widetilde{E})\right\}$), then for any purifying function $\mathtt{Pur}$ for $\widetilde{Q}$ onto $\widetilde{E}$ (Definition 4.2.2), we have

$$h_{\widehat{\alpha}} = \inf_{\mathbf{q}\in S_\Omega} \inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} \left(\frac{1}{\widehat{\alpha}-1} D\left(\mathbf{q}\middle\|\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right) - \sum_{\bar{c}\hat{c}\in\mathrm{supp}\left(\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right)} q(\bar{c}\hat{c}) D_{\widehat{\alpha}}\left(\nu^{\omega}_{SE\widetilde{E}\wedge\bar{c}\hat{c}}\middle\|\mathbb{I}_S \otimes \nu^{\omega}_{E\widetilde{E}\wedge\hat{c}}\right) \right), \quad \text{where} \quad \nu^{\omega} \coloneqq \mathcal{M}\left[\mathtt{Pur}\left(\omega_{\widetilde{Q}}\right)\right], \tag{4.62}$$

and the objective function in the above infimum is jointly convex in ω and $\mathbf{q}$.

It is worth noting that in cases where the set S_Ω is not convex, one can either take its convex hull or use the alternative analysis presented later in Corollary 4.4.1 and Lemma 4.4.5. Furthermore, we note that it is always possible to construct a channel such that Eq. (4.62) holds; this will be demonstrated later in Lemma 4.4.3.

Looking at the RHS of Eq. (4.61), we see that the $h_{\widehat{\alpha}}$ term is the so-called first-order correction term. In the next section, we analyze this first-order term in more detail to gain a better intuition. To that end, we first simplify the $h_{\widehat{\alpha}}$ term using the following lemmas (for ease of readability, we defer their proofs in Appendix A)

Lemma 4.4.1. *In Theorem 4.4.1, the following bound holds on $h_{\widehat{\alpha}}$:*

$$\begin{aligned} h_{\widehat{\alpha}} &\geq \inf_{\mathbf{q}\in S_\Omega} \inf_{\nu\in\Sigma} \left(\frac{1}{\widehat{\alpha}-1} D\left(\mathbf{q}\middle\|\boldsymbol{\nu}_{\overline{C}\widehat{C}}\right) + \sum_{\bar{c}\hat{c}\in\mathrm{supp}\left(\boldsymbol{\nu}_{\overline{C}\widehat{C}}\right)} q(\bar{c}\hat{c}) H_{\widehat{\alpha}}(S|E\widetilde{E})_{\nu_{|\bar{c}\hat{c}}} \right) \\ &= \inf_{\mathbf{q}\in S_\Omega} \inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} \left(\frac{1}{\widehat{\alpha}-1} D\left(\mathbf{q}\middle\|\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right) + \sum_{\bar{c}\hat{c}\in\mathrm{supp}\left(\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right)} q(\bar{c}\hat{c}) H_{\widehat{\alpha}}(S|E\widetilde{E})_{\nu^{\omega}_{|\bar{c}\hat{c}}} \right), \end{aligned} \tag{4.63}$$

and the objective function in the second line is jointly convex in ω and $\mathbf{q}$. If the $\overline{C}$ register

is trivial, then the above bound becomes an equality, i.e., we have

$$h_{\widehat{\alpha}} = \inf_{\mathbf{q}\in S_\Omega} \inf_{\nu\in\Sigma} \left(\frac{1}{\widehat{\alpha}-1} D\left(\mathbf{q}\middle\|\boldsymbol{\nu}_{\widehat{C}}\right) + \sum_{\hat{c}\in\mathrm{supp}(\boldsymbol{\nu}_{\widehat{C}})} q(\hat{c}) H_{\widehat{\alpha}}(S|E\widetilde{E})_{\nu_{|\hat{c}}} \right)$$

$$= \inf_{\mathbf{q}\in S_\Omega} \inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} \left(\frac{1}{\widehat{\alpha}-1} D\left(\mathbf{q}\middle\|\boldsymbol{\nu}^{\omega}_{\widehat{C}}\right) + \sum_{\hat{c}\in\mathrm{supp}\left(\boldsymbol{\nu}^{\omega}_{\widehat{C}}\right)} q(\hat{c}) H_{\widehat{\alpha}}(S|E\widetilde{E})_{\nu^{\omega}_{|\hat{c}}} \right). \tag{4.64}$$

Lemma 4.4.2. *In Theorem 4.4.1, the following bound holds on $h_{\widehat{\alpha}}$, for any $\alpha', \alpha'' \in (1,\infty)$ such that $\frac{\widehat{\alpha}}{\widehat{\alpha}-1} = \frac{\alpha'}{\alpha'-1} + \frac{\alpha''}{\alpha''-1}$:*

$$h_{\widehat{\alpha}} \geq \inf_{\mathbf{q}\in S_\Omega} \inf_{\nu\in\Sigma} \left(\frac{\alpha''}{\alpha''-1} D\left(\mathbf{q}\middle\|\boldsymbol{\nu}_{\overline{C}\widehat{C}}\right) + H_{\alpha'}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu} \right)$$

$$= \inf_{\mathbf{q}\in S_\Omega} \inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} \left(\frac{\alpha''}{\alpha''-1} D\left(\mathbf{q}\middle\|\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right) + H_{\alpha'}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^{\omega}} \right), \tag{4.65}$$

and the objective function in the second line is jointly convex in ω and $\mathbf{q}$.

4.4.1 First-order analysis

We begin by discussing the intuition behind the single-round term $h_{\widehat{\alpha}}$ in Lemmas 4.4.1–4.4.2. Given the simpler structure of the bound in Lemma 4.4.2, we use it for intuition purposes.

To obtain some intuition, we shall compare our bound against a scenario where the global state is an IID state and ignore the conditioning on Ω. In this case, for the conditional entropy of the global state ρ, we have

$$H^{\uparrow}_{\alpha'}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_{\rho^{\otimes n}} = H^{\uparrow}_{\alpha'}(S\overline{C}|\widehat{C}E\widetilde{E})_{\rho}. \tag{4.66}$$

Moreover, for this intuitive comparison, let us focus on the case where the state is such that the single-round probability distribution $\boldsymbol{\nu}_{\overline{C}\widehat{C}}$ lies in S_Ω, because it would be accepted with high probability in the asymptotic limit. In this case, the single-round quantity in the first-order term cannot be better than

$$\begin{aligned} &\inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} H_{\alpha'}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^{\omega}} \\ &\text{s.t.} \quad \boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}} \in S_\Omega, \end{aligned} \tag{4.67}$$

since the state would be accepted with high probability asymptotically. The optimization in Eq. (4.67) differs from the one in Lemma 4.4.2 in only one aspect: instead of imposing a hard constraint $\boldsymbol{\nu}_{\overline{C}\widehat{C}} \in S_\Omega$ as in (4.67), we have a soft constraint, which allows $\boldsymbol{\nu}_{\overline{C}\widehat{C}}$ to move outside of S_Ω at the cost of a penalty term, quantified by $D\left(\mathbf{q}\middle\|\boldsymbol{\nu}_{\overline{C}\widehat{C}}\right)$. Therefore, the optimization in Lemma 4.4.2 can be seen as a relaxation of Eq. (4.67). One important aspect of that lemma is that this relaxation already accounts for all finite-size and non-IID effects in the protocol. Another convenient aspect is that one only needs to specify the acceptance set S_Ω and then evaluate the bounds in those lemmas to find a bound on the key rate, without needing to deal with specifying the min-tradeoff function and the task of optimizing its choice, which were required in earlier versions of entropy accumulation. In particular, previous variations of entropy accumulation (the GEAT [89] and the original EAT [41]) had the nontrivial complication of evaluating and optimizing the choice of a min-tradeoff function, whereas with this method, such a complication no longer arises. On another point, if needed, one can straightforwardly use a coarse-grained version of the distributions $\mathbf{q}, \boldsymbol{\nu}^\omega_{\overline{C}\widehat{C}}$ instead of the fine-grained distribution, since mathematically, coarse-graining is a stochastic map on the distribution. Thus, by the data-processing inequality for KL divergence, we have that if $\mathcal{E}$ is the coarse-graining map, then $D\left(\mathcal{E}[\mathbf{q}]\middle\|\mathcal{E}[\boldsymbol{\nu}^\omega_{\overline{C}\widehat{C}}]\right)$ lower bounds $D\left(\mathbf{q}\middle\|\boldsymbol{\nu}^\omega_{\overline{C}\widehat{C}}\right)$.

Before explaining other implications of having a bound as the one in Lemma 4.4.2, we note that for any reasonable choice of the acceptance set S_Ω, for classical S, the $h_{\widehat{\alpha}}$ term is non-negative. The reason is that any reasonable choice of S_Ω should yield a positive solution to Eq. (4.67). Therefore, for any feasible point of that optimization, either $\boldsymbol{\nu}^\omega_{\overline{C}\widehat{C}} \in S_\Omega$ and so $H_{\alpha'}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega} > 0$, or we have $\boldsymbol{\nu}^\omega_{\overline{C}\widehat{C}} \notin S_\Omega$ and so $D\left(\mathbf{q}\middle\|\boldsymbol{\nu}^\omega_{\overline{C}\widehat{C}}\right) > 0$.

We highlight that another advantage of Lemma 4.4.2 is that the single-round entropy is a Rényi entropy as opposed to the previous variations of entropy accumulation [41, 89]. Therefore, for instance, in principle, we can accumulate min-entropy ($\alpha \to \infty$) or collision entropy ($\alpha = 2$). However, it does not mean we are necessarily required to evaluate single-round Rényi entropies. This is because we can still apply continuity arguments to our single-round term to convert it to von Neumann entropy. More explicitly, let us write $\alpha' = 1 + \mu'$, $\alpha'' = 1 + \mu''$ for some μ', μ'' for Rényi parameters in Lemma 4.4.2, and recalling $\widehat{\alpha} = \frac{1}{2-\alpha}$, then we have:

$$\begin{aligned} \widehat{\alpha} &= 1 + \frac{1}{1 + \frac{1}{\mu'} + \frac{1}{\mu''}} = 1 + \frac{\mu'\mu''}{\mu'\mu'' + \mu' + \mu''}, \\ \alpha &= 1 + \frac{1}{2 + \frac{1}{\mu'} + \frac{1}{\mu''}} = 1 + \frac{\mu'\mu''}{2\mu'\mu'' + \mu' + \mu''}. \end{aligned} \tag{4.68}$$

Therefore, we see that as $\alpha', \alpha'' \to 1$, we have $\widehat{\alpha}, \alpha \to 1$. This is a convenient result, as it allows us to employ continuity approaches from [41] or the more refined bound in [37] to express our result in terms of known results based on the smooth min-entropy of the global state and the von Neumann entropy of the single-round state.

To do so, let us convert the single-round Rényi entropy to the von Neumann entropy using the following continuity argument:

$$H_\alpha(S\overline{C}|\widehat{C}E\widetilde{E}) \geq H(S\overline{C}|\widehat{C}E\widetilde{E}) - g_{\text{cont}}(\alpha), \tag{4.69}$$

where $g_{\text{cont}}(\alpha)$ is such that $\lim_{\alpha\to 1} g_{\text{cont}}(\alpha) = 0$. Moreover, explicit bounds on $g_{\text{cont}}(\alpha)$ have already been provided in [37, 41] for α sufficiently close to 1 as:

$$\begin{aligned} g_{\text{cont}}(\alpha) &\leq (\alpha-1)^2 \log^2\left(1 + 2\dim(S\overline{C})\right) && \text{Proved in [41]} \\ g_{\text{cont}}(\alpha) &\leq (\alpha-1)\frac{\ln 2}{2}\log^2\left(1 + 2\dim^\zeta(S\overline{C})\right) + O\left((\alpha-1)^2\right), && \text{Proved in [37]} \end{aligned} \tag{4.70}$$

where in the second bound $\zeta = 2$ if S is a quantum register, and $\zeta = 1$ if S is a classical register. Note that when S is a classical register, the latter is tighter than the former. On the other hand, using known relations between smooth min-entropies and Rényi entropies, we have the following conversion [41]:

$$H_\infty^{\uparrow,\varepsilon}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n) \geq H_\alpha^\uparrow(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n) - \frac{\vartheta_\varepsilon}{\alpha-1}, \text{ where } \vartheta_\varepsilon \coloneqq \log\left(\frac{1}{1-\sqrt{1-\varepsilon^2}}\right) \leq \log\frac{2}{\varepsilon^2}, \tag{4.71}$$

for any $\alpha \in (1, 2]$ and $\varepsilon \in (0, 1)$.

Substituting Eq. (4.69) into the single-round term in Lemma 4.4.2, and Eq. (4.71) into the global entropy in Theorem 4.4.1, and combining them, we have the following conversion:

$$\begin{aligned} H_\infty^{\uparrow,\varepsilon}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_{\rho_{|\Omega}} \geq \inf_{\mathbf{q}\in S_\Omega}\inf_{\omega\in \mathrm{S}_=(\widetilde{Q})}&\left(\frac{\alpha''}{\alpha''-1}D\left(\mathbf{q}\middle\|\boldsymbol{\nu}^\omega_{\widehat{C}\widehat{C}}\right) + H(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega}\right)n \\ &- g_{\text{cont}}(\alpha')n - \frac{\vartheta_\varepsilon}{\alpha-1} - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega}, \end{aligned} \tag{4.72}$$

where g_{cont} can be bounded using (4.70). With this bound, let us choose $\alpha', \alpha'' = 1 + \Theta(1/\sqrt{n})$; then the result in (4.68) states that the same scaling also holds for $\widehat{\alpha}, \alpha$. (i.e., $\widehat{\alpha}, \alpha = 1 + \Theta(1/\sqrt{n})$, therefore we have:

$$H_\infty^{\uparrow,\varepsilon}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_{\rho_{|\Omega}} \geq \inf_{\mathbf{q}\in S_\Omega}\inf_{\omega\in \mathrm{S}_=(\widetilde{Q})}\left(\Theta(\sqrt{n})D\left(\mathbf{q}\middle\|\boldsymbol{\nu}^\omega_{\widehat{C}\widehat{C}}\right) + H(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega}\right)n - \Theta(\sqrt{n}). \tag{4.73}$$

We highlight that the $\Theta(\sqrt{n})$ correction term has the same scaling as what one would obtain from the AEP in the IID case. Furthermore, the first-order term is again a relaxed optimization, similar to what we had with the Rényi single-round analysis. Moreover, given the $\sqrt{n}$ prefactor on the KL divergence term, we see that as n increases, this term increasingly enforces the condition $\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}} \in S_\Omega$, effectively approaching a hard constraint in the asymptotic limit. Therefore, the bound in (4.73) converges to what is expected asymptotically.

So far, we have predominantly discussed Lemma 4.4.2. However, we now state one of its main drawbacks, which motivates the use of Lemma 4.4.1 instead. Let $\alpha', \alpha'' = 1 + \mu$ for some μ. Then, Eq. (4.68) shows that $\alpha, \widehat{\alpha} = 1 + \mu/2 + \mathcal{O}(\mu^2)$. In other words, the distance between α', α'' and 1 is almost twice the distance of $\alpha, \widehat{\alpha}$ to 1. This implies that, for example, the effect of the KL divergence term roughly halves compared to the case where no such change in the Rényi parameter had occurred.

On the other hand, if we instead use Lemma 4.4.1, the only change in the Rényi parameter that we encounter is the change from α to $\widehat{\alpha}$, which constitutes a third-order correction (with respect to key rate) when α is close to 1. However, using this lemma raises another form of difficulty, which is dealing with the linear combination of entropies $\sum_{\bar{c}\hat{c}} q(\bar{c}\hat{c}) H_{\widehat{\alpha}}(S|E\widetilde{E})$. We can also convert the bound in this lemma to a bound where the global entropy is a smooth min-entropy and the single-round entropies are von Neumann entropies by using the same substitutions from Eqs. (4.69) and (4.71) into Lemma 4.4.1 to obtain the following result:

$$H_\infty^{\uparrow,\varepsilon}(S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\rho_{|\Omega}} \geq \inf_{\mathbf{q}\in S_\Omega} \inf_{\omega \in \mathrm{S}_=(\widetilde{Q})} \left(\frac{1}{\widehat{\alpha}-1} D\left(\mathbf{q} \middle\| \boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right) + \sum_{\bar{c}\hat{c}\in \mathrm{supp}\left(\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right)} q(\bar{c}\hat{c}) H(S|E\widetilde{E})_{\nu^{\omega}_{|\bar{c}\hat{c}}} \right) n$$
$$- g_{\mathrm{cont}}(\widehat{\alpha}) n - \frac{\vartheta_\varepsilon}{\alpha - 1} - \frac{\alpha}{\alpha - 1} \log \frac{1}{p_\Omega}, \tag{4.74}$$

where g_{cont} again can be bounded by either of the bounds in (4.70). We note that if we choose the Rényi parameters $\widehat{\alpha}, \alpha = 1 + \Theta(1/\sqrt{n})$, we would obtain the same scaling as Eq. (4.73).

4.4.2 Infrequent-sampling channels

In many applications of entropy accumulation, the sequence of channels has a particular structure termed "infrequent-sampling channels" [41], for which the single-round bound

can be simplified. In this section, we analyze the bound from the previous section for these channels. We begin by formally defining the infrequent-sampling channels.

Definition 4.4.2. A channel $\mathcal{M}$ with some classical registers $\overline{C}\widehat{C}$ in its output is said to be an **infrequent-sampling channel** if there is a constant **test probability** value $\gamma \in (0,1)$ such that

$$\mathcal{M} = (1-\gamma)\mathcal{M}^{\text{gen}} + \gamma\mathcal{M}^{\text{test}}, \tag{4.75}$$

for some **generation** and **test** channels $\mathcal{M}^{\text{gen}}, \mathcal{M}^{\text{test}}$, where $\mathcal{M}^{\text{gen}}$ satisfies the property that it always sets $(\overline{C}, \widehat{C}) = (\perp, \perp)$ for some special symbol $\perp$, while $\mathcal{M}^{\text{test}}$ never sets $(\overline{C}, \widehat{C}) = (\perp, \perp)$.

In other words, an infrequent-sampling channel probabilistically chooses, with probability γ, whether to implement $\mathcal{M}^{\text{test}}$ or $\mathcal{M}^{\text{gen}}$.

These channels exhibit several desirable properties. For example, when applying Theorem 4.4.1 to them—assuming they all share the same testing probability γ— their convex range is described by states of the form $\nu = (1-\gamma)\nu^{\text{gen}} + \gamma\nu^{\text{test}}$, where $\nu^{\text{gen}}(\perp,\perp) = 1$ and $\nu^{\text{test}}(\perp,\perp) = 0$. Furthermore, to derive results similar to Eq. (4.62), one can construct a single infrequent-sampling channel with the same testing probability, effectively reproducing the outcome of Eq. (4.62) for these channels.

In device-dependent protocols of the form studied in Chapter 3, a fixed value is assigned to the S register in test rounds. Thus, we obtain the following simplification:

$$\sum_{\bar{c}\hat{c}\in\text{supp}\left(\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right)} q(\bar{c}\hat{c}) H_{\widehat{\alpha}}(S|E\widetilde{E})_{\nu^{\omega}_{|\bar{c}\hat{c}}} = q(\perp\perp) H_{\widehat{\alpha}}(S|E\widetilde{E})_{\nu^{\omega}_{|\perp\perp}}, \tag{4.76}$$

where we used the fact that for test rounds, $H_{\widehat{\alpha}}(S|E\widetilde{E})\nu^{\omega}_{|\perp\perp} = 0$.[15] Moreover, if we set $\overline{C}$ to be trivial, which is usually the case for device-dependent protocols, we would get the following simplifications of Lemma 4.4.1:

$$h_{\widehat{\alpha}} = \inf_{\mathbf{q}\in S_{\Omega}} \inf_{\omega\in \mathrm{S}_{=}(\widetilde{Q})} \left(\frac{1}{\widehat{\alpha}-1} D\left(\mathbf{q} \middle\| \boldsymbol{\nu}^{\omega}_{\widehat{C}}\right) + q(\perp) H_{\widehat{\alpha}}(S|E\widetilde{E})_{\nu^{\omega}_{|\perp}} \right), \tag{4.77}$$

with the objective function being jointly convex in $(\mathbf{q}, \omega)$.

[15] We highlight that if we are analyzing a protocol where such an assumption does not hold, we would instead obtain an inequality as $\sum_{\bar{c}\hat{c}\in\text{supp}\left(\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right)} q(\bar{c}\hat{c}) H_{\widehat{\alpha}}(S|E\widetilde{E})\nu^{\omega}_{|\bar{c}\hat{c}} \geq q(\perp\perp) H_{\widehat{\alpha}}(S|E\widetilde{E})\nu^{\omega}_{|\perp\perp}$ for classical S.

Moreover, the KL divergence term $D\left(\mathbf{q}\middle\|\boldsymbol{\nu}^{\omega}_{\widehat{CC}}\right)$ can also be rewritten in a form that makes the testing probability dependency clearer. To do so, we express the distributions $\mathbf{q}, \boldsymbol{\nu}^{\omega}_{\widehat{CC}}$ as follows:

$$\mathbf{q} = \begin{pmatrix} 1-\gamma_q \\ \gamma_q \breve{\mathbf{q}} \end{pmatrix}, \quad \boldsymbol{\nu}^{\omega}_{\widehat{CC}} = \begin{pmatrix} 1-\gamma \\ \gamma \breve{\boldsymbol{\nu}}^{\omega}_{\widehat{CC}} \end{pmatrix}, \tag{4.78}$$

where $\gamma_q \coloneqq 1 - q(\perp\perp)$, and $\breve{\mathbf{q}}, \breve{\boldsymbol{\nu}}^{\omega}_{\widehat{CC}}$ are normalized distributions. We now rewrite the KL divergence in terms of those expressions

$$\begin{aligned} D\left(\mathbf{q}\middle\|\boldsymbol{\nu}^{\omega}_{\widehat{CC}}\right) &= (1-\gamma_q)\log\frac{1-\gamma_q}{1-\gamma} + \sum_{\bar{c}\hat{c}\neq\perp\perp} \gamma_q \breve{q}(\bar{c}\hat{c}) \log \frac{\gamma_q \breve{q}(\bar{c}\hat{c})}{\gamma \breve{\nu}^{\omega}(\bar{c}\hat{c})} \\ &= (1-\gamma_q)\log\frac{1-\gamma_q}{1-\gamma} + \gamma_q \log\frac{\gamma_q}{\gamma} + \gamma_q D\left(\breve{\mathbf{q}}\middle\|\breve{\boldsymbol{\nu}}^{\omega}_{\widehat{CC}}\right) \\ &= D\left(\mathbf{r}\|\mathbf{s}\right) + \gamma_q D\left(\breve{\mathbf{q}}\middle\|\breve{\boldsymbol{\nu}}^{\omega}_{\widehat{CC}}\right) \text{ where } \mathbf{r} = \begin{pmatrix} 1-\gamma_q \\ \gamma_q \end{pmatrix}, \ \mathbf{s} = \begin{pmatrix} 1-\gamma \\ \gamma \end{pmatrix}. \end{aligned} \tag{4.79}$$

4.4.3 Convexity of relaxed lower bounds

We highlight here that even though we proved the convexity of the objective function in Theorem 4.4.1 and Lemmas 4.4.1–4.4.2, when we relax the bounds in those expressions, we may no longer have a convex objective function. For future applications, we analyze an example of such a case and prove its convexity in certain special instances. We take the bound in Lemma 4.4.1, apply the continuity argument in Eq. (4.69), and retain only the generation component, which leads to an inequality version of Eq. (4.76) where $\widehat{\alpha} = 1$. This results in a lower bound on Lemma 4.4.1 with a KL divergence term $D\left(\mathbf{q}\middle\|\boldsymbol{\nu}^{\omega}_{\widehat{CC}}\right)$, an entropic term $q(\perp\perp)H(S|E\widetilde{E})_{\nu^{\omega}_{|\perp\perp}}$, and some correction terms that are irrelevant for convexity considerations. To analyze the convexity of this relaxed bound, we first note that the KL divergence is already convex. Thus, it suffices to show the convexity of the entropic term. To do so, we first note that according to [132], for device-dependent protocols, it is possible to construct $\mathcal{Z}, \mathcal{G}$ maps from $\mathcal{M}^{\text{gen}}$ such that $H(S|E\widetilde{E})_{\nu^{\omega}_{|\perp\perp}} = D\left(\mathcal{G}[\omega]\|\mathcal{Z}\circ\mathcal{G}[\omega]\right)$, which is convex with respect to input states ω [117, Chapter 4.1.2, Property (IXb)]. However, to show joint convexity with respect to ω and $q(\perp,\perp)$, we proceed as follows:

- First, we note that the Umegaki divergence has the property that for any $t > 0$, $D\left(\mathcal{G}[\omega/t]\|\mathcal{Z}\circ\mathcal{G}[\omega/t]\right) = D\left(\mathcal{G}[\omega]\|\mathcal{Z}\circ\mathcal{G}[\omega]\right)$.

- Second, for any convex function $f(\omega)$, its perspective function defined as $g(\omega, t) \coloneqq tf(\omega/t)$ is jointly convex in (ω, t) when $t > 0$ [18, Chapter 3.2.6].

Therefore, for any convex function $f(\omega)$ such that $f(t\omega) = \omega$, we have $g(\omega, t) = tf(\omega)$, which is jointly convex in (ω, t). As a result, $q(\perp\perp)D\left(\mathcal{G}[\omega] \| \mathcal{Z} \circ \mathcal{G}[\omega]\right)$ is jointly convex in $(q(\perp\perp), \omega)$. However, it is unclear how to generalize the above argument to cases where we have linear combinations of entropies, such as $\sum_{\bar{c}\hat{c}} q(\bar{c}\hat{c}) H(S|E\widetilde{E})_{\nu^{\omega}_{|\bar{c}\hat{c}}}$, instead.

4.4.4 Proof of Theorem 4.4.1

We begin by deriving a consequence of Corollary 4.3.1, which, combined with Lemma 4.4.5, can serve as a substitute for Theorem 4.4.1 in the case where the acceptance set S_Ω is not convex.

Corollary 4.4.1. *Consider the same conditions and notation as in Theorem 4.4.1, except without requiring that S_Ω is convex. Then for any QES f on $\overline{C}\widehat{C}$, we have*

$$\begin{aligned} H_\alpha^\uparrow(S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\rho_{|\Omega}} &\geq \inf_{\mathbf{q} \in S_\Omega} \inf_{\nu \in \Sigma} \left(H_{\widehat{\alpha}}^f(S\overline{C}|\widehat{C}E\widetilde{E})_\nu + \mathbf{f} \cdot \mathbf{q} \right) n - \frac{\alpha}{\alpha - 1} \log \frac{1}{p_\Omega} \\ &= \inf_{\mathbf{q} \in S_\Omega} \inf_{\omega \in \mathrm{S}_=(\widetilde{Q})} \left(H_{\widehat{\alpha}}^f(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega} + \mathbf{f} \cdot \mathbf{q} \right) n - \frac{\alpha}{\alpha - 1} \log \frac{1}{p_\Omega}, \end{aligned} \tag{4.80}$$

where $\mathbf{f}$ denotes the function f viewed equivalently as a tuple in $\mathbb{R}^{|\overline{\mathcal{C}} \times \widehat{\mathcal{C}}|}$. Furthermore, in the second line, $H_{\widehat{\alpha}}^f(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega}$ is convex in ω, and hence the objective function in the infimum is jointly convex in ω and $\mathbf{q}$.

Proof. Simply set all the QES-s $f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ in the Corollary 4.3.1 statement to be equal to f.

Then the bound (4.59) becomes

$$
\begin{aligned}
&H_\alpha^\uparrow(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_{\rho_{|\Omega}} \\
&\geq \min_{\bar{c}_1^n\hat{c}_1^n\in\widetilde{\Omega}} \sum_{j=1}^n \left(f(\bar{c}_j\hat{c}_j) + \inf_{\nu\in\Sigma_j} H_{\widehat{\alpha}}^f(S_j\overline{C}_j|\widehat{C}_j E_j\widetilde{E})_\nu \right) - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega} \\
&= \min_{\bar{c}_1^n\hat{c}_1^n\in\widetilde{\Omega}} \sum_{\bar{c}\hat{c}\in\overline{\mathcal{C}}\times\widehat{\mathcal{C}}} f(\bar{c}\hat{c})\,\mathrm{freq}_{\bar{c}_1^n\hat{c}_1^n}(\bar{c}\hat{c})n + \sum_{j=1}^n \inf_{\nu\in\Sigma_j} H_{\widehat{\alpha}}^f(S_j\overline{C}_j|\widehat{C}_j E_j\widetilde{E})_\nu - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega} \\
&\geq \inf_{\mathbf{q}\in S_\Omega} \mathbf{f}\cdot\mathbf{q}\,n + \sum_{j=1}^n \inf_{\nu\in\Sigma_j} H_{\widehat{\alpha}}^f(S_j\overline{C}_j|\widehat{C}_j E_j\widetilde{E})_\nu - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega} \\
&\geq \inf_{\mathbf{q}\in S_\Omega}\inf_{\nu\in\Sigma} \left(H_{\widehat{\alpha}}^f(S\overline{C}|\widehat{C}E\widetilde{E})_\nu + \mathbf{f}\cdot\mathbf{q} \right) n - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega},
\end{aligned}
\tag{4.81}
$$

where in the third line we rewrite the sum of $f(\bar{c}_j\hat{c}_j)$ values in terms of the frequency distribution $\mathrm{freq}_{\bar{c}_1^n\hat{c}_1^n}$, the forth line holds by expressing the minimization over $\bar{c}_1^n\hat{c}_1^n \in \widetilde{\Omega}$ as a minimization over its induced frequency distributions and then relaxing the domain to S_Ω (using the first condition on S_Ω), and the fifth line follows simply from the fact that every state $\nu \in \Sigma_j$ also lies in the convex range Σ by definition.

The second line in (4.80) follows straightforwardly from Lemma 4.3.8; from that lemma, since $\alpha > 1$ and thus $\widehat{\alpha} > 1$, we can conclude that $H_{\widehat{\alpha}}^f(S\overline{C} \mid \widehat{C}E\widetilde{E})_{\nu^\omega}$ is convex in ω, as claimed. □

As the next step toward proving Theorem 4.4.1, we show that for any finite set of channels, there exists a single channel whose range is equal to the convex hull of the ranges of that set.

Lemma 4.4.3. *Let $\{\mathcal{M}_j\}_{j=1}^n$ be any set of channels $\mathcal{M}_j \in \mathrm{CPTP}(\widetilde{Q}_j, Z_j'Z_j''\dots)$ (for some finite number of output registers $Z_j', Z_j'', \dots$), such that the registers Z_j' (resp. $Z_j'', \dots$) can all be embedded in a common register Z' (resp. $Z'', \dots$). Define $\widetilde{Q}$ to be a register with Hilbert space $\mathcal{H}_{\widetilde{Q}} = \bigoplus_{j=1}^n \mathcal{H}_{\widetilde{Q}_j}$, and write $Z_j \coloneqq Z_j'Z_j''\dots$ and $Z \coloneqq Z'Z''\dots$ for brevity.*

Define a channel $\mathcal{M} \in \mathrm{CPTP}(\widetilde{Q}, Z)$ as follows: it first performs a projective measurement where outcome j corresponds to the projector onto $\mathcal{H}_{\widetilde{Q}_j}$, then conditioned on the outcome value j, implements the channel $\mathcal{M}_j$ and embeds the output in Z. Then the convex range Σ of $\{\mathcal{M}_j\}_{j=1}^n$ is equal to the range of $\mathcal{M}$, i.e. we have $\Sigma = \left\{ \mathcal{M}\left[\omega_{\widetilde{Q}}\right] \;\middle|\; \omega \in \mathrm{S}_=(\widetilde{Q}) \right\}$. Furthermore, for any register $\widetilde{E}$, the convex range of $\{\mathcal{M}_j \otimes \mathrm{id}_{\widetilde{E}}\}_{j=1}^n$ is equal to the range of $\mathcal{M} \otimes \mathrm{id}_{\widetilde{E}}$.

Proof. $\Sigma \subseteq \left\{\mathcal{M}\left[\omega_{\widetilde{Q}}\right]; \middle| ; \omega \in \mathrm{S}_{=}(\widetilde{Q})\right\}$: Observe that any state $\sigma \in \Sigma$ has the form $\sigma = \sum_j p_j \mathcal{M}_j\left[\omega^{(j)}_{\widetilde{Q}_j}\right]$ for some probability distribution $\mathbf{p}$ and states $\omega^{(j)}_{\widetilde{Q}_j}$. Therefore, by the definition of $\mathcal{M}$, we have:

$$\sigma = \mathcal{M}\left[\bigoplus_j p_j \omega^{(j)}_{\widetilde{Q}_j}\right]. \tag{4.82}$$

Setting $\omega_{\widetilde{Q}} \coloneqq \bigoplus_j p_j \omega^{(j)}_{\widetilde{Q}_j}$, we have shown $\subseteq$ direction.

$\Sigma \supseteq \left\{\mathcal{M}\left[\omega_{\widetilde{Q}}\right] \;\middle|\; \omega \in \mathrm{S}_{=}(\widetilde{Q})\right\}$: Note that for any state $\sigma \in \left\{\mathcal{M}\left[\omega_{\widetilde{Q}}\right]; \middle| ; \omega \in \mathrm{S}_{=}(\widetilde{Q})\right\}$, due to the first action of $\mathcal{M}$, which is the projective measurement, the state at that step is of the form $\bigoplus_j p_j \omega^{(j)}_{\widetilde{Q}_j}$, which then, by definition, has the form $\sum_j p_j \mathcal{M}_j\left[\omega^{(j)}_{\widetilde{Q}_j}\right]$. Therefore, we have $\sigma \in \Sigma$, and thus the containment holds.

The statement for $\{\mathcal{M}_j \otimes \mathrm{id}_{\widetilde{E}}\}_{j=1}^n$ and $\mathcal{M} \otimes \mathrm{id}_{\widetilde{E}}$ follows a similar argument, noting that the projection operator acts only on the $\widetilde{Q}$ register. □

We continue towards the proof of Theorem 4.4.1 by deriving a lemma that serves as a critical tool in the proof of that theorem.

Lemma 4.4.4. *Consider any $\alpha \in (1,\infty)$ and any $\rho \in \mathrm{S}_{=}(QQ'\overline{C}\widehat{C})$ such that $\overline{C}, \widehat{C}$ are classical with alphabets $\overline{\mathcal{C}}, \widehat{\mathcal{C}}$. Then letting $\mathbb{P}_{\overline{\mathcal{C}}\widehat{\mathcal{C}}}$ denote the set of probability distributions on $\overline{\mathcal{C}} \times \widehat{\mathcal{C}}$, the pair of functions $\left(G_{\alpha,\rho}, G^*_{\alpha,\rho}\right)$ on $\mathbb{R}^{|\overline{\mathcal{C}}\times\widehat{\mathcal{C}}|}$ defined as*

$$\begin{aligned} G_{\alpha,\rho}(\mathbf{f}) &\coloneqq -H^f_\alpha(Q\overline{C}|\widehat{C}Q')_\rho, \\ G^*_{\alpha,\rho}(\boldsymbol{\lambda}) &\coloneqq \begin{cases} \frac{1}{\alpha-1} D\left(\boldsymbol{\lambda}\middle\|\boldsymbol{\rho}_{\overline{C}\widehat{C}}\right) - \sum_{\bar{c}\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})} \lambda(\bar{c}\hat{c}) D_\alpha\left(\rho_{QQ'\wedge\bar{c}\hat{c}} \| \mathbb{I}_Q \otimes \rho_{Q'\wedge\hat{c}}\right) & \text{if } \boldsymbol{\lambda} \in \mathbb{P}_{\overline{\mathcal{C}}\widehat{\mathcal{C}}}, \\ +\infty & \text{otherwise,} \end{cases} \end{aligned} \tag{4.83}$$

are convex functions that are convex conjugates to each other, where in the $G_{\alpha,\rho}$ formula we recall that any tuple $\mathbf{f} \in \mathbb{R}^{|\overline{\mathcal{C}}\times\widehat{\mathcal{C}}|}$ is equivalent to a QES on $\overline{C}\widehat{C}$, and in the $G^*_{\alpha,\rho}$ formula we recall that $\boldsymbol{\rho}_{\overline{C}\widehat{C}}$ denotes the distribution induced by ρ on $\overline{C}\widehat{C}$.[16]

[16]Note that in the $G^*_{\alpha,\rho}$ formula, the first case can still take value $+\infty$ for some values of $\boldsymbol{\lambda}$, due to the $D\left(\boldsymbol{\lambda}\middle\|\boldsymbol{\rho}_{\overline{C}\widehat{C}}\right)$ term. Specifically, this happens whenever $\mathrm{supp}(\boldsymbol{\lambda}) \not\subseteq \mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$.

Proof. We begin by showing that $G_{\alpha,\rho}(\mathbf{f})$ is a convex function of $\mathbf{f}$. Let $\tilde{\mathbf{f}}, \widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\rho}}_{\overline{C}\widehat{C}}$ denote the restrictions of $\mathbf{f}, \boldsymbol{\lambda}, \boldsymbol{\rho}_{\overline{C}\widehat{C}}$, respectively, to the support of $\boldsymbol{\rho}_{\overline{C}\widehat{C}}$, i.e., $\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$. Furthermore, let $\widetilde{\mathbf{D}}_\alpha$ denote the tuple of values $\{D_\alpha\left(\rho_{QQ'\wedge\bar{c}\hat{c}} \| \mathbb{I}_Q \otimes \rho_{Q'\wedge\hat{c}}\right)\}_{\bar{c}\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})}$. Then, note that $H^f_\alpha(Q\overline{C}|\widehat{C}Q')\rho$ can be written in terms of the log-sum-exponential function [18] as:

$$\begin{aligned} G_{\alpha,\rho}(\mathbf{f}) &= \frac{1}{(\alpha-1)\ln 2} \ln\left(\sum_{\bar{c}\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})} \rho(\bar{c}\hat{c}) e^{(\alpha-1)(\ln 2)\left(f(\bar{c}\hat{c})+D_\alpha\left(\rho_{QQ'\wedge\bar{c}\hat{c}}\|\mathbb{I}_Q\otimes\rho_{Q'\wedge\hat{c}}\right)\right)}\right) \\ &= \frac{1}{(\alpha-1)\ln 2} \mathrm{logsumexp}\left(\ln\left(\widetilde{\boldsymbol{\rho}}_{\overline{C}\widehat{C}}\right) + (\alpha-1)(\ln 2)\left(\tilde{\mathbf{f}} + \widetilde{\mathbf{D}}_\alpha\right)\right), \end{aligned} \tag{4.84}$$

where $\ln\left(\widetilde{\boldsymbol{\rho}}_{\overline{C}\widehat{C}}\right)$ is the elementwise logarithm of $\widetilde{\boldsymbol{\rho}}_{\overline{C}\widehat{C}}$. Convexity follows by noting that $\alpha > 1$, the log-sum-exponential function is convex [18], and $G_{\alpha,\rho}(\mathbf{f})$ is constant with respect to any $\mathbf{f}$ component outside $\tilde{\mathbf{f}}$, convexity follows. Since $G_{\alpha,\rho}(\mathbf{f})$ is convex, taking the convex conjugate twice returns the original function [18]. Moreover, by definition its convex conjugate $G^*_{\alpha,\rho}(\boldsymbol{\lambda})$ is a convex function in $\boldsymbol{\lambda}$.

We now turn to proving that the convex conjugate of $G_{\alpha,\rho}(\mathbf{f})$ is the function claimed in 4.83. By Definition 2.3.5, $G^*_{\alpha,\rho}$ is the convex conjugate of $G_{\alpha,\rho}(\mathbf{f})$ if and only if the following relation holds:

$$G^*_{\alpha,\rho}(\boldsymbol{\lambda}) = \sup_{\mathbf{f}\in\mathbb{R}^{|\overline{C}\times\widehat{C}|}} \left(\boldsymbol{\lambda}\cdot\mathbf{f} - G_{\alpha,\rho}(\mathbf{f})\right). \tag{4.85}$$

First, consider the case where $\boldsymbol{\lambda}$ is such that $\mathrm{supp}(\boldsymbol{\lambda}) \not\subseteq \mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$. In that case, the above optimization becomes $+\infty$. To prove this, we first note that the support condition is equivalent to the following statement:

$$\exists \bar{c}\hat{c} \notin \mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}}) \text{ s.t. } \lambda(\bar{c}\hat{c}) \neq 0. \tag{4.86}$$

Observe that $G_{\alpha,\rho}(\mathbf{f})$ is independent of $\bar{c}\hat{c}$ that satisfy the above condition. Hence, without loss of generality, depending on the sign of $\lambda(\bar{c}\hat{c})$, we either take $f(\bar{c}\hat{c}) \to -\infty$ or $f(\bar{c}\hat{c}) \to +\infty$.[17]

We now focus on the cases where $\mathrm{supp}(\boldsymbol{\lambda}) \subseteq \mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$. Let $\widetilde{G}_{\alpha,\rho}(\tilde{\mathbf{f}})$ denote the value

[17] We highlight that this is indeed compatible with the claim in the lemma, since for such λ, either $\boldsymbol{\lambda} \in \mathbb{P}_{\overline{C}\widehat{C}}$, in which case the KL divergence term becomes $+\infty$, or $\boldsymbol{\lambda} \notin \mathbb{P}_{\overline{C}\widehat{C}}$, which by definition is $+\infty$.

of the last line in Eq. (4.84). Then, we have:

$$\begin{aligned} G^*_{\alpha,\rho}(\boldsymbol{\lambda}) &= \sup_{\mathbf{f}\in\mathbb{R}^{|\overline{\mathcal{C}}\times\widehat{\mathcal{C}}|}} \left(\boldsymbol{\lambda}\cdot\mathbf{f} - G_{\alpha,\rho}(\mathbf{f})\right) \\ G^*_{\alpha,\rho}(\boldsymbol{\lambda}) &= \sup_{\mathbf{f}\in\mathbb{R}^{|\overline{\mathcal{C}}\times\widehat{\mathcal{C}}|}} \left(\widetilde{\boldsymbol{\lambda}}\cdot\tilde{\mathbf{f}} - \widetilde{G}_{\alpha,\rho}(\tilde{\mathbf{f}})\right) \\ &= \sup_{\mathbf{g}\in\mathbb{R}^{\left|\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})\right|}} \left(\widetilde{\boldsymbol{\lambda}}\cdot\mathbf{g} - \widetilde{G}_{\alpha,\rho}(\mathbf{g})\right), \end{aligned} \tag{4.87}$$

where, in the second line, we use the fact that by the support condition we have $\boldsymbol{\lambda}\cdot\mathbf{f} = \widetilde{\boldsymbol{\lambda}}\cdot\tilde{\mathbf{f}}$, and $G_{\alpha,\rho}(\mathbf{f}) = \widetilde{G}_{\alpha,\rho}(\tilde{\mathbf{f}})$. The third line holds since the objective depends only on the restricted values $\tilde{\mathbf{f}}$.

The convexity of $G_{\alpha,\rho}(\mathbf{f})$ ensures the concavity of the objective function in the last line of Eq. (4.87). Therefore, any stationary point is a global maximum. Thus, by differentiating the objective function, we see that for any $\mathbf{g}\in\mathbb{R}^{\left|\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})\right|}$, it is a stationary point if and only if it satisfies

$$\forall \bar{c}\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}}),\quad \lambda(\bar{c}\hat{c}) = \frac{\rho(\bar{c}\hat{c})e^{(\alpha-1)(\ln 2)\left(g(\bar{c}\hat{c})+D_\alpha\left(\rho_{QQ'\wedge\bar{c}\hat{c}}\middle\|\mathbb{I}_Q\otimes\rho_{Q'\wedge\hat{c}}\right)\right)}}{\sum_{\bar{c}'\hat{c}'\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})}\rho(\bar{c}'\hat{c}')e^{(\alpha-1)(\ln 2)\left(g(\bar{c}'\hat{c}')+D_\alpha\left(\rho_{QQ'\wedge\bar{c}'\hat{c}'}\middle\|\mathbb{I}_Q\otimes\rho_{Q'\wedge\hat{c}'}\right)\right)}}. \tag{4.88}$$

While the expression above formally defines a candidate solution, it does not guarantee existence or validity in all cases. To ensure correctness, we adopt a case-based analysis with the standing assumption that $\mathrm{supp}(\boldsymbol{\lambda})\subseteq\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$.

- $\boldsymbol{\lambda}\in\mathbb{P}_{\overline{\mathcal{C}}\widehat{\mathcal{C}}}$ and has full support on $\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$: For this case, for all $\bar{c}\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$, $\lambda(\bar{c}\hat{c})>0$, and thus $\log(\lambda(\bar{c}\hat{c})/\rho(\bar{c}\hat{c}))$ is finite for all $\bar{c}\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$. Hence if we set

$$\forall \bar{c}\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}}),\quad g(\bar{c}\hat{c}) = \frac{1}{\alpha-1}\log\frac{\lambda(\bar{c}\hat{c})}{\rho(\bar{c}\hat{c})} - D_\alpha\left(\rho_{QQ'\wedge\bar{c}\hat{c}}\|\mathbb{I}_Q\otimes\rho_{Q'\wedge\hat{c}}\right), \tag{4.89}$$

  then we see this is a valid solution to the above conditions (4.88), noting that for this $\mathbf{g}$, the denominator in those conditions is simply

$$\sum_{\bar{c}'\hat{c}'\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})}\rho(\bar{c}'\hat{c}')e^{(\alpha-1)(\ln 2)\left(g(\bar{c}'\hat{c}')+D_\alpha\left(\rho_{QQ'\wedge\bar{c}'\hat{c}'}\middle\|\mathbb{I}_Q\otimes\rho_{Q'\wedge\hat{c}'}\right)\right)} = \sum_{\bar{c}'\hat{c}'\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})}\lambda(\bar{c}'\hat{c}') = 1, \tag{4.90}$$

where the last equality holds since $\boldsymbol{\lambda} \in \mathbb{P}_{\overline{C}\widehat{C}}$ and we also supposed $\operatorname{supp}(\boldsymbol{\lambda}) \subseteq \operatorname{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$. By substituting this stationary point into (4.87) we obtain

$$\begin{aligned} G^*_{\alpha,\rho}(\boldsymbol{\lambda}) &= \sum_{\bar{c}\hat{c} \in \operatorname{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})} \lambda(\bar{c}\hat{c}) \left(\frac{1}{\alpha - 1} \log \frac{\lambda(\bar{c}\hat{c})}{\rho(\bar{c}\hat{c})} - D_\alpha \left(\rho_{QQ' \wedge \bar{c}\hat{c}} \| \mathbb{I}_Q \otimes \rho_{Q' \wedge \hat{c}} \right) \right) - \widetilde{G}_{\alpha,\rho}(\mathbf{g}) \\ &= \frac{1}{\alpha - 1} D\left(\boldsymbol{\lambda} \middle\| \boldsymbol{\rho}_{\overline{C}\widehat{C}} \right) - \sum_{\bar{c}\hat{c} \in \operatorname{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})} \lambda(\bar{c}\hat{c}) D_\alpha \left(\rho_{QQ' \wedge \bar{c}\hat{c}} \| \mathbb{I}_Q \otimes \rho_{Q' \wedge \hat{c}} \right), \end{aligned} \tag{4.91}$$

where in the second line we used the definition of KL divergence and the fact that for this $\mathbf{g}$ we have $\widetilde{G}_{\alpha,\rho}(\mathbf{g}) = 0$ due to (4.90).

- $\boldsymbol{\lambda} \in \mathbb{P}_{\overline{C}\widehat{C}}$ and does not have full support on $\operatorname{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$: For every $\bar{c}\hat{c} \in \operatorname{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$ such that $\lambda(\bar{c}\hat{c}) = 0$, we note that the supremum with respect to that $g(\bar{c}\hat{c})$ term in (4.87) is approached by taking the $g(\bar{c}\hat{c}) \to -\infty$ limit (because $\widetilde{G}_{\alpha,\rho}$ is monotone increasing with respect to each $g(\bar{c}\hat{c})$ value). Therefore, we have:

$$\lim_{g(\bar{c}\hat{c}) \to -\infty} e^{(\alpha-1)(\ln 2)\left(g(\bar{c}\hat{c}) + D_\alpha\left(\rho_{QQ' \wedge \bar{c}\hat{c}} \middle\| \mathbb{I}_Q \otimes \rho_{Q' \wedge \hat{c}}\right)\right)} = 0. \tag{4.92}$$

  Hence, they do not contribute to the sum, and so we can basically set those values in the sum to zero, and restrict our attention to the supremum over the remaining $\mathbf{g}$ components, which then lets us apply essentially the same analysis as the preceding case.

- $\boldsymbol{\lambda} \notin \mathbb{P}_{\overline{C}\widehat{C}}$ and there is at least one $\bar{c}\hat{c} \in \operatorname{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$ with $\lambda(\bar{c}\hat{c}) < 0$: For such cases, note that as the corresponding $g(\bar{c}\hat{c})$ term decreases towards $-\infty$, the objective goes to $+\infty$, because the $\lambda(\bar{c}\hat{c})g(\bar{c}\hat{c})$ term goes to $+\infty$ and the $-\widetilde{G}_{\alpha,\rho}(g)$ term is monotone increasing. Therefore $G^*_{\alpha,\rho}(\boldsymbol{\lambda}) = +\infty$ for such $\boldsymbol{\lambda}$, as claimed in (4.83).

- $\boldsymbol{\lambda} \notin \mathbb{P}_{\overline{C}\widehat{C}}$ and $\lambda(\bar{c}\hat{c}) \geq 0$ for all $\bar{c}\hat{c} \in \operatorname{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$: With the implicit condition $\operatorname{supp}(\boldsymbol{\lambda}) \subseteq \operatorname{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$, this is only possible by having $\sum_{\bar{c}\hat{c} \in \operatorname{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})} \lambda(\bar{c}\hat{c}) \neq 1$. Consider the choice of $\mathbf{g}$ given by taking some $k \in \mathbb{R}$ and setting $g(\bar{c}\hat{c}) = k$ for all $\bar{c}\hat{c} \in \operatorname{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})$. Then, the last line of (4.87) becomes:

$$\begin{aligned} G^*_{\alpha,\rho}(\boldsymbol{\lambda}) = \sup_{k \in \mathbb{R}} \Bigg(& k \left(\sum_{\bar{c}\hat{c} \in \operatorname{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})} \lambda(\bar{c}\hat{c}) - 1 \right) \\ & + \frac{1}{(\alpha - 1)\ln 2} \operatorname{logsumexp}\left(\ln\left(\widetilde{\boldsymbol{\rho}}_{\overline{C}\widehat{C}}\right) + (\alpha - 1)(\ln 2)\widetilde{\mathbf{D}}_\alpha \right) \Bigg), \end{aligned} \tag{4.93}$$

Since the logsumexp term is independent of k, we see that if we have $\sum_{\bar{c}\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})}\lambda(\bar{c}\hat{c}) > 1$ (resp. $\sum_{\bar{c}\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})}\lambda(\bar{c}\hat{c}) < 1$), then as $k \to +\infty$ (resp. $k \to -\infty$) the optimization increases towards $+\infty$, and thus $G^*_{\alpha,\rho}(\boldsymbol{\lambda}) = +\infty$.

Therefore, the claim is proven. □

With this in hand, we now prove the final lemma, which is used to prove Theorem 4.4.1. We note that for this lemma, we require the existence of a single channel similar to that of Eq. (4.62).

Lemma 4.4.5. *Consider the same conditions and notation as in Corollary 4.4.1. Let $r_{\widehat{\alpha}}(\mathbf{f})$ denote the value of the infimum in the second line of (4.80) as a function of the QES $\mathbf{f}$, i.e.*

$$r_{\widehat{\alpha}}(\mathbf{f}) \coloneqq \inf_{\mathbf{q}\in S_\Omega} \inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} \left(H^f_{\widehat{\alpha}}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega} + \mathbf{f}\cdot\mathbf{q} \right). \tag{4.94}$$

Then the optimization problem

$$\sup_{\mathbf{f}} r_{\widehat{\alpha}}(\mathbf{f}), \tag{4.95}$$

is the Lagrange dual problem of the constrained optimization

$$\inf_{\mathbf{q}\in S_\Omega} \inf_{\boldsymbol{\lambda}\in\mathbb{P}_{\overline{C}\widehat{C}}} \inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} \left(\frac{1}{\widehat{\alpha}-1} D\left(\boldsymbol{\lambda}\middle\|\boldsymbol{\nu}^\omega_{\overline{C}\widehat{C}}\right) - \sum_{\bar{c}\hat{c}\in\mathrm{supp}\left(\boldsymbol{\nu}^\omega_{\overline{C}\widehat{C}}\right)} \lambda(\bar{c}\hat{c}) D_{\widehat{\alpha}}\left(\nu^\omega_{SE\widetilde{E}\wedge\bar{c}\hat{c}}\middle\|\mathbb{I}_S\otimes\nu^\omega_{E\widetilde{E}\wedge\hat{c}}\right) \right), \quad s.t. \quad \mathbf{q}-\boldsymbol{\lambda}=\mathbf{0} \tag{4.96}$$

in which the objective function is jointly convex in $\boldsymbol{\lambda}$ and ω (and $\mathbf{q}$, trivially), with $\mathbb{P}_{\overline{C}\widehat{C}}$ denoting the set of probability distributions on $\overline{C}\widehat{C}$.

Furthermore, if S_Ω is convex, the values of (4.95) and (4.96) are equal.

Proof. Let us identify the registers Q and Q' in Lemma 4.4.4 with S and $E\widetilde{E}$ respectively. Then, we have $H^f_{\widehat{\alpha}}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega} = -G_{\widehat{\alpha},\nu^\omega}(\mathbf{f})$. Therefore, Eq. (4.94), can be rewritten as

$$\begin{aligned} r_{\widehat{\alpha}}(\mathbf{f}) &= \inf_{\mathbf{q}\in S_\Omega} \inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} \left(-G_{\widehat{\alpha},\nu^\omega}(\mathbf{f}) + \mathbf{f}\cdot\mathbf{q}\right) \\ &= \inf_{\mathbf{q}\in S_\Omega} \inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} \left(-\sup_{\boldsymbol{\lambda}\in\mathbb{P}_{\overline{C}\widehat{C}}} \left(\boldsymbol{\lambda}\cdot\mathbf{f} - G^*_{\widehat{\alpha},\nu^\omega}(\boldsymbol{\lambda})\right) + \mathbf{f}\cdot\mathbf{q} \right) \\ &= \inf_{\mathbf{q}\in S_\Omega} \inf_{\boldsymbol{\lambda}\in\mathbb{P}_{\overline{C}\widehat{C}}} \inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} \left(G^*_{\widehat{\alpha},\nu^\omega}(\boldsymbol{\lambda}) + (\mathbf{q}-\boldsymbol{\lambda})\cdot\mathbf{f} \right), \end{aligned} \tag{4.97}$$

where the second line follows from the duality argument in Lemma 4.4.4. With this, we have:

$$\sup_{\mathbf{f}} r_{\widehat{\alpha}}(\mathbf{f}) = \sup_{\mathbf{f}} \inf_{\mathbf{q}\in S_\Omega} \inf_{\boldsymbol{\lambda}\in\mathbb{P}_{\overline{C}\widehat{C}}} \inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} \left(G^*_{\widehat{\alpha},\nu^\omega}(\boldsymbol{\lambda}) + (\mathbf{q}-\boldsymbol{\lambda})\cdot\mathbf{f}\right). \tag{4.98}$$

Observe that $G^*_{\widehat{\alpha},\nu^\omega}(\boldsymbol{\lambda})$ is independent of $\mathbf{f}$. Therefore, the above optimization, by definition, is the Lagrange dual of the following optimization:

$$\begin{aligned} \inf_{\mathbf{q}\in S_\Omega} \inf_{\boldsymbol{\lambda}\in\mathbb{P}_{\overline{C}\widehat{C}}} \inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} & G^*_{\widehat{\alpha},\nu^\omega}(\boldsymbol{\lambda}) \\ \text{s.t.} \quad & \mathbf{q}-\boldsymbol{\lambda} = \mathbf{0} \end{aligned}. \tag{4.99}$$

Substituting $G^*_{\widehat{\alpha},\nu^\omega}(\boldsymbol{\lambda})$ from Eq. (4.83) returns the result in Eq. (4.96).

We now prove that if S_Ω is convex, then the solutions to the optimizations in (4.95) and (4.96) are equal. To do so, we resort to Sion's minimax theorem [109].

We first note that, as proved in Corollary 4.4.1, $G_{\widehat{\alpha},\nu^\omega}(\mathbf{f})$ is concave in ω since $-G_{\widehat{\alpha},\nu^\omega}(\mathbf{f}) = H^f_{\widehat{\alpha}}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega}$.

Moreover, following Lemma 4.4.4, $-G_{\widehat{\alpha},\nu^\omega}(\mathbf{f})$ is concave in $\mathbf{f}$, and thus $\mathbf{f}\cdot\mathbf{q} - G_{\widehat{\alpha},\nu^\omega}(\mathbf{f})$ is concave in $\mathbf{f}$. Hence, we can apply Sion's minimax theorem as follows:

$$\begin{aligned} \sup_{\mathbf{f}} r_{\widehat{\alpha}}(\mathbf{f}) &= \sup_{\mathbf{f}} \inf_{\mathbf{q}\in S_\Omega} \inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} \left(-G_{\widehat{\alpha},\nu^\omega}(\mathbf{f}) + \mathbf{f}\cdot\mathbf{q}\right) \\ &= \inf_{\mathbf{q}\in S_\Omega} \inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} \sup_{\mathbf{f}} \left(-G_{\widehat{\alpha},\nu^\omega}(\mathbf{f}) + \mathbf{f}\cdot\mathbf{q}\right) \\ &= \inf_{\mathbf{q}\in S_\Omega} \inf_{\omega\in \mathrm{S}_=(\widetilde{Q})} G^*_{\widehat{\alpha},\nu^\omega}(\mathbf{q}), \end{aligned} \tag{4.100}$$

where the second line holds by Sion's minimax theorem since the objective function is convex in $(\mathbf{q},\omega)$ and concave in $\mathbf{f}$, and the last line holds by the definition of convex conjugates. This proves the desired claim since the last line is equal to the value of (4.99), by noting that the constraint in that optimization can be substituted into the objective function. □

We now discuss some important implications of Lemma 4.4.5. We begin by noting that the optimal choice of f in Corollary 4.4.1 is exactly the same as the optimal $\mathbf{f}$ in the optimization (4.95). Therefore, in cases where S_Ω is no longer convex, one can instead find the optimal $\mathbf{f}$ by computing the dual solution to Eq. (4.96) instead of Theorem 4.4.1. However, note that in such cases, there is no guarantee that the optimal values for the optimizations in (4.95) and (4.96) match. Therefore, to obtain valid lower bounds in those cases, one must substitute the optimal $\mathbf{f}$ into Corollary 4.4.1 and compute the result.

The second implication is for the case where S_Ω is convex, which yields the proof of Theorem 4.4.1. Observe that if we optimize the Corollary 4.4.1 bound over all f, we have:

$$H_\alpha^\uparrow(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_{\rho_{|\Omega}} \geq \left(\sup_{\mathbf{f}} r_{\widehat{\alpha}}(\mathbf{f})\right) n - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega}. \tag{4.101}$$

If we set $h_{\widehat{\alpha}} \coloneqq \sup_{\mathbf{f}} r_{\widehat{\alpha}}(\mathbf{f})$, then, since S_Ω is convex, the solution to the optimization in (4.96) is equal to $h_{\widehat{\alpha}}$, which, after substituting the constraint in that optimization, yields the expression in (4.62).

We now prove that the expression in (4.62) matches the one in (4.61). First, note that for any choice of $\boldsymbol{q}$, every feasible point $\omega_{\widetilde{Q}}$ in (4.62) is also a feasible point in (4.61) by setting $\nu = \nu^\omega$. Therefore, we have that (4.61) $\leq$ (4.62).

For the other direction, take any feasible point $\nu \in \Sigma$. Then, $\nu = \mathcal{M}\left[\omega_{\widetilde{QE}}\right]$ for some $\omega_{\widetilde{QE}}$. Now, note that since $\texttt{Pur}(\omega_{\widetilde{Q}})$ is a purification of $\omega_{\widetilde{Q}}$, there exists a channel $\mathcal{E}$ on $\widetilde{E}$ such that $\omega_{\widetilde{QE}} = \mathcal{E}\left[\texttt{Pur}(\omega_{\widetilde{Q}})\right]$. Then, we have:

$$\nu = \mathcal{M}\left[\omega_{\widetilde{QE}}\right] = \mathcal{M}\circ\mathcal{E}\left[\texttt{Pur}(\omega_{\widetilde{Q}})\right] = \mathcal{E}\circ\mathcal{M}\left[\texttt{Pur}(\omega_{\widetilde{Q}})\right] = \mathcal{E}\left[\nu^\omega\right]. \tag{4.102}$$

In particular, $\nu_{\widehat{C}\widehat{C}} = \nu^\omega_{\widehat{C}\widehat{C}}$ and $\nu_{\wedge\bar{c}\hat{c}} = \mathcal{E}[\nu^\omega_{\wedge\bar{c}\hat{c}}]$. Therefore, $D\left(\mathbf{q}\middle\|\boldsymbol{\nu}_{\widehat{C}\widehat{C}}\right) = D\left(\mathbf{q}\middle\|\boldsymbol{\nu}^\omega_{\widehat{C}\widehat{C}}\right)$ and also

$$\begin{aligned}\forall \bar{c}\hat{c} \in \operatorname{supp}\left(\boldsymbol{\nu}_{\widehat{C}\widehat{C}}\right), \quad -D_{\widehat{\alpha}}\left(\nu_{SE\widetilde{E}\wedge\bar{c}\hat{c}}\middle\|\mathbb{I}_S\otimes\nu_{E\widetilde{E}\wedge\hat{c}}\right) &= -D_{\widehat{\alpha}}\left(\mathcal{E}\left[\nu^\omega_{SE\widetilde{E}\wedge\bar{c}\hat{c}}\right]\middle\|\mathbb{I}_S\otimes\mathcal{E}\left[\nu^\omega_{E\widetilde{E}\wedge\hat{c}}\right]\right)\\ &\geq -D_{\widehat{\alpha}}\left(\nu^\omega_{SE\widetilde{E}\wedge\bar{c}\hat{c}}\middle\|\mathbb{I}_S\otimes\nu^\omega_{E\widetilde{E}\wedge\hat{c}}\right),\end{aligned} \tag{4.103}$$

where the inequality follows from the data-processing property in Theorem 2.2.1. Thus, we conclude that the reduced state $\omega_{\widetilde{Q}}$ is a feasible point of (4.62) with an objective value that is no larger. Hence, (4.61) $\geq$ (4.62), which completes the proof.

The third implication of Lemma 4.4.5 concerns variable-length protocols. In particular, it provides a procedure for choosing QES-s for variable-length protocols. First, observe that the optimal dual solution to (4.95) is precisely the optimal choice of $\mathbf{f}$ to use in (4.80). Depending on whether one wishes to use a fixed QES across all rounds or update it adaptively, two approaches can be considered:

- Same QES across rounds:

  If we want to use the same QES f in every round, we proceed as follows. Let S_Ω be

$$S_\Omega = \{\boldsymbol{q} \mid \ d(\boldsymbol{q}, \mathbf{q}^{\text{hon}}) \leq \delta\}, \tag{4.104}$$

for some $\delta \geq 0$ and some distance measure d, with $\mathbf{q}^{\text{hon}}$ being the distribution corresponding to the honest behavior. Then, take f to be the dual solution to the constrained optimization in (4.96).

- Adaptively updated QES across rounds:

  If we want to adaptively change QES-s $f_{|\vec{c}_1^{j-1}\hat{c}_1^{j-1}}$, we proceed as follows. For each j at which we want to update the QES, we consider the preceding rounds and compute an estimate $\mathbf{q}^{\text{est}}$ of the upcoming behavior of the device[18]. Let S_Ω be

  $$S_\Omega = \left\{ \boldsymbol{q} \mid \ d(\boldsymbol{q}, \mathbf{q}^{\text{est}}) \leq \delta \right\}, \tag{4.105}$$

  for some $\delta \geq 0$ and some distance measure d. Then, take $f_{|\vec{c}_1^{j-1}\hat{c}_1^{j-1}}$ to be the dual solution to the constrained optimization in (4.96).

Based on Lemma 4.4.5 and the above explanation, this procedure identifies a choice of QES that provides good key rates when applying Theorem 4.3.1.

4.5 Original EAT with improved statistical analysis

In this section, we apply our QES analysis to the original entropy accumulation model [41] (see also Sec. 2.5.2) to obtain tighter results under that model. We note that, using the original EAT, we can avoid changing the Rényi parameter for the single-round quantity. Furthermore, we no longer have to impose the non-signaling condition required in the GEAT model. However, we must impose a specific Markov condition used in the original EAT model. The following lemma essentially states that the various bounds we obtained in the previous section are also valid under the original EAT model.

Lemma 4.5.1. *Let $\rho_{S_1^n T_1^n \overline{C}_1^n \widehat{C}_1^n \widehat{E}}$ be a state of the form $\rho = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1[\omega^0_{R_0\widehat{E}}]$ for some initial state $\omega^0 \in \mathrm{S}_=(R_0\widehat{E})$ and some channels $\mathcal{M}_j \in \mathrm{CPTP}(R_{j-1}, S_j T_j R_j \overline{C}_j \widehat{C}_j)$ such that the output registers $\overline{C}_j\widehat{C}_j$ are always classical (for any input state). Suppose that for each j, ρ satisfies the Markov condition*

$$I(S_1^{j-1}\overline{C}_1^{j-1} : T_j\widehat{C}_j | T_1^{j-1}\widehat{C}_1^{j-1}\widehat{E})_\rho = 0, \tag{4.106}$$

[18]The procedure we describe here is independent of how $\mathbf{q}^{\text{est}}$ is constructed.

i.e. $S_1^{j-1}\overline{C}_1^{j-1} \leftrightarrow T_1^{j-1}\widehat{C}_1^{j-1}\widehat{E} \leftrightarrow T_j\widehat{C}_j$ *forms a Markov chain. Then Theorems 4.3.1 and 4.4.1, Corollaries 4.3.1 and 4.4.1, and Lemmas 4.4.1–4.4.5 all hold with the following minor modifications:*

- *Replace* $\widehat{\alpha}$ *with* α *throughout, and allow* $\alpha \in (1,\infty)$ *instead of* $\alpha \in (1,2)$.
- *Omit the condition that* $\{\mathcal{M}_j\}_{j=1}^n$ *are GEATT channels.*
- *Replace the registers with suitable counterparts, e.g.* $\mathbb{H}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n) \to \mathbb{H}(S_1^n\overline{C}_1^n|\widehat{C}_1^n T_1^n\widehat{E})$, $\mathbb{H}(S_j\overline{C}_j|\widehat{C}_j E_j\widetilde{E}) \to \mathbb{H}(S_j\overline{C}_j|\widehat{C}_j T_j\widetilde{E})$, $\mathbb{H}(S\overline{C}|\widehat{C}E\widetilde{E}) \to \mathbb{H}(S\overline{C}|\widehat{C}T\widetilde{E})$ *and so on, where* $\mathbb{H}$ *denotes any type of entropy.*

Proof. We construct new channels $\mathcal{N}_j$ the same way as the Theorem 4.3.1 proof, except that we also have them output an extra copy of $\widehat{C}_j$ in a register $\widehat{\widehat{C}}_j$. Since the channels $\mathcal{M}_j$ in this case have the form $\mathcal{M}_j \in \mathrm{CPTP}(R_{j-1}, S_jT_jR_j\overline{C}_j\widehat{C}_j)$, we end up with channels $\mathcal{N}_j \in \mathrm{CPTP}(R_{j-1}\overline{C}_1^{j-1}\widehat{C}_1^{j-1}, D_j\overline{\overline{C}}_j S_jT_j\widehat{\widehat{C}}_j R_j\widehat{C}_1^j\overline{C}_1^j)$. Again writing $\rho = \mathcal{N}_n \circ \cdots \circ \mathcal{N}_1[\omega^0]$, this is a state on registers $D_1^nS_1^nT_1^n\overline{\overline{C}}_1^n\widehat{\widehat{C}}_1^n\overline{C}_1^n\widehat{C}_1^n\widetilde{E}$ that is an extension of the state ρ in this lemma statement. We now show that the state produced by each channel $\mathcal{N}_j$ satisfies the following Markov condition:

$$I(S_1^{j-1}\overline{\overline{C}}_1^{j-1}D_1^{j-1} : T_j\widehat{\widehat{C}}_j|T_1^{j-1}\widehat{\widehat{C}}_1^{j-1}\widehat{E})_\rho = 0, \tag{4.107}$$

To do this, we first use the chain rule for conditional mutual information [68, Proposition 7.7] to obtain

$$\begin{aligned}
&I(S_1^{j-1}\overline{\overline{C}}_1^{j-1}D_1^{j-1} : T_j\widehat{\widehat{C}}_j|T_1^{j-1}\widehat{\widehat{C}}_1^{j-1}\widehat{E})_\rho \\
=&I(S_1^{j-1}\overline{C}_1^{j-1}D_1^{j-1} : T_j\widehat{C}_j|T_1^{j-1}\widehat{C}_1^{j-1}\widehat{E})_\rho \\
=&I(S_1^{j-1}\overline{C}_1^{j-1} : T_j\widehat{C}_j|T_1^{j-1}\widehat{C}_1^{j-1}\widehat{E})_\rho + I(D_1^{j-1} : T_j\widehat{C}_j|S_1^{j-1}\overline{C}_1^{j-1}T_1^{j-1}\widehat{C}_1^{j-1}\widehat{E})_\rho \\
=&I(D_1^{j-1} : T_j\widehat{C}_j|S_1^{j-1}\overline{C}_1^{j-1}T_1^{j-1}\widehat{C}_1^{j-1}\widehat{E})_\rho \\
=&0,
\end{aligned} \tag{4.108}$$

where the second line holds since $\overline{\overline{C}}_1^{j-1} = \overline{C}_1^{j-1}$, and $\widehat{\widehat{C}}_1^{j-1} = \widehat{C}_1^{j-1}$. The third line is due to the chain rule for conditional mutual information, the fourth line follows from the Markov condition (4.106), and the last line holds by noting that D_1^{j-1} is produced from $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$. Thus, we conclude that this extended ρ satisfies the Markov condition. We now apply [41, Corollary 3.5] by identifying the registers $A_1A_2B_1B_2R$ in that corollary with the registers here as follows:

- $A_1 \leftrightarrow D_1^{j-1}\overline{\overline{C}}_1^{j-1}S_1^{j-1}$
- $A_2 \leftrightarrow D_j\overline{\overline{C}}_j S_j$
- $B_1 \leftrightarrow \widehat{\widehat{C}}_1^{j-1}T_1^{j-1}\widehat{E}$
- $B_2 \leftrightarrow \widehat{\widehat{C}}_j T_j$
- $R \leftrightarrow R_{j-1}\overline{C}_1^j\widehat{C}_1^j$

With this identification, we obtain for each j:

$$H_\alpha(D_1^jS_1^j\overline{\overline{C}}_1^j|\widehat{\widehat{C}}_1^jT_1^j\widehat{E})_\rho \geq H_\alpha(D_1^{j-1}S_1^{j-1}\overline{\overline{C}}_1^{j-1}|\widehat{\widehat{C}}_1^{j-1}T_1^{j-1}\widehat{E})_\rho + \inf_{\nu'\in\Sigma'_j} H_\alpha(D_jS_j\overline{\overline{C}}_j|\widehat{\widehat{C}}_jT_1^j\widetilde{E})_{\nu'}. \tag{4.109}$$

Applying this for every j, we get:

$$H_\alpha(D_1^jS_1^j\overline{\overline{C}}_1^j|\widehat{\widehat{C}}_1^jT_1^j\widehat{E})_\rho \geq \sum_j \inf_{\nu'\in\Sigma'_j} H_\alpha(D_jS_j\overline{\overline{C}}_j|\widehat{\widehat{C}}_jT_1^j\widetilde{E})_{\nu'}. \tag{4.110}$$

Recall that by definition of $\mathcal{N}_j$, the state ν' contains registers $\overline{C}_1^j\widehat{C}_1^j$ even though they do not appear in entropies in the above expression. Therefore we can apply the same argument as the bound (4.40) in the Theorem 4.3.1 proof, lower bounding the $H_\alpha(D_jS_j\overline{\overline{C}}_j|\widehat{\widehat{C}}_jT_1^j\widetilde{E})_{\nu'}$ term by conditioning on $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$ and taking the worst-case value. Moreover, since $\overline{\overline{C}}_j$ and $\widehat{\widehat{C}}_j$ are just a copy of $\overline{C}_j$ and $\widehat{C}_j$, respectively (in particular $\overline{\overline{C}}_1^n = \overline{C}_1^n$, and $\widehat{\widehat{C}}_1^n = \widehat{C}_1^n$), we arrive at

$$H_\alpha(D_1^nS_1^n\overline{C}_1^n|\widehat{C}_1^nT_1^n\widehat{E})_\rho \geq \sum_j \inf_{\nu'\in\Sigma'_j} \min_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}} H_\alpha(D_jS_j\overline{C}_j|\widehat{C}_jT_1^j\widetilde{E})_{\nu'_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}}. \tag{4.111}$$

With this, we carry on the remainder of all the proofs in the same way, since none of the subsequent steps require the NS condition of GEATT channels. □

4.6 Tightness of bounds and numerical example

In this section, we first discuss the tightness of the bounds obtained in the previous sections and then provide a simple example to demonstrate the strength of our findings.

4.6.1 Remaining steps for improvement

We highlight that if, in Theorem 4.4.1, S_Ω satisfies the conditions in that theorem, then for α close to 1, almost all the bounds used in the proof of that theorem are nearly saturated simultaneously by IID states. More explicitly, starting from Theorem 4.3.1, we have the following:

- As discussed in 4.3.2, Theorem 4.3.1 is tight for IID states up to higher-order changes in the Rényi parameter.
- In the procedure of obtaining Corollary 4.3.1, we used $H_\alpha^\uparrow \geq H_\alpha$, and Lemma 4.3.9. However, the looseness caused by this relaxation is at most what would be caused by a higher-order change of Rényi parameter, since by a converse bound we have $H_{2-1/\alpha} \geq H_\alpha^\uparrow$.
- We simplified further to obtain Corollary 4.4.1, where we assumed all QES-s are identical and relaxed the optimization over $\widetilde{\Omega}$ to S_Ω. However, setting all QES-s to be identical is a natural choice for protocols in which no round is *a priori* different from the others.
- In proceeding from Corollary 4.4.1 to Theorem 4.4.1, all the steps are equalities, and thus no loss of tightness occurred.

We also highlight that, apart from our analysis, a potential source of sub-optimality might lie in the Rényi privacy amplification theorem itself. Additionally, we note that in obtaining Corollary 4.3.1, the term $\frac{\alpha}{\alpha-1}\log\frac{1}{\Pr[\Omega]}$ does not affect the final key rates, as discussed in Sec. 3.3.1. We also note that the suboptimalities arising from the change in the Rényi parameter—however negligible—can be avoided by employing the framework discussed in Chapter 5.

Therefore, to summarize, the main potential sources of suboptimality are the relaxation from $\widetilde{\Omega}$ to S_Ω, and Lemma 4.3.9, which we will revisit in the next section. Given the above explanations, we expect that Theorem 4.4.1 provides strictly better bounds than previous GEAT or EAT results (see [7, Section 7.1] for a rigorous argument).

4.6.2 Entanglement-based BB84

In this section, as an example, we calculate key rates for a "fully qubit" EB-BB84 protocol.

For the analysis in this part, we note that, for simplicity, the numerical methods used are heuristics and thus may not provide certified lower bounds on secure key rates. However, the minimizations we evaluate can be expressed using at most three optimization variables and are convex in each individual variable. Therefore, we believe there is minimal risk of significantly overestimating the true key rates.

We first establish the protocol description.

Protocol 4. EB-BB84 protocol
Parameters:

$n \in \mathbb{N}_0$	:	Total number of rounds
$l \in \mathbb{N}_0$	:	Length of final key
S_Ω	:	Acceptance set of accepted frequencies
λ_{EC}	:	Number of bits exchanged during error correction step

Protocol steps:

1. For each round $i \in \{1, 2, \ldots, n\}$, Alice and Bob perform the following steps:
   (a) Alice and Bob each receive a qubit register. Then, via public communication, with probability $1 - \gamma$ (independently in each round), they jointly declare the round is a generation round, and otherwise it is a test round. Alice also independently generates a uniformly random "symmetrization bit"[19] $F_j \in \{0, 1\}$ and publicly announces it.
   (b) If it is a generation round, they both measure in the Z basis, XOR their outcomes with F_j, and store the resulting values in registers S_j and $\widetilde{S}_j$ for Alice and Bob, respectively. Also, they jointly set $\widehat{C}_j = \perp$.
   (c) If it is a test round, they both measure in the X basis, XOR their outcomes with F_j, and publicly announce the resulting values, then jointly set $\widehat{C}_j = 0$ if their outcomes matched and $\widehat{C}_j = 1$ otherwise. Also, Alice sets $S_j = 0$ and Bob sets $\widetilde{S}_j = \texttt{test}$.
2. **Acceptance test:** Perform an acceptance test, in which Alice and Bob abort if $\mathrm{freq}_{\hat{c}_1^n}$ lies outside some predetermined set S_Ω.

[19] We note that the symmetrization procedure is only used to simplify calculations and is not strictly necessary (see [7, Sec. 7.2] for further details).

3. **Error correction:** Perform one-way error correction in which Alice sends Bob a bitstring of length λ_{EC}, which he uses together with $\widetilde{S}_1^n$ to produce a guess $\mathbf{S}^{\text{guess}}$ for Alice's string S_1^n. Then perform error verification, in which Alice sends Bob a 2-universal hash of S_1^n (together with the choice of hash function), who compares it with the hash of $\mathbf{S}^{\text{guess}}$ and aborts if they do not match.

4. **Privacy amplification:** If neither of the above steps aborted, produce final keys of length ℓ_{key} by performing privacy amplification, in which Alice chooses a 2-universal function $\{0,1\}^n \to \{0,1\}^{\ell_{\text{key}}}$ and publicly announces it, then Alice and Bob apply it to S_1^n and $\mathbf{S}^{\text{guess}}$ respectively.

We note that given that we aim to compare the key rates we obtain with those in [119, 80], we tried to be as similar as possible to those works. For example, we allowed Alice and Bob to jointly decide the basis to measure in. Moreover, we choose the acceptance set S_Ω as follows:

$$S_\Omega = \left\{ \mathbf{q} \in \mathbb{P}_{\widehat{C}} \;\middle|\; q_{\widehat{C}}(1) \leq \gamma Q_{\text{thresh}} \right\}, \tag{4.112}$$

for some value of QBER threshhold Q_{thresh}. Moreover, the length of the error-correction string can be evaluated as

$$\lambda_{\text{EC}} = \xi_{\text{EC}}(1-\gamma) h_{\text{bin}}(Q_{\text{thresh}}) n, \tag{4.113}$$

where $\xi_{\text{EC}} > 1$ is a heuristically chosen value describing the finite-size efficiency of the error-correction procedure. Note that the $(1-\gamma)$ factor is the generation-round probability.

Remark 4.6.1. *Numerical results suggest that Theorem 4.4.1 yields much higher bounds if the acceptance set S_Ω is chosen such that it constrains all entries of $\mathbf{q}$ to a small neighborhood of the honest behavior. In other words,*

$$S_\Omega = \left\{ \mathbf{q} \in \mathbb{P}_{\widehat{C}} \;\middle|\; q_{\widehat{C}}(\hat{c}) \in I_{\hat{c}} \; \forall \; \hat{c} \in \widehat{\mathcal{C}} \right\}, \tag{4.114}$$

for some small intervals $I_{\hat{c}}$.[20] *This seems related to the observation that Lemma 4.3.9 is potentially the loosest step in our proof, and choosing S_Ω such that it only captures the typical IID sequences helps reduce the value of $\max_{\bar{c}_1^n \hat{c}_1^n \in \widetilde{\Omega}} H_\alpha(D)_{\rho|\bar{c}_1^n \hat{c}_1^n}$ there.*

[20] We highlight that, given we are constraining more elements of $\mathbf{q}$, to achieve the same bound on the honest abort probability (completeness parameter), one must have a looser upper bound on $q_{\widehat{C}}(1)$ than the one in (4.112).

Going back to the BB84 example, we construct GEATT channels $\mathcal{M}_j \in \mathrm{CPTP}(E_{j-1}, E_j)$ such that E_j contains all side information the adversary has in each round, including her quantum side information and all public announcements in that round. Furthermore, the $\mathcal{M}_j$ channels are infrequent-sampling channels with testing probability γ. Finally, we apply Theorem 4.4.1 to these channels to obtain the required bounds. However, to provide a deeper comparison with the key rates in [119, 80], we also present multiple relaxations of the bound in Theorem 4.4.1.

Below we summarize the final key rates for a protocol to be $\varepsilon^{\text{secure}}$-secure.[21]

- Employing the Rényi privacy amplification theorem, along with the bounds in Theorem 4.4.1, and setting $\varepsilon^{\text{secure}} = \varepsilon_{\text{PA}} + \varepsilon_{\text{EV}}$, the secret key length is:

$$\ell_{\text{key}} = \left\lfloor \inf_{\mathbf{q}\in S_\Omega} \inf_{\boldsymbol{\nu}_{\widehat{C}}\in\mathbb{P}_{\widehat{C}}} \left(q(\perp) \left(1 - \frac{1}{1-\beta} \log \left(\left(1 - \frac{\nu_{\widehat{C}}(1)}{\gamma} \right)^{\beta} + \left(\frac{\nu_{\widehat{C}}(1)}{\gamma} \right)^{\beta} \right) \right) + \frac{1}{\widehat{\alpha}-1} D\left(\mathbf{q} \middle\| \boldsymbol{\nu}_{\widehat{C}}\right) \right) n - \lambda_{\text{EC}} - \left\lceil \log \frac{1}{\varepsilon_{\text{EV}}} \right\rceil - \frac{\alpha}{\alpha-1} \log \frac{1}{\varepsilon_{\text{PA}}} + 2 \right\rfloor, \quad (4.115)$$

- Converting the global entropy to smooth min-entropy, applying the smooth min-entropy privacy amplification theorem [104], and setting $\varepsilon^{\text{secure}} = \max\left\{\frac{\varepsilon_{\text{PA}}}{2} + 2\varepsilon_{\text{s}}, \varepsilon_{\text{a}}\right\} + \varepsilon_{\text{EV}}$, the secret key length is:

$$\ell_{\text{key}} = \left\lfloor \inf_{\mathbf{q}\in S_\Omega} \inf_{\boldsymbol{\nu}_{\widehat{C}}\in\mathbb{P}_{\widehat{C}}} \left(q(\perp) \left(1 - \frac{1}{1-\beta} \log \left(\left(1 - \frac{\nu_{\widehat{C}}(1)}{\gamma} \right)^{\beta} + \left(\frac{\nu_{\widehat{C}}(1)}{\gamma} \right)^{\beta} \right) \right) + \frac{1}{\widehat{\alpha}-1} D\left(\mathbf{q} \middle\| \boldsymbol{\nu}_{\widehat{C}}\right) \right) n - \frac{\alpha}{\alpha-1} \log \frac{1}{\varepsilon_{\text{a}}} - \frac{1}{\alpha-1} \log \frac{2}{\varepsilon_{\text{s}}^2} - \lambda_{\text{EC}} - \left\lceil \log \frac{1}{\varepsilon_{\text{EV}}} \right\rceil - 2 \log \frac{1}{\varepsilon_{\text{PA}}} \right\rfloor. \quad (4.116)$$

- Performing yet another conversion to the above method to relax the single-round entropy into the von Neumann conditional entropy, applying the smooth min-entropy privacy amplification, and setting $\varepsilon^{\text{secure}} = \max\left\{\frac{\varepsilon_{\text{PA}}}{2} + 2\varepsilon_{\text{s}}, \varepsilon_{\text{a}}\right\} + \varepsilon_{\text{EV}}$, the secret key

[21] For a more comprehensive derivation of these bounds, we refer the reader to [7].

length is:

$$\ell_{\text{key}} = \left\lfloor \inf_{\mathbf{q}\in S_\Omega} \inf_{\boldsymbol{\nu}_{\widehat{C}}\in\mathbb{P}_{\widehat{C}}} \left(q(\perp)\left(1 - h_{\text{bin}}\left(\frac{\nu_{\widehat{C}}(1)}{\gamma}\right) - (\widehat{\alpha}-1)\log^2(5)\right) + \frac{1}{\widehat{\alpha}-1} D\left(\mathbf{q}\middle\|\boldsymbol{\nu}_{\widehat{C}}\right)\right) n \right.$$
$$\left. - \frac{\alpha}{\alpha-1}\log\frac{1}{\varepsilon_{\text{a}}} - \frac{1}{\alpha-1}\log\frac{2}{\varepsilon_{\text{s}}^2} - \lambda_{\text{EC}} - \left\lceil \log\frac{1}{\varepsilon_{\text{EV}}}\right\rceil - 2\log\frac{1}{\varepsilon_{\text{PA}}}\right\rfloor, \tag{4.117}$$

In Fig. 4.2–4.3, we plot the key rates $\frac{\ell_{\text{key}}}{n}$ obtained from the above formulas. We highlight that, for generating the plot, the optimal choice of security parameters for Eq. (4.115) was found to be:

$$\varepsilon_{\text{EV}} = \frac{\alpha-1}{2\alpha-1}\varepsilon^{\text{secure}}, \qquad \varepsilon_{\text{PA}} = \frac{\alpha}{2\alpha-1}\varepsilon^{\text{secure}}, \tag{4.118}$$

and for the formulas in Eqs. (4.116) and (4.117), we obtained the following optimal choices:

$$\varepsilon_{\text{PA}} = \varepsilon_{\text{EV}} = \varepsilon_{\text{small}}, \qquad \varepsilon_{\text{a}} = \varepsilon^{\text{secure}} - \varepsilon_{\text{small}}, \qquad \varepsilon_{\text{s}} = \frac{1}{2}\left(\varepsilon^{\text{secure}} - \frac{3}{2}\varepsilon_{\text{small}}\right), \tag{4.119}$$

with $\varepsilon_{\text{small}} = \frac{3}{4}(\alpha-1)\varepsilon^{\text{secure}}$. Comparing them to the corresponding results in [119, Fig. 7], we find that while the loosest bound (4.117) is slightly worse at small n, we have better key rates everywhere else as compared to the results in that work. However, compared to the tighter finite-size analysis in [80, Fig. 1]. we see that while (4.116) and (4.115) also performed better in almost all cases, our loosest bound (4.117) is worse.

These results indicate that our best approach for bounding the smooth min-entropy outperforms the improvements in [80], although the difference is not very large. However, by employing Rényi privacy amplification—which allows us to directly analyze the Rényi entropy in our approach—we obtain a significant improvement over the result in [80].

We highlight that although we only study a fairly simple example, [66] treated a number of more practical protocols, such as decoy-state and PM-BB84 with loss. Their findings indicate strict improvements over other proof techniques.

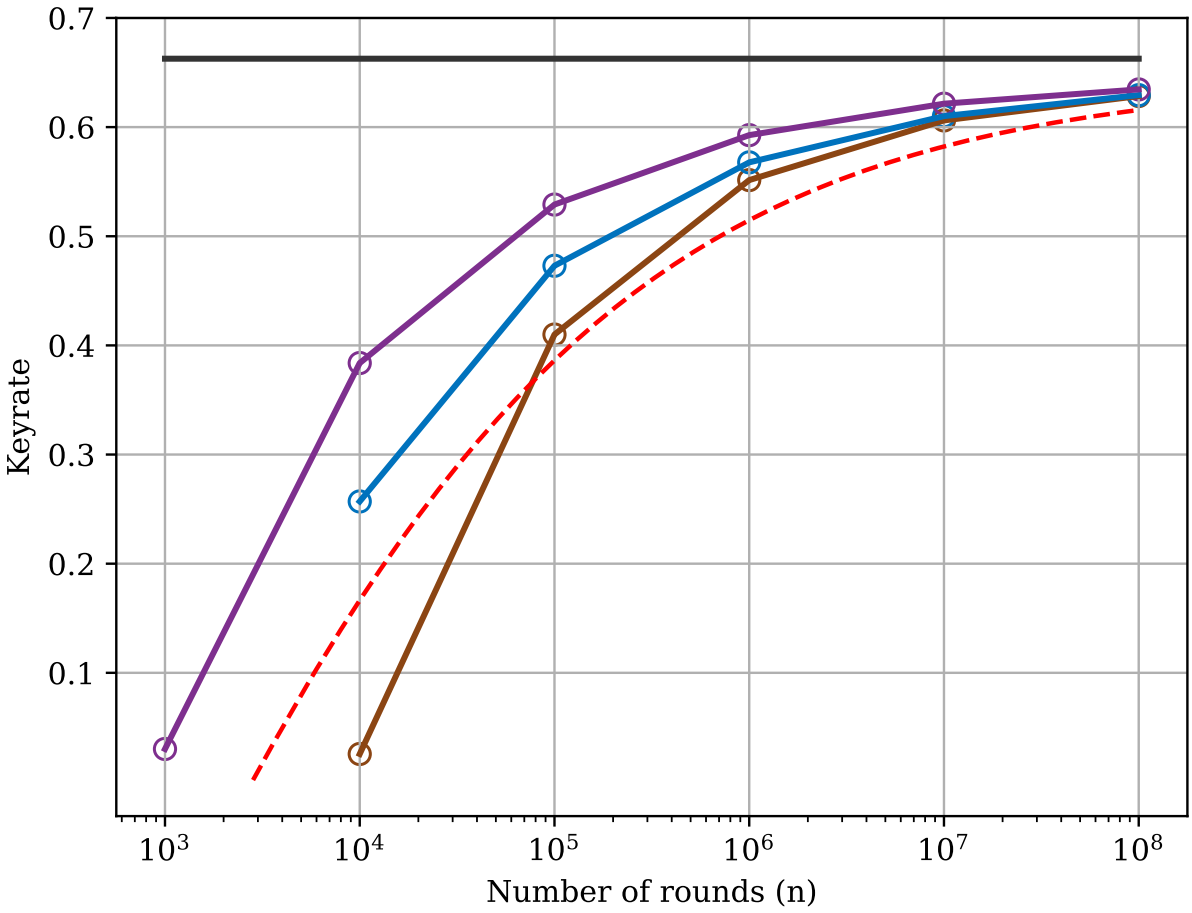


Figure 4.2: Key rates for EB-BB84 from our approach, for an example with QBER threshold $Q_{\text{thresh}} = 0.025$, error-correction efficiency $\xi_{\text{EC}} = 1.1$, and security parameter $\varepsilon^{\text{secure}} = 10^{-10}$. The purple, blue, and brown curves correspond to the key rates given by the formulas (4.115), (4.116) and (4.117), respectively, where we used heuristic numerical methods to evaluate the (fairly simple) minimizations in those formulas. For comparison, the dashed red curve is the corresponding result from [119, Fig. 7], and the black horizontal line is the asymptotic rate. It can be seen that our results are always an improvement over that work, except for the case of the suboptimal formula (4.117) at small n. We highlight that the formula (4.116) already performs better everywhere despite also proceeding via a smooth min-entropy bound (instead of Rényi privacy amplification), i.e., this indicates we genuinely obtained a better bound on smooth min-entropy as compared to [119]. The choices of S_Ω, λ_{EC}, and epsilon parameters we used in our formulas are described in the main text.

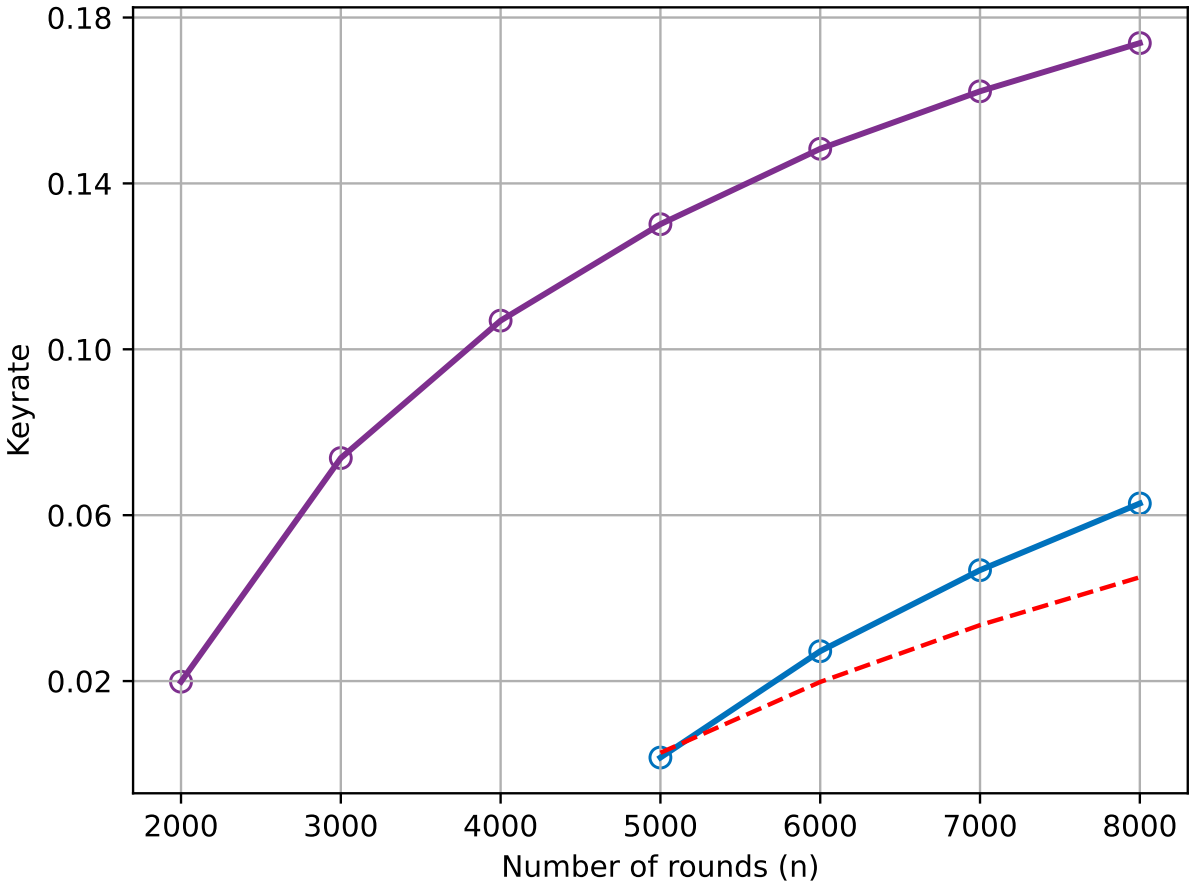


Figure 4.3: Similar to Fig. 4.2, except that the parameter choices are $Q_{\text{thresh}} = 0.0451$, $\xi_{\text{EC}} = 1.19$, $\varepsilon^{\text{secure}} = 10^{-10}$, and the dashed red curve shows the corresponding data points from [80] instead (which improved the finite-size analysis in [119]). Here, there is no curve shown for our loosest bound (4.117), as we could not obtain positive key rates from it for the n values in this range, consistent with the observation in Fig. 4.2 that it performs less well at very small n. These results suggest that our best bound on the smooth min-entropy itself (i.e. that used in (4.116)) is still better than [80] at most n values, though the advantage is not that large. However, in the end, we can achieve much higher actual key rates by instead using the fully Rényi formula (4.115).

4.7 Conclusion and open questions

In this chapter, we established a connection between entropy accumulation and the QEF framework. Through this connection, we obtained several entropic bounds suitable for fixed-length and variable-length QKD protocols. Furthermore, they significantly improved upon previous EAT-based results in [37, 41, 89]. Moreover, our approach is based on analyzing Rényi entropies, yielding a fully Rényi security proof. In addition, for fixed-length protocols, the analysis in this chapter avoids the complication of obtaining the min-tradeoff function, as seen in the previous chapter.

Going forward, it is important to see whether the bounds presented here can be improved. However, as mentioned earlier, there does not appear to be much room left for tightening the bounds. It seems that the only improvement would likely come from refining Lemma 4.3.9. We re-emphasize, however, that with the current improvements, the bounds we obtained already surpass those based on the EUR approach.[22]

[22]Recently, in [66], key rates for some practical protocols such as decoy-state and qubit BB84 with loss were analyzed, and it was shown that even for those protocols, this approach surpasses the key rates from the EUR framework.

Chapter 5

Marginal-constrained entropy accumulation

5.1 Introduction

So far, the underlying framework for the analysis conducted in previous chapters has been either the original EAT or its generalization, GEAT. In particular, our improved statistical analysis in the previous chapter was valid under the original EAT and the GEAT framework. However, as pointed out in Chapter 3, when aiming to analyze PM-QKD protocols, while the GEAT is applicable to both PM and EB protocols, it is restrictive in that the adversary is limited to interacting with only a single quantum register sent by the sender at a time, which would, in practice, limit the repetition rate of the protocol. In this chapter, we aim to overcome this limitation while maintaining compatibility with the improved analysis developed earlier.

A solution to the repetition-rate restriction was given in [126]. However, it is somewhat restrictive in other respects. Essentially, the idea in that work is as follows. As discussed in Subsection 2.4.2, through the source-replacement technique [26, 49], we can analyze a PM protocol as an EB protocol. Recall that employing this technique requires a constraint on Alice's marginal state that reflects the fact that Alice does not send that part of the state through the adversarial channel. The solution proposed in [126] is restricted in the following sense. First, it requires Alice's single-round marginal state to be identical across all rounds. Second, its proof approach involves modeling the protocol via a specific tensor product of channels, in a manner that imposes some technical restrictions on how the side-information registers can be updated.

In another approach introduced in [46], the assumption of identical marginals is relaxed by allowing Alice to have different single-round marginal states—yet in a tensor product form—across rounds. However, this approach is still somewhat restrictive compared to what is presented here. Specifically, the sequence of channels in that work must be in a tensor product with each other, whereas here we allow the composition of channels.

In this chapter, we derive a new variant of the Rényi entropy accumulation theorem, which not only allows imposing different marginal constraints for each channel, but also permits updating the side-information registers in a similar manner to the GEAT. The latter property further enables "fully adaptive" updating of the "statistical estimation" procedure over the course of the protocol, as described in Chapter 4. We obtain this version of EAT, which we term the **marginal-constrained entropy accumulation theorem** (MEAT), by combining the result on superadditivity of channel conditional entropy with the QPE-inspired method developed in Chapter 4 to handle statistical analysis. This enables our new variant of entropy accumulation to be applied directly to PM-QKD protocols while benefiting from the tight analysis in that chapter. We also develop a slight generalization of the notion of **f-weighted Rényi entropies** from [126], which may be useful for some more general protocols.

A central technical challenge underlying our construction is understanding the additivity properties of a certain family of channel entropies. Additivity questions are among the fundamental problems in quantum information theory, concerning whether certain entropic quantities associated with quantum channels are additive when multiple channels are used together. At its core, the question seeks to determine whether the optimal performance of quantum channels, measured under these quantities, can be achieved independently for each channel or if there is an advantage to joint usage. When the entropic quantity is additive for IID copies of the same channel, this is referred to as *weak additivity*, whereas when additivity holds for arbitrary tensor products of different channels, it is known as *strong additivity*. This problem arises in various areas of quantum information theory, for example, in the context of channel capacities [1, 108, 69].

In the setting presented in this chapter, the additivity properties of a suitably defined channeled entropy are the key to establishing the MEAT framework. Therefore, we begin by presenting a family of channel conditional entropies defined based on sandwiched Rényi divergences [93], which essentially quantify the minimum output conditional entropy of a given quantum channel. In other words, given a channel $\mathcal{M} \in \mathrm{CPTP}(\widetilde{AA}, BC)$ and (optionally) some state ψ_A on register A, we denote the channel conditional entropy by

$H_\alpha^\uparrow(\mathcal{M}, B, [\psi_A])$ and define it as

$$H_\alpha^\uparrow(\mathcal{M}, B, [\psi_A]) := \inf_{\rho_{A\widetilde{A}\widetilde{R}}\ \rho_A=\psi_A} H_\alpha^\uparrow(B|C\widetilde{R})_{\mathcal{M}[\rho]}, \tag{5.1}$$

where the optimization is over all density operators $\rho_{A\widetilde{A}\widetilde{R}}$ with a marginal constraint on A, and $\widetilde{R}$ is a stabilizing register.

This family can be viewed as a generalization of the channel entropies studied in [55, 134], in that we can recover the definitions in those works by setting the C register to be trivial and omitting the marginal constraint (equivalently, setting A to be trivial) in the above definition.[1]

We extend some results from those works (see also [70, 33, 131, 56, 111]) by demonstrating that for $\alpha \geq 1$, the above version of channel conditional entropy (Eq.(5.1)) is also strongly additive across tensor products of channels. More broadly, we prove a chain rule that generalizes this property by showing that this quantity is superadditive under channel composition in a particular form. The proofs used here are based on techniques recently developed in [44], which employed a variational characterization of measured divergences to derive various chain rules, followed by a regularization argument to convert the measured divergences to sandwiched divergences. We will then show that these results can be used to construct a marginal-constrained EAT-style framework, which we aim to prove.

We highlight that the result of strong additivity across tensor products has been independently obtained in [46], using different proof techniques. That work also derives a number of other chain rules; however, those results are "incomparable to" (i.e., neither strictly stronger nor weaker than) our chain rule for channel composition. This is mainly because our chain rule incorporates marginal constraints, making it more suitable for prepare-and-measure QKD protocols—for such protocols, it yields a more general form of "adaptive" protocol as compared to the results in [46].

We note, however, that more generally (outside the context of PM-QKD), our result is neither strictly stronger nor weaker than the GEAT. This is because, while we allow for marginal constraints (which the GEAT did not accommodate), we do not allow for "secret" internal memory registers that were present in the GEAT.

This chapter is based on [9].

[1] A different notion of channel entropy was studied in [27]; apart from the marginal constraint, it also differs from our definition in terms of some technical points involving the conditioning and stabilizing registers. However, in special cases the definitions coincide, for instance as described in [27, Theorem 5] for the case of von Neumann entropy.

5.2 Rényi channel conditional entropy

We begin with a slight generalization of the minimized channel divergence in [44] that incorporates some form of marginal-state constraints. It is worth noting that in these definitions, as opposed to the usual channel divergence defined in Definition 2.2.4, the stabilizing register no longer appears, since it involves minimization over input states instead of the maximization that appears in common channel divergence quantities. Therefore, as long as the underlying divergence satisfies data processing, including a stabilizing register makes no difference.

Definition 5.2.1. (Minimized channel divergence) For $\mathcal{M} \in \mathrm{CPTP}(A\widetilde{A}, B), \mathcal{N} \in \mathrm{CP}(A'\widetilde{A'}, B)$, and $\psi_A \in \mathrm{S}_=(A)$, $\phi_{A'} \in \mathrm{S}_=(A')$, the (marginal-constrained) minimized channel divergence between $\mathcal{M}, \mathcal{N}$ is defined as:

$$\mathbb{D}^{\inf}(\mathcal{M}, [\psi_A] \| \mathcal{N}, [\phi_{A'}]) = \inf_{\substack{\rho \in \mathrm{S}_=(A\widetilde{A}) \text{ s.t. } \rho_A = \psi_A \\ \sigma \in \mathrm{S}_=(A'\widetilde{A'}) \text{ s.t. } \sigma_{A'} = \phi_{A'}}} \mathbb{D}(\mathcal{M}[\rho] \| \mathcal{N}[\sigma]), \tag{5.2}$$

where $\mathbb{D}$ can be any of the divergences defined in this thesis. Its regularized version is defined as

$$\mathbb{D}^{\inf,\mathrm{reg}}(\mathcal{M}, [\psi_A] \| \mathcal{N}, [\phi_{A'}]) = \lim_{m \to \infty} \frac{1}{m} \mathbb{D}^{\inf}(\mathcal{M}^{\otimes m}, [\psi_A^{\otimes m}] \| \mathcal{N}^{\otimes m}, [\phi_{A'}^{\otimes m}]), \tag{5.3}$$

if the limit exists. We view the marginal-state specifications in the above notation as "optional arguments", in that if we do not wish to impose a marginal constraint on the registers A, A', or both, then we denote the corresponding minimized divergences as $\mathbb{D}^{\inf}(\mathcal{M} \| \mathcal{N}, [\phi_{A'}])$, $\mathbb{D}^{\inf}(\mathcal{M}, [\psi_A] \| \mathcal{N})$, and $\mathbb{D}^{\inf}(\mathcal{M} \| \mathcal{N})$, respectively, e.g.

$$\mathbb{D}^{\inf}(\mathcal{M}, [\psi_A] \| \mathcal{N}) = \inf_{\substack{\rho \in \mathrm{S}_=(A\widetilde{A}) \text{ s.t. } \rho_A = \psi_A \\ \sigma \in \mathrm{S}_=(A'\widetilde{A'})}} \mathbb{D}(\mathcal{M}[\rho] \| \mathcal{N}[\sigma]), \tag{5.4}$$

and so on. Moreover, the regularized versions for these cases are defined and represented similarly, if the corresponding limits exist.

Generalizing the definitions from [55, 134], we also define conditional entropies with marginal-state constraints. For these entropies, since we instead have minimization of conditional entropies, introducing a stabilizing register does make a difference, and so we define them with a stabilizing register.

Definition 5.2.2. (Rényi channel conditional entropy) For any channel $\mathcal{M} \in \mathrm{CPTP}(A\widetilde{A}, BC)$, any $\psi_A \in \mathrm{S}_=(A)$, and $\alpha \in [0, \infty]$, the (marginal-constrained) Rényi channel conditional entropy is defined as[2]

$$H_\alpha^\uparrow(\mathcal{M}, B, [\psi_A]) = \inf_{\rho \in \mathrm{S}_=(A\widetilde{A}\widetilde{R}) \text{ s.t. } \rho_A = \psi_A} H_\alpha^\uparrow(B|C\widetilde{R})_{\mathcal{M}[\rho]}, \tag{5.5}$$

where $\widetilde{R}$ is isomorphic to $A\widetilde{A}$. Note that the above optimization can be restricted to the set of pure states, since, by data-processing, purifying the input state only decreases the conditional entropy of the states in the above optimization. Its regularized version is defined as

$$H_\alpha^{\uparrow,\mathrm{reg}}(\mathcal{M}, B, [\psi_A]) = \lim_{m\to\infty} \frac{1}{m} H_\alpha^\uparrow(\mathcal{M}^{\otimes m}, B^m, [\psi_A^{\otimes m}]), \tag{5.6}$$

noting that the limit always exists by Fekete's lemma [48, 32].

Note that the regularized Rényi channel conditional entropy is additive across IID copies of the channel, i.e., weakly additive:

$$\begin{aligned} H_\alpha^{\uparrow,\mathrm{reg}}(\mathcal{M}^{\otimes n}, B^n, [\psi_A^{\otimes n}]) &= \lim_{m\to\infty} \frac{1}{m} H_\alpha^\uparrow(\mathcal{M}^{\otimes nm}, B^{nm}, [\psi_A^{\otimes nm}]) \\ &= n \lim_{m'\to\infty} \frac{1}{m'} H_\alpha^\uparrow(\mathcal{M}^{\otimes m'}, B^{m'}, [\psi_A^{\otimes m'}]) \\ &= n H_\alpha^{\uparrow,\mathrm{reg}}(\mathcal{M}, B, [\psi_A]), \end{aligned} \tag{5.7}$$

where the second line holds via a change of indices (and noting that since the $m' \to \infty$ limit exists, any subsequence of it converges to the same limit).

5.3 Chain rule and additivity

In this section, we prove that the (marginal-constrained) Rényi channel conditional entropies are superadditive under channel composition. As a special case, we also show that this quantity is strongly additive under tensor products. We begin by proving weak additivity of these entropies.

[2] As we have assumed all registers are finite-dimensional in this chapter, this infimum is always finite.

5.3.1 Weak additivity under tensor products

Lemma 5.3.1. *(Weak additivity) Consider a channel* $\mathcal{E} \in \mathrm{CPTP}(AY, X\widehat{Y})$ *and a state* $\psi_A \in \mathrm{S}_=(A)$*. For any* $\alpha \in [1, \infty]$*, and any* $m \in \mathbb{N}$*, we have*

$$H_\alpha^\uparrow\left(\mathcal{E}, X, [\psi_A]\right) = H_\alpha^{\uparrow,\mathrm{reg}}\left(\mathcal{E}, X, [\psi_A]\right) = \frac{1}{m} H_\alpha^\uparrow\left(\mathcal{E}^{\otimes m}, X^m, [\psi_A^{\otimes m}]\right). \tag{5.8}$$

Proof. This result was previously proven in [126], but for completeness, we include a proof here. First suppose $\alpha > 1$. The first equality in (5.8) is established using nearly the same argument as in the proof of [89, Lemma 3.5][3], just replacing the de Finetti theorem used in that argument with a generalized version described in [47, Corollary 3.2]. Subsequently, the second equality holds by the weak additivity property of the regularized channel conditional entropy in (5.7).

To prove the case where $\alpha = 1$, we first note that Rényi conditional entropy $H_\alpha^\uparrow(Q|Q')$ uniformly converges to von Neumann conditional entropy $H(Q|Q')$ as $\alpha \to 1^+$. More specifically, by [41, Lemma B.9], we have the following uniform continuity bound:

$$H_\alpha^\uparrow(Q|Q') - H(Q|Q') \geq (\alpha - 1) K_{\dim(Q)} \qquad \forall \alpha \in \left(1, \frac{1}{\log(1 + 2\dim(Q))}\right), \tag{5.9}$$

where $K_{\dim(Q)}$ is a constant depending only on the system dimensions. Hence, we have the

[3] A similar approach has been used in previous work, for instance [55].

following chain of relations:

$$
\begin{aligned}
&\forall \alpha > 1 \qquad H_\alpha^\uparrow\left(\mathcal{E}, X, [\psi_A]\right) = \frac{1}{m} H_\alpha^\uparrow\left(\mathcal{E}^{\otimes m}, X^m, [\psi_A^{\otimes m}]\right) \\
\iff &\forall \alpha > 1 \qquad \inf_{\substack{\rho \in \mathrm{S}_=(AY\widetilde{R}) \\ \text{s.t. } \rho_A = \psi_A}} H_\alpha^\uparrow(X|\widehat{Y}\widetilde{R})_{\mathcal{E}[\rho]} = \inf_{\substack{\rho \in \mathrm{S}_=(A^m Y^m \widetilde{R}) \\ \text{s.t. } \rho_{A^m} = \psi_A^{\otimes m}}} \frac{1}{m} H_\alpha^\uparrow(X^m|\widehat{Y}^m\widetilde{R})_{\mathcal{E}^{\otimes m}[\rho]} \\
\implies &\lim_{\alpha \to 1^+} \inf_{\substack{\rho \in \mathrm{S}_=(AY\widetilde{R}) \\ \text{s.t. } \rho_A = \psi_A}} H_\alpha^\uparrow(X|\widehat{Y}\widetilde{R})_{\mathcal{E}[\rho]} = \lim_{\alpha \to 1^+} \inf_{\substack{\rho \in \mathrm{S}_=(A^m Y^m \widetilde{R}) \\ \text{s.t. } \rho_{A^m} = \psi_A^{\otimes m}}} \frac{1}{m} H_\alpha^\uparrow(X^m|\widehat{Y}^m\widetilde{R})_{\mathcal{E}^{\otimes m}[\rho]} \\
\iff &\inf_{\substack{\rho \in \mathrm{S}_=(AY\widetilde{R}) \\ \text{s.t. } \rho_A = \psi_A}} \lim_{\alpha \to 1^+} H_\alpha^\uparrow(X|\widehat{Y}\widetilde{R})_{\mathcal{E}[\rho]} = \inf_{\substack{\rho \in \mathrm{S}_=(A^m Y^m \widetilde{R}) \\ \text{s.t. } \rho_{A^m} = \psi_A^{\otimes m}}} \lim_{\alpha \to 1^+} \frac{1}{m} H_\alpha^\uparrow(X^m|\widehat{Y}^m\widetilde{R})_{\mathcal{E}^{\otimes m}[\rho]} \\
\iff &\inf_{\substack{\rho \in \mathrm{S}_=(AY\widetilde{R}) \\ \text{s.t. } \rho_A = \psi_A}} H(X|\widehat{Y}\widetilde{R})_{\mathcal{E}[\rho]} = \inf_{\substack{\rho \in \mathrm{S}_=(A^m Y^m \widetilde{R}) \\ \text{s.t. } \rho_{A^m} = \psi_A^{\otimes m}}} \frac{1}{m} H(X^m|\widehat{Y}^m\widetilde{R})_{\mathcal{E}^{\otimes m}[\rho]} \\
\iff &\qquad H\left(\mathcal{E}, X, [\psi_A]\right) = \frac{1}{m} H\left(\mathcal{E}^{\otimes m}, X^m, [\psi_A^{\otimes m}]\right), \qquad (5.10)
\end{aligned}
$$

where the fourth line holds due to the uniform convergence established in Eq. (5.9) (which allows us to swap the infimum and one-sided limit). To obtain the first equality in (5.8) for $\alpha = 1$, we take the limit of $m \to \infty$ from the last line in (5.10). □

5.3.2 Duality

In this part, we use the duality relation in Proposition 2.2.2 to establish a connection between channel conditional entropy and minimized channel divergence. Through the rest of this section, we will frequently be using the notation $\mathbb{I}_{Q'} \otimes \mathrm{id}_Q \in \mathrm{CP}(Q, QQ')$ to denote the CP map that simply outputs the tensor product of the input state on Q with the identity operator $\mathbb{I}_{Q'}$, i.e.[4]

$$(\mathbb{I}_{Q'} \otimes \mathrm{id}_Q)[\rho_Q] \coloneqq \mathbb{I}_{Q'} \otimes \rho_Q. \qquad (5.11)$$

Lemma 5.3.2. *Consider a channel $\mathcal{E} \in \mathrm{CPTP}(AY, X\widehat{Y})$ and a state $\psi_A \in \mathrm{S}_=(A)$. Let $V \in \mathrm{CPTP}(AY, X\widehat{Y}Z)$ be any Stinespring dilation of $\mathcal{E}$. Then, for any $\alpha \in [\frac{1}{2}, \infty]$ we have:*

$$H_\alpha^\uparrow\left(\mathcal{E}, X, [\psi_A]\right) = D_\beta^{\inf}\left(\mathrm{Tr}_{\widehat{Y}} \circ V, [\psi_A] \| \mathbb{I}_X \otimes \mathrm{id}_Z\right), \qquad (5.12)$$

where $\beta = \frac{\alpha}{2\alpha - 1} \in [\frac{1}{2}, \infty]$.

[4] This construction can alternatively be written in terms of replacement channels [68].

Proof. For any pure state $\rho \in \mathrm{S}_{=}(AY\widetilde{R})$, $V[\rho]$ is also a pure state. Thus, we have:

$$\begin{aligned} H_\alpha^\uparrow(X|\widehat{Y}\widetilde{R})_{\mathcal{E}[\rho]} &= -H_\beta^\uparrow(X|Z)_{V[\rho]} \\ &= -H_\beta^\uparrow(X|Z)_{\mathrm{Tr}_{\widehat{Y}} \circ V[\rho]} \\ &= \inf_{\sigma_Z \in \mathrm{S}_{=}(Z)} D_\beta\left(\mathrm{Tr}_{\widehat{Y}} \circ V[\rho] \| \mathbb{I}_X \otimes \sigma_Z\right), \end{aligned} \tag{5.13}$$

where the first line follows from Proposition 2.2.2, and the last line holds by the definition of Rényi conditional entropies in Definition 2.2.5. Taking the infimum over all pure $\rho \in \mathrm{S}_{=}(AY\widetilde{R})$ such that $\rho_A = \psi_A$ proves the claim. □

Having established this lemma, we can now combine its result with the weak additivity property in Lemma 5.3.1 to derive several further relationships that link channel conditional entropy to minimized channel divergence and their regularized counterparts.

Corollary 5.3.1. *Let $\mathcal{E} \in \mathrm{CPTP}(AY, X\widehat{Y})$ be a channel, $\psi_A \in \mathrm{S}_{=}(A)$ be a state, and $V \in \mathrm{CPTP}(AY, X\widehat{Y}Z)$ be any Stinespring dilation of $\mathcal{E}$. Then, for any $\alpha \in [1, \infty]$, we have*

$$H_\alpha^\uparrow(\mathcal{E}, X, [\psi_A]) = D_\beta^{\inf}\left(\mathrm{Tr}_{\widehat{Y}} \circ V, [\psi_A] \| \mathbb{I}_X \otimes \mathrm{id}_Z\right) \tag{5.14}$$

$$= \frac{1}{m} H_\alpha^\uparrow\left(\mathcal{E}^{\otimes m}, X^m, [\psi_A^{\otimes m}]\right) = \frac{1}{m} D_\beta^{\inf}\left(\mathrm{Tr}_{\widehat{Y}}^{\otimes m} \circ V^{\otimes m}, [\psi_A^{\otimes m}] \| \mathbb{I}_X^{\otimes m} \otimes \mathrm{id}_{Z^m}\right) \tag{5.15}$$

$$= H_\alpha^{\uparrow,\mathrm{reg}}(\mathcal{E}, X, [\psi_A]) = D_\beta^{\inf,\mathrm{reg}}\left(\mathrm{Tr}_{\widehat{Y}} \circ V, [\psi_A] \| \mathbb{I}_X \otimes \mathrm{id}_Z\right), \tag{5.16}$$

where $\beta = \frac{\alpha}{2\alpha - 1} \in [\frac{1}{2}, 1]$. In particular, this implies $D_\beta^{\inf,\mathrm{reg}}\left(\mathrm{Tr}_{\widehat{Y}} \circ V, [\psi_A] \| \mathbb{I}_X \otimes \mathrm{id}_Z\right)$ exists, in that the limit in its definition indeed converges.

Proof. The equality between the leftmost expressions in each line is just Lemma 5.3.1. The middle equalities in the first two lines follow from Lemma 5.3.2 (noting $V^{\otimes m}$ is a valid Stinespring dilation of $\mathcal{E}^{\otimes m}$), and the middle equality in the third line, holds by taking the $m \to \infty$ limit on both expressions in the second line. □

5.3.3 Superadditivity under channel composition

Before demonstrating the main result of this part, we show that the minimized measured Rényi channel divergence is superadditive under the composition of channels. Since the proof of this property relies on Lemmas 22 and 24 from [44], we first restate those lemmas.

Fact 5.3.1. *(Lemma 22 and 24 in [44]) Consider any sets* $\mathbf{D}_1(A_1) \subseteq \mathrm{S}_=(A_1)$, $\mathbf{D}_2(A_2) \subseteq \mathrm{S}_=(A_2)$, *and* $\mathbf{D}_3(A_1A_2) \subseteq \mathrm{S}_=(A_1A_2)$ *and also* $\mathbf{D}'_1(A_1) \subseteq \mathrm{Pos}(A_1)$, $\mathbf{D}'_2(A_2) \subseteq \mathrm{Pos}(A_2)$, *and* $\mathbf{D}'_3(A_1A_2) \subseteq \mathrm{Pos}(A_1A_2)$. *Suppose furthermore that all those sets are convex and compact, and the following inclusions hold:*

$$\begin{aligned}&\left\{\Gamma_{A_1} \otimes \Gamma_{A_2} \;\middle|\; \Gamma_{A_j} \in \mathrm{Pos}(A_j) \wedge \sup_{\nu \in \mathbf{D}_j(A_j)} \mathrm{Tr}[\Gamma_{A_j}\nu_{A_j}] \leq 1 \;\text{ for } j \in \{1,2\}\right\} \\ &\subseteq \left\{\Gamma_{A_1A_2} \;\middle|\; \Gamma_{A_1A_2} \in \mathrm{Pos}(A_1A_2) \wedge \sup_{\nu \in \mathbf{D}_3(A_1A_2)} \mathrm{Tr}[\Gamma_{A_1A_2}\nu_{A_1A_2}] \leq 1\right\}, \end{aligned} \tag{5.17}$$

$$\begin{aligned}&\left\{\Gamma_{A_1} \otimes \Gamma_{A_2} \;\middle|\; \Gamma_{A_j} \in \mathrm{Pos}(A_j) \wedge \sup_{\chi \in \mathbf{D}'_j(A_j)} \mathrm{Tr}[\Gamma_{A_j}\chi_{A_j}] \leq 1 \;\text{ for } j \in \{1,2\}\right\} \\ &\subseteq \left\{\Gamma_{A_1A_2} \;\middle|\; \Gamma_{A_1A_2} \in \mathrm{Pos}(A_1A_2) \wedge \sup_{\chi \in \mathbf{D}'_3(A_1A_2)} \mathrm{Tr}[\Gamma_{A_1A_2}\chi_{A_1A_2}] \leq 1\right\}. \end{aligned} \tag{5.18}$$

Then for any $\alpha \in [\frac{1}{2}, 1]$, *we have the following:*

$$\inf_{\substack{\nu \in \mathbf{D}_3(A_1A_2) \\ \chi \in \mathbf{D}'_3(A_1A_2)}} D_{\mathbb{M},\alpha}(\nu\|\chi) \geq \inf_{\substack{\nu \in \mathbf{D}_1(A_1) \\ \chi \in \mathbf{D}'_1(A_1)}} D_{\mathbb{M},\alpha}(\nu\|\chi) + \inf_{\substack{\nu \in \mathbf{D}_2(A_2) \\ \chi \in \mathbf{D}'_2(A_2)}} D_{\mathbb{M},\alpha}(\nu\|\chi). \tag{5.19}$$

With this, we show superadditivity for the minimized measured Rényi channel divergence:

Lemma 5.3.3. *Let* $\mathcal{F}_1 \in \mathrm{CPTP}(A_0Y_0, X_1Y_1Z_1)$, *and* $\mathcal{F}_2 \in \mathrm{CPTP}(A_1Y_1, X_2Y_2Z_2)$ *be quantum channels. Consider states* $\psi_{A_0} \in \mathrm{S}_=(A_0)$ *and* $\phi_{A_1} \in \mathrm{S}_=(A_1)$. *Then, for any* $\alpha \in [\frac{1}{2}, 1]$, *we have*

$$\begin{aligned}&D^{\inf}_{\mathbb{M},\alpha}\left(\mathrm{Tr}_{Y_2} \circ \mathcal{F}_2 \circ \mathcal{F}_1, [\psi_{A_0} \otimes \phi_{A_1}] \| \mathbb{I}_{X_1X_2} \otimes \mathrm{id}_{Z_1Z_2}\right) \\ &\quad \geq D^{\inf}_{\mathbb{M},\alpha}\left(\mathrm{Tr}_{Y_2} \circ \mathcal{F}_2, [\phi_{A_1}] \| \mathbb{I}_{X_2} \otimes \mathrm{id}_{Z_2}\right) + D^{\inf}_{\mathbb{M},\alpha}\left(\mathrm{Tr}_{Y_1} \circ \mathcal{F}_1, [\psi_{A_0}] \| \mathbb{I}_{X_1} \otimes \mathrm{id}_{Z_1}\right). \end{aligned} \tag{5.20}$$

Proof. Consider Fact 5.3.1 with the following choices of sets: let $\mathbf{D}_1(X_1Z_1), \mathbf{D}_2(X_2Z_2)$ and $\mathbf{D}_3(X_1Z_1X_2Z_2)$ be

$$\begin{aligned}\mathbf{D}_1(X_1Z_1) &\coloneqq \{\mathrm{Tr}_{Y_1} \circ \mathcal{F}_1[\rho_{A_0Y_0}] |\; \rho_{A_0Y_0} \in \mathrm{S}_=(A_0Y_0) \wedge \rho_{A_0} = \psi_{A_0}\} \\ \mathbf{D}_2(X_2Z_2) &\coloneqq \{\mathrm{Tr}_{Y_2} \circ \mathcal{F}_2[\rho_{A_1Y_1}] |\; \rho_{A_1Y_1} \in \mathrm{S}_=(A_1Y_1) \wedge \rho_{A_1} = \phi_{A_1}\} \\ \mathbf{D}_3(X_1Z_1X_2Z_2) &\coloneqq \{\mathrm{Tr}_{Y_2} \circ \mathcal{F}_2 \circ \mathcal{F}_1[\rho_{A_0A_1Y_0}] |\; \rho \in \mathrm{S}_=(A_0A_1Y_0) \wedge \rho_{A_0A_1} = \psi_{A_0} \otimes \phi_{A_1}\}, \end{aligned} \tag{5.21}$$

and furthermore, let $\mathbf{D}'_1(X_1Z_1), \mathbf{D}'_2(X_2Z_2)$ and $\mathbf{D}'_3(X_1Z_1X_2Z_2)$ be

$$\begin{aligned}
\mathbf{D}'_1(X_1Z_1) &\coloneqq \{\mathbb{I}_{X_1} \otimes \sigma_{Z_1} |\ \sigma_{Z_1} \in \mathrm{S}_=(Z_1)\} \\
\mathbf{D}'_2(X_2Z_2) &\coloneqq \{\mathbb{I}_{X_2} \otimes \sigma_{Z_2} |\ \sigma_{Z_2} \in \mathrm{S}_=(Z_2)\} \\
\mathbf{D}'_3(X_1Z_1X_2Z_2) &\coloneqq \{\mathbb{I}_{X_1X_2} \otimes \sigma_{Z_1Z_2} |\ \sigma_{Z_1Z_2} \in \mathrm{S}_=(Z_1Z_2)\}\,.
\end{aligned} \tag{5.22}$$

With this, the bound (5.19) in Fact 5.3.1 is equivalent to our desired bound (5.20), so it suffices to show that the conditions in Fact 5.3.1 are satisfied.[5]

First, we note that the sets $\mathbf{D}_1(X_1Z_1), \mathbf{D}_2(X_2Z_2)$, and $\mathbf{D}_3(X_1Z_1X_2Z_2)$ are convex, bounded, and closed; therefore they are compact. Similarly, $\mathbf{D}'_1(X_1Z_1), \mathbf{D}'_2(X_2Z_2)$ and $\mathbf{D}'_3(X_1Z_1X_2Z_2)$ are also convex and compact. Therefore, it now suffices to show the inclusion relations in Eqs. (5.17) and (5.18), for the sets defined in (5.21) and (5.22), respectively.

We first consider Eq. (5.18), as it is more straightforward. We need to show that for any $\Gamma_{X_1Z_1} \in \mathrm{Pos}(X_1Z_1)$ and $\Gamma_{X_2Z_2} \in \mathrm{Pos}(X_2Z_2)$ such that

$$\sup_{\sigma \in \mathrm{S}_=(Z_1)} \mathrm{Tr}\left[\Gamma_{X_1Z_1}\left(\mathbb{I}_{X_1} \otimes \sigma_{Z_1}\right)\right] \leq 1 \quad \text{and} \quad \sup_{\sigma \in \mathrm{S}_=(Z_2)} \mathrm{Tr}\left[\Gamma_{X_2Z_2}\left(\mathbb{I}_{X_2} \otimes \sigma_{Z_2}\right)\right] \leq 1, \tag{5.23}$$

we have

$$\sup_{\sigma \in \mathrm{S}_=(Z_1Z_2)} \mathrm{Tr}\left[\left(\Gamma_{X_0Z_0} \otimes \Gamma_{X_1Z_1}\right)\left(\mathbb{I}_{X_1X_2} \otimes \sigma_{Z_1Z_2}\right)\right] \leq 1. \tag{5.24}$$

The initial two conditions are equivalent to $\Gamma_{Z_1} \leq \mathbb{I}_{Z_1}$ and $\Gamma_{Z_2} \leq \mathbb{I}_{Z_2}$, respectively. Therefore we have

$$\begin{aligned}
&\sup_{\sigma \in \mathrm{S}_=(Z_1Z_2)} \mathrm{Tr}\left[\left(\Gamma_{X_0Z_0} \otimes \Gamma_{X_1Z_1}\right)\left(\mathbb{I}_{X_1X_2} \otimes \sigma_{Z_1Z_2}\right)\right] \\
&= \sup_{\sigma \in \mathrm{S}_=(Z_1Z_2)} \mathrm{Tr}\left[\left(\Gamma_{Z_0} \otimes \Gamma_{Z_1}\right)\left(\sigma_{Z_1Z_2}\right)\right] \\
&\leq \sup_{\sigma \in \mathrm{S}_=(Z_1Z_2)} \mathrm{Tr}\left[\sigma_{Z_1Z_2}\right] \\
&\leq 1,
\end{aligned} \tag{5.25}$$

which is the desired result.

[5]We find that some of the optimizations and SDPs in the following proof do not seem to have a very intuitive interpretation. It may be helpful to look at the proof of Corollary 5.3.3 (strong additivity) below, where the SDPs seem to have a slightly more intuitive tensor-product structure.

Let us now prove the inclusion in (5.17). In other words, we need to show that for any $\Gamma_{X_1Z_1} \in \mathrm{Pos}(X_1Z_1)$ and $\Gamma_{X_2Z_2} \in \mathrm{Pos}(X_2Z_2)$ such that

$$\sup_{\nu \in \mathbf{D}_1(X_1Z_1)} \mathrm{Tr}\left[\Gamma_{X_1Z_1}\nu_{X_1Z_1}\right] \leq 1 \quad \text{and} \quad \sup_{\nu \in \mathbf{D}_2(X_2Z_2)} \mathrm{Tr}\left[\Gamma_{X_2Z_2}\nu_{X_2Z_2}\right] \leq 1, \tag{5.26}$$

we have

$$\sup_{\nu \in \mathbf{D}_3(X_1Z_1X_2Z_2)} \mathrm{Tr}\left[\left(\Gamma_{X_1Z_1} \otimes \Gamma_{X_2Z_2}\right)\nu_{X_1Z_1X_2Z_2}\right] \leq 1. \tag{5.27}$$

The initial two conditions may be equivalently expressed as upper bounding the subsequent SDPs by 1 (note that since ψ_{A_0}, ϕ_{A_1} are normalized states by definition, we do not need to separately impose a normalization constraint):

$$\sup_{\rho \in \mathrm{Pos}(A_0Y_0)} \mathrm{Tr}\left[\rho_{A_0Y_0}\mathcal{F}_1^\dagger \circ \mathrm{Tr}_{Y_1}^\dagger[\Gamma_{X_1Z_1}]\right] \quad \text{s.t.} \quad \rho_{A_0} = \psi_{A_0} \tag{5.28}$$

$$\sup_{\rho \in \mathrm{Pos}(A_1Y_1)} \mathrm{Tr}\left[\rho_{A_1Y_1}\mathcal{F}_2^\dagger \circ \mathrm{Tr}_{Y_2}^\dagger[\Gamma_{X_2Z_2}]\right], \quad \text{s.t.} \quad \rho_{A_1} = \phi_{A_1}. \tag{5.29}$$

Additionally, Eq. (5.27) can be equivalently expressed as upper bounding the following SDP by 1:

$$\sup_{\rho \in \mathrm{Pos}(A_0A_1Y_0)} \mathrm{Tr}\left[\rho_{A_0A_1Y_0}\mathcal{F}_1^\dagger \circ \mathcal{F}_2^\dagger \circ \mathrm{Tr}_{Y_2}^\dagger\left[\Gamma_{X_0Z_0} \otimes \Gamma_{X_1Z_1}\right]\right] \quad \text{s.t.} \quad \rho_{A_0A_1} = \psi_{A_0} \otimes \phi_{A_1}. \tag{5.30}$$

Without loss of generality, we shall suppose that ψ_{A_0} and ϕ_{A_1} have full support.[6]

[6] In case ψ_{A_0} or ϕ_{A_1} do not have full support, we can instead replace all of these optimizations with restricted versions where the Hilbert spaces on A_0 and A_1 are replaced by $\mathrm{supp}(\psi_{A_0})$ and $\mathrm{supp}(\phi_{A_1})$ (including when they appear in the domains of the channels $\mathcal{F}_1, \mathcal{F}_2$). Observe that this restriction does not change the optimal values since any feasible point in the original space corresponds to a feasible solution in the reduced space, and vice versa.

The dual problems of the SDPs in Eq. (5.28), and (5.29) are as follows:

$$\inf_{\Lambda\in\mathrm{Herm}(A_0)} \mathrm{Tr}\left[\psi_{A_0}\Lambda_{A_0}\right] \quad \text{s.t.} \quad \Lambda_{A_0}\otimes\mathbb{I}_{Y_0} \geq \mathcal{F}_1^\dagger[\Gamma_{X_1Z_1}\otimes\mathbb{I}_{Y_1}] \tag{5.31}$$

$$\inf_{\Lambda\in\mathrm{Herm}(A_1)} \mathrm{Tr}\left[\phi_{A_1}\Lambda_{A_1}\right] \quad \text{s.t.} \quad \Lambda_{A_1}\otimes\mathbb{I}_{Y_1} \geq \mathcal{F}_2^\dagger[\Gamma_{X_2Z_2}\otimes\mathbb{I}_{Y_2}], \tag{5.32}$$

and the dual problem to the SDP in Eq. (5.30) is given by

$$\inf_{\Lambda\in\mathrm{Herm}(A_0A_1)} \mathrm{Tr}\left[\left(\psi_{A_0}\otimes\phi_{A_1}\right)\Lambda_{A_0A_1}\right] \quad \text{s.t.} \quad \Lambda_{A_0A_1}\otimes\mathbb{I}_{Y_1} \geq \mathcal{F}_1^\dagger\circ\mathcal{F}_2^\dagger\left[\Gamma_{X_1Z_1}\otimes\Gamma_{X_2Z_2}\otimes\mathbb{I}_{Y_2}\right]. \tag{5.33}$$

Note that since $\Gamma_{X_1Z_1}\geq 0$, $\Gamma_{X_2Z_2}\geq 0$, and the adjoint of a CP map is also CP, the constraints in Eqs. (5.31), (5.32), and (5.33) imply that $\Lambda_{A_0}\geq 0$, $\Lambda_{A_1}\geq 0$, and $\Lambda_{A_0A_1}\geq 0$. Hence without loss of generality, the optimizations in Eqs. (5.31), (5.32), and (5.33) can be restricted to $\Lambda\in\mathrm{Pos}(A_0)$, $\Lambda\in\mathrm{Pos}(A_1)$, and $\Lambda\in\mathrm{Pos}(A_0A_1)$, respectively.

For these SDPs, strong duality holds by the Slater condition [18], recalling we supposed that ψ_{A_0} and ϕ_{A_1} have full support. Specifically, $\rho_{A_0Y_0}=\psi_{A_0}\otimes\mathbb{U}_{Y_0}$, $\rho_{A_1Y_1}=\phi_{A_1}\otimes\mathbb{U}_{Y_1}$, and $\rho_{A_0A_1Y_0}=\psi_{A_0}\otimes\phi_{A_1}\otimes\mathbb{U}_{Y_0}$ are strict feasible points for the SDPs in Eqs. (5.28), (5.29), and (5.30), respectively. Moreover, $\Lambda_{A_0}=c_0\mathbb{I}_{A_0}$, $\Lambda_{A_1}=c_1\mathbb{I}_{A_1}$, and $\Lambda_{A_0A_1}=c_{01}\mathbb{I}_{A_0A_1}$, for large enough values of c_0, c_1, c_{01}, are strict feasible points for the SDPs in Eqs. (5.31), (5.32), and (5.33). Therefore, we see that the Slater condition is satisfied, and we have strong duality.

Since strong duality holds, we instead analyze the dual problems. Observe that it suffices to show that the tensor product of any feasible points of the SDPs in Eq. (5.31) and (5.32) is a feasible point for the SDP in Eq. (5.33).[7] Denote some arbitrary feasible points of the

[7]Explicitly: by fundamental infimum properties, each of the optimizations in Eq. (5.31)–(5.32) has a sequence of feasible points converging to the optimal values, which are at most 1 by hypothesis. If the tensor products of such feasible points are always feasible in the Eq. (5.33) optimization, this yields a sequence of feasible points in the latter that converge to a value of at most 1, as desired. (Alternatively, one can use the fact that by the Slater condition, the optimal values in Eq. (5.31)–(5.32) are indeed attained, and apply the subsequent arguments specifically to some *optimal* solutions for those optimizations.)

SDPs in (5.31) and (5.32) by $\Lambda^*_{A_0}$ and $\Lambda^*_{A_1}$, respectively. We have

$$\begin{aligned}
&\Lambda^*_{A_1} \otimes \mathbb{I}_{Y_1} \geq \mathcal{F}_2^\dagger[\Gamma_{X_2Z_2} \otimes \mathbb{I}_{Y_2}] \\
\iff &\Gamma_{X_1Z_1} \otimes \Lambda^*_{A_1} \otimes \mathbb{I}_{Y_1} \geq \Gamma_{X_1Z_1} \otimes \mathcal{F}_2^\dagger[\Gamma_{X_2Z_2} \otimes \mathbb{I}_{Y_2}] \\
\implies &\mathcal{F}_1^\dagger\left[\Gamma_{X_1Z_1} \otimes \mathbb{I}_{Y_1}\right] \otimes \Lambda^*_{A_1} \geq \mathcal{F}_1^\dagger\left[\Gamma_{X_1Z_1} \otimes \mathcal{F}_2^\dagger[\Gamma_{X_2Z_2} \otimes \mathbb{I}_{Y_2}]\right] \\
\implies &\Lambda^*_{A_0} \otimes \Lambda^*_{A_1} \otimes \mathbb{I}_{Y_1} \geq \mathcal{F}_1^\dagger\left[\Gamma_{X_1Z_1} \otimes \mathcal{F}_2^\dagger[\Gamma_{X_2Z_2} \otimes \mathbb{I}_{Y_2}]\right], \qquad (5.34)
\end{aligned}$$

where the first line is the constraint of the SDP in Eq. (5.32), the second step holds since $\Gamma_{X_1Z_1} \geq 0$, the third line holds because the adjoint of a CP map is also CP, and the last line follows from applying the constraint in Eq. (5.31) to the LHS of the penultimate line, and noting that we have $\Lambda^*_{A_1} \geq 0$ as mentioned above. The desired claim then follows by noting that the RHS of the last line of Eq. (5.34) is equal to the RHS of the constraint in Eq. (5.33). □

We now show that superaddivity holds for the regularization of minimized Rényi channel divergence.

Corollary 5.3.2. *For any channels $\mathcal{F}_1 \in \mathrm{CPTP}(A_0Y_0, X_1Y_1Z_1)$, and $\mathcal{F}_2 \in \mathrm{CPTP}(A_1Y_1, X_2Y_2Z_2)$, let $\psi_{A_0} \in \mathrm{S}_=(A_0)$ and $\phi_{A_1} \in \mathrm{S}_=(A_1)$ be states. Then, for any $\alpha \in [\frac{1}{2}, 1]$, we have*

$$\begin{aligned}
&D_\alpha^{\inf,\mathrm{reg}}\left(\mathrm{Tr}_{Y_2} \circ \mathcal{F}_2 \circ \mathcal{F}_1, [\psi_{A_0} \otimes \phi_{A_1}] \| \mathbb{I}_{X_1X_2} \otimes \mathrm{id}_{Z_1Z_2}\right) \\
&\qquad \geq D_\alpha^{\inf,\mathrm{reg}}\left(\mathrm{Tr}_{Y_2} \circ \mathcal{F}_2, [\phi_{A_1}] \| \mathbb{I}_{X_2} \otimes \mathrm{id}_{Z_2}\right) + D_\alpha^{\inf,\mathrm{reg}}\left(\mathrm{Tr}_{Y_1} \circ \mathcal{F}_1, [\psi_{A_0}] \| \mathbb{I}_{X_1} \otimes \mathrm{id}_{Z_1}\right). \qquad (5.35)
\end{aligned}$$

Proof. Applying Lemma 5.3.3 to $\mathcal{F}_1^{\otimes m} \in \mathrm{CPTP}(A_0^mY_0^m, X_1^mY_1^mZ_1^m)$, and $\mathcal{F}_2^{\otimes m} \in \mathrm{CPTP}(A_1^mY_1^m, X_2^mY_2^mZ_2^m)$, and dividing both sides by m, we have

$$\begin{aligned}
&\frac{1}{m} D^{\inf}_{\mathbb{M},\alpha}\left(\mathrm{Tr}_{Y_2}^{\otimes m} \circ \mathcal{F}_2^{\otimes m} \circ \mathcal{F}_1^{\otimes m}, [\psi_{A_0}^{\otimes m} \otimes \phi_{A_1}^{\otimes m}] \| \mathbb{I}_{X_1X_2}^{\otimes m} \otimes \mathrm{id}_{Z_1Z_2}^{\otimes m}\right) \\
&\quad \geq \frac{1}{m} D^{\inf}_{\mathbb{M},\alpha}\left(\mathrm{Tr}_{Y_2}^{\otimes m} \circ \mathcal{F}_2^{\otimes m}, [\phi_{A_1}^{\otimes m}] \| \mathbb{I}_{X_2}^{\otimes m} \otimes \mathrm{id}_{Z_2}^{\otimes m}\right) + \frac{1}{m} D^{\inf}_{\mathbb{M},\alpha}\left(\mathrm{Tr}_{Y_1}^{\otimes m} \circ \mathcal{F}_1^{\otimes m}, [\psi_{A_0}^{\otimes m}] \| \mathbb{I}_{X_1}^{\otimes m} \otimes \mathrm{id}_{Z_1}^{\otimes m}\right). \qquad (5.36)
\end{aligned}$$

The proof follows by taking the limit as $m \to \infty$ of the above relation, and using the result of Lemma 29, Eq. (96) in [44], noting that the sets of states being optimized over in these terms indeed satisfy the required conditions in that lemma — namely, some convexity,

compactness, and permutation-invariance conditions, along with $O(m)$ max-divergence bounds such as

$$\begin{aligned} D_\infty^{\inf}\left(\mathrm{Tr}_{Y_1}^{\otimes m}\circ\mathcal{F}_1^{\otimes m}, [\psi_{A_0}^{\otimes m}] \| \mathbb{I}_{X_1}^{\otimes m} \otimes \mathrm{id}_{Z_1}^{\otimes m}\right) &\leq D_\infty\left(\rho^*_{X_1^m Z_1^m} \| \mathbb{I}_{X_1^m} \otimes \rho^*_{Z_1^m}\right) \\ &= -H_\infty(X_1^m|Z_1^m)_{\rho^*} \\ &\leq m \log \dim(X) \quad \text{(see e.g. [117, Lemma 5.2])}, \end{aligned} \tag{5.37}$$

where ρ^* is just any state obtained by picking some arbitrary feasible point in the first argument of the $D_\infty^{\inf}$ optimization, and similarly for the other two terms. □

We are now in a position to demonstrate our main result in this section: the superadditivity of Rényi channel conditional entropy under channel composition.

Theorem 5.3.1. *Let $\mathcal{E}_1 \in \mathrm{CPTP}(A_0Y_0, X_1Y_1)$, and $\mathcal{E}_2 \in \mathrm{CPTP}(A_1Y_1, X_2Y_2)$ be quantum channels. Furthermore, consider states $\psi_{A_0} \in \mathrm{S}_=(A_0)$ and $\phi_{A_1} \in \mathrm{S}_=(A_1)$. Then, for any $\alpha \in [1, \infty]$ we have*

$$H_\alpha^\uparrow\left(\mathcal{E}_2 \circ \mathcal{E}_1, X_1X_2, [\psi_{A_0} \otimes \phi_{A_1}]\right) \geq H_\alpha^\uparrow\left(\mathcal{E}_2, X_2, [\phi_{A_1}]\right) + H_\alpha^\uparrow\left(\mathcal{E}_1, X_1, [\psi_{A_0}]\right). \tag{5.38}$$

Proof. Let $V_1 \in \mathrm{CPTP}(A_0Y_0, X_1Y_1Z_1)$, and $V_2 \in \mathrm{CPTP}(A_1Y_1, X_2Y_2Z_2)$ be any Stinespring dilations of $\mathcal{E}_1$, and $\mathcal{E}_2$, respectively. Note that $V_2 \circ V_1 \in \mathrm{CPTP}(A_0A_1Y_0, X_1X_2Y_2Z_1Z_2)$ is a valid Stinespring dilation for $\mathcal{E}_2 \circ \mathcal{E}_1$. Therefore,

$$\begin{aligned} &H_\alpha^\uparrow\left(\mathcal{E}_2 \circ \mathcal{E}_1, X_1X_2, [\psi_{A_0} \otimes \phi_{A_1}]\right) \\ &\quad= D_\beta^{\inf,\mathrm{reg}}\left(\mathrm{Tr}_{Y_2} \circ V_2 \circ V_1, [\psi_{A_0} \otimes \phi_{A_1}] \| \mathbb{I}_{X_1X_2} \otimes \mathrm{id}_{Z_1Z_2}\right) \\ &\quad\geq D_\beta^{\inf,\mathrm{reg}}\left(\mathrm{Tr}_{Y_2} \circ V_2, [\phi_{A_1}] \| \mathbb{I}_{X_2} \otimes \mathrm{id}_{Z_2}\right) + D_\beta^{\inf,\mathrm{reg}}\left(\mathrm{Tr}_{Y_1} \circ V_1, [\psi_{A_0}] \| \mathbb{I}_{X_1} \otimes \mathrm{id}_{Z_1}\right) \\ &\quad= H_\alpha^\uparrow\left(\mathcal{E}_2, X_2, [\phi_{A_1}]\right) + H_\alpha^\uparrow\left(\mathcal{E}_1, X_1, [\psi_{A_0}]\right), \end{aligned} \tag{5.39}$$

where the second and fourth lines follow from Corollary 5.3.1, and the third line is the result of Corollary 5.3.2. □

5.3.4 Strong additivity under tensor products

As a corollary to Theorem 5.3.1, we now show that the Rényi channel conditional entropies are strongly additive under tensor products.

Corollary 5.3.3. *(Strong additivity) Let* $\mathcal{E}_1 \in \mathrm{CPTP}(A_0B_0, X_0Y_0)$*, and* $\mathcal{E}_2 \in \mathrm{CPTP}(A_1B_1, X_1Y_1)$ *be quantum channels, and* $\psi_{A_0} \in \mathrm{S}_=(A_0)$*, and* $\phi_{A_1} \in \mathrm{S}_=(A_1)$ *be states. Then, for any* $\alpha \in [1, \infty]$ *we have*

$$H_\alpha^\uparrow\left(\mathcal{E}_1 \otimes \mathcal{E}_2, X_1X_2, [\psi_{A_0} \otimes \phi_{A_1}]\right) = H_\alpha^\uparrow\left(\mathcal{E}_1, X_1, [\psi_{A_0}]\right) + H_\alpha^\uparrow\left(\mathcal{E}_2, X_2, [\phi_{A_1}]\right). \tag{5.40}$$

Proof. The $\leq$ direction holds simply by choosing tensor product inputs. For the $\geq$ direction, it can be obtained as a corollary of Theorem 5.3.1 by simply taking a suitable choice of channels in that theorem.

However, we find it may be instructive to see how an argument analogous to the Theorem 5.3.1 proof would proceed in this scenario (to prove the $\geq$ direction), and hence we present this in some detail here. Let $V_1 \in \mathrm{CPTP}(A_0B_0, X_0Y_0Z_0)$, and $V_2 \in \mathrm{CPTP}(A_1B_1, X_1Y_1Z_1)$ be any Stinespring dilations of $\mathcal{E}_1$ and $\mathcal{E}_2$, respectively. Then we only need to prove the following holds:

$$\begin{aligned} &D_{\mathbb{M},\alpha}^{\inf}\left(\mathrm{Tr}_{Y_0Y_1} \circ (V_2 \otimes V_1), [\psi_{A_0} \otimes \phi_{A_1}] \| \mathbb{I}_{X_0X_1} \otimes \mathrm{id}_{Z_0Z_1}\right) \\ &\quad \geq D_{\mathbb{M},\alpha}^{\inf}\left(\mathrm{Tr}_{Y_1} \circ V_2, [\phi_{A_1}] \| \mathbb{I}_{X_1} \otimes \mathrm{id}_{Z_1}\right) + D_{\mathbb{M},\alpha}^{\inf}\left(\mathrm{Tr}_{Y_0} \circ V_1, [\psi_{A_0}] \| \mathbb{I}_{X_0} \otimes \mathrm{id}_{Z_0}\right). \end{aligned} \tag{5.41}$$

The reason for this is that once we prove Eq. (5.41), we can derive an analogous result to Corollary 5.3.2. Then, the claim is proven by employing an argument similar to that in Theorem 5.3.1. Therefore, we focus only on proving Eq. (5.41).

We utilize an argument similar to that of Lemma 5.3.3. First, note that the proof of the inclusion relation in Eq. (5.18) is identical to that of Lemma 5.3.3. Then, we only need to prove the inclusion relation in Eq. (5.17). Let us suppose that the following SDPs are upper bounded by 1:

$$\sup_{\rho \in \mathrm{Pos}(A_0B_0)} \mathrm{Tr}\left[\rho_{A_0B_0} V_1^\dagger \circ \mathrm{Tr}_{Y_0}^\dagger[\Gamma_{X_0Z_0}]\right] \quad \text{s.t.} \quad \rho_{A_0} = \psi_{A_0} \tag{5.42}$$

$$\sup_{\rho \in \mathrm{Pos}(A_1B_1)} \mathrm{Tr}\left[\rho_{A_1B_1} V_2^\dagger \circ \mathrm{Tr}_{Y_1}^\dagger[\Gamma_{X_1Z_1}]\right], \quad \text{s.t.} \quad \rho_{A_1} = \phi_{A_1}. \tag{5.43}$$

Then, we only need to show the SDP

$$\sup_{\rho \in \mathrm{Pos}(A_0A_1B_0B_1)} \mathrm{Tr}\left[\rho_{A_0A_1B_0B_1} V_1^\dagger \otimes V_2^\dagger \circ \mathrm{Tr}_{Y_0Y_1}^\dagger\left[\Gamma_{X_0Z_0} \otimes \Gamma_{X_1Z_1}\right]\right] \quad \text{s.t.} \quad \rho_{A_0A_1} = \psi_{A_0} \otimes \phi_{A_1}. \tag{5.44}$$

is bounded from above by 1. Following a similar reasoning as in Lemma 5.3.3, where strong duality is established, we will focus on analyzing the dual problems. The duals corresponding to the SDPs in (5.42) and (5.43) are as follows:

$$\begin{aligned} &\inf_{\Lambda\in\mathrm{Herm}(A_0)} \mathrm{Tr}\left[\psi_{A_0}\Lambda_{A_0}\right] \\ &\text{s.t.} \\ &\Lambda_{A_0}\otimes\mathbb{I}_{B_0} \geq V_1^\dagger[\Gamma_{X_0Z_0}\otimes\mathbb{I}_{Y_0}] \end{aligned} \tag{5.45}$$

$$\begin{aligned} &\inf_{\Lambda\in\mathrm{Herm}(A_1)} \mathrm{Tr}\left[\phi_{A_1}\Lambda_{A_1}\right] \\ &\text{s.t.} \\ &\Lambda_{A_1}\otimes\mathbb{I}_{B_1} \geq V_2^\dagger[\Gamma_{X_1Z_1}\otimes\mathbb{I}_{Y_1}], \end{aligned} \tag{5.46}$$

and the dual problem to the SDP in (5.44) is

$$\begin{aligned} &\inf_{\Lambda\in\mathrm{Herm}(A_0A_1)} \mathrm{Tr}\left[\left(\psi_{A_0}\otimes\phi_{A_1}\right)\Lambda_{A_0A_1}\right] \\ &\text{s.t.} \\ &\Lambda_{A_0A_1}\otimes\mathbb{I}_{B_0B_1} \geq V_1^\dagger\left[\Gamma_{X_0Z_0}\otimes\mathbb{I}_{Y_0}\right]\otimes V_2^\dagger\left[\Gamma_{X_1Z_1}\otimes\mathbb{I}_{Y_1}\right]. \end{aligned} \tag{5.47}$$

Note that similar to the reasoning in Lemma 5.3.3, since $\Gamma_{X_0Z_0}\geq 0$, $\Gamma_{X_1Z_1}\geq 0$, and the adjoint of a CP map is also CP, the constraints in Eqs. (5.45), (5.46), and (5.47) imply that $\Lambda_{A_0}\geq 0$, $\Lambda_{A_1}\geq 0$, and $\Lambda_{A_0A_1}\geq 0$. Hence without loss of generality, the optimization in Eqs. (5.45), (5.46), and (5.47) can be restricted to $\Lambda\in\mathrm{Pos}(A_0)$, $\Lambda\in\mathrm{Pos}(A_1)$, and $\Lambda\in\mathrm{Pos}(A_0A_1)$, respectively. Also, we can again show that the problems are strictly feasible and hence strong duality holds.

Let $\Lambda^*_{A_0}$, and $\Lambda^*_{A_1}$ be arbitrary feasible points of the SDPs in (5.45) and (5.46), respectively. We can show that the product $\Lambda^*_{A_0}\otimes\Lambda^*_{A_1}$ is a feasible solution for the SDP given in (5.47), as follows:

$$\begin{aligned} &\Lambda^*_{A_0}\otimes\mathbb{I}_{B_0} \geq V_1^\dagger[\Gamma_{X_0Z_0}\otimes\mathbb{I}_{Y_0}] \\ \implies &\Lambda^*_{A_1}\otimes\mathbb{I}_{B_1}\otimes\Lambda^*_{A_0}\otimes\mathbb{I}_{B_0} \geq \Lambda^*_{A_1}\otimes\mathbb{I}_{B_1}\otimes V_1^\dagger[\Gamma_{X_0Z_0}\otimes\mathbb{I}_{Y_0}] \\ \implies &\Lambda^*_{A_0}\otimes\Lambda^*_{A_1}\otimes\mathbb{I}_{B_0B_1} \geq V_1^\dagger\left[\Gamma_{X_0Z_0}\otimes\mathbb{I}_{Y_0}\right]\otimes V_2^\dagger\left[\Gamma_{X_1Z_1}\otimes\mathbb{I}_{Y_1}\right], \end{aligned} \tag{5.48}$$

where the first line is the constraint of the SDP in Eq. (5.45), the second line holds by noting that $\Lambda^*_{A_1}\otimes\mathbb{I}_{B_1}\geq 0$ as mentioned above, and last line follows from the constraint in Eq. (5.46) along with the fact that $\Gamma_{X_0Z_0}\geq 0$ and the adjoint of a CP map is also a CP map. Hence, the relation in (5.41) holds; thus, the claim is proven. □

5.4 Marginal-constrained entropy accumulation theorem (MEAT)

In this section, we employ the results from the previous section to construct an entropy accumulation theorem that preserves some form of marginal-state constraint.

5.4.1 f-weighted $H_\alpha^{\uparrow,f}$ entropies

In order to accommodate the tight analysis in Chapter 4, we need to define a similar notion of f-weighted entropies. To do so, we begin by extending the result in Eq. (2.41) of Lemma 2.2.1 to the case where we have classical registers on both sides of the conditioning.

Lemma 5.4.1. *Let $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$ be a state classical on $\overline{C}\widehat{C}$. Then for any $\alpha \in (0,1)\cup(1,\infty]$, we have*

$$H_\alpha^{\uparrow}(Q\overline{C}|\widehat{C}Q')_\rho = \frac{\alpha}{1-\alpha}\log\left(\sum_{\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\widehat{C}})}\rho(\hat{c})\right.$$
$$\left.\inf_{\sigma_{Q'|\hat{c}}\in\mathrm{S}_=(Q')}\left[\sum_{\bar{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}|\hat{c}})}\rho(\bar{c}|\hat{c})^\alpha\, 2^{(\alpha-1)D_\alpha(\rho_{QQ'|\bar{c}\hat{c}}\|\mathbb{I}_Q\otimes\sigma_{Q'|\hat{c}})}\right]^{\frac{1}{\alpha}}\right), \tag{5.49}$$

where $\boldsymbol{\rho}_{\overline{C}|\hat{c}} = \left\{\rho\left(\bar{c}|\hat{c}\right)\,\middle|\,\bar{c}\in\overline{\mathcal{C}}\right\}$, and the $\alpha=\infty$ case is to be understood by taking the $\alpha\to\infty$ limit on both sides.[8]

Proof. We first focus on the $\alpha\in(1,\infty)$ regime, taking the $\alpha\to\infty$ limit at the end to obtain the result for $\alpha=\infty$ (recalling that the LHS of Eq. (5.49) for $\alpha=\infty$ is given by that limit by definition, and hence if the equality holds for all $\alpha<\infty$, then it holds in the limit). Moreover, given any such α, it suffices to prove the equality for the case where ρ has full support on $\overline{C}\widehat{C}$, since it can then be extended to arbitrary ρ by a continuity argument. With these restrictions in mind, we apply analogous techniques to those utilized in the

[8] Some care is needed in taking the limit on the right-hand side, as it may contain expressions of the form $F(\alpha)^{\frac{1}{\alpha}}$ where $\lim_{\alpha\to\infty}F(\alpha)=0$; however, we shall argue that the limit indeed exists.

proof of [93, Proposition 9]. First, we note that by definition, we have

$$H_\alpha^\uparrow(Q\overline{C}|\widehat{C}Q')_\rho = \sup_{\sigma_{\widehat{C}Q'}\in \mathrm{S}_=(\widehat{C}Q')} -D_\alpha\left(\rho_{QQ'\overline{C}\widehat{C}}\|\mathbb{I}_{Q\overline{C}}\otimes\sigma_{\widehat{C}Q'}\right). \tag{5.50}$$

Then, by the data-processing property in Theorem 2.2.1, we can restrict the optimization to be over only σ classical on $\widehat{C}$ (by applying a pinching channel on $\widehat{C}$ in both ρ and σ, which does not change ρ), i.e., $\sigma_{\widehat{C}Q'} = \sum_{\hat{c}} \sigma(\hat{c})\,|\hat{c}\rangle\langle\hat{c}|\otimes\sigma_{Q'}^{(\hat{c})}$ for some normalized state $\sigma_{Q'}^{(\hat{c})}$.[9] Thus, from Eq. (2.39) we have

$$\begin{aligned} H_\alpha^\uparrow(Q\overline{C}|\widehat{C}Q')_\rho &= \sup_{\{\sigma(\hat{c})\},\{\sigma_{Q'}^{(\hat{c})}\}} -\frac{1}{\alpha-1}\log\left(\sum_{\bar{c}\hat{c}}\rho(\bar{c}\hat{c})^\alpha\sigma(\hat{c})^{1-\alpha}2^{(\alpha-1)D_\alpha(\rho_{QQ'|\bar{c}\hat{c}}\|\mathbb{I}_Q\otimes\sigma_{Q'}^{(\hat{c})})}\right) \\ &= \sup_{\{\sigma(\hat{c})\}}\frac{1}{1-\alpha}\log\left(\sum_{\bar{c}\hat{c}}\rho(\bar{c}\hat{c})^\alpha\sigma(\hat{c})^{1-\alpha}\inf_{\sigma^{(\hat{c})}\in\mathrm{S}_=(Q')}2^{(\alpha-1)D_\alpha(\rho_{QQ'|\bar{c}\hat{c}}\|\mathbb{I}_Q\otimes\sigma_{Q'}^{(\hat{c})})}\right). \end{aligned} \tag{5.51}$$

Let $r(\bar{c}\hat{c}) \coloneqq \rho(\bar{c}\hat{c})2^{\frac{\alpha-1}{\alpha}D_\alpha(\rho_{QQ'|\bar{c}\hat{c}}\|\mathbb{I}_Q\otimes\sigma_{Q'}^{(\hat{c})})}$, then by employing Lagrange multiplier technique, it can be seen that the infimum is attained by

$$\sigma(\hat{c}) = \frac{\left(\inf_{\sigma^{(\hat{c})}\in\mathrm{S}_=(Q')}\sum_{\bar{c}} r(\bar{c}\hat{c})^\alpha\right)^{1/\alpha}}{\sum_{\hat{c}}\left(\inf_{\sigma^{(\hat{c})}\in\mathrm{S}_=(Q')}\sum_{\bar{c}} r(\bar{c}\hat{c})^\alpha\right)^{1/\alpha}}. \tag{5.52}$$

Substituting this solution into Eq. (5.51), followed by simple algebraic simplifications, yields the claimed result for $\alpha\in(1,\infty)$. The proof for $\alpha\in[\frac{1}{2},1)$ is identical except that the optimizations over $\sigma^{(\hat{c})}$ do not switch direction when brought inside the logarithm (since $\frac{1}{1-\alpha}>0$). □

We note that in the above lemma, one could instead sum over the full alphabets. However, in that case, any term for which $\hat{c}\notin\mathrm{supp}(\boldsymbol{\rho}_{\widehat{C}})$ would be multiplied by a prefactor of $\rho(\hat{c})=0$, and therefore such terms anyway do not contribute to the sum. A similar argument applies to any pair $(\bar{c},\hat{c})$ such that $\bar{c}\notin\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}|\hat{c}})$.

We now introduce the notion of f-weighted entropies, which slightly generalizes the definition in [126] by allowing for a register $\overline{C}$ on the left side of the conditioning. These entropies also extend the ones defined in 4.3.1, by instead being based on $H_\alpha^\uparrow$.

[9] Here we choose to denote these states as normalized states $\sigma_{Q'}^{(\hat{c})}$ indexed by $\hat{c}$, rather than using the conditional-state notation $\sigma_{Q'|\hat{c}}$ — this is because in the rest of our analysis, they will only form "dummy variables" in the optimizations that may not always have straightforward interpretations as genuine conditional states.

Definition 5.4.1. (f-weighted Rényi entropies) Let $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$ be a state where $\overline{C}$ and $\widehat{C}$ are classical with alphabets $\overline{\mathcal{C}}$ and $\widehat{\mathcal{C}}$ respectively. A **tradeoff function on** $\overline{C}\widehat{C}$ is simply a function $f : \overline{\mathcal{C}} \times \widehat{\mathcal{C}} \to \mathbb{R}$; equivalently, we may denote it as a real-valued tuple $\mathbf{f} \in \mathbb{R}^{|\overline{\mathcal{C}}\times\widehat{\mathcal{C}}|}$ where each term in the tuple specifies the value $f(\bar{c}\hat{c})$. Given a tradeoff function f and a value $\alpha \in (1,\infty)$, the **f-weighted entropy of order** α for ρ is defined as

$$
\begin{aligned}
&H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho := \\
&\frac{\alpha}{1-\alpha}\log\left(\sum_{\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\widehat{C}})}\inf_{\sigma^{(\hat{c})}\in\mathrm{S}_=(Q')}\left[\sum_{\bar{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}|\hat{c}})}\rho(\bar{c}\hat{c})^\alpha\, 2^{(\alpha-1)\left(f(\bar{c}\hat{c})+D_\alpha(\rho_{QQ'|\bar{c}\hat{c}}\|\mathbb{I}_Q\otimes\sigma^{(\hat{c})}_{Q'})\right)}\right]^{\frac{1}{\alpha}}\right) \\
&= \frac{\alpha}{1-\alpha}\log\left(\sum_{\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\widehat{C}})}\rho(\hat{c})\inf_{\sigma^{(\hat{c})}\in\mathrm{S}_=(Q')}\left[\sum_{\bar{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}|\hat{c}})}\rho(\bar{c}|\hat{c})^\alpha\, 2^{(\alpha-1)\left(f(\bar{c}\hat{c})+D_\alpha(\rho_{QQ'|\bar{c}\hat{c}}\|\mathbb{I}_Q\otimes\sigma^{(\hat{c})}_{Q'})\right)}\right]^{\frac{1}{\alpha}}\right),
\end{aligned}
\tag{5.53}
$$

where $\boldsymbol{\rho}_{\overline{C}|\hat{c}} = \left\{\rho\left(\bar{c}|\hat{c}\right) \,\middle|\, \bar{c}\in\overline{\mathcal{C}}\right\}$. Moreover, we extend the definition to $\alpha = \infty$, by taking the limit $\alpha \to \infty$. Note that f-weighted entropies can also be represented by Schatten norms by the following construction: given ρ, α and f, let us write $\mathbf{x}^{\rho,\sigma,\alpha,f}_{|\hat{c}}$ (for any $\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\widehat{C}})$) to denote a vector in $\mathbb{R}^{|\,\mathrm{supp}(\boldsymbol{\rho}_{\overline{C}|\hat{c}})|}$ with components specified by

$$
x^{\rho,\sigma,\alpha,f}_{|\hat{c}}(\bar{c}) := \rho(\bar{c}|\hat{c})2^{\frac{\alpha-1}{\alpha}\left(f(\bar{c}\hat{c})+D_\alpha(\rho_{QQ'|\bar{c}\hat{c}}\|\mathbb{I}_Q\otimes\sigma_{Q'})\right)}. \tag{5.54}
$$

Then, Eq. (5.53) can be written as

$$
\begin{aligned}
H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho &= \frac{\alpha}{1-\alpha}\log\left(\sum_{\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\widehat{C}})}\rho(\hat{c})\inf_{\sigma^{(\hat{c})}\in\mathrm{S}_=(Q')}\left\|\mathbf{x}^{\rho,\sigma^{(\hat{c})},\alpha,f}_{|\hat{c}}\right\|_\alpha\right) \\
&= \frac{\alpha}{1-\alpha}\log\mathbb{E}_{\boldsymbol{\rho}_{\widehat{C}}}\inf_{\sigma\in\mathrm{S}_=(Q')}\left\|\mathbf{x}^{\rho,\sigma,\alpha,f}_{|\widehat{C}}\right\|_\alpha,
\end{aligned}
\tag{5.55}
$$

where in the last expression we have suppressed the index $\hat{c}$ in the "dummy variables" $\sigma^{(\hat{c})}$, to avoid ambiguity in the random variable notation (with the understanding that the value of the infimum will be a function of the random variable $\widehat{C}$ in that expression).

For $\alpha \in [\frac{1}{2}, 1)$, we define $H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho$ analogously, but with the infima over $\sigma^{(\hat{c})}$ replaced by suprema.

Qualitatively, the registers $\overline{C}$ and $\widehat{C}$ can be respectively viewed as "secret" and "public" classical registers, on which we define a function f that captures some properties of the Rényi entropy of the state, as we now describe.

Note that when $f(\bar{c}\hat{c}) = 0$, $H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho$ is exactly equal to the sandwiched Rényi conditional entropy $H_\alpha^\uparrow(Q\overline{C}|\widehat{C}Q')_\rho$. Furthermore, when the $\overline{C}$ register is trivial it reduces to

$$H_\alpha^{\uparrow,f}(Q|\widehat{C}Q')_\rho = \frac{\alpha}{1-\alpha}\log\left(\sum_{\hat{c}}\rho(\hat{c})\,2^{\left(\frac{1-\alpha}{\alpha}\right)\left(H_\alpha^\uparrow(Q|Q')_{\rho_{|\hat{c}}}-f(\hat{c})\right)}\right), \tag{5.56}$$

which is exactly the f-weighted entropy defined in [126]. A similar intuition to that described in Chapter 4 can also be developed here, through the concept of a "log-mean-exponential":

$$\operatorname*{lme}_{P_X}{}_b\left[g(X)\right] \coloneqq \log_b\left(\sum_x P_X(x)\,b^{g(x)}\right), \tag{5.57}$$

which allows us to rewrite (5.56) as

$$H_\alpha^{\uparrow,f}(Q|\widehat{C}Q')_\rho = \operatorname*{lme}_{\boldsymbol{\rho}_{\widehat{C}}}{}_b\left[H_\alpha^\uparrow(Q|Q')_{\rho_{|\widehat{C}}} - f(\widehat{C})\right], \quad \text{where } b = 2^{\frac{1-\alpha}{\alpha}} \in (0,\infty). \tag{5.58}$$

Thus we observe that having a tradeoff function $f(\widehat{C})$ that lower bounds the conditional entropy $H_\alpha^\uparrow(Q|Q')_{\rho_{|\widehat{C}}}$ in a log-mean-exponential manner is equivalent to the non-negativity of f-weighted entropies, expressed as $H_\alpha^{\uparrow,f}(Q|\widehat{C}Q')_\rho \geq 0$.

We now show that these entropies possess similar properties to their counterpart in Definition 4.3.1.

Lemma 5.4.2. *Let f be a tradeoff function on some classical registers $\overline{C}\widehat{C}$, let $M \in \mathbb{R}$ be any real value such that $M - f(\bar{c}\hat{c}) > 0$ for all $\bar{c}\hat{c}$, and take any $\alpha \in (0,1)\cup(1,\infty]$. Consider any read-and-prepare channel $\mathcal{D} \in \mathrm{CPTP}(\overline{C}\widehat{C}, \overline{C}\widehat{C}D)$ such that the state it prepares on D always satisfies*

$$H_\alpha(D)_{\rho_{|\bar{c}\hat{c}}} = H_\alpha^\uparrow(D)_{\rho_{|\bar{c}\hat{c}}} = M - f(\bar{c}\hat{c}). \tag{5.59}$$

(It is always possible to construct such a channel, and construct it such that D is classical.) Then for any $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$ classical on $\overline{C}\widehat{C}$, extending ρ with this channel yields[10]

$$H_\alpha^\uparrow(DQ\overline{C}|\widehat{C}Q')_\rho = M + H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho. \tag{5.60}$$

[10]There is no danger of ambiguity in having used ρ to denote all states in this lemma, since a read-and-prepare channel always simply extends a state without "disturbing" any registers.

Proof. First we briefly verify that such channels can indeed be constructed: note that for any $\alpha > 0$ and $h > 0$, given a register D with dimension at least 2^h, one can straightforwardly construct a state on D with $H_\alpha(D) = H_\alpha^\uparrow(D) = h$ (the first equality holds simply because $H_\alpha = H_\alpha^\uparrow$ when there is no conditioning system), e.g. by taking a mixture between some pure state on D and the maximally mixed state; furthermore this can clearly be achieved using a classical D. Therefore, to obtain a read-and-prepare channel satisfying (5.59), we simply need it to read the classical value $\bar{c}\hat{c}$ and prepare a state on D with the corresponding desired entropy value.

Then Eq. (5.60) follows from the following calculation (assuming $\alpha \neq \infty$; the $\alpha = \infty$ case follows by taking the limit at the end, and the $\alpha \in [\frac{1}{2}, 1)$ cases are analogous with the infima replaced by suprema):

$$
\begin{aligned}
&H_\alpha^\uparrow(DQ\overline{C}|\widehat{C}Q')_\rho = \\
&\frac{\alpha}{1-\alpha}\log\left(\sum_{\hat{c}}\rho(\hat{c})\inf_{\sigma^{(\hat{c})}\in \mathrm{S}_=(Q')}\left[\sum_{\bar{c}}\rho(\bar{c}|\hat{c})^\alpha\, 2^{(\alpha-1)D_\alpha(\rho_{DQQ'|\bar{c}\hat{c}}\|\mathbb{I}_{DQ}\otimes\sigma^{(\hat{c})}_{Q'})}\right]^{\frac{1}{\alpha}}\right) \\
&= \frac{\alpha}{1-\alpha}\log\left(\sum_{\hat{c}}\rho(\hat{c})\inf_{\sigma^{(\hat{c})}\in \mathrm{S}_=(Q')}\left[\sum_{\bar{c}}\rho(\bar{c}|\hat{c})^\alpha\, 2^{(\alpha-1)\left(-H_\alpha^\uparrow(D)_{\rho|\bar{c}\hat{c}}+D_\alpha(\rho_{QQ'|\bar{c}\hat{c}}\|\mathbb{I}_Q\otimes\sigma^{(\hat{c})}_{Q'})\right)}\right]^{\frac{1}{\alpha}}\right) \\
&= \frac{\alpha}{1-\alpha}\log\left(\sum_{\hat{c}}\rho(\hat{c})\inf_{\sigma^{(\hat{c})}\in \mathrm{S}_=(Q')}\left[\sum_{\bar{c}}\rho(\bar{c}|\hat{c})^\alpha\, 2^{(\alpha-1)\left(f(\bar{c}\hat{c})-M+D_\alpha(\rho_{QQ'|\bar{c}\hat{c}}\|\mathbb{I}_Q\otimes\sigma^{(\hat{c})}_{Q'})\right)}\right]^{\frac{1}{\alpha}}\right) \\
&= M + \frac{\alpha}{1-\alpha}\log\left(\sum_{\hat{c}}\rho(\hat{c})\inf_{\sigma^{(\hat{c})}\in \mathrm{S}_=(Q')}\left[\sum_{\bar{c}}\rho(\bar{c}|\hat{c})^\alpha\, 2^{(\alpha-1)\left(f(\bar{c}\hat{c})+D_\alpha(\rho_{QQ'|\bar{c}\hat{c}}\|\mathbb{I}_Q\otimes\sigma^{(\hat{c})}_{Q'})\right)}\right]^{\frac{1}{\alpha}}\right) \\
&= M + H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho, \qquad (5.61)
\end{aligned}
$$

where the second line holds since $\rho_{DQQ'|\bar{c}\hat{c}} = \rho_{D|\bar{c}\hat{c}} \otimes \rho_{QQ'|\bar{c}\hat{c}}$, and the third line holds by substituting from Eq. (5.60). □

We now present a relaxed version of the previous lemma, which is the same as Lemma 4.3.2, but differs by using f-weighted entropies:

Lemma 5.4.3. *Let f be a tradeoff function on some classical registers $\overline{C}\widehat{C}$, let $M > 0$ be any value such that $M - f(\bar{c}\hat{c}) > M/2 > 0$ for all $\bar{c}\hat{c}$. Consider any read-and-prepare channel $\mathcal{D} \in \mathrm{CPTP}(\overline{C}\widehat{C}, \overline{C}\widehat{C}D)$ such that the state it prepares on D always satisfies*

$$
\forall \alpha \in [0,\infty], \quad H_\alpha(D)_{\rho_{|\bar{c}\hat{c}}} = H_\alpha^\uparrow(D)_{\rho_{|\bar{c}\hat{c}}} \in \left[M - f(\bar{c}\hat{c}), M - f(\bar{c}\hat{c}) + 2^{-\frac{M}{2}}\log e\right]. \qquad (5.62)
$$

(It is always possible to construct such a channel, and construct it such that D is classical.) Then for any $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$ classical on $\overline{C}\widehat{C}$, extending ρ with this channel yields

$$M + H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho \leq H_\alpha^{\uparrow}(DQ\overline{C}|\widehat{C}Q')_\rho \leq M + 2^{-\frac{M}{2}} \log e + H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho, \tag{5.63}$$

for all $\alpha \in (0,1) \cup (1,\infty]$.

Proof. The proof is by the same argument as above, except we upper and lower bound the entropies on the D register with (5.62) instead of (5.59). (The existence of such a read-and-prepare channel is established by the same argument as in the proof of Lemma 4.3.2) □

Utilizing these lemmas, we transfer many properties of $H_\alpha^\uparrow$ to $H_\alpha^{\uparrow,f}$, in very similar fashion to the results in Chapter 4. The proof of these lemmas follows exactly the same methodology as their counterparts in Chapter 4.

Lemma 5.4.4. *(Data-processing) Let $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$ be classical on $\overline{C}\widehat{C}$, let f be a tradeoff function on $\overline{C}\widehat{C}$, and take any $\alpha \in [\frac{1}{2},1) \cup (1,\infty]$. Then for any channel $\mathcal{E} \in \mathrm{CPTP}(Q',Q'')$,*

$$H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q'')_{\mathcal{E}[\rho]} \geq H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho. \tag{5.64}$$

If $\mathcal{E}$ is an isometry, then we have equality in the above bound.

Lemma 5.4.5. *(Conditioning on classical registers) Let $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ'Z)$ be classical on $\overline{C}\widehat{C}Z$, let f be a tradeoff function on $\overline{C}\widehat{C}$, and take any $\alpha \in [\frac{1}{2},1) \cup (1,\infty]$. Then*

$$H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q'Z)_\rho = \frac{\alpha}{1-\alpha} \log \left(\sum_z \rho(z)\, 2^{\left(\frac{1-\alpha}{\alpha}\right) H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_{\rho_{|z}}} \right), \tag{5.65}$$

Lemma 5.4.6. *(Monotonicity in α) Let $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$ be classical on $\overline{C}\widehat{C}$, let f be a tradeoff function on $\overline{C}\widehat{C}$, and take any $\alpha, \alpha' \in [\frac{1}{2},1) \cup (1,\infty]$ such that $\alpha \geq \alpha'$. Then,*

$$H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho \leq H_{\alpha'}^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho. \tag{5.66}$$

We now provide the following lemma, which connects $H_\alpha^{\uparrow,f}$ to H_α^f.

Lemma 5.4.7. *(Relating $H_\alpha^{\uparrow,f}$ and H_α^f) Let $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$ be classical on $\overline{C}\widehat{C}$, let f be a tradeoff function on $\overline{C}\widehat{C}$, and take any $\alpha \in [\frac{1}{2},1) \cup (1,\infty]$. Then,*

$$H_{\widehat{\alpha}}^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho \leq H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho \leq H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho, \tag{5.67}$$

where $\widehat{\alpha} = \frac{1}{2-\alpha}$, and H_α^f is defined in Definition 4.3.1.

Proof. We show only the proof of the second inequality, as the proof of the first is analogous but simpler (since the Rényi parameter remains unchanged). Let $M > 0$ be any value such that $M - f(\bar{c}\hat{c}) > M/2 > 0$, and extend ρ with a read-and-prepare channel as described in Lemma 5.4.3; note that this is the same as the channel in Lemma 4.3.2. Then we have

$$\begin{aligned} H_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho &\geq H_\alpha(DQ\overline{C}|\widehat{C}Q') - M - 2^{-M/2}\log e \\ &\geq H_{\widehat{\alpha}}^\uparrow(DQ\overline{C}|\widehat{C}Q') - M - 2^{-M/2}\log e \\ &\geq H_{\widehat{\alpha}}^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho - 2^{-M/2}\log e, \end{aligned} \tag{5.68}$$

where the first line is Lemma 4.3.2, the second line follows from [118, Corollary 4], and the third line is Lemma 5.4.3. Since this inequality holds for arbitrary (sufficiently large) M, we can take the $M \to \infty$ limit to get the claimed inequality. □

Lemma 5.4.8. *(Chain rules) Let f be a tradeoff function on classical registers $\overline{C}\widehat{C}$, take any $\alpha, \alpha', \alpha'' \in (\frac{1}{2}, 1) \cup (1, \infty)$ such that $\frac{\alpha}{\alpha-1} = \frac{\alpha'}{\alpha'-1} + \frac{\alpha''}{\alpha''-1}$. Then we have for any $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}RQQ')$:*

$$\begin{aligned} H_\alpha^{\uparrow,f}(RQ\overline{C}|\widehat{C}Q')_\rho &\leq H_{\alpha'}^{\uparrow,f}(Q\overline{C}|\widehat{C}Q'R)_\rho + H_{\alpha''}^\uparrow(R|\widehat{C}Q')_\rho \quad \text{if } (\alpha-1)(\alpha'-1)(\alpha''-1) < 0, \\ H_\alpha^{\uparrow,f}(RQ\overline{C}|\widehat{C}Q')_\rho &\geq H_{\alpha'}^{\uparrow,f}(Q\overline{C}|\widehat{C}Q'R)_\rho + H_{\alpha''}^\uparrow(R|\widehat{C}Q')_\rho \quad \text{if } (\alpha-1)(\alpha'-1)(\alpha''-1) > 0. \end{aligned} \tag{5.69}$$

Furthermore, we also have for any $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$:

$$\begin{aligned} H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho &\leq H_{\alpha'}^{\uparrow,f}(Q|\overline{C}\widehat{C}Q')_\rho + H_{\alpha''}^\uparrow(\overline{C}|\widehat{C}Q')_\rho \quad \text{if } (\alpha-1)(\alpha'-1)(\alpha''-1) < 0, \\ H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho &\geq H_{\alpha'}^{\uparrow,f}(Q|\overline{C}\widehat{C}Q')_\rho + H_{\alpha''}^\uparrow(\overline{C}|\widehat{C}Q')_\rho \quad \text{if } (\alpha-1)(\alpha'-1)(\alpha''-1) > 0, \end{aligned} \tag{5.70}$$

where by $H_\alpha^{\uparrow,f}(Q|\overline{C}\widehat{C}Q')_\rho$ we mean

$$H_\alpha^{\uparrow,f}(Q|\overline{C}\widehat{C}Q')_\rho = \frac{\alpha}{1-\alpha}\log\left(\sum_{\bar{c}\hat{c}} \rho(\bar{c}\hat{c}) 2^{\left(\frac{1-\alpha}{\alpha}\right)\left(-f(\bar{c}\hat{c}) + H_\alpha^\uparrow(Q|Q')_{\rho_{|\bar{c}\hat{c}}}\right)}\right). \tag{5.71}$$

Lemma 5.4.9. *(Normalization property) Let $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$ be classical on $\overline{C}\widehat{C}$ and let f be a tradeoff function on $\overline{C}\widehat{C}$. Then for any constant $\kappa \in \mathbb{R}$, we have*

$$H_\alpha^{\uparrow,f+\kappa}(Q\overline{C}|\widehat{C}Q')_\rho = H_\alpha^{\uparrow,f}(Q\overline{C}|\widehat{C}Q')_\rho - \kappa, \tag{5.72}$$

where $f + \kappa$ simply denotes the tradeoff function with values $f(\bar{c}\hat{c}) + \kappa$.

5.4.2 MEAT (simplified)

In this part, to build intuition, we first present a simplified version of MEAT and defer the general form and its proofs to the next section.

We begin by demonstrating that if the global state is generated through the application of a suitable sequence of channels (typically representing rounds of a protocol) to an arbitrary input state, then tradeoff functions can accumulate to estimate the conditional entropy of the global state.

Theorem 5.4.1. *For each $j \in \{1, 2, \cdots, n\}$, take a state $\sigma^{(j-1)} \in \mathrm{S}_=(A_{j-1})$, and a channel $\mathcal{M}_j \in \mathrm{CPTP}(A_{j-1}E_{j-1}, S_j\widehat{C}_jE_j)$, such that $\widehat{C}_j$ are classical. Let ρ be a state of the form $\rho_{S_1^n\widehat{C}_1^nE_n} = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1[\omega_{A_0^{n-1}E_0}]$ for some $\omega \in \mathrm{S}_=(A_0^{n-1}E_0)$, such that $\omega_{A_0^{n-1}} = \sigma_{A_0}^{(0)} \otimes \cdots \otimes \sigma_{A_{n-1}}^{(n-1)}$. For each j, suppose that for every value $\hat{c}_1^{j-1}$, we have a tradeoff function $f_{|\hat{c}_1^{j-1}}$ on registers $\widehat{C}_j$. Define the following tradeoff function on $\widehat{C}_1^n$:*

$$f_{\text{full}}(\hat{c}_1^n) \coloneqq \sum_{j=1}^{n} f_{|\hat{c}_1^{j-1}}(\hat{c}_j). \tag{5.73}$$

Then for any $\alpha \in (1, \infty]$ we have

$$H_\alpha^{\uparrow, f_{\text{full}}}(S_1^n|\widehat{C}_1^nE_n)_\rho \geq \sum_j \min_{\hat{c}_1^{j-1}} \kappa_{\hat{c}_1^{j-1}} \qquad \text{where} \quad \kappa_{\hat{c}_1^{j-1}} \coloneqq \inf_{\nu \in \Sigma_j} H_\alpha^{\uparrow, f_{|\hat{c}_1^{j-1}}}(S_j|\widehat{C}_jE_j\widetilde{E})_\nu, \tag{5.74}$$

defining

$$\Sigma_j \coloneqq \left\{ \mathcal{M}_j\left[\omega_{A_{j-1}E_{j-1}\widetilde{E}}\right] \;\middle|\; \omega \in \mathrm{S}_=(A_{j-1}E_{j-1}\widetilde{E}) \;\; s.t.\; \omega_{A_{j-1}} = \sigma_{A_{j-1}}^{(j-1)} \right\}, \tag{5.75}$$

with $\widetilde{E}$ being a register of large enough dimension to serve as a purifying register for any of the $A_{j-1}E_{j-1}$ registers.

Consequently, if we instead define the following "normalized" tradeoff function on $\widehat{C}_1^n$:

$$\hat{f}_{\text{full}}(\hat{c}_1^n) \coloneqq \sum_{j=1}^{n} \hat{f}_{|\hat{c}_1^{j-1}}(\hat{c}_j), \quad \text{where} \quad \hat{f}_{|\hat{c}_1^{j-1}}(\hat{c}_j) \coloneqq f_{|\hat{c}_1^{j-1}}(\hat{c}_j) + \kappa_{\hat{c}_1^{j-1}}, \tag{5.76}$$

then

$$H_\alpha^{\uparrow, \hat{f}_{\text{full}}}(S_1^n|\widehat{C}_1^nE_n)_\rho \geq 0. \tag{5.77}$$

Essentially, the above theorem tells us that for any choice of single-round trade-off functions $f_{|\hat{c}_1^{j-1}}$, their accumulation yields a global trade-off function $\hat{f}_{\text{full}}$ that estimates the Rényi conditional entropy of the global state in a log-mean-exponential sense. Furthermore, this accumulation preserves marginal-state constraints on the input state, making it suitable for applying the source-replacement technique in PM protocols.

We highlight that the optimization in the definition of $\kappa_{\hat{c}_1^{j-1}}$ above is generally not a convex optimization. In the following lemma, we show that for each round, this optimization can be written as an equivalent convex optimization problem.

Lemma 5.4.10. *Let $\mathcal{M} \in \mathrm{CPTP}(Q, SE\overline{C}\widehat{C})$ where the output is always classical on $\overline{C}\widehat{C}$, and let f be a tradeoff function on $\overline{C}\widehat{C}$. Let $\widetilde{E}$ be a register such that $\dim(Q) \leq \dim(\widetilde{E})$, and let $\mathtt{Pur}$ be a purifying function for Q onto $\widetilde{E}$ (Definition 4.2.2). Then for any state $\omega_{\widetilde{Q}\widetilde{E}}$ and $\alpha \in [\frac{1}{2}, 1) \cup (1, \infty]$,*

$$H_\alpha^{\uparrow,f}(S\overline{C}|\widehat{C}E\widetilde{E})_{\mathcal{M}[\omega_{\widetilde{Q}\widetilde{E}}]} \geq H_\alpha^{\uparrow,f}(S\overline{C}|\widehat{C}E\widetilde{E})_{\mathcal{M}[\mathtt{Pur}(\omega_{\widetilde{Q}})]}, \tag{5.78}$$

with equality if $\omega_{\widetilde{Q}\widetilde{E}}$ is pure, and furthermore for $\alpha \in [1, \infty]$, the right-hand-side is a convex function of $\omega_{\widetilde{Q}}$.

Proof. The proof follows from the same steps as in Lemma 4.3.7. □

In many scenarios, we may be interested only in the entropy conditioned on an event (e.g., in QKD protocols that make an accept/abort decision and produce a fixed-length key upon acceptance). In what follows, similar to Chapter 4, we provide a theorem that bounds the Rényi conditional entropy conditioned on such an event. To that end, similar to the notion of convex range in Definition 4.4.1, we introduce the "marginal-constraint convex range", which is essentially the convex range, with additional marginal constraints.

Definition 5.4.2. (Marginal-constrained convex range) For $j \in \{1, 2, \cdots, n\}$, consider channels $\mathcal{E}_j \in \mathrm{CPTP}(A_{j-1}\widetilde{A_{j-1}}, Z_j)$ and states $\sigma^{(j-1)} \in \mathrm{S}_=(A_{j-1})$. Suppose furthermore that $Z_j = Z_j' Z_j'' \ldots$ for some (finite) set of "sub-registers" $Z_j' Z_j'' \ldots$, and we have an embedding of all the output registers Z_j' (resp. $Z_j'', \ldots$) in a common register Z' (resp. $Z'', \ldots$), in which case there is a well-defined notion of taking mixtures of states with distinct Z_j (by embedding them in $Z \coloneqq Z_j' Z_j'' \ldots$). Then the **marginal-constrained convex range** Σ_Z of[11] the channels $\mathcal{E}_j$ and states $\sigma^{(j-1)}$ is defined as the set of all mixtures of states of

[11] Strictly speaking, the marginal-constrained convex range is not fully defined by only the channels $\mathcal{E}_j$ and states $\sigma^{(j-1)}$, since one also has to specify the embeddings. However, all of our subsequent results hold for any choice of embeddings, so we will usually not specify this explicitly.

the form $\mathcal{E}_j\left[\omega_{A_{j-1}\widetilde{A}_{j-1}}\right]$ for some j and some $\omega \in \mathrm{S}_=(A_{j-1}\widetilde{A_{j-1}})$ satisfying $\omega_{A_{j-1}} = \sigma^{(j-1)}_{A_{j-1}}$. Slightly more formally, we can write

$$\Sigma_Z \coloneqq \mathrm{conv}\left(\bigcup_{j=1}^{n}\left\{\mathcal{E}_j\left[\omega_{A_{j-1}\widetilde{A}_{j-1}}\right] \;\middle|\; \omega \in \mathrm{S}_=(A_{j-1}\widetilde{A_{j-1}}) \;\text{ s.t. } \omega_{A_{j-1}} = \sigma^{(j-1)}_{A_{j-1}}\right\}\right), \tag{5.79}$$

where $\mathrm{conv}(\mathcal{S})$ denotes the convex hull [18] of a set $\mathcal{S}$.

Note that if all the channels $\mathcal{E}j \in \mathrm{CPTP}(A_{j-1}\widetilde{A}_{j-1}, Z_j)$ and states $\sigma^{(j-1)} \in \mathrm{S}_=(A_{j-1})$ were isomorphic to a single channel $\mathcal{E} \in \mathrm{CPTP}(A\widetilde{A}, Z)$ and a single state $\sigma \in \mathrm{S}_=(A)$, then the marginal-constrained convex range would form a convex set, and could thus be defined as:

$$\Sigma_Z = \left\{\mathcal{E}\left[\omega_{A\widetilde{A}}\right] \;\middle|\; \omega \in \mathrm{S}_=(A\widetilde{A}) \;\text{ s.t. } \omega_A = \sigma_A\right\}, \tag{5.80}$$

which is exactly the structure assumed in [126]. With this notion, we obtain the following bound on the Rényi conditional entropy, conditioned on an event:

Theorem 5.4.2. *(Marginal-constrained entropy accumulation theorem, special case) For each $j \in \{1, 2, \cdots, n\}$, take a state $\sigma^{(j-1)} \in \mathrm{S}_=(A_{j-1})$, and a channel $\mathcal{M}_j \in \mathrm{CPTP}(A_{j-1}E_{j-1}, S_j\widehat{C}_jE_j)$, such that $\widehat{C}_j$ are classical. Let ρ be a state of the form $\rho_{S_1^n\widehat{C}_1^nE_n} = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1[\omega_{A_0^{n-1}E_0}]$ for some $\omega \in \mathrm{S}_=(A_0^{n-1}E_0)$, such that $\omega_{A_0^{n-1}} = \sigma^{(0)}_{A_0} \otimes \cdots \otimes \sigma^{(n-1)}_{A_{n-1}}$.*

Suppose furthermore that $\rho = p_\Omega\rho_{|\Omega} + (1-p_\Omega)\rho_{|\overline{\Omega}}$ for some $p_\Omega \in (0, 1]$ and normalized states $\rho_{|\Omega}, \rho_{|\overline{\Omega}}$, and that all the $\widehat{C}_j$ registers are isomorphic to a single register $\widehat{C}$ with alphabet $\widehat{\mathcal{C}}$. Let S_Ω be a convex set of probability distributions on the alphabet $\widehat{\mathcal{C}}$, such that for all $\hat{c}_1^n$ with nonzero probability in $\rho_{|\Omega}$, the frequency distribution $\mathrm{freq}_{\hat{c}_1^n}$ lies in S_Ω. Then, for any $\alpha \in (1, \infty]$, we have:

$$H^\uparrow_\alpha(S_1^n|\widehat{C}_1^nE_n)_{\rho_{|\Omega}} \geq nh^\uparrow_\alpha - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega}, \tag{5.81}$$

where (recalling $\boldsymbol{\nu}_{\widehat{C}}$ denotes the distribution on $\widehat{C}$ induced by any state $\nu_{\widehat{C}}$)

$$h^\uparrow_\alpha = \inf_{\mathbf{q}\in S_\Omega}\inf_{\nu\in\Sigma_{S\widehat{C}E\widetilde{E}}}\left(\frac{\alpha}{\alpha-1}D\left(\mathbf{q}\middle\|\boldsymbol{\nu}_{\widehat{C}}\right) + \sum_{\hat{c}\in\mathrm{supp}(\boldsymbol{\nu}_{\widehat{C}})} q(\hat{c})H^\uparrow_\alpha(S|E\widetilde{E})_{\nu_{|\hat{c}}}\right), \tag{5.82}$$

with $\Sigma_{S\widehat{C}E\widetilde{E}}$ being the marginal-constrained convex range (Definition 5.4.2) of the channels $\mathcal{M}_j \otimes \mathrm{id}_{\widetilde{E}}$ and states $\sigma^{(j-1)}$, where $\widetilde{E}$ is a register of large enough dimension to serve as a purifying register for any of the $A_{j-1}E_{j-1}$ registers.

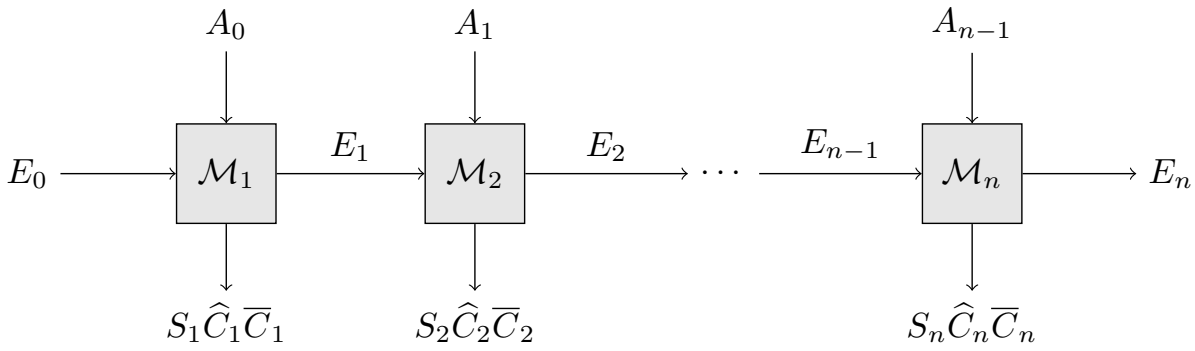


Figure 5.1: The setting for the marginal-constrained entropy accumulation theorem.

We note that the optimization in (5.82), as written, is not convex. However, we will show later that it can be converted into a convex optimization problem.

Theorem 5.4.1 allows for choosing different tradeoff functions in different rounds, where each choice can depend on the classical public values $\hat{c}_1^{j-1}$ from preceding rounds. This enables a fully adaptive procedure in which the choice of tradeoff functions can be updated over the course of the protocol based on the history of public announcements.[12]

5.4.3 MEAT (general statement)

In this part, we state and prove the most general form of f-weighted entropy accumulation, allowing the channels to generate an additional "secret" classical register $\overline{C}_j$ in each round.

Theorem 5.4.3. *For each $j \in \{1, 2, \cdots, n\}$, take a state $\sigma^{(j-1)} \in \mathrm{S}_=(A_{j-1})$, and a channel $\mathcal{M}_j \in \mathrm{CPTP}(A_{j-1}E_{j-1}, S_j\overline{C}_j\widehat{C}_jE_j)$, such that $\overline{C}_j\widehat{C}_j$ are classical. Let ρ be a state of the form $\rho_{S_1^n\overline{C}_1^n\widehat{C}_1^nE_n} = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1[\omega_{A_0^{n-1}E_0}]$ for some $\omega \in \mathrm{S}_=(A_0^{n-1}E_0)$, such that $\omega_{A_0^{n-1}} = \sigma_{A_0}^{(0)} \otimes \cdots \otimes \sigma_{A_{n-1}}^{(n-1)}$ (see Fig. 5.1. For each j, suppose that for every value $\hat{c}_1^{j-1}$, we have a tradeoff function $f_{|\hat{c}_1^{j-1}}$ on registers $\overline{C}_j\widehat{C}_j$. Define the following tradeoff function on $\overline{C}_1^n\widehat{C}_1^n$:*

$$f_{\mathrm{full}}(\bar{c}_1^n\hat{c}_1^n) := \sum_{j=1}^{n} f_{|\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j). \tag{5.83}$$

[12] This is similar to the procedure in 4.3.3.

Then for any $\alpha \in (1,\infty]$ we have

$$H_\alpha^{\uparrow,f_{\text{full}}}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_\rho \geq \sum_j \min_{\hat{c}_1^{j-1}} \kappa_{\hat{c}_1^{j-1}} \qquad \text{where} \quad \kappa_{\hat{c}_1^{j-1}} \coloneqq \inf_{\nu\in\Sigma_j} H_\alpha^{\uparrow,f_{|\hat{c}_1^{j-1}}}(S_j\overline{C}_j|\widehat{C}_j E_j\widetilde{E})_\nu, \tag{5.84}$$

defining

$$\Sigma_j \coloneqq \left\{ \mathcal{M}_j\left[\omega_{A_{j-1}E_{j-1}\widetilde{E}}\right] \;\middle|\; \omega \in \mathrm{S}_=(A_{j-1}E_{j-1}\widetilde{E}) \;\; s.t.\; \omega_{A_{j-1}} = \sigma_{A_{j-1}}^{(j-1)} \right\}, \tag{5.85}$$

with $\widetilde{E}$ being a register of large enough dimension to serve as a purifying register for any of the $A_{j-1}E_{j-1}$ registers.

Consequently, if we instead define the following "normalized" tradeoff function on $\overline{C}_1^n\widehat{C}_1^n$:

$$\hat{f}_{\text{full}}(\bar{c}_1^n\hat{c}_1^n) \coloneqq \sum_{j=1}^n \hat{f}_{|\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j), \quad \text{where} \quad \hat{f}_{|\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j) \coloneqq f_{|\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j) + \kappa_{\hat{c}_1^{j-1}}, \tag{5.86}$$

then

$$H_\alpha^{\uparrow,\hat{f}_{\text{full}}}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_\rho \geq 0. \tag{5.87}$$

Before proving the theorem, we note a subtle difference between this and Theorem 4.3.1. As reflected in the notation, the choice of trade-off function here can depend only on the public classical values $\hat{c}_1^{j-1}$. This restriction is because challenges arise in the proof when attempting to include dependency on $\bar{c}_1^{j-1}$, due to the absence of a "secret" memory register, unlike in the GEAT model.

Proof. The proof is nearly identical to that of Theorem 4.3.1, except we invoke Theorem 5.3.1 instead of the chain rule used in that proof. First, let $M > 0$ be any value such that $M - f_{|\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j) > 0$ for all the $f_{|\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j)$ in the theorem statement (for all j). Now for each j, define a read-and-prepare channel $\mathcal{D}_j \in \mathrm{CPTP}(\overline{C}_j\widehat{C}_1^j, \overline{C}_j\widehat{C}_1^j D_j)$ of the form described in Lemma 5.4.2, so that the state it prepares on D_j satisfies

$$H_\alpha(D)_{\rho_{|\hat{c}_1^j\bar{c}_j}} = H_\alpha^\uparrow(D)_{\rho_{|\hat{c}_1^j\bar{c}_j}} = M - f_{|\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j). \tag{5.88}$$

Now let $\mathcal{N}_j \in \mathrm{CPTP}(A_{j-1}E_{j-1}\widehat{C}_1^{j-1}, D_j S_j\overline{C}_j E_j\widehat{C}_1^j)$ denote a channel that does the following

1. Apply $\mathcal{M}_j \otimes \mathcal{P}_j$, where $\mathcal{P}_j$ is a pinching channel on $\widehat{C}_1^{j-1}$ in its classical basis, hence making the $\widehat{C}_1^{j-1}$ registers classical.

2. Generate a D_j register by applying the above read-and-prepare channel $\mathcal{D}_j$ on $\overline{C}_j \widehat{C}_1^j$.

Note that here, in each round j the tradeoff function only depends on previous public announcements $\hat{c}_1^{j-1}$, as opposed to $\bar{c}_1^{j-1}\hat{c}_1^{j-1}$ in Theorem 4.3.1.[13]

Observing that $\mathcal{N}_n \circ \cdots \circ \mathcal{N}_1[\omega]$ is a valid extension of the state ρ in the theorem, we can use ρ to refer to this state as well without any ambiguity. We now apply Theorem 5.3.1 iteratively to this state. Specifically, by identifying the channels and registers in that theorem with those in this scenario (for each j) as:

- $\mathcal{E}_1 \in \mathrm{CPTP}(A_0Y_0, X_1Y_1) \;\leftrightarrow\; \mathcal{N}_{j-1}\circ\cdots\circ\mathcal{N}_1 \in \mathrm{CPTP}(A_1^{j-2}E_0, D_1^{j-1}S_1^{j-1}\overline{C}_1^{j-1}E_{j-1}\widehat{C}_1^{j-1})$
- $A_0 \leftrightarrow A_0^{j-2}$
- $Y_0 \leftrightarrow E_0$
- $X_1 \leftrightarrow D_1^{j-1}S_1^{j-1}\overline{C}_1^{j-1}$,
- $Y_1 \leftrightarrow E_{j-1}\widehat{C}_1^{j-1}$
- $\mathcal{E}_2 \in \mathrm{CPTP}(A_1Y_1, X_2Y_2) \;\leftrightarrow\; \mathcal{N}_j \in \mathrm{CPTP}(A_{j-1}E_{j-1}\widehat{C}_1^{j-1}, D_jS_j\overline{C}_jE_j\widehat{C}_1^j)$
- $A_1 \leftrightarrow A_{j-1}$
- $X_2 \leftrightarrow D_jS_j\overline{C}_j$
- $Y_2 \leftrightarrow E_j\widehat{C}_1^j$,

[13] One would need to keep a copy of the registers $\overline{C}_1^{j-1}$ in the output registers of $\mathcal{N}_j$ for the tradeoff function to be able to depend on $\bar{c}_1^{j-1}$ as well. This was handled in the analysis done in Chapter 4 by keeping that copy in a "secret" memory register; however, that was under the framework of the GEAT [89], which allowed having a "secret" memory register such that there is no signaling from it to the adversary's register (see the discussion in Sec. 5.6 later). In order to accommodate a memory register in this framework, one would need to construct a chain rule similar to the one in Theorem 5.3.1, but generalized to allow handling a memory register under a non-signaling condition to the adversary. Such a chain rule seems highly non-trivial to obtain, and we leave the analysis to future work.

we have:

$$\begin{aligned}
&H_\alpha^\uparrow(D_1^n S_1^n \overline{C}_1^n|\widehat{C}_1^n E_n)_\rho \\
&\geq H_\alpha^\uparrow\left(\mathcal{N}_n \circ \cdots \circ \mathcal{N}_1, D_1^n S_1^n \overline{C}_1^n, [\omega_{A_0^{n-1}}]\right) \\
&\geq H_\alpha^\uparrow\left(\mathcal{N}_n, D_n S_n \overline{C}_n, [\omega_{A_{n-1}}]\right) + H_\alpha^\uparrow\left(\mathcal{N}_{n-1} \circ \cdots \circ \mathcal{N}_1, D_1^{n-1} S_1^{n-1} \overline{C}_1^{n-1}, [\omega_{A_0^{n-2}}]\right) \\
&\quad \vdots \\
&\geq \sum_j H_\alpha^\uparrow\left(\mathcal{N}_j, D_j S_j \overline{C}_j, [\omega_{A_{j-1}}]\right) \\
&= \sum_j \inf_{\nu' \in \Sigma_j'} H_\alpha^\uparrow(D_j S_j \overline{C}_j|\widehat{C}_1^j E_j \widetilde{E})_{\nu'},
\end{aligned} \tag{5.89}$$

where Σ_j' denotes the set of all states that could be produced by $\mathcal{N}_j$ acting on some initial state $\omega' \in \mathrm{S}_=(A_{j-1}E_{j-1}\overline{C}_{j-1}\widehat{C}_1^{j-1}\widetilde{E})$[14], such that $\omega'_{A_{j-1}} = \sigma_{A_{j-1}}^{(j-1)}$.

The remainder of the proof is identical to the proof of Theorem 4.3.1, i.e., by applying calculations similar to Eq. (5.61) above, we show that

$$H_\alpha^\uparrow(D_1^n S_1^n \overline{C}_1^n|\widehat{C}_1^n E_n)_\rho = nM + H_\alpha^{\uparrow, f_{\text{full}}}(S_1^n \overline{C}_1^n|\widehat{C}_1^n E_n)_\rho, \tag{5.90}$$

$$\inf_{\nu' \in \Sigma_j'} H_\alpha^\uparrow(D_j S_j \overline{C}_j|\widehat{C}_1^j E_j \widetilde{E})_{\nu'} = M + \min_{\hat{c}_1^{j-1}} \inf_{\nu \in \Sigma_j} H_\alpha^{\uparrow, f_{|\hat{c}_1^{j-1}}}(S_j \overline{C}_j|\widehat{C}_j E_j \widetilde{E})_\nu, \tag{5.91}$$

then substitute these relations into Eq. (5.89), which yields the claimed result by canceling nM from both sides. □

We now proceed to prove Theorem 5.4.2. To begin, we first state the following duality result.

Lemma 5.4.11. *Consider any $\alpha \in (1, \infty)$ and any $\rho \in \mathrm{S}_=(QQ'\widehat{C})$ such that $\widehat{C}$ is classical with alphabet $\widehat{\mathcal{C}}$. Then letting $\mathbb{P}_{\widehat{\mathcal{C}}}$ denote the set of probability distributions on $\widehat{\mathcal{C}}$, the pair of functions $\left(G_{\alpha,\rho}, G_{\alpha,\rho}^*\right)$ on $\mathbb{R}^{|\widehat{\mathcal{C}}|}$ defined as*

$$\begin{aligned}
G_{\alpha,\rho}(\mathbf{f}) &:= -H_\alpha^{\uparrow, f}(Q|\widehat{C}Q')_\rho, \\
G_{\alpha,\rho}^*(\boldsymbol{\lambda}) &:= \begin{cases} \frac{\alpha}{\alpha-1} D\left(\boldsymbol{\lambda} \middle\| \boldsymbol{\rho}_{\widehat{C}}\right) + \sum_{\hat{c} \in \mathrm{supp}(\boldsymbol{\rho}_{\widehat{C}})} \lambda(\hat{c}) H_\alpha^\uparrow(Q|Q')_{\rho_{|\hat{c}}} & \boldsymbol{\lambda} \in \mathbb{P}_{\widehat{\mathcal{C}}}, \\ +\infty & \text{else}, \end{cases}
\end{aligned} \tag{5.92}$$

[14] In this step, let us take $\widetilde{E}$ to be of large enough dimension to be a purifying register for the input registers in the $\mathcal{N}_j$ scenario as well; this can be achieved without loss of generality by expanding its dimension as necessary.

are convex functions that are convex conjugates to each other, where in the $G_{\alpha,\rho}$ formula we recall that any tuple $\mathbf{f} \in \mathbb{R}^{|\widehat{\mathcal{C}}|}$ is equivalent to a tradeoff function on $\widehat{C}$, and in the $G^*_{\alpha,\rho}$ formula we recall that $\boldsymbol{\rho}_{\widehat{C}}$ denotes the distribution induced by ρ on $\widehat{C}$.

Proof. First, we establish the convexity of $G_{\alpha,\rho}(\mathbf{f})$. To achieve this, let $\tilde{\mathbf{f}}, \widetilde{\boldsymbol{\lambda}}, \widetilde{\boldsymbol{\rho}}_{\widehat{C}}$ denote the restrictions of $\mathbf{f}, \boldsymbol{\lambda}, \boldsymbol{\rho}_{\widehat{C}}$, respectively, to the support of $\boldsymbol{\rho}_{\widehat{C}}$, i.e., $\text{supp}(\boldsymbol{\rho}_{\widehat{C}})$. Furthermore, let $\widetilde{\mathbf{H}}_\alpha$ denote the tuple of values $\{H^{\uparrow}_\alpha(Q|Q')_{\rho_{|\hat{c}}}\}_{\hat{c}\in\text{supp}(\boldsymbol{\rho}_{\widehat{C}})}$. Similar to the proof of Lemma 4.4.4 we exploit the notion of log-sum-exponential function:

$$\begin{aligned} G_{\alpha,\rho}(\mathbf{f}) &= \frac{\alpha}{(\alpha-1)\ln 2} \ln\left(\sum_{\hat{c}\in\text{supp}(\boldsymbol{\rho}_{\widehat{C}})} \rho(\hat{c}) e^{\left(\frac{\alpha-1}{\alpha}\right)(\ln 2)\left(f(\hat{c})-H^{\uparrow}_\alpha(Q|Q')_{\rho_{|\hat{c}}}\right)}\right) \\ &= \frac{\alpha}{(\alpha-1)\ln 2} \text{logsumexp}\left(\ln\left(\widetilde{\boldsymbol{\rho}}_{\widehat{C}}\right) + \left(\frac{\alpha-1}{\alpha}\right)(\ln 2)\left(\tilde{\mathbf{f}} - \widetilde{\mathbf{H}}_\alpha\right)\right), \end{aligned} \tag{5.93}$$

where $\ln\left(\widetilde{\boldsymbol{\rho}}_{\widehat{C}}\right)$ is the elementwise logarithm of $\widetilde{\boldsymbol{\rho}}_{\widehat{C}}$. Convexity follows by noting that $\alpha > 1$, the log-sum-exponential function is convex [18], and $G_{\alpha,\rho}(\mathbf{f})$ is constant with respect to any $\mathbf{f}$ components outside of $\tilde{\mathbf{f}}$.

Since $G_{\alpha,\rho}$ is convex and finite over the entire set $\mathbb{R}^{|\widehat{\mathcal{C}}|}$, taking the convex conjugate twice returns the original function [18], so to prove the claimed result, it suffices to show its convex conjugate has the formula in (5.92). The convex conjugate of $G_{\alpha,\rho}(\mathbf{f})$ is, by Definition 2.3.5,

$$G^*_{\alpha,\rho}(\boldsymbol{\lambda}) = \sup_{\mathbf{f}\in\mathbb{R}^{|\widehat{\mathcal{C}}|}} \left(\boldsymbol{\lambda}\cdot\mathbf{f} - G_{\alpha,\rho}(\mathbf{f})\right). \tag{5.94}$$

Note that if $\boldsymbol{\lambda}$ is such that $\text{supp}(\boldsymbol{\lambda}) \nsubseteq \text{supp}(\boldsymbol{\rho}_{\widehat{C}})$, the optimization is $+\infty$. This is because for those values of $\hat{c}$ such that $\lambda(\hat{c}) \neq 0$ and $\rho(\hat{c}) = 0$, $G_{\alpha,\rho}(\mathbf{f})$ becomes independent of the corresponding $f(\hat{c})$, and therefore, depending on the sign of $\lambda(\hat{c})$ we either set $f(\hat{c}) \to \infty$ or $f(\hat{c}) \to -\infty$ to make the objective diverge to $+\infty$. This behavior matches the formula given in Eq. (5.92), since for those $\boldsymbol{\lambda}$ that $\text{supp}(\boldsymbol{\lambda}) \nsubseteq \text{supp}(\boldsymbol{\rho}_{\widehat{C}})$, either $\boldsymbol{\lambda} \notin \mathbb{P}_{\widehat{C}}$, which the optimization results $+\infty$, which is reflected in the formula in (5.92), or $\boldsymbol{\lambda} \in \mathbb{P}_{\widehat{C}}$, in which case $D(\boldsymbol{\lambda}\|\boldsymbol{\rho}_{\widehat{C}})$ becomes $+\infty$.

Thus, we only focus on the case where $\text{supp}(\boldsymbol{\lambda}) \subseteq \text{supp}(\boldsymbol{\rho}_{\widehat{C}})$. Hence, for such $\boldsymbol{\lambda}$, the

optimization for $G^*_{\alpha,\rho}$ reduces to

$$\begin{aligned} G^*_{\alpha,\rho}(\boldsymbol{\lambda}) &= \sup_{\mathbf{f}\in\mathbb{R}^{|\widehat{C}|}} \left(\widetilde{\boldsymbol{\lambda}}\cdot\tilde{\mathbf{f}} - G_{\alpha,\rho}(\tilde{\mathbf{f}})\right) \\ &= \sup_{\mathbf{g}\in\mathbb{R}^{\left|\mathrm{supp}(\boldsymbol{\rho}_{\widehat{C}})\right|}} \left(\widetilde{\boldsymbol{\lambda}}\cdot\mathbf{g} - G_{\alpha,\rho}(\mathbf{g})\right), \end{aligned} \tag{5.95}$$

Since the above optimization is concave, any stationary point would be a global maximum. We differentiate the objective to find that such a stationary point would have to satisfy:

$$\forall \hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\widehat{C}}),\quad \lambda(\hat{c}) = \frac{\rho(\hat{c})e^{\left(\frac{\alpha-1}{\alpha}\right)(\ln 2)\left(g(\hat{c})-H^{\uparrow}_{\alpha}(Q|Q')_{\rho_{|\hat{c}}}\right)}}{\sum_{\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\hat{c}})}\rho(\hat{c})e^{\left(\frac{\alpha-1}{\alpha}\right)(\ln 2)\left(g(\hat{c})-H^{\uparrow}_{\alpha}(Q|Q')_{\rho_{|\hat{c}}}\right)}}. \tag{5.96}$$

We first note that the above equations can only have a solution if $\sum_{\hat{c}}\lambda(\hat{c}) = 1$ (this can be seen by summing the equations over $\hat{c}$). Therefore, let us first focus on the case where $\boldsymbol{\lambda}\in\mathbb{P}_{\widehat{C}}$ and has support equal to $\mathrm{supp}(\boldsymbol{\rho}_{\widehat{C}})$. In this case, $\log\lambda(\hat{c})$ and $\log\rho(\hat{c})$ are finite for all $\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\widehat{C}})$, which allows us to take the log on both sides of Eq. (5.96) and show that a valid solution exists. Evaluating the objective function at that solution yields

$$G^*_{\alpha,\rho}(\boldsymbol{\lambda}) = \frac{\alpha}{\alpha-1}D\left(\boldsymbol{\lambda}\middle\|\boldsymbol{\rho}_{\widehat{C}}\right) + \sum_{\hat{c}\in\mathrm{supp}(\boldsymbol{\rho}_{\widehat{C}})}\lambda(\hat{c})H^{\uparrow}_{\alpha}(Q|Q')_{\rho_{|\hat{c}}}, \tag{5.97}$$

where the first term is the KL divergence. Other remaining cases can be handled similarly to the proof of Lemma 4.4.4. □

With this, we have the following duality result, which is similar to Lemma 4.4.5.

Lemma 5.4.12. *Let $\mathcal{M}\in\mathrm{CPTP}(Q, SE\widehat{C})$ where the output is always classical on $\widehat{C}$, let $\widetilde{E}$ be a register such that $\dim(Q)\leq\dim(\widetilde{E})$, and let* Pur *be a purifying function for Q onto $\widetilde{E}$ (Definition 4.2.2). Let $\mathbb{P}_{\widehat{C}}$ denote the set of probability distributions on $\widehat{C}$, and consider any $S_\Omega\subseteq\mathbb{P}_{\widehat{C}}$ and any $\Theta\subseteq\mathrm{S}_{=}(Q)$. For any $\omega\in\mathrm{S}_{=}(Q)$, define a corresponding state $\nu^\omega := \mathcal{M}\left[\mathtt{Pur}\left(\omega\right)\right]$. Then, the optimization*

$$\sup_{\mathbf{f}}\inf_{\mathbf{q}\in S_\Omega}\inf_{\omega\in\Theta}\left(H^{\uparrow,f}_{\alpha}(S|\widehat{C}E\widetilde{E})_{\nu^\omega} + \mathbf{f}\cdot\mathbf{q}\right), \tag{5.98}$$

is the Lagrange dual problem of the following constrained optimization:

$$\begin{aligned} &\inf_{\mathbf{q}\in S_\Omega}\inf_{\boldsymbol{\lambda}\in\mathbb{P}_{\widehat{C}}}\inf_{\omega\in\Theta}\left(\frac{\alpha}{\alpha-1}D\left(\boldsymbol{\lambda}\middle\|\boldsymbol{\nu}^\omega_{\widehat{C}}\right) + \sum_{\hat{c}\in\mathrm{supp}(\boldsymbol{\nu}^\omega_{\widehat{C}})}\lambda(\hat{c})H^{\uparrow}_{\alpha}(S|E\widetilde{E})_{\nu^\omega_{|\hat{c}}}\right), \\ &\qquad s.t.\quad \mathbf{q}-\boldsymbol{\lambda} = \mathbf{0} \end{aligned} \tag{5.99}$$

in which the objective function is jointly convex in $\boldsymbol{\lambda}$ *and* ω *(and* $\mathbf{q}$*, trivially).*

Furthermore, if S_Ω *is convex and* Θ *is convex and compact, the values of* (5.98) *and* (5.99) *are equal, i.e., strong duality holds.*

Proof. The proof is identical to that of Lemma 4.4.4, except with the convex-conjugate result in Lemma 5.4.11 instead. □

Aided by the above lemmas, we now prove Theorem 5.4.2.

Proof. We proceed in a similar fashion to the proof of Theorem 4.4.1. We begin by considering the same channels $\mathcal{N}_j$ and state as in the proof of Theorem 5.4.3. Let $\widetilde{\Omega}$ be the set of all values $\hat{c}_1^n$ with nonzero probability in $\rho_{|\Omega}$. Then, we have the following result, which is analogous to Corollary 4.3.1:

$$\begin{aligned}
H_\alpha^\uparrow(S_1^n|\widehat{C}_1^n E_n)_{\rho_{|\Omega}} &\geq H_\alpha^\uparrow(D_1^n S_1^n|\widehat{C}_1^n E_n)_{\rho_{|\Omega}} - \max_{\hat{c}_1^n \in \widetilde{\Omega}} H_\alpha(D_1^n)_{\rho_{|\hat{c}_1^n}} \\
&\geq H_\alpha^\uparrow(D_1^n S_1^n|\widehat{C}_1^n E_n)_{\rho} - \max_{\hat{c}_1^n \in \widetilde{\Omega}} H_\alpha(D_1^n)_{\rho_{|\hat{c}_1^n}} - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega} \\
&= H_\alpha^\uparrow(D_1^n S_1^n|\widehat{C}_1^n E_n)_{\rho} - nM + \min_{\hat{c}_1^n \in \widetilde{\Omega}} f_{\text{full}}(\hat{c}_1^n) - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega} \\
&\geq \left(\sum_j \inf_{\nu \in \Sigma_j} H_\alpha^{\uparrow, f_{|\hat{c}_1^{j-1}}}(S_j|\widehat{C}_j E_j \widetilde{E})_\nu\right) + \min_{\hat{c}_1^n \in \widetilde{\Omega}} f_{\text{full}}(\hat{c}_1^n) - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega},
\end{aligned} \tag{5.100}$$

where the first line is Lemma 5.4.10, the second line is [41, Lemma B.5], the third line is (5.59), and the fourth line is by (5.89) and (5.91).

Let f be an arbitrary tradeoff function on $\widehat{C}$ such that for all tradeoff functions we have $f_{|\hat{c}_1^{j-1}} = f$. Then, the above bound yields

$$\begin{aligned}
H_\alpha^\uparrow(S_1^n|\widehat{C}_1^n E_n)_{\rho_{|\Omega}} &\geq \left(\sum_j \inf_{\nu \in \Sigma_j} H_\alpha^{\uparrow, f}(S_j|\widehat{C}_j E_j \widetilde{E})_\nu\right) + \inf_{\mathbf{q} \in S_\Omega} \mathbf{f}\cdot\mathbf{q}\, n - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega} \\
&\geq \inf_{\mathbf{q} \in S_\Omega} \inf_{\nu \in \Sigma_{S\widehat{C}E\widetilde{E}}} \left(H_\alpha^{\uparrow, f}(S|\widehat{C}E\widetilde{E})_\nu + \mathbf{f}\cdot\mathbf{q}\right) n - \frac{\alpha}{\alpha-1}\log\frac{1}{p_\Omega},
\end{aligned} \tag{5.101}$$

where the first line rearranges the summation in f_{full} the same way as the Corollary 4.4.1 proof, and the second line holds since the marginal-constrained convex range $\Sigma_{S\widehat{C}E\widetilde{E}}$ contains every state $\nu \in \Sigma_j$.

We now reparametrize the optimization over $\Sigma_{S\widehat{C}E\widetilde{E}}$ in terms of a specifically constructed channel $\mathcal{M}$. We begin with embedding all the registers A_{j-1} and E_{j-1} into some register A and E of large enough dimensions. secondly, extend all the channels $\mathcal{M}_j$ to act on $\mathrm{S}_=(AE)$. Then, we let J to be a register of dimension n, and define $\mathcal{M} \in \mathrm{CPTP}(AEJ, S\widehat{C}E)$ to be a channel that does the following

1. Measure J in some basis $|j\rangle$ (we refer to this basis as its classical basis).
2. Conditioned on outcome value j, implement the channel $\mathcal{M}_j$ on the registers AE, producing an output state on $S\widehat{C}E$ (via the embedding chosen in the definition of $\Sigma_{S\widehat{C}E\widetilde{E}}$).

Furthermore, we define the following convex and compact set:

$$\Theta_{AEJ} \coloneqq \left\{ \omega \in \mathrm{S}_=(AEJ) \;\middle|\; \omega \text{ is classical on } J, \text{ with } \omega_{AE|j} \in \mathrm{S}_=(A_{j-1}E_{j-1}) \text{ and } \omega_{A|j} = \sigma^{(j-1)}_{A_{j-1}} \right\}, \tag{5.102}$$

where the conditions on $\omega_{|j}$ are interpreted with respect to the embeddings of $A_{j-1}E_{j-1}$ in AE.

We now show that the channel $\mathcal{M}$ and set Θ_{AEJ} have the following property: This channel and set have the following critical property:

$$\Sigma_{S\widehat{C}E\widetilde{E}} = \left\{ \mathcal{M}\left[\omega_{AEJ\widetilde{E}}\right] \;\middle|\; \omega \in \mathrm{S}_=(AEJ\widetilde{E}) \;\text{ s.t. } \omega_{AEJ} \in \Theta_{AEJ} \right\}, \tag{5.103}$$

$\subseteq$ direction: Let $\sigma_{S\widehat{C}E\widetilde{E}} \in \Sigma_{S\widehat{C}E\widetilde{E}}$. Then,

$$\exists\, \omega_{AEJ\widetilde{E}} \quad \text{s.t.} \quad \sigma_{S\widehat{C}E\widetilde{E}} = \mathcal{M}\left(\omega_{AEJ\widetilde{E}}\right) \;\wedge \omega_{AEJ\widetilde{E}} = \sum_j p_j \omega_{A_{j-1}E_{j-1}\widetilde{E}|j} \otimes |j\rangle\langle j|_J, \tag{5.104}$$

where p_j is a probability distribution and $\omega_{A_{j-1}E_{j-1}\widetilde{E}|j}$ are normalized states satisfying $\omega_{A|j} = \sigma^{(j-1)}A_{j-1}$. Therefore, $\omega_{AEJ} \in \Theta_{AEJ}$.

$\supseteq$ direction: The first operation performed by $\mathcal{M}$ is to measure J in its classical basis, which will collapse the input state $\omega_{AEJ\widetilde{E}}$ into a mixture of the form $\overline{\omega}_{AEJ\widetilde{E}} = \sum_j p_j \overline{\omega}_{AE\widetilde{E}|j} \otimes |j\rangle\langle j|_J$ for some probabilities p_j and normalized conditional states $\overline{\omega}_{AE\widetilde{E}|j}$. Furthermore, given that the input state satisfies $\omega_{AEJ} \in \Theta_{AEJ}$, by definition this means ω_{AEJ} is classical on J, thus it is unchanged by the measurement and we have $\overline{\omega}_{AEJ} = \omega_{AEJ} \in \Theta_{AEJ}$. This implies $\overline{\omega}_{AE|j} \in \mathrm{S}_=(A_{j-1}E_{j-1})$ and $\overline{\omega}_{A|j} = \sigma^{(j-1)}_{A_{j-1}}$, thus the second action of $\mathcal{M}$

produces a mixture of states $\mathcal{M}_j\left[\overline{\omega}_{A_{j-1}E_{j-1}\widetilde{E}|j}\right]$ for some $\overline{\omega}_{A_{j-1}E_{j-1}\widetilde{E}|j}$ satisfying the marginal constraints, as desired.

Therefore, with this parametrization, we can express Eq. (5.101) as follows:

$$\begin{aligned}\inf_{\mathbf{q}\in S_\Omega}\inf_{\nu\in\Sigma_{S\widehat{C}E\widetilde{E}}}\left(H_\alpha^{\uparrow,f}(S|\widehat{C}E\widetilde{E})_\nu+\mathbf{f}\cdot\mathbf{q}\right)&=\inf_{\mathbf{q}\in S_\Omega}\inf_{\substack{\omega\in\mathrm{S}_=(AEJ\widetilde{E})\\ \text{s.t. }\omega_{AEJ}\in\Theta_{AEJ}}}\left(H_\alpha^{\uparrow,f}(S|\widehat{C}E\widetilde{E})_{\mathcal{M}[\omega_{AEJ\widetilde{E}}]}+\mathbf{f}\cdot\mathbf{q}\right)\\&=\inf_{\mathbf{q}\in S_\Omega}\inf_{\omega\in\Theta_{AEJ}}\left(H_\alpha^{\uparrow,f}(S|\widehat{C}E\widetilde{E})_{\nu^\omega}+\mathbf{f}\cdot\mathbf{q}\right),\end{aligned}\tag{5.105}$$

where $\nu^\omega := \mathcal{M}\left[\mathtt{Pur}\left(\omega_{AEJ}\right)\right]$ for some choice of purifying function $\mathtt{Pur}$, as given in Definition 4.2.2, for AEJ onto $\widetilde{E}$. The first line above holds by (5.103), and the second line holds by data-processing in Lemma 5.4.4. Moreover, the objective function in the last line is jointly convex in $\mathbf{q}$ and ω by Lemma 5.4.10.

Since the result of (5.105) holds for any choice of tradeoff functions, the bound can be optimized by maximizing over $\mathbf{f}$, i.e.,

$$\sup_{\mathbf{f}}\inf_{\mathbf{q}\in S_\Omega}\inf_{\omega\in\Theta_{AEJ}}\left(H_\alpha^{\uparrow,f}(S|\widehat{C}E\widetilde{E})_{\nu^\omega}+\mathbf{f}\cdot\mathbf{q}\right).\tag{5.106}$$

By Lemma 5.4.12, the optimal value of the above optimization is the same as Eq. (5.99). The claim of the theorem then follows by observing that the value of (5.99) is equal to $h_\alpha^\uparrow$, since we can substitute the $\mathbf{q}-\boldsymbol{\lambda}=\mathbf{0}$ constraint into the objective and use the same reasoning as in (5.105) to again reparametrize the optimization over input states $\omega\in\Theta_{AEJ}$ into an optimization over $\Sigma_{S\widehat{C}E\widetilde{E}}$. □

We highlight that, similar to the discussion after Lemma 4.4.5 in the previous chapter, the optimal dual solution to (5.98) is precisely the optimal choice of $\mathbf{f}$ to use in (5.101). Therefore, similarly, by extracting the dual solution when solving (5.99), one obtains a choice of $\mathbf{f}$ that yields good key rates when applying Theorems 5.4.1 and 5.4.3.

Before demonstrating how to prove security for PM-QKD under the framework in this chapter, we present the following corollary, which states that the bounds that were obtained under previous variants of EAT (especially, the bound obtained in Chapter 3) are also valid (while suboptimal) under the MEAT framework.

Corollary 5.4.1. *For each $j\in\{1,2,\cdots,n\}$, take a state $\sigma^{(j-1)}\in\mathrm{S}_=(A_{j-1})$, and a channel $\mathcal{M}_j\in\mathrm{CPTP}(A_{j-1}E_{j-1},S_j\overline{C}_j\widehat{C}_jE_j)$, such that $\overline{C}_j\widehat{C}_j$ are classical. Let ρ be a state of the form $\rho_{S_1^n\overline{C}_1^n\widehat{C}_1^nE_n}=\mathcal{M}_n\circ\cdots\circ\mathcal{M}_1[\omega_{A_0^{n-1}E_0}]$ for some $\omega\in\mathrm{S}_=(A_0^{n-1}E_0)$, such that $\omega_{A_0^{n-1}}=\sigma_{A_0}^{(0)}\otimes\cdots\otimes\sigma_{A_{n-1}}^{(n-1)}$.*

Suppose furthermore that $\rho = p_\Omega \rho_{|\Omega} + (1-p_\Omega)\rho_{|\overline{\Omega}}$ *for some* $p_\Omega \in (0,1]$ *and normalized states* $\rho_{|\Omega}, \rho_{|\overline{\Omega}}$. *Let* S_Ω *be a convex set of probability distributions on the alphabet* $\overline{\mathcal{C}} \times \widehat{\mathcal{C}}$, *such that for all* $\bar{c}_1^n \hat{c}_1^n$ *with nonzero probability in* $\rho_{|\Omega}$, *the frequency distribution* $\mathrm{freq}_{\bar{c}_1^n \hat{c}_1^n}$ *lies in* S_Ω. *Moreover, let* f *be a min-tradeoff function, i.e., for all* j*:*

$$\forall \nu \in \Sigma_j, \quad H(S_j \overline{C}_j | \widehat{C}_j E_j \widetilde{E})_\nu \geq f(\boldsymbol{\nu}_{\overline{C}_j \widehat{C}_j}), \tag{5.107}$$

where Σ_j *is defined as in* (5.85) *(i.e. the set of output states of* $\mathcal{M}_j$ *under a marginal constraint and with a purifying register on the input state). Let* h_{vN} *be a value satisfying* $f(\mathrm{freq}_{\bar{c}_1^n \hat{c}_1^n}) \geq h_{\mathrm{vN}}$ *for all* $\bar{c}_1^n \hat{c}_1^n$ *with nonzero probability in* $\rho_{|\Omega}$. *Then for* $\alpha \in (1,2)$ *and* $\widehat{\alpha} = 1/(2-\alpha)$, *we have*

$$\begin{aligned} H_\alpha^\uparrow(S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\rho_{|\Omega}} &\geq n h_{\mathrm{vN}} + n T_\alpha(f) - n(\alpha-1)^2 K_\alpha(f) - \frac{\alpha}{\alpha-1} \log \frac{1}{p_\Omega}, \\ H_\alpha^\uparrow(S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\rho_{|\Omega}} &\geq n h_{\mathrm{vN}} + n T_{\widehat{\alpha}}(f) - n(\widehat{\alpha}-1)^2 K_{\widehat{\alpha}}(f) - \frac{\alpha}{\alpha-1} \log \frac{1}{p_\Omega}, \end{aligned} \tag{5.108}$$

where (letting $\kappa \coloneqq 1$ *if* S *is always classical and* $\kappa \coloneqq 2$ *otherwise, and letting* $\delta_{\bar{c}\hat{c}}$ *denote the distribution with all its weight on the value* $\bar{c}\hat{c}$*):*

$$\begin{aligned} T_\alpha(f) &\coloneqq \min_j \inf_{\nu \in \Sigma_j} \left(H(S_j \overline{C}_j | \widehat{C}_j E_j \widetilde{E})_\nu - f(\boldsymbol{\nu}_{\overline{C}\widehat{C}}) - (\alpha-1) V(f, \boldsymbol{\nu}_{\overline{C}\widehat{C}}) \right), \\ V(f, \mathbf{p}) &\coloneqq \frac{\ln 2}{2} \left(\log\left(1 + 2 d_{SC}^\kappa\right) + \sqrt{2 + \mathrm{Var}(f, \mathbf{p})} \right)^2, \ \text{where } d_{SC} \coloneqq \max_j \dim(S_j \overline{C}_j), \\ \mathrm{Var}(f, \mathbf{p}) &\coloneqq \sum_{\bar{c}\hat{c}} p(\bar{c}\hat{c}) \left(f(\delta_{\bar{c}\hat{c}}) - \sum_{\bar{c}'\hat{c}'} p(\bar{c}'\hat{c}') f(\delta_{\bar{c}'\hat{c}'}) \right)^2, \\ K_\alpha(f) &\coloneqq \frac{2^{(\alpha-1)(\kappa \log d_{SC} + \mathrm{Max}(f) - \mathrm{Min}_{\mathcal{Q}}(f))}}{6(2-\alpha)^3 \ln 2} \ln^3\left(2^{\kappa \log d_{SC} + \mathrm{Max}(f) - \mathrm{Min}_{\mathcal{Q}}(f)} + e^2\right), \\ \mathrm{Max}(f) &\coloneqq \sup_{\mathbf{p} \in \mathbb{P}_{\overline{C}\widehat{C}}} f(\mathbf{p}), \qquad \mathrm{Min}_{\mathcal{Q}}(f) \coloneqq \min_j \inf_{\nu \in \Sigma_j} f(\boldsymbol{\nu}_{\overline{C}\widehat{C}}). \end{aligned} \tag{5.109}$$

Proof. Let $\mathcal{D}_j \in \mathrm{CPTP}(\overline{C}_j \widehat{C}_j, \overline{C}_j \widehat{C}_j D_j)$ be read-and-prepare channels, as defined in the proof of [37, PropositionV.3], such that the entropies of the D_j register depend on f, analogous to the dependency of tradeoff functions in Lemma 5.4.2. Defining $\mathcal{N}_j \coloneqq \mathcal{D}_j \circ \mathcal{M}_j$,

we can apply Theorem 5.3.1 iteratively to obtain

$$\begin{aligned} H_\alpha^\uparrow(D_1^n S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\mathcal{N}_n \circ \cdots \circ \mathcal{N}_1[\omega]} &= H_\alpha^\uparrow(D_1^n S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\left(\mathcal{N}_n \otimes \mathrm{id}_{\widehat{C}_1^{n-1}}\right) \circ \cdots \circ \mathcal{N}_1[\omega]} \\ &\geq \sum_j H_\alpha^\uparrow \left(\mathcal{N}_j \otimes \mathrm{id}_{\widehat{C}_1^{j-1}}, D_j S_j \overline{C}_j, [\omega_{A_{j-1}}]\right) \\ &= \sum_j H_\alpha^\uparrow \left(\mathcal{N}_j, D_j S_j \overline{C}_j, [\omega_{A_{j-1}}]\right) \\ &= \sum_j \inf_{\nu' \in \Sigma_j'} H_\alpha^\uparrow (D_j S_j \overline{C}_j | \widehat{C}_j E_j \widetilde{E})_{\nu'}, \end{aligned} \tag{5.110}$$

where Σ_j' denotes the set of all states that could be produced by $\mathcal{N}_j$ acting on some initial state $\omega' \in \mathrm{S}_=(A_{j-1} E_{j-1} \widetilde{E})$ such that $\omega'_{A_{j-1}} = \sigma^{(j-1)}_{A_{j-1}}$. The remainder of the proof then follows the same steps as in the proof of [81, Supplement, Theorem 2]. □

5.5 Security of PM-QKD protocols via the MEAT

In this section, we show how the MEAT statements from the previous section can be used to bound the secure key rate in a PM-QKD protocol. We begin with the simple cases—namely, protocols where the measurements are all completed before any public announcements are made. In this case, by employing the source-replacement technique [26, 49], the state produced in the protocol before the error-correction and privacy amplification procedure can be written as $\widetilde{\mathcal{M}}_n \otimes \cdots \otimes \widetilde{\mathcal{M}}_1[\omega^0]$, for some measurement channels $\widetilde{\mathcal{M}}_j \in \mathrm{CPTP}(A_j B_j, S_j \widehat{C}_j I_j)$ acting on the joint state of Alice, Bob, and Eve $\omega^0 \in \mathrm{S}_=(A_1^n B_1^n \widehat{E})$, such that it satisfies the marginal constraints on A_1^n. Then, a bound on the desired entropic quantity, which in turn leads to a bound on the securely extractable key length, is obtained by applying the result of Theorems 5.4.1 and 5.4.2 (depending on whether the protocol is a fixed-length protocol or a variable-length protocol) when identifying registers as follows:

- $E_j \coloneqq B_j^{n-1} I_1^j \widehat{E}$,
- $\mathcal{M}_j \coloneqq \widetilde{\mathcal{M}}_j \otimes \mathrm{id}_{B_j^{n-1} I_1^{j-1} \widehat{E}}$,

and observing that it is not necessary to explicitly account for the $\mathrm{id}_{B_j^{n-1} I_1^{j-1} \widehat{E}}$ terms in these channels when computing the single-round quantities, because all those computations

are stabilized by the purifying register $\widetilde{E}$, which can be used to absorb the optimization over the input registers to the identity channel $\mathrm{id}_{B_j^{n-1} I_1^{j-1} \widehat{E}}$.

Moving on to the more general situation, the MEAT framework can also handle scenarios where some announcements are made before the completion of all measurements. This structure of announcements is termed "on-the-fly" announcements, which requires a more sophisticated analysis. To that end, we begin by stating the structure of protocols compatible with this framework.

Protocol 5. Prepare-and-measure protocol.

Parameters:

$n \in \mathbb{N}_0$	:	Total number of rounds
$\{\rho^{(j)}_{X_j Q_j}\}_{j=1}^n$	:	Classical-quantum states prepared by Alice
$\{t_j^A, t_j^B, t_j^{\mathrm{ann}}\}_{j=1}^n$	:	Timings of various steps (see below)

We require the times listed above to satisfy the following conditions:[15]

- For each $j \in \{1, 2, \ldots, n\}$, we have $t_j^A < t_j^B < t_j^{\mathrm{ann}}$.
- Each of the sequences $\{t_j^A\}_{j=1}^n$, $\{t_j^B\}_{j=1}^n$, and $\{t_j^{\mathrm{ann}}\}_{j=1}^n$ is monotonically increasing.

Protocol steps:

1. Alice and Bob perform the following steps for every $j \in \{1, 2, \ldots, n\}$.
   (a) **State preparation and transmission:** At time t_j^A, independently for each round, Alice prepares some state $\rho^{(j)}_{X_j Q_j}$ where X_j is a classical register. She then sends out Q_j to Bob through a quantum channel that is controlled by Eve.
   (b) **Measurements:** At time t_j^B, Bob receives some quantum register from Eve and acts on it with some POVM. Then, he stores the outcome in a register Y_j.
   (c) **Public announcement:** At time t_j^{ann}, Alice and Bob publicly announce (and potentially communicate) some classical information based only on the values $X_j Y_j$. Denote all such public information as a classical register I_j, and suppose they contain a "public test data" register $\widehat{C_j}$ for use in later steps.

[15] We note that these requirements are minimal. The first condition, $t_j^B > t_j^A$, is the fundamental physical requirement, and the second inequality, $t_j^{\mathrm{ann}} > t_j^B$, prevents Eve from attacking the state based on the basis choices. Essentially, the second condition is a time-ordering constraint in the sequence of preparations, measurements, and announcements *individually*. Notably, these conditions do *not* restrict the repetition rate, unlike existing GEAT-based proofs.

(d) **Raw data processing:** At any time after t_j^{ann}, Alice computes a classical value S_j based only on her raw data X_j and the public announcements I_j . Furthermore, if desired, at this point Alice can also compute some "secret test data" register $\overline{C}_j$ that is not announced.

2. Alice and Bob apply various further classical procedures, such as an acceptance test or variable-length decision (depending on whether it is a fixed-length or variable-length protocol [101]), error correction, and privacy amplification.

To prove security for a QKD protocol, one needs to bound the Rényi conditional entropy or f-weighted entropy, depending on whether the goal is to prove the security of a variable-length or fixed-length protocol, respectively, of the state produced at the end of Step 1. However, the structure of the MEAT statement provided in this chapter does not align well with the protocol description. Therefore, we reduce the analysis to another model that is compatible with the MEAT statement. We note that this reduction entails no loss of generality.

We begin by considering a modified protocol in which all of Alice's preparation steps, as well as the announcement steps, are shifted to an earlier time in the protocol. To be more precise, we define new timings $\tilde{t}_j^A$ and $\tilde{t}_j^{\text{ann}}$ such that they satisfy

$$\tilde{t}_1^A < \cdots < \tilde{t}_n^A < t_1^B < \tilde{t}_1^{\text{ann}} < t_2^B < \tilde{t}_2^{\text{ann}} < \cdots < \tilde{t}_n^{\text{ann}}. \tag{5.111}$$

In other words, Bob's timing remains unchanged, whereas Alice's timings are all shifted to occur before Bob performs his first measurement. Additionally, the announcement timings are moved forward so that, in each round, they occur immediately after Bob performs his measurement for that round. In other words, in this modified protocol, Alice first prepares and sends out all the states. Then, Bob begins receiving the first state and performs a measurement on it. This step is followed by Alice and Bob making announcements to generate I_1, and they proceed to the next round, continuing this process until all measurements and announcements are completed.[16]

Given that the security of the modified protocol implies the security of the original protocol, we focus on analyzing the former. In the modified version, all of Alice's states are prepared before Bob performs his measurements. With this in mind, we apply the source-replacement technique, wherein Alice's preparation step, which generates the CQ

[16]We note that no generality is lost in focusing on this modified protocol, since, compared to Protocol 5, we are essentially providing various registers to Eve at an earlier time. She can perform any action she could have in the original protocol by simply ignoring the fact that she has access to more states earlier.

state $\bigotimes_{j=1}^{n} \rho_{X_j Q_j}^{(j)}$, is equivalent to an alternative preparation in which Alice prepares a pure state $\bigotimes_{j=1}^{n} \left|\psi^{(j)}\right\rangle_{A_j Q_j}$ and then performs a suitable measurement on the A_j systems. Note that Alice's measurement on A_j commutes with all other operations except the public announcements involving X_j, which means it can be deferred until that step. Therefore, we consider the following entanglement-based protocol, which produces exactly the same state as the modified prepare-and-measure protocol described above.

Protocol 6. Virtual entanglement-based protocol.

Parameters:

$$\begin{aligned} n \in \mathbb{N}_0 \quad &: \quad \text{Total number of rounds} \\ \left\{\left|\psi^{(j)}\right\rangle_{A_j Q_j}\right\}_{j=1}^{n} \quad &: \quad \text{Quantum states prepared by Alice} \\ \{t_j^B, \tilde{t}_j^{\text{ann}}\}_{j=1}^{n} \quad &: \quad \text{Timings of various steps (see below)} \end{aligned}$$

Protocol steps:

0. **State preparation and transmission:** Alice prepares a pure state $\bigotimes_{j=1}^{n} \left|\psi^{(j)}\right\rangle_{A_j Q_j}$. She then sends out Q_1^n to Bob through a quantum channel that is controlled by Eve.

1. For each round $j \in \{1, 2, \ldots, n\}$, Alice and Bob perform the following steps.

   (a) **Measurements:** At time t_j^B, Bob receives some quantum register from Eve and acts on it with some POVM. Then, he stores the outcome in a register Y_j. At the same time, Alice measures A_j and stores the outcome in a register X_j.

   (b) **Public announcement:** At time $\tilde{t}_j^{\text{ann}}$, Alice and Bob publicly announce (and potentially communicate) some classical information based only on the values $X_j Y_j$. Denote all such public information as a classical register I_j, and suppose that they contain a "public test data" register $\widehat{C}_j$ for use in later steps.

   (c) **Raw data processing:** At time $\tilde{t}_j^{\text{ann}}$, Alice computes a classical value S_j based solely on her raw data X_j and the public announcements I_j. If desired, she may also compute a "secret test data" register $\overline{C}_j$ at this point, which is not announced.[17]

[17] Note that this step differs in time ordering compared to Protocol 5. However, since the processing operations Alice uses to generate the S_j register commute with all other operations occurring after $\tilde{t}_j^{\text{ann}}$, this change does not affect the protocol.

2. Alice and Bob apply various further classical procedures, such as an acceptance test or variable-length decision (depending on whether it is a fixed-length or variable-length protocol [101]), error correction, and privacy amplification.

The virtual protocol described above is now compatible with the channel structure required for MEAT. Specifically, we identify the registers A_1^n (resp. Q_1^n) in the above protocol with the registers A_0^{n-1} (resp. E_0) in the MEAT theorems in Secs. 5.4.2–5.4.3.[18] Let $\mathcal{M}_j$ denote the channel that performs all physical processes occurring during the time interval $\left(\tilde{t}_{j-1}^{\text{ann}}, \tilde{t}_j^{\text{ann}}\right]$. That is, it includes Eve performing any operations on her side-information based on I_{j-1} and preparing a state to send to Bob, followed by Alice and Bob measuring their j^{th} systems, announcing the public information I_j (which Eve immediately incorporates into her side-information), and generating Alice's value S_j. With this structure in place, we can apply various MEAT statements from Secs. 5.4.2–5.4.3 to derive a secure bound on the extractable key rate.

5.6 Conclusion and open questions

In this chapter, we derived a chain rule for channel conditional entropies defined through sandwiched Rényi entropies, for the region $\alpha \geq 1$. Furthermore, we showed that these channel conditional entropies are strongly additive across a tensor product of channels. Motivated by cryptography, we combined this chain rule with techniques developed in Chapter 4 and [126] to prove the MEAT, a new EAT result that is compatible with having some form of marginal constraint on the input. Finally, we explained how one can use the analysis in this chapter to obtain a security statement for prepare-and-measure protocols, without imposing any of the repetition-rate constraints required in the GEAT approach, and covering a wider range of scenarios than [126].

Going forward, several directions can be explored. Notably, as briefly mentioned earlier, one limitation of this framework compared to the GEAT is its handling of a "secret" memory register. In particular, as noted in previous chapters, the GEAT considers channels of the form

$$\mathcal{M}_j \in \text{CPTP}(R_{j-1}E_{j-1}, S_jR_jE_j\overline{C}_j\widehat{C}_j), \tag{5.112}$$

[18] We highlight that if the protocol is such that it keeps a classical secret register $\overline{C}$, due to some technical difficulties, we were unable to prove a theorem analogous to Theorem 5.4.2 such that it accounts for $\overline{C}$ as well. However, we emphasize that this limitation is specific to Theorem 5.4.2; as we proved, the accumulation of f-weighted entropies in Theorem 5.4.3 remains valid even in the presence of a classical secret register $\overline{C}$.

where the R_j registers serve as internal memory registers that can be updated in each round and passed to the next channel. The single-round entropies analyzed in the GEAT have the form $H_\alpha(S_j\overline{C}_j|\widehat{C}_jE_j\widetilde{E})$, without the R_j registers appearing, as long as the channels satisfy a technical "non-signaling" condition from R_{j-1} to E_j (see Definition 4.2.1).[19]

The MEAT does not accommodate such a property, since all the output registers of each channel appear somewhere in the single-round entropies in our results. Conversely, the GEAT channel structure can in principle incorporate the MEAT channel structure by setting the R_j registers in the GEAT channels to be trivial (in which case the NS condition automatically holds) and identifying each E_j register with the registers $A_j^{n-1}E_j$ in the MEAT channel structure, taking the j^{th} GEAT channel to act as identity on the registers A_j^{n-1}.[20] However, this comes at the cost that the GEAT *does not* impose any marginal constraints on the input states, which makes it difficult to use for analyzing PM-QKD in the manner we described in Sec. 5.5.

A natural question is whether the GEAT and the MEAT frameworks can be unified. The GEAT was obtained by deriving a chain rule similar to our Theorem 5.3.1, but where each channel can output registers R_j that do not appear in the conditioning, via a series of arguments based on the NS condition. Unfortunately, the series of arguments used in [89] does not directly translate to the MEAT framework. For instance, it is unclear how to adapt the statement of Lemma 3.3 in [89] to one that aligns with the structure of the MEAT framework. This remains an important question for future work.

Another potential direction for future work is to explore whether a similar entropy accumulation statement can be formulated with different marginal constraints. For example, instead of requiring the global marginal state in Theorem 5.4.1 to be a tensor product of individual states, one could consider a less restrictive class, such as certain forms of separable states that are not necessarily tensor products. In a completely different direction, it might be valuable to investigate whether the techniques used in this chapter to prove the strong additivity of channel conditional entropy could provide insights into additivity problems in other areas of quantum information theory.

[19] We note that without such a non-signaling assumption, the memory register would, in fact, affect the entropies by appearing in the conditioning register of the single-round conditional entropy.

[20] Note that this means the registers A_j^{n-1} would appear in the conditioning systems for the j^{th}-round entropy term in the GEAT; however, in the MEAT they were effectively also present as part of the stabilizing register, in some sense.

Chapter 6

Mutual information chain rules for security proofs robust against device imperfections

So far, all the analyses have been carried out under the assumption of perfect sources and detectors. Here, we propose a model to account for memory-based imperfections at both the source and detector levels.

This chapter is based on [8].

6.1 Introduction

Conventional security proofs in quantum cryptographic protocols—such as quantum key distribution (QKD)—typically assume that, following measurement, the resulting raw key data remains secure and independent of subsequent processes. However, this assumption often breaks down in practical implementations, where real devices may exhibit imperfections. Examples include photon leakage from detectors [75, 100] or source imperfections and correlations [24, 96, 86, 25], where the emitted states may deviate from their ideal forms or display memory effects across rounds.

To more accurately reflect practical setups, it is essential to incorporate such imperfections into the security analysis. We also emphasize that a suitable security framework capable of handling these imperfections should be robust against small deviations from the

ideal case. While this may seem intuitive, phenomena like information locking [73, 36, 133] demonstrate that even minimal leakage can lead to severe security vulnerabilities.

To address this issue, in this chapter, we propose a variety of chain rules involving mutual information quantities to quantify the amount of information leakage from imperfect devices. While most intuitive in the context of photon leakage or similar mechanisms, we demonstrate that these rules are also applicable to other imperfections, such as source correlations. From an information-theoretic standpoint, our bounds are natural in that they involve mutual information quantities, capturing the correlation between the leakage and the secret data. Moreover, they are straightforward to apply in practice, as our final results involve simply subtracting easily computed corrections from the key rates obtained through existing entropic proof techniques.

We begin by proving a chain rule that connects smooth min-entropy and smooth max-information, characterizing the information leakage caused by an additional conditioning register in a one-shot setting. In prior works, the usual method was to subtract the logarithm of the leakage register's dimension. In [127], a tighter bound was established using smooth max-entropy to quantify the leakage. However, that bound still falls short of the tightest result one might hope for. Intuitively, when leakage occurs, what should matter is the amount of its correlation with the secret data. This intuition is supported by the chain rules for von Neumann entropies, where the impact of removing a conditioning register is captured by the conditional quantum mutual information (CQMI) [20].

One of the main challenges in obtaining the tightest possible bound is that there is currently no consensus on a suitable definition of smooth max-CQMI, and in this thesis, we do not aim to introduce such a definition. However, it is well known that an upper bound on CQMI in the von Neumann regime is the mutual information between appropriately chosen registers, which is also a measure of correlation [68, 29, 30]; furthermore, unlike smooth max-CQMI, there are well-behaved definitions of smooth max-information [22]. Here, we show a chain rule in which the effect of removing a conditioning register is measured by mutual information, thus improving over previous bounds by being based on a more natural quantity.

We highlight that a chain rule for the Rényi generalization of the above (without smoothing) was derived in [87]. However, the grouping of registers in that chain rule is such that it is harder to apply in contexts such as QKD security proofs. Moving on, given that in QKD applications the mutual information measure is typically evaluated over an n-round global state, we prove an "information bounding theorem" that provides a method to decompose Rényi mutual information into a sum of mutual information terms evaluated over the so-called "single-round" state.

However, these results still have some limitations in how leakage processes can be handled, as they essentially analyze the leakage registers as a "separate" process from the main protocol. This is restrictive, since for some device imperfections, a more realistic description of the protocol would allow the leakage to occur during the protocol itself, and it does not seem straightforward to apply the above tools in such cases. To address these scenarios, we modify the result of the GEAT [88] to incorporate leakage within that framework, allowing the adversary to act on the leakage register during the protocol.

We note that the question of quantum cryptography with leaky devices has been considered in various works: [75, 100, 115, 114, 95] studied various forms of leakage behavior in IID scenarios; [64] considered a leakage model accounting for a bounded number of qubits; and [112] explored various forms of "probabilistic" constrained leakage. Furthermore, various source imperfection models were studied in [24, 96, 25]. In particular, an approximate version of entropy accumulation was developed in [85], and a related approach was used in [86] to analyze some forms of source imperfection. However, it is unclear whether the class of imperfections considered here is comparable to that of [86].

6.2 One-shot chain rule

In this section, we derive a chain rule relating the smooth min-entropy to the smooth max-information, suitable for one-shot characterizations of information leakage introduced by an additional conditioning register. This is achieved through an operational argument based on the task of quantum state redistribution. To clarify the discussion, we first define state redistribution protocols:

Definition 6.2.1. (δ-Quantum state redistribution; see [35, 3, 16]) Consider any state $\rho \in \mathrm{S}_{=}(ABCD)$. The protocol consists of three parties, Alice, Bob, and a reference party, where Alice begins with registers AC, Bob has B, and the reference party has D. In addition, Alice and Bob have access to any shared entangled state $\rho_{A'B'}$. The goal is for Alice to transmit the C register to Bob, possibly assisted by the shared entanglement, such that the final state on $ABCD$ after the protocol is δ-close in purified distance to the initial state. More formally, let Q denote a quantum register Alice sends to Bob in the protocol. A **δ-quantum state redistribution protocol for** ρ consists of a pair of encoding and decoding channels $\mathcal{E} \in \mathrm{CPTP}(AA'C, AQ)$ and $\mathcal{D} \in \mathrm{CPTP}(QBB', BC)$, such that

$$P\left(\rho_{ABCD}, (\mathcal{D} \circ \mathcal{E})\left(\rho_{ABCD} \otimes \rho_{A'B'}\right)\right) \leq \delta. \tag{6.1}$$

We refer to $\log |Q|$ as the **quantum communication cost** of the protocol.

We now employ this task to derive a one-shot chain rule.

Theorem 6.2.1. *For any $\rho \in \mathrm{S}_{=}(SLE)$, consider a purification $|\rho\rangle\langle\rho| \in \mathrm{S}_{=}(SLEP)$. Suppose there exists a δ-quantum state redistribution protocol for ρ (Definition 6.2.1) in which Alice holds PL, Bob's initial register is E, the reference holds S, and the register to be transmitted is L[1]. Let the quantum communication cost of the protocol be $\log|Q|$ with Q being the encoded register as mentioned in that definition. Then for any $\epsilon \in [0,1)$, the following holds:*

$$H_{\infty}^{\uparrow,\varepsilon+\delta}(S|LE)_{\rho} \geq H_{\infty}^{\uparrow,\varepsilon}(S|E)_{\rho} - 2\log|Q|. \tag{6.2}$$

If the register Q in the state redistribution protocol is classical, then in the above bound, $2\log|Q|$ can be replaced with $\log|Q|$.

Proof. Consider a δ-quantum state redistribution protocol where Alice, Bob, and the reference hold onto the registers indicated in the statement of the theorem. We write the initial state of the protocol as $\rho_{SLEPXY} = |\rho\rangle\langle\rho|_{SLEP} \otimes \rho_{XY}$, where ρ_{XY} is the shared entanglement between Alice and Bob. Then, by Definition 6.2.1 there exist an encoding $\mathcal{E} \in \mathrm{CPTP}(LPX, PQ)$, and decoding $\mathcal{D} \in \mathrm{CPTP}(QEY, LE)$, such that

$$P\left(\rho''_{SLEP}, \rho_{SLEP}\right) \leq \delta \tag{6.3}$$

where (leaving identity maps implicit) $\rho''_{SLEP} = (\mathcal{D} \circ \mathcal{E})\left(\rho_{SLEPXY}\right)$. Denoting the state after the encoding map as $\rho'_{SEPQY} = \mathcal{E}(\rho_{SLEPXY})$, we have

$$\begin{aligned}
H_{\infty}^{\uparrow,\varepsilon+\delta}(S|LE)_{\rho} &\geq H_{\infty}^{\uparrow,\varepsilon}(S|LE)_{\mathcal{D}\circ\mathcal{E}[\rho]} \\
&\geq H_{\infty}^{\uparrow,\varepsilon}(S|QEY)_{\mathcal{E}[\rho]} \\
&\geq H_{\infty}^{\uparrow,\varepsilon}(S|EY)_{\mathcal{E}[\rho]} - 2\log|Q| \\
&= H_{\infty}^{\uparrow,\varepsilon}(S|EY)_{\rho} - 2\log|Q| \\
&\geq H_{\infty}^{\uparrow,\varepsilon}(S|E)_{\rho} - 2\log|Q|,
\end{aligned} \tag{6.4}$$

where the first line holds by Eq. (6.3), the second line is an application of data-processing (Theorem 2.2.1), the third line follows from Proposition. 3.3.5 in [77], the fourth line follows from the fact that the encoding channel $\mathcal{E}$ acts as identity on SEY registers, and the last line holds since by construction $\rho_{SEY} = \rho_{SE} \otimes \rho_{Y}$. Note that by the isometric equivalence

[1]Note that while the ordering of the assigned registers here seems unintuitive at first glance, it is in fact in the correct ordering to give a proper CQMI-like term.

of purifications and the isometric invariance property of min-entropies, the above proof holds for all purifications of ρ_{SLE}.

The version for classical Q holds by simply noting that we instead have $H_\infty^{\uparrow,\varepsilon}(S|QEY)_{\mathcal{E}[\rho]} \geq H_\infty^{\uparrow,\varepsilon}(S|EY)_{\mathcal{E}[\rho]} - \log|Q|$ in that case, or alternatively by noting that any such state redistribution protocol can be converted to one where Q is fully quantum but halved in size, via dense coding (since we are allowing arbitrary entanglement assistance). □

So far, all existing achievability and converse bounds on one-shot state redistribution protocols, as shown in [5, 4, 2, 3], indicate that the optimal quantum communication cost should take the form of a one-shot CQMI-like quantity. However, the quantities introduced in those works are not easy to analyze. It is also worth pointing out that neither of the definitions provided in [17] is very well-behaved. For this reason, we resort to a simplified version of the above result, based on a looser term involving smooth max-information. Specifically, [3] provides an achievable rate for the quantum communication cost of a δ-quantum state redistribution task, which involves a looser quantity than smooth CQMI—namely, the smooth max-information between suitable registers. We use that result to derive the following bound:

Corollary 6.2.1. *For any $\rho \in \mathrm{S}_=(SLE)$, $\delta \in (0,1)$, and any $\epsilon \in [0,1)$ the following holds:*

$$H_\infty^{\uparrow,\varepsilon+\delta}(S|LE)_\rho \geq H_\infty^{\uparrow,\varepsilon}(S|E)_\rho - I_\infty^{\downarrow,\delta}(SE;L)_\rho - \log\left(\frac{4}{\delta^2}\right). \tag{6.5}$$

Proof. According to the result in [3], there exists a δ-quantum state redistribution protocol in which Alice holds PL, Bob's registers are trivial[2], the reference is ES, and the register to be transmitted is L, such that the quantum communication cost is[3] $\log|Q| = \frac{1}{2} I_\infty^{\downarrow,\delta}(SE;L)_\rho + \log\left(\frac{2}{\delta}\right)$. Note, however, that this can also be validly viewed as a δ-quantum state redistribution protocol where Alice holds PL, Bob holds E, and the reference is S; in which we just constrained Bob's decoding channel to act as the identity on E. Therefore, we can apply Theorem 6.2.1 on this result, which gives

$$\begin{aligned} H_\infty^{\uparrow,\varepsilon+\delta}(S|LE)_\rho &\geq H_\infty^{\uparrow,\varepsilon}(S|E)_\rho - 2\log|Q| \\ &\geq H_\infty^{\uparrow,\varepsilon}(S|E)_\rho - I_\infty^{\downarrow,\delta}(SE;L)_\rho - \log\left(\frac{4}{\delta^2}\right), \end{aligned} \tag{6.6}$$

as claimed. □

[2] This scenario is usually referred to as **quantum state splitting**.

[3] Strictly speaking the proof in that work only straightforwardly applies when the smoothing in $I_\infty^{\downarrow,\delta}$ is taken over normalized states, but as discussed bellow Definition 2.2.10, this is not an issue for this version of smooth max-information.

It is also possible to derive a similar chain rule for the case where the smooth max-information is defined based on the unoptimized definition of Rényi mutual information in Definition 2.2.7 (i.e., I_∞^ϵ instead of $I_\infty^{\downarrow,\epsilon}$). However, we note that the derivation in this case is more straightforward, proceeding from the SDP characterization of min-entropy combined with the simultaneous smoothing technique, and does not rely on any operational characterization. We start by first presenting a version without smoothing:

Theorem 6.2.2. *For any $\rho \in \mathrm{S}_\leq(SLE)$, the following holds:*

$$H_\infty^\uparrow(S|LE)_\rho \geq H_\infty^\uparrow(S|E)_\rho - I_\infty(L:SE)_\rho\,. \tag{6.7}$$

Proof. Consider the SDP for D_∞ [117, Chapter 4.2.4]:

$$\begin{aligned} 2^{D_\infty(\rho_{SLE}\|\rho_L\otimes\rho_{SE})} = \underset{\mu}{\text{minimize}} \quad & \mu \\ \text{subject to} \quad & \rho_{SLE} \leq \mu\rho_L\otimes\rho_{SE} \\ & \mu \geq 0\,. \end{aligned} \tag{6.8}$$

From this, it is clear that

$$\rho_{SLE} \leq 2^{D_\infty(\rho_{SLE}\|\rho_L\otimes\rho_{SE})}\rho_L\otimes\rho_{SE}\,. \tag{6.9}$$

Now consider the SDP for $H_\infty^\uparrow$ [117, Definition 6.2] (note that this is consistent with the $H_\infty^\uparrow$ definition we used; Definition 2.2.5):

$$\begin{aligned} 2^{-H_\infty^\uparrow(S|LE)_\rho} = \quad & \underset{\sigma_{LE}}{\text{minimize}} \quad && \mathrm{Tr}[\sigma_{LE}] \\ & \text{subject to} && \mathbb{I}_S\otimes\sigma_{LE} \geq \rho_{SLE} \\ & && \sigma_{LE} \geq 0 \\ \leq \quad & \underset{\sigma_{LE}}{\text{minimize}} && \mathrm{Tr}[\sigma_{LE}] \\ & \text{subject to} && \mathbb{I}_S\otimes\sigma_{LE} \geq 2^{D_\infty(\rho_{SLE}\|\rho_L\otimes\rho_{SE})}\rho_L\otimes\rho_{SE} \\ & && \sigma_{LE} \geq 0\,. \end{aligned} \tag{6.10}$$

The inequality holds because by Eq. (6.9), the feasible set in the second SDP is smaller.

We further consider

$$\begin{aligned} 2^{-H_\infty^\uparrow(S|E)_\rho} = \underset{\omega_E}{\text{minimize}} \quad & \mathrm{Tr}\left[\omega_E\right] \\ \text{subject to} \quad & \mathbb{I}_S\otimes\omega_E \geq \rho_{SE} \\ & \omega_E \geq 0\,, \end{aligned} \tag{6.11}$$

and denote an optimizer for this by ω_E^*, i.e. ω_E^* is feasible and $\mathrm{Tr}\left[\omega_E^*\right] = 2^{-H_\infty^\uparrow(S|E)_\rho}$.

For the SDP in Eq. (6.10), we now pick

$$\sigma_{LE} = 2^{D_\infty(\rho_{SLE}||\rho_L\otimes\rho_{SE})}\rho_L \otimes \omega_E^* \,. \tag{6.12}$$

This choice of σ_{LE} is feasible since

$$\mathbb{I}_S \otimes \sigma_{LE} = 2^{D_\infty(\rho_{SLE}||\rho_L\otimes\rho_{SE})}\mathbb{I}_S \otimes \rho_L \otimes \omega_E^* \geq 2^{D_\infty(\rho_{SLE}||\rho_L\otimes\rho_{SE})}\rho_L \otimes \rho_{SE} \,, \tag{6.13}$$

where the inequality holds because ω_E^* is feasible, i.e., $\mathbb{I}_S \otimes \omega_E^* \geq \rho_{SE}$. Therefore we get

$$\begin{aligned} 2^{-H_\infty^\uparrow(S|LE)_{\rho_{SLE}}} &\leq \mathrm{Tr}\left[2^{D_\infty(\rho_{SLE}||\rho_L\otimes\rho_{SE})}\rho_L \otimes \omega_E^*\right] \\ &\leq 2^{D_\infty(\rho_{SLE}||\rho_L\otimes\rho_{SE})} \cdot \mathrm{Tr}\left[\omega_E^*\right] \\ &= 2^{D_\infty(\rho_{SLE}||\rho_L\otimes\rho_{SE}) - H_\infty^\uparrow(S|E)_\rho} \,. \end{aligned} \tag{6.14}$$

Taking the logarithm yields the result. □

We now proceed to prove a chain rule involving the smoothed quantities.

Corollary 6.2.2. *For any $\rho \in \mathrm{S}_\leq(SLE)$ and $\varepsilon \in [0,1)$, the following holds:*

$$H_\infty^{\uparrow,\varepsilon}(S|LE)_\rho \geq H_\infty^{\uparrow,\varepsilon/3}(S|E)_\rho - I_\infty^{\varepsilon/3}(L:SE)_\rho \,. \tag{6.15}$$

Proof. The proof technique we use here is similar to the one used in [22, 15, 40]. For compactness of notation, let us write $B_\varepsilon(\rho) \subset \mathrm{S}_\leq(SLE)$ to denote the set of subnormalized states ε-close to ρ in purified distance. The proof follows from a sequence of lower bounds on $H_\infty^{\uparrow,\varepsilon}(X|BC)_\rho$. First, from Theorem 6.2.2 we have

$$H_\infty^{\uparrow,\varepsilon}(S|LE)_\rho \geq \max_{\omega\in B_\varepsilon(\rho)} \left(H_\infty^\uparrow(S|E)_\omega - I_\infty(L:SE)_\omega\right) \tag{6.16}$$

We can lower-bound the maximization by splitting it into two parts as follows:

$$(6.16) \geq \max_{\hat{\omega}\in B_{\varepsilon/3}(\rho)} \quad \max_{\substack{T_{SE}\in\mathrm{End}(SE) \text{ s.t.} \\ T_{SE}\hat{\omega}_{SLE}T_{SE}^\dagger \in B_{2\varepsilon/3}(\hat{\omega})}} \left(H_\infty^\uparrow(S|E)_{T_{SE}\hat{\omega}_{SLE}T_{SE}^\dagger} - I_\infty(L:SE)_{T_{SE}\hat{\omega}_{SLE}T_{SE}^\dagger}\right) \tag{6.17}$$

The above is a lower bound because the maximization in Eq. (6.17) is over a subset of the maximization in Eq. (6.16). We now observe that since D_∞ satisfies data-processing for

any completely positive map [117], it follows that $I_\infty(L:SE)_{T_{SE}\hat{\omega}_{SLE}T_{SE}^\dagger} \leq I_\infty(L:SE)_{\hat{\omega}_{SLE}}$. Therefore, it is, in turn, lower bounded by:

$$(6.17) \geq \max_{\hat{\omega}\in B_{\varepsilon/3}(\rho)} \max_{\substack{T_{SE}\in \mathrm{End}(SE) \text{ s.t.} \\ T_{SE}\hat{\omega}_{SLE}T_{SE}^\dagger \in B_{2\varepsilon/3}(\hat{\omega})}} \left(H_\infty^\uparrow(S|E)_{T_{SE}\hat{\omega}_{SLE}T_{SE}^\dagger} - I_\infty(L:SE)_{\hat{\omega}_{SLE}}\right) \tag{6.18}$$

We now pick $\hat{\omega} \in B_{\varepsilon/3}(\rho)$ such that $I_\infty(L:SE)_{\hat{\omega}_{SLE}} = I_\infty^{\varepsilon/3}(L:SE)_\rho$, obtaining another lower bound:

$$(6.18) \geq \max_{\substack{T_{SE}\in \mathrm{End}(SE) \text{ s.t.} \\ T_{SE}\hat{\omega}_{SLE}T_{SE}^\dagger \in B_{2\varepsilon/3}(\hat{\omega})}} \left(H_\infty^\uparrow(S|E)_{T_{SE}\hat{\omega}_{SLE}T_{SE}^\dagger}\right) - I_\infty^{\varepsilon/3}(L:SE)_\rho \tag{6.19}$$

Let $\tilde{\rho}_{SE} \in B_{\varepsilon/3}(\rho_{SE})$ be such that $H_\infty^{\uparrow,\varepsilon/3}(S|E)_\rho = H_\infty^\uparrow(S|E)_{\tilde{\rho}}$. By [40, Lemma B.3] there exists a $\hat{T}_{SE}$ such that $\hat{T}_{SE}\hat{\omega}_{SLE}\hat{T}_{SE}^\dagger \in \mathrm{S}_\leq(SLE)$ is an extension of $\tilde{\rho}_{SE}$ and $P(\hat{\omega}_{SLE}, \hat{T}_{SE}\hat{\omega}_{SLE}\hat{T}_{SE}^\dagger) = P(\hat{\omega}_{SE}, \tilde{\rho}_{SE})$. We can bound the latter quantity using the triangle inequality for the purified distance, obtaining $P(\hat{\omega}_{SE}, \tilde{\rho}_{SE}) \leq P(\hat{\omega}_{SE}, \rho_{SE}) + P(\rho_{SE}, \tilde{\rho}_{SE}) \leq \varepsilon/3 + \varepsilon/3 = 2\varepsilon/3$. Therefore, this $\hat{T}_{SE}$ lies inside the set of T_{SE} over which we optimize, so we can pick $\hat{T}_{SE}$ to get a final lower bound:

$$H_\infty^\uparrow(S|E)_{\tilde{\rho}} - I_\infty^{\varepsilon/3}(L:SE)_\rho = H_\infty^{\uparrow,\varepsilon/3}(S|E)_\rho - I_\infty^{\varepsilon/3}(L:SE)_\rho\,, \tag{6.20}$$

as claimed. □

The chain rules derived in Corollary 6.2.1 and 6.2.2 can be used in cryptographic protocols that account for device imperfections. For instance, let the S register hold the raw secret data, E be the adversary's side-information register, and L be a register containing leakage caused by imperfections in the measurement devices. In such cases, one approach is to remove the leakage term from the conditioning register. This allows the use of standard security proofs on the conditional entropy without the leakage register, while subtracting a term that quantifies the leaked information, measured here by the smooth max-information between suitable registers.

6.3 Information bounding theorem

In this section, we derive a theorem that bounds the Rényi mutual information produced by a sequence of channels, similar to the entropy accumulation theorem [41]. For this purpose, we first prove the following lemma, which can be used to construct a chain rule for Rényi mutual information.

Lemma 6.3.1. *Let $\rho \in \mathrm{S}_=(ABC)$, $\sigma \in \mathrm{Pos}(C)$, $\omega' \in \mathrm{S}_=(A)$, and $\omega'' \in \mathrm{S}_=(B)$. Then, for any $\alpha \in [0, \infty)$ we have*

$$\inf_{\omega_{AB} \in \mathrm{S}_=(AB)} D_\alpha(\rho_{ABC} \| \omega_{AB} \otimes \sigma_C) \leq D_\alpha(\rho_{AC} \| \omega'_A \otimes \sigma_C) + D_\alpha(\nu_{ABC} \| \omega''_B \otimes \nu_{AC}), \tag{6.21}$$

where $\nu_{ABC} = \nu_{AC}^{\frac{1}{2}} \rho_{B|AC} \nu_{AC}^{\frac{1}{2}}$, with

$$\nu_{AC} = \frac{\left(\rho_{AC}^{\frac{1}{2}} \left(\omega'_A \otimes \sigma_C\right)^{\frac{1-\alpha}{\alpha}} \rho_{AC}^{\frac{1}{2}}\right)^\alpha}{\mathrm{Tr}\left[\left(\rho_{AC}^{\frac{1}{2}} \left(\omega'_A \otimes \sigma_C\right)^{\frac{1-\alpha}{\alpha}} \rho_{AC}^{\frac{1}{2}}\right)^\alpha\right]}. \tag{6.22}$$

In particular

$$I_\alpha^\downarrow(C; AB)_\rho \leq I_\alpha^\downarrow(C; A)_\rho + I_\alpha^\downarrow(AC; B)_\nu \tag{6.23}$$

Proof. Let $|\psi\rangle_{ABCE}$ be any purification of ρ_{ABC}. Then, the proof follows from the following

crucial relation:

$$
\begin{aligned}
&D_\alpha\left(\rho_{ABC}\|\omega'_A\otimes\omega''_B\otimes\sigma_C\right)\\
&=\sup_{\tau_E}\frac{\alpha}{\alpha-1}\log\left\|\omega_A'^{\frac{1-\alpha}{2\alpha}}\otimes\omega_B''^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_C^{\frac{1-\alpha}{2\alpha}}\otimes\tau_E^{-\frac{1-\alpha}{2\alpha}}\left|\psi\right\rangle\right\|_2^2\\
&=\sup_{\tau_E}\frac{\alpha}{\alpha-1}\log\left\|\omega_A'^{\frac{1-\alpha}{2\alpha}}\otimes\omega_B''^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_C^{\frac{1-\alpha}{2\alpha}}\rho_{AC}^{\frac{1}{2}}\otimes\tau_E^{-\frac{1-\alpha}{2\alpha}}\rho_{AC}^{-\frac{1}{2}}\left|\psi\right\rangle\right\|_2^2\\
&=\sup_{\tau_E}\frac{\alpha}{\alpha-1}\log\operatorname{Tr}\left[\omega_A'^{\frac{1-\alpha}{2\alpha}}\otimes\omega_B''^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_C^{\frac{1-\alpha}{2\alpha}}\rho_{AC}^{\frac{1}{2}}\otimes\tau_E^{-\frac{1-\alpha}{2\alpha}}\rho_{AC}^{-\frac{1}{2}}\left|\psi\right\rangle\left\langle\psi\right|\rho_{AC}^{-\frac{1}{2}}\rho_{AC}^{\frac{1}{2}}\right.\\
&\qquad\left.\omega_A'^{\frac{1-\alpha}{2\alpha}}\otimes\omega_B''^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_C^{\frac{1-\alpha}{2\alpha}}\otimes\tau_E^{\frac{1-\alpha}{2\alpha}}\right]\\
&=\sup_{\tau_E}\frac{\alpha}{\alpha-1}\log\operatorname{Tr}\left[\left(\rho_{AC}^{\frac{1}{2}}\omega_A'^{\frac{1-\alpha}{\alpha}}\otimes\omega_B''^{\frac{1-\alpha}{\alpha}}\otimes\sigma_C^{\frac{1-\alpha}{\alpha}}\rho_{AC}^{\frac{1}{2}}\otimes\tau_E^{-\frac{1-\alpha}{2\alpha}}\right)\rho_{AC}^{-\frac{1}{2}}\left|\psi\right\rangle\left\langle\psi\right|\rho_{AC}^{-\frac{1}{2}}\right.\\
&\qquad\left.\left(\mathbb{I}_{ABC}\otimes\tau_E^{\frac{1-\alpha}{2\alpha}}\right)\right]\\
&=\sup_{\tau_E}\frac{\alpha}{\alpha-1}\log\operatorname{Tr}\left[\left(\nu_{AC}^{\frac{1}{\alpha}}\otimes\omega_B''^{\frac{1-\alpha}{\alpha}}\otimes\tau_E^{-\frac{1-\alpha}{2\alpha}}\right)\rho_{AC}^{-\frac{1}{2}}\left|\psi\right\rangle\left\langle\psi\right|\rho_{AC}^{-\frac{1}{2}}\left(\mathbb{I}_{ABC}\otimes\tau_E^{\frac{1-\alpha}{2\alpha}}\right)\right]\\
&\quad+\frac{1}{\alpha-1}\log\operatorname{Tr}\left[\left(\rho_{AC}^{\frac{1}{2}}\omega_A'^{\frac{1-\alpha}{\alpha}}\otimes\sigma_C^{\frac{1-\alpha}{\alpha}}\rho_{AC}^{\frac{1}{2}}\right)^\alpha\right]\\
&=\sup_{\tau_E}\frac{\alpha}{\alpha-1}\log\operatorname{Tr}\left[\left(\nu_{AC}^{\frac{1-\alpha}{2\alpha}}\otimes\omega_B''^{\frac{1-\alpha}{2\alpha}}\otimes\tau_E^{-\frac{1-\alpha}{2\alpha}}\right)\nu_{AC}^{\frac{1}{2}}\rho_{AC}^{-\frac{1}{2}}\left|\psi\right\rangle\left\langle\psi\right|\rho_{AC}^{-\frac{1}{2}}\nu_{AC}^{\frac{1}{2}}\right.\\
&\qquad\left.\left(\nu_{AC}^{\frac{1-\alpha}{2\alpha}}\otimes\omega_B''^{\frac{1-\alpha}{2\alpha}}\otimes\tau_E^{\frac{1-\alpha}{2\alpha}}\right)\right]+\frac{1}{\alpha-1}\log\operatorname{Tr}\left[\left(\rho_{AC}^{\frac{1}{2}}\omega_A'^{\frac{1-\alpha}{\alpha}}\otimes\sigma_C^{\frac{1-\alpha}{\alpha}}\rho_{AC}^{\frac{1}{2}}\right)^\alpha\right]\\
&=\sup_{\tau_E}\frac{\alpha}{\alpha-1}\log\left\|\nu_{AC}^{\frac{1-\alpha}{2\alpha}}\otimes\omega_B''^{\frac{1-\alpha}{2\alpha}}\otimes\tau_E^{-\frac{1-\alpha}{2\alpha}}\left|\nu\right\rangle\right\|_2^2+\frac{1}{\alpha-1}\log\operatorname{Tr}\left[\left(\rho_{AC}^{\frac{1}{2}}\omega_A'^{\frac{1-\alpha}{\alpha}}\otimes\sigma_C^{\frac{1-\alpha}{\alpha}}\rho_{AC}^{\frac{1}{2}}\right)^\alpha\right]\\
&=D_\alpha\left(\nu_{ABC}\|\nu_{AC}\otimes\omega''_B\right)+D_\alpha\left(\rho_{AC}\|\omega'_A\otimes\sigma_C\right),
\end{aligned}
\tag{6.24}
$$

where in the first equality we invoked [41, Lemma B.2], in the second equality we insert $\rho_{AC}^{\frac{1}{2}}\rho_{AC}^{-\frac{1}{2}}$, the third equality holds by noting that $\|A\|_2^2=\operatorname{Tr}[A^\dagger A]$, the fourth equality is just a reshuffling, the fifth inequality follows from substituting Eq. (6.22), the sixth equality is also a reshuffling, in the penultimate equality we defined $|\nu\rangle:=\nu_{AC}^{\frac{1}{2}}\rho_{AC}^{-\frac{1}{2}}|\psi\rangle$ to be the purification of ν_{ABC}, and the last equality follows by invoking [41, Lemma B.2], and noting that $\nu_{AC}=\operatorname{Tr}_B[\nu_{ABC}]$.

From Eq. (6.24) we immediately obtain Eq. (6.21) as a consequence, by simply picking $\omega_{AB} = \omega'_A \otimes \omega''_B$ as a feasible point in the infimum over ω_{AB}. Now to prove Eq. (6.23), we pick $\sigma_C = \rho_C$, and $\omega''_B = \omega^*_B, \omega'_A = \omega^*_A$ such that the following relations hold:

$$\begin{aligned} D_\alpha(\nu_{ABC}||\omega^*_B \otimes \nu_{AC}) &= \inf_{\sigma_B \in \mathrm{S}_=(B)} D_\alpha(\nu_{ABC}||\sigma_B \otimes \nu_{AC}) \\ D_\alpha(\rho_{AC}||\omega^*_A \otimes \rho_C) &= \inf_{\sigma_A \in \mathrm{S}_=(A)} D_\alpha(\rho_{AC}||\sigma_A \otimes \rho_C). \end{aligned} \tag{6.25}$$

With these choices, Eq. (6.21) states:

$$\begin{aligned} I^\downarrow_\alpha(C;AB)_\rho &= \inf_{\sigma_{AB} \in \mathrm{S}_=(AB)} D_\alpha(\rho_{ABC}||\sigma_{AB} \otimes \rho_C) \\ &\leq D_\alpha(\rho_{AC}||\omega^*_A \otimes \rho_C) + D_\alpha(\nu_{ABC}||\omega^*_B \otimes \nu_{AC}) \\ &= I^\downarrow_\alpha(C;A)_\rho + I^\downarrow_\alpha(AC;B)_\nu, \end{aligned} \tag{6.26}$$

where the second line follows from definition 2.2.7. □

We note that by setting $\omega'_A = \rho_A$, $\omega''_B = \rho_B$, and $\sigma_C = \rho_C$, Eq. (6.24) can be viewed as a Rényi analogue of the following relation for von Neumann mutual information:

$$D(\rho_{ABC}||\rho_A \otimes \rho_B \otimes \rho_C) = D(\rho_{AC}||\rho_A \otimes \rho_C) + D(\rho_{ABC}||\rho_B \otimes \rho_{AC}) = I(A:C) + I(AC:B), \tag{6.27}$$

where the LHS is sometimes termed as **multipartite mutual information** $I(A:B:C) \coloneqq D(\rho_{ABC}||\rho_A \otimes \rho_B \otimes \rho_C)$, which serves as an upper bound on $I(AB:C)$. Therefore, roughly speaking, the bound in Eq. (6.23) is loose up to the difference between Rényi version of $I(A:B:C)$ and $I(AB;C)$, ignoring the fact that one of the terms in Eq. (6.23) is evaluated on a different state. Due to this intrinsic looseness, deriving a converse bound to Eq. (6.23) is not straightforward[4]. However, we later show that the bound is sufficient for various applications.

When it comes to applying Eq. (6.23), a drawback is that the term $I^\downarrow_\alpha(AC;B)_\nu$ must be evaluated on a complicated state. To circumvent this difficulty, we show that this term can instead be replaced by an optimization over input states to some channel.

[4]Another way to understand this looseness is to observe that, for the von Neumann mutual information, the following chain rule holds:

$$I(AB:C) = I(A:C) + I(A:B|C). \tag{6.28}$$

In contrast, in Eq. (6.23), the term $I(A:B|C)$ in the limit $\alpha \to 1$ is relaxed to its upper bound $I(AC:B)$.

Lemma 6.3.2. *Let $\rho' \in \mathrm{S}_=(ARC)$ and $\mathcal{N} \in \mathrm{CPTP}(R, B)$. Let $\rho_{ABC} = \mathcal{N}(\rho'_{ARC})$, then for any $\alpha \in [0, \infty)$ we have:*

$$I^{\downarrow}_{\alpha}(C; AB)_{\mathcal{N}(\rho')} \leq I^{\downarrow}_{\alpha}(C; A)_{\rho'} + \sup_{\omega \in \mathrm{S}_=(ARC)} I^{\downarrow}_{\alpha}(AC; B)_{\mathcal{N}(\omega)}. \tag{6.29}$$

Proof. Defining $S_\rho \coloneqq \left\{\sigma \in \mathrm{S}_=(ABC) \middle| \ \sigma_{B|AC} = \rho_{B|AC}\right\}$, we have

$$I^{\downarrow}_{\alpha}(C; AB)_{\rho} \leq I^{\downarrow}_{\alpha}(C; A)_{\rho} + \sup_{\sigma \in S_\rho} I^{\downarrow}_{\alpha}(AC; B)_{\sigma}, \tag{6.30}$$

where the inequality follows from Eq. (6.23), and by noting that ν_{ABC} in that equation is a feasible solution to the above optimization. Since by definition $\rho'_{AC} = \rho_{AC}$, it is now only needed to show that the optimization in Eq. (6.29) contains the set S_ρ defined above (i.e., $\nu_{ABC} \in S_\rho$). In order to do so let $\rho_{ABC} = \mathcal{N}(\rho'_{ARC})$, Defining $\omega \in \mathrm{S}_=(ARC)$ such that

$$\omega_{ARC} = \nu^{\frac{1}{2}}_{AC} \rho^{-\frac{1}{2}}_{AC} \rho'_{ARC} \rho^{-\frac{1}{2}}_{AC} \nu^{\frac{1}{2}}_{AC}. \tag{6.31}$$

Then we have

$$\begin{aligned} \mathcal{N}(\omega_{ARC}) &= \nu^{\frac{1}{2}}_{AC} \rho^{-\frac{1}{2}}_{AC} \mathcal{N}(\rho'_{ARC}) \rho^{-\frac{1}{2}}_{AC} \nu^{\frac{1}{2}}_{AC} \\ &= \nu^{\frac{1}{2}}_{AC} \rho_{B|AC} \nu^{\frac{1}{2}}_{AC} \\ &= \nu_{ABC}. \end{aligned} \tag{6.32}$$

where the last line follows from definition of ν_{ABC} in lemma 6.3.1. □

We can now apply the result in the above lemma iteratively to obtain an information bounding theorem, which bounds the Rényi mutual information of a state produced by a sequence of channels in terms of the Rényi mutual information of the individual channel outputs. Furthermore, by employing techniques from [22, Theorem 3], we can use this theorem to derive bounds on the smooth max-information chain rules introduced in Corollaries 6.2.1–6.2.2.

Theorem 6.3.1 (Information bounding theorem)**.** *Consider a sequence of channels $\{\mathcal{M}_i \in \mathrm{CPTP}(R_{i-1}, R_i L_i)\}_{i=1}^n$. Let $\epsilon \in [0, 1)$. Then, for any $\alpha > 1$, and any $\rho \in \mathrm{S}_=(R_0 SE)$ we have:*

$$I^{\downarrow\downarrow,\epsilon}_{\infty}(SE : L^n_1)_{\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 SE})} \leq \sum_{i=1}^{n} \sup_{\sigma \in \mathrm{S}_=(R_{i-1} SE)} I^{\downarrow}_{\alpha}(L^{i-1}_1 SE; L_i)_{\mathcal{M}_i(\sigma_{R_{i-1} SE})} - \frac{g(\epsilon)}{\alpha - 1}, \tag{6.33}$$

where $g(\epsilon) = \log\left(1 - \sqrt{1 - \epsilon^2}\right)$.

Proof. First note that by combining Proposition 2.2.1 with the generic inequality $I_\alpha^{\downarrow\downarrow} \leq I_\alpha^{\downarrow}$, we have

$$I_\infty^{\downarrow\downarrow,\epsilon}(SE : L_1^n)_{\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 SE})} \leq I_\alpha^{\downarrow}(SE; L_1^n)_{\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 SE})} - \frac{g(\epsilon)}{\alpha - 1}. \tag{6.34}$$

Then to bound the Rényi mutual information term, we employ Lemma 6.3.2:

$$\begin{aligned} I_\alpha^{\downarrow}(SE; L_1^n)_{\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 SE})} \leq\ & I_\alpha^{\downarrow}(SE; L_1^{n-1})_{\mathcal{M}_{n-1} \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 SE})} \\ & + \sup_{\sigma \in \mathrm{S}_=(R_{n-1} SE)} I_\alpha^{\downarrow}(L_1^{n-1} SE; L_n)_{\mathcal{M}_n(\sigma_{R_{n-1} SE})}. \end{aligned} \tag{6.35}$$

Repeating this step n times we get:

$$I_\alpha^{\downarrow}(SE; L_1^n)_{\mathcal{M}_n \circ \cdots \circ \mathcal{M}_1(\rho_{R_0 SE})} \leq \sum_{i=1}^{n} \sup_{\sigma \in \mathrm{S}_=(R_{i-1} SE)} I_\alpha^{\downarrow}(L_1^{i-1} SE; L_i)_{\mathcal{M}_i(\sigma_{R_{i-1} SE})}. \tag{6.36}$$

Combining Eq. (6.36) with (6.34) proves the Theorem. □

We point out that, when the registers have bounded dimensions, continuity arguments such as those in [37, 41] can be used to bound the Rényi mutual information as $I_\alpha^{\downarrow}(L_1^{i-1} SE; L_i) \leq I(L_1^{i-1} SE; L_i) + O(\alpha - 1)$. However, we will show later that working directly with the Rényi mutual information yields tighter bounds.

Remark 6.3.1. *Note that another work [23] has also analyzed Rényi mutual information quantities in some "sequential" processes, though mainly those focused on contexts such as quantum hypothesis testing and quantum-feedback-assisted channel capacities. We believe that the proof techniques in that work (specifically, Sec. 5) can be used to obtain a bound with a somewhat similar form to Theorem 6.3.1, but where the global state is instead $\mathcal{M}_n \otimes \cdots \otimes \mathcal{M}_1(\rho_{R_1^n SE})$ for some channels $\{\mathcal{M}_i : R_i \to L_i\}_{i=1}^n$; in other words, it is produced by a tensor product of channels rather than a sequence of channels. However, a detailed investigation of this point is left for future work.*

Combining the results in this section with the chain rules obtained in the previous section (specifically, Theorem 6.3.1 together with Corollaries 6.2.1–6.2.2) yields a fairly straightforward treatment of the problem of leakage in QKD. However, we point out that one drawback of this method is that it requires the leakage registers L_1^n to be handled separately from the rest of the protocol. This implies, for example, that under this model, the adversary cannot interact actively with the leaked information while the states are being sent and measured during the protocol. Furthermore, the leakage channel is restricted

to the specific structure imposed by Theorem 6.3.1, thereby limiting the applicability of the model to certain leakage scenarios. In the next section, we introduce an alternative model that resolves this limitation by incorporating the leakage process directly into an entropy accumulation framework.

6.4 Entropy accumulation with leakage

One of the drawbacks of the EAT-style frameworks is that they do not naturally accommodate certain device imperfections, such as source correlations or leakage from measurement devices. This limitation arises due to the non-signaling constraint in the GEAT framework [89] and the Markov condition in the original EAT framework [41], as briefly discussed earlier in Secs. 2.5.2–2.5.3. More precisely, to model such scenarios, one must introduce an additional register—referred to as the leakage register—which transmits information to the adversary in each round. However, this transmission inherently violates the non-signaling and Markov conditions required by the GEAT and the original EAT, respectively. In this section, we derive a chain rule that overcomes this limitation in the GEAT framework and enables the treatment of such imperfections within that framework.

6.4.1 Relating Rényi conditional entropies to Rényi mutual information

We begin with the following corollary, which establishes a relation analogous to those in Corollaries 6.2.1 and 6.2.2, but formulated in terms of Rényi entropies.

Corollary 6.4.1. *Let $\rho \in \mathrm{S}_{=}(ABC)$ and $\nu_{ABC} = \nu_{AC}^{\frac{1}{2}} \rho_{B|AC} \nu_{AC}^{\frac{1}{2}}$, where ν_{AC} is defined in (6.22). Then, for any $\alpha \in [0, \infty)$ we have*

$$H_\alpha^\uparrow(C|AB)_\rho \geq H_\alpha^\uparrow(C|A)_\rho - I_\alpha^\downarrow(AC;B)_\nu. \tag{6.37}$$

Moreover, let $\rho' \in \mathrm{S}_{=}(RAC)$, and $\mathcal{M} \in \mathrm{CPTP}(R, B)$ be a channel such that $\rho_{ABC} = \mathcal{M}(\rho'_{RAC})$. Then, the above relation can be relaxed to the following

$$H_\alpha^\uparrow(C|AB)_{\mathcal{M}(\rho')} \geq H_\alpha^\uparrow(C|A)_{\rho'} - \sup_{\omega \in \mathrm{S}_{=}(RAC)} I_\alpha^\downarrow(AC;B)_{\mathcal{M}(\omega)}. \tag{6.38}$$

Proof. The proof of Eq. (6.37) follows directly by applying Lemma 6.3.1, setting the states in that lemma to be $\sigma_C = \mathbb{I}_C$ and ω'_A, ω''_B to be such that

$$\begin{aligned} D_\alpha(\rho_{AC}||\omega'_A \otimes \mathbb{I}_C) &= \inf_{\sigma_A \in \mathrm{S}_=(A)} D_\alpha(\rho_{AC}||\sigma_A \otimes \mathbb{I}_C), \\ D_\alpha(\nu_{ABC}||\omega''_B \otimes \nu_{AC}) &= \inf_{\sigma_B \in \mathrm{S}_=(B)} D_\alpha(\nu_{ABC}||\sigma_B \otimes \nu_{AC}). \end{aligned} \tag{6.39}$$

Then we have

$$\begin{aligned} H^\uparrow_\alpha(C|AB)_\rho &= -\inf_{\sigma_{AB} \in \mathrm{S}_=(AB)} D_\alpha(\rho_{ABC}||\sigma_{AB} \otimes \mathbb{I}_C) \\ &\geq -D_\alpha(\rho_{AC}||\omega'_A \otimes \mathbb{I}_C) - D_\alpha(\nu_{ABC}||\omega''_B \otimes \nu_{AC}) \\ &= H^\uparrow_\alpha(C|A)_\rho - I^\downarrow_\alpha(AC;B)_\nu, \end{aligned} \tag{6.40}$$

where the second line follows by applying Lemma 6.3.1, and the last line holds by definition. To prove Eq. (6.38), let us define $S_\rho = \left\{\sigma \in \mathrm{S}_=(ABC) \middle| \ \sigma_{B|AC} = \rho_{B|AC}\right\}$. Since by definition $\nu_{B|AC} = \rho_{B|AC}$, we have $\nu_{ABC} \in S_\rho$, and thus we can relax Eq. (6.37) to the following:

$$H^\uparrow_\alpha(C|AB)_\rho \geq H^\uparrow_\alpha(C|A)_\rho - \sup_{\sigma \in S_\rho} I^\downarrow_\alpha(AC;B)_\sigma. \tag{6.41}$$

It is now sufficient to show that the optimization in Eq. (6.37) contains the set S_ρ. Defining $\omega \in \mathrm{S}_=(RAC)$ as

$$\omega_{RAC} = \nu_{AC}^{\frac{1}{2}} \rho_{AC}^{-\frac{1}{2}} \rho'_{RAC} \rho_{AC}^{-\frac{1}{2}} \nu_{AC}^{\frac{1}{2}} \tag{6.42}$$

Then we have

$$\begin{aligned} \mathcal{M}(\omega_{RAC}) &= \nu_{AC}^{\frac{1}{2}} \rho_{AC}^{-\frac{1}{2}} \mathcal{M}(\rho'_{RAC}) \rho_{AC}^{-\frac{1}{2}} \nu_{AC}^{\frac{1}{2}} \\ &= \nu_{ABC}. \end{aligned} \tag{6.43}$$

It is clear that $\mathcal{M}(\omega_{RAC}) \in S_\rho$, and thus the lemma is proven. □

While the above corollary allows for removing a conditioning register, the resulting penalty term is expressed as an *unconditional* mutual information. However, one might expect—particularly in the limit as $\alpha \to 1$—that the penalty should instead be a conditional mutual information. In the following, we establish a tighter bound in which the penalty term is indeed a conditional mutual information, for a particular choice of definition. We begin by stating a simpler intermediate result.

Lemma 6.4.1. *Let $\rho \in \mathrm{S}_{=}(AB)$, $\sigma \in \mathrm{Pos}(B)$, and $\omega \in \mathrm{Pos}(A)$. Then, for any $\alpha \in [0, \infty)$ we have*

$$D_\alpha(\rho_{AB}||\omega_A \otimes \sigma_B) = D_\alpha(\rho_A||\omega_A) + D_\alpha(\eta_{AB}||\eta_A \otimes \sigma_B), \tag{6.44}$$

where $\eta_{AB} = \eta_A^{\frac{1}{2}} \rho_{B|A} \eta_A^{\frac{1}{2}}$, with

$$\eta_A = \frac{\left(\rho_A^{\frac{1}{2}} \omega_A^{\frac{1-\alpha}{\alpha}} \rho_A^{\frac{1}{2}}\right)^\alpha}{\mathrm{Tr}\left[\left(\rho_A^{\frac{1}{2}} \omega_A^{\frac{1-\alpha}{\alpha}} \rho_A^{\frac{1}{2}}\right)^\alpha\right]}. \tag{6.45}$$

In particular

$$H_\alpha^\uparrow(A|B)_\rho = H_\alpha(A)_\rho - I_\alpha^\downarrow(A;B)_\eta. \tag{6.46}$$

Proof. Let $|\psi\rangle_{ABC}$ be any purification of ρ_{AB}. Then, the proof holds by the following chain

of equalities:

$$
\begin{aligned}
&D_\alpha\left(\rho_{AB}\|\omega_A\otimes\sigma_B\right)\\
&=\sup_{\tau_C}\frac{\alpha}{\alpha-1}\log\left\|\omega_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\otimes\tau_C^{-\frac{1-\alpha}{2\alpha}}\left|\psi\right\rangle\right\|_2^2\\
&=\sup_{\tau_C}\frac{\alpha}{\alpha-1}\log\left\|\omega_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\rho_A^{\frac{1}{2}}\otimes\tau_C^{-\frac{1-\alpha}{2\alpha}}\rho_A^{-\frac{1}{2}}\left|\psi\right\rangle\right\|_2^2\\
&=\sup_{\tau_C}\frac{\alpha}{\alpha-1}\log\operatorname{Tr}\left[\omega_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\rho_A^{\frac{1}{2}}\otimes\tau_C^{-\frac{1-\alpha}{2\alpha}}\rho_A^{-\frac{1}{2}}\left|\psi\right\rangle\left\langle\psi\right|\rho_A^{-\frac{1}{2}}\rho_A^{\frac{1}{2}}\omega_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\otimes\tau_C^{\frac{1-\alpha}{2\alpha}}\right]\\
&=\sup_{\tau_C}\frac{\alpha}{\alpha-1}\log\operatorname{Tr}\left[\left(\rho_A^{\frac{1}{2}}\omega_A^{\frac{1-\alpha}{\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{\alpha}}\rho_A^{\frac{1}{2}}\otimes\tau_C^{-\frac{1-\alpha}{2\alpha}}\right)\rho_A^{-\frac{1}{2}}\left|\psi\right\rangle\left\langle\psi\right|\rho_A^{-\frac{1}{2}}\left(\mathbb{I}_{AB}\otimes\tau_C^{\frac{1-\alpha}{2\alpha}}\right)\right]\\
&=\sup_{\tau_C}\frac{\alpha}{\alpha-1}\log\operatorname{Tr}\left[\left(\eta_A^{\frac{1}{\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{\alpha}}\otimes\tau_C^{-\frac{1-\alpha}{2\alpha}}\right)\rho_A^{-\frac{1}{2}}\left|\psi\right\rangle\left\langle\psi\right|\rho_A^{-\frac{1}{2}}\left(\mathbb{I}_{AB}\otimes\tau_C^{\frac{1-\alpha}{2\alpha}}\right)\right]\\
&\quad+\frac{1}{\alpha-1}\log\operatorname{Tr}\left[\left(\rho_A^{\frac{1}{2}}\omega_A^{\frac{1-\alpha}{\alpha}}\rho_A^{\frac{1}{2}}\right)^\alpha\right]\\
&=\sup_{\tau_C}\frac{\alpha}{\alpha-1}\log\operatorname{Tr}\left[\left(\eta_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\otimes\tau_C^{-\frac{1-\alpha}{2\alpha}}\right)\eta_A^{\frac{1}{2}}\rho_A^{-\frac{1}{2}}\left|\psi\right\rangle\left\langle\psi\right|\rho_A^{-\frac{1}{2}}\eta_A^{\frac{1}{2}}\left(\eta_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\otimes\tau_C^{\frac{1-\alpha}{2\alpha}}\right)\right]\\
&\quad+\frac{1}{\alpha-1}\log\operatorname{Tr}\left[\left(\rho_A^{\frac{1}{2}}\omega_A^{\frac{1-\alpha}{\alpha}}\rho_A^{\frac{1}{2}}\right)^\alpha\right]\\
&=\sup_{\tau_C}\frac{\alpha}{\alpha-1}\log\left\|\eta_A^{\frac{1-\alpha}{2\alpha}}\otimes\sigma_B^{\frac{1-\alpha}{2\alpha}}\otimes\tau_C^{-\frac{1-\alpha}{2\alpha}}\left|\eta\right\rangle\right\|_2^2+\frac{1}{\alpha-1}\log\operatorname{Tr}\left[\left(\rho_A^{\frac{1}{2}}\omega_A^{\frac{1-\alpha}{\alpha}}\rho_A^{\frac{1}{2}}\right)^\alpha\right]\\
&=D_\alpha\left(\eta_{AB}\|\eta_A\otimes\sigma_B\right)+D_\alpha\left(\rho_A\|\omega_A\right),
\end{aligned}
\tag{6.47}
$$

where the first equality follows from [41, Lemma B.2], in the second equality we insert $\rho_{AC}^{\frac{1}{2}}\rho_{AC}^{-\frac{1}{2}}$, the third equality holds by noting that $\|A\|_2^2=\operatorname{Tr}[A^\dagger A]$, the fourth equality is just a reshuffling, the fifth equality follows from Eq. (6.45), the sixth equality is another reshuffling, the penultimate line holds by taking $\left|\eta\right\rangle=\eta_A^{\frac{1}{2}}\rho_A^{\frac{1}{2}}\left|\psi\right\rangle$ to be a purification of η_{AB}, and the last line holds due to [41, Lemma B.2] and noting that $\eta_A=\operatorname{Tr}_B[\eta_{AB}]$. To prove Eq. (6.46), we take $\omega_A=\mathbb{I}_A$ in (6.44), and take an infimum from both sides. Thus, we have:

$$
\begin{aligned}
H_\alpha^\uparrow(A|B)_\rho&=-\inf_{\sigma_B\in\mathrm{S}_=(B)}D_\alpha(\rho_{AB}||\omega_A\otimes\sigma_B)\\
&=-D_\alpha(\rho_A||\mathbb{I}_A)-\inf_{\sigma_B\in\mathrm{S}_=(B)}D_\alpha(\eta_{AB}||\eta_A\otimes\sigma_B)\\
&=H_\alpha(A)_\rho-I_\alpha^\downarrow(A;B)_\eta,
\end{aligned}
\tag{6.48}
$$

which completes the proofs. □

We now prove a bound involving a form of Rényi conditional mutual information:

Corollary 6.4.2. *Let $\rho \in \mathrm{S}_=(ABC)$, and set $\eta_{ABC} \coloneqq \eta_A^{\frac{1}{2}} \rho_{BC|A} \eta_A^{\frac{1}{2}}$ where η_A is the state in Eq. (6.45). Then, for any $\alpha \in [0, \infty)$:*

$$H_\alpha^\uparrow(A|BC)_\rho = H_\alpha^\uparrow(A|B)_\rho - I_\alpha^{\downarrow,\mathrm{diff}}(A;C|B)_\eta. \tag{6.49}$$

Moreover, let $\rho' \in \mathrm{S}_=(RA)$, and $\mathcal{N} \in \mathrm{CPTP}(R, BC)$ be a channel such that $\rho_{ABC} = \mathcal{N}(\rho'_{RA})$. Then, Eq. (6.49) can be relaxed to

$$\begin{aligned} H_\alpha^\uparrow(A|B)_{\mathcal{N}(\rho')} - \sup_{\omega \in \mathrm{S}_=(RA)} I_\alpha^{\downarrow,\mathrm{diff}}(A;C|B)_{\mathcal{N}(\omega)} &\leq H_\alpha^\uparrow(A|BC)_{\mathcal{N}(\rho')} \\ &\leq H_\alpha^\uparrow(A|B)_{\mathcal{N}(\rho')} - \inf_{\omega \in \mathrm{S}_=(RA)} I_\alpha^{\downarrow,\mathrm{diff}}(A;C|B)_{\mathcal{N}(\omega)}. \end{aligned} \tag{6.50}$$

Proof. We start by applying Lemma 6.4.1 to the left side of the above equality:

$$\begin{aligned} H_\alpha^\uparrow(A|BC)_\rho &= H_\alpha(A)_\rho - I_\alpha^\downarrow(A;BC)_\eta \\ &= H_\alpha^\uparrow(A|B)_\rho + I_\alpha^\downarrow(A;B)_\eta - I_\alpha^\downarrow(A;BC)_\eta \\ &= H_\alpha^\uparrow(A|B)_\rho - I_\alpha^{\downarrow,\mathrm{diff}}(A;C|B)_\eta, \end{aligned} \tag{6.51}$$

where the second line follows from applying Eq. (6.45) and noting $\eta_{AB} = \mathrm{Tr}_C\, \eta_{ABC} = \eta_A^{\frac{1}{2}} \rho_{B|A} \eta_A^{\frac{1}{2}}$, and the last line follows from the definition of Rényi conditional mutual information in Definition 2.2.8.
To prove the relaxed version, let us first define $S_\rho = \left\{\sigma \in \mathrm{S}_=(ABC) \middle|\ \sigma_{BC|A} = \rho_{BC|A}\right\}$. Then, we have:

$$H_\alpha^\uparrow(A|B)_\rho - \sup_{\sigma \in S_\rho} I_\alpha^{\downarrow,\mathrm{diff}}(A;C|B)_\sigma \leq H_\alpha^\uparrow(A|BC)_\rho \leq H_\alpha^\uparrow(A|B)_\rho - \inf_{\sigma \in S_\rho} I_\alpha^{\downarrow,\mathrm{diff}}(A;C|B)_\sigma, \tag{6.52}$$

where the inequalities hold by noting that by definition, $\eta_{BC|A} = \rho_{BC|A}$, therefore, $\eta_{ABC} \in S_\rho$. Thus, we only need to show that the optimizations in Eq. (6.50) contain the set S_ρ. Let $\rho_{ABC} = \mathcal{N}(\omega_{RA})$ and define $\omega \in \mathrm{S}_=(RA)$ such that the following holds:

$$\omega_{RA} = \eta_A^{\frac{1}{2}} \rho_A^{-\frac{1}{2}} \rho'_{RA} \rho_A^{-\frac{1}{2}} \eta_A^{\frac{1}{2}}. \tag{6.53}$$

Then,

$$\begin{aligned}\mathcal{N}(\omega_{RA}) &= \eta_A^{\frac{1}{2}}\rho_A^{-\frac{1}{2}}\mathcal{N}(\rho'_{RA})\rho_A^{-\frac{1}{2}}\eta_A^{\frac{1}{2}} \\ &= \eta_{ABC}.\end{aligned} \tag{6.54}$$

It is then clear that $\mathcal{N}(\omega_{RA}) \in S_\rho$, which completes the proof. □

The improved chain rule presented in Corollary 6.4.2 offers several advantages. First, in the limit as $\alpha \to 1$, it converges to its von Neumann counterpart, ensuring consistency with the standard entropy framework. Second, unlike the setting in Corollary 6.4.1, the channels involved here can generate both the B and C registers, providing more structural flexibility. However, a notable drawback is that the register ordering in this result differs inherently from that in Corollary 6.4.1, which prevents the latter from being recovered as a special case of Corollary 6.4.2. More specifically, the correction term one would get from Corollary 6.4.1 is of the form $I_\alpha^\downarrow(C;AB)$, whereas the result in Corollary 6.4.2 can be relaxed to terms of the form $I_\alpha^\downarrow(A;BC)$. Therefore, depending on the application, one formulation may be more suitable than the other.

6.4.2 Incorporating the GEAT

The first step in incorporating the chain rules from the previous section into the GEAT framework is to construct a modification that is compatible with Rényi conditional entropies $H_\alpha^\uparrow$. We begin by restating a generic chain rule for sandwiched divergences.

Fact 6.4.1. *(Corollary 5.1 in [45], with Theorem 3.1 and Lemma 3.5 in [89]) Let $\rho \in \mathrm{S}_=(AA'R)$, $\sigma \in \mathrm{Pos}(AA'R)$, $\mathcal{M} \in \mathrm{CPTP}(AA'R, BB'R')$, and $\mathcal{N} = \mathcal{R}_B \circ \mathcal{M}$ with $\mathcal{R}_B$ being a replacer map to identity (i.e., $\forall \omega \in \mathrm{S}_=(AB)$; $\mathcal{R}_B(\omega) = \omega_A \otimes \mathbb{I}_B$). Then for any $\alpha \in (1,2)$, letting $\widehat{\alpha} := 1/(2-\alpha)$ we have*

$$D_\alpha(\mathcal{M}(\rho)||\mathcal{N}(\sigma)) \leq D_\alpha(\rho||\sigma) + D_{\widehat{\alpha}}(\mathcal{M}||\mathcal{N}). \tag{6.55}$$

Furthermore, if $\mathcal{N}$ is such that there exists a channel $\mathcal{E} \in \mathrm{CPTP}(AA', BB'R')$ such that $\mathcal{N} = \mathcal{E} \circ \mathrm{Tr}_R$, then we have

$$D_\alpha(\mathcal{M}(\rho)||\mathcal{N}(\sigma)) \leq D_\alpha(\rho_{AA'}||\sigma_{AA'}) + D_{\widehat{\alpha}}(\mathcal{M}||\mathcal{N}). \tag{6.56}$$

We now state a simple modification of Lemma 3.6 in [89], where we have $H_\alpha^\uparrow$ instead of H_α.[5]

[5]Similar techniques were used in [85] to prove a similar result for $H_\alpha^\uparrow$ entropies in Theorem 3.2 of [41].

Lemma 6.4.2. *Let $\rho \in \mathrm{S}_=(SRE)$, $\mathcal{M} \in \mathrm{CPTP}(RE, S'R'E')$, and $\widetilde{E}$ be a purifying register for RE. Then for any $\alpha \in (1,2)$, letting $\widehat{\alpha} \coloneqq 1/(2-\alpha)$ we have*

$$H_\alpha^\uparrow(SS'|E')_{\mathcal{M}(\rho)} \geq H_\alpha^\uparrow(S|RE)_\rho + \inf_{\omega \in \mathrm{S}_=(RE\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathrm{Tr}_{R'} \circ \mathcal{M}(\omega)}. \tag{6.57}$$

Furthermore, if there exists a channel $\mathcal{R} \in \mathrm{CPTP}(E, E')$ such that $\mathrm{Tr}_{S'R'} \circ \mathcal{M} = \mathcal{R} \circ \mathrm{Tr}_R$ (i.e., the channel $\mathcal{M}$ is non-signaling from R to E'), then we have[6]*:*

$$H_\alpha^\uparrow(SS'|E')_{\mathcal{M}(\rho)} \geq H_\alpha^\uparrow(S|E)_\rho + \inf_{\omega \in \mathrm{S}_=(RE\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathrm{Tr}_{R'} \circ \mathcal{M}(\omega)}. \tag{6.58}$$

Proof. Let $\tilde{\mathcal{M}}, \tilde{\mathcal{N}} \in \mathrm{CPTP}(SRE, SS'E')$ be such that $\tilde{\mathcal{M}} = \mathrm{id}_S \otimes \mathrm{Tr}_{R'} \circ \mathcal{M}$ and $\tilde{\mathcal{N}} = \mathcal{R}_{S'} \circ \tilde{\mathcal{M}}$, with $\mathcal{R}_{S'}$ being a replacer channel, then for any $\rho \in \mathrm{S}_=(SRE)$, define $\sigma \in \mathrm{Pos}(SRE)$ to be $\sigma_{SRE} \coloneqq \mathbb{I}_S \otimes \sigma^*_{RE}$, where σ^*_{RE} is the optimal solution to the following optimization

$$D_\alpha(\rho_{SRE} || \mathbb{I}_S \otimes \sigma^*_{RE}) = \inf_{\sigma \in \mathrm{S}_=(RE)} D_\alpha(\rho_{SRE} || \mathbb{I}_S \otimes \sigma_{RE}). \tag{6.59}$$

Then, by definition, it is clear that

$$\begin{aligned}
&D_\alpha\left(\tilde{\mathcal{M}}(\rho_{SRE}) \middle|\middle| \tilde{\mathcal{N}}(\sigma_{SRE})\right) \geq \inf_{\omega \in \mathrm{S}_=(SE')} D_\alpha\left(\tilde{\mathcal{M}}(\rho_{SRE}) \middle|\middle| \mathbb{I}_{S'} \otimes \omega_{SE'}\right) = -H_\alpha^\uparrow(SS'|E')_{\mathcal{M}(\rho)} \\
&D_{\widehat{\alpha}}(\tilde{\mathcal{M}} || \tilde{\mathcal{N}}) = \sup_{\omega \in \mathrm{S}_=(RE\widetilde{E})} -H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathrm{Tr}_{R'} \circ \mathcal{M}(\omega)} \\
&D_\alpha(\rho_{SRE} || \sigma_{SRE}) = -H_\alpha^\uparrow(S|RE)_\rho,
\end{aligned} \tag{6.60}$$

where $\widetilde{E}$ is isomorphic to RE. Combining Eq. (6.55) with (6.60), results in the following:

$$H_\alpha^\uparrow(SS'|E')_{\mathcal{M}(\rho)} \geq H_\alpha^\uparrow(S|RE)_\rho + \inf_{\omega \in \mathrm{S}_=(RE\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathrm{Tr}_{R'} \circ \mathcal{M}(\omega)}, \tag{6.61}$$

where we used the fact that $H_\alpha^\uparrow \geq H_\alpha$.

To prove Eq. (6.58), we use the same channels $\tilde{\mathcal{M}}, \tilde{\mathcal{N}}$, and for any $\rho \in \mathrm{S}_=(SRE)$, let $\sigma^*_E \in \mathrm{S}_=(E)$ be the optimizer in the following optimization

$$D_\alpha(\rho_{SE} || \mathbb{I}_S \otimes \sigma^*_E) = \inf_{\sigma \in \mathrm{S}_=(E)} D_\alpha(\rho_{SE} || \mathbb{I}_S \otimes \sigma_E). \tag{6.62}$$

[6] Recently, [46] found a similar chain rule where all the terms are based on $H_\alpha^\uparrow$ entropies, and the stabilizing register $\widetilde{E}$ no longer appears. However, the drawback of that chain rule is that it does not have any analogs to Eq. (6.58).

Define $\sigma \in \mathrm{Pos}(SRE)$ to be $\sigma_{SRE} \coloneqq \mathbb{I}_S \otimes \sigma^*_{RE}$, such that σ^*_{RE} is any extension of σ^*_E to R. Then, the first two lines in Eq. (6.60) remain unchanged. Then, furthermore, to employ Eq. (6.58), we need to show that there exists a channel $\mathcal{E} \in \mathrm{CPTP}(SE, SS'E')$ such that $\tilde{\mathcal{N}} = \mathcal{E} \circ \mathrm{Tr}_R$. For any input state ω_{SE} we define $\mathcal{E}$ as follows:

$$\begin{aligned}\mathcal{E}(\omega_{SE}) &\coloneqq \mathbb{I}_{S'} \otimes \mathrm{id}_S \otimes \mathcal{R}(\omega_{SE}) \\ &= \mathbb{I}_{S'} \otimes \mathrm{id}_S \otimes \mathcal{R} \circ \mathrm{Tr}_R(\omega_{SRE}) \\ &= \mathbb{I}_{S'} \otimes \mathrm{Tr}_{S'} \circ \tilde{\mathcal{M}}(\omega_{SRE}) \\ &= \tilde{\mathcal{N}}(\omega_{SRE}),\end{aligned} \tag{6.63}$$

where the third line follows from the non-signaling assumption on $\mathcal{M}$. Applying the chain rule in Eq. (6.56) using the first two terms in Eq. (6.60), and (6.62) we have:

$$H^{\uparrow}_{\alpha}(SS'|E')_{\mathcal{M}(\rho)} \geq H^{\uparrow}_{\alpha}(S|E)_{\rho} + \inf_{\omega \in \mathrm{S}_{=}(RE\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathrm{Tr}_{R'} \circ \mathcal{M}(\omega)}, \tag{6.64}$$

where we used $H^{\uparrow}_{\alpha} \geq H_{\alpha}$. □

With the above chain rule in place, we can now leverage the results from the previous section to incorporate a leakage process into the entropy accumulation framework. Qualitatively, the idea is that prior to each application of the channel $\mathcal{N}$, a leakage channel $\mathcal{L}$ transmits a register from memory to the side-information register, thereby modeling information leakage. By applying the chain rules developed earlier in Section 6.4, we can account for this leakage by introducing a penalty term, typically expressed as a mutual information or conditional mutual information involving the leakage register L.

Theorem 6.4.1. *Let $\rho \in \mathrm{S}_{=}(SRE)$, $\mathcal{L} \in \mathrm{CPTP}(RE, L\widehat{LRE})$ and $\mathcal{N} \in \mathrm{CPTP}(L\widehat{LRE}, S'R'E')$ be channels, such that there exists a channel $\mathcal{R} \in \mathrm{CPTP}(L\widehat{E}, E')$, where $\mathrm{Tr}_{S'R'} \circ \mathcal{N} = \mathcal{R} \circ \mathrm{Tr}_{\widehat{R}}$ (i.e., $\mathcal{N}$ is non-signaling from $\widehat{R}$ to E'). Let $\widetilde{E}$ be a purifying register for $L\widehat{LRE}$. Then, for any $\alpha \in (1, 2)$, letting $\widehat{\alpha} \coloneqq 1/(2-\alpha)$ we have:*

$$H^{\uparrow}_{\alpha}(SS'|E')_{\mathcal{N} \circ \mathcal{L}(\rho)} \geq H^{\uparrow}_{\alpha}(S|E)_{\rho} - \sup_{\omega \in \mathrm{S}_{=}(SRE)} I^{\downarrow,\mathrm{diff}}_{\alpha}(S; L|\widehat{E})_{\mathcal{L}(\omega)} + \inf_{\omega \in \mathrm{S}_{=}(L\widehat{LRE}\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathcal{N}(\omega)}. \tag{6.65}$$

Furthermore, if the registers $E, \widehat{E}$ are isomorphic and we restrict $\mathcal{L}$ to act non-trivially only on register R (formally: there exists a channel $\widetilde{\mathcal{L}} \in \mathrm{CPTP}(R, L\widehat{R})$ such that $\mathcal{L} = \widetilde{\mathcal{L}} \otimes \mathrm{id}_{E \to E}$),

then we have:

$$H_\alpha^\uparrow(SS'|E')_{\mathcal{N}\circ\mathcal{L}(\rho)} \geq H_\alpha^\uparrow(S|E)_\rho - \sup_{\omega\in \mathrm{S}_=(SRE)} I_\alpha^\downarrow(S\widehat{E};L)_{\mathcal{L}(\omega)} + \inf_{\omega\in \mathrm{S}_=(L\widehat{R}\widetilde{E}\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathcal{N}(\omega)}$$
$$= H_\alpha^\uparrow(S|E)_\rho - \sup_{\omega\in \mathrm{S}_=(\widetilde{Z}R)} I_\alpha^\downarrow(\widetilde{Z};L)_{\widetilde{\mathcal{L}}(\omega)} + \inf_{\omega\in \mathrm{S}_=(L\widehat{R}\widetilde{E}\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathcal{N}(\omega)}, \tag{6.66}$$

where $\widetilde{Z}$ is a purifying register for R.

Proof. Since $\mathcal{N}$ is non-signaling from $\widehat{R}$ to E', by applying Lemma 6.4.2, we have

$$\begin{aligned} H_\alpha^\uparrow(SS'|E')_{\mathcal{N}\circ\mathcal{L}(\rho)} &\geq H_\alpha^\uparrow(S|L\widehat{E})_{\mathcal{L}(\rho)} + \inf_{\omega\in \mathrm{S}_=(L\widehat{R}\widetilde{E}\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathcal{N}(\omega)} \\ &\geq H_\alpha^\uparrow(S|\widehat{E})_{\mathcal{L}(\rho)} - \sup_{\omega\in \mathrm{S}_=(SRE)} I_\alpha^{\downarrow,\mathrm{diff}}(S;L|\widehat{E})_{\mathcal{L}(\omega)} + \inf_{\omega\in \mathrm{S}_=(L\widehat{R}\widetilde{E}\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathcal{N}(\omega)} \\ &\geq H_\alpha^\uparrow(S|E)_\rho - \sup_{\omega\in \mathrm{S}_=(SRE)} I_\alpha^{\downarrow,\mathrm{diff}}(S;L|\widehat{E})_{\mathcal{L}(\omega)} + \inf_{\omega\in \mathrm{S}_=(L\widehat{R}\widetilde{E}\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathcal{N}(\omega)}, \end{aligned} \tag{6.67}$$

where the first line follows from Eq. (6.58), the second line is an application of Corollary 6.4.2, and the last line holds by data-processing (Theorem 2.2.1). To prove Eq. (6.66), we use Corollary 6.4.1:

$$\begin{aligned} H_\alpha^\uparrow(SS'|E')_{\mathcal{N}\circ\mathcal{L}(\rho)} &\geq H_\alpha^\uparrow(S|L\widehat{E})_{\mathcal{L}(\rho)} + \inf_{\omega\in \mathrm{S}_=(L\widehat{R}\widetilde{E}\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathcal{N}(\omega)} \\ &\geq H_\alpha^\uparrow(S|E)_\rho - \sup_{\omega\in \mathrm{S}_=(SRE)} I_\alpha^\downarrow(SE;L)_{\mathcal{L}(\omega)} + \inf_{\omega\in \mathrm{S}_=(L\widehat{R}\widetilde{E}\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathcal{N}(\omega)}, \end{aligned} \tag{6.68}$$

where the first line holds by Eq. (6.58), and the second line follows from Eq. (6.38) together with the fact that $\mathcal{L}$ essentially acts as identity on E. □

6.4.3 Incorporating generalized entropy accumulation with testing

In the previous section, we incorporated the leakage process into the GEAT framework. In this section, we show that this construction is compatible with the full GEAT-with-testing procedure. This is important, as testing is required in QKD scenarios. In particular, we

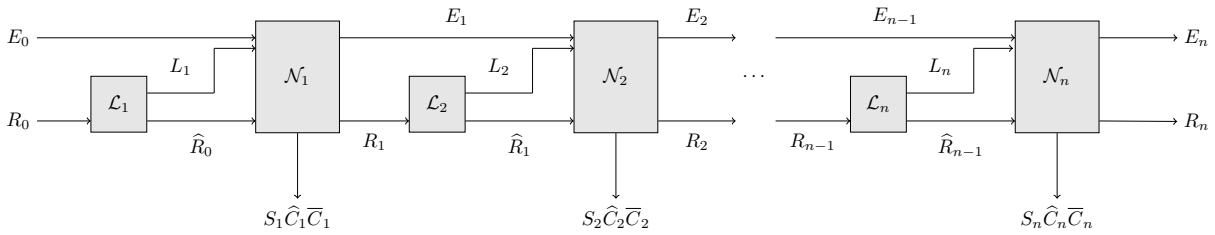


Figure 6.1: Schematic setup of a sequence of GEATTL channels. The registers E_i are the adversary's side information, L_i represent some "leaked" information registers in each round, R_i are memory registers stored between rounds without being available to the adversary except via the leakage, S_i can be seen as the secret registers (possibly quantum), and the classical registers $\overline{C}_i$, $\widehat{C}_i$ respectively contain secret and public information that will be used to estimate the accumulated entropy. The channels $\mathcal{L}_i$ describe the leakage processes, while the channels $\mathcal{N}_i$ represent all other actions in each round that satisfy a non-signaling condition from the memory registers to the side-information registers as Eq. (6.69).

demonstrate that the result remains compatible with the improved statistical analysis presented in Chapter 4.

Similar to Definition 4.2.1, we first define the sequence of GEAT-with-testing-and-leakage (GEATTL) channels (Fig 6.1).

Definition 6.4.1. $\{\mathcal{M}_j\}_{j=1}^n$ is called a **sequence of GEAT-with-testing-and-leakage (GEATTL) channels** if each $\mathcal{M}_j \in \mathrm{CPTP}(R_{j-1}E_{j-1}, S_jR_jE_j\overline{C}_j\widehat{C}_j)$ is such that the output registers $\overline{C}_j\widehat{C}_j$ are always classical (for any input states), and it can be written as $\mathcal{M}_j = \mathcal{N}_j \circ \mathcal{L}_j$, where $\mathcal{L}_j \in \mathrm{CPTP}(R_{j-1}, \widehat{R}_{j-1}L_j)$ and $\mathcal{N}_j \in \mathrm{CPTP}(\widehat{R}_{j-1}L_jE_{j-1}, S_jR_jE_j\overline{C}_j\widehat{C}_j)$, with $\mathcal{N}_j$ satisfying the following **non-signaling (NS) condition**:

$$\exists \text{ a channel } \mathcal{R}_j \in \mathrm{CPTP}(L_jE_{j-1}, E_j\widehat{C}_j) \text{ such that } \mathrm{Tr}_{S_jR_j\overline{C}_j} \circ \mathcal{N}_j = \mathcal{R}_j \circ \mathrm{Tr}_{\widehat{R}_{j-1}}. \quad (6.69)$$

We refer to the channels $\mathcal{L}_j$ as the **leakage channels**. If a state $\rho \in \mathrm{S}_=(S_1^n\overline{C}_1^n\widehat{C}_1^nE_nR_n)$ is of the form $\rho = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1[\omega^0]$ (leaving some identity channels implicit) for some initial state $\omega^0 \in \mathrm{S}_=(R_0E_0)$, we say it is **generated by the sequence of GEATTL channels** $\{\mathcal{M}_j\}_{j=1}^n$.

Theorem 6.4.2 (QES-entropy accumulation with leakage)**.** *Let ρ be a state generated by a sequence of GEATTL channels $\{\mathcal{M}_j\}_{j=1}^n$ (Definition 6.4.1), so they have the form*

$\mathcal{M}_j = \mathcal{N}_j \circ \mathcal{L}_j$. *For each* j, *suppose that for every value* $\bar{c}_1^{j-1}\hat{c}_1^{j-1}$, *we have a QES* $f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}$ *on registers* $\overline{C}_j\widehat{C}_j$. *Define the following QES on* $\overline{C}_1^n\widehat{C}_1^n$:

$$f_{\text{full}}(\bar{c}_1^n\hat{c}_1^n) \coloneqq \sum_{j=1}^{n} f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j). \tag{6.70}$$

Take any $\alpha \in (1,2)$ *and let* $\widehat{\alpha} = 1/(2-\alpha)$. *Then we have*

$$\begin{aligned} H_{2-\frac{1}{\alpha}}^{f_{\text{full}}}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_\rho &\geq \sum_j \left(\min_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}} \kappa_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}} - \xi_j \right), \\ \text{where} \quad \kappa_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}} &\coloneqq \inf_{\nu \in \Sigma_j^{\mathcal{N}}} H_{\widehat{\alpha}}^{f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}}(S_j\overline{C}_j|\widehat{C}_j E_j \widetilde{E})_\nu, \\ \xi_j &\coloneqq \sup_{\nu \in \Sigma_j^{\mathcal{L}}} I_\alpha^{\downarrow}(\widetilde{Z}; L_j)_\nu, \end{aligned} \tag{6.71}$$

where $\Sigma_j^{\mathcal{N}}$ *denotes the set of all states of the form* $\mathcal{N}_j\left[\omega_{\widehat{R}_{j-1}L_jE_{j-1}\widetilde{E}}\right]$ *for some initial state* $\omega \in \mathrm{S}_=(\widehat{R}_{j-1}L_jE_{j-1}\widetilde{E})$, *with* $\widetilde{E}$ *being a register of large enough dimension to serve as a purifying register for any of the* $\widehat{R}_{j-1}L_jE_{j-1}$ *registers, and* $\Sigma_j^{\mathcal{L}}$ *denotes the set of all states of the form* $\mathcal{L}_j\left[\omega''_{R_{j-1}\widetilde{Z}}\right]$ *for some initial state* $\omega'' \in \mathrm{S}_=(R_{j-1}\widetilde{Z})$, *with* $\widetilde{Z}$ *being a register of large enough dimension to serve as a purifying register for any of the* $\widehat{R}_{j-1}$ *registers.*

Consequently, if we instead define the following "normalized" QES on $\overline{C}_1^n\widehat{C}_1^n$:

$$\hat{f}_{\text{full}}(\bar{c}_1^n\hat{c}_1^n) \coloneqq \sum_{j=1}^{n} \hat{f}_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j), \quad \text{where } \hat{f}_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j) \coloneqq f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j) + \kappa_{\bar{c}_1^{j-1}\hat{c}_1^{j-1}} - \xi_j, \tag{6.72}$$

then

$$H_{2-\frac{1}{\alpha}}^{\hat{f}_{\text{full}}}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_\rho \geq 0. \tag{6.73}$$

Proof. Let $M > 0$ be any value such that $M - f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j) > \frac{M}{2} > 0$ for all the $f_{|\bar{c}_1^{j-1}\hat{c}_1^{j-1}}(\bar{c}_j\hat{c}_j)$ in the theorem statement (for all j). Define a read-and-prepare channel $\mathcal{D}_j \in \mathrm{CPTP}(\overline{C}_1^j\widehat{C}_1^j, \overline{C}_1^j\widehat{C}_1^j D_j)$ for each j, the same way as the proof Theorem 4.3.1. Let $\mathcal{E}_j \in \mathrm{CPTP}(\widehat{R}_{j-1}L_jE_{j-1}\overline{C}_1^{j-1}\widehat{C}_1^{j-1}, D_j\overline{\overline{C}}_jS_jR_jE_j\overline{C}_1^j\widehat{C}_1^j)$ denote a channel that does the following:

1. Apply $\mathcal{N}_j \otimes \mathcal{P}_j$, where $\mathcal{P}_j$ is a pinching channel on $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$ (in its classical basis).

2. Copy the classical register $\overline{C}_j$ onto another classical register $\overline{\overline{C}}_j$.

3. Generate a D_j register by applying the above read-and-prepare channel $\mathcal{D}_j$ on $\overline{C}_1^j\widehat{C}_1^j$.

As argued in the proof of Theorem 4.3.1, these channels $\mathcal{E}_j$ inherit the NS condition of the $\mathcal{N}_j$ channels, in that they are non-signaling from $\widehat{R}_{j-1}\overline{C}_1^{j-1}$ to $E_j\widehat{C}_j$. Also, the state $\mathcal{E}_n \circ \mathcal{L}_n \circ \ldots \mathcal{E}_1 \circ \mathcal{L}_1[\omega^0]$ is a valid extension of the state ρ in the theorem statement, so we can use ρ to refer to this state as well without ambiguity. By now applying Theorem 6.4.1 to this state (since the $\mathcal{E}_j$ channels satisfy the appropriate NS condition), we obtain:

$$H_\alpha^\uparrow(D_1^n S_1^n \overline{\overline{C}}_1^n | \widehat{C}_1^n E_n)_\rho \geq \sum_j \left(\inf_{\nu' \in \Sigma_j^{\mathcal{E}}} H_{\widehat{\alpha}}(D_j S_j \overline{\overline{C}}_j | \widehat{C}_1^j E_j \widetilde{E})_{\nu'} - \sup_{\nu' \in \Sigma_j^{\mathcal{L}}} I_\alpha^\downarrow(\widetilde{Z}; L_j)_{\nu'} \right) \tag{6.74}$$

where $\Sigma_j^{\mathcal{E}}$ is the set of all states that could be produced by $\mathcal{E}_j$ acting on some initial state $\omega'_{\widehat{R}_{j-1}L_jE_{j-1}\overline{C}_1^{j-1}\widehat{C}_1^{j-1}\widetilde{E}}$, and $\Sigma_j^{\mathcal{L}}$ is the set of all states that could be produced by $\mathcal{L}_j$ for some initial state $\omega''_{R_{j-1}\widetilde{Z}}$. Furthermore, by the bound $H_{2-\frac{1}{\alpha}} \geq H_\alpha^\uparrow$ [118, Corollary 4], this immediately also gives

$$H_{2-\frac{1}{\alpha}}(D_1^n S_1^n \overline{\overline{C}}_1^n | \widehat{C}_1^n E_n)_\rho \geq \sum_j \left(\inf_{\nu' \in \Sigma_j^{\mathcal{E}}} H_{\widehat{\alpha}}(D_j S_j \overline{\overline{C}}_j | \widehat{C}_1^j E_j \widetilde{E})_{\nu'} - \sup_{\nu' \in \Sigma_j^{\mathcal{L}}} I_\alpha^\downarrow(\widetilde{Z}; L_j)_{\nu'} \right). \tag{6.75}$$

The claim then follows by applying the remaining arguments in the proof of Theorem 4.3.1 to relate $H_{2-\frac{1}{\alpha}}(D_1^n S_1^n \overline{\overline{C}}_1^n | \widehat{C}_1^n E_n)_\rho$ and $H_{\widehat{\alpha}}(D_j S_j \overline{\overline{C}}_j | \widehat{C}_1^j E_j \widetilde{E})_{\nu'}$ to $H_{2-\frac{1}{\alpha}}^{f_{\text{full}}}(S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_\rho$ and $H_{\widehat{\alpha}}^{f_{|\widehat{c}_1^{j-1}\widehat{c}_1^{j-1}}}(S_j \overline{C}_j | \widehat{C}_j E_j \widetilde{E})_\nu$ respectively. □

Since the leakage term is independent of the choice of the QES, all subsequent simplifications presented in Chapter 4 remain valid when using GEATTL channels, provided the leakage term is subtracted appropriately. In particular, we state the following theorem, which generalizes Theorem 6.4.3 to incorporate leakage.

Theorem 6.4.3. *Let ρ be a state generated by a sequence of GEATTL channels $\{\mathcal{M}_j\}_{j=1}^n$ (Definition 6.4.1), so they have the form $\mathcal{M}_j = \mathcal{N}_j \circ \mathcal{L}_j$. Suppose all the registers $\overline{C}_j$ (resp. $\widehat{C}_j$) are isomorphic to a single register $\overline{C}$ (resp. $\widehat{C}$) with alphabet $\overline{\mathcal{C}}$ (resp. $\widehat{\mathcal{C}}$). Take any $\alpha \in (1, 2)$ and let $\widehat{\alpha} = 1/(2-\alpha)$. Suppose furthermore that $\rho = p_\Omega \rho_{|\Omega} + (1 - p_\Omega)\rho_{|\overline{\Omega}}$*

for some $p_\Omega \in (0,1]$ *and normalized states* $\rho_{|\Omega}, \rho_{|\overline{\Omega}}$. *Let* S_Ω *be a convex set of probability distributions on the alphabet* $\overline{\mathcal{C}} \times \widehat{\mathcal{C}}$, *such that for all* $\bar{c}_1^n \hat{c}_1^n$ *with nonzero probability in* $\rho_{|\Omega}$, *the frequency distribution* $\mathrm{freq}_{\bar{c}_1^n \hat{c}_1^n}$ *lies in* S_Ω. *Then letting* $\boldsymbol{\sigma}_{\overline{C}\widehat{C}}$ *denote the distribution on* $\overline{C}\widehat{C}$ *induced by any state* $\sigma_{\overline{C}\widehat{C}}$, *we have*

$$H_\alpha^\uparrow(S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\rho_{|\Omega}} \geq n h_{\widehat{\alpha}} - \left(\sum_j \xi_j \right) - \frac{\alpha}{\alpha - 1} \log \frac{1}{p_\Omega},$$

$$\text{where} \quad h_{\widehat{\alpha}} = \inf_{\mathbf{q} \in S_\Omega} \inf_{\nu \in \Sigma} \left(\frac{1}{\widehat{\alpha} - 1} D\left(\mathbf{q} \middle\| \boldsymbol{\nu}_{\overline{C}\widehat{C}}\right) - \sum_{\bar{c}\hat{c} \in \mathrm{supp}(\boldsymbol{\nu}_{\overline{C}\widehat{C}})} q(\bar{c}\hat{c}) D_{\widehat{\alpha}} \left(\nu_{S_j E_j \widetilde{E} \wedge \bar{c}\hat{c}} \middle\| \mathbb{I}_{S_j} \otimes \nu_{E_j \widetilde{E} \wedge \hat{c}} \right) \right), \tag{6.76}$$

where ξ_j *is as defined in* (4.32), *and* Σ *denotes the convex range (Definition 4.4.1) of channels* $\{\mathcal{N}_j \otimes \mathrm{id}_{\widetilde{E}}\}_{j=1}^n$, *with* $\widetilde{E}$ *being a register of large enough dimension to serve as a purifying register for any of the* $\widehat{R}_{j-1} L_j E_{j-1}$ *registers.*

Proof. As mentioned above, the proof of this result follows by simply applying the same chain of arguments used to prove Theorem 4.4.1, as none of them further relied on the NS conditions. A small technical difference is that rather than starting from the final Theorem 6.4.2 statement here, in this case we would instead start from the intermediate bound (4.38) in the above proof, to avoid the change of Rényi parameter (the proof of Theorem 4.4.1 can work directly with $H_\alpha^\uparrow(D_1^n S_1^n \overline{\overline{C}}_1^n | \widehat{C}_1^n E_n)_\rho$ rather than $H_\alpha(D_1^n S_1^n \overline{\overline{C}}_1^n | \widehat{C}_1^n E_n)_\rho$, unlike the proof of Theorem 4.3.1). □

It is worth noting that the lower bounds introduced in Lemmas 4.4.1 and 4.4.2 remain valid in the presence of leakage. Furthermore, the convexity of the function $h_{\hat{\alpha}}$ continues to hold after the introduction of the purifying function. Therefore, all the benefits of those simplifications can be retained and applied in the current setting as well.

6.4.4 GEAT without non-signaling

In this section, we present a variation of the GEAT, modifying Lemma 6.4.2, in which the leakage channel does not send a leakage register to the side-information register, but instead acts directly on the memory register, and modifies it. If this modification is such

that the resulting mutual information penalty term remains bounded, then we can obtain non-trivial bounds without requiring the non-signaling condition on the channels[7].

Corollary 6.4.3. *Let $\rho \in \mathrm{S}_=(SRE)$, let $\mathcal{L} \in \mathrm{CPTP}(RE, \widehat{RE})$ and $\mathcal{N} \in \mathrm{CPTP}(\widehat{RE}, S'R'E')$ be channels. Let $\widetilde{E}$ be a purifying register for $\widehat{RE}$. Then for any $\alpha \in (1,2)$, letting $\widehat{\alpha} \coloneqq 1/(2-\alpha)$ we have:*

$$H_\alpha^\uparrow(SS'|E')_{\mathcal{N}\circ\mathcal{L}(\rho)} \geq H_\alpha^\uparrow(S|E)_\rho - \sup_{\omega \in \mathrm{S}_=(RSE)} I_\alpha^{\downarrow,\mathrm{diff}}(S;\widehat{R}|\widehat{E})_{\mathcal{L}(\omega)} + \inf_{\omega \in \mathrm{S}_=(\widehat{RE}\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathcal{N}(\omega)}. \tag{6.77}$$

Furthermore, if the registers $E, \widehat{E}$ are isomorphic and we restrict $\mathcal{L}$ to act non-trivially only on register R (formally: there exists a channel $\widetilde{\mathcal{L}} \in \mathrm{CPTP}(R, \widehat{R})$ such that $\mathcal{L} = \widetilde{\mathcal{L}} \otimes \mathrm{id}_{E\to\widehat{E}}$), then we have:

$$\begin{aligned} H_\alpha^\uparrow(SS'|E')_{\mathcal{N}\circ\mathcal{L}(\rho)} &\geq H_\alpha^\uparrow(S|E)_\rho - \sup_{\omega \in \mathrm{S}_=(RSE)} I_\alpha^{\downarrow}(SE;\widehat{R})_{\mathcal{L}(\omega)} + \inf_{\omega \in \mathrm{S}_=(\widehat{RE}\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathcal{N}(\omega)} \\ &= H_\alpha^\uparrow(S|E)_\rho - \sup_{\omega \in \mathrm{S}_=(\widetilde{Z}R)} I_\alpha^{\downarrow}(\widetilde{Z};\widehat{R})_{\widetilde{\mathcal{L}}(\omega)} + \inf_{\omega \in \mathrm{S}_=(\widehat{RE}\widetilde{E})} H_{\widehat{\alpha}}(S'|E'\widetilde{E})_{\mathcal{N}(\omega)}. \end{aligned} \tag{6.78}$$

where $\widetilde{Z}$ is a purifying register for R.

Proof. The proof is essentially the same as Theorem 6.4.1, by replacing the L register with $\widehat{R}$, and noting that since we no longer have a non-signaling condition, we must use Eq. (6.57) instead. □

We note that, using a similar analysis, one can also derive a version of the GEAT with testing analogous to Theorems 6.4.2 and 6.4.3, applicable to the case where the channels no longer satisfy the non-signaling condition.

6.5 Example applications

The results obtained so far indicate that if we have a sequence of GEAT-style channels $\{\mathcal{N}_i\}_{i=1}^N$ along with leakage channels $\{\mathcal{L}_i\}_{i=1}^N$, where each leakage channel acts prior to the corresponding $\mathcal{N}_i$ and leaks to the side-information register, then the effect of such leakage

[7]It should be noted, however, that there is currently no clear application for this modification.

can be compensated by subtracting the corresponding mutual information terms—summed over all n channels $\mathcal{L}_j$. For simplicity, let us focus on the bound in Eq. (6.66). In this case, the only quantity that needs to be computed per round (omitting the subscript j for brevity) is the following:

$$\sup_{\omega \in \mathrm{S}_{=}(\widetilde{Z}R)} I_\alpha^{\downarrow}(\widetilde{Z}; L)_{\widetilde{\mathcal{L}}(\omega)}. \tag{6.79}$$

In other words, it suffices to upper bound the Rényi mutual information between the channel output and an arbitrary purification of its input. In the following, we outline several device-imperfection scenarios in which such a leakage model is relevant and can be meaningfully applied.

6.5.1 Physical leakage

The most natural application of this model is if we have an actual physical device that leaks information to the adversary in each round. For example, this could be a photonic system that we constrain by requiring it to have a large vacuum component, as considered in [75, 100] for device-dependent protocols and in [112] for device-independent protocols. In particular, this resolves an issue regarding a "restricted adaptiveness" condition imposed on Eve in [112] — under the model we consider here, Eve is free to modify her attack in every round based on the leakage registers gathered from previous rounds. Hence, our result allows us to handle such leakage mechanisms even against the most general adaptive attacks.

6.5.2 Source correlation

Here, we present a less straightforward application of the leakage model: source correlation. We first summarize the key ideas. Essentially, within the state generation process in a prepare-and-measure protocol, Alice generates a secret classical register and, through some source device, prepares a state intended to be sent to Bob. This procedure depends only on the value of that secret register for the particular round. The GEAT framework can already handle this scenario, as the procedure satisfies the NS condition. However, in more realistic scenarios, the source device might be imperfect, and the state it prepares might not only depend on the value of the classical register in that round but also on previous rounds. Here, we propose a way to handle this situation within the GEAT framework.

More specifically, consider a model where, after each round, the source keeps a memory R for the next round. Then, every time Alice uses the source device to prepare a state, it first prepares an ideal state on some register A', then applies a channel $\widetilde{\mathcal{L}} \in \mathrm{CPTP}(R, L)$ that extracts some information from memory, and finally applies another channel $\widehat{\mathcal{L}} \in \mathrm{CPTP}(A'L, A')$ to modify the ideal state.

We now present an explicit example of a source correlation scenario that fits into this model. Suppose in each round, the source behaves as follows: with probability $1 - \delta_{\text{leak}}$, the source prepares an ideal state on A', otherwise it modifies the ideal state on A' with some channel $\widehat{\mathcal{L}} \in \mathrm{CPTP}(A'L, A')$.

To incorporate this example into the model we described above, let us consider a register L that has one additional dimension to R, and on the subspace excluding that additional dimension, it is isomorphic to R. Let the basis of the extra dimension be $|\perp\rangle_L$. Then, the source first applies a channel $\widetilde{\mathcal{L}} \in \mathrm{CPTP}(R; L)$ such that with probability $1-\delta_{\text{leak}}$ it ignores the input and outputs $|\perp\rangle\langle\perp|_L$, otherwise, it embeds the input into the output register L.

Next, it prepares an ideal state on A'. Then, it performs a projective measurement on L with elements $\{\Pi_\perp, \mathbb{I} - \Pi_\perp\}$. Finally, conditioned on the outcome, it either leaves the prepared ideal state on A' untouched, or acts on it with some channel $\widehat{\mathcal{L}} \in \mathrm{CPTP}(A'L, A')$.

6.5.3 Bounding the mutual information

We now explain how to bound the mutual information term in (6.79). In particular, we focus on the example model we described in the previous part. We begin by noting that in the above example, we can more explicitly write down the channel model $\widetilde{\mathcal{L}} \in \mathrm{CPTP}(R, L)$ as:

$$\widetilde{\mathcal{L}} = (1 - \delta_{\text{leak}})\widetilde{\mathcal{L}}^{\perp} + \delta_{\text{leak}}\widetilde{\mathcal{L}}^{\text{leak}}, \tag{6.80}$$

where $\widetilde{\mathcal{L}}^{\perp}, \widetilde{\mathcal{L}}^{\text{leak}} \in \mathrm{CPTP}(R, L)$ are some channels, such that $\widetilde{\mathcal{L}}^{\perp}$ only outputs a fixed state $|\perp\rangle\langle\perp|_L$. Let $\nu_{\widetilde{Z}L} \coloneqq \widetilde{\mathcal{L}}(\omega_{\widetilde{Z}R})$ for some input state $\omega_{\widetilde{Z}R}$. Moreover, let $\nu'_{\widetilde{Z}L} \coloneqq \widetilde{\mathcal{L}}^{\text{leak}}(\omega_{\widetilde{Z}R})$, then we have:

$$\nu_{\widetilde{Z}L} = (1 - \delta_{\text{leak}})\nu_{\widetilde{Z}} \otimes |\perp\rangle\langle\perp|_L + \delta_{\text{leak}}\nu'_{\widetilde{Z}L}, \text{ where } \langle\perp|\nu'_L|\perp\rangle = 0 \text{ and } \nu'_{\widetilde{Z}} = \nu_{\widetilde{Z}}, \tag{6.81}$$

where the channels act as identity on the $\widetilde{Z}$ register. We can now bound the mutual information term as follows:

$$\begin{aligned}
I^{\downarrow}_{\alpha}(\widetilde{Z};L)_{\nu} &\leq I_{\alpha}(\widetilde{Z}:L)_{\nu} \\
&= D_{\alpha}((1-\delta_{\text{leak}})\nu_{\widetilde{Z}}\otimes|\perp\rangle\langle\perp|_{L}+\delta_{\text{leak}}\nu'_{\widetilde{Z}L}||\nu_{\widetilde{Z}}\otimes((1-\delta_{\text{leak}})|\perp\rangle\langle\perp|_{L}+\delta_{\text{leak}}\nu'_{L})) \\
&= \frac{1}{\alpha-1}\log\left((1-\delta_{\text{leak}})2^{(\alpha-1)(0)}+\delta_{\text{leak}}2^{(\alpha-1)D_{\alpha}(\nu'_{\widetilde{Z}L}||\nu_{\widetilde{Z}}\otimes\nu'_{L})}\right) \\
&= \frac{1}{\alpha-1}\log\left((1-\delta_{\text{leak}})+\delta_{\text{leak}}2^{(\alpha-1)D_{\alpha}(\nu'_{\widetilde{Z}L}||\nu'_{\widetilde{Z}}\otimes\nu'_{L})}\right) \\
&\leq \frac{1}{\alpha-1}\log\left((1-\delta_{\text{leak}})+\delta_{\text{leak}}2^{(\alpha-1)\kappa\log\min\{\dim(\widetilde{Z}),\dim(L)\}}\right) \quad \forall\alpha\in(1,3/2),
\end{aligned} \tag{6.82}$$

where $\kappa \coloneqq 1$ if either $\widetilde{Z}$ or L is classical, and $\kappa \coloneqq 2$ otherwise. Here, the third line follows from (2.39) (and the fact that all the terms with ν' have disjoint support from the terms with $|\perp\rangle\langle\perp|_{L}$), the fourth line uses the critical property $\nu'_{\widetilde{Z}} = \nu_{\widetilde{Z}}$, and the last line is from [37, proof of Corollary III.5].[8]

The bound we obtained above has some appealing characteristics. Firstly, in the limit of $\delta_{\text{leak}} \to 0$, it becomes zero, as expected. Moreover, in the limit of $\alpha \to 1$, it converges to

$$\delta_{\text{leak}}\kappa\log\min\{\dim(\widetilde{Z}),\dim(L)\}, \tag{6.83}$$

which is exactly the bound one would obtain had the von Neumann mutual information $I(\widetilde{Z}:L)_{\nu}$ been considered in the first place.

We plot the bound in Eq. (6.79) for the example we considered in this part in Fig. 6.2. We note that, as stated, the leakage analysis in Sec. 6.4 is such that the only change to the key rate occurs via the mutual information term in (6.79). Therefore, analyzing that term alone essentially tells us how much the key rates decrease when considering leakage in the GEAT framework.

Thus, the plots in Fig. 6.2 precisely show the amount by which the key rate decreases. Moreover, we note that, given the analysis done in [65], the GEAT-based security proofs typically take $\alpha < 1.1$ in standard parameter regimes. Hence, this choice of α provides a conservative upper bound on how much key rates decrease in typical applications.

Alternatively, one can use continuity bounds in [37, 41] to first convert the Rényi mutual information to the von Neumann mutual information as $I^{\downarrow}_{\alpha}(\widetilde{Z};L)_{\nu} \leq I(\widetilde{Z}:L)_{\nu} + O(\alpha-1)$,

[8] Note that in our example $\dim(\widetilde{Z}) < \dim(L)$.

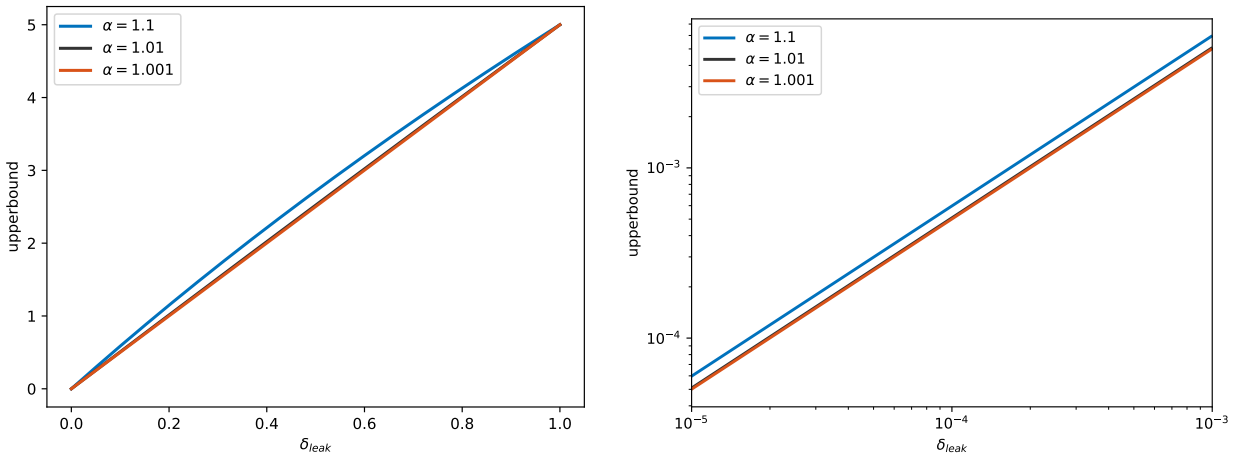


Figure 6.2: Left: the upper bound on Rényi mutual information in Eq (6.82) versus leakage probability δ_{leak}, for three different values of Rényi parameter $\alpha \in \{1.1, 1.01, 1.001\}$, when the memory register is taken to be a classical register of 5 bits (so $\dim(\widetilde{Z}) = \dim(R) = 2^5$ and $\kappa = 1$. Right: the same plot focused on the small-δ_{leak} regime, and on log scales for the axes. A plot of the von Neumann case is omitted because it was visually indistinguishable from the $\alpha = 1.001$ case; it can be seen from the plots that the $\alpha = 1.01$ case is also already very close to this case. Note that existing GEAT security proofs usually choose α significantly smaller than 1.1 at typical protocol sample sizes, and thus the $\alpha = 1.1$ plots in the above figures serve as a conservative upper bound on how much the key rates decrease when applying our techniques to account for leakage.

and then bound the von Neumann mutual information. However, this approach would lead to worse bounds, since even for $\delta_{\text{leak}} \to 0$, the leakage contribution is not zero but rather equal to the $O(\alpha - 1)$ term—unless we are in the large-n regime, in which case that correction factor approaches zero.

6.6 Conclusion and open questions

Overall, in this chapter, we have derived a variety of chain rules involving mutual information quantities, including both one-shot chain rules and Rényi chain rules suitable for repeated application over a sequence of channels. In particular, the latter yield modifications of entropy accumulation that can accommodate some leakage, which are fairly straightforward to incorporate into security proofs. These results should be useful in rigorously quantifying the effects of device imperfections in protocol implementations.

Moreover, they are also compatible with the improved analysis in Chapter 4.

As for future directions, it would be interesting to explore what other explicit imperfections can be addressed using the analysis in this chapter. In particular, the examples we analyzed here are fairly simple and straightforward; one can investigate what other forms of imperfections—either on the source side or the detector side—can be studied using the techniques presented here. Moreover, it would also be interesting to construct a more direct definition of the Rényi CQMI and provide a tighter bound than the one in Corollary 6.2.1. Such a development could also potentially be used to improve the information bounding theorem in 6.3.1.

Chapter 7

Concluding remarks

In this thesis, we studied the security of QKD protocols against coherent attacks by analyzing and extending EAT-style security proofs. In particular, employing the modular security framework introduced in [104], we focused on bounding a suitable conditional entropy of the states generated at the end of the protocol. Given the importance of EAT-style proofs in bounding such a conditional entropy, we analyzed them for a class of QKD protocols and addressed their limitations.

To this end, we developed a flexible framework to analyze PM-QKD and decoy-state protocols under the GEAT framework [89, 90]. However, as mentioned, the GEAT framework has two main drawbacks. First, statistical analysis is suboptimal and involves some extremely challenging steps. Second, the application of the GEAT in PM-QKD suffers from a restriction on the repetition rate of the emitted signals, which limits its practicality.

To address these issues, we extended the connection between the QPE framework [136] and the GEAT and developed an improved statistical analysis that yields extremely tight bounds. Moreover, we introduced a variant of EAT, termed MEAT, that exploits the source-replacement technique to convert the PM-QKD protocol of interest into an equivalent EB-QKD such that it is "compatible" with an EAT-style proof. This allows us to effectively remove the repetition-rate restriction and broaden the practical applicability of the framework.

In addition to these improvements, we investigated scenarios where there are potential imperfections and correlations in the source and measurement devices. In particular, we introduced various chain rules to analyze such cases. For instance, we derived a chain rule between smooth min-entropy and smooth max-information, which improves over previous chain rules for characterizing one-shot information leakage caused by an additional

conditioning register. In addition, we developed a variation of GEAT that can accommodate imperfections in both the source and the measurement devices.

Moving forward, it is important to see if the analysis in Chapter 4 can be further improved, as it would yield better key rates. However, it seems that there are not many steps left to improve (see Sec. 4.7).

On a different front, in practical QKD, imperfections in devices are inevitable. A challenging situation is when device imperfections exhibit memory effects. Thus, it is of significant importance to adapt the MEAT framework in Chapter 5 to accommodate such imperfections. This can be done by modifying the framework to accommodate a secret memory register (see Sec. 5.6). This would allow us to construct a unified framework that simultaneously handles imperfections while benefiting from the tight and broadly applicable framework of marginal-constrained entropy accumulation.

On a more protocol-oriented level, it would be interesting to explore which additional QKD protocols can be analyzed using the frameworks developed in this thesis. Furthermore, it remains an open direction to investigate what models for imperfections can be studied with the framework we introduced here in Chapter 6.

On the information-theoretic side, it would be an interesting direction of exploration to investigate whether the techniques we used in this thesis, especially those used in Chapter 5, to prove the strong additivity of channel conditional entropy, could shed light on the additivity problems in other areas of quantum information theory. In another direction, one can explore obtaining tighter results for the treatment we proposed in Chapter 6 to handle imperfections. This can be done by improving the chain rules presented in that chapter. However, such an improvement would potentially require a satisfactory definition of Rényi conditional mutual information, which would also be of interest in other areas of quantum information theory.

We also highlight that the results in this thesis mostly involve bounding a suitable form of Rényi conditional entropy, which is motivated by Rényi privacy amplification [42]. An important question is how tight such a privacy amplification theorem is. While [107] provides a converse bound, there is still a gap between achievability and the converse bound, and it would be of interest to close this gap.

References

[1] C. Adami and N. J. Cerf. von Neumann capacity of noisy quantum channels. *Phys. Rev. A*, 56:3470–3483, 1997.

[2] Anurag Anshu, Shima Bab Hadiashar, Rahul Jain, Ashwin Nayak, and Dave Touchette. One-Shot Quantum State Redistribution and Quantum Markov Chains. *IEEE Transactions on Information Theory*, 69(9):5788–5804, 2023.

[3] Anurag Anshu, Vamsi Krishna Devabathini, and Rahul Jain. Quantum Communication Using Coherent Rejection Sampling. *Phys. Rev. Lett.*, 119(12):120506, September 2017.

[4] Anurag Anshu, Rahul Jain, and Alexander Streltsov. Quantum state redistribution with local coherence. *arXiv:1804.04915*, April 2018.

[5] Anurag Anshu, Rahul Jain, and Naqueeb Ahmad Warsi. A One-Shot Achievability Result for Quantum State Redistribution. *IEEE Transactions on Information Theory*, 64(3):1425–1435, 2018.

[6] Rotem Arnon-Friedman, Renato Renner, and Thomas Vidick. Simple and Tight Device-Independent Security Proofs. *SIAM Journal on Computing*, 48(1):181–225, January 2019.

[7] Amir Arqand, Thomas A. Hahn, and Ernest Y.-Z. Tan. Generalized Rényi entropy accumulation theorem and generalized quantum probability estimation. *arXiv:2405.05912v4*, 2024.

[8] Amir Arqand, Tony Metger, and Ernest Y.-Z. Tan. Mutual information chain rules for security proofs robust against device imperfections. *arXiv.2407.20396*, July 2024.

[9] Amir Arqand and Ernest Y.-Z. Tan. Marginal-constrained entropy accumulation theorem. *arXiv.2502.02563*, February 2025.

[10] Normand J. Beaudry, Tobias Moroder, and Norbert Lütkenhaus. Squashing Models for Optical Measurements in Quantum Communication. *Physical Review Letters*, 101(9):093601, August 2008.

[11] Salman Beigi. Sandwiched Rényi divergence satisfies data processing inequality. *Journal of Mathematical Physics*, 54(12), December 2013.

[12] J. S. Bell. On the einstein podolsky rosen paradox. *Physics*, 1(3):195–200, November 1964.

[13] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of International Conference on Computers, Systems and Signal Processing, Bangalore, India, 1984*, page 175. IEEE Press, New York, 1984.

[14] Charles H. Bennett, François Bessette, Gilles Brassard, Louis Salvail, and John Smolin. Experimental Quantum Cryptography. *J. Cryptology*, 5(1):3–28, January 1992.

[15] Mario Berta, Matthias Christandl, and Renato Renner. The Quantum Reverse Shannon Theorem Based on One-Shot Information Theory. *Commun. Math. Phys.*, 306(3):579–615, September 2011.

[16] Mario Berta, Matthias Christandl, and Dave Touchette. Smooth Entropy Bounds on One-Shot Quantum State Redistribution. *IEEE Transactions on Information Theory*, 62(3):1425–1439, 2016.

[17] Mario Berta, Kaushik P. Seshadreesan, and Mark M. Wilde. Rényi generalizations of the conditional quantum mutual information. *Journal of Mathematical Physics*, 56(2):022205, 02 2015.

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

[19] Gilles Brassard, Norbert Lütkenhaus, Tal Mor, and Barry C. Sanders. Limitations on Practical Quantum Cryptography. *Physical Review Letters*, 85(6):1330–1333, August 2000.

[20] Nicolas J. Cerf and Chris Adami. Information theory of quantum entanglement and measurement. *Physica D: Nonlinear Phenomena*, 120(1):62–81, 1998. Proceedings of the Fourth Workshop on Physics and Consumption.

[21] Matthias Christandl, Robert König, and Renato Renner. Postselection Technique for Quantum Channels with Applications to Quantum Cryptography. *Physical Review Letters*, 102:020504, January 2009.

[22] Nikola Ciganović, Normand J. Beaudry, and Renato Renner. Smooth Max-Information as One-Shot Generalization for Mutual Information. *IEEE Transactions on Information Theory*, 60(3):1573–1581, 2014.

[23] Tom Cooney, Milán Mosonyi, and Mark M. Wilde. Strong Converse Exponents for a Quantum Channel Discrimination Problem and Quantum-Feedback-Assisted Communication. *Communications in Mathematical Physics*, 344(3):797–829, June 2016.

[24] Guillermo Currás-Lorenzo, Margarida Pereira, Go Kato, Marcos Curty, and Kiyoshi Tamaki. A security framework for quantum key distribution implementations. *arXiv.2305.05930*, May 2023.

[25] Guillermo Currás-Lorenzo, Shlok Nahar, Norbert Lütkenhaus, Kiyoshi Tamaki, and Marcos Curty. Security of quantum key distribution with imperfect phase randomisation. *Quantum Science and Technology*, 9(1):015025, dec 2023.

[26] Marcos Curty, Maciej Lewenstein, and Norbert Lütkenhaus. Entanglement as a Precondition for Secure Quantum Key Distribution. *Phys. Rev. Lett.*, 92(21):217903, May 2004.

[27] Siddhartha Das, Kaumudibikash Goswami, and Vivek Pandey. Conditional entropy and information of quantum processes, 2024.

[28] Nilanjana Datta. Min- and Max-Relative Entropies and a New Entanglement Monotone. *IEEE Transactions on Information Theory*, 55(6):2816–2826, 2009.

[29] Nilanjana Datta and Min-Hsiu Hsieh. The apex of the family tree of protocols: optimal rates and resource inequalities. *New Journal of Physics*, 13(9):093042, sep 2011.

[30] Nilanjana Datta, Min-Hsiu Hsieh, and Mark M. Wilde. Quantum Rate Distortion, Reverse Shannon Theorems, and Source-Channel Separation. *IEEE Transactions on Information Theory*, 59(1):615–630, 2013.

[31] Nilanjana Datta and Renato Renner. Smooth Entropies and the Quantum Information Spectrum. *IEEE Transactions on Information Theory*, 55(6):2807–2815, 2009.

[32] E. Brian Davies. *Linear Operators and their Spectra*. Cambridge University Press, Cambridge, England, UK, April 2007.

[33] Igor Devetak, Marius Junge, Christoper King, and Mary Beth Ruskai. Multiplicativity of Completely Bounded p-Norms Implies a New Additivity Result. *Commun. Math. Phys.*, 266(1):37–63, August 2006.

[34] Igor Devetak and Andreas Winter. Distillation of secret key and entanglement from quantum states. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 461(2053):207–235, 2005.

[35] Igor Devetak and Jon Yard. Exact Cost of Redistributing Multipartite Quantum States. *Phys. Rev. Lett.*, 100(23):230501, June 2008.

[36] David P. DiVincenzo, Michał Horodecki, Debbie W. Leung, John A. Smolin, and Barbara M. Terhal. Locking Classical Correlations in Quantum States. *Physical Review Letters*, 92:067902, 2004.

[37] F. Dupuis and O. Fawzi. Entropy accumulation with improved second-order term. *IEEE Transactions on Information Theory*, pages 1–1, 2019.

[38] Frédéric Dupuis. Chain rules for quantum Rényi entropies. *Journal of Mathematical Physics*, 56(2):022203, 2015.

[39] Frédéric Dupuis. Privacy Amplification and Decoupling Without Smoothing. *IEEE Trans. Inf. Theory*, 69(12):7784–7792, August 2023.

[40] Frédéric Dupuis, Mario Berta, Jürg Wullschleger, and Renato Renner. One-shot decoupling. *Communications in Mathematical Physics*, 328(1):251–284, 2014.

[41] Frédéric Dupuis, Omar Fawzi, and Renato Renner. Entropy accumulation. *Communications in Mathematical Physics*, 379(3):867–913, September 2020.

[42] Frédéric Dupuis. Privacy Amplification and Decoupling Without Smoothing. *IEEE Trans. Inf. Theory*, 69(12):7784–7792, August 2023.

[43] Artur K. Ekert. Quantum cryptography based on Bell's theorem. *Physical Review Letters*, 67(6):661–663, 1991.

[44] Kun Fang, Hamza Fawzi, and Omar Fawzi. Generalized quantum asymptotic equipartition. *arXiv.2411.04035*, November 2024.

[45] Hamza Fawzi and Omar Fawzi. Defining quantum divergences via convex optimization. *Quantum*, 5:387, January 2021.

[46] Omar Fawzi, Jan Kochanowski, Cambyse Rouzé, and Thomas Van Himbeeck. Additivity and chain rules for quantum entropies via multi-index Schatten norms. *arXiv:2502.01611v1 [quant-ph]*, 2025.

[47] Omar Fawzi and Renato Renner. Quantum Conditional Mutual Information and Approximate Markov Chains. *Commun. Math. Phys.*, 340(2):575–611, December 2015.

[48] M. Fekete. Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. *Math. Z.*, 17(1):228–249, December 1923.

[49] Agnes Ferenczi and Norbert Lütkenhaus. Symmetries in quantum key distribution and the connection between optimal attacks and optimal cloning. *Phys. Rev. A*, 85(5):052310, May 2012.

[50] Rupert L. Frank and Elliott H. Lieb. Monotonicity of a relative Rényi entropy. *Journal of Mathematical Physics*, 54(12), December 2013.

[51] Christopher A. Fuchs. *Distinguishability and Accessible Information in Quantum Theory*. PhD thesis, University of New Mexico, 1996.

[52] Ian George, Jie Lin, and Norbert Lütkenhaus. Numerical calculations of the finite key rate for general quantum key distribution protocols. *Phys. Rev. Res.*, 3:013274, Mar 2021.

[53] Ian George, Jie Lin, Thomas van Himbeeck, Kun Fang, and Norbert Lütkenhaus. Finite-Key Analysis of Quantum Key Distribution with Characterized Devices Using Entropy Accumulation. *arXiv:2203.06554v1 [quant-ph]*, 2022.

[54] O. Gittsovich, N. J. Beaudry, V. Narasimhachar, R. Romero Alvarez, T. Moroder, and N. Lütkenhaus. Squashing Model for Detectors and Applications to Quantum-Key-Distribution Protocols. *Physical Review A*, 89(1):012325, January 2014.

[55] Gilad Gour and Mark M. Wilde. Entropy of a quantum channel. *Phys. Rev. Res.*, 3:023096, 2021.

[56] Manish K. Gupta and Mark M. Wilde. Multiplicativity of Completely Bounded p-Norms Implies a Strong Converse for Entanglement-Assisted Capacity. *Commun. Math. Phys.*, 334(2):867–887, March 2015.

[57] Thomas A. Hahn, Ernest Y.-Z. Tan, and Peter Brown. Bounds on Petz-Rényi Divergences and their Applications for Device-Independent Cryptography. *arXiv:2408.12313v1 [quant-ph]*, 2024.

[58] D. Haussler and M. Opper. Mutual information, metric entropy and cumulative relative entropy risk. *The Annals of Statistics*, 25(6):2451–2492, 1997.

[59] Patrick Hayden, Richard Jozsa, Dénes Petz, and Andreas Winter. Structure of States Which Satisfy Strong Subadditivity of Quantum Entropy with Equality. *Commun. Math. Phys.*, 246(2):359–374, April 2004.

[60] Matthias Heid and Norbert Lütkenhaus. Efficiency of coherent-state quantum cryptography in the presence of loss: Influence of realistic error correction. *Phys. Rev. A*, 73(5):052316, May 2006.

[61] Carl W. Helstrom. Quantum detection and estimation theory. *J. Stat. Phys.*, 1(2):231–252, June 1969.

[62] Karol Horodecki, Michal Horodecki, Pawel Horodecki, and Jonathan Oppenheim. General Paradigm for Distilling Classical Key From Quantum States. *IEEE Transactions on Information Theory*, 55(4):1898–1929, 2009.

[63] W. Y. Hwang. Quantum Key Distribution with High Loss: Toward Global Secure Communication. *Physical Review Letters*, 91(5):057901, November 2002.

[64] Rahul Jain and Srijita Kundu. A direct product theorem for quantum communication complexity with applications to device-independent QKD. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, February 2022.

[65] Lars Kamin, Amir Arqand, Ian George, Norbert Lütkenhaus, and Ernest Y.-Z. Tan. Finite-size analysis of prepare-and-measure and decoy-state quantum key distribution via entropy accumulation. *PRX Quantum*, 6:020342, Jun 2025.

[66] Lars Kamin, John Burniston, and Ernest Y.-Z. Tan. Rényi security framework against coherent attacks applied to decoy-state QKD. *arXiv.2504.12248*, April 2025.

[67] Lars Kamin, Devashish Tupkary, and Norbert Lütkenhaus. Improved finite-size effects in QKD protocols with applications to decoy-state QKD. *arXiv:2502.05382v1*, 2025.

[68] Sumeet Khatri and Mark M. Wilde. Principles of Quantum Communication Theory: A Modern Approach. *arXiv:2011.04672*, November 2020.

[69] C. King. The capacity of the quantum depolarizing channel. *IEEE Transactions on Information Theory*, 49(1):221–229, 2003.

[70] Christopher King, Keiji Matsumoto, Michael Nathanson, and Mary Beth Ruskai. Properties of Conjugate Channels with Applications to Additivity and Multiplicativity. *arXiv.quant-ph/0509126*, September 2005.

[71] M Koashi. Simple security proof of quantum key distribution based on complementarity. *New Journal of Physics*, 11(4):045018, 2009.

[72] Masato Koashi and Nobuyuki Imoto. Operations that do not disturb partially known quantum states. *Phys. Rev. A*, 66:022318, Aug 2002.

[73] Robert König, Renato Renner, Andor Bariska, and Ueli Maurer. Small Accessible Quantum Information Does Not Imply Security. *Physical Review Letters*, 98:140502, 2007.

[74] S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.

[75] Christian Kurtsiefer, Patrick Zarda, Sonja Mayer, and Harald Weinfurter. The breakdown flash of silicon avalanche photodiodes-back door for eavesdropper attacks? *Journal of Modern Optics*, 48(13):2039–2047, November 2001.

[76] Simon Lacoste-Julien and Martin Jaggi. On the Global Linear Convergence of Frank-Wolfe Optimization Variants. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[77] Felix Leditzky. Relative entropies and their use in quantum information theory. *arXiv:1611.08802*, November 2016.

[78] Felix Leditzky, Eneet Kaur, Nilanjana Datta, and Mark M. Wilde. Approaches for approximate additivity of the Holevo information of quantum channels. *Phys. Rev. A*, 97:012332, Jan 2018.

[79] Charles Ci Wen Lim, Marcos Curty, Nino Walenta, Feihu Xu, and Hugo Zbinden. Concise Security Bounds for Practical Decoy-State Quantum Key Distribution. *Physical Review A*, 89(2):022307, February 2014.

[80] Charles Ci-Wen Lim, Feihu Xu, Jian-Wei Pan, and Artur Ekert. Security Analysis of Quantum Key Distribution with Small Block Length and Its Application to Quantum Space Communications. *Physical Review Letters*, 126:100501, 2021.

[81] Wen-Zhao Liu, Ming-Han Li, Sammy Ragy, Si-Ran Zhao, Bing Bai, Yang Liu, Peter J. Brown, Jun Zhang, Roger Colbeck, Jingyun Fan, Qiang Zhang, and Jian-Wei Pan. Device-independent randomness expansion against quantum side information. *Nature Physics*, 17(4):448–451, February 2021.

[82] Xiongfeng Ma, Bing Qi, Yi Zhao, and Hoi-Kwong Lo. Practical Decoy State for Quantum Key Distribution. *Physical Review A*, 72(1):012326, July 2005.

[83] Vaisakh Mannalath, Víctor Zapatero, and Marcos Curty. Sharp finite statistics for quantum key distribution. *arXiv:2410.04095v3 [quant-ph]*, 2024.

[84] Y. Mansour, M. Mohri, and A. Rostamizadeh. Multiple source adaptation and the Rényi divergence. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 367–374, 2009.

[85] A. Marwah and F. Dupuis. Smooth min-entropy lower bounds for approximation chains. *Communications in Mathematical Physics*, 405(1):211–254, 2024.

[86] Ashutosh Marwah and Frédéric Dupuis. Proving security of BB84 under source correlations. *arXiv:2402.12346*, February 2024.

[87] Alexander McKinlay and Marco Tomamichel. Decomposition rules for quantum Rényi mutual information with an application to information exclusion relations. *Journal of Mathematical Physics*, 61(7):072202, 07 2020.

[88] Tony Metger, Omar Fawzi, David Sutter, and Renato Renner. Generalised entropy accumulation. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 844–850, 2022.

[89] Tony Metger, Omar Fawzi, David Sutter, and Renato Renner. Generalised Entropy Accumulation. *Communications in Mathematical Physics*, 405(11), October 2024.

[90] Tony Metger and Renato Renner. Security of quantum key distribution from generalised entropy accumulation. *Nat. Commun.*, 14(5272):1–13, August 2023.

[91] D. Morales, L. Pardo, and I. Vajda. Rényi statistics in directed families of exponential experiments. *Statistics*, 34:151–174, 2000.

[92] Milán Mosonyi and Tomohiro Ogawa. Quantum Hypothesis Testing and the Operational Interpretation of the Quantum Rényi Relative Entropies. *Communications in Mathematical Physics*, 334(3):1617–1648, December 2014.

[93] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum Rényi entropies: A new generalization and some properties. *Journal of Mathematical Physics*, 54(12), December 2013.

[94] Shlok Nahar, Devashish Tupkary, Yuming Zhao, Norbert Lütkenhaus, and Ernest Y.-Z. Tan. Postselection technique for optical Quantum Key Distribution with improved de Finetti reductions. *arXiv:2403.11851v2 [quant-ph]*, 2024.

[95] Jef Pauwels, Stefano Pironio, Erik Woodhead, and Armin Tavakoli. Almost Qudits in the Prepare-and-Measure Scenario. *Physical Review Letters*, 129:250504, Dec 2022.

[96] Margarida Pereira, Guillermo Currás-Lorenzo, Álvaro Navarrete, Akihiro Mizutani, Go Kato, Marcos Curty, and Kiyoshi Tamaki. Modified BB84 quantum key distribution protocol robust to source imperfections. *Phys. Rev. Res.*, 5:023065, Apr 2023.

[97] Dénes Petz. Sufficiency of channels over von Neumann algebras. *Q. J. Math.*, 39(1):97–108, March 1988.

[98] Dénes Petz. Quasi-entropies for States of a von Neumann algebra. *Publications of the Research Institute for Mathematical Sciences*, 21(4):787–800, 1985.

[99] Dénes Petz. Quasi-entropies for finite quantum systems. *Reports on Mathematical Physics*, 23(1):57–65, February 1986.

[100] Paulo Vinicius Pereira Pinheiro, Poompong Chaiwongkhot, Shihan Sajeed, Rolf T. Horn, Jean-Philippe Bourgoin, Thomas Jennewein, Norbert Lütkenhaus, and Vadim Makarov. Eavesdropping and countermeasures for backflash side channel in quantum cryptography. *Optics Express*, 26(16):21020, July 2018.

[101] Christopher Portmann and Renato Renner. Security in quantum cryptography. *Rev. Mod. Phys.*, 94:025008, 2022.

[102] Bartosz Regula, Ludovico Lami, and Nilanjana Datta. Tight relations and equivalences between smooth relative entropies. *arXiv.2501.12447*, January 2025.

[103] J. M. Renes and R. Renner. One-Shot Classical Data Compression With Quantum Side Information and the Distillation of Common Randomness or Secret Keys. *IEEE Transactions on Information Theory*, 58(3):1985–1991, March 2012.

[104] Renato Renner. Security of Quantum Key Distribution, 2005.

[105] Alfréd Rényi. On measures of entropy and information. In *1961 Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 547–561. University of California Press, 1961.

[106] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[107] Yu-Chen Shen, Li Gao, and Hao-Chung Cheng. Strong Converse for Privacy Amplification against Quantum Side Information. *arXiv.2202.10263*, February 2022.

[108] Peter W. Shor. Additivity of the classical capacity of entanglement-breaking quantum channels. *J. Math. Phys.*, 43(9):4334–4340, September 2002.

[109] Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8:171–176, 1958.

[110] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory*, 19(4):471–480, July 1973.

[111] Sohail, Vivek Pandey, Uttam Singh, and Siddhartha Das. Fundamental limitations on the recoverability of quantum processes, 2024.

[112] Ernest Y.-Z. Tan. Robustness of implemented device-independent protocols against constrained leakage. *arXiv:2302.13928v2*, 2023.

[113] Vincent Y. F. Tan and Oliver Kosut. On the Dispersions of Three Network Information Theory Problems. *IEEE Trans. Inf. Theory*, 60(2):881–903, November 2013.

[114] Armin Tavakoli, Jef Pauwels, Erik Woodhead, and Stefano Pironio. Correlations in Entanglement-Assisted Prepare-and-Measure Scenarios. *PRX Quantum*, 2:040357, 2021.

[115] Armin Tavakoli, Emmanuel Zambrini Cruzeiro, Erik Woodhead, and Stefano Pironio. Informationally restricted correlations: a general framework for classical and quantum systems. *Quantum*, 6:620, January 2022.

[116] M. Tomamichel, R. Colbeck, and R. Renner. A Fully Quantum Asymptotic Equipartition Property. *IEEE Transactions on Information Theory*, 55(12):5840–5847, 2009.

[117] Marco Tomamichel. *Quantum Information Processing with Finite Resources*. Springer International Publishing, 2016.

[118] Marco Tomamichel, Mario Berta, and Masahito Hayashi. Relating different quantum generalizations of the conditional Rényi entropy. *Journal of Mathematical Physics*, 55(8), August 2014.

[119] Marco Tomamichel and Anthony Leverrier. A largely self-contained and complete security proof for quantum key distribution. *Quantum*, 1:14, 2017.

[120] Marco Tomamichel, Jesus Martinez-Mateo, Christoph Pacher, and David Elkouss. Fundamental finite key limits for one-way information reconciliation in quantum key distribution. *Quantum Information Processing*, 16(11), October 2017.

[121] Marco Tomamichel and Renato Renner. Uncertainty Relation for Smooth Entropies. *Physical Review Letters*, 106:110506, Mar 2011.

[122] Devashish Tupkary, Shlok Nahar, Pulkit Sinha, and Norbert Lütkenhaus. Phase error rate estimation in QKD with imperfect detectors. *arXiv:2408.17349v2*, 2024.

[123] Devashish Tupkary, Ernest Y.-Z. Tan, and Norbert Lütkenhaus. Security proof for variable-length quantum key distribution. *Physical Review Research*, 6:023002, 2024.

[124] Hisaharu Umegaki. Conditional Expectations in an Operator Algebra IV (Entropy and Information). *Kodai Mathematical Seminar Reports*, 14(2):59–85, 1962.

[125] T. van Erven. *When Data Compression and Statistics Disagree: Two Frequentist Challenges for the Minimum Description Length Principle*. Phd thesis, Leiden University, 2010.

[126] Thomas van Himbeeck and Peter Brown. Tight and general finite-size security of quantum key distribution, 2025. Presented at QCrypt 2023 and QIP 2024, manuscript in preparation.

[127] Alexander Vitanov, Frédéric Dupuis, Marco Tomamichel, and Renato Renner. Chain Rules for Smooth Min- and Max-Entropies. *IEEE Transactions on Information Theory*, 59(5):2603–2612, May 2013.

[128] Wenyuan Wang and Norbert Lütkenhaus. Numerical security proof for the decoy-state BB84 protocol and measurement-device-independent quantum key distribution resistant against large basis misalignment. *Physical Review Research*, 4:043097, 2022.

[129] Xiang-Bin Wang. Beating the Photon-Number-Splitting Attack in Practical Quantum Cryptography. *Phys. Rev. Lett.*, 94(23):230503, June 2005.

[130] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.

[131] Mark M. Wilde, Andreas Winter, and Dong Yang. Strong Converse for the Classical Capacity of Entanglement-Breaking and Hadamard Channels via a Sandwiched Rényi Relative Entropy. *Commun. Math. Phys.*, 331(2):593–622, October 2014.

[132] A. Winick, N. Lütkenhaus, and P. J. Coles. Reliable numerical key rates for quantum key distribution. *Quantum*, 2:77, 2018.

[133] Andreas Winter. Weak Locking Capacity of Quantum Channels Can be Much Larger Than Private Capacity. *Journal of Cryptology*, 30(1):1–21, 2017.

[134] Xiao Yuan. Hypothesis testing and entropies of quantum channels. *Phys. Rev. A*, 99(3):032317, March 2019.

[135] T. Zhang. From ϵ-entropy to KL-entropy: Analysis of minimum information complexity density estimation. *The Annals of Statistics*, 34(5):2180–2210, 2006.

[136] Yanbao Zhang, Honghao Fu, and Emanuel Knill. Efficient randomness certification by quantum probability estimation. *Physical Review Research*, 2:013016, January 2020.

APPENDICES

Appendix A

Missing proofs in Chapter 4

A.1 Proof of Lemma 4.4.1

Lemma 4.4.1 has almost the same proof as Theorem 4.4.1, except we introduce a single inequality $H_{\widehat{\alpha}}(DS\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega} \geq H_{\widehat{\alpha}}(DS|\overline{C}\widehat{C}E\widetilde{E})_{\nu^\omega}$ into the proof, which simplifies the final results.

We begin with the second line of the bound (4.80). Taking any read-and-prepare channel of the form described in Lemma 4.3.1, extending any ν^ω in the infimum with this channel yields

$$\begin{aligned}
H^f_{\widehat{\alpha}}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega} &= H_{\widehat{\alpha}}(DS\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega} - M \\
&\geq H_{\widehat{\alpha}}(DS|\overline{C}\widehat{C}E\widetilde{E})_{\nu^\omega} - M \\
&= \frac{1}{1-\widehat{\alpha}} \log \left(\sum_{\bar{c}\hat{c} \in \operatorname{supp}\left(\boldsymbol{\nu}^\omega_{\overline{C}\widehat{C}}\right)} \nu^\omega(\bar{c}\hat{c}) 2^{(1-\widehat{\alpha}) H_{\widehat{\alpha}}(DS|E\widetilde{E})_{\nu^\omega_{|\bar{c}\hat{c}}}} \right) - M \\
&= \frac{1}{1-\widehat{\alpha}} \log \left(\sum_{\bar{c}\hat{c} \in \operatorname{supp}\left(\boldsymbol{\nu}^\omega_{\overline{C}\widehat{C}}\right)} \nu^\omega(\bar{c}\hat{c}) 2^{(1-\widehat{\alpha})\left(H_{\widehat{\alpha}}(D)_{\nu^\omega_{|\bar{c}\hat{c}}} + H_{\widehat{\alpha}}(S|E\widetilde{E})_{\nu^\omega_{|\bar{c}\hat{c}}} \right)} \right) - M \\
&= \frac{1}{1-\widehat{\alpha}} \log \left(\sum_{\bar{c}\hat{c} \in \operatorname{supp}\left(\boldsymbol{\nu}^\omega_{\overline{C}\widehat{C}}\right)} \nu^\omega(\bar{c}\hat{c}) 2^{(1-\widehat{\alpha})\left(-f(\bar{c}\hat{c}) + H_{\widehat{\alpha}}(S|E\widetilde{E})_{\nu^\omega_{|\bar{c}\hat{c}}} \right)} \right), \qquad \text{(A.1)}
\end{aligned}$$

where the first line is due to Eq. (4.10), the second line is an application of [117, Proposition 5.5] by noting $\nu^{\omega}_{\overline{C}\widehat{C}E\widetilde{E}}$ is separable across the registers $\overline{C}$ and $\widehat{C}E\widetilde{E}$, the third line follows from Lemma 2.2.1, the fourth line holds since $\nu^{\omega}_{DSE\widetilde{E}_{|\bar{c}\hat{c}}} = \nu^{\omega}_{D_{|\bar{c}\hat{c}}} \otimes \nu^{\omega}_{SE\widetilde{E}_{|\bar{c}\hat{c}}}$, and the last follows from Eq. (4.9). Furthermore, note that $H_{\widehat{\alpha}}(DS|\overline{C}\widehat{C}E\widetilde{E})_{\nu^{\omega}} - M$ is convex in ω by following the same steps as Eq. (4.51) for $H_{\widehat{\alpha}}(DS|\overline{C}\widehat{C}E\widetilde{E})_{\nu^{\omega}}$ and all lines after that point are equalities; therefore, the last line is also convex in ω.

With this, we can apply exactly the same analysis as in the rest of the Theorem 4.4.1 proof, replacing $H^{f}_{\widehat{\alpha}}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^{\omega}}$ with the lower bound given in the last line above, except that we would modify the definition of $G_{\widehat{\alpha},\nu^{\omega}}$ to

$$G_{\widehat{\alpha},\nu^{\omega}}(\mathbf{f}) \coloneqq -\frac{1}{1-\widehat{\alpha}} \log \left(\sum_{\bar{c}\hat{c} \in \operatorname{supp}\left(\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right)} \nu^{\omega}(\bar{c}\hat{c}) 2^{(1-\widehat{\alpha})\left(-f(\bar{c}\hat{c}) + H_{\widehat{\alpha}}(S|E\widetilde{E})_{\nu^{\omega}_{|\bar{c}\hat{c}}}\right)} \right), \tag{A.2}$$

and carry out the subsequent calculations with $-H_{\widehat{\alpha}}(S|E\widetilde{E})_{\nu^{\omega}_{|\bar{c}\hat{c}}}$ in place of $D_{\widehat{\alpha}}\left(\nu^{\omega}_{SE\widetilde{E}\wedge\bar{c}\hat{c}} \middle\| \mathbb{I}_S \otimes \nu^{\omega}_{E\widetilde{E}\wedge\hat{c}}\right)$.

To show that equality holds for the case where $\overline{C}$ is trivial, we could note that the only point in the above calculations that differs from Sec. 4.4.4 is the inequality in the second line of (A.1), which is immediately redundant if $\overline{C}$ is trivial.

A.2 Proof of Lemma 4.4.2

We begin with the second line of the bound (4.80). Now instead consider a read-and-prepare channel $\overline{C}\widehat{C} \to \overline{C}\widehat{C}D$ of the form described in Lemma 4.3.2, so that for any $\alpha \in [0,\infty]$ the bound (4.12) holds, i.e. all Rényi entropy values of the D register (conditioned on $\overline{C}\widehat{C} = \bar{c}\hat{c}$) lie in the interval $\left[M - f(\bar{c}\hat{c}), M - f(\bar{c}\hat{c}) + 2^{-M/2}\log e\right]$. If we extend any ν^{ω} in the infimum with this channel, we have

$$\begin{aligned} H^{f}_{\widehat{\alpha}}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^{\omega}} &\geq H_{\widehat{\alpha}}(DS\overline{C}|\widehat{C}E\widetilde{E})_{\nu^{\omega}} - M - 2^{-\frac{M}{2}}\log e \\ &\geq H_{\alpha'}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^{\omega}} + H^{\uparrow}_{\alpha''}(D|S\overline{C}\widehat{C}E\widetilde{E})_{\nu^{\omega}} - M - 2^{-\frac{M}{2}}\log e, \end{aligned} \tag{A.3}$$

where the first line holds by (4.13), and the second line is a special case of Proposition 7 in [38]. Before proceeding further, we note that the second term in the last line of Eq. (A.3)

can be bounded as:

$$
\begin{aligned}
H^{\uparrow}_{\alpha''}(D|S\overline{C}\widehat{C}E\widetilde{E})_{\nu^\omega} &= \frac{\alpha''}{1-\alpha''}\log\left(\sum_{\bar{c}\hat{c}\in\mathrm{supp}\left(\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right)} \nu^\omega(\bar{c}\hat{c})2^{\left(\frac{1-\alpha''}{\alpha''}\right)H^{\uparrow}_{\alpha''}(D|SE\widetilde{E})_{\nu^\omega_{|\bar{c}\hat{c}}}}\right) \\
&= \frac{\alpha''}{1-\alpha''}\log\left(\sum_{\bar{c}\hat{c}\in\mathrm{supp}\left(\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right)} \nu^\omega(\bar{c}\hat{c})2^{\left(\frac{1-\alpha''}{\alpha''}\right)H_{\alpha''}(D)_{\nu^\omega_{|\bar{c}\hat{c}}}}\right) \\
&\geq M + \frac{\alpha''}{1-\alpha''}\log\left(\sum_{\bar{c}\hat{c}\in\mathrm{supp}\left(\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right)} \nu^\omega(\bar{c}\hat{c})2^{-\left(\frac{1-\alpha''}{\alpha''}\right)f(\bar{c}\hat{c})}\right), \qquad \text{(A.4)}
\end{aligned}
$$

where the first line is due to Lemma 2.2.1, the second line holds since $\nu^{\omega}_{DSE\widetilde{E}_{|\bar{c}\hat{c}}} = \nu^{\omega}_{D_{|\bar{c}\hat{c}}} \otimes \nu^{\omega}_{SE\widetilde{E}_{|\bar{c}\hat{c}}}$ and $H^{\uparrow}_{\alpha''}(D)_{\nu^\omega_{|\bar{c}\hat{c}}} = H_{\alpha''}(D)_{\nu^\omega_{|\bar{c}\hat{c}}}$ when there is no conditioning register, and the third line follows from the lower bound in (4.12). Combining Eq. (A.4) with Eq. (A.3) we have:

$$
H^{f}_{\widehat{\alpha}}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega} \geq H_{\alpha'}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega} + \frac{\alpha''}{1-\alpha''}\log\left(\sum_{\bar{c}\hat{c}\in\mathrm{supp}\left(\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right)} \nu^\omega(\bar{c}\hat{c})2^{-\left(\frac{1-\alpha''}{\alpha''}\right)f(\bar{c}\hat{c})}\right) - 2^{-\frac{M}{2}}\log e. \qquad \text{(A.5)}
$$

Since the above bound holds for arbitrary (sufficiently large) M, we can take the $M \to \infty$ limit so the $2^{-\frac{M}{2}}$ term vanishes. Taking this into account and using the bound in Eq. (A.5), our desired result follows by applying exactly the same analysis as in the rest of the Theorem 4.4.1 proof, except that we replace $G_{\widehat{\alpha},\rho}(\mathbf{f})$ with

$$
G_{\alpha',\alpha'',\nu^\omega}(\mathbf{f}) \coloneqq -H_{\alpha'}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega} - \frac{\alpha''}{1-\alpha''}\log\left(\sum_{\bar{c}\hat{c}\in\mathrm{supp}\left(\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right)} \nu^\omega(\bar{c}\hat{c})2^{-\left(\frac{1-\alpha''}{\alpha''}\right)f(\bar{c}\hat{c})}\right). \qquad \text{(A.6)}
$$

Note that the above expression is concave with respect to ω, as we would require for those remaining proof steps. This follows by observing that $H_{\alpha'}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega}$ is convex in ω by following same steps as Eq. (4.51) for $H_{\alpha'}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega}$, and we have $\alpha'' > 1$ so $-\frac{\alpha''}{1-\alpha''}\log(\cdot)$ is a concave function, while the values $\nu^\omega(\bar{c}\hat{c})$ are affine functions of ω. In fact, to make the

correspondence with the Theorem 4.4.1 proof even more explicit, we can rewrite the above expression as

$$G_{\alpha',\alpha'',\nu^\omega}(\mathbf{f}) = -\frac{1}{1-\widetilde{\alpha}''}\log\left(\sum_{\bar{c}\hat{c}\in\mathrm{supp}\left(\boldsymbol{\nu}^{\omega}_{\overline{C}\widehat{C}}\right)} \nu^\omega(\bar{c}\hat{c})2^{(1-\widetilde{\alpha}'')\left(-f(\bar{c}\hat{c})+H_{\alpha'}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega}\right)}\right), \quad \text{(A.7)}$$

where $\widetilde{\alpha}'' \coloneqq \frac{2\alpha''-1}{\alpha''}$, so we are carrying out the remaining calculations with $\widetilde{\alpha}''$ in place of $\widehat{\alpha}$ and a single value $H_{\alpha'}(S\overline{C}|\widehat{C}E\widetilde{E})_{\nu^\omega}$ in place of all the $-D_{\widehat{\alpha}}\left(\nu^{\omega}_{SE\widetilde{E}\wedge\bar{c}\hat{c}}\middle\|\mathbb{I}_S\otimes\nu^{\omega}_{E\widetilde{E}\wedge\hat{c}}\right)$ terms.

Appendix B

Generalized Rényi entropy accumulation for $\alpha < 1$

In this appendix, we construct an upper bound on the global Rényi entropy $H_\alpha^\uparrow(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_{\rho|\Omega}$ in the region where $\alpha < 1$. Using standard techniques [68], one can then obtain bounds on the smooth max-entropy $H_{1/2}^{\uparrow,\varepsilon}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_{\rho|\Omega}$, similar to the approach taken in [41, 89].

We begin by defining a Petz version of QES entropies.

Definition B.0.1. Let $\rho \in \mathrm{S}_=(\overline{C}\widehat{C}QQ')$ be a state where $\overline{C}$ and $\widehat{C}$ are classical with alphabets $\overline{\mathcal{C}}$ and $\widehat{\mathcal{C}}$ respectively. Given a QES f on $\overline{C}\widehat{C}$ and a value $\alpha \in (0,1)\cup(1,\infty)$, the **QES-Petz-entropy (QESP-entropy) of order** α for ρ is defined as

$$\begin{aligned}\bar{H}_\alpha^f(Q\overline{C}|\widehat{C}Q')_\rho &:= \frac{1}{1-\alpha}\log\left(\sum_{\bar{c}\hat{c}}\rho(\bar{c}\hat{c})^\alpha\rho(\hat{c})^{1-\alpha}\,2^{(1-\alpha)\left(-f(\bar{c}\hat{c})-\bar{D}_\alpha\left(\rho_{QQ'|\bar{c}\hat{c}}\middle\|\mathbb{I}_Q\otimes\rho_{Q'|\hat{c}}\right)\right)}\right)\\ &= \frac{1}{1-\alpha}\log\left(\sum_{\bar{c}\hat{c}}\rho(\bar{c}\hat{c})2^{(1-\alpha)\left(-f(\bar{c}\hat{c})-\bar{D}_\alpha\left(\rho_{QQ'\wedge\bar{c}\hat{c}}\middle\|\mathbb{I}_Q\otimes\rho_{Q'\wedge\hat{c}}\right)\right)}\right)\\ &= \frac{1}{1-\alpha}\log\left(\sum_{\bar{c}\hat{c}}2^{-(1-\alpha)f(\bar{c}\hat{c})}\operatorname{Tr}\left[\left(\rho_{QQ'\wedge\bar{c}\hat{c}}\right)^\alpha\left(\rho_{Q'\wedge\hat{c}}\right)^{1-\alpha}\right]\right),\end{aligned} \tag{B.1}$$

where the sum is over all $\bar{c}\hat{c}$ values such that $\rho(\bar{c}\hat{c}) > 0$, and we leave some tensor factors of identity implicit in the last expression.

With this, we present the $\alpha < 1$ version of our bound in Theorem 4.3.1. This version exhibits some subtle differences compared to its $\alpha > 1$ counterpart.

First, in the single-round quantity, the extra purifying register $\widetilde{E}$ no longer appears. This is because the single-round expression now involves a supremum, rather than an infimum, over input states. By the data-processing inequality, the optimal input state can be taken to be trivial on the purifying register.

Second, in this direction of the bound, the choice of QES can no longer depend on the values $\bar{c}_1^{j-1}\hat{c}_1^{j-1}$ from previous rounds. As the reason behind this restriction is somewhat technical, we defer the explanation to the middle of the proof, where the obstruction naturally arises.

Theorem B.0.1. *Let $\{\mathcal{M}_j\}_{j=1}^n$ be a sequence of channels with $\mathcal{M}_j \in \mathrm{CPTP}(R_{j-1}E_{j-1}, S_jR_jE_j\overline{C}_j\widehat{C}_j)$ such that the output registers $\overline{C}_j\widehat{C}_j$ are always classical (for any input state), and let $\rho \in \mathrm{S}_=(S_1^n\overline{C}_1^n\widehat{C}_1^nE_nR_n)$ be a state of the form $\rho = \mathcal{M}_n \circ \cdots \circ \mathcal{M}_1[\omega^0]$ (leaving some identity channels implicit) for some initial state $\omega^0 \in \mathrm{S}_=(R_0E_0)$. Suppose that each $\mathcal{M}_j$ has some Stinespring dilation $V_j \in \mathrm{CPTP}(R_{j-1}E_{j-1}, S_jR_jE_j\overline{C}_j\widehat{C}_jF_j)$ (with "environment" system F_j), such that V_j satisfies the following NS condition:*[1]

$$\exists \text{ a channel } \mathcal{R}_j \in \mathrm{CPTP}(R_{j-1}, R_j\overline{C}_j\widehat{C}_jF_j) \text{ such that } \mathrm{Tr}_{S_jE_j} \circ V_j = \mathcal{R}_j \circ \mathrm{Tr}_{E_{j-1}}. \tag{B.2}$$

For each j, let $f_{|j}$ be a QES on registers $\overline{C}_j\widehat{C}_j$. Define the following QES on $\overline{C}_1^n\widehat{C}_1^n$:

$$f_{\mathrm{full}}(\bar{c}_1^n\hat{c}_1^n) \coloneqq \sum_{j=1}^n f_{|j}(\bar{c}_j\hat{c}_j). \tag{B.3}$$

Take any $\alpha \in (2/3, 1)$ and let $\widetilde{\alpha} = \frac{3\alpha-2}{2\alpha-1}$. Then we have

$$\begin{aligned} \bar{H}_\alpha^{f_{\mathrm{full}}}(S_1^n\overline{C}_1^n|\widehat{C}_1^nE_n)_\rho &\leq \sum_j \kappa_j, \\ \text{where} \quad \kappa_j &\coloneqq \sup_{\nu \in \Sigma_j} \bar{H}_{\widetilde{\alpha}}^{f_{|j}}(S_j\overline{C}_j|\widehat{C}_jE_j)_\nu, \end{aligned} \tag{B.4}$$

where Σ_j denotes the set of all states of the form $\mathcal{M}_j\left[\omega_{R_{j-1}E_{j-1}}\right]$ for some initial state $\omega \in \mathrm{S}_=(R_{j-1}E_{j-1})$.

Consequently, if we instead define the following "normalized" QES on $\overline{C}_1^n\widehat{C}_1^n$:

$$\hat{f}_{\mathrm{full}}(\bar{c}_1^n\hat{c}_1^n) \coloneqq \sum_{j=1}^n \hat{f}_{|j}(\bar{c}_j\hat{c}_j), \quad \text{where} \quad \hat{f}_{|j}(\bar{c}_j\hat{c}_j) \coloneqq f_{|j}(\bar{c}_j\hat{c}_j) + \kappa_j, \tag{B.5}$$

[1] Qualitatively, this states that V_j does not signal from E_{j-1} to $R_j\overline{C}_j\widehat{C}_jF_j$. Also note that the choice of Stinespring dilation can be arbitrary here, due to the isometric equivalence of purifications.

then

$$H_\alpha^{\hat{f}_{\text{full}}}(S_1^n\overline{C}_1^n|\widehat{C}_1^n E_n)_\rho \leq 0. \tag{B.6}$$

Before we present the proof, we highlight some remarks on this theorem. First, note that if we set $\alpha = 1-\mu$ for some $\mu > 0$, then $\widetilde{\alpha} = 1-\mu-O(\mu^2)$. Therefore, the change in Rényi parameter is again of a higher-order amount. In addition, as observed in [89], if the action of $\mathcal{M}_j$ is such that it first apply some channel $\widetilde{\mathcal{M}}_j \in \text{CPTP}(R_{j-1}, S_j\overline{C}_j\widehat{C}_j T_j R_j)$, then incorporate $\widehat{C}_j T_j$ into the side-information E_j without "disturbing" the state already on the register E_{j-1} (i.e., $\mathcal{M}_j = \text{id}_{\widehat{C}_j T_j E_{j-1}\to E_j} \circ \widetilde{\mathcal{M}}_j$ for some $\widetilde{\mathcal{M}}_j$ with $E_j \equiv \widehat{C}_j T_j E_{j-1}$), then, the NS condition automatically holds.

Proof. Let $M > 0$ be any value such that $M - f_{|j}(\bar{c}_j\hat{c}_j) > \frac{M}{2} > 0$ for all $f_{|j}(\bar{c}_j\hat{c}_j)$ in the theorem statement; then for each j, define a read-and-prepare channel $\mathcal{D}_j \in \text{CPTP}(\overline{C}_j\widehat{C}_j, \overline{C}_j\widehat{C}_j D_j)$ of the form described in Lemma 4.3.2, so that the state it prepares on D_j always satisfies

$$\forall \alpha \in [0,\infty], \quad H_\alpha(D_j)_{\rho_{|\bar{c}_j\hat{c}_j}} \in \left[M - f_{|j}(\bar{c}_j\hat{c}_j), M - f_{|j}(\bar{c}_j\hat{c}_j) + 2^{-\frac{M}{2}}\log e\right]. \tag{B.7}$$

Consider the channels $\mathcal{N}_j \in \text{CPTP}(R_{j-1}E_{j-1}, D_j S_j R_j E_j \overline{C}_j\widehat{C}_j)$ defined by $\mathcal{N}_j := \mathcal{D}_j \circ \mathcal{M}_j$. View ρ in the theorem statement as the reduced state of an extended version $\rho = \mathcal{N}_n \circ \cdots \circ \mathcal{N}_1[\omega^0]$. We will now only prove the following bound[2] (where again Σ'_j is the set of possible output states of $\mathcal{N}_j$, though here we do not need the purifying system $\widetilde{E}$):

$$\bar{H}_\alpha^\uparrow(D_1^n S_1^n \overline{C}_1^n|\widehat{C}_1^n E_n)_\rho \leq \sum_j \sup_{\nu'\in\Sigma'_j} \bar{H}^\uparrow_{2-\frac{1}{\alpha}}(D_j S_j \overline{C}_j|\widehat{C}_1^j E_j)_{\nu'}. \tag{B.8}$$

Once the above relation is proven, the remainder of the proof will then follow in the same way as before, with some adjustments to the Rényi parameters.

Let $U_j \in \text{CPTP}(\overline{C}_j\widehat{C}_j, \overline{C}_j\widehat{C}_j D_j G_j)$ be a Stinespring dilation of $\mathcal{D}_j$, in which case $U_j \circ V_j \in \text{CPTP}(R_{j-1}E_{j-1}, D_j S_j R_j E_j \overline{C}_j\widehat{C}_j F_j G_j)$ is a Stinespring dilation of $\mathcal{N}_j$. We extend

[2]We highlight that, if we wish to allow QES-s to depend on the past $\overline{C}_j\widehat{C}_j$ values, we would need to allow the $\mathcal{N}_j$ channels to act non-trivially on $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$. However, for the relation in (B.8) to hold, each $\mathcal{N}_j$ must satisfy a no-signaling condition from $E_{j-1}\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$ to $R_j F_1^j G_1^j$. This condition would never hold, since such channels, by construction, would allow nontrivial signaling from the $\overline{C}_1^{j-1}\widehat{C}_1^{j-1}$ registers to the Stinespring dilation F_j.

this trivially to another isometry $W_j \coloneqq (U_j \circ V_j) \otimes \mathrm{id}_{\widehat{C}_1^{j-1} F_1^{j-1} G_1^{j-1}}$. Then, we can view the state ρ in the above bounds as the reduced state of a *pure* state $|\rho\rangle_{D_1^n S_1^n R_n E_n \overline{C}_1^n \widehat{C}_1^n F_1^n G_1^n \widetilde{E}_0} \coloneqq W_n \circ \cdots \circ W_1 |\omega^0\rangle_{R_0 E_0 \widetilde{E}_0}$, where $|\omega^0\rangle_{R_0 E_0 \widetilde{E}_0}$ is a purification of $\omega^0_{R_0 E_0}$. Similarly, any $\nu' \in \Sigma'_j$ can be viewed as the reduced state of a pure state $|\nu'\rangle_{D_j S_j R_j E_j \overline{C}_j \widehat{C}_1^j F_1^j G_1^j \widetilde{E}} \coloneqq W_j |\omega\rangle_{R_{j-1} E_{j-1} \widehat{C}_1^{j-1} F_1^{j-1} G_1^{j-1} \widetilde{E}}$, where $|\omega\rangle_{R_{j-1} E_{j-1} \widehat{C}_1^{j-1} F_1^{j-1} G_1^{j-1} \widetilde{E}}$ is a pure state (for some register $\widetilde{E}$ of sufficiently large dimension). Then by a duality relation in Proposition 2.2.2 (see also [118, Theorem 2]) between the Rényi entropies $\bar{H}^{\uparrow}_{\alpha}$ and $H_{1/\alpha}$ (for all $\alpha \in [0, \infty]$), we have:

$$\begin{aligned} \bar{H}^{\uparrow}_{\alpha}(D_1^n S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n \widetilde{E}_0)_{\rho} &= -H_{\beta}(D_1^n S_1^n \overline{C}_1^n | R_n F_1^n G_1^n)_{\rho}, \text{ where } \beta = \frac{1}{\alpha}, \\ \bar{H}^{\uparrow}_{2-\frac{1}{\alpha}}(D_j S_j \overline{C}_j | \widehat{C}_1^j E_j)_{\nu'} &= -H_{\widehat{\beta}}(D_j S_j \overline{C}_j | R_j F_1^j G_1^j \widetilde{E})_{\nu'}, \text{ where } \widehat{\beta} = \frac{1}{2 - \frac{1}{\alpha}} = \frac{1}{2-\beta}. \end{aligned} \tag{B.9}$$

Note that since $\alpha \in (1/2, 1)$, we have $\beta \in (1, 2)$. We also highlight that we placed the purifications $\widetilde{E}_0$ and $\widetilde{E}$ on opposite sides in the above two equations, to simplify some subsequent arguments.

We now show that W_j satisfy the required conditions in Theorem 2.5.3. Recall each W_j channel has input registers $R_{j-1} E_{j-1} \widehat{C}_1^{j-1} F_1^{j-1} G_1^{j-1}$, and partition them into "memory" registers $E_{j-1} \widehat{C}_1^{j-1}$ (which do not appear in these entropy terms) and "side-information" registers $R_{j-1} F_1^{j-1} G_1^{j-1}$ (which appear in the conditioning registers of these entropy terms). Critically, the NS condition on V_j implies that the W_j channels are non-signaling from $E_{j-1} \widehat{C}_1^{j-1}$ to $R_j F_1^j G_1^j$; more precisely, if we define a channel $\widetilde{\mathcal{R}}_j \coloneqq \left(\mathrm{Tr}_{D_j \overline{C}_j \widehat{C}_j} \circ U_j \circ \mathcal{R}_j\right) \otimes \mathrm{id}_{F_1^{j-1} G_1^{j-1}}$ where $\mathcal{R}_j$ is the channel in the theorem condition (B.2), then this is a channel $\widetilde{\mathcal{R}}_j : R_{j-1} F_1^{j-1} G_1^{j-1} \to R_j F_1^j G_1^j$ such that (leaving various identity channels implicit):

$$\begin{aligned} \mathrm{Tr}_{D_j S_j \overline{C}_j E_j \widehat{C}_1^j} \circ W_j &= \mathrm{Tr}_{D_j S_j \overline{C}_j E_j \widehat{C}_1^j} \circ U_j \circ V_j \\ &= \mathrm{Tr}_{D_j \overline{C}_j \widehat{C}_j} \circ U_j \circ \mathrm{Tr}_{S_j E_j} \circ V_j \circ \mathrm{Tr}_{\widehat{C}_1^{j-1}} \\ &= \mathrm{Tr}_{D_j \overline{C}_j \widehat{C}_j} \circ U_j \circ \mathcal{R}_j \circ \mathrm{Tr}_{E_{j-1}} \circ \mathrm{Tr}_{\widehat{C}_1^{j-1}} \\ &= \widetilde{\mathcal{R}}_j \circ \mathrm{Tr}_{E_{j-1} \widehat{C}_1^{j-1}}, \end{aligned} \tag{B.10}$$

where in the second line we commute some partial traces with channels that act purely as identity on the corresponding registers (namely, U_j acts as identity on $S_j E_j \widehat{C}_1^{j-1}$ and V_j acts as identity on $\widehat{C}_1^{j-1}$), in the third line we apply the NS condition (B.2) imposed in the theorem, and in the fourth line we substitute the definition of $\widetilde{\mathcal{R}}_j$. Hence W_j form a

sequence of GEAT channels, and with this we can finally apply [89, Lemma 3.6] to conclude that (since $\beta \in (1,2)$ and $\widehat{\beta} = 1/(2-\beta)$)

$$H_\beta(D_1^n S_1^n \overline{C}_1^n | R_n F_1^n G_1^n)_\rho \geq \sum_j \inf_{\nu' \in \Sigma_j''} H_{\widehat{\beta}}(D_j S_j \overline{C}_j | R_j F_1^j G_1^j \widetilde{E})_{\nu'}, \tag{B.11}$$

where Σ_j'' denotes the set of all states of the form $W_j \left|\omega\right\rangle_{R_{j-1}E_{j-1}\widehat{C}_1^{j-1}F_1^{j-1}G_1^{j-1}\widetilde{E}}$ for some ω.

Substituting the duality relations (B.9), and noting that the data-processing inequality implies $\bar{H}_\alpha^\uparrow(D_1^n S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_\rho \leq \bar{H}_\alpha^\uparrow(D_1^n S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n \widetilde{E}_0)_\rho$, we get the desired bound (B.8). To get the claimed results in the theorem, we first use the inequalities $\bar{H}_\alpha \leq \bar{H}_\alpha^\uparrow$ and $\bar{H}_\alpha^\uparrow \leq \bar{H}_{2-\frac{1}{\alpha}}$ [118, Corollary 4] on the left- and right-hand-sides respectively, to obtain

$$\begin{aligned}\bar{H}_\alpha(D_1^n S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_\rho &\leq \sum_j \sup_{\nu' \in \Sigma_j'} \bar{H}_{\widetilde{\alpha}}(D_j S_j \overline{C}_j | \widehat{C}_1^j E_j)_{\nu'}, \text{ where } \widetilde{\alpha} = 2 - \frac{1}{2-\frac{1}{\alpha}} = \frac{3\alpha - 2}{2\alpha - 1} \\ &\leq \sum_j \sup_{\nu' \in \Sigma_j'} \bar{H}_{\widetilde{\alpha}}(D_j S_j \overline{C}_j | \widehat{C}_j E_j)_{\nu'}, \end{aligned} \tag{B.12}$$

where the second line holds by data-processing. These entropies can then be related to the QESP-entropies $\bar{H}_\alpha^{f_{\text{full}}}(S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_\rho$ and $\bar{H}_{\widetilde{\alpha}}^{f_{|j}}(S_j \overline{C}_j | \widehat{C}_j E_j)_\nu$ the same way as in Lemma 4.3.2. This yields the claimed bound (B.4) by taking $M \to \infty$. The other claimed bound (B.5) is also obtained a similar way to before: by Lemma 4.3.7, each $\hat{f}_{|j}$ satisfies

$$\sup_{\nu \in \Sigma_j} \bar{H}_{\widetilde{\alpha}}^{\hat{f}_{|j}}(S_j \overline{C}_j | \widehat{C}_j E_j)_\nu = \sup_{\nu \in \Sigma_j} \bar{H}_{\widetilde{\alpha}}^{f_{|j}}(S_j \overline{C}_j | \widehat{C}_j E_j)_\nu - \kappa_j = 0, \tag{B.13}$$

hence applying (B.4) to the QES-s $\hat{f}_{|j}$ yields (B.5). □

We now state the analogue of Theorem 4.4.1

Theorem B.0.2. *Consider a state ρ and channels $\{\mathcal{M}_j\}_{j=1}^n$ fulfilling the conditions described in Theorem B.0.1. Let Σ_j denotes the set of all states of the form $\mathcal{M}_j\left[\omega_{R_{j-1}E_{j-1}}\right]$ for some initial state $\omega \in \mathrm{S}_=(R_{j-1}E_{j-1})$. Furthermore, let Σ denote the convex range of channels $\{\mathcal{M}_j\}_{j=1}^n$, and let $\mathcal{M} \in \mathrm{CPTP}(\widetilde{Q}, SE\overline{C}\widehat{C})$ be a channel such that Σ is equal to the set of output states of $\mathcal{M}$ (i.e. $\left\{\mathcal{M}\left[\omega_{\widetilde{Q}E}\right] \,\middle|\, \omega \in \mathrm{S}_=(\widetilde{Q}\widetilde{E})\right\}$)[3]. Take any $\alpha \in (3/4, 1)$ and let $\overline{\alpha} = \frac{4\alpha-3}{3\alpha-2}$. Suppose furthermore that $\rho = p_\Omega \rho_{|\Omega} + (1-p_\Omega)\rho_{|\overline{\Omega}}$ for some $p_\Omega \in (0,1]$ and*

[3]Note that such a channel always exists (see Lemma 4.4.3 for an explicit construction).

normalized states $\rho_{|\Omega}, \rho_{|\overline{\Omega}}$. Let S_Ω be a convex set of probability distributions on the alphabet $\overline{\mathcal{C}} \times \widehat{\mathcal{C}}$, such that for all $\bar{c}_1^n \hat{c}_1^n$ with nonzero probability in $\rho_{|\Omega}$, the frequency distribution $\mathrm{freq}_{\bar{c}_1^n \hat{c}_1^n}$ lies in S_Ω. Then letting $\boldsymbol{\sigma}_{\overline{C}\widehat{C}}$ denote the distribution on $\overline{C}\widehat{C}$ induced by any state $\sigma_{\overline{C}\widehat{C}}$, we have

$$H_\alpha^\uparrow(S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\rho_{|\Omega}} \leq n \bar{h}_{\overline{\alpha}} + \frac{\alpha}{1-\alpha} \log \frac{1}{p_\Omega},$$

$$\text{where} \quad \bar{h}_{\overline{\alpha}} = \sup_{\mathbf{q} \in S_\Omega} \sup_{\nu \in \Sigma} \left(-\frac{1}{1-\overline{\alpha}} D\left(\mathbf{q} \middle\| \boldsymbol{\nu}_{\overline{C}\widehat{C}}\right) - \sum_{\bar{c}\hat{c} \in \mathrm{supp}(\boldsymbol{\nu}_{\overline{C}\widehat{C}})} q(\bar{c}\hat{c}) \bar{D}_{\overline{\alpha}}\left(\nu_{SE \wedge \bar{c}\hat{c}} \| \mathbb{I}_S \otimes \nu_{E \wedge \hat{c}}\right) \right), \tag{B.14}$$

and the objective function in the above supremum is jointly concave in ν and $\mathbf{q}$. Furthermore, for any $\alpha', \alpha'' \in (1/2, 1)$ such that $\frac{\overline{\alpha}}{\overline{\alpha}-1} = \frac{\alpha'}{\alpha'-1} + \frac{\alpha''}{\alpha''-1}$, we have

$$\bar{h}_{\overline{\alpha}} \leq \sup_{\mathbf{q} \in S_\Omega} \sup_{\nu \in \Sigma} \left(-\frac{\alpha''}{1-\alpha''} D\left(\mathbf{q} \middle\| \boldsymbol{\nu}_{\overline{C}\widehat{C}}\right) + H_{\alpha'}(S\overline{C}|\widehat{C}E)_\nu \right), \tag{B.15}$$

and the objective function in the above supremum is jointly concave in ν and $\mathbf{q}$.

Proof. The proof is analogous to the proof of Theorem 4.4.1; we highlight the key steps. Starting from the intermediate bound (B.8), we have

$$\begin{aligned}
H_\alpha^\uparrow(D_1^n S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_\rho &\leq \bar{H}^\uparrow_{2-\frac{1}{\alpha}}(D_1^n S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_\rho \\
&\leq \sum_j \sup_{\nu' \in \Sigma'_j} \bar{H}^\uparrow_{\widetilde{\alpha}}(D_j S_j \overline{C}_j | \widehat{C}_1^j E_j)_{\nu'}, \ \text{where } \widetilde{\alpha} = 2 - \frac{1}{2-\frac{1}{\alpha}} = \frac{3\alpha - 2}{2\alpha - 1} \\
&\leq \sum_j \sup_{\nu' \in \Sigma'_j} \bar{H}_{\overline{\alpha}}(D_j S_j \overline{C}_j | \widehat{C}_1^j E_j)_{\nu'}, \ \text{where } \overline{\alpha} = 2 - \frac{1}{\widetilde{\alpha}} = \frac{4\alpha - 3}{3\alpha - 2} \\
&\leq \sum_j \sup_{\nu' \in \Sigma'_j} \bar{H}_{\overline{\alpha}}(D_j S_j \overline{C}_j | \widehat{C}_j E_j)_{\nu'},
\end{aligned} \tag{B.16}$$

where the first follows from $H_\alpha^\uparrow \leq \bar{H}^\uparrow_{2-\frac{1}{\alpha}}$ [118, Corollary 4], the second line is (B.8), third lines is due to $H_\alpha^\uparrow \leq \bar{H}^\uparrow_{2-\frac{1}{\alpha}}$ [118, Corollary 4], and the fourth line is by data-processing.

We now relate the $H_\alpha^\uparrow(D_1^n S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_\rho$ to $H_\alpha^\uparrow(S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\rho_{|\Omega}}$ in similar fashion to the

Corollary 4.3.1 proof. Let $\widetilde{\Omega}$ denote the set of all $\bar{c}_1^n \hat{c}_1^n$ values such that $\mathrm{freq}_{\bar{c}_1^n \hat{c}_1^n} \in S_\Omega$:

$$\begin{aligned} H_\alpha^\uparrow(S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\rho_{|\Omega}} &\leq H_\alpha^\uparrow(D_1^n S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\rho_{|\Omega}} - \min_{\bar{c}_1^n \hat{c}_1^n \in \widetilde{\Omega}} H_\alpha(D_1^n)_{\rho_{|\bar{c}_1^n \hat{c}_1^n}} \\ &\leq H_\alpha^\uparrow(D_1^n S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\rho} - \min_{\bar{c}_1^n \hat{c}_1^n \in \widetilde{\Omega}} H_\alpha(D_1^n)_{\rho_{|\bar{c}_1^n \hat{c}_1^n}} + \frac{\alpha}{1-\alpha} \log \frac{1}{p_\Omega} \\ &\leq H_\alpha^\uparrow(D_1^n S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\rho} + \max_{\bar{c}_1^n \hat{c}_1^n \in \widetilde{\Omega}} -H_\alpha(D_1^n)_{\rho_{|\bar{c}_1^n \hat{c}_1^n}} + \frac{\alpha}{1-\alpha} \log \frac{1}{p_\Omega}, \end{aligned} \tag{B.17}$$

where the first line holds by similar arguments to the Lemma 4.3.9 proof, and the second line is [41, Lemma B.5] in the $\alpha < 1$ regime.

Picking all the QES-s $f_{|j}$ to be equal to a single QES f, we can relate the entropies of the D_j registers to the QES values the same way as in Lemma 4.3.2. Taking the $M \to \infty$ limit in Lemma 4.3.2, we get an analogue of Corollary 4.4.1:

$$\begin{aligned} H_\alpha^\uparrow(S_1^n \overline{C}_1^n | \widehat{C}_1^n E_n)_{\rho_{|\Omega}} &\leq n \sup_{\nu \in \Sigma} \bar{H}_{\bar{\alpha}}^f (S\overline{C}|\widehat{C}E)_\nu + \max_{\bar{c}_1^n \hat{c}_1^n \in \widetilde{\Omega}} \sum_{j=1}^n f(\bar{c}_j \hat{c}_j) + \frac{\alpha}{1-\alpha} \log \frac{1}{p_\Omega}, \\ &\leq \sup_{\mathbf{q} \in S_\Omega} \sup_{\nu \in \Sigma} \left(\bar{H}_{\bar{\alpha}}^f (S\overline{C}|\widehat{C}E)_\nu + \mathbf{f} \cdot \mathbf{q} \right) n + \frac{\alpha}{1-\alpha} \log \frac{1}{p_\Omega}, \end{aligned} \tag{B.18}$$

The best upper bound would be given by minimizing over the choice of QES f on the right-hand-side; it hence only remains to show that this yields the claimed formula for $\bar{h}_{\bar{\alpha}}$, analogous to the strong duality proof in Lemma 4.4.5. To do so, we define a sign-flipped version of $G_{\alpha,\rho}$ from Lemma 4.4.4, with Petz instead of sandwiched divergences:

$$\widehat{G}_{\alpha,\rho}(\mathbf{f}) := \bar{H}_\alpha^f (Q\overline{C}|\widehat{C}Q')_\rho. \tag{B.19}$$

From the log-sum-exponential formulation (4.84) (with a sign flip) we see that $\widehat{G}_{\alpha,\rho}(\mathbf{f})$ is again convex in $\mathbf{f}$ since $\alpha < 1$; by analogous calculations we find that its convex conjugate is

$$\widehat{G}^*_{\alpha,\rho}(\boldsymbol{\lambda}) := \begin{cases} \frac{1}{1-\alpha} D\left(-\boldsymbol{\lambda} \middle\| \boldsymbol{\rho}_{\overline{C}\widehat{C}}\right) - \sum_{\bar{c}\hat{c} \in \mathrm{supp}(\boldsymbol{\rho}_{\overline{C}\widehat{C}})} \lambda(\bar{c}\hat{c}) \bar{D}_\alpha \left(\rho_{QQ' \wedge \bar{c}\hat{c}} \| \mathbb{I}_Q \otimes \rho_{Q' \wedge \hat{c}}\right) & \text{if } -\boldsymbol{\lambda} \in \mathbb{P}_{\overline{C}\widehat{C}}, \\ +\infty & \text{otherwise.} \end{cases} \tag{B.20}$$

Again, since $\widehat{G}_{\alpha,\rho}$ is a convex function with domain $\mathbb{R}^{|\overline{C} \times \widehat{C}|}$, this means $\widehat{G}_{\alpha,\rho}$ is also the convex conjugate of $\widehat{G}^*_{\alpha,\rho}$, thus we have $\widehat{G}_{\alpha,\rho}(\mathbf{f}) = \sup_{\boldsymbol{\lambda} \in \mathbb{R}^{|\overline{C} \times \widehat{C}|}} \left(\boldsymbol{\lambda} \cdot \mathbf{f} - \widehat{G}^*_{\alpha,\rho}(\boldsymbol{\lambda}) \right)$. With this,

we again write

$$\begin{aligned}
&\bar{H}^f_{\overline{\alpha}}(S\overline{C}|\widehat{C}E)_\nu + \mathbf{f}\cdot\mathbf{q} \\
=&\widehat{G}_{\overline{\alpha},\nu}(\mathbf{f}) + \mathbf{f}\cdot\mathbf{q} \\
=&\sup_{\boldsymbol{\lambda}\in\mathbb{R}^{|\overline{\mathcal{C}}\times\widehat{\mathcal{C}}|}} \Big(\boldsymbol{\lambda}\cdot\mathbf{f} - \widehat{G}^*_{\overline{\alpha},\nu}(\boldsymbol{\lambda}) + \mathbf{f}\cdot\mathbf{q}\Big) \\
=&\sup_{\boldsymbol{\lambda}'\in\mathbb{P}_{\overline{C}\widehat{C}}} \Big(-\widehat{G}^*_{\overline{\alpha},\nu}(-\boldsymbol{\lambda}') + \mathbf{f}\cdot(\mathbf{q}-\boldsymbol{\lambda}')\Big) \text{ via a reparametrization } \boldsymbol{\lambda}' = -\boldsymbol{\lambda}.
\end{aligned} \tag{B.21}$$

Furthermore, by Lemma 4.3.1 (adapted to the Petz case) we know there exists a read-and-prepare channel that can extend any state ν so that we have $\bar{H}^f_{\overline{\alpha}}(S\overline{C}|\widehat{C}E)_\nu = \bar{H}_{\overline{\alpha}}(DS\overline{C}|\widehat{C}E)_\nu - M$ for some fixed M; it thus follows from the convexity of Petz divergences for $\alpha \in (0,1]$ that $\bar{H}^f_{\overline{\alpha}}(S\overline{C}|\widehat{C}E)_\nu$ is concave in ν. This lets us proceed similarly to the Lemma 4.4.5 proof: we see $\widehat{G}^*_{\overline{\alpha},\nu}(\boldsymbol{\lambda})$ is jointly convex in $(\boldsymbol{\lambda},\nu)$, since by definition it is a supremum over a family of functions $\boldsymbol{\lambda}\cdot\mathbf{f} - \widehat{G}_{\overline{\alpha},\nu}(\mathbf{f})$ that are each jointly convex in $(\boldsymbol{\lambda},\nu)$. Therefore, $-\widehat{G}^*_{\overline{\alpha},\nu}(-\boldsymbol{\lambda}') + \mathbf{f}\cdot(\mathbf{q}-\boldsymbol{\lambda}')$ is jointly concave in $(\mathbf{q},\nu,\boldsymbol{\lambda}')$, which allows us to apply the same strong duality arguments to conclude that

$$\begin{aligned}
&\inf_{\mathbf{f}} \sup_{\mathbf{q}\in S_\Omega} \sup_{\nu\in\Sigma} \sup_{\boldsymbol{\lambda}'\in\mathbb{P}_{\overline{C}\widehat{C}}} \Big(-\widehat{G}^*_{\overline{\alpha},\nu}(-\boldsymbol{\lambda}') + \mathbf{f}\cdot(\mathbf{q}-\boldsymbol{\lambda}')\Big) \\
=&\sup_{\mathbf{q}\in S_\Omega} \sup_{\nu\in\Sigma} -\widehat{G}^*_{\overline{\alpha},\nu}(-\mathbf{q}).
\end{aligned} \tag{B.22}$$

This gives the first claimed result (B.14), by just substituting the definition of $-\widehat{G}^*_{\overline{\alpha},\nu}(-\mathbf{q})$. To get the looser bound (B.15), we perform similar calculations except with the relaxation

$$H_{\overline{\alpha}}(DS\overline{C}|\widehat{C}E)_\nu \le H_{\alpha'}(S\overline{C}|\widehat{C}E)_\nu + H^\uparrow_{\alpha''}(D|S\overline{C}\widehat{C}E)_\nu, \tag{B.23}$$

which is a special case of [38, Proposition 8]; this lets us apply the same arguments as the Lemma 4.4.2 proof. □