From Hypothesis Testing of Quantum Channels to Secret Sharing

by

Farzin Salek Shishavan

under supervision of
Prof. Dr. Javier Rodríguez Fonollosa
and
Prof. Dr. Andreas Winter

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in

Universitat Politècnica de Catalunya (UPC)
Departamento de Teoría de la Señal y Comunicaciones

Committee in charge:
Prof. Dr. Antonio Acin
Prof. Dr. Renato Renner
Prof. Dr. Mark M. Wilde

Barcelona 2020

Abstract

The present thesis has three major thrusts: the first thrust presents a broad investigation of asymptotic binary hypothesis testing, when each hypothesis represents asymptotically many independent instances of a quantum channel. Unlike the familiar setting of quantum states as hypotheses, there is a fundamental distinction between adaptive and non-adaptive strategies with respect to the channel uses, and we introduce a number of further variants of the discrimination tasks by imposing different restrictions on the test strategies. The following results are obtained: (1) The first separation between adaptive and non-adaptive symmetric hypothesis testing exponents for quantum channels, which we derive from a general lower bound on the error probability for non-adaptive strategies. (2) We prove that for classical-quantum channels, adaptive and non-adaptive strategies lead to the same error exponents both in the symmetric (Chernoff) and asymmetric (Hoeffding) settings. (3) We prove that, in some sense generalizing the previous statement, for general channels adaptive strategies restricted to classical feed-forward and product state channel inputs are not superior to non-adaptive product state strategies. (4) As application of our findings, we address the discrimination power of quantum channels and show that neither adaptive strategies nor input quantum memory can increase the discrimination power of an entanglement-breaking channel.

In the second thrust, we construct new protocols for the tasks of converting noisy multipartite quantum correlations into noiseless classical and quantum ones using local operations and classical communications (LOCC). For the former, known as common randomness (CR) distillation, two new lower bounds are obtained. Our proof relies on a generalization of communication for omniscience (CO). Our contribution here is a novel simultaneous decoder for the compression of correlated classical sources by random binning with quantum side information at the decoder. For the latter, we derive two new lower bounds on the rate at which Greenberger-Horne-Zeilinger (GHZ) states can be asymptotically distilled from any given pure state under LOCC. Our approach consists in "making coherent" the proposed CR distillation protocols and recycling of resources.

The final thrust studies communication over a single-serving two-receiver quantum broadcast channel with legitimate receiver and eavesdropper. We find inner and outer boundary regions for the tradeoff between common, individualized, confidential messages as well as the rate of the dummy randomness used for obfuscation. As applications, we find one-shot capacity bounds on the simultaneous transmission of classical and quantum information and re-derive a number of asymptotic results in the literature.

Resum

Aquesta tesi està estructurada en tres eixos: A la primera part es presenta una investigació extensa dels test d'hipòtesis binaris asimptòtics quan cada hipòtesi representa diferents instàncies independents d'un canal quàntic. A diferència del cas habitual ja conegut en el qual els estats quàntics es modelen com a hipòtesi, en aquest treball es distingeix entre estratègies adaptatives i no adaptatives pel que fa a l'ús del canal, i es presenten una sèrie de variants addicionals de les tasques de discriminació imposant diferents restriccions a la estratègies de test. S'obtenen els següents resultats: (1) S'obté per primera vegada una separació entre els exponents de test d'hipòtesis simètrics adaptatius i no adaptatius, derivant per primera vegada una fita inferior de la probabilitat d'error per estratègies no adaptatives. (2) Es demostra que per a canals quàntics clàssics, les estratègies adaptatives i no adaptatives condueixen als mateixos exponents d'error tant en el cas simètric (Chernoff) com en configuracions asimètriques (Hoeffding); (3) Es demostra, generalitzant el resultat anterior, que en general estratègies adaptatives restringides a feed-forward clàssic i entrades de tipus producte no són superiors a estratègies de tipus producte no adaptatives; (4) Com a aplicació dels resultats anteriors s'aborda la discriminació de potència en canals quàntics. Es demostra que, ni estratègies adaptatives, ni la utilització d'entrades amb memòria, permeten millorar la discriminació de potència de canals de tipus entanglement-breaking.

A la segona part de la tesi es construeixen nous protocols per convertir correlacions quàntiques sorolloses en correlacions clàssiques, o bé en correlacions quàntiques, ambdues lliures de soroll, mitjançant la utilització d'operacions locals i comunicacions clàssiques (LOCC). Per a la primera tasca, coneguda com destil·lació d'aleatorietat comú (CR), s'obtenen dos nous límits inferiors de l'aleatorietat comuna destil·lable. Aquest treball suposa una generalització de la comunicació per a l'omnisciència. En la segona tasca, s'obtenen dos nous límits inferiors de la taxa a la qual els estats Greenberger-Horne-Zeilinger (GHZ) poden destil·lar asimptòticament, des de qualsevol estat pur, utilitzant LOCC. L'enfocament consisteix a fer coherent el protocol de destil·lació CR proposat així com en la reutilització de recursos.

L'última part de la tesi final estudia la comunicació mitjançant un sol ús de canal quàntic en presència d'un receptor legítim i d'un observador no autoritzat. S'obtenen regions interiors i exteriors associades a el compromís entre la taxa de transmissió comú, confidencial, individualitzada i de la font aleatòria.

Resumen

Esta tesis está estructurada en tres ejes: En la primera parte se presenta una investigación extensa de los test de hipótesis binarios asintóticos cuando cada hipótesis representa diferentes instancias independientes de un canal cuántico. A diferencia del caso habitual ya conocido en el que los estados cuánticos se modelan como hipótesis, en este trabajo se distingue entre estrategias adaptativas y no adaptativas con respecto al uso del canal, y se presentan una serie de variantes adicionales de las tareas de discriminación imponiendo diferentes restricciones a la estrategias de test. Se obtienen los siguientes resultados: (1) Se obtiene por primera vez una separación entre los exponentes de test de hipótesis simétricos adaptativos y no adaptativos, derivándose por primera vez una cota inferior de la probabilidad de error para estrategias no adaptativas. (2) Se demuestra que para canales cuánticos clásicos, las estrategias adaptativas y no adaptativas conducen a los mismos exponentes de error tanto en el caso simétrico (Chernoff) como en configuraciones asimétricas (Hoeffding); (3) Se demuestra, generalizando el resultado anterior, que en general estrategias adaptativas restringidas al feedforward clásico y entradas de tipo producto no son superiores a estrategias de tipo producto no adaptativas; (4) Como aplicación de los resultados anteriores se aborda la discriminación de potencia en canales cuánticos. Se demuestra que, ni estrategias adaptativas, ni la utilización de entradas con memoria, permiten mejorar la discriminación de potencia de canales del tipo entanglement-breaking.

En la segunda parte de la tesis se construyen nuevos protocolos para convertir correlaciones cuánticas ruidosas en correlaciones clásicas ,o bien en correlaciones cuánticas, ambas libres de ruido, mediante la utilización de operaciones locales y comunicaciones clásicas (LOCC). Para la primera tarea, conocida como destilación de aleatoriedad común (CR), se obtienen dos nuevos límites inferiores de la aleatoriedad común destilable. Este trabajo supone una generalización de la comunicación para la omnisciencia. En la segunda tarea, se obtienen dos nuevos límites inferiores de la tasa a la que los estados Greenberger-Horne-Zeilinger (GHZ) pueden destilarse asintóticamente, desde cualquier estado puro, utilizando LOCC. El enfoque consiste en hacer coherente el protocolo de destilación CR propuesto así como en la reutilización de recursos.

La última parte de la tesis final estudia la comunicación mediante un sólo uso del canal cuántico en presencia de un receptor legítimo y de un observador no autorizado. Se obtienen regiones interiores y exteriores asociadas al compromiso entre la tasa de transmisión común, confidencial, individualizada y de la fuente aleatoria.

Acknowledgements

First, I would like to thank my two supervisors, Javier R. Fonollosa and Andreas Winter. Many thanks are due to Javier R. Fonollosa for accepting to be my supervisor, which allowed me to quickly enter the PhD program, who has always impressed me by his kindness and modest disposition. I am indebted to Andreas Winter whose door has always been open to me, especially in the last 2 years, I have been recklessly taking a lot of his time. It is from him whatever I have learned about quantum information; on the personal side, Andreas has passed on life lessons to me. I cannot forget his passionate support during difficult times. I also have worked under the supervision of Min-Hsiu Hsieh, although we never did the necessary paperwork to have him as an official supervisor. In the beginning of my PhD that I only barely knew the basics of quantum information, Min-Hsiu offered me a problem and walked me through things helping me to solve it; during which he supported my visit to his group in Sydney, I am very grateful to him for all his support and kindness.

Special thank you goes to Masahito Hayashi: During my time in Shenzhen, he was always available to discuss problems and answer my questions kindly and patiently no matter how many times I repeated the same question! And working with him remotely, he immediately replied to my questions and was up for online meetings. I should also thank his support of my visit to China, which happened to be funded by Peng Cheng Laboratories.

I feel honored and privileged to present this thesis before distinguished scientists of the field: Profs. Antonio Acin, Renato Renner and Mark M. Wilde; I thank them for accepting to examine this thesis.

During my time in Shenzhen, my good friend Kun Wang always offered his help in every single possible way, I am immensely thankful to him. I would like to express my gratitude to Anurag Anshu, Mario Berta, Marco Tomamichel, Shun Watanabe and Mark M. Wilde for fruitful discussions. I want to extend my gratitude to all members of our groups in UPC and UAB, past and present, for making my time very enjoyable. I must also thank many friends who made life in Barcelona so enjoyable, you know who you are, I love you!

به پدر و مادرم که سرگردانی و ترس در پناهشان به شجاعت می گراید

which is in Persian, to English it translates: to my parents, my refuge and strength.

Finally and above all, there is Zahra, my constant rock throughout the whole ordeal.

Contents

Chapter 1

Introduction

1.1 Background and motivation

In 1948, Bell Labs scientist Claude E. Shannon published a paper entitled "A Mathematical Theory of Communication" [1] underlying the foundation of a major field known today as *information theory*. Information theory expresses in very simple-looking terms what is possible to communicate, what is possible to do with information, with data. The field intersects a number of other fields including electrical engineering, computer science, physics and mathematics. In Shannon's definition of information, it is not a physical entity but an abstract mathematical concept, and therefore hard to quantify in general.

When human factors come into play, a piece of information that is to be communicated over or be stored on certain media, must follow the laws of physics. On the other hand, existence of a piece of information lies on its physical support: while Shannon's theory essentially assumes a classical physical support, once the media are on the scales of microscopic physical systems, the communicated or stored information behaves much differently; this requires translation of information-theoretic language to another language featuring all those specific features of microscopic physical systems. Discovered in the early twentieth century, quantum theory is the name of this language, essentially and fundamentally consisting of a set of postulates that governs the physics of elementary particles, i.e. how particles on the scale of atoms evolve dynamically in nature as time progresses. The quantum theory has a number of aspects with no classical counterparts. The objects of classical theory consist of bits, i.e. combination of zeros and ones; however, quantum counterpart of a bit, a *qubit*, can be a zero and a one at the same time. In more technical terms, a quantum particle can be in a lin-

ear combination of two or more allowable quantum states. The other central issue in quantum theory is that non-orthogonal states cannot be perfectly discriminated. Another striking aspect of quantum theory is entanglement, referring to the strong quantum correlations between multiple parties. Common or secret bits shared by multiple parties are perhaps the closest counterparts to this quantum phenomena; however, entanglement offers such strong correlations that do not exit in classical world.

The synthesis of information theory and quantum mechanics, i.e. a mathematical theory of communication common to information theory and quantum theory, is now called "quantum Shannon theory", basically occupied with generalizing and applying (Shannon's) classical contributions to the quantum setting. The difficulties usually are faced in these extensions stem from non-commutativity of quantum mechanics. However, it is important to note that there are intrinsically quantum information-processing tasks that are not solely non-commuting versions of certain classical tasks. As such, quantum Shannon theory covers a broader field than a non-commutative analogue of classical information theory. The key to this distinction is entanglement, whose presence gives rise to several quantum information processing protocols.

This thesis is mostly built around studying the quantum counterparts of several classical contributions; the three major thrusts of the thesis can be cast as follows:

- Hypothesis testing of quantum channels
- Distillation of common randomness and entanglement
- Transmission over single-serving quantum broadcast channel

While all three topics are generically applications and explanations of von Neumann entropy, there is no connection between them from the perspective of problem solving. Therefore I provide an introduction for each topic separately below.

1.1.1 Hypothesis Testing of Quantum Channels

Arguably, hypothesis testing is one of the most fundamental primitives in both classical and quantum information processing. It is such a central task because a variety of other information processing problems can be cast in the framework of hypothesis testing; both direct coding theorems and converses can be reduced to it. In binary hypothesis testing, the two hypotheses are usually referred to as null and alternative hypotheses and accordingly,

two error probabilities are defined: type-I error due to a wrong decision in favour of the alternative hypothesis (while the truth corresponds to the null hypothesis) and type-II error due to the alternative hypothesis is rejected despite being correct. The overall objective of the hypothesis testing is to minimize the error probability in identifying the hypotheses. Depending on the significance attributed to two types of errors, several settings can be distinguished. An historical distinction is between the symmetric and the asymmetric hypothesis testing: in symmetric hypothesis testing, the goal is to minimize both error probabilities simultaneously, while in asymmetric hypothesis testing, the goal is to minimize one type of error probability subject to a constraint on the other type of error probability. This description of the problem presupposes that the two hypotheses correspond to objects in a probabilistic framework, in which also the possible tests (decision rules) are phrased, so as to give unambiguous meaning to the type-I and type-II error probabilities. The traditionally studied framework is that each hypothesis represents a probability distribution on a given set, and more generally a state on a given quantum system.

The setting of state discrimination is naturally extended to the discrimination of quantum channels, i.e. the two hypotheses are two quantum channels and the task follows by feeding quantum states into the unknown channel and performing measurements on the outputs. Despite the inherent similarities between state and channel discrimination, the setting of the latter problem is fundamentally richer and considerably more difficult. This complexity basically stems from the more general "adaptive" strategies that can be devised in the channel setting: if only product-state inputs are allowed to the unknown channel and measurement is performed at the end, the problem follows trivially from the traditional state discrimination setting; however, allowing entangled state inputs or feed-forward information after performing each measurement, gives rise to strategies whose error performance have been widely open until recently.

1.1.2 Common Randomness and Entanglement Distillation

Interconversion between various resources is one of the big ongoing programs of quantum and classical information theory for a considerable time [2]. Within that broad class of questions, the transformations of multipartite quantum states into other forms has provided considerable inspiration. A particularly prototypical example of this is bipartite entanglement of pure states: in the asymptotic setting of many copies of a pure state, not only can

each pure state $|\psi\rangle^{AB}$ be converted to EPR states $|\phi\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle)$ at rate $E(\psi) = S(A)_\psi$ by local operations and classical communication (LOCC), where $S(A)_\rho = -\operatorname{Tr}\rho^A \log \rho^A$ is the von Neumann entropy of the reduced state of a quantum state ρ^{AB}, the same rate governs the reverse transformation from ϕ to ψ [3]. The story is far less satisfying for mixed states [4], nevertheless this raised certain expectations for multipartite pure states: while it is clear that there cannot be a single "gold standard" like the EPR state in the bipartite setting – as EPR states between any pair of m parties are inequivalent to EPR states between any other pair –, the question arose whether there is a "minimal reversible entanglement generating set" (MREGS) [5]. In the latter paper, it was shown that for $m \geq 4$ parties, also the GHZ state $|\Gamma_m\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes m} + |1\rangle^{\otimes m})$ needs to be part of an MREGS, and in [6] this was extended to $m = 3$. Since then, increasing lower bounds on the size of an MREGS have been proved, and it is conceivable that any MREGS is infinitely large. For a broad overview over the history and state of the art in multipartite entanglement, see the review [7].

In any case, the frustrated hope of the MREGS programme has made researchers reevaluate what we actually want from our theory of state conversions. One big component, rather than a universal normal form, is knowledge how, and how efficiently, to transform a given m-partite pure state $|\psi\rangle^{A_1\ldots A_m}$ into a specific desired target state. In the multipartite setting, this presents a problem of choice. There seem to be at least two canonical options: first, aim for EPR states between designated pairs of parties, and second, an m-party GHZ state. The first problem has an elegant solution, based on quantum state merging [8]. If EPR states are to be distilled between a specific pair of parties, say i and j, then the optimal rate (capacity) is the following number [9]:

$$C_{\mathrm{EPR}(i:j)}(\psi) = \min_I S(A_I)_\psi \text{ s.t. } i \in I \subseteq [m] \setminus j. \tag{1.1}$$

If we want to distill EPR states between different pairs of parties simultaneously, there are partial results, for example outer rate bounds from the subgroup entropies, all of which are monotones [5, Lemma 1 & Thm. 2], i.e. each $S(A_I)$, for $I \subseteq [m]$, is a monotone under asymptotic LOCC; furthermore, [6] gives asymptotic monotones for certain state conversions based on the quantum relative entropy. And there is the "entanglement combing" protocol that yields EPR pairs between a single party and each of the other $m-1$ [10]. These tasks of creating pairwise (EPP type) entanglement between nodes, assisted by the others, is very much tied to the objectives of the so-called quantum internet [11, 12]. As for GHZ distillation, also this is evidently relevant for the quantum internet, but has received considerably

less attention; we review some of the relevant prior work later in this chapter.

1.1.3 Transmission over Single-Serving Quantum Broadcast Channel

Shannon modeled a noisy (classical) channel as a stochastic map $\mathcal{W}_{X\to Y}$ taking classical inputs to classical outputs according to some probability distribution, $p_{Y|X}(y|x)$ [1]. In his paper, he defined and computed the fundamental feature of a channel, its capacity: the amount of classical information, i.e., bits, that can be reliably transmitted from a sender to a remote receiver over a classical channel. In the limit of many independent uses of a stationary memoryless channel, Shannon showed that its capacity in bits per use of the channel is equal to the mutual information between the input and output.

Since nature is fundamentally quantum, it seemed necessary to enhance or replace Shannon's channel model with a *quantum channel* model that takes quantum mechanics into account. Many years after Shannon in the context of quantum information theory, a quantum channel is modelled by a completely-positive trace-preserving map (cptp) with possibly different input and output Hilbert spaces. Denoted by $\mathcal{N}_{A\to B}$, a quantum channel with input and output systems A and B respectively, can now be used to accomplish a variety of information-processing tasks and accordingly different capacities can be defined. One such task is communication of secret information and the utmost ability of a channel to communicate information in a secure, secret fashion is called its secrecy capacity.

Security is one of the most important issues in communications. The information theoretic approach to achieving secure communication was initiated by Wyner under the name "wiretap channel" [13]: a communication system in which a sender aims to transmit a message reliably to a receiver while hiding it from an eavesdropper. The basic idea underlying Wyner's coding scheme is to generate a sufficiently large number of random sequences and position them into bins labeled with the messages to be transmitted. To send a message, a sequence from the message bin is randomly selected and transmitted. In the original model of Wyner, the eavesdropper is assumed to be at a physical disadvantage with respect to the legitimate receiver, meaning that, upon transmission over the channel, the eavesdropper only receives what can be regarded as a noisy version of the information received by the legitimate receiver. This model is usually referred to as the physically degraded wiretap channel. This channel description was later enhanced by Csiszár and Körner [14] by introducing a public message that is piggybacked on top of the confidential message and that is to be reliably decoded by both

receivers. Furthermore, in this new model called broadcast channel with confidential messages (BCC), the legitimate receiver has no specific physical advantage over the eavesdropper. The coding scheme of the BCC consists of superposition coding [15] to encode the confidential message on top of the common message in combination with Wyner's codebook structure with local randomness for equivocation. The most important contribution of Csiszár and Körner consists of prepending a prefixing stochastic map to the channel and using a then-new single-letterization trick in the converse proof.

The implementation of the prefixing stochastic map is performed from random numbers using a method such as channel simulation [16]. This means that two sources of randomness are required for BCC coding, one for random codeword selection and another for channel simulation. Traditionally randomness has been assumed to be an unlimited resource. In practice, however, a limited randomness rate is reasonable, potentially compromising the simultaneous reliability and secrecy criteria. Csiszár and Körner [17] later proposed an alternative description of the BCC where the message to be transmitted consists of two *independent* parts, a confidential part defined in the same sense as the original BCC and a non-secret or individualized part, i.e. a message without any secrecy requirement placed on it, potentially and partially playing the role of a source of randomness.

The quantum generalisation of the wiretap channel was studied in [18] and [19], where the capacity for the transmission of confidential classical information was given by a regularized formula. The ability of the quantum channels to preserve quantum superpositions gives rise to purely quantum information processing tasks with no classical counterparts. The quantum capacity, i.e. the ability of a quantum channel to transmit qubits, is one such example. The unified task of transmission of the classical and quantum information was studied in [20] and simultaneously achievable rates were proven. The protocol of [20] is conceptually related to the superposition coding, where, for each classical message a different quantum code is used and the capacity region is given in the form of a regularized rate region.

1.2 Notation and preliminaries

In this section, we introduce some conventions, notation and facts that we use throughout this thesis. A, B, C, etc. denote quantum systems, but also their corresponding Hilbert space. We identify states ρ with their density operators and use superscripts to denote the systems on which the mathematical objects are defined. The set of linear operators on A is denoted by $\mathcal{L}^A$, the set of positive semi-definite operators acting on A is denoted by $\mathcal{P}^A$, the set of

density matrices on A is written as $\mathcal{S}^A$ and the set of subnormalized states, i.e. $\{\rho \in \mathcal{P}^A | \operatorname{Tr}\rho \leq 1\}$, as $\mathcal{S}_{\leq}^A$. When talking about tensor products of spaces, we may habitually omit the tensor sign, so $A \otimes B = AB$, etc. The capital letters X, Y, etc. denote random variables whose realizations and the alphabets will be shown by the corresponding small and calligraphic letters, respectively: $X = x \in \mathcal{X}$. All Hilbert spaces and ranges of variables may be infinite; the dimension of a Hilbert space A is denoted $|A|$, as is the cardinality $|\mathcal{X}|$ of a set $\mathcal{X}$. For a state $\rho \in \mathcal{S}^{AB}$ of the composite system AB, the partial trace over system A (resp. B) is denoted by Tr_A (resp. Tr_B). We denote the identity operator by I. We use log and ln to denote base 2 and natural logarithms, respectively.

Moving on to quantum channels, these are linear, completely positive and trace preserving maps $\mathcal{N} : \mathcal{S}^A \to \mathcal{S}^B$ for two quantum systems A and B; $\mathcal{N}$ extends uniquely to a linear map from trace class operators on A to those on B. We often denote quantum channels, by slight abuse of notation, as $\mathcal{N} : A \to B$. According to Stinespring's factorization theorem [21], if $\mathcal{N} : A \longrightarrow B$ is a cptp map, then it can be dilated to the isometry $U_{\mathcal{N}} : A \hookrightarrow BW$ with W as the environment system such that $\mathcal{N}(\rho) = \operatorname{Tr}_W(U_{\mathcal{N}}\rho U_{\mathcal{N}}^\dagger)$ The ideal, or identity, channel on A is denoted id_A. Note furthermore that a state ρ^A on a system A can be viewed as a quantum channel $\rho : 1 \to A$, where 1 denotes the canonical one-dimensional Hilbert space, isomorphic to the complex numbers $\mathbb{C}$, which interprets a state operationally consistently as a state preparation procedure. For any positive integer m, we use the notation $[m] = \{1, \ldots, m\}$. For conciseness, we denote the tuple $(X_1, \ldots, X_m)$ by $X_{[m]}$ and also $\vec{X}_m \coloneqq (X_1, \cdots, X_m)$; while the former is used in Chapter 3, the latter is more convenient to use in Chapter 2. More generally, for a set L, we write $X_L = (X_i : i \in L)$, subsequently for a pair of integers $i \leq j$, $[i : j] \coloneqq \{i, i+1, ..., j\}$.

The normalized trace distance between two states ρ and σ is given as $\frac{1}{2}\|\rho - \sigma\|_1$ and the fidelity between them is defined as:

$$F(\rho, \sigma) \coloneqq \|\sqrt{\rho}\sqrt{\sigma}\|_1$$

The fidelity relates to the quantum relative entropy in the following way [22]:

$$F^2(\rho, \sigma) \geq 2^{-D(\rho\|\sigma)}. \tag{1.2}$$

The definition of the fidelity can be extended to subnormalized states, where the generalized fidelity is defined for subnormalized states $\tau, \nu \in \mathcal{S}_{\leq}$ as follows [23]

$$\bar{F}(\tau, \nu) = F(\tau, \nu) + \sqrt{(1 - \operatorname{Tr}\tau)(1 - \operatorname{Tr}\nu)}.$$

It is easily seen that the generalized fidelity reduces to the fidelity if at least one of the states is normalized.

We frequently use the following properties of the trace distance:

- The (normalized) trace distance (res. fidelity) is a convex (res. concave) function. For two ensembles $\{p(x), \rho_A^x\}$ and $\{p(x), \sigma_A^x\}$, where $\rho_A^x, \sigma_A^x \in \mathcal{D}(\mathcal{H}_A)$ for all x, let $\rho_{XA} \coloneqq \sum_x p(x) |x\rangle\langle x| \otimes \rho_A^x$ and $\sigma_{XA} \coloneqq \sum_x p(x) |x\rangle\langle x| \otimes \sigma_A^x$ be the associated classical-quantum (CQ) states, respectively. Then,

$$\left\| \sum_x p(x)\rho_A^x - \sum_x p(x)\sigma_A^x \right\|_1 \leq \sum_x p(x) \|\rho_A^x - \sigma_A^x\|_1.$$

  Moreover, the following property can be easily checked:

$$\|\rho_{XA} - \sigma_{XA}\|_1 = \sum_x p(x) \|\rho_A^x - \sigma_A^x\|_1. \tag{1.3}$$

- Trace distance is monotone non-increasing with respect to cptp maps. That is, for quantum states ρ and σ and the map $\mathcal{N}$, the following inequality holds:

$$\|\mathcal{N}(\rho) - \mathcal{N}(\sigma)\|_1 \leq \|\rho - \sigma\|_1.$$

- Trace distance is invariant with respect to tensor-product states, meaning that for quantum states ρ, σ and τ, we have that:

$$\|\rho \otimes \tau - \sigma \otimes \tau\|_1 = \|\rho - \sigma\|_1.$$

- Trace distance fulfills the triangle inequality; That is, for any three quantum states ρ, σ and τ, the following inequality holds:

$$\|\rho - \sigma\|_1 \leq \|\rho - \tau\|_1 + \|\tau - \sigma\|_1.$$

The generalized fidelity is used to define the purified distance as follows:

$$\mathrm{P}(\rho, \sigma) \coloneqq \sqrt{1 - \bar{F}^2(\rho, \sigma)}.$$

We use the purified distance to specify an ε-ball around $\rho_A \in \mathcal{S}$, that is $\mathcal{B}^\varepsilon(\rho_A) \coloneqq \{\rho'_A \in \mathcal{S} : P(\rho'_A, \rho_A) \leq \varepsilon\}$. Purified distance relates to the trace distance in the following way:

$$\frac{1}{2}\|\rho - \sigma\|_1 \leq \mathrm{P}(\rho, \sigma) \leq \sqrt{\|\rho - \sigma\|_1}.$$

The purified distance satisfies several properties similar to those of the trace distance, we list some of them below see for example [24] [1].

[1] In this thesis, without loss of generality, we work with normalized quantum states and the definition of the generalized fidelity is mentioned for completeness.

- Monotonicity: For quantum states ρ, σ and any completely positive trace-preserving map $\mathcal{E}$,

$$\mathrm{P}(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq \mathrm{P}(\rho, \sigma).$$

- Triangle inequality: For quantum states ρ, σ and ω, it holds that

$$\mathrm{P}(\rho, \sigma) \leq \mathrm{P}(\rho, \omega) + \mathrm{P}(\omega, \sigma).$$

- Invariance with respect to tensor product states: For quantum states ρ, σ and ω, it holds that:

$$\mathrm{P}(\rho \otimes \omega, \sigma \otimes \omega) = \mathrm{P}(\rho, \sigma).$$

  The following can also be easily verified:

$$\begin{aligned}\mathrm{P}\big(\sum_x p(x)\,|x\rangle\langle x| \otimes \rho_x^A \otimes \omega_x^B, \sum_x q(x)\,|x\rangle\langle x| \otimes \sigma_x^A \otimes \omega_x^B\big) \\ = \mathrm{P}\big(\sum_x p(x)\,|x\rangle\langle x| \otimes \rho_x^A, \sum_x q(x)\,|x\rangle\langle x| \otimes \sigma_x^A\big).\end{aligned}$$

The von Neumann entropy and the quantum relative entropy are defined as:

$$\begin{aligned}S(A)_\rho &:= S(\rho^A) := -\operatorname{Tr}\rho^A \log \rho^A \\ D(\rho\|\sigma) &:= \operatorname{Tr}(\rho\log\rho - \rho\log\sigma), \text{ if } \operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma) \text{ and } +\infty \text{ otherwise,}\end{aligned}$$

respectively, where $\operatorname{supp}(\rho)$ is the support of ρ. Conditional entropy, mutual information and conditional mutual information, $S(A|B)_\rho$, $I(A;B)_\rho$ and $I(A;B|C)_\rho$, are defined as:

$$\begin{aligned}S(A|B)_\rho &:= S(AB)_\rho - S(B)_\rho, \\ I(A;B)_\rho &:= S(A)_\rho + S(B)_\rho - S(AB)_\rho, \\ I(A;B|C)_\rho &:= S(A|C)_\rho - S(A|BC)_\rho = S(AC)_\rho + S(BC)_\rho - S(ABC)_\rho - S(C)_\rho.\end{aligned}$$

The von Neumann entropy and the mutual information can be defined as special cases of the quantum relative entropy; for instance it can be seen that $D(\rho^{AB}\|\rho^A \otimes \rho^B) = I(A;B)_\rho$. The coherent information of a bipartite state $\rho_{AB} \in \mathcal{S}^{AB}$ is defined as follows:

$$I(A\rangle B)_\rho := S(B)_\rho - S(AB)_\rho. \tag{1.4}$$

The conditional coherent information of a tripartite state ρ_{ABC} is defined as $I(A\rangle B|C)_\rho \coloneqq S(B|C)_\rho - S(AB|C)_\rho$ and it can be shown that $I(A\rangle B|C)_\rho = I(A\rangle BC)_\rho$. In particular, for the CQ state $\rho_{XAB} = \sum_x p_X(x)|x\rangle\langle x|_X \otimes \rho^x_{AB}$, we have $I(A\rangle BX)_\rho = \sum_x p_X(x) I(A\rangle B)_{\rho^x_{AB}}$. For classical systems (random variables), the von Neumann entropy reduces to the Shannon entropy, denoted $H(X)$.

Let $\{\Lambda, \mathbb{1} - \Lambda\}$ be the elements of a POVM that distinguishes between quantum states ρ and σ such that the probability of a correct guess on input ρ equals $\operatorname{Tr}\Lambda\rho$ and a wrong guess on σ is made with probability $\operatorname{Tr}\Lambda\sigma$. Let $\varepsilon \in (0,1)$. Then, the hypothesis testing relative entropy is defined as follows [25], [26]:

$$D^\varepsilon_{\mathrm{H}}(\rho\|\sigma) \coloneqq \max\{-\log_2 \operatorname{Tr}\Lambda\sigma : 0 \le \Lambda \le \mathbb{1} \wedge \operatorname{Tr}\Lambda\rho \ge 1-\varepsilon\}. \tag{1.5}$$

For all state ρ_A and σ_A and $\varepsilon \in [0,1)$, the following inequality holds [25]

$$D^\varepsilon_{\mathrm{H}}(\rho_A\|\sigma_A) \le \frac{1}{1-\varepsilon}\left[D(\rho_A\|\sigma_A) + h_b(\varepsilon)\right], \tag{1.6}$$

where $h_b(\varepsilon) \coloneqq -\varepsilon\log_2\varepsilon - (1-\varepsilon)\log_2(1-\varepsilon)$ is the binary entropy function. Another connection between the two relative entropies is due to the quantum Stein's lemma as given below for $\varepsilon \in (0,1)$ [27], [28]:

$$\lim_{n\to\infty}\frac{1}{n}D^\varepsilon_{\mathrm{H}}(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma). \tag{1.7}$$

From Eq. (1.5), the hypothesis testing mutual information for a bipartite state ρ^{AB} is defined as follows:

$$I^\varepsilon_{\mathrm{H}}(A;B)_\rho \coloneqq D^\varepsilon_{\mathrm{H}}(\rho^{AB}\|\rho^A\otimes\rho^B).$$

The followings are simple consequences of the above definitions. For a bipartite state $\rho^{AB} \in \mathcal{S}^{AB}$ and $\varepsilon \in (0,1)$, we have

$$I^\varepsilon_{\mathrm{H}}(A;B)_\rho \le \frac{1}{1-\varepsilon}(I(A;B)_\rho + h_b(\varepsilon)), \tag{1.8}$$

$$\lim_{n\to\infty}\frac{1}{n}I^\varepsilon_{\mathrm{H}}(A^n;B^n)_{\rho^{\otimes n}} = D(\rho^{AB}\|\rho^A\otimes\rho^B) = I(A;B)_\rho. \tag{1.9}$$

Max-relative entropy for $\rho_A, \sigma_A \in \mathcal{D}(\mathcal{H}_A)$ is defined as follows [29]:

$$D_{\max}(\rho_A\|\sigma_A) \coloneqq \inf\left\{\lambda \in \mathbb{R} : \rho_A \le 2^\lambda\sigma_A\right\}, \tag{1.10}$$

where it is well-defined if $\mathrm{supp}(\rho_A) \subseteq \mathrm{supp}(\sigma_A)$. And it relates to quantum relative entropy as follows [29]:

$$D(\rho_A \| \sigma_A) \leq D_{\max}(\rho_A \| \sigma_A). \tag{1.11}$$

An important property of the max-relative entropy is its monotonically non-increasing behavior with cptp maps, i.e., for quantum states ρ, σ and any cptp map $\mathcal{E}$, the following holds [29]:

$$D_{\max}(\mathcal{E}(\rho) \| \mathcal{E}(\sigma)) \leq D_{\max}(\rho \| \sigma). \tag{1.12}$$

For a parameter $\varepsilon \in (0,1)$ and quantum states ρ and σ, the smooth max-relative entropy is defined as follows [29]:

$$D^{\varepsilon}_{\max}(\rho \| \sigma) \coloneqq \min_{\rho' \in \mathcal{B}^{\varepsilon}(\rho)} D_{\max}(\rho' \| \sigma). \tag{1.13}$$

Further relation between the two entropies is given by the quantum Stein's lemma, usually referred to as asymptotic equipartition property (AEP): for $\varepsilon \in (0,1)$ [30]:

$$\lim_{n \to \infty} \frac{1}{n} D^{\varepsilon}_{\max}(\rho^{\otimes n} \| \sigma^{\otimes n}) = D(\rho \| \sigma). \tag{1.14}$$

From the smooth max-relative entropy, one can define a mutual information-like quantity for a bipartite state ρ^{AB} as follows:

$$D^{\varepsilon}_{\max}(A;B)_{\rho} \coloneqq D^{\varepsilon}_{\max}(\rho^{AB} \| \rho^A \otimes \rho^B) = \min_{\rho' \in \mathcal{B}^{\varepsilon}(\rho)} D_{\max}(\rho'^{AB} \| \rho^A \otimes \rho^B), \tag{1.15}$$

and its relation to quantum relative entropy can be seen from Eq. (1.14):

$$\lim_{n \to \infty} \frac{1}{n} D^{\varepsilon}_{\max}(A^n; B^n)_{\rho^{\otimes n}} = D(\rho^{AB} \| \rho^A \otimes \rho^B) = I(A;B)_{\rho},$$

holding for $\varepsilon \in (0,1)$.

For a bipartite state ρ^{AB} and a parameter $\varepsilon \in (0,1)$, a mutual information-like quantity can be defined as follows [31]:

$$\widetilde{I}^{\varepsilon}_{\max}(A;B)_{\rho} \coloneqq \inf_{\rho'^{AB} \in \mathcal{B}^{\varepsilon}(\rho^{AB})} D_{\max}(\rho'^{AB} \| \rho'^{A} \otimes \rho^B). \tag{1.16}$$

The following inequality from [31] relates the mutual information-like quantity above and the quantity defined in (1.15).

$$\widetilde{I}^{2\varepsilon}_{\max}(A;B)_{\rho} \leq D^{\varepsilon}_{\max}(A;B)_{\rho} + \log_2\left(\frac{3}{\varepsilon^2}\right). \tag{1.17}$$

where $\rho \in \mathcal{S}^{AB}$ and $\varepsilon \in (0,1)$.

The quantum relative entropy variance for $\rho_A, \sigma_A \in \mathcal{D}(\mathcal{H}_A)$ is given by [32]:

$$V(\rho_A \| \sigma_A) \coloneqq \mathrm{Tr}\{\rho_A [\log_2 \rho_A - \log_2 \sigma_A - D(\rho_A \| \sigma_A)]^2\}, \tag{1.18}$$

whenever $\mathrm{supp}(\rho_A) \subseteq \mathrm{supp}(\sigma_A)$ and $D(\rho_A \| \sigma_A)$ is the quantum relative entropy.

Fact 1 ([32] and [30]). *Let $\varepsilon \in (0,1)$ and n be an integer. For any pair of states ρ_A and σ_A and their n-fold products, i.e., $\rho_A^{\otimes n}$ and $\sigma_A^{\otimes n}$, the following equations hold:*

$$\begin{aligned} D_{\mathrm{H}}^{\varepsilon}(\rho_A^{\otimes n} \| \sigma_A^{\otimes n}) &= nD(\rho_A \| \sigma_A) \\ &\quad + \sqrt{nV(\rho_A \| \sigma_A)}\Phi^{-1}(\varepsilon) + O(\log n), \end{aligned}$$

$$\begin{aligned} D_{\max}^{\varepsilon}(\rho_A^{\otimes n} \| \sigma_A^{\otimes n}) &= nD(\rho_A \| \sigma_A) \\ &\quad - \sqrt{nV(\rho_A \| \sigma_A)}\Phi^{-1}(\varepsilon^2) + O(\log n), \end{aligned}$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} exp(-\frac{x^2}{2})dx$ is the cumulative distribution function for a standard Gaussian random variable and its inverse is defined as $\Phi^{-1}(\varepsilon) \coloneqq \sup\{\alpha \in \mathbb{R} | \Phi(\alpha) \leq \varepsilon\}$.

Lemma 1.1 (Hayashi and Nagaoka [33]). *For a POVM element $0 \leq T \leq \mathbb{1}$ and an operator $S \geq 0$,*

$$\mathbb{1} - (S+T)^{-\frac{1}{2}} T (S+T)^{-\frac{1}{2}} \leq 2(\mathbb{1} - T) + 4S. \qquad \blacksquare$$

Lemma 1.2 (Gentle measurement lemma [34]). *Let $\rho_A \in \mathcal{D}(\mathcal{H}_A)$ and $0 \leq \Lambda_A \leq \mathbb{1}$ be a measurement operator. If the measurement operator decides in favor of ρ_A with high probability, $\mathrm{Tr}\{\Lambda_A \rho_A\} \geq 1 - \varepsilon$ for $\varepsilon \in [0,1]$, then*

$$\left\| \rho_A - \sqrt{\Lambda_A} \rho_A \sqrt{\Lambda_A} \right\|_1 \leq 2\sqrt{\varepsilon}.$$

1.3 The structure of the thesis

In this subsection we delve into the details of the problems studied and the organisation of the thesis.

1.3.1 Chapter 2

Despite the existence of classical channels that can be better discriminated by adaptive ("sequential") strategies than by non-adaptive ("parallel") strategies, Hayashi [35] showed that adaptive strategies for classical channel discrimination do not give any advantage over non-adaptive strategies on the asymptotic rate at which the error probability exponentially decays with the number of channel uses. Verifying the potential advantages of the adaptive strategies in discrimination of quantum channels have been the focus of many studies. The second chapter of the thesis studies this problem further, where we first show an exponential lower bound on the discrimination of two channels by non-adaptive strategies, which is applicable to a large number of cases. In particular, it applies to a certain pair of entanglement-breaking channels for which it is known that two adaptive uses result in perfect discrimination [36]. Thus we get a separation between the adaptive Chernoff exponent of the two channels, which is infinite, and the non-adaptive one, which is finite. Next we show that the asymptotic error exponent rate of discrimination of classical-quantum channels cannot be improved by adaptive strategies. This conclusion follows by proving a generalised Chernoff bound and the Hoeffding error exponent; that is, we show that a lower bound on the error exponent rate of the alternative hypothesis leads to the error probability of the null hypothesis decreasing to zero exponentially fast at a rate determined by a function of the difference between the imposed lower bound and the maximum quantum Rényi divergence registered between the channels. Combining our results with the recent findings of [37] in the settings of Stein's lemma and Han-Kobayashi indicates that adaptive strategies for discrimination of classical-quantum channels are not necessary. The improved proof strategy in our work is also novel for the classical case. We also prove the non-adaptive strategies to be optimal for discrimination of quantum channels under a subclass of strategies which only allow classical feedback and product input states. As an application of our findings, we address the discrimination power of quantum channels and show that neither adaptive strategies nor input quantum memory can increase the discrimination power of entanglement-breaking channels.

1.3.2 Chapter 3

In chapter three, we address the second-tier type of question posed in subsection 1.1.2 via a two-pronged strategy. The first resource conversion we study is the task of converting noisy multipartite quantum correlations, i.e. an m-partite quantum state ($m \geq 2$), into noiseless m-partite classical corre-

lations, i.e. common randomness (CR), under local operations and classical communications (LOCC). Intuitively, CR is a random variable that is uniformly distributed and known to all m parties. It is known that distillation of CR without additional classical communication is generically impossible [38]. On the other hand, since classical communication and CR are not "orthogonal" resources, allowing free classical communications is not appropriate, because it can be used to create unlimited CR. However, one can consider two interesting directions: imposing a secrecy requirement on CR, or limiting the classical communication. In this chapter, we are concerned with the second direction; the first one, known as key distillation, was studied by Maurer [39], Ahlswede and Csiszár [40] and its quantum generalization in [41]. The problem of distilling CR from two correlated random variables under one-way classical communication of R bits per source observation was studied by Ahlswede and Csiszár [38] (see the paper for other models). Subsequently, their model was generalized in [42], introducing the *distillable CR*, the amount of CR generated in excess of the consumed classical communication. When the classical communication is one-way, the distillable CR is still an (asymmetric) measure of the total classical correlations in the state [43]. For a recent review of multi-party key distillation see [44].

In Section 3.2, we prove two lower bounds on the distillable CR from multipartite mixed quantum states. We do this by offering a generalization of a result in multi-terminal distributed lossless source coding and secret key agreement due to Csiszár and Narayan [45] known as *communication for omniscience* (CO). There, m parties observe a correlated discrete memoryless multiple source $X_{[m]} = (X_1, \ldots, X_m)$, the i-th node obtaining X_i. The nodes are allowed to communicate interactively over a public noiseless broadcast channel so that at the end they attain omniscience: each node reconstructs the whole vector of observations $X_{[m]}$. The objective is to minimise the overall communication to achieve this goal.

We first apply the main result of [45] to the outcomes of local measurements on m-partite quantum states, and then generalize this result to partial measurements, modelled as instruments, such that each party not only has classical information X_i but also a quantum register A_i' containing containing correlated quantum side information. It uses a novel random binning coding and decoding strategy for the problem of correlated source compression with quantum side information at the decoder, presented in a concise way in the Appendix. The reason for the secrecy rate being exactly the difference between the entropy of $X_{[m]}$ and the total communication rate R_{CO} is that this is attained by privacy amplification. We note that the same rate is also an achievable rate for the distillable CR by the recycling of resources idea; for more on their relation see [46].

Our second problem concerns converting multipartite quantum correlation into noiseless quantum correlation, i.e. the so-called entanglement distillation task (Section 3.3). The theory of asymptotic manipulation of multipartite entanglement is very complex, even in the pure state case a simple theory as is known for bipartite pure states, is probably forever beyond reach; for mixed states, already the bipartite case defies complete analysis, so much so that it is even open whether there are bound entangled states with non-positive partial transpose (NPT). For these reasons, for the task of entanglement distillation, we focus on the Greenberger-Horne-Zeilinger (GHZ) distillation problem, and on pure initial states. Very little previous work has concerned itself with the asymptotic rate of GHZ distillation, despite such states being evidently useful for cryptography [47]. The important exceptions are Smolin *et al.* [48], Fortescue and Lo [49] and Streltsov *et al.* [50]; furthermore [51] for stabilizer states and exact distillation. For the entropy and relative entropy upper rate bounds see [5, 6].

Motivated by the recent paper [52], which treats the distillation of multipartite GHZ states from many copies of a given multipartite pure state and presents an achievable rate based on a combinatorial construction, we realised that the same rate can be obtained and improved using off-the-shelf techniques of quantum Shannon theory from the early 2000s, namely the coherification of protocols for CR distillation.

The first lower bound reproduces the result of Vrana and Christandl [52], and the second protocol improves upon this lower bound. To the best of our knowledge it is the best available bound, subsuming a number of other previous results.

1.3.3 Chapter 4

In this chapter, we aim to study the problem of simultaneous transmission of classical and quantum information over a single use of a quantum channel. In other words, we are interested in the one-shot tradeoff between the number of bits and qubits that are simultaneously achievable. The root of our approach is the well-known quantum capacity theorem via private classcial communication [19]. The basic intuition underlying the quantum capacity is the no-cloning theorem which states that it is impossible to create an identical copy of an arbitrary unknown quantum state. We know well that associated to every quantum channel there is an environment (Eve). If Eve can learn anything about the quantum information that the transmitter is trying to send to the receiver, the receiver will not be able to retrieve this information, otherwise the no-cloning theorem would be violated. Hence, to transmit quantum information, the transmitter needs to store her quantum

information in such subspaces of her input space that Eve does not have access to. By using this idea, Devetak [19] proved that a code for private classical communication can be readily translated into a code for quantum communication. This approach to quantum information transmission underpins our approach in which the focus is put on transmission of public and private classical information, and the latter can be translated into quantum information directly.

1.3.4 Chapter 5

The problem addressed in this chapter is a continuation of the simultaneous transmission of classical and quantum information studied in the previous chapter. It is known that existence of local randomness in encoding is a prerequisite for transmission of secret messages. It was also shown that a non-secret message may also play the role of the dummy randomness [17]. The interesting question arose here is that of the tradeoff between the non-secret message and dummy randomness. Moreover, traditionally, local randomness in encoding was assumed to be available infinitely, undermining its effect and value in secret transmission. In this chapter, we consider transmission of secret messages over a single-serving quantum broadcast channel with rate-limited randomness and will find inner and outer bound regions for the tradeoff between the secret message, dummy randomness and also the non-secret message that help compensate for the lack of enough dummy randomness.

In summary, the last four chapters are essentially based on the following publications and preprints:

- Chapter 2:
  - Farzin Salek, Masahito Hayashi and Andreas Winter, 'When are Adaptive Strategies in Asymptotic Quantum Channel Discrimination Useful?", *in preparation*, 2020.
- Chapter 3:
  - [53]: Farzin Salek and Andreas Winter, "Multi-User Distillation of Common Randomness and Entanglement from Quantum States," *pre-print (2019)*, arXiv: 2008.04964.
  - [54]: Farzin Salek and Andreas Winter, "Multi-User Distillation of Common Randomness and Entanglement from Quantum States," 2020 IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 2020, pp. 1949-1954.

- Chapter 4:
  - [55] F. Salek, A. Anshu, M. Hsieh, R. Jain and J. R. Fonollosa, "One-Shot Capacity Bounds on the Simultaneous Transmission of Classical and Quantum Information," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2141-2164, 2020. arXiv: 1811.09177.
  - [56]: F. Salek, A. Anshu, M. Hsieh, R. Jain and J. R. Fonollosa, "One-shot Capacity Bounds on the Simultaneous Transmission of Public and Private Information Over Quantum Channels," 2018 IEEE International Symposium on Information Theory (ISIT), Vail, CO, 2018, pp. 296-300.
- Chapter 5:
  - [57] : F. Salek, M. Hsieh, J. R. Fonollosa, "Single-Serving Quantum Broadcast Channel with Common, Individualized and Confidential Messages," *IEEE Trans. Inf. Theory*, 2020. arXiv:1903.04463.
  - [58]: F. Salek, M. Hsieh and J. R. Fonollosa, "Publicness, Privacy and Confidentiality in the Single-Serving Quantum Broadcast Channel," 2019 IEEE International Symposium on Information Theory (ISIT), Paris, France, 2019, pp. 1712-1716.

Chapter 2

When are adaptive strategies in asymptotic quantum channel discrimination useful?

This chapter concerns the separation between adaptive and non-adaptive strategies in discrimination of quantum channels. To show that adaptive strategies in general are superior to non-adaptive ones, we find an exponential lower bound on the discrimination of two channels by non-adaptive strategies, which is applicable to a large number of cases. In particular, it applies to a certain pair of entanglement-breaking channels for which it is known that two adaptive uses result in perfect discrimination. Thus we get a separation between the adaptive Chernoff exponent of the two channels, which is infinite, and the non-adaptive one, which is finite. Next we show that for classical-quantum channels the most general adaptive strategy does not give any advantage over non-adaptive strategies on the asymptotic rate at which the error probability exponentially decays with the number of channel uses. This conclusion follows by proving a generalised Chernoff bound and the Hoeffding error exponent. This result generalises to arbitrary quantum channels under a restricted class of protocols where only product state inputs are allowed and only classical information for the feed-forward in the adaptive strategy. We also show that under the same class of protocols, the non-adaptive strategies are optimal for discrimination of binary hypotheses with an arbitrary quantum channel. However, for a slightly larger class where entangled state inputs are allowed, we only show the optimality of non-adaptive strategies for the entanglement-breaking channels.

2.1 Background

Hypothesis testing is one of the most important problems in information theory, and in quantum information, one of the simplest problems where the difficulties stemming from non-commutativity of operators appear. A particular hypothesis testing setting that we study in this chapter is that of quantum channel discrimination, in which the hypotheses are described by two quantum channels, i.e. completely positive and trace preserving (cptp) maps, acting on a given quantum system, and more precisely $n \gg 1$ independent realizations of the unknown channel. It is not hard to see that both the type-I and type-II error probabilities can be made to go to 0 exponentially fast, just as in the case of hypotheses described by quantum states, and hence the fundamental question is the characterization of the possible error exponents.

To spell out the precise questions, let us introduce a bit of notation. Throughout the chapter, A, B, C, etc, denote quantum systems, but also their corresponding Hilbert space. We identify states ρ with their density operators and use superscripts to denote the systems on which the mathematical objects are defined. The set of density matrices on A is written as $\mathcal{S}^A$. When talking about tensor products of spaces, we may habitually omit the tensor sign, so $A \otimes B = AB$, etc. The capital letters X, Y, etc. denote random variables whose realizations and the alphabets will be shown by the corresponding small and calligraphic letters, respectively: $X = x \in \mathcal{X}$. All Hilbert spaces and ranges of variables may be infinite; the dimension of a Hilbert space A is denoted $|A|$, as is the cardinality $|\mathcal{X}|$ of a set $\mathcal{X}$. For any positive integer m, we define $\vec{x}_m \coloneqq (x_1, \cdots, x_m)$. For the state $\rho \in \mathcal{S}^{AB}$ in the composite system AB, the partial trace over system A (resp. B) is denoted by Tr_A (resp. Tr_B). We denote the identity operator by I. We use log and ln to denote base 2 and natural logarithms, respectively. Moving on to quantum channels, these are linear, completely positive and trace preserving maps $\mathcal{M} : \mathcal{S}^A \to \mathcal{S}^B$ for two quantum systems A and B; $\mathcal{M}$ extends uniquely to a linear map from trace class operators on A to those on B. We often denote quantum channels, by slight abuse of notation, as $\mathcal{M} : A \to B$. The ideal, or identity, channel on A is denoted id_A. Note furthermore that a state ρ^A on a system A can be viewed as a quantum channel $\rho : 1 \to A$, where 1 denotes the canonical one-dimensional Hilbert space, isomorphic to the complex numbers $\mathbb{C}$, which interprets a state operationally consistently as a state preparation procedure.

The most general operationally justified strategy to distinguish two channels $\mathcal{M}, \overline{\mathcal{M}} : A \to B$ is to prepare a state ρ^{RA}, apply the unknown channel to A (and the identity channel id_R to R), and then apply a binary measurement

POVM $(T, I-T)$ on BR, so that

$$\alpha(\mathcal{M}\|\overline{\mathcal{M}}) = \mathrm{Tr}\big((\mathrm{id}_R \otimes \mathcal{M})\rho\big)(I-T),$$
$$\beta(\mathcal{M}\|\overline{\mathcal{M}}) = \mathrm{Tr}\big((\mathrm{id}_R \otimes \overline{\mathcal{M}})\rho\big)T,$$

are the error probabilities of type I and type II, respectively. It is easy to see that whatever state ρ^{AR} is considered as input, it can be purified to $\psi^{ARR'}$, with a suitable Hilbert space, and the latter state can be used to get the same error probabilities. Then, once there is a pure state, one only needs a subspace of $R \otimes R'$ of dimension $|A|$, namely the support of $\psi^{RR'}$, which by the Schmidt decomposition is at most $|A|$-dimensional. Therefore, the state ρ is without loss of generality pure and that hence R has dimension at most that of A. The strategy is entirely described by the pair $\big(\rho, (T, I-T)\big)$ consisting of the initial state and the final measurement, and we denote it $\mathcal{T}$. Consequently, the above error probabilities are more precisely denoted $\alpha(\mathcal{M}\|\overline{\mathcal{M}}|\mathcal{T})$ and $\beta(\mathcal{M}\|\overline{\mathcal{M}}|\mathcal{T})$, respectively.

These strategies use the unknown channel exactly once; to use it $n > 1$ times, one could simply consider that $\mathcal{M}^{\otimes n}$ and $\overline{\mathcal{M}}^{\otimes n}$ are quantum channels themselves and apply the above recipe. While for states this indeed leads to the most general possible discrimination strategy, for general channels other, more elaborate procedures are possible. The most general strategy we shall consider in this chapter is the *adaptive* strategy, applying the n channel instances sequentially, using quantum memory and quantum feed-forward, and a measurement at the end. It is defined as follows.

Definition 1. *Concretely, the strategy $\mathcal{T}_n$ is given by an $(n+1)$-tuple*

$$\big(\rho_1^{R_1A}, \mathcal{F}_1, \ldots, \mathcal{F}_{n-1}, (T, I-T)\big),$$

consisting of an auxiliary system R_1 and a state ρ_1 on R_1A_1, quantum channels $\mathcal{F}_m : R_mB_m \to R_{m+1}A_{m+1}$ and a binary POVM $(T, I-T)$ on R_nB. It encodes the following procedure: in the m-th round ($1 \le m \le n$), apply the unknown channel $\Xi \in \{\mathcal{M}, \overline{\mathcal{M}}\}$ to $\rho_m = \rho_m^{R_mA_m}$, obtaining

$$\omega_m^{R_mB_m} = \omega_m^{R_mB_m}(\Xi) = (\mathrm{id}_{R_t} \otimes \Xi)\rho_m^{R_mA_m}.$$

Then, as long as $m < n$, use $\mathcal{F}_m$ to prepare the state for the next channel use:

$$\rho_{m+1}^{R_{m+1}A_{m+1}} = \mathcal{F}_m(\omega_m^{R_mB_m}).$$

When $m = n$, measure the state $\omega_n^{R_nB_n}$ with $(T, I-T)$, where the first outcome corresponds to declaring the unknown channel to be $\mathcal{M}$, the second $\overline{\mathcal{M}}$. Thus,

the n-copy error probabilities of type I and type II are given by

$$\alpha_n(\mathcal{M}\|\overline{\mathcal{M}}|\mathcal{T}_n) = \mathrm{Tr}\big(\omega_n^{R_nB_n}(\mathcal{M})\big)(I-T),$$
$$\beta_n(\mathcal{M}\|\overline{\mathcal{M}}|\mathcal{T}_n) = \mathrm{Tr}\big(\omega_n^{R_nB_n}(\overline{\mathcal{M}})\big)T,$$

respectively. As in the case of a single use of the channel, one can without loss of generality (w.l.o.g.) simplify the strategy, by purifying the initial state ρ_1, hence $|R_1| \leq |A|$, and for each $m > 1$ going to the Stinespring isometric extension of the channel $\mathrm{Tr}_{R_{m+1}} \circ \mathcal{F}_m : R_mB_m \to A_{m+1}$, which requires a system R_{m+1} with dimension $|R_{m+1}| \leq |R_m||A||B|$.

The set of all adaptive strategies of n sequential channel uses is denoted $\mathbb{A}_n$. It quite evidently includes the n parallel uses described at the beginning of this paragraph, when a single-use strategy is applied to $\xi^{\otimes n}$; the set of these non-adaptive or parallel strategies is denoted $\mathbb{P}_n$. Among those again, we can distinguish the subclass of parallel strategies without quantum memory, meaning that $R = 1$ is trivial and that the input state ρ^{A^n} at the input system $A^n = A_1 \dots A_n$ is a product state, $\rho^{A^n} = \rho_1^{A_1} \otimes \cdots \otimes \rho_n^{A_n}$; this set is denoted $\mathbb{P}_n^0$. Another restricted set of strategies we are considering in the present chapter is that of adaptive strategies with classical feed-forward and no quantum memory, which we will define formally in Section 2.6, and denote by $\mathbb{A}_n^{c,0}$.

For a given class $\mathbb{S}_n \subset \mathbb{A}_n$ of adaptive strategies for any number n of channel uses, the fundamental problem is now to characterize the possible pairs of error exponents for two channels $\mathcal{M}$ and $\overline{\mathcal{M}}$:

$$\mathfrak{E}(\mathcal{M}\|\mathcal{N}|\mathbb{S}) \coloneqq \Bigg\{(r,s) : \exists \mathcal{T}_n \in \mathbb{S}_n \; 0 \leq r \leq \liminf_{n\to\infty} -\frac{1}{n}\log\beta_n(\mathcal{M}\|\overline{\mathcal{M}}|\mathcal{T}_n), \\ 0 \leq s \leq \liminf_{n\to\infty} -\frac{1}{n}\log\alpha_n(\mathcal{M}\|\overline{\mathcal{M}}|\mathcal{T}_n)\Bigg\}. \tag{2.1}$$

In particular, we are interested, for each $s \geq 0$, in the largest r such that $(r,s) \in \mathfrak{E}(\mathcal{M}\|\overline{\mathcal{M}}|\mathbb{S})$. To this end, we define the error rate tradeoff

$$B_e^{\mathbb{S}}(r|\mathcal{M}\|\overline{\mathcal{M}}) \coloneqq \sup s \text{ s.t. } \exists \mathcal{T}_n \in \mathbb{S}_n \; r \leq \liminf_{n\to\infty} -\frac{1}{n}\log\beta_n(\mathcal{M}\|\overline{\mathcal{M}}|\mathcal{T}_n), \\ s \leq \liminf_{n\to\infty} -\frac{1}{n}\log\alpha_n(\mathcal{M}\|\overline{\mathcal{M}}|\mathcal{T}_n), \tag{2.2}$$

as well as the closely related function

$$C^{\mathbb{S}}(a,b|\mathcal{M}\|\overline{\mathcal{M}}) \coloneqq \inf_{\mathcal{T}_n\in\mathbb{S}_n} \liminf_{n\to\infty} -\frac{1}{n}\log\left(2^{na}\alpha_n(\mathcal{M}\|\overline{\mathcal{M}}|\mathcal{T}_n) + 2^{nb}\beta_n(\mathcal{M}\|\overline{\mathcal{M}}|\mathcal{T}_n)\right). \tag{2.3}$$

Note that $\mathfrak{E}(\mathcal{M}\|\overline{\mathcal{M}}|\mathbb{S})$ is a closed set by definition, and for most 'natural' restrictions $\mathbb{S}$ it is also convex. In the latter case, the graph of $B_e^{\mathbb{S}}(r|\mathcal{M}\|\overline{\mathcal{M}})$ traces the upper boundary of $\mathfrak{E}(\mathcal{M}\|\overline{\mathcal{M}}|\mathbb{S})$, and it can be reconstructed from $C^{\mathbb{S}}(a,b|\mathcal{M}\|\overline{\mathcal{M}})$ by a Legendre transform.

Historically, two extreme regimes are of special interest: the maximally asymmetric error exponent,

$$\max r \text{ s.t. } \exists s\ (r,s) \in \mathfrak{E}(\mathcal{M}\|\overline{\mathcal{M}}|\mathbb{S}) = \max r \text{ s.t. } (r,0) \in \mathfrak{E}(\mathcal{M}\|\overline{\mathcal{M}}|\mathbb{S}),$$

together with the opposite one of maximisation of s, which are known as Stein's exponents, and the symmetric error exponent

$$\begin{aligned}\xi^{\mathbb{S}}(\mathcal{M},\overline{\mathcal{M}}) &= \max r \text{ s.t. } (r,r) \in \mathfrak{E}(\mathcal{M}\|\overline{\mathcal{M}}|\mathbb{S}) \\ &= C^{\mathbb{S}}(0,0|\mathcal{M}\|\overline{\mathcal{M}}),\end{aligned}$$

which is generally known as Chernoff exponent or Chernoff bound.

In the present chapter we are primarily occupied with the performance of adaptive strategies. Naturally, the first question in this search would be to investigate the existence of quantum channels for which some class $\mathbb{S}_n \subset \mathbb{A}_n$ outperforms the parallel strategy when $n \to \infty$; in other words, if there exists a separation between adaptive and non-adaptive strategies. We then shift the focus to a particular class of channels, namely, classical-quantum channels (cq-channels) and investigate if the most general strategy offers any benefit over the most weak strategy $\mathbb{P}_n^0$. To see whether our cq-channels study can be generalized to other channels, we study adaptive strategies that only allow for classical feed-forward, i.e. $\mathbb{A}_n^c$ and $\mathbb{A}_n^{c,0}$. These findings are then applied to discrimination power of a quantum channel: how well two given states in $A^{\otimes n}$ can be discriminated after passing through a quantum channel and whether employment of adaptive strategies can be beneficiary.

The rest of the chapter is organized as follows. In the following Section 2.2 we review some of the background history of hypothesis testing and previous work, this should also motives the problems we solve. In Section 2.3 we show the first asymptotic separation between adaptive and non-adaptive strategies via proving a lower bound on the Chernoff error for non-adaptive strategies and analyzing an example where adaptive strategies achieve error zero even with two copies of the channels. Section 2.5 contains our analysis of cq-channel discrimination, where we start by describing the most general adaptive strategy in this case, and mathematically define the specific quantities that we address. In Section 2.6, we study the discrimination of quantum channels with classical feed-forward, with and without entangled inputs, and in Section 2.7 we apply our results to the discrimination power of an arbitrary quantum channel. We conclude in Section 2.8.

2.2 History and previous work

In classical information theory, discriminating two distributions has been studied by many researchers; Stein, Chernoff [59], Hoeffding [60] and Han-Kobayashi [61] formulated asymptotic hypothesis testing of two distributions as optimization problems and subsequently found optimum expressions. As generalizations of these settings to quantum realm, discrimination of two quantum states has been studied extensively in quantum information theory, albeit the complications stemming from the noncommutativity of quantum mechanics appear in the most visible way among these problems. The direct part and weak converse of the quantum extension of Stein's lemma were proven by Hiai and Petz [62]; the proof combines the classical case of Stein's lemma and the fact that for a properly chosen measurement, the classical relative entropy of the measurement outcome approaches the quantum relative entropy of the initial states. Subsequently, Ogawa and Nagaoka [27] proved the strong converse of quantum Stein's lemma, that is, they showed that if the error exponent of type II goes to zero exponentially fast at a rate higher than the relative entropy registered between the states, the probability of correctly identifying the null hypothesis decays to zero with a certain speed, where they found the exact expression. The Chernoff bound was settled by Nussbaum and Szkoła [63] and Audenaert *et al.* [64], where the former proved the converse and the latter showed its attainability (See [65] for earlier significant progress). Concerning the quantum extension of the Hoeffding bound, in [66] a lower bound was proved suggesting the existence of a tighter lower bound. Later, [67] proved the suggested tighter lower bound and subsequently, Nagaoka [68] showed the optimality of the above quantum Hoeffding lower bound.

Discrimination of two (quantum) channels appears as a natural extension of the state discrimination problem. However, despite inherent mathematical links between the channel and state discrimination problems, due to the additional degrees of freedom introduced by the adaptive strategies, discrimination of channels is more complicated. Many papers have been dedicated to study the potential advantages of adaptive strategies over non-adaptive strategies in channel discrimination.

The seminal classical work [35] showed that in the asymptotic regime, the exponential error rate for classical channel discrimination cannot be improved by adaptive strategies for any of the symmetric or asymmetric settings. In thw classical case, where the classical channels $W : x \to W_x$ and $\overline{W} : x \to \overline{W}_x$ with common input ($\mathcal{X}$) and output ($\mathcal{Y}$) alphabets, and output distributions $\{W_x\}_{x\in\mathcal{X}}$ and $\{\overline{W}_x\}_{x\in\mathcal{X}}$, resp., are given, Ref. [35, Thm. 1] proved the strong

converse

$$B_e^{\mathbb{A}}(r|W\|\overline{W}) = 0 \quad \text{if} \quad r > D(W\|\overline{W}) \coloneqq \sup_x D(W_x\|\overline{W}_x), \tag{2.4}$$

where $D(W_x\|\overline{W}_x) \coloneqq \sum_{y\in\mathcal{Y}} W_x(y)\log\big(W_x(y)/\overline{W}_x(y)\big)$ is the relative entropy; and for $0 \le r \le D(W\|\overline{W})$, [35, Thm. 2] showed that

$$B_e^{\mathbb{A}}(r|W\|\overline{W}) = B_e^{\mathbb{P}^0}(r|W\|\overline{W}) = \sup_x \sup_{0\le\alpha\le 1} \frac{\alpha-1}{\alpha}\big(r - D_\alpha(W_x\|\overline{W}_x)\big), \tag{2.5}$$

where $D_\alpha(W_x\|\overline{W}_x) \coloneqq \frac{1}{\alpha-1}\log\sum_{y\in\mathcal{Y}}(W_x(y)/\overline{W}_x(y))^\alpha \overline{W}_x(y)$ is the Rényi relative entropy.[1] Moreover, from the relation between the Hoeffding and Chernoff exponents

$$\xi^{\mathbb{S}}(W,\overline{W}) = \sup_r \left\{r|B_e^{\mathbb{S}}(r|\mathcal{N}\|\overline{\mathcal{N}}) \ge r\right\},$$

it was shown in [35, Cor. 2] that

$$\xi^{\mathbb{A}}(W,\overline{W}) = \xi^{\mathbb{P}^0}(W,\overline{W}) = \sup_x \sup_{0\le\alpha\le 1} (1-\alpha)D_\alpha(W_x\|\overline{W}_x). \tag{2.6}$$

Since the publication of this seminal work, significant amount of research has focused on showing the potential advantages of adaptive strategies in discrimination of quantum channels. Significant progress was reported in [69] concerning classical-quantum (cq) channels. Let $\mathcal{N} : x \to \rho_x$ and $\overline{\mathcal{N}} : x \to \sigma_x$ be two cq-channels (these channels will be formally defined in Sec. 2.5). One may expect the same relations as (2.4), (2.5) and (2.6) to hold for cq-channels, replacing the Rényi relative entropy with a quantum extension of it. For Stein's lemma and its strong converse this was indeed shown to be the case in [69, Cor. 28], namely

$$B_e^{\mathbb{A}}(r|\mathcal{N}\|\overline{\mathcal{N}}) = 0 \quad \text{if} \quad r > D(\mathcal{N}\|\overline{\mathcal{N}}) \coloneqq \sup_x D(\rho_x\|\sigma_x), \tag{2.7}$$

where $D(\rho_x\|\sigma_x) \coloneqq \operatorname{Tr}\rho_x(\log\rho_x - \log\sigma_x)$ is the quantum relative entropy. Thus, one can assume $0 \le r \le D(\mathcal{N}\|\overline{\mathcal{N}})$ for the Hoeffding and Chernoff bounds.

A number of upper bounds for $B_e^{\mathbb{A}}(r|\mathcal{N}\|\overline{\mathcal{N}})$ are reported in the literature but finding a compact form meeting $B_e^{\mathbb{P}^0}(r|\mathcal{N}\|\overline{\mathcal{N}})$ has been an open problem. Two such upper bounds were reported by Berta *et al.* [69]; the first upper

[1] In the original notation in [35], $D_{1+\alpha}(W_x\|\overline{W}_x) \coloneqq \frac{\phi(-\alpha|W_x\|\overline{W}_x)}{\alpha}$. Moreover, the definition of $B_e^{\mathbb{S}_n}(r|W\|\overline{W})$ implicitly follows from (2.15) by replacing the quantum channels with respective classical channels.

bound follows the similar reasoning as in the classical Hoeffding bound [35], that is, considering an intermediate channel and using the strong and weak Stein's lemma. However, unlike the classical case, this line of reasoning could not yield a tight bound. Note that besides (2.5), in the classical case there is another compact expression for $B_e^{\mathbb{A}}(r|W\|\overline{W})$

$$B_e^{\mathbb{A}}(r|W\|\overline{W}) = \sup_{x\in\mathcal{X}} \min_{Q:D(Q\|\overline{W}_x)\leq r} D(Q\|W_x).$$

The reason that the classical approach of [35] does not yield a tight bound in quantum case is [65, Sec. 3.8]

$$B_e^{\mathbb{A}}(r|\mathcal{N}\|\overline{\mathcal{N}}) \leq \max_{x\in\mathcal{X}} \min_{\tau:D(\tau\|\sigma_x)\leq r} D(\tau\|\rho_x).$$

The second upper bound of Berta *et al.* [69] employs the fact that cq-channels are *environment-parameterized*: Due to the structure of the environment-parametrized channels, any n-round adaptive channel discrimination protocol can be understood as a particular kind of state discrimination protocol for the environment states of each channel. This development reduces the cq-channel discrimination problem to that of state discrimination between $(\bigotimes_x \rho_x)^{\otimes n}$ and $(\bigotimes_x \sigma_x)^{\otimes n}$. However, plugging the states into the well-known state discrimination bounds does not lead to tight characterisation.

In general quantum channel discrimination, it is known that adaptive strategies offer an advantage in the non-asymptotic regime for discrimination in symmetric Chernoff setting [36, 70–72]. In particular, Harrow *et al.* [36] demonstrated the advantage of adaptive strategies in discriminating a pair of entanglement-breaking channels that requires just two channel evaluations to distinguish them perfectly, but such that no non-adaptive strategy can give perfect distinguishability using any finite number of channel evaluations. However, it was open whether the same holds in the asymptotic setting.

This question in the asymmetric regime was recently settled by Wang and Wilde: In [73, Thm. 3], they found an exponent in Stein's setting for non-adaptive strategies in terms of channel max-relative entropy, also in the same paper [73, Thm. 6], they found an exponent in Stein's setting for the adaptive strategies in terms of amortized channel divergence, a quantity introduced in [69] to quantify the largest distinguishability between two channels. However, the fact that adaptive strategies do not offer an advantage in the setting of Stein's lemma for quantum channels, i.e. the equality of the aforementioned exponents of Wang and Wilde, was later shown in [74] via a chain rule for the quantum relative entropy proven therein.

Cooney *et al.* [75] proved quantum Stein's lemma for discriminating between an arbitrary quantum channel and a "replacer channel" that discards

its input and replaces it with a fixed state. This work led to the conclusion that at least in the asymptotic regime, a non-adaptive strategy is optimal in the setting of Stein's lemma. However, in the Hoeffding and Chernoff settings, the question of potential advantages of adaptive strategies involving replacer channels remains open.

Hirche *et al.* [76] studied the maximum power of a fixed quantum detector, i.e. a POVM, in discriminating two possible states. This problem is dual to the state discrimination scenario considered so far in that, while in the state discrimination problem the state pair is fixed and optimization is over all measurements, in this problem a measurement POVM is fixed and the question is how powerful this discriminator is, and then whatever criterion considered for quantifying the power of the given detector, it should be optimized over all input states. In particular, if $n \geq 2$ uses of the detector are available, the optimization takes place over all n-partite entangled states and also all adaptive strategies that may help improve the performance of the measurement. The main result of [76] states that when asymptotically many uses (i.e. $n \to \infty$) of a given detector is available, its performance does not improve by considering general input states or adopting any kind of adaptive strategy in any of the symmetric or asymmetric settings described before. The main ingredient in this paper is the classical result in the previous paper [35]; that is, it is shown that adaptive improvement of the measurement results with general entangled states can be cast as discriminating two classical channels, which is known to be improve by adaptive strategies.

2.3 Asymptotic separation between adaptive and non-adaptive strategies

In this section we exhibit an asymptotic separation between the Chernoff error exponents of discriminating between two channels by adaptive versus non-adaptive strategies. Concretely, we will show that two channels described in [36], and shown to be perfectly distinguishable by adaptive strategies of $n \geq 2$ copies, hence having infinite Chernoff exponent, nevertheless have a finite error exponent under non-adaptive strategies.

The separation is based on a general lower bound on non-adaptive strategies for an arbitrary pair of channels. Consider two quantum channels, i.e. cptp maps, $\mathcal{M}, \overline{\mathcal{M}} : A \to B$. To fix notation, we can write their Kraus de-

compositions as

$$\mathcal{M}(\rho) = \sum_i E_i \rho E_i^\dagger,$$
$$\overline{\mathcal{M}}(\rho) = \sum_j F_j \rho F_j^\dagger.$$

The most general strategy to distinguish them consists in the preparation of a, w.l.o.g. pure, state φ on $A \otimes R$, where $|R| \simeq |A|$, send it through the unknown channel, and make a binary measurement $(T, I - T)$ on $B \otimes R$:

$$p = \mathrm{Tr}\left((\mathrm{id}_R \otimes \mathcal{M})\varphi\right) T,$$
$$q = \mathrm{Tr}\left((\mathrm{id}_R \otimes \overline{\mathcal{M}})\varphi\right) T,$$

and likewise $1 - p$ and $1 - q$ by replacing T in the above formulas with $I - T$. Note that for uniform prior probabilities on the two hypotheses, the error probability in inferring the true channel from the measurement output is $\frac{1}{2}(1 - |p - q|)$.

The maximum of $|p - q|$ over state preparations and measurements gives rise to the (normalized) diamond norm distance of the channels:

$$\max_{\varphi, T} |p - q| = \frac{1}{2} \|\mathcal{M} - \overline{\mathcal{M}}\|_\diamond,$$

which in turn quantifies the minimum discrimination error under the most general quantum strategy:

$$P_e = \frac{1}{2}\left(1 - \frac{1}{2}\|\mathcal{M} - \overline{\mathcal{M}}\|_\diamond\right).$$

We are interested in the asymptotics of this error probability when the discrimination strategy has access to $n \gg 1$ many instances of the unknown channel in parallel, or in other words, in a non-adaptive way. This means effectively that the two hypotheses are the simple channels $\mathcal{M}^{\otimes n}$ and $\overline{\mathcal{M}}^{\otimes n}$, so that the error probability is

$$P_{e,\mathbb{P}}^{(n)} = \frac{1}{2}\left(1 - \frac{1}{2}\left\|\mathcal{M}^{\otimes n} - \overline{\mathcal{M}}^{\otimes n}\right\|_\diamond\right).$$

The (non-adaptive) Chernoff exponent is then defined as

$$\xi^{\mathbb{P}}(\mathcal{M}, \overline{\mathcal{M}}) := \lim_{n \to \infty} -\frac{1}{n} \log P_{e,\mathbb{P}}^{(n)},$$

the existence of the limit being guaranteed by general principles. Note that the limit can be $+\infty$, which happens in all cases where there is an n such that

$P_{e,\mathbb{P}}^{(n)} = 0$. It is currently unknown whether this is the only case; cf. the case of the more flexible adaptive strategies, for which there is a simple criterion to determine whether there exists an n such that the adaptive error probability $P_{e,\mathbb{A}}^{(n)} = 0$ [70], and then evidently $\xi_{\mathbb{A}}(\mathcal{M}, \overline{\mathcal{M}}) = +\infty$; conversely, we know that in all other cases, the adaptive Chernoff exponent is $\xi_{\mathbb{A}}(\mathcal{M}, \overline{\mathcal{M}}) < +\infty$ [77].

Duan *et al.* [71] have attempted a characterization of the channel pairs such that there exists an n with $P_{e,\mathbb{P}}^{(n)} = 0$, and have given a simple sufficient condition for the contrary. Namely, the existing result [71, Cor. 1] states that if $\operatorname{span}\{E_i^\dagger F_j\}$ contains a positive definite element, then for all n we have $P_{e,\mathbb{P}}^{(n)} > 0$. The following proposition, which makes the result of [71] quantitative, is the main result of this section.

Proposition 2.1. *Let $\alpha_{ij} \in \mathbb{C}$ be such that $\sum_{ij} |\alpha_{ij}|^2 = 1$ and $P := \sum_{ij} \alpha_{ij} E_i^\dagger F_j > 0$, i.e. P is assumed to be positive definite. Then for all n,*

$$P_{e,\mathbb{P}}^{(n)} \geq \frac{1}{4} \lambda_{\min}(P)^{4n},$$

where $\lambda_{\min}(A)$ denotes the smallest eigenvalue of the Hermitian operator A. Consequently,

$$\xi^{\mathbb{P}}(\mathcal{M}, \overline{\mathcal{M}}) \leq 4 \log \|P^{-1}\|_\infty.$$

Proof. We begin with a test state φ as in the above description of the most general non-adaptive strategy for the channels $\mathcal{M}$ and $\overline{\mathcal{M}}$, so that the two output states are $\rho = (\mathrm{id}_R \otimes \mathcal{M})\varphi$, $\sigma = (\mathrm{id}_R \otimes \overline{\mathcal{M}})\varphi$. By well-known inequalities [78], it holds

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

where $F(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1$ is the fidelity. Thus, it will be enough to lower bound the fidelity between the output states of the two channels. With $\tau = \operatorname{Tr}_R |\varphi\rangle\langle\varphi|$, we have:

$$\begin{aligned}
F(\rho,\sigma) &= \|\sqrt{\rho}\sqrt{\sigma}\|_1 \\
&\geq \operatorname{Tr} \sqrt{\rho}\sqrt{\sigma} \\
&\geq \operatorname{Tr} \rho\sigma \\
&= \sum_{ij} |\operatorname{Tr} E_i^\dagger F_j \tau|^2 \\
&\geq \left| \sum_{ij} \alpha_{ij} \operatorname{Tr} E_i^\dagger F_j \tau \right|^2 \\
&= |\operatorname{Tr} \tau P|^2.
\end{aligned}$$

Here, the second line is by standard inequalities for the trace norm, the third is because of $\rho \leq \sqrt{\rho}$, the fourth is a formula from [71, Sec. II], in the fifth we used Cauchy-Schwarz inequality and in the last line the definition of P. Since τ, like φ, ranges over all states, we get

$$F(\rho,\sigma)^2 \geq \lambda_{\min}(P)^4,$$

and so

$$P_e \geq \frac{1}{4}\lambda_{\min}(P)^4.$$

We can apply the same reasoning to $\mathcal{M}^{\otimes n}$ and $\overline{\mathcal{M}}^{\otimes n}$, for which the vector $(\alpha_{ij})^{\otimes n}$ is eligible and leads to the positive definite operator $P^{\otimes n}$. Thus,

$$P^{(n)}_{e,\mathbb{P}} \geq \frac{1}{4}\lambda_{\min}\left(P^{\otimes n}\right)^4 = \frac{1}{4}\lambda_{\min}(P)^{4n}.$$

Taking the limit and noting $\lambda_{\min}(P)^{-1} = \left\|P^{-1}\right\|_\infty$ concludes the proof. ∎

Next we show that two channels defined by Harrow *et al.* [36] yield an example of a pair with $\xi^{\mathbb{P}}(\mathcal{M},\overline{\mathcal{M}}) < +\infty$, yet $\xi^{\mathbb{A}}(\mathcal{M},\overline{\mathcal{M}}) = +\infty$ because indeed $P^{(2)}_{e,\mathbb{A}} = 0$.

Example 1. *Harrow* et al. *[36] proposed the following two entanglement-breaking channels, from $A \otimes C = \mathbb{C}^2 \otimes \mathbb{C}^2$ (two qubits) to $B = \mathbb{C}^2$ (one qubit):*

$$\mathcal{M}(\rho^A \otimes \gamma^C) = |0\rangle\langle 0| \langle 0|\gamma|0\rangle + |0\rangle\langle 0| \langle 1|\gamma|1\rangle \langle 0|\rho|0\rangle + \frac{1}{2}I \langle 1|\gamma|1\rangle \langle 1|\rho|1\rangle,$$
$$\overline{\mathcal{M}}(\rho^A \otimes \gamma^C) = |+\rangle\langle +| \langle 0|\gamma|0\rangle + |1\rangle\langle 1| \langle 1|\gamma|1\rangle \langle +|\rho|+\rangle + \frac{1}{2}I \langle 1|\gamma|1\rangle \langle -|\rho|-\rangle,$$

extended by linearity to all states. Here, $|0\rangle, |1\rangle$ is the computational basis (Z eigenbasis) of the qubits, while $|+\rangle, |-\rangle$ is the Hadamard basis (X eigenbasis).

In words, both channels measure the qubit C in the computational basis. If the outcome is '0', they each prepare a pure state on B (ignoring the input in A): $|0\rangle\langle 0|$ for $\mathcal{M}$, $|+\rangle\langle +|$ for $\mathcal{N}$. If the outcome is '1', they each make a measurement on A and prepare an output state on B depending on its outcome: standard basis measurement for $\mathcal{M}$ with $|0\rangle\langle 0|$ on outcome '0' and the maximally mixed state $\frac{1}{2}I$ on outcome '1'; Hadamard basis measurement for $\mathcal{N}$ with $|1\rangle\langle 1|$ on outcome '+' and the maximally mixed state $\frac{1}{2}I$ on outcome '-'. In [36], a simple adaptive strategy for $n = 2$ uses of the channel is given that discriminates $\mathcal{M}$ and $\overline{\mathcal{M}}$ perfectly: The first instance of the channel is fed with $|0\rangle\langle 0| \otimes |0\rangle\langle 0|$, resulting in an output state ρ_1; the second instance of the channel is fed with $|1\rangle\langle 1| \otimes \rho_1$; the output state ρ_2 of the second instance

is $|0\rangle\langle 0|$ *if the unknown channel is* $\mathcal{M}$*, and* $|1\rangle\langle 1|$ *if the unknown channel is* $\overline{\mathcal{M}}$*, so a computational basis measurement reveals it. Note that no auxiliary system* R *is needed, but the feed-forward nevertheless requires a qubit of quantum memory for the strategy to be implemented. In any case, this proves that* $P_{e,\mathbb{A}}^{(2)} = 0$*. In [36], it is furthermore proved that for all* $n \geq 1$*,* $P_{e,\mathbb{P}}^{(n)} > 0$*.*

We now show that Proposition 2.1 is applicable to yield an exponential lower bound on the non-adaptive error probability. The Kraus operators of the two channels can be chosen as follows:

$$\begin{aligned}
\mathcal{M} : E_i \in \Big\{ & |0\rangle^B \langle 00|^{AC}, & \overline{\mathcal{M}} : F_j \in \Big\{ & |+\rangle^B \langle 00|^{AC}, \\
& |0\rangle^B \langle 10|^{AC}, & & |+\rangle^B \langle 10|^{AC}, \\
& |0\rangle^B \langle 01|^{AC}, & & |1\rangle^B \langle +1|^{AC}, \\
& |0\rangle^B \langle 11|^{AC}/\sqrt{2}, & & |0\rangle^B \langle -1|^{AC}/\sqrt{2}, \\
& |1\rangle^B \langle 11|^{AC}/\sqrt{2} \Big\}, & & |1\rangle^B \langle -1|^{AC}/\sqrt{2} \Big\}.
\end{aligned}$$

Thus, the products $E_i^\dagger F_j$ *include the matrices*

$$\begin{aligned}
E_1^\dagger F_1 &= \sqrt{\frac{1}{2}}\, |00\rangle\langle 00|, \\
E_2^\dagger F_2 &= \sqrt{\frac{1}{2}}\, |10\rangle\langle 10|, \\
E_5^\dagger F_3 &= \sqrt{\frac{1}{2}}\, |11\rangle\langle +1|, \\
E_5^\dagger F_5 &= \frac{1}{2}\, |11\rangle\langle -1|, \\
E_3^\dagger F_4 &= \sqrt{\frac{1}{2}}\, |01\rangle\langle -1|,
\end{aligned}$$

from which we can form, by linear combination, the projectors

$$\begin{aligned}
E_1^\dagger F_1 &= \sqrt{\frac{1}{2}}\, |0\rangle\langle 0| \otimes |0\rangle\langle 0|, \\
E_2^\dagger F_2 &= \sqrt{\frac{1}{2}}\, |1\rangle\langle 1| \otimes |0\rangle\langle 0|, \\
E_5^\dagger F_3 - E_5^\dagger F_5 &= \sqrt{\frac{1}{2}}\, |1\rangle\langle 1| \otimes |1\rangle\langle 1|, \\
E_3^\dagger F_4 - E_5^\dagger F_5 &= \sqrt{\frac{1}{2}}\, |-\rangle\langle -| \otimes |1\rangle\langle 1|,
\end{aligned}$$

whose sum is indeed positive definite, so we get an exponential lower bound on $P_{e,\mathbb{P}}^{(n)}$ *and hence a finite upper bound on* $\xi_{\mathbb{P}}(\mathcal{M},\overline{\mathcal{M}})$.

Note that a lower bound is the Chernoff bound of the two pure output states $|0\rangle\langle 0| = \mathcal{M}(|00\rangle\langle 00|)$ *and* $|+\rangle\langle +| = \overline{\mathcal{M}}(|00\rangle\langle 00|)$*, which is* $\log 2 = 1$*, so* $\xi^{\mathbb{P}}(\mathcal{M},\overline{\mathcal{M}}) \geq 1$*. It seems reasonable to conjecture that this is optimal, but we do not have at present a proof of it.*

To get a concrete upper bound on $\xi_{\mathbb{P}}(\mathcal{M},\overline{\mathcal{M}})$ *from the above method, we make the ansatz*

$$\begin{aligned} P &= \alpha E_1^\dagger F_1 + \alpha E_2^\dagger F_2 + \beta\sqrt{\frac{1}{2}} E_5^\dagger F_3 + \beta\sqrt{\frac{1}{2}} E_3^\dagger F_4 - 2\beta E_5^\dagger F_5 \\ &= \alpha\sqrt{\frac{1}{2}} I \otimes |0\rangle\langle 0| + \beta\sqrt{\frac{1}{2}}(|1\rangle\langle 1| + |-\rangle\langle -|) \otimes |1\rangle\langle 1|, \end{aligned}$$

where $\alpha, \beta > 0$ *and* $2\alpha^2 + 5\beta^2 = 1$*. Now* P *consists of two orthogonal pieces with easily calculated minimum eigenvalues, which are* $\alpha\sqrt{\frac{1}{2}}$ *and* $\beta\sqrt{2}\sin^2\frac{\pi}{8}$*. Since* $\lambda_{\min}(P)$ *will be the smaller of the two, we optimize it by making the two values equal, i.e. we want* $\alpha = 2\beta\sin^2\frac{\pi}{8}$*. Inserting this in the normalization condition and solving for* β *yields* $\beta^2 = \frac{1}{8\sin^4\frac{\pi}{8}+5}$*, thus*

$$\lambda_{\min}(P) = \sqrt{\frac{2}{8\sin^4\frac{\pi}{8}+5}}\sin^2\frac{\pi}{8} = \frac{2-\sqrt{2}}{4\sqrt{4-\sqrt{2}}} \approx 0.091,$$

where we have used the identity $\sin^2\frac{\pi}{8} = \frac{1}{2}(1-\sqrt{\frac{1}{2}})$*. Thus,*

$$\xi^{\mathbb{P}}(\mathcal{M},\overline{\mathcal{M}}) \leq 4\log\frac{4\sqrt{4-\sqrt{2}}}{2-\sqrt{2}} \approx 13.83.$$

Thus, we learn that there are channels, entanglement-breaking channels at that, for which the adaptive and the non-adaptive Chernoff exponents are different; in fact, the separation is maximal, in that the former is $+\infty$ while the latter is finite. It should be noted that this separation is a robust phenomenon, and not for example related to the perfect finite-copy distinguishability. Namely, by simply mixing our example channels with the same small fraction of the fully depolarizing channel, we get two new channels $\mathcal{M}'$ and $\overline{\mathcal{M}}'$ with only smaller non-adaptive Chernoff bound, $\xi^{\mathbb{P}}(\mathcal{M}',\overline{\mathcal{M}}') \leq \xi^{\mathbb{P}}(\mathcal{M},\overline{\mathcal{M}}) < +\infty$, but the fully general adaptive strategies yield arbitrarily large $\xi^{\mathbb{A}}(\mathcal{M}',\overline{\mathcal{M}}')$, as it is based on a two-copy strategy.

Furthermore, since the error rate tradeoff function $B_e^{\mathbb{P}}(r|\mathcal{M}\|\overline{\mathcal{M}})$ is continuous near $r = \xi^{\mathbb{P}}(\mathcal{M},\overline{\mathcal{M}})$, whereas the adaptive variant $B_e^{\mathbb{A}}(r|\mathcal{M}\|\overline{\mathcal{M}})$ is infinite, we automatically get separations in the Hoeffding setting, as well.

2.4 Preliminaries on quantum measurements

A general quantum state evolution from A to B is written as a cptp map $\mathcal{M}$ from the space $\mathcal{T}^A$ to the space $\mathcal{T}^B$, where $\mathcal{T}^A$ and $\mathcal{T}^B$ are the sets of Hermitian matrices on the Hilbert spaces A and B, respectively. When we make a measurement on the initial system A, we obtain the measurement outcome K and the resultant state on the output system B. To describe this situation, we use a set $\{\kappa_k\}_{k\in\mathcal{K}}$ of cp maps from the space $\mathcal{T}^A$ to the space $\mathcal{T}^B$ such that $\sum_{k\in\mathcal{K}}\kappa_k$ is trace preserving. (For simplicity, here we assume the set $\mathcal{K}$ to be discrete.) Since it is a decomposition of a cptp map, it is often called a cp-map valued measure. In this case, when the initial state on A is ρ and the outcome k is observed with probability $\operatorname{Tr}\kappa_k(\rho)$, where the resultant state on B is $\kappa_k(\rho)/\operatorname{Tr}\kappa_k(\rho)$. A state on the composite system of the classical system K and the quantum B is written as $\sum_{k\in\mathcal{K}}|k\rangle\langle k|\otimes\rho_{B|k}$, which belongs to the vector space $\mathcal{T}^{KB} := \sum_{k\in\mathcal{K}}|k\rangle\langle k|\otimes\mathcal{T}^B$. The above measurement process can be written as the following cptp $\mathcal{E}$ map from $\mathcal{T}^A$ to $\mathcal{T}^{KB}$.

$$\mathcal{E}(\rho) := \sum_{k\in\mathcal{K}} |k\rangle\langle k|\otimes\kappa_k(\rho). \tag{2.8}$$

In the following, both of the above cptp map $\mathcal{E}$ and a cp-map valued measure are called a quantum instrument. A cp-map valued measure has the following form.

Lemma 2.1 (Cf. [65, Thm. 7.2])**.** *Let $\boldsymbol{\kappa} = \{\kappa_\omega : A \to B\}_\omega$ be an instrument (i.e. a cp-map valued measure) with an input system A and an output system B. Then there exist a POVM $\boldsymbol{M} = \{M_\omega\}$ on a Hilbert space A and cptp maps κ'_ω from A to B for each outcome ω, such that for any density operator ρ,*

$$\kappa_\omega(\rho) = \kappa'_\omega\left(\sqrt{M_\omega}\rho\sqrt{M_\omega}\right). \qquad \blacksquare$$

A general POVM can be lifted to a projective valued measure (PVM), as follows.

Lemma 2.2 (Naimark's theorem [79])**.** *Given a positive operated-valued measure (POVM) $\boldsymbol{M} = \{M_\omega\}_{\omega\in\Omega}$ on A with a discrete measure space Ω, there exists a larger Hilbert space C including A and a projection-valued measure (PVM) $\boldsymbol{E} = \{E_\omega\}_{\omega\in\Omega}$ on C such that*

$$\operatorname{Tr}\rho M_\omega = \operatorname{Tr}\rho E_\omega \quad \forall\rho\in S^A,\ \omega\in\Omega. \qquad \blacksquare$$

Combining these two lemmas, we have the following corollary.

Corollary 2.1. *Let $\boldsymbol{\kappa} = \{\kappa_\omega : A \to B\}_\omega$ be an instrument (i.e. a cp-map valued measure) with an input system A and an output system B. Then there exist a PVM $\boldsymbol{E} = \{E_\omega\}$ on a larger Hilbert space C including A and cptp maps κ''_ω from C to B for each outcome ω, such that for any density operator ρ,*

$$\kappa_\omega(\rho) = \kappa''_\omega\left(E_\omega \rho E_\omega\right). \tag{2.9}$$

Proof. First, using Lemma 2.1, we choose a POVM $\boldsymbol{M} = \{M_\omega\}$ on a Hilbert space A and cptp maps $\boldsymbol{\kappa}'_\omega$ from A to B for each outcome ω. Next, using Lemma 2.2, we choose a larger Hilbert space C including A and a projection-valued measure (PVM) $\boldsymbol{E} = \{E_\omega\}_{\omega\in\Omega}$ on C. We denote the projection from C to A by P. Then, we have

$$(E_\omega P)^\dagger E_\omega P = P E_\omega P = M_\omega = \sqrt{M_\omega}\sqrt{M_\omega} \tag{2.10}$$

for any $\omega \in \Omega$. Thus, there exists a partial isomerty V_ω from C to A such that $\sqrt{M_\omega} = V_\omega E_\omega P$. Hence, we have

$$\kappa_\omega(\rho) = \kappa'_\omega\left(\sqrt{M_\omega}\rho\sqrt{M_\omega}\right) = \kappa'_\omega\left(V_\omega E_\omega P \rho P E_\omega V_\omega^\dagger\right) = \kappa'_\omega\left(V_\omega E_\omega \rho E_\omega V_\omega^\dagger\right).$$

Defining cptp maps κ''_ω by $\kappa''_\omega(\rho) = \kappa'_\omega(V_\omega \rho V_\omega^\dagger)$. This completes the proof. ∎

2.5 Discrimination of classical-quantum channels

A cq-channel is defined with respect to a set $\mathcal{X}$ of input signals and the Hilbert space B of the output states. In this case, the channel from $\mathcal{X}$ to B is described by the map from the set $\mathcal{X}$ to the set of density operators in B; as such, a cq-channel is given as $\mathcal{N} : x \to \rho_x$, where ρ_x denotes the output state when the input is $x \in \mathcal{X}$. Our goal is to distinguish between two cq-channels, $\mathcal{N} : x \to \rho_x$ and $\overline{\mathcal{N}} : x \to \sigma_x$. Here, we do not assume any condition for the set $\mathcal{X}$, except that it is a measurable space and that the channels are measurable maps (with the usual Borel sets on the state space $\mathcal{S}^B$). In particular, it might be an infinite set. We consider the scenario when $n \to \infty$ uses of the unknown channel are provided. The task is to discriminate two hypotheses, the null hypothesis $H_0 : \mathcal{N}$ versus the alternative hypothesis $H_1 : \overline{\mathcal{N}}$ where n (independent) uses of the unknown channel are provided. Then, the challenge we face is to make a decision in favor of the true channel based on n inputs

$\vec{x}_n = (x_1, \ldots, x_n)$ and corresponding output states on $B^n = B_1 \cdots B_n$; note that the input $\vec{x}_n = (x_1, \ldots, x_n)$ is generated by a very complicated joint distribution of n random variables, which – except for x_1 – depend on the actual channel. Hence, they are written with the capitals as $X^n = X_1, \ldots, X_n$ when they are treated as random variables.

2.5.1 Adaptive method

General protocol for cq-channels

To study the adaptive discrimination of cq-channels, the general strategy for discrimination of qq-channels in Definition 1 should be tailored to the cq-channels. We argue that the general strategy of Definition 1 can w.l.o.g. be replaced by the kind of strategy with the instrument and only classical feed-forward when the hypotheses are a pair of cq-channels. This in particular will turn out to be crucial since we consider general cq-channels with arbitrary (continuous) input alphabet.

The first input is chosen subject to the distribution $p_{X_1}(x_1)$. The receiver receives the output ρ_{x_1} or σ_{x_1} on B_1. Dependently of the input x_1, the receiver applies the first quantum instrument $B_1 \to K_1 R_2$, where R is the quantum memory system and K_1 is the classical outcome. The receiver sends the outcome K_1 to the sender. Then, the sender choose the second input x_2 according to the conditional distribution $p_{X_2|X_1,K_1}(x_2|x_1, k_1)$. The receiver receives the second output ρ_{x_2} or σ_{x_2} on B_2. Dependently of the previous outcome k_1 and the previous inputs x_1, x_2, the receiver applies the second quantum instrument $B_2 R_2 \to K_2 R_3$, and sends the outcome K_2 to the sender. The third input is chosen as the distribution $p_{X_3|X_1,X_2,K_1,K_2}(x_2|x_1, x_2, k_1, k_2)$.

In the same way as the above, the m-th step is given as follows. The sender chooses the m-th input x_m according to the conditional distribution $p_{X_m|\vec{X}_{m-1},\vec{K}_{m-1}}(x_m|\vec{x}_{m-1}, \vec{k}_{m-1})$. The receiver receives the second output ρ_{x_m} or σ_{x_m} on B_m. The remaining processes need the following divided cases. For $m < n$, dependently of the previous outcomes $\vec{k}_{m-1} := (k_1, \ldots, k_{m-1})$ and the previous inputs $\vec{x}_m := (x_1, \ldots, x_m)$, the receiver applies the m-th quantum instrument $\mathcal{E}_m : R_m B_m \to K_m R_{m+1}$, and sends the outcome k_m to the sender. For $m = n$, dependently of the previous outcomes $\vec{K}_{n-1}$ and the previous inputs $\vec{Y}_n$, the receiver measures the final state on $R_n B_n$ with the binary POVM $(T_{n|\vec{k}_{n-1},\vec{x}_n}, I - T_{n|\vec{k}_{n-1},\vec{x}_n})$, where hypothesis $\mathcal{N}$ (res. $\overline{\mathcal{N}}$) is accepted if and only if the first (res. second) outcome clicks.

Remark 2.1 (Relation to general setting with qq-channels)**.** *Here, we discuss how to derive the above setting from the general setting presented in*

introduction for cq-channels. In the case with cq-channel, the input needs to be a classical element in the discrete set $\mathcal{X}$. To decide the classical input, we need to apply measurement after the application of the m-th cptp map $\mathcal{F}_m$. That is, we need to replace the m-th cptp map $\mathcal{F}_m$ by a quantum instrument $\mathcal{E}_m : R_m B_m \to K_m R_{m+1}$, and sends the outcome K_m to the sender. Hence, the obtained procedure is equivalent to the procedure given above.

Note however that this way of thinking of a cq-channel as a special type of qq-channel is restricted to discrete input alphabets; for general, in particular continuous input alphabet to the channels $\mathcal{N}$ and $\overline{\mathcal{N}}$, we directly use the description above

Protocol with PVM for cq-channels

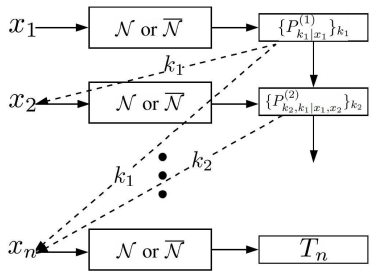

Figure 2.1: Adaptive strategy for cq-channel discrimination. The classical outputs of the PVMs are employed to decide the inputs adaptively while the post-measurement state remains in disposal of the next measurement; that is, after the m-th step, a quantum system on the subspace spanned by $\left(\operatorname{Im}\Pi^{(1)}_{k_1|x_1} \otimes I^{\otimes(m-1)}\right)\cdots\left(\operatorname{Im}\Pi^{(m)}_{\vec{k}_m|\vec{x}_m}\right)$ passes to the next measurement.

The general procedure for discriminating cq-channels can be rewritten as follows using PVMs.

To start, Fig. 5.1 illustrates the general protocol with PVMs, which we shall describe now In the following, according to Naimarks dilation theorem, in each m-th step, we choose a sufficiently large space B_m including the original space B_m such that the measurement is a PVM.

The first input is chosen subject to the distribution $p_{X_1}(x_1)$. Then the output state is measured by a projection-valued measure (PVM) $\left\{\Pi^{(1)}_{k_1|x_1}\right\}_{k_1}$ on B_1. The second input is then chosen according to the distribution $p_{X_2|X_1,K_1}(x_2|x_1,k_1)$. Then, a PVM $\left\{\Pi^{(2)}_{k_2,k_1|x_1,x_2}\right\}_{k_1,k_2}$ is made on B_1B_2, which

satisfies $\sum_{k_2} \Pi^{(2)}_{k_2,k_1|x_1,x_2} = \Pi^{(1)}_{k_1|x_1} \otimes I$. The third input is chosen as the distribution $p_{X_3|X_1,X_2,K_1,K_2}(x_3|x_1,x_2,k_1,k_2)$, etc. Continuing, the m-th step is given as follows. the sender chooses the m-th input x_m according to the conditional distribution $p_{X_m|\vec{X}_{m-1},\vec{K}_{m-1}}(x_m|\vec{x}_{m-1},\vec{k}_{m-1})$. The receiver receives the m-th output ρ_{x_m} or σ_{x_m} on B_m.

The description of the remaining processing requires that we distinguish two cases.

- For $m < n$, depending on the previous outcomes $\vec{k}_{m-1} = (k_1, \ldots, k_{m-1})$ and the previous inputs $\vec{x}_m = (x_1, \ldots, x_m)$, as the m-th projective measurement, the receiver applies a PVM $\left\{\Pi^{(m)}_{\vec{k}_m|\vec{x}_m}\right\}_{\vec{k}_m}$ on $B_1B_2\cdots B_m$, which satisfies the condition $\sum_{k_m} \Pi^{(m)}_{\vec{k}_m|\vec{x}_m} = \Pi^{(m-1)}_{\vec{k}_{m-1}|\vec{x}_{m-1}} \otimes I$. He sends the outcome k_m to the sender.

- For $m = n$, dependently of the inputs $\vec{x}_n$, the receiver measures the final state on $B_1B_2\cdots B_n$ with the binary POVM $(T_{\vec{x}_n}, I - T_{\vec{x}_n})$ on $B_1B_2\cdots B_n$, where hypothesis $\mathcal{N}$ (res. $\overline{\mathcal{N}}$) is accepted if and only if the first (res. second) outcome clicks.

Proposition 2.2. *Any general procedure given in Subsubsection 2.5.1 can be rewritten in the above form.*

Proof. Recall Corollary 2.1 given in Section 2.4. Due to Corollary 2.1, when the Hilbert space B can be chosen sufficiently large, any state reduction written by a cp-map valued measure $\{\Gamma_{k_1|x_1}\}_{k_1}$ can also be written as the combination of a PVM $\{\Pi^{(1)}_{k_1|x_1}\}_{k_1}$ and a state change by a cptp map Λ_{k_1,x_1} depending on the measurement outcome k_1 such that $\Gamma_{k_1|x_1}(\rho) = \Lambda_{k_1,x_1}(\Pi^{(1)}_{k_1|x_1}\rho\Pi^{(1)}_{k_1|x_1})$ for k_1, x_1. Hence, we have $\Gamma_{k_1|x_1}(\rho) = \Lambda_{k_1,x_1}(\Pi_{k_1|x_1}\rho\Pi_{k_1|x_1})$ for k_1, x_1.

Then, we treat the cptp map Λ_{k_1,x_1} as a part of the next measurement. Let $\{\Gamma_{k_2|x_1,x_2,k_1}\}_{k_2}$ be the quantum instrument to describe the second measurement. We define the quantum instrument $\{\overline{\Gamma}_{k_2|x_1,x_2,k_1}\}_{k_2}$ as $\overline{\Gamma}_{k_2|x_1,x_2,k_1}(\rho) := \Gamma_{k_2|x_1,x_2,k_1}(\Lambda_{k_1,x_1}(\rho))$. Applying Corollary 2.1 to the quantum instrument $\{\overline{\Gamma}_{k_2|x_1,x_2,k_1}\}_{k_2}$, we choose the PVM $\{\Pi^{(2)}_{k_2|x_1,x_2,k_1}\}_{k_2}$ on $\operatorname{Im}\Pi^{(1)}_{k_1|x_1} \otimes B_2$ and the state change by a cptp map $\Lambda_{k_1,k_2,x_1,x_2}$ depending on the measurement outcome k_2 to satisfy 2.9. Since $\sum_{k_1,k_2} \Pi^{(2)}_{k_2|x_1,x_2,k_1}$ is the identity on B_1B_2, setting $\Pi^{(2)}_{k_1,k_2|x_1,x_2} := \Pi^{(2)}_{k_2|x_1,x_2,k_1}$, we define the PVM $\{\Pi^{(2)}_{k_1,k_2|x_1,x_2}\}_{k_1,k_2}$ on B_1B_2. In the same way, for the m-th step, using a quantum instrument $\{\Gamma_{k_m|\vec{x}_m,\vec{k}_{m-1}}\}_{k_m}$, cptp maps $\Lambda_{\vec{k}_{m-1},\vec{x}_{m-1}}$, and Corollary 2.1, we define the PVM $\{\Pi^{(m)}_{k_m|\vec{x}_m,\vec{k}_{m-1}}\}_{k_m}$

on $\mathrm{Im}\,\Pi^{(m-1)}_{\vec{k}_{m-1}|\vec{x}_{m-1}} \otimes B_m$ and the state change by cptp maps $\Lambda_{\vec{k}_m,\vec{x}_m}$. Then, setting $\Pi^{(m)}_{\vec{k}_m|\vec{x}_m} := \Pi^{(m)}_{k_m|\vec{x}_m,\vec{k}_{m-1}}$, we define the PVM $\{\Pi^{(m)}_{\vec{k}_m|\vec{x}_m}\}_{\vec{k}_m}$ on $B_1B_2\cdots B_m$.

In the n-th step, i.e., the final step, using the binary POVM $(T_{n|\vec{k}_{n-1},\vec{x}_n}, I - T_{n|\vec{k}_{t-1},\vec{x}_n})$ and cptp maps $\Lambda_{\vec{k}_{n-1},\vec{x}_{n-1}}$, we define the binary POVM $(T_{\vec{x}_n}, I - T_{\vec{x}_n})$ on $B_1B_2\cdots B_n$ as follows.

$$T_{\vec{x}_n} := \sum_{\vec{k}_n} \Lambda^\dagger_{\vec{k}_{n-1},\vec{x}_{n-1}}(T_{n|\vec{k}_{n-1},\vec{x}_n}), \tag{2.11}$$

where $\Lambda^\dagger_{\vec{k}_{n-1},\vec{x}_{n-1}}$ is defined as $\mathrm{Tr}\,\Lambda_{\vec{k}_{n-1},\vec{x}_{n-1}}(\rho)X = \mathrm{Tr}\,\rho\Lambda^\dagger_{\vec{k}_{n-1},\vec{x}_{n-1}}(X)$. In this way, the general protocol in given in Subsubsection 2.5.1 has been converted a protocol given in this subsection. ■

It is implicit that the projective measurement $\{\Pi^{(m)}_{\vec{k}_m|\vec{x}_m}\}_{\vec{k}_m}$ includes first projecting the output from the quantum memory onto a subspace spanned by $\{\Pi^{(m-1)}_{\vec{k}_{m-1}|\vec{x}_{m-1}}\}_{\vec{k}_{m-1}}$, and then finding $\vec{k}_m$ in the entire subspace of $\mathrm{Im}\,\Pi^{(m-1)}_{\vec{k}_{m-1}|\vec{x}_{m-1}} \otimes B$. Hence, $\{\Pi^{(m)}_{\vec{k}_m|\vec{k}_{m-1},\vec{x}_{m-1}}\}_{\vec{k}_m}$ can be regarded as a PVM on $B^{\otimes m}$ and from the construction

$$\sum_{\vec{k}_m} \Pi^{(m)}_{\vec{k}_m|\vec{x}_m} = (\Pi^{(1)}_{k_1|x_1} \otimes I^{\otimes(m-1)})\cdots(\Pi^{(m-1)}_{\vec{k}_{m-1}|\vec{x}_{m-1}} \otimes I),$$

which shows that the PVMs commute.

Notice also that

$$\Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}} \le \Pi^{(n-2)}_{\vec{k}_{n-2}|\vec{x}_{n-2}} \otimes I \le \cdots \le \Pi^{(1)}_{\vec{k}_1|\vec{x}_1} \otimes I^{\otimes(n-2)}.$$

When the true channel is $\mathcal{N} : x \to \rho_x$, the state before the final measurement is

$$\begin{aligned}\rho^{(n)} := \sum_{\vec{x}_n,\vec{k}_{n-1}} & p_{X_1}(x_1)\cdots p_{X_n|\vec{X}_{n-1},\vec{K}_{n-1}}(x_n|\vec{x}_{n-1},\vec{k}_{n-1}) \\ & \Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}}\big(\rho_{x_1}\otimes\cdots\otimes\rho_{x_n}\big)\Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}} \otimes |\vec{x}_n\rangle\langle\vec{x}_n|,\end{aligned} \tag{2.12}$$

where here we need to store the information for inputs $\vec{x}_n$. Similarly, when the true channel is $\overline{\mathcal{N}} : x \to \sigma_x$

$$\begin{aligned}\sigma^{(n)} := \sum_{\vec{x}_n,\vec{k}_{n-1}} & p_{X_1}(x_1)\cdots p_{X_n|\vec{X}_{n-1},\vec{K}_{n-1}}(x_n|\vec{x}_{n-1},\vec{k}_{n-1}) \\ & \Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}}\big(\sigma_{x_1}\otimes\cdots\otimes\sigma_{x_n}\big)\Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}} \otimes |\vec{x}_n\rangle\langle\vec{x}_n|.\end{aligned} \tag{2.13}$$

A test of the hypotheses $\{\mathcal{N}, \overline{\mathcal{N}}\}$ on the true channel is a two-valued POVM $\{T_n, I - T_n\}$, where T_n is given as a Hermitian operator $\sum_{\vec{x}_n} T_{\vec{x}_n} \otimes |\vec{x}_n\rangle\langle\vec{x}_n|$ on $B^{\otimes n} \otimes X^{\otimes n}$ satisfying $0 \leq T_n \leq I$. Overall, our strategy to distinguish the channels $\{\mathcal{N}, \overline{\mathcal{N}}\}$ when n independent uses of each are available, is given by the triple $\mathcal{T}_n \coloneqq (\{\Pi^{(m)}_{\vec{k}_m|\vec{x}_m}\}_{m=1}^{n-1}, \{p_{X_m|\vec{X}_{m-1},\vec{K}_{m-1}}\}_{m=1}^{n}, T_n)$. The n-copy error probabilities of type I and type II are respectively as follows

$$\alpha_n(\mathcal{N}\|\overline{\mathcal{N}}|\mathcal{T}_n) \coloneqq \operatorname{Tr} \rho^{(n)}(I - T_n),$$
$$\beta_n(\mathcal{N}\|\overline{\mathcal{N}}|\mathcal{T}_n) \coloneqq \operatorname{Tr} \sigma^{(n)} T_n.$$

The generalized Chernoff and Hoeffding quantities introduced in the introduction read as follows in the present cq-channel case for a given class $\mathbb{S}_n \subset \mathbb{A}_n$:

$$C^{\mathbb{S}_n}(a, b|\mathcal{N}\|\overline{\mathcal{N}}) \coloneqq \sup_{\{\mathcal{T}_n\}} \left\{ \liminf_{n\to\infty} -\frac{1}{n} \log\left(2^{an}\alpha_n(\mathcal{N}\|\overline{\mathcal{N}}|\mathcal{T}_n) + 2^{bn}\beta_n(\mathcal{N}\|\overline{\mathcal{N}}|\mathcal{T}_n)\right)\right\}, \tag{2.14}$$

$$B_e^{\mathbb{S}_n}(r|\mathcal{N}\|\overline{\mathcal{N}}) \coloneqq \sup_{\{\mathcal{T}_n\}} \left\{ \liminf_{n\to\infty} -\frac{1}{n} \log\left(\alpha_n(\mathcal{N}\|\overline{\mathcal{N}}|\mathcal{T}_n)\right) \,\middle|\, \liminf_{n\to\infty} -\frac{1}{n} \log\left(\beta_n(\mathcal{N}\|\overline{\mathcal{N}}|\mathcal{T}_n)\right) \geq r \right\}, \tag{2.15}$$

where a, b, are arbitrary real numbers and r is an arbitrary non-negative number.

2.5.2 Auxiliary results and techniques

We set $\rho_x \coloneqq \mathcal{N}(x)$ and $\sigma_x \coloneqq \overline{\mathcal{N}}(x)$, and define

$$C(a, b) \coloneqq \sup_x \sup_{0\leq\alpha\leq 1} (1-\alpha) D_\alpha(\rho_x\|\sigma_x) - \alpha a - (1-\alpha) b \tag{2.16}$$
$$= \sup_{0\leq\alpha\leq 1} (1-\alpha) D_\alpha(\mathcal{N}\|\overline{\mathcal{N}}) - \alpha a - (1-\alpha) b, \tag{2.17}$$
$$B(r) \coloneqq \sup_x \sup_{0\leq\alpha\leq 1} \frac{\alpha-1}{\alpha}\big(r - D_\alpha(\rho_x\|\sigma_x)\big) = \sup_{0\leq\alpha\leq 1} \frac{\alpha-1}{\alpha}\big(r - D_\alpha(\mathcal{N}\|\overline{\mathcal{N}})\big), \tag{2.18}$$

where $D_\alpha(\mathcal{N}\|\overline{\mathcal{N}}) \coloneqq \sup_x D_\alpha(\rho_x\|\sigma_x)$ and $D_\alpha(\rho_x\|\sigma_x) \coloneqq \frac{1}{\alpha-1} \log \operatorname{Tr} \rho_x^\alpha \sigma_x^{1-\alpha}$ is a quantum extension of the Rényi relative entropy.

Since $D_\alpha(\rho_x\|\sigma_x)$ is monotonically increasing for α, $D_\alpha(\mathcal{N}\|\overline{\mathcal{N}})$ is monotonically increasing for α. Thus,

$$\begin{aligned}\lim_{\alpha\to 1} D_\alpha(\mathcal{N}\|\overline{\mathcal{N}}) &= \sup_{0\le\alpha\le 1} D_\alpha(\mathcal{N}\|\overline{\mathcal{N}}) = \sup_{0\le\alpha\le 1}\sup_x D_\alpha(\rho_x\|\sigma_x)\\ &= \sup_x \sup_{0\le\alpha\le 1} D_\alpha(\rho_x\|\sigma_x) = \sup_x D(\rho_x\|\sigma_x) = D(\mathcal{N}\|\overline{\mathcal{N}}).\end{aligned}$$

Before stating the main results of this section we shall study the $B(r)$ function further. Since the $B(r)$ function is monotonically decreasing in r, $B(D(\mathcal{N}\|\overline{\mathcal{N}})) = 0$. To find $B(0)$, since $\frac{1-\alpha}{\alpha} D_\alpha(\mathcal{N}\|\overline{\mathcal{N}}) = D_{1-\alpha}(\overline{\mathcal{N}}\|\mathcal{N})$, we infer that $\frac{1-\alpha}{\alpha} D_\alpha(\mathcal{N}\|\overline{\mathcal{N}})$ is monotonically decreasing for α, and $D(\overline{\mathcal{N}}\|\mathcal{N}) = \lim_{\alpha\to 0} \frac{1-\alpha}{\alpha} D_\alpha(\mathcal{N}\|\overline{\mathcal{N}})$. Hence, $B(0) = D(\overline{\mathcal{N}}\|\mathcal{N})$, and $B(r) < D(\overline{\mathcal{N}}\|\mathcal{N})$ for $r > 0$.

The following lemma states the continuity of the $B(r)$ function, of which we give two different proofs. The first proof uses the known facts for the case of two states, and the cq-channel case is reduced to the former by general statements from convex analysis. The second proof is rather more ad-hoc and relies on peculiarities of the functions at hand.

Lemma 2.3. *The Hoeffding expoent $B(r)$ is continuous in r, i.e. for any non-negative real number r_0,*

$$\lim_{r\to r_0} B(r) = B(r_0). \tag{2.19}$$

Proof. The crucial difficulty in this lemma is that unlike previous works, here we allow that $|\mathcal{X}|$ is infinite. Note that in the case of a finite alphabet, we just need to note the role of the channel (as opposed to states): it is a supremum over channel inputs $x \in \mathcal{X}$, so a preliminary task is to prove that for a fixed x, i.e. a pair of states ρ_x and σ_x, the Hoeffding function is continuous. This is already known [80, Lemma 1] and follows straightforwardly from the convexity and monotonicity of the Hoeffding function. After that, the channel's Hoeffding function is the maximum over finitely many continuous functions and so continuous. However, when the alphabet size is infinite, the supremum of infinitely many continuous functions is not necessarily continuous. Nevertheless, it inherits the convexity of the functions for each x, cf. [81, Cor. 3.2.8]. Since the function is defined on the non-negative reals $\mathbb{R}_{\ge 0}$, it is continuous for all $r_0 > 0$, by the well-known and elementary fact that a convex function on an interval is continuous on its interior. It only remains to prove the continuity at $r_0 = 0$; to this end consider swapping null and alternative hypotheses and denote the corresponding Hoeffding exponent by $\overline{B}(r)$. We then find that $\overline{B}(r)$ is the inverse function of $B(r)$. Since $\overline{B}(r)$ is continuous even when it is equal to zero, i.e. at $r = D(\overline{\mathcal{N}}\|\mathcal{N})$, we conclude $B(r)$ is continuous at $r = 0$ and $B(0) = D(\overline{\mathcal{N}}\|\mathcal{N})$. ∎

Alternative proof of Lemma 2.3. Given $r_0 > 0$, there exist $\alpha_0 \in (0,1)$ and a sequence α_n such that $\alpha_n \to \alpha_0$ and $\lim_{n\to\infty} \frac{\alpha_n - 1}{\alpha_n}\big(r_0 - D_{\alpha_n}(\mathcal{N}\|\overline{\mathcal{N}})\big) = B(r_0)$. Hence, we have $\sup_{\frac{\alpha_0}{2}\le\alpha\le 1} \frac{\alpha-1}{\alpha}\big(r_0 - D_\alpha(\mathcal{N}\|\overline{\mathcal{N}})\big) = B(r_0)$. For $r > r_0$, the above supremum with r is realized by $\alpha \in [\frac{\alpha_0}{2}, 1]$. That is, $\sup_{\frac{\alpha_0}{2}\le\alpha\le 1} \frac{\alpha-1}{\alpha}\big(r - D_\alpha(\mathcal{N}\|\overline{\mathcal{N}})\big) = B(r)$. In the range $[\frac{\alpha_0}{2}, 1]$, $\frac{\alpha-1}{\alpha}\big(r - D_\alpha(\mathcal{N}\|\overline{\mathcal{N}})\big)$ is uniformly continuous for r, we have 2.19. The proof implies that larger r corresponds to larger α. To show this, we introduce $k(\alpha)$ as the first derivative of $D_\alpha(\mathcal{N}\|\overline{\mathcal{N}})$, which crucially does not depend on r. The other term, $\frac{\alpha-1}{\alpha}r$, has derivative $\frac{-r}{\alpha^2}$, so the condition for the maximum is $\frac{-r}{\alpha^2} + k(\alpha) = 0$. Now consider the optimal value α_0 for a certain $r_0 > 0$, so the above equation is satisfied for r_0 and α_0. If we now consider $r > r_0$, the same $\alpha = \alpha_0$ gives a negative derivative, which means that we make the objective function larger by increasing $\alpha \ge \alpha_0$, which is where the optimal value must lie. Continuity at $r = 0$ follows similar to the previous proof. ∎

The combination of Lemma 2.3 and the above observation guarantees that the map $r \mapsto B(r) - r$ is a continuous and strictly increasing function from $[0, D(\mathcal{N}\|\overline{\mathcal{N}})]$ to $[-D(\mathcal{N}\|\overline{\mathcal{N}}), D(\overline{\mathcal{N}}\|\mathcal{N})]$. Hence, when real numbers a, b satisfy $-D(\mathcal{N}\|\overline{\mathcal{N}}) \le a - b \le D(\overline{\mathcal{N}}\|\mathcal{N})$, there exists $r_{a,b} \in [0, D(\mathcal{N}\|\overline{\mathcal{N}})]$ such that $B(r_{a,b}) - r_{a,b} = a - b$.

Lemma 2.4. *When real numbers a, b satisfy $-D(\mathcal{N}\|\overline{\mathcal{N}}) \le a - b \le D(\overline{\mathcal{N}}\|\mathcal{N})$, then we have*

$$C(a,b) = r_{a,b} - b = B(r_{a,b}) - a. \tag{2.20}$$

Proof. Definition of $C(a,b)$, Eq. 2.16, implies that $C(a-c, b-c) = C(a,b) + c$. Hence, it is sufficient to show that for $r \in [0, D(\mathcal{N}\|\overline{\mathcal{N}})]$:

$$C(B(r), r) = 0. \tag{2.21}$$

$$\begin{aligned} C(B(r), r) &= \sup_{0\le\alpha\le 1} (1-\alpha) D_\alpha(\mathcal{N}\|\overline{\mathcal{N}}) - \alpha B(r) - (1-\alpha) r \\ &= \sup_{0\le\alpha\le 1} \alpha\Big(\frac{\alpha - 1}{\alpha}\big(r - D_\alpha(\mathcal{N}\|\overline{\mathcal{N}})\big) - B(r)\Big) = 0, \end{aligned}$$

where the last equality follows since $\frac{\alpha-1}{\alpha}\big(r - D_\alpha(\mathcal{N}\|\overline{\mathcal{N}})\big) \le B(r)$ for $0 \le \alpha \le 1$. ∎

Our approach consists of associating suitable classical channels to the given cq-channels, and noting the lessons learned about adaptive strategy

for discrimination of classical channels in [35]. Our proof methodology however, is also novel for the classical case. The following Lemmas 2.5 and 2.6 addresses these matters; the former is verified easily and its proof is omitted, and the latter is more involved and is the key to our developments.

Lemma 2.5. *Consider the cq-channels $\mathcal{N} : x \to \rho_x$ and $\overline{\mathcal{N}} : x \to \sigma_x$ with input alphabet $\mathcal{X}$ and output density operators on Hilbert space B. Let the eigenvalue decompositions of the output operators be as follows:*

$$\rho_x = \sum_i \lambda_i^x \left|u_i^x\right\rangle\!\left\langle u_i^x\right|, \tag{2.22}$$

$$\sigma_x = \sum_j \mu_j^x \left|v_j^x\right\rangle\!\left\langle v_j^x\right|. \tag{2.23}$$

Define

$$\Gamma_x(i,j) \coloneqq \lambda_i^x \left|\left\langle v_j^x | u_i^x\right\rangle\right|^2, \tag{2.24}$$

$$\overline{\Gamma}_x(i,j) \coloneqq \mu_i^x \left|\left\langle v_j^x | u_i^x\right\rangle\right|^2. \tag{2.25}$$

First, $\Gamma_x(i,j)$ and $\overline{\Gamma}_x(i,j)$ form (conditional) probability distributions on the range $\{(i,j)\}$ of the pairs (i,j), i.e. for all pair of indexes (i,j), we have $\Gamma_x(i,j) \geq 0, \overline{\Gamma}_x(i,j) \geq 0$ and $\sum_{(i,j)} \Gamma_x(i,j) = \sum_{(i,j)} \overline{\Gamma}_x(i,j) = 1$. One can think of these distributions as classical channels. Second, we have

$$D_\alpha(\rho_x \| \sigma_x) = D_\alpha(\Gamma_x \| \overline{\Gamma}_x),$$

which implies [see Eq. (2.18)]

$$B(r) = \sup_x \sup_{0 \leq \alpha \leq 1} \frac{\alpha - 1}{\alpha} \big(r - D_\alpha(\Gamma_x \| \overline{\Gamma}_x)\big). \quad \blacksquare \tag{2.26}$$

Note that any extensions of the operators $\{\rho_x, \sigma_x\}$ (not just i.i.d.) correspond to the classical extensions by distributions $\Gamma_x(i,j)$ and $\overline{\Gamma}_x(i,j)$. Define

$$\Gamma^n_{\vec{x}_n}(\vec{i}_n, \vec{j}_n) \coloneqq \Gamma_{x_1}(i_1, j_1) \cdots \Gamma_{x_n}(i_n, j_n),$$
$$\overline{\Gamma}^n_{\vec{x}_n}(\vec{i}_n, \vec{j}_n) \coloneqq \overline{\Gamma}_{x_1}(i_1, j_1) \cdots \overline{\Gamma}_{x_n}(i_n, j_n).$$

Then, we have the following lemma.

Lemma 2.6.

$$E^Q_{a,b,n} \coloneqq \min_T 2^{an} \operatorname{Tr}(I - T)\rho^{(n)} + 2^{bn} \operatorname{Tr} T\sigma^{(n)} \geq \frac{1}{2} E^C_{a,b,n},$$

where

$$E^C_{a,b,n} \coloneqq \min_{q_{X_1},\dots,q_{X_n K_{n-1}|\vec{K}_{n-2}\vec{X}_{n-1}\vec{I}_{n-1}\vec{J}_{n-1}}} \sum_{\vec{x}_n,\vec{j}_n,\vec{i}_n,\vec{k}_{n-1}} q_{X_1}(x_1)\dots q_{X_n K_{n-1}|\vec{K}_{n-2}\vec{X}_{n-1}\vec{I}_{n-1}\vec{J}_{n-1}}(x_n,k_{n-1}|\vec{k}_{n-2},\vec{x}_{n-1},\vec{i}_{n-1},\vec{j}_{n-1}) \cdot \min\left\{2^{an}\Gamma^n_{\vec{x}_n}(\vec{i}_n,\vec{j}_n), 2^{bn}\overline{\Gamma}^n_{\vec{x}_n}(\vec{i}_n,\vec{j}_n)\right\}.$$

Proof. Let

$$\left|u^{\vec{x}_n}_{\vec{i}_n}\right\rangle \coloneqq \left|u^{x_1}_{i_1},\dots,u^{x_n}_{i_n}\right\rangle,\ \lambda^{\vec{x}_n}_{\vec{i}_n} \coloneqq \lambda^{x_1}_{i_1},\cdots\lambda^{x_n}_{i_n},$$
$$\left|v^{\vec{x}_n}_{\vec{j}_n}\right\rangle \coloneqq \left|v^{x_1}_{j_1},\dots,v^{x_n}_{j_n}\right\rangle,\ \mu^{\vec{x}_n}_{\vec{j}_n} \coloneqq \mu^{x_1}_{j_1},\cdots\mu^{x_n}_{j_n}.$$

Consider $\min_T 2^{an}\operatorname{Tr}(I-T)\rho^{(n)} + 2^{bn}\operatorname{Tr}T\sigma^{(n)}$; it is sufficient to consider T to a projective measurement because the minimum can be attained when T is a projection onto the subspace that is given as the linear span of eigenspaces corresponding to negative eigenvalues of $(-2^{an}\rho^{(n)} + 2^{bn}\sigma^{(n)})$. For a given $\vec{x}_n$, the final decision is given as the projection $T_{\vec{x}_n}$ on the image of the projection $\Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}}$ on $B^{\otimes n}$ depending on $\vec{x}_n$. Since $\rho^{(n)}$ and $\sigma^{(n)}$ both commute with the projection $\Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}}$, without loss of generality, we can assume that the projection $T_{\vec{x}_n}$ is also commutative with $\Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}}$. Then, the final decision operator T_n is given as the projection $T_n \coloneqq \sum_{\vec{x}_n} T_{\vec{x}_n}\otimes|\vec{x}_n\rangle\langle\vec{x}_n|$.

We expand the first term as follows:

$$\begin{aligned}
&\operatorname{Tr}(I-T_n)\rho^{(n)} \\
&= \sum_{\vec{x}_n,\vec{k}_{n-1}} \operatorname{Tr}(I-T_{\vec{x}_n})p_{X_1}(x_1)\cdots p_{X_n|\vec{X}_{n-1},\vec{K}_{n-1}}(x_n|\vec{x}_{n-1},\vec{k}_{n-1}) \\
&\qquad \Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}}\left(\rho_{x_1}\otimes\cdots\otimes\rho_{x_n}\right)\Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}} \\
&= \sum_{\vec{x}_n,\vec{k}_{n-1}} \operatorname{Tr}(I-T_{\vec{x}_n})^2 p_{X_1}(x_1)\cdots p_{X_n|\vec{X}_{n-1},\vec{K}_{n-1}}(x_n|\vec{x}_{n-1},\vec{k}_{n-1}) \\
&\qquad \Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}}\left(\rho_{x_1}\otimes\cdots\otimes\rho_{x_n}\right)\Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}} \\
&= \sum_{\vec{x}_n,\vec{k}_{n-1}} \operatorname{Tr}(I-T_{\vec{x}_n})\sum_{\vec{j}_n}\left|v^{\vec{x}_n}_{\vec{j}_n}\right\rangle\!\left\langle v^{\vec{x}_n}_{\vec{j}_n}\right|(I-T_{\vec{x}_n})p_{X_1}(x_1)\cdots p_{X_n|\vec{X}_{n-1},\vec{K}_{n-1}}(x_n|\vec{x}_{n-1},\vec{k}_{n-1}) \\
&\qquad \cdot\Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}}\left(\rho_{x_1}\otimes\cdots\otimes\rho_{x_n}\right)\Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}} \\
&= \sum_{\vec{x}_n,\vec{j}_n,\vec{i}_n,\vec{k}_{n-1}} p_{X_1}(x_1)\cdots p_{X_n|\vec{X}_{n-1},\vec{K}_{n-1}}(x_n|\vec{x}_{n-1},\vec{k}_{n-1}) \\
&\qquad \lambda^{\vec{x}_n}_{\vec{i}_n}\left|\left\langle u^{\vec{x}_n}_{\vec{i}_n}\right|(I-T_{\vec{x}_n})\Pi^{(n-1)}_{\vec{k}_{n-1}|\vec{x}_{n-1}}\left|v^{\vec{x}_n}_{\vec{j}_n}\right\rangle\right|^2,
\end{aligned}$$

where the first line follows from the definition of T, the second line is due to the fact that the final measurement can be chosen as a projective measurement, the third line follows because $\sum_{\vec{j}_n} \left|v_{\vec{j}_n}^{\vec{x}_n}\right\rangle\!\left\langle v_{\vec{j}_n}^{\vec{x}_n}\right| = I^{\otimes n}$ and the last line is simple manipulation.

Similarly, we have

$$\operatorname{Tr} T\sigma^{(n)} =$$

$$\sum_{\vec{x}_n,\vec{j}_n,\vec{i}_n,\vec{k}_{n-1}} p_{X_1}(x_1)\cdots p_{X_n|\vec{X}_{n-1},\vec{K}_{n-1}}(x_n|\vec{x}_{n-1},\vec{k}_{n-1})\mu_{\vec{j}_n}^{\vec{x}_n}\left|\left\langle u_{\vec{i}_n}^{\vec{x}_n}\right|T_{\vec{x}_n}\Pi_{\vec{k}_{n-1}|\vec{x}_{n-1}}^{(n-1)}\left|v_{\vec{j}_n}^{\vec{x}_n}\right\rangle\right|^2.$$

For $m \in [1:n]$, define

$$q_{X_mK_{m-1}|\vec{K}_{m-2}\vec{X}_{m-1}\vec{I}_{m-1}\vec{J}_{m-1}}(x_m,k_{m-1}|\vec{x}_{m-1},\vec{k}_{m-2},\vec{i}_{m-1},\vec{j}_{m-1})$$

$$\coloneqq p_{X_m|\vec{X}_{m-1},\vec{K}_{m-1}}(x_m|\vec{x}_{m-1},\vec{k}_{m-1})\frac{\left|\left\langle u_{\vec{i}_{m-1}}^{\vec{x}_{m-1}}\right|\Pi_{\vec{k}_{m-1}|\vec{x}_{m-1}}^{(m-1)}\left|v_{\vec{j}_{m-1}}^{\vec{x}_{m-1}}\right\rangle\right|^2}{\left|\left\langle u_{\vec{i}_{m-2}}^{\vec{x}_{m-2}}\right|\Pi_{\vec{k}_{m-2}|\vec{x}_{m-2}}^{(m-2)}\left|v_{\vec{j}_{m-2}}^{\vec{x}_{m-2}}\right\rangle\right|^2\cdot\left|\left\langle u_{i_{m-1}}^{x_{m-1}}|v_{j_{m-1}}^{x_{m-1}}\right\rangle\right|^2}.$$

Hence,

$$\begin{aligned}
&\min_T e^{bn}2^{an}\operatorname{Tr}(I-T)\rho^{(n)} + 2^{an}\operatorname{Tr}T\sigma^{(n)}\\
&= \sum_{\vec{x}_n,\vec{j}_n,\vec{i}_n,\vec{k}_{n-1}} p_{X_1}(x_1)\cdots p_{X_n|\vec{X}_{n-1},\vec{K}_{n-1}}(x_n|\vec{x}_{n-1},\vec{k}_{n-1})\\
&\cdot\left(2^{an}\lambda_{\vec{j}_n}^{\vec{x}_n}\left|\left\langle u_{\vec{i}_n}^{\vec{x}_n}\right|(I-T_{\vec{x}_n})\Pi_{\vec{k}_{n-1}|\vec{x}_{n-1}}^{(n-1)}\left|v_{\vec{j}_n}^{\vec{x}_n}\right\rangle\right|^2 + 2^{bn}\mu_{\vec{i}_n}^{\vec{x}_n}\left|\left\langle u_{\vec{i}_n}^{\vec{x}_n}\right|T_{\vec{x}_n}\Pi_{\vec{k}_{n-1}|\vec{x}_{n-1}}^{(n-1)}\left|v_{\vec{j}_n}^{\vec{x}_n}\right\rangle\right|^2\right)\\
&\geq \sum_{\vec{x}_n,\vec{j}_n,\vec{i}_n,\vec{k}_{n-1}} p_{X_1}(x_1)\cdots p_{X_n|\vec{X}_{n-1},\vec{K}_{n-1}}(x_n|\vec{x}_{n-1},\vec{k}_{n-1})\min\left\{2^{an}\lambda_{\vec{i}_n}^{\vec{x}_n},2^{bn}\mu_{\vec{j}_n}^{\vec{x}_n}\right\}\\
&\cdot\left(\left|\left\langle u_{\vec{i}_n}^{\vec{x}_n}\right|T_{\vec{x}_n}\Pi_{\vec{k}_{n-1}|\vec{x}_{n-1}}^{(n-1)}\left|v_{\vec{j}_n}^{\vec{x}_n}\right\rangle\right|^2 + \left|\left\langle u_{\vec{i}_n}^{\vec{x}_n}\right|(I-T_{\vec{x}_n})\Pi_{\vec{k}_{n-1}|\vec{x}_{n-1}}^{(n-1)}\left|v_{\vec{j}_n}^{\vec{x}_n}\right\rangle\right|^2\right)\\
&\overset{(a)}{\geq} \sum_{\vec{x}_n,\vec{j}_n,\vec{i}_n,\vec{k}_{n-1}} p_{X_1}(x_1)\cdots p_{X_n|\vec{X}_{n-1},\vec{K}_{n-1}}(x_n|\vec{x}_{n-1},\vec{k}_{n-1})\min\left\{e^{an}\lambda_{\vec{i}_n}^{\vec{x}_n},2^{bn}\mu_{\vec{j}_n}^{\vec{x}_n}\right\}\\
&\cdot\frac{1}{2}\left|\left\langle u_{\vec{i}_n}^{\vec{x}_n}\right|\Pi_{\vec{k}_{n-1}|\vec{x}_n}^{(n-1)}\left|v_{\vec{j}_n}^{\vec{x}_n}\right\rangle\right|^2\\
&= \frac{1}{2}\sum_{\vec{x}_n,\vec{j}_n,\vec{i}_n,\vec{k}_{n-1}} q_{X_1}(x_1)\cdots q_{X_nK_{n-1}|\vec{K}_{n-2}\vec{X}_{n-1}\vec{I}_{n-1}\vec{J}_{n-1}}(x_nk_{n-1}|\vec{k}_{n-2}\vec{x}_{n-1}\vec{i}_{n-1}\vec{j}_{n-1})\\
&\cdot\min\left\{2^{an}\Gamma_{\vec{x}_n}(\vec{i}_n,\vec{j}_n),2^{bn}\overline{\Gamma}_{\vec{x}_n}(\vec{i}_n,\vec{j}_n)\right\},
\end{aligned}$$

where (a) follows from the relation $|\alpha|^2+|\beta|^2 \geq \frac{1}{2}|\alpha+\beta|^2$. ∎

2.5.3 Main results

We are now in a position to present and prove our main result, the generalized Chernoff bound, as follows:

Theorem 2.1 (Generalized Chernoff bound)**.** *For real numbers a, b satisfying $-D(\mathcal{N}\|\overline{\mathcal{N}}) \le a - b \le D(\overline{\mathcal{N}}\|\mathcal{N})$,*

$$C^{\mathbb{A}}(a,b|\mathcal{N}\|\overline{\mathcal{N}}) = C^{\mathbb{P}^0}(a,b|\mathcal{N}\|\overline{\mathcal{N}}) = C(a,b) = r_{a,b} - b = B(r_{a,b}) - a.$$

Proof. For the direct part, i.e. that strategies in $\mathbb{P}^0$ achieve this exponent, the following non-adaptive strategy achieves $C(a,b)$. Consider the transmission of a letter x on every channel use. Define the test T_n as the projection to the eigenspace of the positive eigenvalues of $2^{na}\rho_x^{\otimes n} - 2^{nb}\sigma_x^{\otimes n}$. Audenaert *et al.* [64] showed that

$$\begin{aligned} 2^{na}\operatorname{Tr}[\rho_x^{\otimes n}(I - T_n)] + 2^{nb}\operatorname{Tr}[\sigma_x^{\otimes n}T_n] &\le \inf_{0\le\alpha\le 1} \operatorname{Tr}(2^{na}\rho_x^{\otimes n})^{\alpha}(2^{nb}\sigma_x^{\otimes n})^{1-\alpha} \\ &= 2^{-n\sup_{0\le\alpha\le 1}\left((1-\alpha)D_\alpha(\rho_x\|\sigma_x)-\alpha a-(1-\alpha)b\right)}. \end{aligned} \tag{2.27}$$

Considering the optimization for x, we obtain the direct part.

For the converse part, since

$$\begin{aligned} C^{\mathbb{A}}(a,b|\mathcal{N}\|\overline{\mathcal{N}}) &= C^{\mathbb{A}}(B(r_{a,b}), r_{a,b}|\mathcal{N}\|\overline{\mathcal{N}}) + B(r_{a,b}) - a \\ &= C^{\mathbb{A}}(B(r_{a,b}), r_{a,b}|\mathcal{N}\|\overline{\mathcal{N}}) + r_{a,b} - b, \end{aligned}$$

it is sufficient to show $C^{\mathbb{A}}(B(r), r|\mathcal{N}\|\overline{\mathcal{N}}) \ge 0$ for $r \in [0, D(\mathcal{N}\|\overline{\mathcal{N}})]$. Observe that

$$E^C_{a,b,n} = 2^{an}\alpha_n(\Gamma\|\overline{\Gamma}|\mathcal{T}_{a,b,n}) + 2^{bn}\beta_n(\Gamma\|\overline{\Gamma}|\mathcal{T}_{a,b,n}),$$

where we let $\mathcal{T}_{a,b,n}$ be the optimal test to achieve $E^C_{a,b,n}$. We choose $a = B(r)$ and $b = r$ in Lemma 2.6. The combination of 2.26 and [35, Eq. (16)] guarantees that

$$B^{\mathbb{A}}_e(r|\Gamma\|\overline{\Gamma}) = B(r). \tag{2.28}$$

Notice that the analysis in [35] does not assume any condition on the set $\mathcal{X}$. When

$$\liminf_{n\to\infty} \frac{1}{n}\log 2^{rn}\beta_n(\Gamma\|\overline{\Gamma}|\mathcal{T}_{B(r),r,n}) < 0,$$

then Eq. 2.28 implies

$$\liminf_{n\to\infty} \frac{1}{n}\log 2^{B(r)n}\alpha_n(\Gamma\|\overline{\Gamma}|\mathcal{T}_{B(r),r,n}) \ge 0.$$

Hence, we have

$$\liminf_{n\to\infty}\frac{1}{n}\log E^{C}_{B(r),r,n} = \max\left\{\liminf_{n\to\infty}\frac{1}{n}\log 2^{rn}\beta_n(\Gamma\|\overline{\Gamma}|\mathcal{T}_{B(r),r,n})\ ,\right.$$
$$\left.\liminf_{n\to\infty}\frac{1}{n}\log 2^{B(r)n}\alpha_n(\Gamma\|\overline{\Gamma}|\mathcal{T}_{B(r),r,n})\right\} \geq 0. \tag{2.29}$$

Therefore, the combination of Lemma 2.6 and 2.29 implies that

$$C^{\mathbb{A}}(B(r), r|\mathcal{N}\|\overline{\mathcal{N}}) = \liminf_{n\to\infty}\frac{1}{n}\log E^{Q}_{B(r),r,n} \geq 0. \tag{2.30}$$

This completes the proof. ■

As corollary, we obtain the Hoeffding exponent.

Corollary 2.1 (Hoeffding bound)**.** *For any* $0 \leq r \leq D(\mathcal{N}\|\overline{\mathcal{N}})$*,*

$$B_e^{\mathbb{A}}(r|\mathcal{N}\|\overline{\mathcal{N}}) = B_e^{\mathbb{P}^0}(r|\mathcal{N}\|\overline{\mathcal{N}}) = B(r).$$

Proof. For the direct part, note that a non-adaptive strategy following the Hoeffding bound for state discrimination developed in [67] suffices to show the achievability. More precisely, sending the letter x optimizing the expression on the right-hand side to every channel use and invoking the result by [67] for state discrimination shows the direct part of the theorem.

For the converse part, note first that from Theorem 2.1, for any $r \in [0, D(\mathcal{N}\|\overline{\mathcal{N}})]$

$$C^{\mathbb{A}}(B(r), r|\mathcal{N}\|\overline{\mathcal{N}}) = 0.$$

When a sequence of tests T_n satisfies $\liminf_{n\to\infty}\frac{1}{n}\log\beta_n[\mathcal{N}\|\overline{\mathcal{N}}|T_n] \leq -r_0 < -r_0+\epsilon$, Eq. 2.30 with $r = r_0 - \epsilon$ implies that $\liminf_{n\to\infty}\frac{1}{n}\log\alpha_n[\mathcal{N}\|\overline{\mathcal{N}}|T_n] \geq -B(r_0-\epsilon)$. Hence, we have

$$B_e^{\mathbb{A}}(r_0|\mathcal{N}\|\overline{\mathcal{N}}) \leq B(r_0-\epsilon). \tag{2.31}$$

Due to Lemma 2.3, taking the limit $\epsilon \to 0$ leads to the following inequality

$$B_e^{\mathbb{A}}(r_0|\mathcal{N}\|\overline{\mathcal{N}}) \leq B(r_0). \tag{2.32}$$

This completes the proof. ■

2.6 Discrimination of quantum channels with classical feed-forward

We showed in Section 2.3 that quantum feed-forward generally can improve the error exponent in the symmetric setting. This result followed by investigating a pair of entanglement-breaking channels introduced in [36]; this example, however, did not clue in on the profitability of input quantum memory, i.e. whether entangled state inputs have any beneficial effect when the given pair consists of entanglement-breaking channels. To shed more light on this problem, in this section we back off from quantum feed-forward information and consider discrimination of qq-channels under the class of adaptive strategies that only allow for classical feed-forward information in the presence and absence of input entanglement, i.e. $\mathbb{A}_n^c$ and $\mathbb{A}_n^{c,0}$, respectively. We will show that without use of entanglement in the input, the adaptive strategies offers no gain over non-adaptive strategies. However, when entangled state inputs are allowed, unlike cq-channels, we show that the optimality of non-adaptive strategies for entanglement-breaking channels cannot be inferred from that for cq-channels. In the following we use the notation introduced in Section 2.5: a pair of qq-quantum channels $\mathcal{M}$ and $\overline{\mathcal{M}}$ with common input A and output B systems; we shall identify $\mathcal{M}$ and $\overline{\mathcal{M}}$ as the null and alternative hypotheses, respectively.

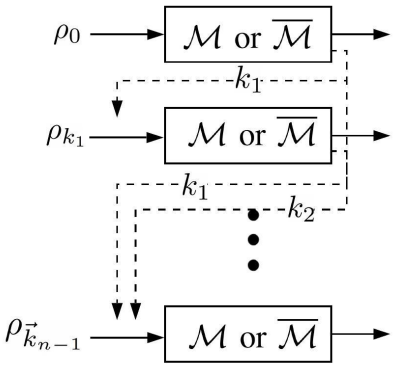

Figure 2.2: Adaptive quantum channel discrimination with classical feed-forward. Solid and dashed arrow denote the flow of quantum and classical information, respectively. At step m, Alice sends the state ρ_{x_m} which she has prepared using Bob's $(m-1)$ classical feed-forward, and sends it via either $\mathcal{M}$ or $\overline{\mathcal{M}}$ to Bob.

2.6.1 Quantum memory is not allowed at the input: $\mathbb{A}_n^{c,0}$

In this setting, the protocol is similar to the adaptive protocol described in Section 2.5, see Fig. (2.2): after each transmission, the input state ρ is chosen adaptively from the classical feed-forward. Denoting this adaptive choice of input states as $\vec{x}_m = (x_1, \ldots, x_n)$, the m-th input is chosen conditioned on the feed-forward information $\vec{k}_{m-1}$ and $\vec{x}_{m-1}$ from the conditional distribution $p_{X_m|\vec{X}_{m-1},\vec{K}_{m-1}}(x_m|\vec{x}_{m-1},\vec{k}_{m-1})$.

For two quantum channels $\mathcal{M}$ and $\overline{\mathcal{M}}$, define

$$D_\alpha(\mathcal{M}\|\overline{\mathcal{M}}) \coloneqq \sup_\rho D_\alpha(\mathcal{M}(\rho)\|\overline{\mathcal{M}}(\rho)), \text{ and}$$
$$D(\overline{\mathcal{M}}\|\mathcal{M}) \coloneqq \sup_\rho D_\alpha(\overline{\mathcal{M}}(\rho)\|\mathcal{M}(\rho)).$$

Theorem 2.2. *Let $\mathcal{M}$ and $\overline{\mathcal{M}}$ be qq-channels. Then, for real numbers a, b satisfying $-D(\mathcal{M}\|\overline{\mathcal{M}}) \le a - b \le D(\overline{\mathcal{M}}\|\mathcal{M})$ and any $0 \le r \le D(\mathcal{M}\|\overline{\mathcal{M}})$,*

$$\begin{aligned} C^{\mathbb{A}^{c,0}}(a,b|\mathcal{M}\|\overline{\mathcal{M}}) &= C^{\mathbb{P}^0}(a,b|\mathcal{M}\|\overline{\mathcal{M}}) \\ &= \sup_\rho \sup_{0\le\alpha\le1} (1-\alpha)D_\alpha(\mathcal{M}(\rho)\|\overline{\mathcal{M}}(\rho)) - \alpha a - (1-\alpha)b \\ &= \sup_{0\le\alpha\le1} (1-\alpha)D_\alpha(\mathcal{M}\|\overline{\mathcal{M}}) - \alpha a - (1-\alpha)b, \end{aligned}$$

$$\begin{aligned} B_e^{\mathbb{A}^{c,0}}(r|\mathcal{M}\|\overline{\mathcal{M}}) &= B_e^{\mathbb{P}^0}(r|\mathcal{M}\|\overline{\mathcal{M}}) \\ &= \sup_\rho \sup_{0\le\alpha\le1} \frac{\alpha-1}{\alpha}\big(r - D_\alpha(\mathcal{M}(\rho)\|\overline{\mathcal{M}}(\rho))\big) \\ &= \sup_{0\le\alpha\le1} \frac{\alpha-1}{\alpha}\big(r - D_\alpha(\mathcal{M}\|\overline{\mathcal{M}})\big) \end{aligned}$$

Proof. Since only classical feed-forward is allowed, it is seen that this discrimination problem can be cast in the framework of cq-channel discrimination problem. To show Theorem 2.2, we apply Theorem 2.1 and Corollary 2.1 to the case when the cq-channel has input alphabet $\mathcal{X} = \mathcal{S}^A$, i.e. it equals the set of states on the input systems. In other words, we choose the classical (continuous) input alphabet as $\mathcal{X}$, where each letter $x \in \mathcal{X}$ is a classical description of a state ρ on the input system A. In this application, ρ_x and σ_x are given as $\mathcal{M}(\rho)$ and $\overline{\mathcal{M}}(\rho)$, respectively, for $x = \rho$. Hence, $\sup_x D_\alpha(\rho_x\|\sigma_x)$ equals $D_\alpha(\mathcal{M}\|\overline{\mathcal{M}}) = \sup_\rho D_\alpha(\mathcal{M}(\rho)\|\overline{\mathcal{M}}(\rho))$. Hence, the desired relation is obtained. ∎

Remark 2.2. *The above theorem concludes that in the absence of entangled inputs, no adaptive strategy built upon classical feed-forward can outperform the best non-adaptive or fix strategy. In other words, the optimal error rate can be achieved by the simple i.i.d. sequence where all n input states are chosen to be the same, i.e.* $\rho \otimes \cdots \otimes \rho$.

2.6.2 Quantum memory is allowed at the input: $\mathbb{A}_n^c$

The most general class $\mathbb{A}_n^c$ of strategies to distinguish two qq-channels $\mathcal{M}$ and $\overline{\mathcal{M}}$ is the set of strategies given in Definition 1. For this class, we denote the generalized Chernoff and Hoeffding quantities as $C^{\mathbb{A}_n^c}(a,b|\mathcal{M}\|\overline{\mathcal{M}})$ and $B_e^{\mathbb{A}_n^c}(r|\mathcal{M}\|\overline{\mathcal{M}})$, respectively. In this subsection, we discuss the effect of input entanglement for our cq-channel discrimination strategy. To this aim, we consider the special case when the two qq-quantum channels $\mathcal{M}$ and $\overline{\mathcal{M}}$ are entanglement-breaking and thus have the following form:

$$\mathcal{M}(\xi) = \sum_x \rho_x \operatorname{Tr} \xi E_x, \quad \overline{\mathcal{M}}(\xi) = \sum_x \sigma_x \operatorname{Tr} \xi E_x, \tag{2.33}$$

where $\{E_x\}_{x\in\mathcal{X}}$ is a PVM and the rank of E_x is one.

In this case, the most general strategy stated in Definition 1 for the discrimination of two qq-channels $\mathcal{M}$ and $\overline{\mathcal{M}}$ can be converted to the strategy stated in Subsection 2.5.1 for the discrimination of two cq-channels $\mathcal{N} : x \mapsto \rho_x$ and $\overline{\mathcal{N}} : x \mapsto \sigma_x$ as follows. In the former strategy, the m-step operation is given as a quantum channel $\mathcal{F}_m : R_m B_m \to R_{m+1} A_{m+1}$. As the latter strategy, we define the quantum instrument $\mathcal{E}_m : R_m B_m \to X_m R_{m+1}$ as

$$\mathcal{E}_m(\xi) := \sum_{x_m} \operatorname{Tr}_{A_{m+1}} E_{x_m} \mathcal{F}_m(\xi) \otimes |x_m\rangle\langle x_m|. \tag{2.34}$$

Then, we choose the obtained outcome x_m as the input of the cq-channel to be discriminated. The final states in the former strategy is the same as the final state in the latter strategy. That is, the performance of the general strategy for these two qq-channels is the same as the performance of the general strategy for the above defined cq-channels. This fact means that the adaptive method does not improve the performance of the discrimination of the channels 2.33.

Theorem 2.3. *Assume that two qq-quantum channels* $\mathcal{M}$ *and* $\overline{\mathcal{M}}$ *are given by Eq. 2.33. For* $0 \le r \le D(\mathcal{M})$ *[see Eq. 2.35] and real* a *and* b *with* $-D(\mathcal{M}\|\overline{\mathcal{M}}) \le a - b \le D(\mathcal{M}\|\overline{\mathcal{M}})$*, then the following holds:*

$$C^{\mathbb{A}^c}(a,b|\mathcal{M}\|\overline{\mathcal{M}}) = C^{\mathbb{A}^{c,0}}(a,b|\mathcal{M}\|\overline{\mathcal{M}}) = C(a,b|\mathcal{N}\|\overline{\mathcal{N}})$$
$$B_e^{\mathbb{A}^c}(r|\mathcal{M}\|\overline{\mathcal{M}}) = B_e^{\mathbb{A}^{c,0}}(r|\mathcal{M}\|\overline{\mathcal{M}}) = B_e(r|\mathcal{N}\|\overline{\mathcal{N}}).$$

■

Furthermore, when the quantum channel $\mathcal{F}_m$ in the strategy is replaced by the channel $\mathcal{F}'_m$ defined as $\mathcal{F}'_m(\xi) := \sum_{x_m} E_{x_m} \mathcal{F}_m(\xi) E_{x_m}$, we do not change the statistics of the protocol for either channel. Since the output of $\mathcal{F}'_m$ has no entanglement between X_m and R_{m+1}, the presence of input entanglement does not improve the performance in this case.

Note that it was essential not only that the channels are entanglement-breaking, but that the measurement $\{E_x\}$ is a PVM, and in fact the same PVM for both channels. The discussion fails already when the channels have each their own PVM, which are non-commuting (such channels were essential to the counterexample in Section 2.3). In this case, the construction of the channel $\mathcal{F}'_m$ depends on the choice of the hypothesis. Therefore, the condition 2.33 is essential for this discussion.

Furthermore, if the channels are entanglement-breaking, but with only a POVM in Eq. 2.33, i.e. the E_x are not orthogonal projectors, the above discussion does not hold, either. In this case, the output state is separable, but it cannot be necessarily simulated by a separable input state.

Remark 2.3. *It is worth mentioning that the adaptive strategy in Section 2.3 that is applied to a pair of entanglement-breaking channels and shown to be better than non-adaptive strategies, was actually using quantum feed-forward, however no entangled inputs nor indeed quantum memory at the channel input. In the extension of our cq-channel formalism to other settings one has to consider that by the nature of cq-channels, neither entangled inputs nor quantum feed-forward can help adaptive strategies. In case of entanglement-breaking channels, the optimality of adaptive strategies in the absence of quantum feed-forward remains a question. The examples of Harrow* et al. *remain inspiring in that respect, as the two channels are each individually classical-quantum, but unlike the form in Eq. 2.33, they have different orthogonal measurements*

Remark 2.4. *The discussion of this section shows that without loss of generality, we can assume that the measurement outcome equals the next input when $\mathcal{X}$ is discrete. That is, it is sufficient to consider the case when $k_m = x_m$. This fact can be shown as follows. Given two cq-channels $x \mapsto \rho_x$ and $x \mapsto \sigma_x$, we define two entanglement breaking channels $\mathcal{M}$ and $\overline{\mathcal{M}}$ by 2.33. For the case with two qq-channel, the most general strategy is given in Definition 1. The most general strategy can be simulated by the strategy with $k_m = x_m$ for two cq-channels $x \mapsto \rho_x$ and $x \mapsto \sigma_x$.*

However, when $\mathcal{X}$ is not discrete, the above discussion does not hold. Hence, to cover the case with continuous $\mathcal{X}$, we need to address the case with general outcome k_m as in Section 2.5.

2.7 Discrimination power of a quantum channel

In this section we study how well a pair of quantum states can be distinguished after passing through a quantum channel. This quantifies the power of a quantum channel when it is seen as a measurement device. In some sense this scenario is dual to the state discrimination problem in which a pair of states are given and the optimization is taken over all measurements, whilst in the current scenario a quantum channel is given and the optimization takes place over all pairs of states passing through the channel. The reference [76] studies the special case of quantum-classical (qc) channels, that is investigation of the power of a quantum detector given by a specific POVM in discriminating two quantum states. It was shown in the paper that when the qc-channel is available asymptotically many times, neither entangled state inputs nor classical feed-forward and adaptive choice of inputs can improve the performance of the channel. We extend the model of the latter paper to general quantum channels, see Fig. 2.3. It is useful to cast this hypothesis testing setting as a communication problem as follows: Assume a quantum channel $\mathcal{M}_o = \mathcal{M}_o^{A_o \to B}$ connects Alice and Bob, where Alice posses A_o and Bob has B. Alice wants to send one bit of information $Z \in \{0, 1\}$ to Bob by using this channel. The strategy allows Alice to use the channel n times and also allows Bob, who has access to quantum memory, to perform any measurement of his desire on its received systems and send back classical information to Alice; then Alice's encoder can choose a suitable state in $A_o^{\otimes n}$ adaptively based on the feedback it receives after each transmission. This problem resembles quantum channel discrimination problem, we slight abuse of notation, we use the same notation introduced for channel discrimination; this way, the above corresponds to the classes $\mathbb{A}_n^c$ and $\mathbb{A}_n^{c,0}$, based on whether entangled inputs are or are not allowed, respectively. We investigate if these classes offer any advantage over the class $\mathbb{P}_n^0$ in any of symmetric and asymmetric settings. We freely use the notation introduced in Section 2.5. In particular, for a given class $\mathbb{S}_n \subset \mathbb{A}_n$ of adaptive strategies for n channel uses, the generalized Chernoff and Hoeffding quantities are denoted by $C^{\mathbb{S}}(a, b|\mathcal{M})$ and $B_e^{\mathbb{S}}(r|\mathcal{M})$, respectively.

2.7.1 Alice has no quantum memory: class $\mathbb{A}_n^{c,0}$

To quantify the power of the quantum channel $\mathcal{M}_o$, we apply channel discrimination to the case when $\mathcal{X} = \mathcal{S}^{A_o} \times \mathcal{S}^{A'_o}$, i.e. $\mathcal{X} = \{(\rho, \sigma)\}_{\rho,\sigma}$, the cq-channel $\mathcal{N}$ of the null hypothesis maps (ρ, σ) to $\mathcal{M}_o(\rho)$, and the cq-channel $\overline{\mathcal{N}}$ of

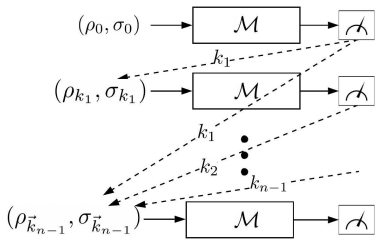

Figure 2.3: Discrimination with a cptp map $\mathcal{M}_o$. At step m, Alice prepares a state, either ρ_{x_m} or σ_{x_m}, which she has prepared using Bob's $m-1$ feedbacks (dashed arrows), and sends it via the channel $\mathcal{M}$ to Bob. Bob's measurements resembles the PVM's of Section 2.5.

the alternative hypothesis maps (ρ, σ) to $\mathcal{M}_o(\sigma)$. That is, we distinguish which of A_o and A'_o is the true input system. Here, the event $Z = 0$ ($Z = 1$) corresponds to the cq-channel $\mathcal{N}$ ($\overline{\mathcal{N}}$). With full generality, Bob's final decision is the measurement device defined by a two-outcome POVM $\{T_n, I - T_n\}$ on n-tensor product system, where T_n and $I - T_n$ correspond to the event supporting the null and alternative hypotheses, respectively.

While within class $\mathbb{P}_n^0$ Alice's choice of the pair of input states is just tensor-product states $(\rho^{\otimes n}, \sigma^{\otimes n})$, in class $\mathbb{A}_n^{c,0}$, the adaptive strategy follows the cq-channel discrimination strategy: denoting the input generically as $x_1, \ldots, x_n$, the sequence of Bob's measurements is given as $\{P^{(m)}_{\vec{k}_m|\vec{x}_m}\}_{m=1}^{n-1}$ and the classical feedback depends on the previous information $\vec{x}_m, \vec{k}_{m-1}$, and Alice's adaptive choice of the input states $(x_1, \ldots, x_n)$ is given as the sequence of conditional randomized choice $\{p_{X_m|\vec{X}_{m-1},\vec{K}_{m-1}}\}_{m=1}^{n}$ of the pair of the input states (ρ, σ). In this formulation, as the feedback, Bob sends information x_m to Alice. That is, Bob decides which pair is used for Alice's inputs according to the conditional distributions $\{p_{X_m|\vec{X}_{m-1},\vec{K}_{m-1}}\}_{m=1}^{n}$. The reason why the feedback information is x_m and not k_{m-1} is that Bob's choice of the measurement depends on x_m and classical information transmission from Alice to Bob is not allowed in the above setting, the feedback information from Bob to Alice needs to contain x_m at least. Note that the communication problem developed here should not put more restrictions on the protocol inferred from cq-channel discrimination. If there was an additional classical channel from Alice to Bob that Alice can generate x_m from k_{m-1} and send it to Bob, then we did not need to assume that Bob generates x_m. We set

$$D(\mathcal{M}_o) := \max_{\rho,\sigma} D(\mathcal{M}_o(\rho)\|\mathcal{M}_o(\sigma)) = \max_{\rho,\sigma} D(\mathcal{M}_o(\sigma)\|\mathcal{M}_o(\rho)). \tag{2.35}$$

Theorem 2.4. *Let $0 \le r \le D(\mathcal{M}_o)$ and real numbers a and b satisfy $-D(\mathcal{M}_o) \le a - b \le D(\mathcal{M}_o)$, then we have*

$$\begin{aligned} C^{\mathbb{A}^{c,0}}(a,b|\mathcal{M}_o) &= C^{\mathbb{P}^0}(a,b|\mathcal{M}_o) \\ &= \sup_{\rho,\sigma} \sup_{0\le\alpha\le 1} (1-\alpha) D_\alpha(\mathcal{M}(\rho)\|\mathcal{M}_o(\sigma)) - \alpha a - (1-\alpha) b, \end{aligned}$$

$$B_e^{\mathbb{A}_n^{c,0}}(r|\mathcal{M}_o) = B_e^{\mathbb{P}^0}(r|\mathcal{M}_o) = \sup_{\rho,\sigma} \sup_{0\le\alpha\le 1} \frac{\alpha-1}{\alpha}\big(r - D_\alpha(\mathcal{M}_o(\rho)\|\mathcal{M}_o(\sigma))\big).$$

Proof. Here we only need to consider the set $\mathcal{S}^{A_o} \times \mathcal{S}^{A'_o}$ of pairs of input states as the set $\mathcal{X}$. In other words, we choose the classical (continuous) input alphabet as $\mathcal{X} = \mathcal{S}^{A_o} \times \mathcal{S}^{A'_o}$, where each letter $x = (\rho, \sigma) \in \mathcal{X}$ is a classical description of the pair of states (ρ, σ). Then the result follows from the adaptive protocol in Section 2.5. See also the proof of Theorem 2.2. ∎

2.7.2 Alice has quantum memory: class $\mathbb{A}_n^c$

We denote the most general class of strategies by $\mathbb{A}_n^c$. This class is given as as the strategy given in Definition 1 for two qq-channels $\mathcal{M}$ and $\overline{\mathcal{M}}$ from $A = A_o A'_o$ to B when they are defined as

$$\mathcal{M}(\rho) := \mathrm{Tr}_2(\mathcal{M}_o \otimes \mathcal{M}_o)(\rho) \tag{2.36}$$

$$\overline{\mathcal{M}}(\rho) := \mathrm{Tr}_1(\mathcal{M}_o \otimes \mathcal{M}_o)(\rho) \tag{2.37}$$

for $\rho \in \mathcal{S}^{A_o A'_o}$. In this class, we denote the generalized Chernoff and Hoeffding quantities as $C^{\mathbb{A}_n^c}(a,b|\mathcal{M}_o)$ and $B_e^{\mathbb{A}_n^c}(r|\mathcal{M}_o)$, respectively.

As a corollary of Theorem 2.4, we obtain the following.

Corollary 2.1. *Assume that the qq-channel $\mathcal{M}_o$ has the form*

$$\mathcal{M}_o(\rho) = \sum_x \rho_x \,\mathrm{Tr}\, E_x \rho \tag{2.38}$$

where $\{E_x\}_{x\in\mathcal{X}}$ is a PVM and the rank of E_x is one. For $0 \le r \le D(\mathcal{M}_o)$ (see Eq. 2.35) and real a and b with $-D(\mathcal{M}_o) \le a - b \le D(\mathcal{M}_o)$, when $\mathcal{M}_o$ is entanglement-breaking, the following holds

$$\begin{aligned} C^{\mathbb{A}^c}(a,b|\mathcal{M}_o) &= C^{\mathbb{A}^{c,0}}(a,b|\mathcal{M}_o) = C^{\mathbb{P}^0}(a,b|\mathcal{M}_o), \\ B_e^{\mathbb{A}^c}(r|\mathcal{M}_o) &= B_e^{\mathbb{A}^{c,0}}(r|\mathcal{M}_o) = B_e^{\mathbb{P}^0}(r|\mathcal{M}_o). \end{aligned}$$

Proof. When the condition 2.38, two qq-channels $\mathcal{M}$ and $\overline{\mathcal{M}}$ satisfy the condition 2.33. Hence, Theorem 2.4 implies this corollary. ∎

Remark 2.5. *The above result states that the optimal error rates for discrimination with a quantum channel can be achieved by i.i.d. state pairs* $(\rho^{\otimes n}, \sigma^{\otimes n})$ *only when no quantum memory is allowed on the sender side. Also, when entangled state inputs are allowed, we could only show the optimality of non-adaptive tensor-product strategy* $\mathbb{P}_n^0$ *for entanglement-breaking channel with the form 2.38. The same conclusion holds for the Chernoff bound, Stein's lemma and Han-Kobayashi bound.*

2.7.3 Examples

In this subsection we derive the generalized Chernoff and Hoeffding bounds for three qubit channels, namely, we study the discrimination power of depolarizing, Pauli and amplitude damping channels. In each case, the key is identifying the structure of the output states of each channel by employing the lessons learned in [82]. Here we briefly summarize the basics. A quantum state ρ in two-level systems can be parametrized as $\rho = \frac{1}{2}(I + \vec{r}\cdot\vec{\sigma})$, where $\vec{r} = (r_x, r_y, r_z) \in \mathbb{R}^3$ is the Bloch vector which satisfies $r_x^2 + r_y^2 + r_z^2 \leq 1$ and $\vec{\sigma}$ denotes the vector of Pauli matrices $\{\sigma_x, \sigma_y, \sigma_z\}$ such that $\vec{r}\cdot\vec{\sigma} \coloneqq r_x\sigma_x + r_y\sigma_y + r_z\sigma_z$. Any cptp map $\mathcal{M}_o$ on qubits can be represented as follows:

$$\mathcal{M}_o(\rho) = \frac{1}{2}\left(I + (\vec{t} + T\vec{r})\cdot\vec{\sigma}\right) \quad \text{for} \quad \rho = \frac{1}{2}(I + \vec{r}\cdot\vec{\sigma}),$$

where $\vec{t}$ is a vector and T is a real 3×3 matrix. For each channel, we first need to identify these parameters. The following lemma comes in handy in simplifying the optimisation problem.

Lemma 2.7 (Cf. [81, Thm. 3.10.11])**.** *A continuous convex function* f *on a compact convex set attains its global maximum at an extreme point of its domain.* ∎

Lemma 2.8. *For any quantum channel* $\mathcal{M}_o$ *we have*

$$\sup_{\rho,\sigma} D_\alpha(\mathcal{M}_o(\rho)\|\mathcal{M}_o(\sigma)) = \sup_{|\psi\rangle,|\phi\rangle} D_\alpha(\mathcal{M}_o(|\psi\rangle\langle\psi|)\|\mathcal{M}_o(|\phi\rangle\langle\phi|)),$$

that is, pure states are sufficient for the maximisation of the Rényi divergence with channel $\mathcal{M}_o$*.*

Proof. This is a consequence of Lemma 2.7. Note that the space of quantum states is a convex set; on the other hand, the Rényi divergence is a convex function, and we actually need convexity separately in each argument. Therefore the optimal states are extreme points of the set, i.e. pure states. ∎

Remark 2.6. *Since we will focus on 2-level systems, we should recall that a special property of the convex set of qubits which is not shared by n-level systems with $n \geq 3$ is that every boundary point of the set is an extreme point. Since the states on the surface of the Bloch sphere are mapped onto the states on the surface of the ellipsoid, the global maximum will be achieved by a pair of states on the surface of the output ellipsoid.*

Example 2 (Depolarizing channel)**.** *For $0 \leq q \leq 1$, the depolarizing channel is defined as follows:*

$$\rho \mapsto (1-q)\rho + q\frac{I}{2},$$

that is, the depolarizing channel transmits the state with probability $(1-q)$ or replaces it with the maximally mixed state with probability q. In both generalised Chernoff and Hoeffding exponents, we should be dealing with two optimisations, one over (ρ, σ) and the other over $0 \leq \alpha \leq 1$. We can take the supremum over the state pair inside each expression and deal with α next. Hence, we start with the supremum of the Rényi divergence employing Lemma 2.8.

For the depolarizing channel, it can be easily seen that

$$\vec{t} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \textit{and} \quad T = \begin{pmatrix} 1-q & 0 & 0 \\ 0 & 1-q & 0 \\ 0 & 0 & 1-q \end{pmatrix}.$$

Therefore, the set of output states consists of a sphere of radius $1-q$ centered at the origin, i.e. $r_x^2 + r_y^2 + r_z^2 = (1-q)^2$. Note that we only consider the states on the surface of the output sphere. Because of the symmetry of the problem and the fact that divergence is larger on orthogonal states, we can choose any two states at the opposite sides of a diameter. Here for simplicity we choose the states corresponding to $\vec{r}_1 = (r_x = 0, r_y = 0, r_z = 1-q)$ and $\vec{r}_2 = (r_x = 0, r_y = 0, r_z = -(1-q))$ leading to the following states, respectively:

$$\rho' = (1 - \frac{q}{2}) \left|0\right\rangle\!\left\langle 0\right| + \frac{q}{2} \left|1\right\rangle\!\left\langle 1\right|, \tag{2.39}$$

$$\sigma' = \frac{q}{2} \left|0\right\rangle\!\left\langle 0\right| + (1 - \frac{q}{2}) \left|1\right\rangle\!\left\langle 1\right|. \tag{2.40}$$

Then it can be easily seen that

$$\sup_{\rho,\sigma} D_\alpha(\mathcal{M}_o(\rho) \| \mathcal{M}_o(\sigma)) = \frac{1}{\alpha - 1} \log Q(q, \alpha), \tag{2.41}$$

where $Q(q,\alpha) = (1-\frac{q}{2})^{\alpha}(\frac{q}{2})^{1-\alpha} + (1-\frac{q}{2})^{1-\alpha}(\frac{q}{2})^{\alpha}$. *By plugging back into the respective equations, we have for* $0 \le r \le -(1-q)\log\frac{q}{2-q}$ *and* $(1-q)\log\frac{q}{2-q} \le a-b \le -(1-q)\log\frac{q}{2-q}$,

$$C^{\mathbb{A}_n^{c,0}}(a,b|\mathcal{M}_o) = \sup_{0\le\alpha\le 1} -\log Q(q,\alpha) - \alpha a - (1-\alpha)b,$$
$$B_e^{\mathbb{A}_n^{c,0}}(r|\mathcal{M}_o) = \sup_{0\le\alpha\le 1} \frac{\alpha-1}{\alpha}\big(r - \frac{1}{\alpha-1}\log Q(q,\alpha)\big).$$

The function $Q(q,\alpha)$ *introduced above is important and will also appear in later examples; below we see some of its basic behaviour.*

$$\frac{\partial Q(q,\alpha)}{\partial\alpha} = \ln\frac{q}{2-q}\big((\frac{q}{2})^{\alpha}(1-\frac{q}{2})^{1-\alpha} - (\frac{q}{2})^{1-\alpha}(1-\frac{q}{2})^{\alpha}\big),$$
$$\frac{\partial^2 Q(q,\alpha)}{\partial\alpha^2} = (\ln\frac{q}{2-q})^2 Q(q,\alpha),$$

where $\frac{\partial}{\partial\alpha}$ *and* $\frac{\partial^2}{\partial\alpha^2}$ *denote the first and second-order partial derivatives with respect to the variable* α. *It can also be easily checked that* $\log\frac{q}{2-q} \le 0$, $0 \le Q(q,\alpha) \le 1$.

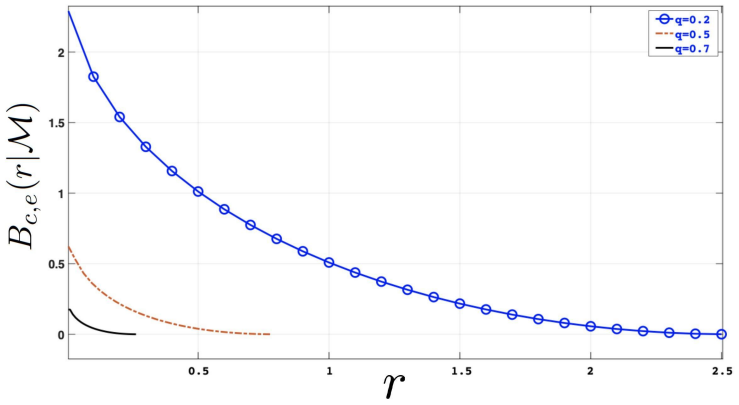

Figure 2.4: Hoeffding bound for depolarizing channel when entangled inputs are not allowed. The legitimate values of r for each exponent are imposed by Strong Stein's lemma and differ with q as $r = (q-1)\log\frac{q}{2-q}$.

Let $\overline{C}(\alpha)$ *denote the expression inside the supremum in* $C^{\mathbb{A}_n^{c,0}}(a,b|\mathcal{M}_o)$. *For the generalised Chernoff bound, from the observations above and some*

algebra, it can be seen that

$$\frac{\partial \overline{C}(\alpha)}{\partial \alpha} = 0 \implies \alpha = \frac{1}{2} - \frac{\log \frac{\log \frac{q}{2-q} + (a-b)}{\log \frac{q}{2-q} - (a-b)}}{2 \log \frac{q}{2-q}}. \tag{2.42}$$

On the other hand, it can be checked that $\frac{\partial^2 \overline{C}(\alpha)}{\partial \alpha^2} \geq 0$*, making sure that the generalized Chernoff bound is a convex function and also that the above zero is unique. Note that the generalized Chernoff bound is not a monotonic function since* $\frac{\partial \overline{C}(\alpha)}{\partial \alpha}$ *obviously changes sign, hence the zero is not necessarily at the ends of the interval.*

For the Hoeffding exponent $B_e^{\mathbb{A}_n^{c,0}}(r|\mathcal{M}_o)$*, finding a compact formula for the global maximum is not possible. However, numerical simulation guarantees that* $B_e^{\mathbb{A}_n^{c,0}}(r|\mathcal{M}_o)$ *is a convex function that the first derivative has a unique zero. We solved the optimisation numerically for depolarizing channel with three different parameter, see Fig. 2.4.*

Example 3 (Pauli channel)**.** *Let* $\{p_I, p_x, p_y, p_z\}$ *be a probability distribution. The Pauli channel is defined as follows:*

$$\rho \to p_I \rho + \sum_{i=x,y,z} p_i \sigma_i \rho \sigma_i,$$

that is, it returns the state with probability p_I *or applies Pauli operators* $\sigma_x, \sigma_y, \sigma_z$ *with probabilities* p_x, p_y *and* p_z*, respectively.*

For this channel, it can be seen by some algebra that (see e.g. [65, Sec. 5.3] and [83])

$$\vec{t} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} p_I + p_x - p_y - p_z & 0 & 0 \\ 0 & p_I - p_x + p_y - p_z & 0 \\ 0 & 0 & p_I - p_x - p_y + p_z \end{pmatrix}. \tag{2.43}$$

Therefore, the states on the surface of the Bloch sphere are mapped into the surface of the following ellipsoid:

$$\left(\frac{r_x}{p_I + p_X - p_Y - p_z}\right)^2 + \left(\frac{r_y}{p_I - p_X + p_Y - p_z}\right)^2 + \left(\frac{r_z}{p_I - p_X - p_Y + p_z}\right)^2 = 1. \tag{2.44}$$

Note that the Pauli channel shrinks the unit sphere with different magnitudes along each axis, and the two states on the surface of the ellipsoid that have

the largest distance depends on the lengths of the coordinates on each axis. We need to choose the states along the axis that is shrunk the least. We define the following:

$$p_{max} = \max\{|p_I + p_X - p_Y - p_Z|, |p_I - p_X + p_Y - p_Z|, |p_I - p_X - p_Y + p_Z|\}, \tag{2.45}$$

then from the symmetry of the problem and the fact that the eigenvalues of the state $\vec{r} = (r_x, r_y, r_z)$ *are* $\{\frac{1-|\vec{r}|}{2}, \frac{1+|\vec{r}|}{2}\}$*, the following can be seen after some algebra:*

$$\sup_{\rho,\sigma} D_\alpha(\mathcal{M}_o(\rho)\|\mathcal{M}_o(\sigma)) = \frac{1}{\alpha - 1}\log Q(1 - p_{max}, \alpha). \tag{2.46}$$

From this, for $0 \le r \le -p_{max}\log\frac{1-p_{max}}{1+p_{max}}$ *and* $p_{max}\log\frac{1-p_{max}}{1+p_{max}} \le a-b \le -p_{max}\log\frac{1-p_{max}}{1+p_{max}}$*, we have*

$$C^{\mathbb{A}_n^{c,0}}(a, b|\mathcal{M}) = \sup_{0\le\alpha\le1} -\log Q(1 - p_{max}, \alpha) - \alpha a - (1-\alpha)b,$$
$$B_e^{\mathbb{A}_n^{c,0}}(r|\mathcal{M}) = \sup_{0\le\alpha\le1} \frac{\alpha - 1}{\alpha}\Big(r - \frac{1}{\alpha - 1}\log Q(1 - p_{max}, \alpha)\Big).$$

Similar to our findings in the Example 2, we can show that the generalised Hoeffding bound is maximised at the following point

$$\alpha = \frac{1}{2} - \frac{\log\frac{\log\frac{1-p_{max}}{1+p_{max}} + (a-b)}{\log\frac{1-p_{max}}{1+p_{max}} - (a-b)}}{2\log\frac{1-p_{max}}{1+p_{max}}},$$

and this point is unique. The same conclusion using numerical optmisations indicates that the Hoeffding bound of the Pauli channel resembles that of the depolarizing channel. Note that a depolarizing channel with parameter q *is equivalent to Pauli channel with parameters* $\{p_I = 1 - 3q/4, p_x = q/4, p_y = q/4, p_z = q/4\}$ *[65, Ex.5.3].*

Example 4 (Amplitude damping channel)**.** *The amplitude damping channel with parameter* $0 \le \gamma \le 1$ *is defined as follows:*

$$\rho \mapsto \sum_{i=0,1} A_i \rho A_i, \tag{2.47}$$

where the Kraus operators are given as $A_0 = \sqrt{\gamma}\,|0\rangle\langle 1|$ *and* $A_1 = |0\rangle\langle 0| + \sqrt{1-\gamma}\,|1\rangle\langle 1|$*.*

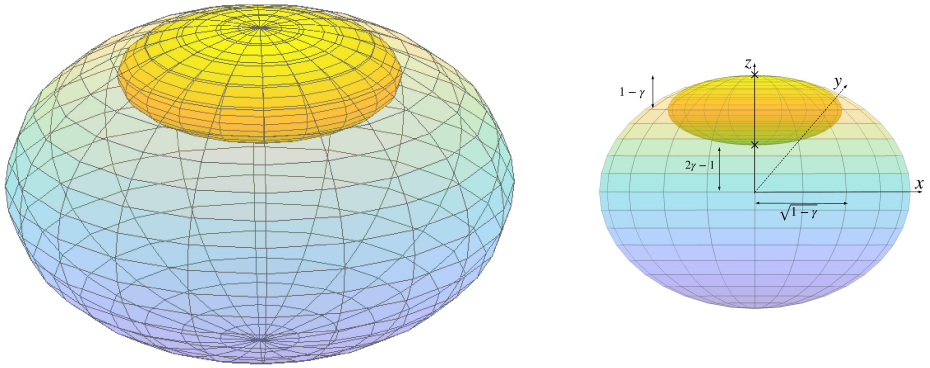

Figure 2.5: The Bloch sphere and its image under the amplitude damping channel with parameter γ. There are two large and one small semi-axes. As indicated in the right-hand-side figure, the points $(0,0,1)$ and $(0,0,2\gamma-1)$ are the intersection points of the surface of the displaced ellipsoid with the z axis; the former point also is its intersection point with the Bloch sphere.

For this channel, simple algebra shows that

$$\vec{t} = \begin{pmatrix} 0 \\ 0 \\ \gamma \end{pmatrix} \quad \textit{and} \quad T = \begin{pmatrix} \sqrt{1-\gamma} & 0 & 0 \\ 0 & \sqrt{1-\gamma} & 0 \\ 0 & 0 & 1-\gamma \end{pmatrix}. \tag{2.48}$$

Note unlike depolarizing and Pauli channels, $\vec{t}$ has a non-zero element for the amplitude damping channel, i.e. amplitude damping channel is not unital. The non-zero $\vec{t}$ indicates shifting the center of the ellipsoid. The output states of the amplitude damping channel is depicted in Fig. 2.5. Some algebra reveals the equation of the image to be as follows:

$$(\frac{r_x}{\sqrt{1-\gamma}})^2 + (\frac{r_y}{\sqrt{1-\gamma}})^2 + (\frac{r_z-\gamma}{1-\gamma})^2 = 1. \tag{2.49}$$

To calculate the divergence, we choose the optimal states on $x-z$ plane as $\vec{r}_1 = (\sqrt{1-\gamma}, 0, \gamma)$ and $\vec{r}_2 = (-\sqrt{1-\gamma}, 0, \gamma)$. It can be numerically checked that these points lead to maximum divergence. These two points correspond to the following states, respectively:

$$\rho_1 = \frac{1}{2}\begin{pmatrix} 1+\gamma & \sqrt{1-\gamma} \\ \sqrt{1-\gamma} & 1-\gamma \end{pmatrix} \quad \textit{and} \quad \rho_2 = \frac{1}{2}\begin{pmatrix} 1+\gamma & -\sqrt{1-\gamma} \\ -\sqrt{1-\gamma} & 1-\gamma \end{pmatrix}.$$

Since $|\vec{r}_1| = |\vec{r}_2| = \sqrt{\gamma^2 - \gamma + 1}$*, both states have the following eigenvalues:*

$$\lambda_1, \lambda_2 = \frac{1 \pm \sqrt{\gamma^2 - \gamma + 1}}{2},$$

and since ρ_1 *and* ρ_2 *obviously do not commute, we find the eigenvectors for* ρ_1 *and* ρ_2 *respectively as follows:*

$$|\nu_1\rangle = \frac{1}{\sqrt{1 + (\frac{2\lambda_1 - 1 - \gamma}{\sqrt{1-\gamma}})^2}} \begin{pmatrix} 1 \\ \frac{2\lambda_1 - 1 - \gamma}{\sqrt{1-\gamma}} \end{pmatrix}, \quad |\nu_2\rangle = \frac{1}{\sqrt{1 + (\frac{2\lambda_2 - 1 - \gamma}{\sqrt{1-\gamma}})^2}} \begin{pmatrix} 1 \\ \frac{2\lambda_2 - 1 - \gamma}{\sqrt{1-\gamma}} \end{pmatrix},$$

and

$$|\mu_1\rangle = \frac{1}{\sqrt{1 + (\frac{2\lambda_1 - 1 - \gamma}{\sqrt{1-\gamma}})^2}} \begin{pmatrix} 1 \\ -\frac{2\lambda_1 - 1 - \gamma}{\sqrt{1-\gamma}} \end{pmatrix}, \quad |\mu_2\rangle = \frac{1}{\sqrt{1 + (\frac{2\lambda_2 - 1 - \gamma}{\sqrt{1-\gamma}})^2}} \begin{pmatrix} 1 \\ -\frac{2\lambda_2 - 1 - \gamma}{\sqrt{1-\gamma}} \end{pmatrix}.$$

The following can be seen after some algebra:

$$\sup_{\rho,\sigma} D_\alpha(\mathcal{M}_o(\rho) \| \mathcal{M}_o(\sigma)) = \frac{1}{\alpha - 1} \log W(\gamma, \alpha),$$

where

$$W(\gamma,\alpha) = \lambda_1 \left(\frac{1 - (\frac{2\lambda_1 - 1 - \gamma}{\sqrt{1-\gamma}})^2}{1 + (\frac{2\lambda_1 - 1 - \gamma}{\sqrt{1-\gamma}})^2} \right)^2 + \lambda_2 \left(\frac{1 - (\frac{2\lambda_2 - 1 - \gamma}{\sqrt{1-\gamma}})^2}{1 + (\frac{2\lambda_2 - 1 - \gamma}{\sqrt{1-\gamma}})^2} \right)^2 + \frac{(Q(1 - \sqrt{\gamma^2 - \gamma + 1}, \alpha))(1 - \frac{(2\lambda_1 - 1 - \gamma)(2\lambda_2 - 1 - \gamma)}{(\sqrt{1-\gamma})^2})^2}{(1 + (\frac{2\lambda_1 - 1 - \gamma}{\sqrt{1-\gamma}})^2)(1 + (\frac{2\lambda_2 - 1 - \gamma}{\sqrt{1-\gamma}})^2)}.$$

We also have

$$D(\mathcal{M}) = \lambda_1 \log \lambda_1 + \lambda_2 \log \lambda_2 - \lambda_1 \log \lambda_1 \left(\frac{1 - (\frac{2\lambda_1 - 1 - \gamma}{\sqrt{1-\gamma}})^2}{1 + (\frac{2\lambda_1 - 1 - \gamma}{\sqrt{1-\gamma}})^2} \right)^2 - \lambda_2 \log \lambda_2 \left(\frac{1 - (\frac{2\lambda_2 - 1 - \gamma}{\sqrt{1-\gamma}})^2}{1 + (\frac{2\lambda_2 - 1 - \gamma}{\sqrt{1-\gamma}})^2} \right)^2 - \frac{(\lambda_1 \log \lambda_2 + \lambda_2 \log \lambda_1)(1 - \frac{(2\lambda_1 - 1 - \gamma)(2\lambda_2 - 1 - \gamma)}{(\sqrt{1-\gamma})^2})^2}{(1 + (\frac{2\lambda_1 - 1 - \gamma}{\sqrt{1-\gamma}})^2)(1 + (\frac{2\lambda_2 - 1 - \gamma}{\sqrt{1-\gamma}})^2)}.$$

The cumbersome expressions reflect the complexity of analytically solving the optimisations; however, it can be seen numerically that the first derivative of the generalised Chernoff bound has a unique zero and its second derivative is positive ensuring the convexity. We calculate the Hoeffding exponent for three different parameters of the amplitude damping channel, see Fig. 2.6.

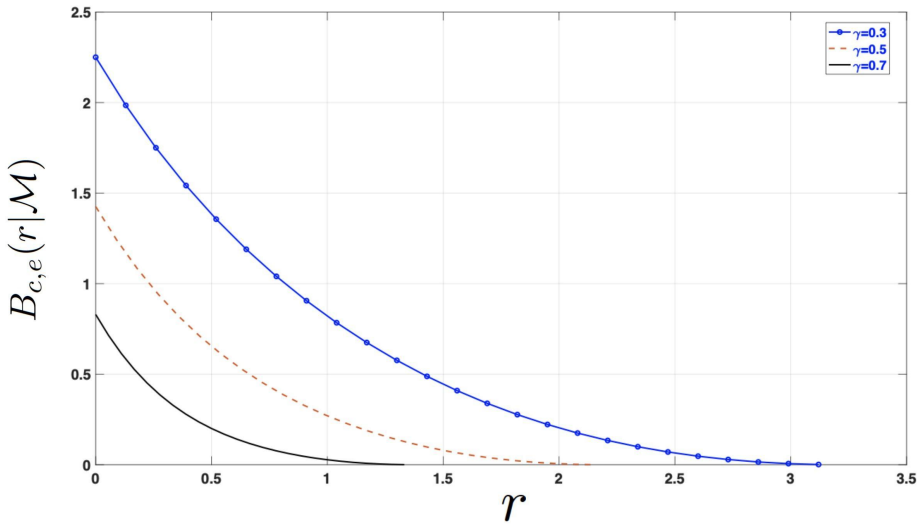

Figure 2.6: Hoeffding exponent for amplitude damping channel when entangled inputs are not allowed. The legitimate values of r for each exponent are imposed by the strong Stein's lemma and differ as a function of γ, i.e. $D(\mathcal{M})$.

2.8 Conclusion

In an attempt to further extend the classical results [35] to quantum channels, showed that for the discrimination of a pair of cq-channels, adaptive strategies cannot offer any advantage over non-adaptive strategies concerning the asymmetric Hoeffding and the symmetric Chernoff problems in the asymptotic limit of error exponents. Our approach consists in associating to the cq-channels a pair of classical channels. This latter finding led us to prove the optimality of non-adaptive strategies for discriminating quantum channels via a subclass of protocols which only use classical feed-forward and product inputs. For a broader subclass of protocols that allow classical feed-forward and entangled inputs, we leave open the question of optimality of non-adaptive protocols for entanglement-breaking channels.

Beyond cq-channels, by proving a lower bound on the non-adaptive discrimination error of quantum channels, we showed an asymptotic separation between the Chernoff and Hoeffding exponents of adaptive and non-adaptive strategies for a pair of entanglement-breaking channels.

We also studied the hypothesis testing of binary information via a noisy quantum channel and show that when entangled inputs are not allowed, non-adaptive strategies using the channel are optimal and when entangled

inputs are allowed, we showed the optimality of non-adaptive strategies for entanglement-breaking channels.

Chapter 3

Multi-User Distillation of Common Randomness and Entanglement from Quantum States

We study the rates at which noisy quantum correlations, i.e. quantum states, can be asymptotically distilled into noiseless classical and quantum ones using local operations and classical communications (LOCC). Concerning the first task, we find two lower bounds on distillable common randomness (CR), i.e. the amount of distilled classical bits in excess of consumed communications, based on two protocols. While the first lower bound is constructed from the communication for omniscience (CO) protocol, the protocol of the second lower bound requires finding a simultaneous decoder for compression of correlated classical sources with quantum side information at the decoder, a long standing open problem in quantum information theory. For the second task we focus on constructing protocols for distillation of Greenberger-Horne-Zeilinger (GHZ) states from multipartite pure quantum states. We use the idea of *making coherent* to construct our protocols from the protocols of the previous task. Each protocol leads to a lower bound; while the first lower bound re-derives an existing result, the second lower bound improves on it and generalises a number of other existing bounds.

3.1 Background

Restricting the operations of distinct parties in a network signifies the importance of certain resources shared by the parties. When multiple parties are

allowed to exchange classical messages through rate-limited channels, random bits shared between them become a resource. On the other hand, imposing locality conditions, i.e. each party can act locally on its system but allowing for free classical communications, introduces entanglement as a resource: perhaps the most important feature of quantum mechanics that renders it different from the classical theory. A number of important tasks in quantum information theory are accomplished by virtue of entanglement. Given its importance in variety of the information-processing tasks, one of the ongoing programs of the quantum information theory is to understand what the basically different types of the entanglement are and what the possibility of conversion between them allowing only local operations and exchanging classical messages (LOCC) is. The entanglement content of pure bipartite states can be considered fully understood: under LOCC, two parties sharing a pure bipartite state $|\psi\rangle^{AB}$ can convert a large number of copies of it into as many copies as possible of another bipartite pure state $|\phi\rangle^{AB}$ and the conversion rate is determined by a single number, the ratio of the entanglement entropy of $|\psi\rangle^{AB}$ to that of $|\phi\rangle^{AB}$, i.e., $E(|\psi\rangle)/E(|\phi\rangle)$, where the entanglement entropy is defined as the von Neumann entropy of each reduced state. However, the theory of asymptotic manipulation of pure multi-partite entanglement is far less understood: unlike (pure) bipartite entanglement, the lack of a "gold standard" unique multipartite state into which an arbitrary pure state can be reversibly transformed makes the study of multipartite entanglement much more intricate. However, the multipartite entanglement does not cease to have a resource character. The Greenberger-Horne-Zeilinger (GHZ) state

$$|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{2}}(|00...0\rangle + |11...1\rangle),$$

which is the important constituent of the quantum secret sharing protocol [47], is perhaps the simplest state to contain genuine multipartite entanglement.

The first problem we study in this chapter is distillation of common randomness (CR) from mutlipartite mixed quantum states. It is obvious that allowing free classical communications leads to infinite random bits shared between parties; our approach consists of restricting the classical communications between parties. More precisely, we are interested in *distillable CR* [42], the CR that we get after subtracting the consumed communication from the total achieved CR. We tend to call the achieved shared randomness CR rather than secret key; however, if we envision an eavesdropper who observes the classical communications between parties only, the same rate for distillable CR is also a rate for achievable secret keys. This result can be shown via privacy amplification, but perhaps more easily, this conclusion fol-

lows from the resource calculus of Shannon theory: roughly speaking, parties initially use certain amount of secret keys to communicate with each other; it is well-known that secret communications is possible at the rate of the initial keys. Finally the consumed initial rate is subtracted from the achieved CR rate; for more their relation we refer the reader to [46]. Our main results in CR distillation problem include two lower bounds on the distillable CR from mixed states. While the first lower bound requires only tools from classical information theory, for the second lower bound, which improves upon the first, we construct a novel simultaneous decoder for the compression of correlated classical sources by random binning with quantum side information at the decoder.

The second problem we study in this chapter is the distillation task of converting asymptotically many copies of a multipartite pure quantum state into as many GHZ states as possible under LOCC. The key to the proof of our results on GHZ distillation is to replace the protocols for CR distillation by quantum protocols by *making coherent*. We have basically used lessons learned in [19], [84], [41] and [85] and observations of [86] regarding making protocols coherent. The first idea of making protocols coherent is that classical words/letters x become basis states $|x\rangle$ of the Hilbert space. Functions $f : x \to f(x)$ thus induce linear operators on Hilbert space, but only permutations (one-to-one functions) are really interesting, since they give rise to unitaries (isometries, resp.). The second idea is thus to make classical computations first reversible, by extending them into one-to-one functions. The last step is to use the local decodings, which exist by the classical theorems. In summary, "making coherent" means we can take a classical protocol working on letters and turn it into a sequence of unitaries acting as permutations on the basis states, and that we can run perfectly well on superpositions.

This way, our two lower bounds on CR distillation translate into two lower bounds on distillable GHZ state. The first lower bound was already proven by Vrana and Christandl [52]. However, their work is unsatisfactory because of at least two reasons: first, it seems an ad hoc method, and is not clear how to adapt it to other target states or multiple target states and their tradeoffs; second, the expression obtained is in terms of classical entropies, of random variables obtained from local measurements of the state. Our second lower bound improves this result and unifies a number of results in the literature.

Notation. Capital letters X, Y, etc. denote random variables, whose realizations and the alphabets are shown by the corresponding small and calligraphic letters, respectively: $X = x \in \mathcal{X}$. Quantum systems A, B, etc. are associated with (finite-dimensional) Hilbert spaces A, B, etc. whose dimensions are denoted by $|A|$, $|B|$, etc. Multipartite systems $AB \dots Z$ are

described by tensor product Hilbert space $A \otimes B \otimes \cdots \otimes Z$. We identify states with their density operators and use superscripts to denote the systems on which the mathematical objects are defined. For any positive integer m, we use the notation $[m] = \{1, \ldots, m\}$. For conciseness, we denote the tuple $(X_1, \ldots, X_m)$ by $X_{[m]}$. More generally, for a set L, we write $X_L = (X_i : i \in L)$. Throughout the chapter, log denotes by default the binary logarithm.

Beyond the von Neumann entropy of a state, we also use the conditional von Neumann entropy of a bipartite state ρ^{AB}, defined as $S(A|B) = S(AB) - S(B)$, and the quantum mutual information $I(A:B) = S(A) + S(B) - S(AB)$. For classical systems (random variables), the von Neumann entropy reduces to the Shannon entropy, denoted $H(X)$.

3.2 Common randomness distillation and omniscience

We shall consider *common randomness distillation* (in the source model). This means that we have m spatially separated parties sharing $n \gg 1$ copies of an m-partite quantum state $\rho^{A_1 \cdots A_m}$, i.e. party i holds the subsystem A_i^n. All parties can communicate to each other through a public noiseless classical broadcast channel of unlimited capacity. The following definition is a generalization of the bipartite case in [24].

Definition 2 (Common randomness distillation protocol)**.** *Let ρ be a state on $A_{[m]} = A_1 \otimes \cdots \otimes A_m$, and consider the initial state $\rho^{\otimes n}$. Let r be the total number of rounds; for $i \in [m]$, let B_i be a local quantum system used by party i to store quantum information, originally in state $|0\rangle\langle 0|$; for all $j \in [r]$, let $U_{i_j}^j$ be classical systems to store the classical communication of party i_j after round j.*

Step 1) *Terminal $i_1 \in [m]$ applies the completely positive instrument*

$$\Phi_{i_1}^1 : A_{i_1}^n \otimes B_{i_1} \to A_{i_1}^n \otimes B_{i_1} \otimes U_{i_1}^1,$$

and broadcasts $U_{i_1}^1$ to the other parties. This means that the shared state $\rho^{\otimes n}$ is mapped to the state

$$\rho^{(1)} = \sum_u (\mathrm{id}_{A^n_{[m]\setminus i_1}} \otimes \Phi_{i_1}^1(u))(\rho^{\otimes n} \otimes |0\rangle\langle 0|^{B_{i_1}}) \otimes |0\rangle\langle 0|^{B_{[m]\setminus i_1}} \otimes |u\rangle\langle u|^{U_{i_1}^1}$$

on $A_{[m]}^n \otimes B_{[m]} \otimes U_{i_1}^1$.

Step j) *Terminal $i_j \in [m]$ applies a completely positive map*

$$\Phi_{i_j}^j : A_{i_j}^n \otimes B_{i_j} \otimes U^{[j-1]} \to A_{i_j}^n \otimes B_{i_j} \otimes U^{[j]},$$

where we use the shorthand $U^{[j-1]} = U_{i_1}^1 U_{i_2}^2 \dots U_{i_{j-1}}^{j-1}$, and broadcasts $U_{i_j}^j$ to the rest of the parties. This maps the previous state $\rho^{(j-1)}$ to the new state $\rho^{(j)}$ on $A_{[m]}^n \otimes B_{[m]} \otimes U^{[j]}$.

Step r+1) *After the last communication, each party $i \in [m]$, measures its systems by means of a POVM on $A_i^n \otimes B_i \otimes U^{[r]}$ and indexed by $\{1, \dots, |V|\}$, giving rise to a random variable V_i with distribution $p_i(v)$.*

Let R_i denote the total rate of classical communication by the i-th party.

Remark 3.1. *This CR distillation protocol is a general LOCC procedure, in which we explicitly keep track of the classical communication.*

Definition 3. *A number $R = \frac{1}{n} \log |V| - \sum_{i=1}^m R_i$ will be called an achievable distillable CR rate for common randomness distillation if for every $\varepsilon > 0$ and sufficiently large n, there exists a common randomness distillation protocol where the total communication of party i is bounded by nR_i bits, such that $\{V_i\}_{i=1}^m$ satisfy*

$$\Pr\{V_1 = \dots = V_m\} \geq 1 - \varepsilon, \tag{3.1}$$

$$\frac{1}{2} \|p_1 - u_V\|_1 = \frac{1}{2} \sum_v \left| p_1(v) - \frac{1}{|V|} \right| \leq \varepsilon, \tag{3.2}$$

where u_V denotes the uniform distribution. The maximal achievable rate for distillable CR is called the distillable CR capacity *$D_{CR}(\rho)$.*

Now, we prove two achievability results for the distillable CR rate, and in the next section two achievability results for the distillable GHZ rate, all based on a subclass of protocols with "non-interactive communication", which are called this way because each party broadcasts only one message to all others that depends only on their own local state. The proofs of the distillable CR results is based on our generalization of the communication for omniscience (CO) [45]. We present two protocols for our achievability bounds. The first protocol uses full local measurements and communication; the second uses instruments that initially turn the state into a classical-quantum state, and thus generalizes the first.

Theorem 3.1. *Let $\rho^{A_1\dots A_m}$ be a quantum state and let $\{M^i_{x_i}\}_{x_i\in\mathcal{X}_i}$ denote a POVM used by party i. Define $p(x_{[m]})$ as the joint distribution of m random variables X_i recording the measurement outcomes on ρ:*

$$p(x_1,\dots,x_m) = \operatorname{Tr}\rho(M^1_{x_1}\otimes\cdots\otimes M^m_{x_m}).$$

The following is an achievable rate for the distillable CR:

$$R = H(X_{[m]}) - R^c_{CO},$$

where $R^c_{CO} = \min_{R_{[m]}\in\mathcal{R}_c}\sum_{i=1}^m R_i$, and $\mathcal{R}_c$ is the rate region of tuples $R_{[m]} = (R_1,\dots,R_m)$ given as follows:

$$\forall L\subsetneq[m]\quad \sum_{j\in L} R_j \geq H(X_L|X_{[m]\setminus L}).$$

Proof. This really is an instance of the results of Csiszár and Narayan [45], who prove precisely that for the RVs $X_1,\dots,X_m$, the set $\mathcal{R}_c$ is precisely the rate region of communication for omniscience, i.e. protocols at the end of which all users know $X_{[m]}$ up to arbitrarily small error probability. This shows that $R = H(X_{[m]}) - R^c_{\mathrm{CO}}$ is an achievable rate for distillable CR. Incidentally, in [45] it is actually shown to be the optimal CR rate for the given RVs. However, this is of less relevance for us, as different choices of local measurements lead to different tuples of RVs. The theorem is also a special case of Theorem 3.2 below. The basic idea of the coding procedure, referred to as *random binning*, is not much different than that of hash functions. Each classical sequence obtained from the local measurements is randomly and uniformly assigned a bin index; if the number of bins (the range of the hash function) is large enough compared to the jointly entropy-typical sets, a randomly selected mapping of classical sequences will suffer a collision with small probability. This means that the classical information can be extracted from their index set with high probability.

In detail, the i-th party assigns each sequence $x_i^n\in\mathcal{X}_i^n$ to one of 2^{nR_i} bins; all parties broadcast the bin index associated to their obtained sequence, $(\mu_1,\dots,\mu_m)\in[M_1]\times\cdots\times[M_m]$, to the other parties. Then, the parties use joint typicality decoding to extract the sequences of other parties from their local information and $\mu_{[m]}$. That is, having received $\mu_{[m]\setminus i}$, the i-th party looks into the bins indexed $\mu_{[m]\setminus i}$ to find a unique tuple $\hat{x}^n_{[m]\setminus i}$ that is jointly typical with their observed x_i^n. An error occurs when one of the following events happen: the obtained sequence of tuples $x^n_{[m]}$ is not typical, or there is no jointly typical sequence $\hat{x}^n_{[m]\setminus i}x_i^n$, or there are two different jointly typical candidates $\hat{x}^n_{[m]\setminus i}x_i^n$ and $\check{x}^n_{[m]\setminus i}x_i^n$ in the correct bins. These in

fact are the same conditions as for correct decodability in the Slepian-Wolf problem [87, Ch. 15.4], in the special case that $R_i = H(X_i) + \delta$, for some $\delta > 0$. The analysis there shows that the error probability indeed goes to zero, with high probability for a randomly chosen binning strategy, if for all $L \subseteq [m] \setminus i$ it holds $\sum_{j \in L} R_j \geq H(X_L | X_{[m] \setminus L}) + \delta$, for some $\delta > 0$.

As $\mathcal{R}_c$ consists of the rate tuples satisfying these conditions for all $i \in [m]$, it means then that all parties can decode $x^n_{[m]}$ with high probability correctly, as $n \to \infty$. ∎

Theorem 3.2. *Let $\rho^{A_1 \dots A_m}$ be a quantum state and let $\mathcal{E}^i : A_i \to A'_i \otimes X_i$ be an instrument used by party i, with quantum registers A'_i and classical registers X_i. Define $\omega^{X_1 A'_1 \dots X_m A'_m}$ as the cq-state after applying the local instruments:*

$$\begin{aligned} \omega^{X_1 A'_1 \dots X_m A'_m} &= (\mathcal{E}^1 \otimes \cdots \otimes \mathcal{E}^m)\rho \\ &= \sum_{x_{[m]}} \left| x_{[m]} \right\rangle\!\left\langle x_{[m]} \right|^{X_{[m]}} \otimes (\mathcal{E}^1_{x_1} \otimes \cdots \otimes \mathcal{E}^m_{x_m})\rho. \end{aligned}$$

The following is an achievable rate for the distillable CR:

$$R = H(X_{[m]}) - R^{cq}_{CO},$$

where $R^{cq}_{CO} = \min_{R_{[m]} \in \mathcal{R}_{cq}} \sum_{i=1}^m R_i$, and $\mathcal{R}_{cq}$ is the rate region given as follows:

$$\forall j \in [m] \ \forall L \subseteq [m] \setminus j \quad \sum_{i \in L} R_i \geq S(X_L | X_{[m] \setminus L} A'_j). \tag{3.3}$$

Proof. Each party j evaluates a function $U_j := f_j(X^n_j) \in \{0,1\}^{nR_j}$ of their input, and broadcasts U_j to all other parties. The objective for party j is then, knowing $U_{[m] \setminus j}$, that they can decode $X^n_{[m]}$ from $B^n_j := X^n_j A'^n_j$ by a suitable measurement.

Thus it is unsurprising that the answer should be given by a quantum version of Slepian-Wolf coding. Indeed, for each fixed j, the necessity and sufficiency of the rate conditions in Eq. (3.3) is proved in [88, Thm. IV.14 & Cor. IV.16], generalising [89]. However, this is not enough because we need a code (i.e. a set of encoding functions, one for each party) that works for all parties simultaneously, allowing each of them to recover $X^n_{[m]}$ from their A'_j and $U_{[m]}$. To achieve this, we use random binning: each party j uses a random function $F_j : \mathcal{X}^n_j \to \{0,1\}^{nR_j}$ (to be precise, we draw them independently from m 2-universal families). In the case of classical A'_i, it is well-known that this strategy works as long as the rate conditions in Eq. (3.3) are satisfied, by using a joint typicality decoder; see the proof sketch of Theorem 3.1; cf. the discussion of Slepian-Wolf data compression in [87,

Ch. 15.4]. In the general quantum case, joint typicality decoding presents considerable technical difficulties, but they were eventually overcome by Sen [90].

In Lemma 3.4 in the Appendix, we show how to use Sen's joint typicality construction to build a joint decoder that achieves small expected decoding error for party j, $\mathbb{E}_{F_{[m]\setminus j}} P_e(j) \leq \varepsilon$, for all $\varepsilon > 0$ and sufficiently large n, if the rates satisfy

$$\forall\ \varnothing \neq I \subset [m] \setminus j \quad \sum_{i \in I} R_i \geq H(X_I | X_{[m]\setminus j \setminus I} B_j) + \delta,$$

where $\delta > 0$ is an arbitrary constant. Thus, summing over all j, and recalling that $B_j = X_j A'_j$, we get $\mathbb{E}_{F_{[m]}} \big(P_e(1) + \ldots + P_e(m) \big) \leq m\varepsilon$ for all sufficiently large n, if the rates satisfy

$$\forall j \in [m]\ \forall L \subseteq [m] \setminus j \quad \sum_{i \in L} R_i \geq S(X_L | X_{[m]\setminus L} A'_j) + \delta.$$

Since $\varepsilon, \delta > 0$ are arbitrary, the claim follows.

This shows that the rate tuples $(R_1, \ldots, R_m) \in \mathcal{R}_{cq}$ are all achievable to provide omniscience of the $X_{[m]}^n$ among all m parties. Concentrating the randomness in the shared random variables into uniform randomness, yielding a rate of $H(X_{[m]})$, and subtracting the communication $\sum_i R_i$, completes the proof that $R = H(X_{[m]}) - R_{\mathrm{CO}}^{cq}$ is an achievable rate for distillable CR. ∎

Remark 3.2. *It is easy to see, via the Slepian-Wolf connection made in the above proof, that given the cq-state $\omega^{X_{[m]} A'_{[m]}}$, any non-interactive protocol to achieve omniscience of $X_{[m]}$, by which party j broadcasts at asymptotic rate R_j, must necessarily satisfy $(R_1, \ldots, R_m) \in \mathcal{R}_{cq}$.*

Indeed, focusing on party j for the moment, for them to be able to reconstruct $X_1^n, \ldots, X_{j-1}^n, X_{j+1}^n, \ldots, X_m^n$ using $X_j^n A_j'^n$ and communications U_i from party $i \in [m] \setminus j$ at rate R_i, is precisely the task of correlated classical source coding with quantum side information at the decoder [88, 89]. For this, the conditions in Eq. 3.3 for the given j are necessary and sufficient. Since they have to hold for all j, it follows that $\mathcal{R}_{cq}$ is precisely the achievable region of rates for CO.

3.3 GHZ distillation from pure states

Now, we move on to using the above results on distillable CR to derive two lower bounds for the distillable entanglement from pure quantum states. The

first, Theorem 3.3, re-derives the result of [52], with a different, information theoretic proof, by making the protocol of Theorem 3.1 coherent. The second, which improves upon the preceding result, is obtained by making the protocol of Theorem 3.2 coherent. We use lessons learned in [19, 41, 84, 85], and observations of [86] regarding making protocols coherent.

In short, the first idea of making protocols coherent is that classical symbols x become basis states $|x\rangle$ of the Hilbert space. Functions $f : x \to f(x)$ thus induce linear operators on Hilbert space, but only permutations (resp. one-to-one functions) are really interesting, since they give rise to unitaries (resp. isometries). The second idea is thus to make classical computations first reversible, by extending them to one-to-one functions. The last step is to use the local decoding operations that exist by the "classical" theorems, which are cptp maps, in the form of their isometric Stinespring dilations [91]. In summary, "making coherent" means that we can take a classical protocol working on letters and turn it into a bunch of unitaries acting as permutations on the basis states, and that we can run perfectly well on superpositions.

As in CR distillation, we have m spatially separated parties, now sharing $n \gg 1$ copies of an m-partite *pure* quantum state $|\psi\rangle^{A_1 \dots A_m}$, i.e. party i holds the subsystem A_i^n. All parties can communicate to each other through a public noiseless classical broadcast channel of unlimited capacity.

Definition 4 (GHZ distillation protocol)**.** *The m parties, to convert the state $\psi^{\otimes n}$ to k copies of the GHZ state $|\Gamma_m\rangle$, they perform LOCC channels interactively in r rounds. Let $\sigma^{B_1^k \dots B_m^k}$ denote the final state after LOCC channels, where B_i denotes qubit systems. If*

$$\frac{1}{2} \left\| \sigma^{B_1^k \dots B_m^k} - |\Gamma_m\rangle\langle\Gamma_m|^{\otimes k} \right\|_1 \leq \varepsilon,$$

we call the protocol ε-accurate and the GHZ conversion rate is k/n. We call a number R an achievable rate for GHZ distillation if for all $\varepsilon > 0$ and sufficiently large n, there exists a sequence of ε-accurate protocols with conversion rate $R - \varepsilon$. The supremum of all achievable rates is the GHZ distillation capacity, $C_{GHZ}(\psi)$.

At the time of writing, there is no formula known for $C_{\mathrm{GHZ}}(\psi)$ for a general state; however various protocols (giving lower bounds) and upper bounds have been developed. Regarding the latter, this involves finding LOCC monotones that have certain requisite additivity and continuity properties. For example, in [5, Lemma 1 & Thm. 2] it was shown that for multipartite pure state transformation, all the $S(A_I)_\psi$, $I \subset [m]$, are such monotones, thus limiting the conversion rate for any target state. In the case of a GHZ state,

which has $S(A_I)_{\Gamma_m} = 1$ for all $\varnothing \neq I \subsetneq [m]$, this leads to

$$C_{\mathrm{GHZ}}(\psi) \leq \min_{\varnothing \neq I \subsetneq [m]} S(A_I)_\psi. \tag{3.4}$$

Incidentally, the right hand side equals the minimum of $C_{\mathrm{EPR}(i:j)}(\psi)$ over all $i \neq j$, according to Eq. (1.1), which even gives an operational meaning to the bound, since from a GHZ-state between m parties an EPR-state between any pair of parties can be obtained by LOCC.

In the introduction we have already referenced several GHZ distillation protocols. Here we briefly review a protocol based on *entanglement combing* [10], which results in a simple protocol and basic lower bound on the rate of GHZ distillation. The following lemma is also going to be invoked in the proofs of our main results.

Lemma 3.1. *Let $|\psi\rangle^{B_1 \ldots B_m}$ be a pure state shared among m parties. The following rate of GHZ state is distillable from $|\psi\rangle$ under LOCC:*

$$R_{comb} = \max_{i \in [m]} \left\{ \min_{I \subseteq [m] \setminus i} \frac{S(B_I)}{|I|} \right\}. \tag{3.5}$$

In particular, if $|\psi\rangle$ is genuinely multi-party entangled (i.e. it is not a product state w.r.t. any bipartite cut), then $R_{comb} > 0$.

Proof. The entanglement combing protocol [10] turns the given state into bipartite entanglement shared between a distinguished party, say i, and each of the other parties $j \in [m] \setminus i$. Let R_j denote the rate of the EPR pairs distilled between the distinguished party B_i and another party B_j. The following rate region is proven optimal for this task:

$$\forall I \subseteq [m] \setminus i \quad \sum_{j \in I} R_j \leq S(B_I). \tag{3.6}$$

By means of LOCC one can turn the combed entanglement into GHZ states shared between all parties. This can be done by letting party i teleport their information using the EPR pairs. In this case, the rates have to be equal, i.e. $R_1 = \ldots = R_m =: R_{\mathrm{comb}}$. Thus, from the rate region for combing Eq. (3.6), we have as a necessary and sufficent condition

$$\forall I \subseteq [m] \setminus i \quad |I| R_{\mathrm{comb}} \leq S(B_I), \tag{3.7}$$

which is satisfied by $R_{\mathrm{comb}} := \min_{I \subseteq [m] \setminus i} \frac{S(B_I)}{|I|}$. Finally, we optimise over the choice of distinguished party. ∎

Remark 3.3. *The preceding result shows that unless the state is a product state across some bipartite cut, the GHZ-rate is always positive. Such states are called "bi-separable", in which case evidently no GHZ states can be distilled, cf. Eq. (3.4). The rate R_{comb} is the baseline against which to compare any new protocol.*

It can be far from optimal, for example even if the initial $|\psi\rangle = |\Gamma_m\rangle$ is a GHZ state, then $R_{comb} = \frac{1}{m-1}$, while obviously $C_{GHZ}(\Gamma_m) = 1$.

In the proofs of our GHZ distillation protocols, we shall use the following rules from the resource calculus of quantum Shannon theory [2], where '$\geq$' means that the resources on the left hand side can be transformed asymptotically to the resources on the right hand side by local operations only; o is an arbitrarily small positive number.

Lemma 3.2 (Cancellation lemma [2, Lemma 4.6])**.** *For resources α, β, γ, if $\alpha + \gamma \geq \beta + \gamma$, then $\alpha + o\gamma \geq \beta$.* ■

Lemma 3.3 (Removal of o terms [2, Lemma 4.5])**.** *For resources, α, β, γ, if $\alpha + o\gamma \geq \beta$ and $\alpha \geq z\gamma$ for some real $z > 0$, then $\alpha \geq \beta$.* ■

Theorem 3.3 (Vrana and Christandl [52, Thm. 1])**.** *Let $|\psi\rangle = \sum \psi_{x_1 \ldots x_m} \left|x_{[m]}\right\rangle$ be a pure state written in the computational basis, and define $p(x_1, \ldots, x_m) = |\psi_{x_1 \ldots x_m}|^2$, the probability distribution of measuring ψ in the computational bases locally. Define the region $\mathcal{R}_c$ as the set of rate tuples $R_{[m]} = (R_1, \ldots, R_m)$ satisfying the following conditions,*

$$\forall I \subsetneq [m] \quad \sum_{j \in I} R_j \geq H(X_I | X_{[m] \setminus I}). \tag{3.8}$$

Finally, let $R^c_{CO} := \min_{R_{[m]} \in \mathcal{R}_c} \sum_{j=1}^m R_j$. Then,

$$C_{GHZ}(\psi) \geq H(X_{[m]}) - R^c_{CO}.$$

Proof. The m terminals share n copies of the pure state $|\psi\rangle = \sum_{x_1 \ldots x_m} \psi_{x_1 \cdots x_m} |x_1\rangle \cdots |x_m\rangle$, i.e.

$$\begin{aligned} |\psi\rangle^{\otimes n} &= \sum_{x_1^n \ldots x_m^n} \psi_{x_1^n \ldots x_m^n} |x_1^n\rangle \cdots |x_m^n\rangle, \text{ where} \\ \psi_{x_1^n \ldots x_m^n} &= \prod_{t=1}^n \psi_{x_{1,t} \cdots x_{m,t}} \text{ and} \\ \left|x_j^n\right\rangle &= |x_{j,1}\rangle \otimes \cdots \otimes |x_{j,n}\rangle. \end{aligned}$$

Let $f_j : \mathcal{X}_j^n \to \mathcal{U}_j$ be the Slepian-Wolf hash function used by party j in the classical part of the protocol of Theorem 3.1 (omniscience), and $(\Delta_{x_{[m]}^n}^{(j,u_{[m]})} : x_{[m]}^n)$ the POVM (decision rule) that they use to recover $x_{[m]}^n$ when the classical messages $u_{[m]}$ are broadcast.

In the first step, each party j will apply an isometry based on the mappings $x_j^n \longmapsto (f_j(x_j^n), x_j^n)$ for $j \in [m]$, namely

$$V_j = \sum_{x_j^n} \left|f_j(x_j^n), x_j^n\right\rangle\!\left\langle x_j^n\right|,$$

where $\{|u\rangle = \left|f_j(x_j^n)\right\rangle\}$ is computational basis for some Hilbert space $U_j = \mathrm{span}\{|u\rangle : u \in \mathcal{U}_j\}$. The state at the end of the first step is

$$|\psi'\rangle = \sum_{x_1^n \ldots x_m^n} \psi_{x_1^n \ldots x_m^n} |x_1^n, f_1(x_1^n)\rangle \cdots |x_m^n, f_m(x_m^n)\rangle\,.$$

Next comes the coherent transmission of the hash value u_j to other parties, which in fact is implementing a multi-receiver cobit channel [86], i.e. party j wishes to implement the isometry $|u_j\rangle \longmapsto |u_j\rangle^{\otimes m}$. This multi-receiver cobit channel can be implemented by teleportation through GHZ states. In order to coherently transmit nR_j bits, where $R_j \coloneqq \frac{1}{n}\log|U_j|$, nR_j GHZ states are needed, i.e. the following state:

$$|\Gamma_m\rangle^{\otimes nR_j} = \left(\frac{1}{\sqrt{2}}(|0\rangle^{\otimes m} + |1\rangle^{\otimes m})\right)^{\otimes nR_i}.$$

After implementing the multi-receiver cobit channel, the j-th party owns its initial share $\left|x_j^n\right\rangle$ as well as all the hash values broadcast to it. Thus, the overall state is

$$\left|\widetilde{\psi}\right\rangle = \sum_{x_1^n \ldots x_m^n} \psi_{x_1^n \ldots x_m^n} |x_1^n, f_1(x_1^n) \ldots f_m(x_m^n)\rangle \cdots |x_m^n, f_1(x_1^n) \ldots f_m(x_m^n)\rangle\,.$$

Having received the hash values, each party proceeds to recovering $x_{[m]}^n$. Each party locally runs its Slepian-Wolf decoder in a coherent fashion to work out the $\left|x_j^n\right\rangle$ of the other $m-1$ parties. More precisely, the j-th party applies the following controlled isometry on its corresponding systems:

$$\sum_{u_{[m]}} \left|u_{[m]}\right\rangle\!\left\langle u_{[m]}\right| \otimes V_D^{(j,u_{[m]})},$$

where the coherent measurement isometry of the j-th party is defined as:

$$V_D^{(j,u_{[m]})} = \sum_{\forall i \in [m]\, x_i^n \in f_i^{-1}(u_i)} \sqrt{\Delta_{x_{[m]}^n}^{(j,u_{[m]})}} \otimes \left|x_{[m]}^n\right\rangle, \tag{3.9}$$

with $\Delta_{x^n_{[m]}}^{(j,u_{[m]})}$ the POVM elements of the j-th decoder acting on A_j^n. The classical result of Csiszár and Narayan [45], i.e. Theorem 3.1 in the diagonal case, ensures successful decoding if the rates $R_{[m]}$ satisfy the conditions (3.8). The state after each party has applied their decoding isometry is as follows:

$$\begin{aligned}\left|\overline{\psi}\right\rangle = \sum_{x_1^n \ldots x_m^n} \psi_{x_m^n \ldots x_m^n} &\left(\sum_{\forall i \in [m]\, \xi_i^n \in f_i^{-1}(u_i)} \sqrt{\Delta_{\xi^n_{[m]}}^{(1,u_{[m]})}} \left|x_1^n\right\rangle \left|f_1(x_1^n) \ldots f_m(x_m^n)\right\rangle \left|\xi^n_{[m]}\right\rangle \right) \\ &\otimes \cdots \\ &\otimes \left(\sum_{\forall i \in [m]\, \xi_i^n \in f_i^{-1}(u_i)} \sqrt{\Delta_{\xi^n_{[m]}}^{(m,u_{[m]})}} \left|x_m^n\right\rangle \left|f_1(x_1^n) \ldots f_m(x_m^n)\right\rangle \left|\xi^n_{[m]}\right\rangle \right).\end{aligned}$$

After decoding, by the coherent gentle measurement lemma [34, 92], the state will be $\sqrt{2m\varepsilon}$-close in trace distance to the following state:

$$\begin{aligned}\left|\widehat{\psi}\right\rangle = \sum_{x^n_{[m]}} \psi_{x_1^n \ldots x_m^n} &\left|x_1^n, f_1(x_1^n) \ldots f_m(x_m^n)\right\rangle \left|x^n_{[m]}\right\rangle \\ &\otimes \cdots \otimes \left|x_m^n, f_1(x_1^n) \ldots f_m(x_m^n)\right\rangle \left|x^n_{[m]}\right\rangle.\end{aligned}$$

The details of the application of the coherent gentle measurement lemma are as follows. The coherent gentle measurement lemma ensures that for all parties $j \in [m]$

$$\sum_{\forall i \in [m]\, x_i^n \in f_i^{-1}(u_i)} \sqrt{\Delta_{x^n_{[m]}}^{(j,u_{[m]})}} \left|x_j^n\right\rangle \otimes \left|x^n_{[m]}\right\rangle$$

is $2\sqrt{\varepsilon(2-\varepsilon)}$ close in trace distance to $\left|x_j^n\right\rangle \otimes \left|x^n_{[m]}\right\rangle$ provided that the decoding error is not bigger than ε This implies

$$\begin{aligned}\left\langle \widehat{\psi} | \overline{\psi} \right\rangle &= \sum_{x_1^n \ldots x_m^n} |\psi_{x_1^n \ldots x_m^n}|^2 \left\langle x_1^n\right| \sqrt{\Delta_{x^n_{[m]}}^{(1,u_{[m]})}} \left|x_1^n\right\rangle \cdots \left\langle x_m^n\right| \sqrt{\Delta_{x^n_{[m]}}^{(m,u_{[m]})}} \left|x_m^n\right\rangle \\ &\geq \sum_{x_1^n \ldots x_m^n} |\psi_{x_1^n \ldots x_m^n}|^2 \left\langle x_1^n\right| \Delta_{x^n_{[m]}}^{(1,u_{[m]})} \left|x_1^n\right\rangle \cdots \left\langle x_m^n\right| \Delta_{x^n_{[m]}}^{(m,u_{[m]})} \left|x_m^n\right\rangle \\ &\geq (1-\varepsilon)^m \geq 1 - m\varepsilon.\end{aligned}$$

where the equality follows by substitution, the first inequality follows since $\sqrt{\Delta_{x^n_{[m]}}^{(m,u_{[m]})}} \geq \Delta_{x^n_{[m]}}^{(m,u_{[m]})}$ for $\Delta_{x^n_{[m]}}^{(m,u_{[m]})} \leq \mathbb{1}$ and the second inequality follows from the assumption. Then, for the trace distance of pure states,

$$\begin{aligned}\left\| \widehat{\psi} - \overline{\psi} \right\|_1 &= 2\sqrt{1 - \left|\left\langle \widehat{\psi} | \overline{\psi} \right\rangle\right|^2} \\ &\leq 2\sqrt{1 - (1-\varepsilon)^{2m}} \leq \sqrt{2m\varepsilon}.\end{aligned}$$

All parties now clean up their $U_{[m]}$-registers and their original A_j^n-register by virtue of local unitaries, to arrive at the following state, up to trace norm error $\sqrt{2m\varepsilon}$:

$$|\widehat{\gamma}\rangle = \sum_{x^n_{[m]}} \psi_{x_1^n \dots x_m^n} \left|x^n_{[m]}\right\rangle \cdots \left|x^n_{[m]}\right\rangle \tag{3.10}$$

To do that, note that the partial Slepian-Wolf isometries $V_j : \left|x_j^n\right\rangle |0\rangle^E \mapsto \left|x_j^n\right\rangle \left|f_j(x_j^n)\right\rangle$ can be made a unitary by declaring $\left|x_j^n\right\rangle |i\rangle^E \mapsto \left|x_j^n\right\rangle \left|i + f_j(x_j^n)\right\rangle$, where the addition is that of an abelian group on the ancillary register (e.g. integers modulo $|U_j|$). Once we have a unitary, the inverse is also a unitary, and can be applied locally.

The above state can now be turned into a standard GHZ state at rate $nH(X_{[m]})$ via the well-known entanglement concentration protocol, just like the bipartite case [3]. This involves measuring the *type* t of $x^n_{[m]}$, and noting that the phase and amplitude factors are constant along each type class, resulting in GHZ-type states after the measurement. To see that, let $\mathcal{T}_t^n$ denote the set of sequences of the same type t, and let Π_t be the projector onto the subspace spanned by $\mathcal{T}_t^n$, i.e.

$$\Pi_t = \sum_{x^n \in \mathcal{T}_t^n} |x^n\rangle\langle x^n| .$$

If the type resulting from the measurement does not belong to a typical type, then the protocol ends; with the properties of the type projectors, this happens with asymptotically small probability. Finally, we thus obtain approximately the following state resulting from the type-class measurement (which is close to the initial state)

$$\begin{aligned} \frac{\Pi_t \otimes \cdots \otimes \Pi_t |\widehat{\gamma}\rangle}{\sqrt{p^n(\mathcal{T}_t^n)}} &= \sum_{x^n_{[m]} \in \mathcal{T}_t^n} \sqrt{\widetilde{p}(x^n_{[m]})} \left|x^n_{[m]}\right\rangle \cdots \left|x^n_{[m]}\right\rangle \\ &= \frac{1}{\sqrt{|\mathcal{T}_t^n|}} \sum_{x^n_{[m]} \in \mathcal{T}_t^n} \left|x^n_{[m]}\right\rangle \cdots \left|x^n_{[m]}\right\rangle , \end{aligned}$$

where $p^n(\mathcal{T}_t^n) = |\langle\widehat{\gamma}| \Pi_t \otimes \cdots \otimes \Pi_t |\widehat{\gamma}\rangle|$, $\widetilde{p}(x^n_{[m]}) = \frac{p(x^n_{[m]})}{p^n(\mathcal{T}_t^n)}$ and $|\mathcal{T}_t^n| \sim 2^{nH(X_{[m]})}$ for large n.

The protocol so far proves the following resource inequality:

$$\psi + R_{\mathrm{CO}}[\mathrm{GHZ}] + \infty[c \to c] \geq H(X_{[m]})[\mathrm{GHZ}], \tag{3.11}$$

where R_{CO} is the minimum of the sum of all rates of GHZ states used by parties to communication hash values. By using the Cancellation Lemma 3.2, this implies now

$$\psi + o[\mathrm{GHZ}] + \infty[c \to c] \geq (H(X_{[m]}) - R_{\mathrm{CO}})[\mathrm{GHZ}]. \tag{3.12}$$

In order to remove the o term from the left-hand side of the resource inequality, we need Lemma 3.3, which demands the following resource inequality to be true, for some $\alpha > 0$:

$$\psi + \infty[c \to c] \geq \alpha[\mathrm{GHZ}]. \tag{3.13}$$

Note that we need the asymptotic resource inequality, not some single-copy transformation (which might or might not imply the former), as prerequisite of the cancellation lemma. In Lemma 3.1 we have actually proven this inequality by virtue of entanglement combing. Therefore, we can remove the o term and we have the result as desired. ∎

Theorem 3.4. *Let $|\psi\rangle^{A_1 \ldots A_m}$ be a pure state shared by m spatially separated parties, and let $\mathcal{E}^i : A_i \to A_i \otimes X_i$ denote an instrument of party i, consisting of pure CP maps $\mathcal{E}^i_x(\sigma) = E^i_x \sigma (E^i_x)^\dagger$ (which is why we may assume $A'_i = A_i$). Then, with the notation of Theorem 3.2,*

$$C_{GHZ}(\psi) \geq H(X_{[m]}) - R^{cq}_{CO},$$

where $R^{cq}_{CO} = \min_{R_{[m]} \in \mathcal{R}_{cq}} \sum_{i=1}^m R_i$, and $\mathcal{R}_{cq}$ is the rate region given as follows:

$$\forall j \in [m] \;\; \forall L \subseteq [m] \setminus j \quad \sum_{i \in L} R_i \geq S(X_L | X_{[m] \setminus L} A'_j).$$

Proof. The proof follows from the techniques used in Theorem 3.3, and the result of Theorem 3.2: making the protocol coherent and recycling.

Starting with a pure state, as in the proof of Theorem 3.3, each party applies their instrument coherently on its system, resulting in isometries $V_i : A_i \hookrightarrow A_i \otimes X_i$ defined as $V_i = \sum_{x \in \mathcal{X}_i} E^i_x \otimes |x\rangle$. The isometries act as follows on a single copy:

$$\begin{aligned} \left|\widehat{\psi}\right\rangle &= (V_1 \otimes \cdots \otimes V_m) \, |\psi\rangle^{A_{[m]}} \\ &= \sum_{x_{[m]}} (E^1_{x_1} \otimes \cdots \otimes E^m_{x_m}) \, |\psi\rangle^{A_{[m]}} \otimes \left|x_{[m]}\right\rangle \\ &= \sum_{x_{[m]}} \sqrt{p(x_{[m]})} \left|\widetilde{\psi}_{x_{[m]}}\right\rangle^{A_{[m]}} \otimes \left|x_{[m]}\right\rangle, \end{aligned}$$

where

$$p(x_{[m]}) = \langle\psi| \left(E^1_{x_1} \otimes \cdots \otimes E^m_{x_m}\right)^\dagger \left(E^1_{x_1} \otimes \cdots \otimes E^m_{x_m}\right) |\psi\rangle ,$$

and

$$\left|\widetilde{\psi}_{x_{[m]}}\right\rangle^{A_{[m]}} = \frac{\left(E^1_{x_1} \otimes \cdots \otimes E^m_{x_m}\right) |\psi\rangle^{A_{[m]}}}{\sqrt{p(x_{[m]})}}.$$

With n copies of the initial pure state, we want to distill GHZ states from n copies of $\left|\widehat{\psi}\right\rangle$, i.e.

$$\left|\widehat{\psi}\right\rangle^{\otimes n} = \sum_{x^n_{[m]}} \sqrt{p^n(x^n_{[m]})}\, |x^n_1\rangle \cdots |x^n_m\rangle \otimes \left|\widetilde{\psi}_{x^n_{[m]}}\right\rangle^{A^n_{[m]}},$$

where $\left|\widetilde{\psi}_{x^n_{[m]}}\right\rangle^{A^n_{[m]}}$ is the quantum side information at the disposal of the parties to help them with their decodings.

Similar to Theorem 3.3, in the first step each party coherently computes its hash value and broadcasts it coherently to the other parties via GHZ states. By applying the decoder of Theorem 3.2 in a coherent fashion, each party decodes $\left|x^n_{[m]}\right\rangle$ where the minimum rate of initial GHZ states is R^{cq}_{CO}. After the uncomputing of the hash value information and the local X^n_j, the state is approximately

$$\left|\widehat{\theta}\right\rangle = \sum_{x^n_{[m]}} \sqrt{p(x^n_{[m]})} \left|x^n_{[m]}\right\rangle \cdots \left|x^n_{[m]}\right\rangle \otimes \left|\psi_{x^n_{[m]}}\right\rangle^{A^n_{[m]}}, \tag{3.14}$$

with residual states $\left|\psi_{x^n_{[m]}}\right\rangle$ on $A^n_{[m]}$. At the end, the parties implement the entanglement concentration protocol to get a standard GHZ state. That is, each one measures the joint type t of $x^n_{[m]}$, i.e. they apply the projectors Π_t from the proof of Theorem 3.3. If the result is a non-typical type, they abort the protocol; if it is typical, they proceed as follows to decouple the $A^n_{[m]}$-registers: all sequences $x^n_{[m]}$ from the type class $\mathcal{T}^n_t$ are obtained by a permutation $\pi(x^n_{[m]}) \in S_n$ of a fiducial string $x^n_t \in \mathcal{T}^n_t \subset \mathcal{X}^n_{[m]}$. The unitary $U_{\pi(x^n_{[m]})}$ permuting the n systems of $A^n_{[m]}$ do the same with a fiducial vector $|\psi_t\rangle = \left|\psi_{x^n_t}\right\rangle$, i.e. $\left|\psi_{x^n_{[m]}}\right\rangle = U_{\pi(x^n_{[m]})} |\psi_t\rangle$. Party j now applies the controlled permutation

$$U_j = \sum_{x^n_{[m]} \in \mathcal{T}^n_t} |x^n_{[m]}\rangle\langle x^n_{[m]}| \otimes \left(U_{\pi(x^n_{[m]})}\right)^{\dagger A^n_j},$$

which maps the state to an approximation of

$$\left|\widehat{\theta}\right\rangle = \frac{1}{\sqrt{|T_t^n|}} \sum_{x_{[m]}^n \in \mathcal{T}_t^n} \left|x_{[m]}^n\right\rangle \cdots \left|x_{[m]}^n\right\rangle \otimes |\psi_t\rangle^{A_{[m]}^n}, \tag{3.15}$$

The last part, $|\psi_t\rangle^{A_{[m]}^n}$, is decoupled, as it only depends on t, and the remaining state is the desired GHZ state. ■

Remark 3.4. *The above protocol typically leaves some entanglement behind, in the form of the states $|\psi_t\rangle$. This entanglement could potentially be still useful for m-party GHZ distillation, but a more common situation is that it contains only entanglement between fewer ($\leq m-1$) parties, perhaps even only EPR states between a pair of parties.*

To distill it, essentially the same kind of protocol as in Theorem 3.4 can be applied, because $|\psi_t\rangle = \left|\psi_{x_t^n}\right\rangle$ is a product state across the n m-partite systems, and by grouping identical states we can treat it as a collection of i.i.d. states.

3.4 Conclusion

We have derived two achievability bounds for the distillable common randomness from a mixed multipartite state and by making them coherent, we found two achievability bounds for the rate of GHZ distillation from a multipartite pure state. The first bound reproduces a recent result by Vrana and Christandl with genuinely quantum Shannon theoretic methods, and the second improves on it in a truly quantum way.

To our knowledge, it is the best currently known general bound. Note that it includes the lower bound from [48], which was formulated for a tripartite state ψ^{ABC}, and is obtained by choosing a measurement basis $\{|x\rangle\}$ for A and trivial (identity) instruments for B and C in Theorem 3.4; this gives a pure state decomposition $\psi^{BC} = \sum_x \lambda_x |\psi_x\rangle\langle\psi_x|^{BC}$. Let $\overline{E}_{BC} = \sum_x \lambda_x E(|\psi_x\rangle\langle\psi_x|)$ be the average bipartite entanglement of the pure state decomposition. Define finally

$$\chi = \min\{S(B), S(C)\} - \overline{E}_{BC},$$

Then χ is an achievable rate of three-party GHZ distillation, but in addition also EPR pairs between B and C at rate $\overline{E}_{BC}$ are distilled [48]. This is consistent with our Theorem 3.4 and Remark 3.4, too: following through the proof, the leftover state, there denoted $|\psi_t\rangle$, is precisely a tensor product of $|\psi_x\rangle$, with x appearing $\sim n\lambda_x$ times.

Example 5. *Consider the three-qubit W-state*

$$|W\rangle = \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle).$$

Entanglement combing (Lemma 3.1) results in a GHZ rate of $R_{comb} = \frac{1}{2}H(\frac{2}{3}, \frac{1}{3}) \approx 0.4591$*, but already the very simple yet ingenious protocol of [49] achieves* $R_{FL} = 0.5$*, because it extracts an EPR pair deterministically from every copy of the W-state, albeit randomly distributed over the three possible pairs. Theorem 3.3, applied with the local computational bases, gets up to* $R_{VC} = \log 3 - 1 \approx 0.585$*. Namely, note that the omniscience information* $X_1X_2X_3$ *is jointly uniformly distributed over the set* $\{001, 010, 100\}$*, and so the conditions for communication for omniscience in Theorem 3.3 are* $R_1 \geq H(X_1|X_2X_3) = 0$ *and cyclic, and* $R_1 + R_2 \geq H(X_1X_2|X_3) = \frac{2}{3}$ *and cyclic. Thus,* $R^c_{CO} = \min R_1 + R_2 + R_3 = 3 \cdot \frac{1}{2} \cdot \frac{2}{3} = 1$*.*

The result from [48] (recall that it is a special case of Theorem 3.4) however yields the seemingly very bad $R_{SVW} = \log 3 - \frac{4}{3} \approx 0.2516$*, until we remember that as a bonus we get a rate of* $\frac{2}{3}$ *of EPR states – by the symmetry of the W-state between any prescribed pair of parties,* AB *or* BC *or* AC*. Pairs of these, from different pairs, can be fused to get an additional rate of* $\frac{1}{3}$ *for GHZ generation, thus matching the total of* $R_{VC} = \log 3 - 1$*.*

We do not know, however, if this rate is optimal under general LOCC procedures, or even restricted to non-interactive communication protocols.

Example 6. *Consider the tripartite fully antisymmetric state, also known as "determinant state",*

$$|\alpha_3\rangle = \frac{1}{\sqrt{6}}(|123\rangle + |231\rangle + |312\rangle - |132\rangle - |213\rangle - |321\rangle).$$

Similar to the previous example, we can evaluate the rate resulting from entanglement combing (Lemma 3.1), $R_{comb} = \frac{1}{2}\log 3 \approx 0.7925$*, because all three marginal qutrit states are maximally mixed. But Theorem 3.3, applied with the local computational bases, yields the much better* $R_{VC} = \log 3 - \frac{1}{2} \approx 1.085$*. This is straightforward after realising that the computational bases measurements result in the uniform distribution of* $X_1X_2X_3$ *over all* 6 *permutations* $\{123, 231, 312, 132, 213, 321\}$*. The conditions for communication for omniscience in Theorem 3.3 are* $R_1 \geq H(X_1|X_2X_3) = 0$ *and cyclic, and* $R_1 + R_2 \geq H(X_1X_2|X_3) = 1$ *and cyclic. Thus,* $R^c_{CO} = \min R_1 + R_2 + R_3 = \frac{3}{2}$*.*

The result from [48] gives the seemingly disappointing value $R_{SVW} = \log 3 - 1 \approx 0.585$*; but as before, we can salvage a rate of* 1 *of EPR states*

between any prescribed pair of parties, thus contributing an additional rate of $\frac{1}{2}$ *for GHZ generation, and once again matching the total of* $R_{VC} = \log 3 - \frac{1}{2}$.

Again, we do not know whether this is optimal, in particular whether there is a better way of applying Theorem 3.4.

Example 7. *The* flower state *[93],*

$$|\varphi\rangle = \frac{1}{\sqrt{2d}} \sum_{i=1}^{d} \sum_{j=0}^{1} |ij\rangle^A |ij\rangle^B (H^j |i\rangle)^C,$$

where $H^0 = \mathbb{1}$ *and* H^1 *is the* d*-dimensional quantum Fourier transform, provides an example where Theorem 3.4 is better than Vrana-Christandl protocol. The former, by simply letting* A *or* B *measure and broadcast* j*, so that* C *can undo the unitary* H^j*, yields the clearly optimal* $R_{SVW} = \log d$ *(it is the local entropy of* C*, which is an upper bound on the distillable GHZ rate under arbitrary LOCC protocols).*

On the other hand, Theorem 3.3 with the computational bases for A *and* B *(which seems like the evident choice, but we have no full proof that it is optimal), and* any *measurement of* C*, results in a rate* $R_{VC} \leq \frac{1}{2} \log d$*. This follows from Maassen-Uffink's entropic uncertainty relation [94], which reads as* $I(X_1X_2; X_3) = I(X_1; X_3) \leq \frac{1}{2} \log d$ *(cf. [93]) and some elementary algebraic manipulations. In detail, we have* $H(X_1) = H(X_2) = 1 + \log d$*, and* $H(X_3) \geq \log d$*. On the other hand, the conditions for communication for omniscience in Theorem 3.3 are* $R_1 \geq H(X_1|X_2X_3) = 0$*,* $R_2 \geq H(X_2|X_1X_3) = 0$ *and* $R_3 \geq H(X_3|X_1X_2) \geq \frac{1}{2} \log d$*; furthermore* $R_1 + R_2 \geq H(X_1X_2|X_3) \geq 1 + \frac{1}{2} \log d$*, and the now redundant* $R_1 + R_3 \geq H(X_1X_3|X_2) = H(X_3|X_2) \geq \frac{1}{2} \log d$ *and* $R_2 + R_3 \geq H(X_2X_3|X_1) = H(X_3|X_1) \geq \frac{1}{2} \log d$*. Now for the net rate, we can reason*

$$\begin{aligned} H(X_1X_2X_3) - (R_1 + R_2 + R_3) &\leq H(X_1X_2X_3) - H(X_3|X_1X_2) - H(X_1X_2|X_3) \\ &= H(X_1X_2) - H(X_1X_2|X_3) \\ &= I(X_1X_2; X_3) \leq \frac{1}{2} \log d \end{aligned}$$

using the lower bounds for R_3 *and* $R_1 + R_2$ *in the first line, the chain rule for the entropy in the second line, and finally the entropic uncertainty relation.*

In future work we are going to apply the machinery developed in this chapter to secret key distillation against an adversary who is initially correlated and eavesdrops on the public classical communication between the

parties, and to the distillation of GHZ states from *mixed* initial states. Regarding the former, we can quite evidently apply Theorem 3.2 to a general state $\rho^{A_1\dots A_m E}$ and local instruments $\mathcal{E}_i : A_i \to A_i' X_i$, to first attain omniscience $X_{[m]}$ at all legal parties, and then hashing this information down using privacy amplification [95], resulting in a lower bound

$$C_S(\rho) \geq S(X_{[m]}|E) - R_{\mathrm{CO}}^{cq}$$

on the distillable secret key. Regarding GHZ distillation, we would apply these protocols to a purification $|\psi\rangle^{A_1\dots A_m E}$ of $\rho^{A_1\dots A_m}$, and for pure instruments as in Theorem 3.4 we expect to obtain the lower bound

$$C_{\mathrm{GHZ}}(\rho) \geq S(X_{[m]}|E) - R_{\mathrm{CO}}^{cq}$$

on the distillable GHZ rate. This will require a generalization of the techniques from [41] to the multi-party setting with non-interactive communication, of turning a privacy amplification step into a decoupling procedure.

Furthermore, note that we have focused our attention on non-interactive protocols, but it seems evident that in general there is an advantage in protocols using interactive communication, i.e. of fully general CR distillation, cf. [96, 97]. In this context it is an important question to determine which class of interactive communication protocols, when applied to a quantum state, can be made coherent and thus yields achievable rates for GHZ distillation.

Appendix

Classical correlated source coding with side information at the decoder

The analysis of multi-party common randomness distillation via our omniscience protocol leads quite naturally to the consideration of classical source coding with quantum side information at the decoder [88,89]. Here we present the necessary definitions, and prove a new coding theorem for achieving all points of the rate region directly by random binning and a quantum joint typicality decoder, rather than successive decoding and time sharing as in the cited previous works.

A multipartite correlated classical-quantum (cq-)source is given by a cq-state

$$\rho^{X_1\dots X_k B} = \sum_{x_{[k]}} p(x_{[k]}) |x_1\rangle\langle x_1|^{X_1} \otimes \cdots \otimes |x_k\rangle\langle x_k|^{X_k} \otimes \rho_{x_{[k]}}^B, \tag{3.16}$$

where X_i (which we can identify with a classical random variable) is observed by the i-th encoder, who sends a function of X_i to the decoder. The decoder has the quantum system B and by measuring it, depending on all the messages received from the k encoders, attempts to reconstruct $X_{[k]}$ with high probability.

Definition 5. *An n-block coding scheme with quantum side information at the decoder for the cq-source $\rho^{X_{[k]}B}$ consists of k encoding functions $f_i : \mathcal{X}_i^n \to [M_i]$ and decoding POVMs $\Lambda^{(\mu_{[k]})}$ on B^n, one for each $\mu_{[k]} = \mu_1 \dots \mu_k \in [M_1] \times \cdots \times [M_k]$, and indexed by $\mathcal{X}_1^n \otimes \cdots \otimes \mathcal{X}_k^n$. Its rates are the numbers $\frac{1}{n} \log M_i$, and its average error probability is*

$$P_e := 1 - \sum_{x_{[k]}^n} p^n(x_{[k]}^n) \operatorname{Tr} \rho_{x_{[k]}^n}^{B^n} \Lambda_{x_{[k]}^n}^{\left(f_{[k]}(x_{[k]}^n)\right)}.$$

Here, $f_{[k]}(x_{[k]}^n) = f_1(x_1^n) \dots f_k(x_k^n)$ is the k-tuple of compressed data.

A k-tuple $(R_1, \dots, R_k)$ is called an achievable rate tuple if there exist n-block coding schemes for all n, such that their error probability converges to zero, $P_e \to 0$, and the rates $\frac{1}{n} \log |M_i|$ converge to R_i. The set of achievable rate tuples is called the rate region of the compression problem described by $\rho^{X_{[k]}B}$.

By definition, the rate region is a closed subset of the positive orthant $\mathbb{R}_{\geq 0}^k$, that is closed under increasing individual vector components. By the time-sharing principle, it is also convex. Necessary and sufficient conditions for the rate region were proved in [88, Thm. IV.14 & Cor. IV.16], which are the ones expected from Slepian-Wolf coding:

$$\forall\ I \subseteq [k] \quad \sum_{i \in I} R_i \geq S(X_I | X_{[k] \setminus I} B). \tag{3.17}$$

While the necessity of these conditions is rather straightforward, we will be concerned here with their sufficiency. In the cited PhD thesis, this is obtained by showing that the extreme points of the polytope (3.17) can be achieved, which in turn is done by successive decoding of the j-th sender's information X_j^n, in an order given by the extreme point in question, of which there are $k!$, one for each permutation of the parties $[k]$. The rest follows by the convexity and openness-above of the rate region.

The following lemma shows that it is possible to construct a code by random binning and with a simultaneous decoding scheme that achieves directly every point in the rate region. This is essential in applications, such as ours, where there are multiple decoders with different side informations for the same compressed data.

Lemma 3.4 (Simultaneous quantum decoder)**.** *With the above notation, suppose the rates* $R_i = \frac{1}{n}\log|M_i|$ *satisfy the following inequalities for some* $\delta > 0$*,*

$$\forall\ \varnothing \neq I \subseteq [k] \quad \sum_{i\in I} R_i \geq S(X_I|X_{[k]\setminus I}B) + \delta,$$

where the entropies are with respect to the state (3.16).

Then, for independent 2-universal random functions $F_i : \mathcal{X}_i^n \to [M_i]$*, there exist simultaneous decoding POVMs* $\left(\Lambda_{x_{[k]}^n}^{(\mu_{[k]})}\right)$ *such that the expectation of the average error probability over all codes converges to zero:* $\mathbb{E}_{F_1\dots F_k} P_e \to 0$*, as* $n \to \infty$*.*

Proof. We will use Sen's construction of jointly typical POVM elements [90, Sec. 5], which is stated as Lemma 3.5 below, in the simplified form in which we need it.

Consider $\left(\rho^{X_{[k]}B}\right)^{\otimes n} = \rho^{X_{[k]}^n B^n}$ and for the RVs $X_1^n, \dots, X_k^n$ denote the set of jointly entropy-typical sequences by $\mathcal{T}$. This means that $\Pr\{X_{[k]}^n \in \mathcal{T}\} \geq 1-\eta \to 1$ as $n \to \infty$ and that for every $x_{[k]}^n \in \mathcal{T}$ and all $I \subseteq [k]$,

$$2^{-nH(X_I)-n\beta} \leq p^n(x_I^n) \leq 2^{-nH(X_I)+n\beta},$$

with an arbitrarily chosen $\beta > 0$.

Next we apply Lemma 3.5 to the $(k+1)$-party state $\rho^{X_{[k]}^n B^n}$ to obtain first an "augmented" state $\rho^{X_{[k]}^n B^n} \otimes \tau^{C^n}$ for a suitable system C and a universal state τ^C (actually the maximally mixed state), where we think of BC as a new quantum system $\widetilde{B}$, so that the augmented state is still a $(k+1)$-party cq-state. Note that τ^C can be created locally at B. Lemma 3.5 then gives us an approximation $\widetilde{\rho}^{X_1^n\dots X_k^n\widetilde{B}^n}$ and a POVM element E with the properties stated in the lemma. Importantly, both this state and the POVM element share the original cq-structure:

$$\widetilde{\rho}^{X_1^n\dots X_k^n\widetilde{B}^n} = \sum_{x_{[k]}^n} p^n(x_{[k]}^n)|x_{[k]}^n\rangle\langle x_{[k]}^n|^{X_{[k]}^n} \otimes \widetilde{\rho}_{x_{[k]}^n}^{\widetilde{B}^n},$$
$$E = \sum_{x_{[k]}^n} |x_{[k]}^n\rangle\langle x_{[k]}^n|^{X_{[k]}^n} \otimes E_{x_{[k]}^n}.$$

By restricting the latter to typical $x_{[k]}^n$, we obtain

$$E' := \sum_{x_{[k]}^n\in\mathcal{T}} |x_{[k]}^n\rangle\langle x_{[k]}^n|^{X_{[k]}^n} \otimes E_{x_{[k]}^n},$$

which does not affect property 1 in Lemma 3.5, and preserves property 3, while property 2 becomes the only slightly worse $\operatorname{Tr}\widetilde{\rho}^{X_{[k]}^n\widetilde{B}^n}E' \geq 1-2\gamma-\gamma'-\eta$.

Finally, for the encoding by independent 2-universal functions F_j, after the receiver obtains $\mu_1 \dots \mu_k$, we need a decoding POVM for recovering $x_{[k]}^n \in \mathcal{T} \cap F_1^{-1}(\mu_1) \times \cdots \times F_k^{-1}(\mu_k)$ from $\rho_{x_{[k]}^n}^{B^n} \otimes \tau^{C^n}$. We use the square-root measurement $(\Lambda_{x_{[k]}^n})$ constructed from the $E_{x_{[k]}^n}$, $x_{[k]}^n \in \mathcal{T} \cap F_{[k]}^{-1}(\mu_{[k]})$:

$$\Lambda_{x_{[k]}^n} = \left(\sum_{x'^n_{[k]} \in \mathcal{T} \cap F_{[k]}^{-1}(F_{[k]}(x_{[k]}^n))} E_{x'^n_{[k]}} \right)^{-\frac{1}{2}} E_{x_{[k]}^n} \left(\sum_{x'^n_{[k]} \in \mathcal{T} \cap F_{[k]}^{-1}(F_{[k]}(x_{[k]}^n))} E_{x'^n_{[k]}} \right)^{-\frac{1}{2}}.$$

To upper bound its error probability, we employ the Hayashi-Nagaoka operator inequality, stated as Lemma 1.1:

$$\begin{aligned}
P_e &\leq 1 - p^n(\mathcal{T}) + \sum_{x_{[k]}^n \in \mathcal{T}} p^n(x_{[k]}^n) \operatorname{Tr}(\rho_{x_{[k]}^n}^{B^n} \otimes \tau^{C^n}) \Lambda_{x_{[k]}^n} \\
&\leq \eta + \gamma + \sum_{x_{[k]}^n \in \mathcal{T}} p^n(x_{[k]}^n) \operatorname{Tr} \widetilde{\rho}_{x_{[k]}^n}^{\widetilde{B}^n} \Lambda_{x_{[k]}^n} \\
&\leq \eta + \gamma + \sum_{x_{[k]}^n \in \mathcal{T}} p^n(x_{[k]}^n) \left(2 \operatorname{Tr} \widetilde{\rho}_{x_{[k]}^n}^{\widetilde{B}^n} (\mathbb{1} - E_{x_{[k]}^n}) + 4 \sum_{x'^n_{[k]} \in \mathcal{T} \cap F_{[k]}^{-1}(F_{[k]}(x_{[k]}^n)) \setminus x_{[k]}^n} \operatorname{Tr} \widetilde{\rho}_{x_{[k]}^n}^{\widetilde{B}^n} E_{x'^n_{[k]}} \right) \\
&\leq \eta + 5\gamma + 2\gamma' + 4 \sum_{x'^n_{[k]} \in \mathcal{T}} \operatorname{Tr} E_{x'^n_{[k]}} \left(\sum_{x_{[k]}^n \in \mathcal{T} \cap F_{[k]}^{-1}(F_{[k]}(x'^n_{[k]})) \setminus x'^n_{[k]}} p^n(x_{[k]}^n) \widetilde{\rho}_{x_{[k]}^n} \right),
\end{aligned}$$

where in the first line we declare an error for non-typical $x_{[k]}^n$, and in the second line have used property 1 in Lemma 3.5; in the third line, we used Lemma 1.1, applied to $T = E_{x_{[k]}^n}$ and $S = \sum_{x'^n_{[k]} \in \mathcal{T} \cap F_{[k]}^{-1}(F_{[k]}(x_{[k]}^n)) \setminus x_{[k]}^n} E_{x'^n_{[k]}}$; finally, in the fourth line we use property 2 in Lemma 3.5 for the first term in the bracket, and for the second term simply reorganised the double sum.

Thus, to bound the expected error probability, over the random choice of the F_j, we need a bound on the expected state in the round brackets in the last line of the above chain of inequalities. To do so, we distinguish the different cases of coordinates $\varnothing \neq I \subseteq [k]$ in which $x_{[k]}^n$ and $x'^n_{[k]}$ differ:

$$\begin{aligned}
\mathbb{E}_{F_{[k]}} \left(\sum_{x_{[k]}^n \in \mathcal{T} \cap F_{[k]}^{-1}(F_{[k]}(x'^n_{[k]})) \setminus x'^n_{[k]}} p^n(x_{[k]}^n) \widetilde{\rho}_{x_{[k]}^n} \right) &\leq \sum_{\varnothing \neq I \subseteq [k]} \frac{1}{\prod_{i \in I} M_i} \sum_{\substack{x_{[k]}^n \in \mathcal{T} \\ \text{s.t. } x_{I^c}^n = x'^n_{I^c}}} p^n(x_{[k]}^n) \widetilde{\rho}_{x_{[k]}^n} \\
&=: \sum_{\varnothing \neq I \subseteq [k]} \frac{1}{\prod_{i \in I} M_i} p(x'^n_{I^c}) \widetilde{\rho}_{x'^n_{I^c}},
\end{aligned}$$

with the shorthand notation $I^c = [k] \setminus I$ for the set complement. Furthermore, in the first line we have used the 2-universality of the F_j, as well as their

independence, and in the second line note that the probabilities and states $p(x'^n_{I^c})\widetilde{\rho}_{x'^n_{I^c}}$ appear in the marginal

$$\widetilde{\rho}^{X^n_{I^c}\widetilde{B}^n} = \sum_{x^n_{I^c}} p(x^n_{I^c})|x^n_{I^c}\rangle\langle x^n_{I^c}|^{X^n_{I^c}} \otimes \widetilde{\rho}^{\widetilde{B}^n}_{x^n_{I^c}}.$$

This means that

$$\begin{aligned}
\mathbb{E}_{F_{[k]}} P_e &\leq \eta + 5\gamma + 2\gamma' + 4 \sum_{\varnothing \neq I \subseteq [k]} \frac{1}{\prod_{i\in I} M_i} \sum_{x'^n_{[k]} \in \mathcal{T}} \operatorname{Tr}\, p(x'^n_{I^c})\widetilde{\rho}_{x'^n_{I^c}} E_{x'^n_{[k]}} \\
&\leq \eta + 5\gamma + 2\gamma' + 4 \sum_{\varnothing \neq I \subseteq [k]} \frac{2^{nH(X_I)+n\beta}}{\prod_{i\in I} M_i} \sum_{x'^n_{[k]} \in \mathcal{T}} \operatorname{Tr}\, p(x'^n_I)p(x'^n_{I^c})\widetilde{\rho}_{x'^n_{I^c}} E_{x'^n_{[k]}} \\
&= \eta + 5\gamma + 2\gamma' + 4 \sum_{\varnothing \neq I \subseteq [k]} \frac{2^{nH(X_I)+n\beta}}{\prod_{i\in I} M_i} \operatorname{Tr}\left(\widetilde{\rho}^{X^n_I} \otimes \widetilde{\rho}^{X^n_{I^c}\widetilde{B}^n}\right) E' \\
&\leq \eta + 5\gamma + 2\gamma' + 4 \sum_{\varnothing \neq I \subseteq [k]} \frac{2^{nH(X_I)+n\beta}}{\prod_{i\in I} M_i} 2^{-D^{\varepsilon}_h\left(\rho^{X^n_{[k]}B^n} \middle\| \rho^{X^n_I} \otimes \rho^{X^n_{I^c}B^n}\right)} \\
&\leq \eta + 5\gamma + 2\gamma' + 4 \sum_{\varnothing \neq I \subseteq [k]} \frac{2^{nH(X_I)+n\beta}}{\prod_{i\in I} M_i} 2^{-nI(X_I:X_{I^c}B)+n\beta} \\
&\leq \eta + 5\gamma + 2\gamma' + 4 \sum_{\varnothing \neq I \subseteq [k]} 2^{n(H(X_I|X_{I^c}B)+2\beta-\sum_{i\in I} R_i)},
\end{aligned}$$

where in the second line we use entropy typicality of the $x'^n_{[k]}$; to get the third line simply insert the forms of $\widetilde{\rho}$ and E above; in the fourth line we use property 3 in Lemma 3.5, and in the fifth we invoke the asymptotic equipartition property (AEP) for the hypothesis testing relative entropy, stated in Eq. 1.7.

Hence, choosing $\beta = \delta/3$, we obtain as an upper bound on the expected error probability $\mathbb{E}_{F_{[k]}} P_e \leq \eta + 5\gamma + 2\gamma' + 2^{k+2}2^{-n\delta/3}$, which converges to 0 as $n \to \infty$ (and $\varepsilon \to 0$ sufficiently slowly). ∎

Here follow the technical lemmas from the literature invoked in the proof.

Lemma 3.5 (Sen's jointly typical operators [90, Lemma 1 in Sec. 5, cf. Sec. 1.3])**.** *Let $X_1 \otimes \cdots \otimes X_k \otimes B$ be a $(k+1)$-partite classical-quantum system with finite-dimensional classical system X_i and a finite-dimensional quantum system B, and $\varepsilon > 0$. Then there exists a Hilbert space C and a state τ^C on it such that for any cq-state $\sigma^{X_1\ldots X_k B}$, there is a cq-state $\widetilde{\sigma}^{X_1\ldots X_k\widetilde{B}}$ and a POVM element E (also of cq-form) on $X_1 \ldots X_k\widetilde{B}$, where $\widetilde{B} = B \otimes C$, with the following properties:*

1. $\frac{1}{2}\left\|\widetilde{\sigma}^{X_{[k]}\widetilde{B}} - \sigma^{X_{[k]}B} \otimes \tau^{C}\right\|_1 \leq \gamma$,

2. $\operatorname{Tr}\widetilde{\sigma}^{X_{[k]}\widetilde{B}}E \geq 1 - 2\gamma - \gamma'$,

3. *for all* $\varnothing \neq I \subseteq [k]$, $\operatorname{Tr}\left(\widetilde{\sigma}^{X_I} \otimes \widetilde{\sigma}^{X_{[k]\setminus I}\widetilde{B}}\right)E \leq 2^{-D_h^{\varepsilon}\left(\sigma^{X_{[k]}B}\middle\|\sigma^{X_I}\otimes\sigma^{X_{[k]\setminus I}B}\right)}$.

Here, $\gamma = \sqrt{2}^{k+1}\sqrt[4]{\varepsilon}$ *and* $\gamma' = 2^{k+2^{k+5}}\sqrt{\varepsilon}$. ∎

Using the joint decoder for independent random binning we obtain a new proof for the achievability of the rate region (3.17) for correlated classical source coding with quantum side information at the decoder [88, Thm. IV.14 & Cor. IV.16], which does away with the successive decoding of the different parts of the source. This detail allows the solution of a more demanding problem that was out of reach of the methods in [88], correlated source coding for multiple decoders with quantum side information. Rather than giving the formal definition, let us just indicate the changes to Definition 5: the source is given by a cq-state

$$\rho^{X_{[k]}B_{[q]}} = \sum_{x_{[k]}} p(x_{[k]})|x_1\rangle\langle x_1|^{X_1} \otimes \cdots \otimes |x_k\rangle\langle x_k|^{X_k} \otimes \rho_{x_{[k]}}^{B_{[q]}} \tag{3.18}$$

with q quantum systems $B_1, \ldots, B_q$. A block code for this system is still given by encoding function f_i for each user $i \in [k]$, such that $\mu_i = f_i(x_i^n)$ is broadcast to all q decoders; but now we need a decoding POVM $\Lambda^{(j;\mu_{[k]})}$ on B_j^n for each decoder $j \in [q]$ that satisfy all the decoding error probability criterion for the cq-source $\rho^{X_{[k]}B_j}$. The random binning protocol of Lemma 3.4 then shows that the region

$$\forall\ j \in [q]\ \forall\ I \subseteq [k] \quad \sum_{i\in I} R_i \geq S(X_I|X_{[k]\setminus I}B_j) \tag{3.19}$$

is achievable for rates at which all decoders can successfully decode $X_{[k]}$ simultaneously. That the above conditions are necessary is also evident, so Eq. (3.19) is precisely the rate region.

In [90,98] it was shown that the joint typicality Lemma 3.5 leads to simultaneous, joint-typicality decoders for the classical-quantum multiple access channel (cq-MAC), in fact essentially optimal one-shot bounds. Using a well-known reduction of MAC to Slepian-Wolf, we can also derive the iid rate region from the present result Eq. (3.19), even in the presence of multiple receivers, cf. [99]. Namely, for the k-sender, q-receiver cq-MAC that takes input $x_{[k]} = x_1 \ldots x_k$ to $\rho_{x_{[k]}}^{B_{[q]}}$, and in the simplest case a product distribution $p(x_{[k]}) = p_1(x_1)\cdots p_k(x_k)$, consider the cq-state as in Eq. (3.18).

For block length n and the random code as in Lemma 3.4, consider the bins restricted to the typical sequences, for sender i this is $\mathcal{T}_i$, the sequences typical for the probability distribution p_i, and denote their respective cardinalities by $N_i = 2^{nR'_i}$. Then, we have with high probability that most of the bins are good codes for all decoders and that for most of the bins in turn $\left|R'_i - (H(X_i) - R_i)\right| \leq \frac{1}{k}\delta$, and so $\sum_{i\in I} R'_i \leq \min_j I(X_I : B_j|X_{[k]\setminus I}) - 2\delta$ for all $\varnothing \neq I \subseteq [k]$. For any rate tuple satisfying these constraints there exists thus asymptotically good codes.

To get the full rate region, we also need an auxiliary random variable U such that $X_1, \ldots, X_k$ are independent conditionally on U; then, every tuple of rates R'_i such that

$$\forall\ I \subseteq [k] \quad \sum_{i\in I} R'_i \leq \min_j I(X_I : B_j|X_{[k]\setminus I}U),$$

is asymptotically achievable for transmitting k independent messages from the separate senders to all receivers B_j, $j \in [q]$. The proof is quite similar to the sketch above and is omitted.

Chapter 4

One-shot Capacity bounds on the Simultaneous Transmission of Classical and Quantum Information

We study the communication capabilities of a quantum channel under the most general channel model known as the *one-shot* model. Unlike classical channels that can only be used to transmit classical information (bits), a quantum channel can be used for transmission of classical information, quantum information (qubits) and simultaneous transmission of classical and quantum information. In this work, we investigate the one-shot capabilities of a quantum channel for simultaneously transmitting bits and qubits. This problem was studied in the asymptotic regime for a memoryless channel where a regularized characterization of the capacity region was reported. It is known that the transmission of private classical information is closely related to the problem of quantum information transmission. We resort to this idea and find achievable and converse bounds on the simultaneous transmission of the public and private classical information. Then shifting the classical private rate to the quantum information rate leads to a rate region for simultaneous transmission of classical and quantum information. In the case of asymptotic i.i.d. setting, our one-shot result is evaluated to the known results of the literature. Our main tools used in the achievability proofs are position-based decoding and the convex-split lemma [100], [31].

4.1 Background

Describing a channel by a stochastic map [1], and the classical information theory in general, are not rich enough to take quantum effects into account. This urged quantum information theorists to repeal and replace Shannon's channel model with a *quantum channel* model that is consistent with and encompasses quantum mechanical effects. Many years after Shannon, in the context of quantum information theory, the notion of a quantum channel, a completely-positive trace-preserving (cptp) map with possibly different input and output Hilbert spaces was introduced (for a formal definition of a cptp map see [101]). The ability of a quantum channel to run well on superpositions enables transmission of quantum information and accordingly the highest achievable rate of quantum information is called its capacity for transmission of quantum information. Of course it is always possible to transmit classical information over a quantum channel; this immediately gives rise to the question of what is the tradeoff between the rates of classical and quantum information when their simultaneous transmission is aimed. In the asymptotic regime of many channel uses, this problem was studied by Devetak and Shor [20]. A detailed look into the latter paper reveals that classical information is piggybacked on top of quantum information. The idea of piggybacking information on top of each other was introduced in classical information theory in the context of broadcast channel [87] by using superposition of codewords. We realised that the result of [20] can be derived by considering simultaneous transmission of common and private classical messages using superposition coding [14] and then shifting the private information rate to the quantum information via the well-known quantum capacity theorem [19]. To find a one-shot analogue of this result, we again resorted to the idea of superposition coding, however, this time in the one-shot regime. We see that position-based decoding and convex-split lemma are natural one-shot analogues of packing and the covering lemmas, respectively and can be used to realise superposition coding in the one-shot setting. In the next two subsections, we review some concepts and prior works in the asymptotic and one-shot regimes.

4.1.1 Memoryless and stationary channels: Asymptotic Regime

Perhaps the most direct analogue of the capacity of a classical channel, $C(\mathcal{W})$, is the classical capacity of a quantum channel, $C(\mathcal{N})$, i.e., the highest rate (in bits per use of the channel) at which a sender can transmit classical information faithfully to a remote receiver through a quantum channel with

general quantum inputs and quantum outputs. The classical capacity[1] was independently studied in [102] and [103] where an achievability bound, i.e., $C(\mathcal{N}) \geq \chi(\mathcal{N})$, known as HSW theorem was reported, where $\chi(\mathcal{N})$ is the celebrated Holevo Information [104] defined as follows:

$$\chi(\mathcal{N}) \coloneqq \max_{p(x),\rho} I(X;B)_\rho,$$

where $p(x)$ is a probability distribution, $\rho_{XB} = \sum_x p_X(x)|x\rangle\langle x|_X \otimes \mathcal{N}_{A\to B}(\rho_A^x)$ is a bipartite quantum state and $I(X;B)_\rho$ is the quantum mutual information (see Sec. 1.2). The classical capacity equals the regularized Holevo information, taking a limit over many copies of the channel. So unlike the classical channel, we don't fully know the capabilities of a quantum channel for transmitting classical information.

In certain scenarios, a sender may wish to communicate classical information to a receiver by means of a quantum channel such that the information must remain secret from some third party surrounding the legitimate receiver. This information-processing task gives rise to the notion of *private capacity* of a quantum channel. Cai-Winter-Yeung [18] and Devetak [19] showed that the achievable rates for classical private capacity can be formulated as the difference between the Holevo information of the sender and the legitimate receiver and that of the sender and the eavesdropper(s) as given below:

$$\mathcal{P}(\mathcal{N}) \coloneqq \max_{\rho} \left[I(X;B)_\rho - I(X;E)_\rho \right],$$

where $\rho_{XBE} = \sum_x p_X(x)|x\rangle\langle x|_X \otimes \mathcal{U}^{\mathcal{N}}_{A\to BE}(\rho_A^x)$ and $\mathcal{U}^{\mathcal{N}}_{A\to BE}$ is an isometric extension of the channel $\mathcal{N}_{A\to B}$. They also showed that the private capacity equals the regularized form of $\mathcal{P}(\mathcal{N})$ meaning that this ability of the quantum channel is still not fully understood.

The capacity of a quantum channel to transmit quantum information is called the quantum capacity of the channel and we represent it by $Q_{reg}(\mathcal{N})$. For a given quantum channel, one would like to understand the best rates (in terms of qubits per use of the channel) at which quantum information can be transmitted over the channel. The quantum capacity theorem was first studied in [105] and later in [106]. Subsequently, by taking advantage of the properties of the private classical codes, Devetak [19] showed that the quantum capacity is given by the regularized coherent information of the channel:

$$Q_{reg}(\mathcal{N}) \coloneqq \lim_{k\to\infty} \frac{1}{k} Q(\mathcal{N}^{\otimes k})$$

[1]Hereafter in this chapter, we talk about quantum channels unless otherwise specified, hence we drop the term quantum.

where the coherent information is defined as $Q(\mathcal{N}) \coloneqq \max_{\phi_{RA}} I(R\rangle B)_\sigma$ (see Definition 1.4) and the optimization is with respect to all pure, bipartite states ϕ_{RA} and $\sigma_{RB} = \mathcal{N}_{A\to B}(\phi_{RA})$.

Devetak and Shor [20] unified the classical and quantum capacities and introduced a new information-processing task studying the simultaneously achievable rates for transmission of classical and quantum information over a quantum channel. Since we will follow the results of [20] closely in this chapter, we mention its main theorem:

Theorem 4.1 ([20]). *The capacity region of $\mathcal{N}$ for simultaneous transmission of classical and quantum information is as follows:*

$$S_{reg}(\mathcal{N}) \coloneqq \lim_{k\to\infty} \frac{1}{k} S(\mathcal{N}^{\otimes k}),$$

where $S(\mathcal{N})$ is the union, over all states $\rho_{XRB} = \sum_x p(x)\,|x\rangle\langle x| \otimes \rho^x_{BR}$ arising from the channel $\mathcal{N}_{A\to B}$, i.e., for $x \in \mathrm{supp}(p(x))$, $\rho^x_{RB} = \mathcal{N}(\phi^x_{RA})$ for pure states $|\phi^x\rangle_{RA}$, of the (r, R) pairs obeying

$$\begin{aligned} 0 \le r &\le I(X;B)_\rho, \\ 0 \le R &\le I(R\rangle BX)_\rho. \end{aligned}$$

where r and R are the rates of the classical and quantum[2] information, respectively.

The result of Devetak and Shor is generalized in [107] such that the rate of a secret key that used to achieve noiseless private capacity, enters the trade-off. It is known that the interplay between public classical communication, private classical communication and secret key is rather analogous to how classical communication, quantum communication and entanglement interact with one another. This interaction was studied in [108] from an information-theoretic perspective and the corresponding rate regions for several realistic channels were computed.

4.1.2 General channels: One-shot Regime

All the aforementioned capacities are originally evaluated under the assumptions that the channels were memoryless and stationary and they were available to be used arbitrarily many times. However, in many real-world scenarios, we encounter channels which are neither stationary nor memoryless.

[2]It is the same for various information-processing tasks: subspace transmission, entanglement transmission or entanglement generation.

Therefore, it is of fundamental importance to think of coding schemes for the channels which fail to satisfy these assumptions. The independent channel uses are relaxed in [109] and [33] and general channels with memory are studied in [110] and [111], albeit these results are derived in the form of a limit such that the error probability vanishes as the number of channel uses goes to infinity. Later researchers considered *single-serving* scenarios where a given channel is used only once. This approach gives rise to a high level of generality that no assumptions are made on the structure of the channel and the associated capacity is usually referred to as *one-shot* capacity.

The one-shot capacity of a classical channel was characterized in terms of min- and max-entropies in [112]. The one-shot classical capacity of a quantum channel is addressed by a hypothesis testing approach in [113] and [25], yielding expressions in terms of the generalized (Rényi) relative entropies and a smooth relative entropy quantity, respectively. By taking advantage of two primitive information-theoretic protocols, privacy amplification and information reconciliation, the authors of [114] constructed coding schemes for one-shot transmission of public and private classical information. Their results come in terms of the min- and max-entropies. Two new tools, namely, position-based decoding [31] and the convex-split lemma [115], were employed in [100] where one-shot achievability bounds on the public and private transmission rates were reported (note that prior to this work, one-shot bounds on the public transmission rates on both assisted and unassisted cases were reported in [31] and [25], respectively). Recently, [116] reported tight upper and lower bounds for the one-shot capacity of the wiretap channel. This was done by proving a one-shot version of the quantum covering lemma (see [117]) along with an operator Chernoff bound for non-square matrices. Inner and outer bounds on the one-shot quantum capacity of an arbitrary channel are studied in [26]. The general scenario of [26] leads to the evaluation of the quantum capacity of a channel with arbitrary correlated noise in the repeated uses of the channel.

In this chapter, we aim to study the problem of simultaneous transmission of classical and quantum information over a single use of a quantum channel. In other words, we are interested in the one-shot tradeoff between the number of bits and qubits that are simultaneously achievable. The root of our approach is the well-known quantum capacity theorem via private classcial communication [19]. The basic intuition underlying the quantum capacity is the no-cloning theorem which states that it is impossible to create an identical copy of an arbitrary unknown quantum state. We know well that associated to every quantum channel there is an environment (Eve). If Eve can learn anything about the quantum information that Alice is trying to send to Bob, Bob will not be able to retrieve this information; otherwise the no-cloning

theorem would be violated. Hence, to transmit quantum information, Alice needs to store her quantum information in such subspaces of her input space that Eve does not have access to. By using this idea, Devetak [19] proved that a code for private classical communication can be readily translated into a code for quantum communication. Note that Devetak's proof shows the aforementioned translation in the asymptotic regime; however, one can easily check that the same holds true in the one-shot regime and the proof follows along the same lines. We provide a proof sketch in appendix B. Therefore, if we can come up with a protocol for simultaneously transmitting public and private classical information, we are able to adapt it for the simultaneous transmission of classical and quantum information.

4.1.3 Techniques and Tools

Main tools in our achievability bounds are position-based decoding and the convex-split lemma. Our technique is a simple application of superposition coding in classical information theory (not to be confused with the concept of superposition in the quantum mechanics), along with the convex-split lemma and position-based decoding. In this manner, we significantly differ from the technique of Devetak and Shor [20], whose method was inherently asymptotic i.i.d. and could not have been adapted in the one-shot setting.

We briefly review position-based decoding and the convex-split lemma. Assume Alice and Bob have a way of creating the following state shared between them (in other words, they have this resource at their disposal before any communication takes place):

$$\rho_{XA}^{\otimes|\mathcal{M}|} = \rho_{XA}^{1} \otimes ... \otimes \rho_{XA}^{m} \otimes ... \otimes \rho_{XA}^{|\mathcal{M}|},$$

where Alice possesses A systems and Bob has X systems. Here, the positions of states are denoted by superscripts. Alice wishes to transmit the m-th copy of the state above through the channel $\mathcal{N}_{A\to B}$ to Bob. This induces the following state on Bob's side:

$$\rho_{X^{|\mathcal{M}|}B}^{m} = \rho_{X}^{1} \otimes ... \otimes \rho_{XB}^{m} \otimes ... \otimes \rho_{X}^{|\mathcal{M}|}.$$

If Bob has a means by which he can distinguish between the induced states for different values of m (hypotheses), which happens to be reduced to the problem of distinguishing between states ρ_{XB} and $\rho_X \otimes \rho_B$, he is able to learn about the transmitted message m. Position-based decoding, in fact, relates the communication problem to a problem in binary hypothesis testing. On the other hand, once Alice chooses the m-th system uniformly and sends

it over the channel, the induced state on the receiver side can generally be considered as:

$$\frac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|}\rho_X^1\otimes\ldots\otimes\rho_{XB}^m\otimes\ldots\otimes\rho_X^{|\mathcal{M}|}.$$

The convex-split lemma argues that if the number of systems, $|\mathcal{M}|$, is almost equal to a quantity known as max-mutual information, the induced state is close to the following state

$$\rho_X^1\otimes\ldots\otimes\rho_X^m\otimes\ldots\otimes\rho_X^{|\mathcal{M}|}\otimes\rho_B,$$

meaning that the receiver will not be able to distinguish between the induced states and the product state above, resulting in its ignorance about the chosen message m.

The rest of the chapter is organized as follows. In Section II, we give preliminaries and definitions. A code for simultaneous transmission of public and private information is formally discussed in Section III. This section also includes our main results. Section IV is devoted to the description of the protocol as well as our achievability proof. Converse bounds are proven in section V. In Section VI, we argue how the well-known asymptotic bounds can be quickly recovered by many independent uses of a memoryless channel. We conclude the chapter by a discussion in Section VII.

4.2 Preliminaries

In the following we will define new entropic quantities that the analysis of their asymptotic behaviour requires Fact 1 as well as a useful result in information theory known as the *asymptotic equipartition property* (AEP), and in particular its quantum generalization [118]. Let $X^n = (X_1, X_2, ..., X_n)$ be a sequence of independent and identically distributed (i.i.d.) random variables. The AEP states that for all $0 < \varepsilon < 1$, any $\delta > 0$ and for large enough n, a randomly chosen i.i.d. sequence x^n is with probability more than $1-\varepsilon$ in a *δ-typical set* of sequences that satisfy

$$\left|\frac{1}{n}N(x_i|x^n) - p(x_i)\right| \leq \delta,$$

where $N(x_i|x^n)$ is the number of occurrences of x_i in the sequence x^n. To use these concepts in quantum information, the notion of *typical subspace* is defined. Consider the state $\rho_X = \sum_x p(x)\,|x\rangle\langle x|$. The δ-typical subspace is a subspace of the full Hilbert space $X_1\otimes\ldots\otimes X_n$, associated with many

copies of the density operator, i.e., $\rho_X^{\otimes n} = \sum_{x^n} p(x^n) |x^n\rangle\langle x^n|$, that is spanned by states $|x^n\rangle$ whose corresponding classical sequences are δ-typical. For an introduction to quantum typicality and more on the properties of the typical subspace, we refer the reader to [101].

Definition 6 (Max-mutual information [119]). *For a bipartite state $\rho_{AB} \in \mathcal{S}^{AB}$ and a parameter $\varepsilon \in (0,1)$, from the max-relative entropy (Definition 1.10), the max-mutual information can be defined as follow:*

$$I_{\max}(A;B)_\rho \coloneqq D_{\max}(\rho_{AB} \| \rho_A \otimes \rho_B)_\rho.$$

Definition 7 (Smooth max-mutual information [119]). *For a bipartite state ρ_{AB} and a parameter $\varepsilon \in (0,1)$, from the max-mutual information (Definition 6), we define smooth max-mutual information as follows:*

$$I^\varepsilon_{\max}(A;B)_\rho \coloneqq \inf_{\rho'_{AB} \in \mathcal{B}^\varepsilon(\rho_{AB})} I_{\max}(A;B)_{\rho'}.$$

The following quantity is similar to the smooth max-mutual information.

Definition 8 (Smooth max-mutual information (alternative definition) [100]). *For a bipartite state ρ_{AB} and a parameter $\varepsilon \in (0,1)$, the smooth max-mutual information alternately can be defined as follows:*

$$\tilde{I}^\varepsilon_{\max}(B;A)_\rho \coloneqq \inf_{\rho'_{AB} \in \mathcal{B}^\varepsilon(\rho_{AB})} D_{\max}(\rho'_{AB} \| \rho_A \otimes \rho'_B).$$

In the definition above, the prime happens to be on ρ_B rather than ρ_A, where exactly the same quantity was defined in the introductory section with prime on ρ_A. This has to do with the system that will be separated in the convex split lemma. Since in the journal paper underlying the material of this chapter we had defined the max mutual information as above, we stick to this definition in this chapter only.

Fact 2 (Relation between two definitions of the smooth max-mutual information, [120] and see lemma 2 in [31]). *Let $\varepsilon \in (0,1)$ and $\gamma \in (0,\varepsilon)$. For a bipartite state ρ_{AB}, it holds that:*

$$\tilde{I}^\varepsilon_{\max}(B;A)_\rho \leq I^{\varepsilon-\gamma}_{\max}(A;B)_\rho + \log_2\left(\frac{3}{\gamma^2}\right).$$

Definition 9 (Conditional smooth hypothesis testing-mutual information). *Let $\rho_{ABX} \coloneqq \sum_x p_X(x) |x\rangle\langle x|_X \otimes \rho^x_{AB}$ be a CQ state and $\varepsilon \in [0,1)$. We define*

$$\tilde{I}^\varepsilon_{\mathrm{H}}(A;B|X)_\rho \coloneqq \max_{\rho'} \min_{x \in \mathrm{supp}(\rho'_X)} I^\varepsilon_{\mathrm{H}}(A;B)_{\rho^x_{AB}},$$

where maximization is over all $\rho'_X = \sum_x p'_X(x) |x\rangle\langle x|_X$ satisfying $P(\rho'_X, \rho_X) \leq \varepsilon$.

Definition 10 (Conditional smooth max-mutual information[3]). *Let* $\rho_{ABX} \coloneqq \sum_x p_X(x)|x\rangle\langle x|_X \otimes \rho^x_{AB}$ *be a CQ state and* $\varepsilon \in [0,1)$. *The conditional smooth max-mutual information is defined as follows:*

$$\tilde{I}^{\varepsilon}_{\max}(A;B|X)_\rho \coloneqq \min_{\rho'} \max_{x \in \mathrm{supp}(\rho'_X)} I^{\varepsilon}_{\max}(A;B)_{\rho^x_{AB}},$$

where minimization is over all $\rho'_X = \sum_x p'_X(x)|x\rangle\langle x|_X$ *satisfying* $P(\rho'_X, \rho_X) \leq \varepsilon$.

Lemma 4.1. *Let* $\rho_{XAB} = \sum_x p(x)|x\rangle\langle x|_X \otimes \rho^x_{AB}$. *Then the following holds:*

$$\lim_{n\to\infty} \frac{1}{n} \tilde{I}^{\varepsilon}_{\mathrm{H}}(A^{\otimes n}; B^{\otimes n}|X^n)_{\rho^{\otimes n}} = I(A;B|X)_\rho,$$

Proof. The following is easily seen from the definition,

$$\begin{aligned}&\tilde{I}^{\varepsilon}_{\mathrm{H}}(A^{\otimes n}; B^{\otimes n}|X^n)_{\rho^{\otimes n}} \\ &\qquad \coloneqq \max_{\rho'_{X^n}} \min_{x^n \in supp(\rho'_{X^n})} D^{\varepsilon}_{\mathrm{H}}(\rho^{x^n}_{A^nB^n} \| \rho^{x^n}_{A^n} \otimes \rho^{x^n}_{B^n}).\end{aligned}$$

In order to be able to apply the asymptotic results given in Fact 1, we first produce ρ'_{X^n} by projecting $\rho^{\otimes n}_X$ onto its typical subspace and properly normalize it. We know that the resulting state is close to the initial product state. Conditioned on a particular typical sequence x^n, the state $\rho^{x^n}_{A^nB^n}$ is in fact a tensor-product state that can be written as $\rho^{x(1)}_{AB} \otimes \ldots \otimes \rho^{x(i)}_{AB} \otimes \ldots \otimes \rho^{x(n)}_{AB}$ in which $x(i)$, $i \in [1:n]$ indicates the i-th index in the sequence x^n. From the definition of the typical sequences, we know that for n large enough, each realization x appears almost $np(x)$ times in each sequence. Hence, for any $\delta \geq 0$, as $n \to \infty$, by using Fact 1 for each chosen sequence, the multi-letter formula above can be written as shown by (4.1) where x_i, $i \in [1:|\mathcal{X}|]$ denotes an element of the alphabet $\mathcal{X}$ and the second equality follows from Fact 1 and the fully quantum AEP [30]. ∎

Lemma 4.2. *Let* $\rho_{XAB} = \sum_x p(x)|x\rangle\langle x|_X \otimes \rho^x_{AB}$. *Then the following holds.*

$$\lim_{n\to\infty} \frac{1}{n} \tilde{I}^{\varepsilon}_{\max}(A^{\otimes n}; B^{\otimes n}|X^n)_{\rho^{\otimes n}} = I(A;B|X)_\rho$$

Proof. The proof is very similar to that of Lemma 4.1. It employs the properties of the typical sequences as well as the fully quantum asymptotic equipartition property (AEP) for smooth max-mutual information [30]. ∎

[3]Conditional alternate smooth max-information can be defined in the same way.

$$
\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}D_{\mathrm{H}}^{\varepsilon}(\rho_{AB}^{x^n}\|\rho_{A}^{x^n}\otimes\rho_{B}^{x^n})\\
&=\lim_{n\to\infty}\frac{1}{n}D_{\mathrm{H}}^{\varepsilon}\left(\rho_{AB}^{np(x_1)\pm\delta}\otimes\ldots\otimes\rho_{AB}^{np(x_{|\mathcal{X}|})\pm\delta}\|(\rho_{A}^{x_1}\otimes\rho_{B}^{x_1})^{\otimes np(x_1)\pm\delta}\otimes\ldots\otimes(\rho_{A}^{x_{|\mathcal{X}|}}\otimes\rho_{B}^{x_{|\mathcal{X}|}})^{\otimes np(x_{|\mathcal{X}|})\pm\delta}\right)\\
&=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{|\mathcal{X}|}\big(np(x_i)\pm\delta\big)D(\rho_{AB}^{x_i}\|\rho_{A}^{x_i}\otimes\rho_{B}^{x_i})=\sum_{x=1}^{|\mathcal{X}|}p(x)D(\rho_{AB}^{x}\|\rho_{A}^{x}\otimes\rho_{B}^{x})\coloneqq I(A;B|X)_{\rho}.
\end{aligned}
\tag{4.1}
$$

Lemma 4.3. *For a cq-state $\rho_{XAB}=\sum_x p(x)|x\rangle\langle x|_X\otimes\rho_{AB}^x$, the following inequality is true.*

$$
\tilde{I}_{\mathrm{H}}^{\varepsilon}(A;B|X)_{\rho}\leq\frac{1}{1-\varepsilon}\left(I(A;B|X)+h_b(\varepsilon)\right)_{\rho}.
$$

Proof. Considering the definition of the conditional hypothesis testing-mutual information and the fact that

$$
\min_x D_{\mathrm{H}}^{\varepsilon}(\rho_{AB}^{x}\|\rho_{A}^{x}\otimes\rho_{B}^{x})\leq\sum_x p(x)D_{\mathrm{H}}^{\varepsilon}(\rho_{AB}^{x}\|\rho_{A}^{x}\otimes\rho_{B}^{x}),
$$

and also from Fact 1.6 for all x, we have:

$$
D_{\mathrm{H}}^{\varepsilon}(\rho_{AB}^{x}\|\rho_{A}^{x}\otimes\rho_{B}^{x})\leq\frac{1}{1-\varepsilon}\left(D(\rho_{AB}^{x}\|\rho_{A}^{x}\otimes\rho_{B}^{x})+h_b(\varepsilon)\right),
$$

by plugging into the the aforementioned inequality, we can get the result. We note than in order for the above to be true, we should have $\rho'_X\subseteq\rho_X$. However, in case ρ'_X goes beyond the support of ρ_X, it can be projected onto the support of ρ_X. Since $P(\rho'_X,\rho_X)\leq\varepsilon$, from the monotonicity of the purified distance, it can be seen that the state after being projected will remain ε-close to the initial state. ∎

Lemma 4.4. *Let $\rho_{XAB}=\sum_x p(x)|x\rangle\langle x|_X\otimes\rho_{AB}^x$. The following inequality holds.*

$$
\begin{aligned}
\tilde{I}_{\max}^{\varepsilon}(A;B|X)_{\rho}\geq\ & I(A;B|X)_{\rho}\\
&-2\varepsilon\log|A|-2(1+\varepsilon)h_b\left(\frac{\varepsilon}{1+\varepsilon}\right).
\end{aligned}
$$

Proof. In the the following simple inequality:

$$
\max_x D_{\max}^{\varepsilon}(\rho_{AB}^{x}\|\rho_{A}^{x}\otimes\rho_{B}^{x})\geq\sum_x p(x)D_{\max}^{\varepsilon}(\rho_{AB}^{x}\|\rho_{A}^{x}\otimes\rho_{B}^{x}),
\tag{4.2}
$$

we have to deal with $D^{\varepsilon}_{\max}(\rho^x_{AB}\|\rho^x_A\otimes\rho^x_B)$ and try to bound it from below. Let $\bar{\rho}^x_{AB}$ be the state achieving the minimum in the definition of $D^{\varepsilon}_{\max}(\rho^x_{AB}\|\rho^x_A\otimes\rho^x_B)$, hence

$$D^{\varepsilon}_{\max}(\rho^x_{AB}\|\rho^x_A\otimes\rho^x_B)\geq D_{\max}(\bar{\rho}^x_{AB}\|\bar{\rho}^x_A\otimes\bar{\rho}^x_B)$$

where $P(\rho^x_{AB},\bar{\rho}^x_{AB})\leq\varepsilon$. From Fact 1.11 we further know that $D_{\max}(\bar{\rho}^x_{AB}\|\bar{\rho}^x_A\otimes\bar{\rho}^x_B)\geq D(\bar{\rho}^x_{AB}\|\bar{\rho}^x_A\otimes\bar{\rho}^x_B)$. Now we deploy Alicki-Fannes-Winter (AFW) inequality [121] (an improvement over [122]) for the quantum mutual information saying that: (from the relation between the purified and trace distances, we know that $\frac{1}{2}\|\rho^x_{AB}-\bar{\rho}^x_{AB}\|\leq\varepsilon$)

$$\begin{aligned}D(\bar{\rho}^x_{AB}\|\bar{\rho}^x_A\otimes\bar{\rho}^x_B)\geq D(\rho^x_{AB}\|\rho^x_A\otimes\rho^x_B)\\ -2\varepsilon\log|A|-2(1+\varepsilon)h_2(\frac{\varepsilon}{1+\varepsilon}).\end{aligned}$$

Therefore,

$$\begin{aligned}D^{\varepsilon}_{\max}(\rho^x_{AB}\|\rho^x_A\otimes\rho^x_B)\geq D(\rho^x_{AB}\|\rho^x_A\otimes\rho^x_B)\\ -2\varepsilon\log|A|-2(1+\varepsilon)h_2(\frac{\varepsilon}{1+\varepsilon}),\end{aligned}$$

and plugging back into the right-hand side of (4.2), we well get the desired result. ∎

Lemma 4.5 (Convex-split lemma [115])**.** *Fix $\varepsilon\in(0,1)$ and $\delta\in(0,\varepsilon)$. Let $\rho_{AB}\in\mathcal{S}^{AB}$ and define the state $\tau_{A_1\ldots A_KB}$ as follows:*

$$\begin{aligned}&\tau_{A_1\ldots A_{|\mathcal{K}|}B}\\&=\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\rho_{A_1}\otimes\ldots\otimes\rho_{A_{k-1}}\otimes\rho_{A_kB}\otimes\rho_{A_{k+1}}\otimes\ldots\otimes\rho_{A_{|\mathcal{K}|}}.\end{aligned}$$

If

$$\log_2|\mathcal{K}|\geq\tilde{I}^{\sqrt{\varepsilon}-\delta}_{\max}(B;A)_\rho+2\log_2\left(\frac{1}{\delta}\right),$$

then

$$P(\tau_{A_1\ldots A_{|\mathcal{K}|}B},\rho_{A_1}\otimes\ldots\otimes\rho_{A_k}\otimes\ldots\otimes\rho_{A_{|\mathcal{K}|}}\otimes\tilde{\rho}_B)\leq\sqrt{\varepsilon},$$

where $\tilde{\rho}_B$ is the marginal of some state $\tilde{\rho}_{AB}\in\mathcal{B}^{\sqrt{\varepsilon}-\delta}(\rho_{AB})$. The above smooth version of convex-split lemma is taken from [100], which improved the error parameters in the smooth version given in [31].

We note that here we consider quantum communication channels with quantum input and outputs. One may consider channels with classical inputs and quantum outputs, i.e., CQ channels. In this case, an encoder has to be prepended to the CQ channel such that it associates a particular input state to every classical input.

4.3 Problem Statement And Main Results

In this section, we first define a simultaneous public-private one-shot code, then we present our main results. Latter, we discuss the translation of the public-private code to a classical-quantum code. Two classical messages $(m, \ell) \in \mathcal{M} \times \mathcal{L}$ are to be transmitted from a sender to a receiver in the presence of an eavesdropper by using a quantum channel only once, i.e., one-shot communication is considered. The sender Alice, wishes to reliably communicate a public message m and (simultaneously) a private message ℓ to the legitimate receiver Bob such that ℓ must not be leaked to the eavesdropper Eve. The quantum (wiretap) channel to be used by three parties is denoted by $\mathcal{N}_{A\to BE}$ and it takes quantum states from A to $B \otimes E$ where Alice is assumed to control the input system A and systems B and E are outputs received by Bob and Eve, respectively. Let M and L be the random variables[4] corresponding to Alice's choices of the public and private messages, respectively[5]. We formally define a one-shot simultaneous public-private code in the following.

Definition 11. *Fix ε, $\varepsilon' \in (0,1)$ and let r and R be the rates of the public and private messages, respectively (i.e., $|\mathcal{M}| = 2^r$ and $|\mathcal{L}| = 2^R$). A one-shot $(r, R, \varepsilon, \varepsilon')$- simultaneous public-private code for the channel $\mathcal{N}_{A\to BE}$ consists of*

- *An encoding operation by Alice $\mathcal{E} : ML \to \mathcal{S}^A$ such that*

$$\forall m \in \mathcal{M}, \qquad \frac{1}{2}\|\rho^m_{LE} - \rho_L \otimes \tilde{\rho}^m_E\|_1 \leq \varepsilon', \tag{4.3}$$

  where for each message m, ρ^m_{LE} and ρ_L are appropriate marginals of the state $\rho^m_{LBE} = \frac{1}{|\mathcal{L}|}\sum_{\ell=1}^{|\mathcal{L}|} |\ell\rangle\langle\ell| \otimes \mathcal{N}(\mathcal{E}(m,\ell))$ and $\tilde{\rho}^m_E$ can be any arbitrary state.

- *A decoding operation by Bob $\mathcal{D} : \mathcal{S}^B \to \hat{M}\hat{L}$ such that*

$$Pr\left((\hat{M}, \hat{L}) \neq (M, L)\right) \leq \varepsilon, \tag{4.4}$$

  where $\hat{M}$ and $\hat{L}$ denote the estimates of the public and private messages, respectively.

[4] M and L basically are registers which hold the public and private messages, respectively. Here with slightly abuse of notation, we refer to them as random variables to which, corresponding classical states can be tied.

[5] In the literature, for example [14], the public and private messages are referred to as the common and confidential messages, respectively. If Eve were to receive the common message, it could have been considered without jeopardizing the confidential message. Indeed, as we will see, the secrecy analysis is guaranteed assuming Eve has detected the common (or the public) message.

A rate pair (r, R) is said to be $(\varepsilon, \varepsilon')$-achievable if there exist encoding and decoding maps $(\mathcal{E}, \mathcal{D})$ such that (4.3) and (4.4) are fulfilled. For a given $(\varepsilon, \varepsilon')$, the one-shot capacity region for the simultaneous transmission of public and private information of the channel $\mathcal{N}$, $\mathcal{C}^{\varepsilon,\varepsilon'}(\mathcal{N})$, is the closure of all achievable rate pairs in a $(r, R, \varepsilon, \varepsilon')$ coding scheme. In this work, our aim is to find upper and lower bounds on $\mathcal{C}^{\varepsilon,\varepsilon'}(\mathcal{N})$.

In the following, we first have Theorem 4.2 that establishes a lower bound on $\mathcal{C}^{\varepsilon,\varepsilon'}(\mathcal{N})$ referred to as achievability and then Theorem 4.3 that states an upper bound on $\mathcal{C}^{\varepsilon,\varepsilon'}(\mathcal{N})$, i.e., the converse. This section ends with a discussion about the translation of the private classical capacity to the quantum capacity in one-shot regime.

Theorem 4.2 (Achievability)**.** *For any fixed $\varepsilon \in (0,1), \varepsilon' \in (0,1)$, and δ, δ' such that $\delta \in (0, \varepsilon), \delta' \in (0, \varepsilon')$, there exists a one-shot $(r, R, 3\varepsilon+2\sqrt{\varepsilon}+\sqrt{\varepsilon'}, 2(\varepsilon+\sqrt{\varepsilon})+\sqrt{\varepsilon'})$ code for the channel $\mathcal{N}_{A\to BE}$ if the twin (r, R) satisfies the following bounds:*

$$\begin{aligned} r &\leq I_{\mathrm{H}}^{\varepsilon-\delta}(X;B)_\rho - \log_2(\frac{4\varepsilon}{\delta^2}), \\ R &\leq \tilde{I}_{\mathrm{H}}^{\varepsilon-\delta}(Y;B|X)_\rho - \tilde{I}_{\max}^{\sqrt{\varepsilon'}-\delta'}(Y;E|X)_\rho \\ &\qquad - \log_2(\frac{4\varepsilon}{\delta^2}) - 2\log_2(\frac{1}{\delta'}), \end{aligned}$$

for some quantum state ρ arising from the channel. We call the region above $\mathcal{C}_a(\mathcal{N})$, therefore, we have

$$\mathcal{C}_a(\mathcal{N}) \subseteq \mathcal{C}^{3\varepsilon+2\sqrt{\varepsilon}+\sqrt{\varepsilon'}, 2(\varepsilon+\sqrt{\varepsilon})+\sqrt{\varepsilon'}}(\mathcal{N}).$$

Theorem 4.3 (Converse)**.** *For any fixed $\varepsilon \in (0,1), \varepsilon' \in (0,1)$, every one-shot $(r, R, \varepsilon, \varepsilon')$ public-private code for the channel $\mathcal{N}_{A\to BE}$, must satisfy the following inequalities:*

$$\begin{aligned} r &\leq I_{\mathrm{H}}^{\varepsilon}(X;B)_\rho, \\ R &\leq \tilde{I}_{\mathrm{H}}^{\sqrt{\varepsilon}}(Y;B|X)_\rho - \tilde{I}_{\max}^{\sqrt{2\varepsilon'}}(Y;E|X)_\rho, \end{aligned}$$

for some state $\rho_{XYBE} = \sum_{x,y} p(x,y)|x\rangle\langle x| \otimes |y\rangle\langle y| \otimes \rho_{BE}^{x,y}$. We refer to this region as $\mathcal{C}_c(\mathcal{N})$. In fact, we have $\mathcal{C}^{\varepsilon,\varepsilon'}(\mathcal{N}) \subseteq \mathcal{C}_c(\mathcal{N})$.

Once there is a code for simultaneous transmission of public and private classical information, this code can be translated into a coherent code that is capable of transmitting classical and quantum information simultaneously. In other words, the rate pair (public classical, private classical) can be shifted to

the rate pair (public classical, quantum) (or simply (classical, quantum)). We can then translate our one-shot (public, private) code to a one-shot (classcial, quantum) code. Note that the proof is implicit in findings of Devetak [19] such that one can mimic his procedure to see the result in one-shot setting. Henceforth, we have a one-shot code for simultaneous transmission of classical and quantum information.

By evaluating the asymptotic behaviour of the rate region given by Theorem 4.2 and Theorem 4.3 (Section VI), we recover Theorem 1 of [20], the well-known result of Devetak and Shor, as a corollary.

4.4 Achievability

We consider a general quantum channel which is prepended by an encoder (modulator) that associates a particular input state to every classical input pair. In this sense, Alice can be thought of as being in possession of an ensemble $\{p_{X,Y}(x,y), \omega_A^{x,y}\}$ such that the input distribution $p(x,y)$ and the encoder need to be optimized over to get our capacity results. In our protocol, Bob runs two successive decodings, his first decoder has $|\mathcal{M}|$ possible classical outputs as well as a post-measurement quantum state. His second decoder takes the resulted states of the first decoder and its output is a classical system of dimension $|\mathcal{L}|$. Before we get into achievability proof, we describe our protocol.

4.4.1 Protocol description

Fix a joint probability distribution $p_{X,Y}(x,y)$ over the finite alphabets $\{\mathcal{X} \times \mathcal{Y}\}$, $\varepsilon, \varepsilon' \in (0,1)$, $\delta \in (0,\varepsilon)$, $\delta' \in (0,\sqrt{\varepsilon'})$ and $\rho_{XYBE} = \sum_{x,y} p(x,y)|x\rangle\langle x| \otimes |y\rangle\langle y| \otimes \rho_{BE}^{x,y}$. Let

$$
\begin{aligned}
r &\leq I_{\mathrm{H}}^{\varepsilon-\delta}(X;B)_\rho - \log_2(\frac{4\varepsilon}{\delta^2}),\\
R + \tilde{R} &\leq \tilde{I}_{\mathrm{H}}^{\varepsilon-\delta}(Y;B|X)_\rho - \log_2(\frac{4\varepsilon}{\delta^2}),\\
\tilde{R} &\geq \tilde{I}_{\max}^{\sqrt{\varepsilon'}-\delta'}(E;Y|X)_\rho + 2\log_2(\frac{1}{\delta'}).
\end{aligned}
$$

We choose $|\mathcal{M}| = 2^r$, $|\mathcal{L}| = 2^R$ and $|\mathcal{K}| = 2^{\tilde{R}}$ implying that r and R denote our public and private rates, respectively and $|\mathcal{K}|$ stands for the size of a local key, a uniformly distributed random variable K, used by Alice for obfuscation purpose. Let the sender Alice, legitimate receiver Bob and Eve be connected by means of a quantum (wiretap) channel $\mathcal{N}^{A\to BE}$.

$$\rho_{XX'(AYY')^{\otimes|\mathcal{L}||\mathcal{K}|}}^{\otimes|\mathcal{M}|}$$
$$\coloneqq \left(\sum_x p(x)|x\rangle\langle x|_X \otimes |x\rangle\langle x|_{X'} \otimes \left(\sum_y p(y|x)|y\rangle\langle y|_Y \otimes |y\rangle\langle y|_{Y'} \otimes \rho_A^{x,y}\right)^{\otimes|\mathcal{L}||\mathcal{K}|}\right)^{\otimes|\mathcal{M}|}. \tag{4.5}$$

Alice wants to convey to Bob, in a single use of a quantum channel, a classical message $m \in \mathcal{M}$ and simultaneously, a private classical message $\ell \in \mathcal{L}$ where both messages are uniformly distributed on their corresponding sets. The message m is public, meaning that Bob has to be able to decode it correctly with small probability of error. On the other hand, message ℓ is private and while Bob has to receive it with negligible error probability, it must be kept secret from Eve. We clarify that our definition of public and private messages is the same as in [20] and these correspond respectively to common and confidential messages defined in [14]. The position-based decoding is employed in order to accomplish this information-processing task, therefore before communication begins, Alice, Bob and Eve share the state given in (4.5), where Alice controls the system A, Bob has systems (X, Y) and Eve is in possession of (X', Y') systems. Our coding scheme is, in spirit, inferred from the well-known superposition coding in classical information theory [15]. We can think of the state (4.5) as the superposition of two states, each of which is use to accomplish a certain part of the task. There are $|\mathcal{M}|$ bins in the first place, inside each of them, there are $|\mathcal{L}||\mathcal{K}|$ states that are divided into $|\mathcal{L}|$ bins, again inside each one there are $|\mathcal{K}|$ states.

Upon receiving the message pair (m, ℓ), Alice goes to the m-th copy of $\rho_{XX'(AYY')^{\otimes|\mathcal{L}||\mathcal{K}|}}^{\otimes|\mathcal{M}|}$. There she runs the protocol for the private capacity, by considering $|\mathcal{L}||\mathcal{K}|$ copies and choosing a system A uniformly at random from the ℓ-th bin. Upon receiving B, Bob performs a position-based decoding to obtain the public message m (and hence the correct copy of $\rho_{XX'(AYY')^{\otimes|\mathcal{L}||\mathcal{K}|}}$). The choice of the rate for public message r ensures that this is possible and gentle measurement lemma ensures that the quantum state of the correct copy of $\rho_{XX'(AYY')^{\otimes|\mathcal{L}||\mathcal{K}|}}$ is almost unchanged after Bob's decoding.

To decode ℓ, Bob performs another position-based decoding conditioned on X, meaning that having found the correct copy of $\rho_{XX'(AYY')^{\otimes|\mathcal{L}||\mathcal{K}|}}$ used in the transmission, Bob applies a decoder that depends on X, and it works for all $x \in \mathcal{X}$. For this strategy, Bob first appeals to the definition of the conditional smooth hypothesis testing-mutual information, to assume that the distribution over X was $p'(x)$ (achieving the infimum in the definition)

with negligible error. Then for $x \in \text{supp}(\rho_{X'})$, he performs position-based decoding. The choice of $R + \tilde{R}$ guarantees the successful decoding for every x and at the same time, the security criterion is ensured from the fact that even if Eve is aware of the correct copy of $\rho_{XX'(AYY')^{\otimes|\mathcal{L}\|\mathcal{K}|}}$, the condition that convex-split lemma imposes on $|\mathcal{K}|$, gives her very small information about ℓ for every $x \in \text{supp}(\rho_{X'})$ (where here $\rho_{X'} = \sum_x p_{X'}(x)|x\rangle\langle x|_{X'}$ and $p_{X'}(x)$ is the distribution achieving the infimum in the alternate definition of conditional smooth max-mutual information). Now we can derandomize the protocol by fixing the values in corresponding systems. Upon derandomization, the code is publicly available.

Before we proceed to the error analysis of the direct part, we make the following remarks. The state that is fed into the second decoder differs from the original state although negligibly, this adds to the error probability of the private message. Moreover, since successive cancellation decoding is being performed, in the event of a failure of the first decoder, the second decoder will fail as well. We also take the contribution of this event into account. Moreover, note that there is just one decoding map in general, Bob's (two) separate decodings are just a property of our protocol.

4.4.2 Achievability Proof

As is learned in the preceding subsection, we start with a randomness assisted protocol and derandomize it later. We get started on our proof by introducing the encoder and the decoders. We then analyze the average error probability of the public message. Likewise, we inspect the second decoder and analyze the average error probability of the private message. Finally, we study the secrecy requirement.

In the achievability part of our randomness assisted code, for the private message, we stick to a single criterion known as *privacy error* introduced in [123], [124] and [100]. The general idea is to merge the secrecy of the private message (4.3) as well as its error probability (4.4) into one single criterion. While this idea was used in [123] and [124] in understanding upper bounds for private communication protocols, it had not been used in an achievability proof prior to [100]. We should note that the main advantage of dealing with single criterion reveals when the code is to be derandomized. Our procedure is that we analyze the error probability of Bob in detecting the private message separately from keeping Eve ignorant. This leads to two separate criteria and then the separate criteria are merged into one single criterion. It is clear that if the joined criterion is satisfied, each of the single criteria is also fulfilled. After we prove the correctness of these criteria for the randomness assisted code, we immediately proceed to derandomize the code

$$\rho^{m,(\ell,k)}_{XX'(YY')^{\otimes|\mathcal{L}||\mathcal{K}|}(A)^{\otimes|\mathcal{L}||\mathcal{K}|-1}BE} \coloneqq \sum_x p_X(x)|x\rangle\langle x|_X \otimes |x\rangle\langle x|_{X'} \otimes$$
$$\left\{\rho^{x,m,(1,1)}_{YY'A} \otimes \ldots \otimes \rho^{x,m,(\ell,k-1)}_{YY'A} \otimes \mathcal{N}_{A\to BE}\left(\rho^{x,m,(\ell,k)}_{YY'A}\right) \otimes \rho^{x,m,(\ell,k+1)}_{YY'A} \ldots \otimes \rho^{x,m,(|\mathcal{L}|,|\mathcal{K}|)}_{YY'A}\right\}. \tag{4.7}$$

---

in the succeeding step that the unassisted criteria set out by Definition 11 can be inferred. The derandomization involves some procedures that appeared in [100] and [125].

Alice, Bob and Eve are allowed to share some quantum state among themselves. Moreover, Alice has access to a source of uniform dummy randomness given in random variable K. Further, let $\tilde{R} = \log_2|\mathcal{K}|$. The state initially shared between three parties is given by equation (4.5), where Alice possesses the quantum systems A, Bob possesses the classical systems (X, Y) and Eve has the classical systems (X', Y'). For ease of notation, we further define $\Upsilon_{T_X T_{X'} T_A T_Y T_{Y'}} \coloneqq \rho^{\otimes|\mathcal{M}|}_{XX'(AYY')^{\otimes|\mathcal{L}||\mathcal{K}|}}$ with it being clear that for example $\Upsilon_{T_A} = \rho^{\otimes|\mathcal{M}|}_{A^{\otimes|\mathcal{L}||\mathcal{K}|}}$[6].

The encoding and decoding pairs are as follows:

- Alice performs some encoding operation $\mathcal{E} : MLA \to A$. Let us denote the state in (4.5) after channel transmission as:

$$\left(\rho_{XX'(AYY')^{\otimes|\mathcal{L}||\mathcal{K}|}}\right)^{\otimes|\mathcal{M}|-1} \otimes \rho^{m,(\ell,k)}_{XX'(YY')^{\otimes|\mathcal{L}||\mathcal{K}|}(A)^{\otimes|\mathcal{L}||\mathcal{K}|-1}BE}, \tag{4.6}$$

  where $(m, \ell, k) \in [1 : 2^r] \times [1 : 2^R] \times [1 : 2^{\tilde{R}}]$ are the public message, the private message and a dummy index drawn uniformly at random by the encoder and $\rho^{m,(\ell,k)}_{XX'(YY')^{\otimes|\mathcal{L}||\mathcal{K}|}(A)^{\otimes|\mathcal{L}||\mathcal{K}|-1}BE}$ is given by equation (4.7).

- After the channel action, Bob performs a decoding operation (quantum instrument) $\mathcal{D}^1 : BX \to \hat{M}B$ on his Υ_{T_X} systems as well as the received system, whose outputs are a classical system $\hat{M}$ and a quantum system in $\mathcal{S}^B$ (the decoder will be defined formally later, see (4.13)). The action of the quantum decoder $\mathcal{D}^1_{BX\to\hat{M}B}$ on Bob's corresponding systems is

[6]Due to the cumbersome notations we face, the tensor product states are shown for example as either $\rho^{\otimes|\mathcal{M}|}_X$ or $\rho_{X^{\otimes|\mathcal{M}|}}$.

as follows:

$$\mathcal{D}^1_{BX\to\hat{M}B}(\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}) \coloneqq \sum_{m'=1}^{|\mathcal{M}|} |m'\rangle\langle m'|_{\hat{M}} \otimes \mathcal{D}^{1,m'}_{BX\to B}(\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}), \tag{4.8}$$

where $\{|m\rangle\}_{m=1}^{|\mathcal{M}|}$ is some orthonormal basis and $\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}$ can be seen from (4.6) by tracing out uninvolved systems. Moreover, tracing out the classical system $\hat{M}$ gives the induced quantum operation $\mathcal{D}^1_{BX\to B} = \sum_m \mathcal{D}^{1,m}_{BX\to B}$ such that its sum is trace preserving, i.e.

$$\mathrm{Tr}\left\{\sum_{m'=1}^{|\mathcal{M}|} \mathcal{D}^{1,m'}_{BX\to B}(\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B})\right\} = 1.$$

Let $\sigma^{m,(\ell,k)}_{XX'(YY')^{\otimes|\mathcal{L}||\mathcal{K}|}BE}$ denote the *disturbed* state after Bob applied his first decoder (this state will be defined formally later, see (4.19)).

- Bob's second decoder is another quantum map $\mathcal{D}^2 : \hat{M}BY \to \hat{L}$ which is input the classical output of the first decoder, the disturbed quantum output, Bob's Υ_{T_Y} systems and outputs a classical system $\hat{L}$.

$$\mathcal{D}^2_{\hat{M}BY\to\hat{L}}(\sigma^{m,(\ell,k)}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}B}) \coloneqq \sum_{\ell=1}^{|\mathcal{L}|} p_{\hat{L}}(\ell)|\ell\rangle\langle\ell|_{\hat{L}}, \tag{4.9}$$

where $\{|\ell\rangle\}_{\ell=1}^{|\mathcal{L}|}$ is some orthonormal basis and $\sigma^{m,(\ell,k)}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}B}$ comes about by tracing out uninvolved systems in the disturbed state.

Having defined the decoders, it is seen that the phrase in (4.10) indicates the probability of an erroneous detection of the public message, while the expression in (4.11) captures the notions of an erroneous detection of the private message as well as the secrecy condition of the eavesdropper (the latter is clarified below). After we derandomize the code, we see that the criteria mentioned in Definition 11 can be set out from these criteria by using the monotonicity of the trace distance and properly adjusting the constants.

Correctness of Public Message: Eq. (4.10)

All systems are assumed to be traced out except those used by Bob's first decoder (we could have considered multiplying those systems by identity operator as well). To decode the public message m, Bob employs the following

$$P_e = \{\hat{M} \neq M\} \coloneqq \frac{1}{M}\sum_{m=1}^{|\mathcal{M}|}\frac{1}{2}\left\|\mathcal{D}^1_{BX\to\hat{M}}(\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}) - |m\rangle\langle m|_{\hat{M}}\right\|_1 \leq \varepsilon, \tag{4.10}$$

$$P_{priv} \coloneqq \tag{4.11}$$

$$\frac{1}{|\mathcal{L}|}\sum_{l=1}^{|\mathcal{L}|}\frac{1}{2}\left\|\mathcal{D}^2_{\hat{M}BY\to\hat{L}}(\sigma^{m,(\ell,k)}_{XX'(YY')^{\otimes|\mathcal{L}||\mathcal{K}|}BE}) - |l\rangle\langle l|_{\hat{L}}\otimes\hat{\sigma}_{X'Y'^{\otimes|\mathcal{L}||\mathcal{K}|}E}\right\|_1 \leq 2(\varepsilon+\sqrt{\varepsilon}) + \sqrt{\varepsilon'}, \tag{4.12}$$

where

$$\hat{\sigma}_{X'Y'^{\otimes|\mathcal{L}||\mathcal{K}|}E} \coloneqq \sum_x p_X(x)|x\rangle\langle x|_{X'}\otimes\sigma^{x,m,(\ell,k)}_{Y'^{\otimes|\mathcal{L}||\mathcal{K}|}}\otimes\tilde{\sigma}^{x,m}_E \quad \text{and} \quad P(\sigma^x_{YE},\tilde{\sigma}^x_{YE}) \leq \sqrt{\varepsilon'}.$$

$$\Lambda^m_{X^{|\mathcal{M}|}B} \coloneqq \left(\sum_{m'=1}^{|\mathcal{M}|}\Gamma^{m'}_{X^{|\mathcal{M}|}B}\right)^{-\frac{1}{2}}\Gamma^m_{X^{|\mathcal{M}|}B}\left(\sum_{m'=1}^{|\mathcal{M}|}\Gamma^{m'}_{X^{|\mathcal{M}|}B}\right)^{-\frac{1}{2}}. \tag{4.14}$$

---

decoding instrument:

$$\begin{aligned}\mathcal{D}^1_{BX\to\hat{M}B}(\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}) &\coloneqq \sum_{m=1}^{|\mathcal{M}|}\mathrm{Tr}\{\Lambda^m_{X^{|\mathcal{M}|}B}\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}\}|m\rangle\langle m|_{\hat{M}} \\ &\quad\otimes\frac{\sqrt{\Lambda^m_{X^{|\mathcal{M}|}B}}\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}\sqrt{\Lambda^m_{X^{|\mathcal{M}|}B}}}{\mathrm{Tr}\{\Lambda^m_{X^{|\mathcal{M}|}B}\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}\}},\end{aligned} \tag{4.13}$$

where $\Lambda^m_{X^{|\mathcal{M}|}B}$ is given in (4.14), and for $m \in [1:|\mathcal{M}|]$:

$$\Gamma^m_{X^{|\mathcal{M}|}B} = \mathbb{1}^1_X\otimes\mathbb{1}^2_X\otimes\ldots\otimes T^m_{XB}\otimes\ldots\otimes\mathbb{1}^{|\mathcal{M}|}_X,$$

in which, T^m_{XB} is a test operator distinguishing between two hypotheses, ρ_{XB} and $\rho_X\otimes\rho_B$ and $\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}$ can be seen from (4.5). In fact, Bob needs to discriminate between the following states for different values of $m \in \mathcal{M}$

$$\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B} \coloneqq \rho_X^{\otimes|\mathcal{M}|-1}\otimes\rho^{m,(\ell,k)}_{XB}.$$

Note that to decode the public message m, Bob's decoder does not care about the copy selected by Alice among $|\mathcal{L}||\mathcal{K}|$ copies (no matter which one is selected). In other words, to accomplish the protocol for transmitting the

public message, it suffices to consider $|\mathcal{M}|$ copies of $\rho_{XA} = \sum_x P_X(x)|x\rangle\langle x| \otimes \omega_A^x$ shared between Alice and Bob, where $\omega_A^x = \sum_y p(y|x)\omega_A^{x,y}$. Besides, as is clear from the former discussion, Bob's first decoder faces an $|\mathcal{M}|$-ary hypothesis testing problem. This $|\mathcal{M}|$-ary hypothesis testing problem can be reduced to a binary hypothesis testing problem, in which a binary test operator discriminates between two hypotheses. However, it should not be confused with the fact that in general we deal with an $|\mathcal{M}|$-ary problem.

Let T_{XB} be a test operator in a binary hypothesis testing scenario with null and alternative hypotheses being ρ_{XB} and $\rho_X \otimes \rho_B$, respectively. Discriminator employed by Bob succeeds in guessing null and alternative hypotheses with probabilities $\mathrm{Tr}\{T_{XB}\rho_{XB}\}$ and $\mathrm{Tr}\{(\mathbb{1}_{XB} - T_{XB})(\rho_X \otimes \rho_B)\}$, respectively. And accordingly, the error probabilities associated to the type I and II errors are $\mathrm{Tr}\{(\mathbb{1}_{XB} - T_{XB})\rho_{XB}\}$ and $\mathrm{Tr}\{T_{XB}(\rho_X \otimes \rho_B)\}$, respectively.

It is notation-wise useful to assume that the error probability of the hypothesis tester is $\varepsilon - \delta$ where $\delta \in (0, \varepsilon)$ implying that overall probability of error (ε) is greater than or equal to that of the hypothesis tester. Having introduced the test operator, we can define the following measurement operator for all $m \in [1 : |\mathcal{M}|]$:

$$\Gamma^m_{X^{|\mathcal{M}|}B} = \mathbb{1}^1_X \otimes ... \otimes T^m_{XB} \otimes ... \otimes \mathbb{1}^{|\mathcal{M}|}_X.$$

If Alice sends the m-th message (copy), the probability of producing the correct message at the output equals:

$$\begin{aligned}
&\mathrm{Tr}\{\Gamma^m_{X^{|\mathcal{M}|}B}\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}\} \\
&= \mathrm{Tr}\{(\mathbb{1}_X{}^1 \otimes \mathbb{1}^2_X \otimes ... \otimes T^m_{XB} \otimes ... \otimes \mathbb{1}^{|\mathcal{M}|}_X) \\
&\qquad\qquad (\rho^1_X \otimes ... \otimes \rho^{m,(\ell,k)}_{XB} \otimes ... \otimes \rho^{|\mathcal{M}|}_X)\} \\
&= \mathrm{Tr}\{T^m_{XB}\rho^m_{XB}\} = \mathrm{Tr}\{T_{XB}\rho_{XB}\},
\end{aligned} \tag{4.15}$$

where in the last equality we drop the dependence on m since it is the same for all messages. And probability of deciding in favor of $m' \neq m$ when m was sent is equal to:

$$\begin{aligned}
&\mathrm{Tr}\{\Gamma^{m'}_{X^{|\mathcal{M}|}B}\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}\} \\
&= \mathrm{Tr}\{(\mathbb{1}^m_X \otimes T^{m'}_{XB})(\rho^{m,(\ell,k)}_{XB} \otimes \rho^{m'}_X)\} \\
&= \mathrm{Tr}\{T^{m'}_{XB}(\rho^{m,(\ell,k)}_B \otimes \rho^{m'}_X)\} = \mathrm{Tr}\{T_{XB}(\rho_B \otimes \rho_X)\},
\end{aligned} \tag{4.16}$$

where in the last equality we remove the index m' because this quantity is the same for all $m' \neq m$. This endorses our claim saying that we are facing a binary hypothesis testing problem. From the aforementioned measurement

operators, the square-root measurement given in (4.14) is formed acting as Bob's POVM to detect the public message m. The mentioned POVM construction and the coding scheme, known as position-based coding, first appeared in [115] and [31].

We now focus on the analysis of the error probability of the position-based decoder. The POVM elements above are unitary permutations of one another. In particular, it can be easily shown that all of the elements can be reached by a unitary permutation of the first one, i.e., $\Lambda^{m}_{X^{|\mathcal{M}|}B} = U^{\pi(m)}_{X^{|\mathcal{M}|}B}\Lambda^{1}_{X^{|\mathcal{M}|}B}U^{\pi(m)\dagger}_{X^{|\mathcal{M}|}B}$ in which $\pi(.)$ denotes the permutatin operator [100]. Having said this, we find the probability of error for the first message, i.e., Alice received $m = 1$ and has chosen and sent one of the $|\mathcal{L}||\mathcal{K}|$ A subsystems of the first copy over the channel. We emphasize again that although Alice selects a particular A subsystem out of $|\mathcal{L}||\mathcal{K}|$ copies based on reliability and security of the private message, at this point, when Bob aims to estimate the public message, no matter which A was chosen by Alice, it does not affect Bob's decision about the public message.

We begin by applying the Hayashi-Nagaoka operator inequality (Lemma 1.1) with $S = \Gamma^{1}_{X^{|\mathcal{M}|}B}$ and $T = \sum_{m\neq 1}\Gamma^{m}_{X^{|\mathcal{M}|}B}$ (This T should not be confused with the test operator T^{m}_{XB}):

$$
\begin{aligned}
&Pr(\hat{M}\neq 1|M=1)\\
&\quad=\mathrm{Tr}\{(\mathbb{1}_{X^{|\mathcal{M}|}B}-\Lambda^{1}_{X^{|\mathcal{M}|}B})\rho^{1,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}\}\\
&\quad\leq\mathrm{Tr}\{((1+c)(\mathbb{1}_{X^{|\mathcal{M}|}B}-\Gamma^{1}_{X^{|\mathcal{M}|}B})\rho^{1,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}\}\\
&\qquad\qquad+(2+c+c^{-1})\mathrm{Tr}\{(\sum_{m\neq 1}\Gamma^{m}_{X^{|\mathcal{M}|}B})\rho^{1,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}\}\\
&\quad=(1+c)\mathrm{Tr}\{(\mathbb{1}_{X^{|\mathcal{M}|}B}-\Gamma^{1}_{X^{|\mathcal{M}|}B})\rho^{1,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}\}\\
&\qquad\qquad+(2+c+c^{-1})\mathrm{Tr}\{(\sum_{m\neq 1}\Gamma^{m}_{X^{|\mathcal{M}|}B})\rho^{1,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}\}
\end{aligned}
$$

$$
\begin{aligned}
&=(1+c)\mathrm{Tr}\{(\mathbb{1}^{1}_{X}-T^{1}_{XB})\rho^{1,(\ell,k)}_{XB}\}\\
&\qquad\qquad+(2+c+c^{-1})\sum_{m\neq 1}\mathrm{Tr}\{T^{m}_{XB}(\rho^{m,(\ell,k)}_{B}\otimes\rho^{m}_{X})\}\\
&=(1+c)\mathrm{Tr}\{(\mathbb{1}_{XB}-T_{XB})\rho_{XB}\}\\
&\qquad\qquad+(2+c+c^{-1})(|\mathcal{M}|-1)\mathrm{Tr}\{T_{XB}(\rho_{B}\otimes\rho_{X})\},
\end{aligned}
$$

where in the second last equality, the first and second terms follow from (4.15) and (4.16), respectively. Let Π_{XB} be the optimal test operator in the

following optimization: (see Sec. 1.2)

$$I_{\mathrm{H}}^{\varepsilon-\delta}(X;B)_{\rho_{XB}} \coloneqq -\log_2 \inf_{\substack{0\leq T_{XB}\leq \mathbb{1},\\ \alpha(T_{XB},\rho_{XB})\leq \varepsilon-\delta}} \beta(T_{XB},\rho_X\otimes\rho_B),$$

then,

$$\begin{aligned}
&\mathrm{Tr}\{(\mathbb{1}_{X^{|\mathcal{M}|}B}-\Lambda^1_{X^{|\mathcal{M}|}B})\rho^{1,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}\}\\
&\quad\leq (1+c)\mathrm{Tr}\{(\mathbb{1}_{XB}-\Pi_{XB})\rho_{XB}\}\\
&\qquad\qquad + (2+c+c^{-1})(|\mathcal{M}|-1)\mathrm{Tr}\{\Pi_{XB}(\rho_B\otimes\rho_X)\}\\
&\quad\leq (1+c)(\varepsilon-\delta)+(2+c+c^{-1})|\mathcal{M}|2^{-I_{\mathrm{H}}^{\varepsilon-\delta}(X;B)_\rho}.
\end{aligned}$$

The last term above is set equal to ε, if we solve for $|\mathcal{M}|$, we end up with the following term

$$\log_2|\mathcal{M}| = I_{\mathrm{H}}^{\varepsilon-\delta}(X;B)_\rho + \log_2\left(\frac{\varepsilon-(1+c)(\varepsilon-\delta)}{2+c+c^{-1}}\right),$$

the expression inside the log has a global maximum with respect to c, i.e., the parabola is down-side. We put first derivative equal to zero and pick $c=\frac{\delta}{2\varepsilon-\delta}$ and by doing so finally the following bound holds:

$$\log_2|\mathcal{M}| \leq I_{\mathrm{H}}^{\varepsilon-\delta}(X;B)_\rho - \log_2(\frac{4\varepsilon}{\delta^2}), \tag{4.17}$$

and average probability of error of the public message for the one-shot assisted code is

$$\frac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|}\mathrm{Tr}\{(\mathbb{1}_{X^{|\mathcal{M}|}B}-\Lambda^m_{X^{|\mathcal{M}|}B})\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}\}\leq\varepsilon. \tag{4.18}$$

In the following, we deal with the private message and the second decoder. Before we move on to the privacy analysis, we make a couple of remarks. If the first decoder fails, the second decoder breaks down completely since as is intuitively clear, it ends up with a state having zero information about the position of the sent message. We precisely evaluate the contribution of the first decoder to the error of the second decoder. Moreover, Bob's first decoder acts on his X systems as well as the output of the channel. The Y systems remain intact and in fact, when Bob applies the first decoder, one can assume that the uninvolved systems are being multiplied by the identity operators. Considering this point and the action of the POVM, the resulting state on systems X and B are (up to normalization) $\sqrt{\Lambda^m_{X^{|\mathcal{M}|}B}}\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}\sqrt{\Lambda^m_{X^{|\mathcal{M}|}B}}$, and by taking uninvolved systems into account, we define the state that passes to the second decoder as in (4.19).

$$\sigma^{m,(\ell,k)}_{XX'(YY')^{\otimes|\mathcal{L}\|\mathcal{K}|}BE} \coloneqq \sum_x p_X(x)|x\rangle\langle x|_X \otimes |x\rangle\langle x|_{X'} \otimes \left\{\sigma^{x,m,(1,1)}_{YY'} \otimes \ldots \otimes \sigma^{x,m,(\ell,k-1)}_{YY'} \otimes \sigma^{x,m,(\ell,k)}_{YY'BE} \otimes \sigma^{x,m,(\ell,k+1)}_{YY'} \ldots \otimes \sigma^{x,m,(|\mathcal{L}|,|\mathcal{K}|)}_{YY'}\right\}. \tag{4.19}$$

$$P^{(\ell,k)}_{XY^{|\mathcal{L}\|\mathcal{K}|}B} \coloneqq \left(\sum_{\ell'=1}^{|\mathcal{L}|}\sum_{k'=1}^{|\mathcal{K}|} N^{(\ell',k')}_{XY^{|\mathcal{L}\|\mathcal{K}|}B}\right)^{-\frac{1}{2}} N^{(\ell,k)}_{XY^{|\mathcal{L}\|\mathcal{K}|}B} \left(\sum_{\ell'=1}^{|\mathcal{L}|}\sum_{k'=1}^{|\mathcal{K}|} N^{(\ell',k')}_{XY^{|\mathcal{L}\|\mathcal{K}|}B}\right)^{-\frac{1}{2}}, \tag{4.21}$$

where for all $\ell \in [1:|\mathcal{L}|]$, and $k \in [1:|\mathcal{K}|]$,

$$N^{(\ell,k)}_{XY^{|\mathcal{L}\|\mathcal{K}|}B} \coloneqq \tag{4.22}$$

$$|x\rangle\langle x|_X \otimes \mathbb{1}_Y^{(1,1)} \otimes \ldots \otimes \mathbb{1}_Y^{(1,|\mathcal{K}|)} \otimes \ldots \otimes \mathbb{1}_Y^{(\ell,k-1)} \otimes Z^{(\ell,k)}_{YB} \otimes \mathbb{1}_Y^{(\ell,k+1)} \ldots \otimes \mathbb{1}_Y^{(|\mathcal{L}|,|\mathcal{K}|)}. \tag{4.23}$$

Correctness and secrecy of Private Message, (Privacy error) Eq. (4.11)

Reconsider the state in (4.19) showing the state resulted from transmitting the (ℓ,k)-th A subsystem through the channel (for a given m) after Bob applies his first decoder. Remember that in the first part of the protocol it did not matter which copy out of $|\mathcal{L}\|\mathcal{K}|$ copies was chosen but now it does matter as Bob and Eve try to decode the private message. Bob's decoder for the private message ℓ is constructed as follows:

$$\mathcal{D}^2_{\hat{M}BY\to\hat{L}}(\sigma^{m,(\ell,k)}_{XY^{\otimes|\mathcal{L}\|\mathcal{K}|}B}) \coloneqq \sum_{l=1}^{|\mathcal{L}|} \mathrm{Tr}\{P^{\ell}_{XY^{|\mathcal{L}\|\mathcal{K}|}B}\sigma^{m,(\ell,k)}_{XY^{\otimes|\mathcal{L}\|\mathcal{K}|}B}\}|\ell\rangle\langle\ell|_{\hat{L}}, \tag{4.20}$$

where for all $x \in \mathcal{X}$

$$P^{\ell}_{XY^{|\mathcal{L}\|\mathcal{K}|}B} = \sum_{k=1}^{|\mathcal{K}|} P^{(\ell,k)}_{XY^{|\mathcal{L}\|\mathcal{K}|}B},$$

and $P^{x,(\ell,k)}_{XY^{|\mathcal{L}\|\mathcal{K}|}B}$ is given in (4.21), in which, for all $x \in \mathcal{X}$, Z_{YB} is a binary test operator distinguishing between two hypotheses σ^x_{YB} and $\sigma^x_Y \otimes \sigma^x_B$ with an

error of $\varepsilon - \delta$, i.e.,

$$\mathrm{Tr}\{Z_{YB}\sigma^x_{YB}\} \geq 1 - (\varepsilon - \delta),$$

where $\varepsilon \in (0,1)$ and $\delta \in (0,\varepsilon)$. Note that the variable x appearing in the operator indicates the fact that the decoding works for all $x \in \mathcal{X}$.

Bob has to be able to distinguish between states $\sigma^{m,(1,1)}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}B}, \sigma^{m,(1,2)}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}B}, \ldots,$ $\sigma^{m,(l,k)}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}B}$ and $\sigma^{m,(|\mathcal{L}|,|\mathcal{K}|)}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}B}$; we will see that this amounts to Bob being able to distinguish between the following states:

$$\sum_x p_X(x)|x\rangle\langle x|_X \otimes \sigma^x_{YB},$$
$$\sum_x p_X(x)|x\rangle\langle x|_X \otimes \sigma^x_Y \otimes \sigma^x_B,$$

or more precisely, between state σ^x_{YB} and $\sigma^x_Y \otimes \sigma^x_B$ for all $x \in \mathcal{X}$. We importantly note that after detecting the public message m, Bob is faced a $|\mathcal{L}||\mathcal{K}|$-ary hypothesis testing problem. This scenario should not be confused by the binary hypothesis testing above, i.e., Alice distinguishes between σ^x_{YB} and $\sigma^x_Y \otimes \sigma^x_B$ for all $x \in \mathcal{X}$, the latter happens to be a byproduct of the general scenario once we go into the error analysis. Now see that if the pair (ℓ,k) was chosen, the action of the operator $N^{(\ell,k)}_{Y^{|\mathcal{L}||\mathcal{K}|}B}$ would be as follows:

$$\mathrm{Tr}\{N^{(\ell,k)}_{XY^{|\mathcal{L}||\mathcal{K}|}B}\sigma^{m,(\ell,k)}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}B}\} = \sum_x p_X(x)\mathrm{Tr}\{Z^{(\ell,k)}_{YB}\sigma^{x,m,(\ell,k)}_{YB}\},$$

and for any other operator, i.e., the private message-local key pair (ℓ,k) is confused by (ℓ',k'), either $k \neq k'$, $l \neq \ell'$ or $(k \neq k', \ell \neq \ell')$:

$$\begin{aligned}
&\mathrm{Tr}\{N^{(\ell',k')}_{XY^{|\mathcal{L}||\mathcal{K}|}B}\sigma^{m,(\ell,k)}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}B}\} \\
&= \mathrm{Tr}\Big\{|x\rangle\langle x|_X \otimes (Z^{(\ell',k')}_{YB} \otimes \mathbb{1}^{(\ell,k)}_Y) \\
&\qquad (\sum_x p_X(x)|x\rangle\langle x|_X \otimes \sigma^{x,m,(\ell',k')}_Y \otimes \sigma^{x,m,(\ell,k)}_{YB})\Big\} \\
&= \sum_x p_X(x)\mathrm{Tr}\{Z^{(\ell',k')}_{YB}(\sigma^{x,m,(\ell',k')}_Y \otimes \sigma^{x,m,(\ell,k)}_B)\}.
\end{aligned}$$

We can think of the states $\sigma^{x,m}_{YB}$ and $\sigma^{x,m}_Y \otimes \sigma^{x,m}_B$ as the null and alternative hypotheses, respectively. As a typical procedure in quantum error analysis, Bob forms the square-root measurement operators given in (4.21) acting as his POVMs to detect the private message-local key pair (ℓ,k). It can be

shown that each measurement operator $P^{(\ell,k)}_{Y^{|\mathcal{L}\|\mathcal{K}|}B}$ is related to the first one $P^{(1,1)}_{Y^{|\mathcal{L}\|\mathcal{K}|}B}$ by a unitary permutation of $Y^{|\mathcal{L}\|\mathcal{K}|}$ systems for all $x \in \mathcal{X}$. This fact gives rise to the following identity, for all $\ell \in [1:|\mathcal{L}|]$ and $k \in [1:|\mathcal{K}|]$:

$$\begin{aligned}\mathrm{Tr}\{(\mathbb{1}_{XY^{|\mathcal{L}\|\mathcal{K}|}B} - P^{(1,1)}_{XY^{|\mathcal{L}\|\mathcal{K}|}B})\sigma^{m,(1,1)}_{XY^{\otimes|\mathcal{L}\|\mathcal{K}|}B}\} \\ = \mathrm{Tr}\{(\mathbb{1}_{XY^{|\mathcal{L}\|\mathcal{K}|}B} - P^{(\ell,k)}_{XY^{|\mathcal{L}\|\mathcal{K}|}B})\sigma^{m,(\ell,k)}_{XY^{\otimes|\mathcal{L}\|\mathcal{K}|}B}\},\end{aligned}$$

meaning that the error probability is the same for all private messages, in other words, it is independent from a particular chosen twin (ℓ, k); And again this implies that average error probability equals individual error probabilities. In what follows, we deploy Hayashi-Nagaoka operator inequality (Lemma 1.1) to analyze the error probability. Let's assume $(\ell = 1, k = 1)$ was sent. Moreover, let's choose $S = N^{(1,1)}_{Y^{|\mathcal{L}\|\mathcal{K}|}B}$ and $T = \sum_{\ell'\neq 1}\sum_{k'\neq 1} N^{(\ell',k')}_{Y^{|\mathcal{L}\|\mathcal{K}|}B}$ in Hayashi-Nagaoka inequality. We have

$$\begin{aligned}&\mathrm{Tr}\{(\mathbb{1}_{Y^{|\mathcal{L}\|\mathcal{K}|}B} - P^{(1,1)}_{XY^{|\mathcal{L}\|\mathcal{K}|}B})\sigma^{m,(1,1)}_{XY^{\otimes|\mathcal{L}\|\mathcal{K}|}B}\} \\ &\leq (1+c)\mathrm{Tr}\{(\mathbb{1}_{Y^{|\mathcal{L}\|\mathcal{K}|}B} - N^{(1,1)}_{XY^{|\mathcal{L}\|\mathcal{K}|}B})\sigma^{m,(1,1)}_{XY^{\otimes|\mathcal{L}\|\mathcal{K}|}B}\} \\ &\qquad + (2+c+c^{-1})\sum_{l'\neq 1}\sum_{k'\neq 1}\mathrm{Tr}\{N^{(\ell',k')}_{XY^{|\mathcal{L}\|\mathcal{K}|}B}\sigma^{m,(1,1)}_{XY^{\otimes|\mathcal{L}\|\mathcal{K}|}}\} \\ &= (1+c)\sum_x p_X(x)\mathrm{Tr}\{(\mathbb{1}_{YB} - Z^{(1,1)}_{YB})\sigma^{x,m,(1,1)}_{YB}\} \\ &+ (2+c+c^{-1})\sum_x p_X(x)\Big(\\ &\qquad \sum_{l'\neq 1}\sum_{k'\neq 1}\mathrm{Tr}\{Z^{(\ell',k')}_{YB}(\sigma^{x,m,(\ell',k')}_{Y}\otimes\sigma^{x,m,(1,1)}_{B})\}\Big) \\ &= (c+1)\sum_x p_X(x)\mathrm{Tr}\{(\mathbb{1}_{YB} - Z_{YB})\sigma^{x,m}_{YB}\} \\ &+ (2+c+c^{-1})(|\mathcal{L}\|\mathcal{K}|-1)(\\ &\qquad \sum_x p_X(x)\mathrm{Tr}\{Z_{YB}(\sigma^{x,m}_{Y}\otimes\sigma^{x,m}_{B})\})\end{aligned}$$

For each realization x, let Θ^x_{YB} denote the measurement operator that is the answer to the optimization corresponding to the hypothesis testing relative entropy with $\alpha(Z_{YB}, \sigma^x_{YB}) \coloneqq \mathrm{Tr}\{(\mathbb{1} - Z_{YB})\sigma^x_{YB}\}$ and $\beta(Z_{YB}, \sigma^x_Y \otimes \sigma^x_B) \coloneqq \mathrm{Tr}\{Z_{YB}(\sigma^x_Y \otimes \sigma^x_B)\}$ where by assumption it detects the joint state with an error probability of $\varepsilon - \delta$ where $\delta \in (0, \varepsilon)$. This optimization can be done for all x, but from the definition of the conditional hypothesis testing mutual information (Definition 9), the x minimizing the expression given in equation (4.24) over a nearby distribution is of particular interest in error analysis; The error probability simplifies as follows:

$$\tilde{I}_{\mathrm{H}}^{\varepsilon-\delta}(Y;B|X)_{\sigma_{XYB}} \coloneqq \max_{\sigma'_X} \min_{x\in\mathrm{supp}(\sigma'_X)} \left\{ -\log_2 \inf_{\substack{0\leq Z^x_{YB}\leq \mathbb{1},\\ \alpha(Z^x_{YB},\sigma^x_{YB})\leq\varepsilon-\delta}} \beta(Z^x_{YB},\sigma^x_Y\otimes\sigma^x_B) \right\}, \tag{4.24}$$

$$\text{where} \qquad \sigma_{XYB} = \sum_x p_X(x)|x\rangle\langle x|_X \otimes \sigma^x_{YB} \qquad \text{and} \qquad P(\sigma'_X,\sigma_X)\leq\varepsilon''.$$

$$\begin{aligned}
&\mathrm{Tr}\{(\mathbb{1}_{Y^{|\mathcal{L}||\mathcal{K}|}B} - P^{(1,1)}_{XY^{|\mathcal{L}||\mathcal{K}|}B})\sigma^{m,(1,1)}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}B}\}\\
&\quad\leq (c+1)\sum_x p_X(x)\mathrm{Tr}\{(\mathbb{1}_{YB}-\Theta_{YB})\sigma^{x,m}_{YB}\}\\
&\quad+ (2+c+c^{-1})|\mathcal{L}||\mathcal{K}|\sum_x p_X(x)\mathrm{Tr}\{\Theta_{YB}(\sigma^{x,m}_Y\otimes\sigma^{x,m}_B)\}\\
&\quad\leq (c+1)(\varepsilon-\delta)\\
&\qquad\qquad+(2+c+c^{-1})|\mathcal{L}||\mathcal{K}|2^{-\tilde{I}^{\varepsilon-\delta}_{\mathrm{H}}(Y;B|X)_{\sigma_{XYB}}},
\end{aligned}$$

where in the last line, the first expression is derived from the assumption that for all x, $\mathrm{Tr}\{\Theta_{YB}\sigma^{x,m}_{YB}\}\geq 1-(\varepsilon-\delta)$, and the second expression follows from (4.24). By putting the last line above equal to ε (Bob's error in detecting private message is ε) and solving it for $|\mathcal{L}||\mathcal{K}|$, we get:

$$\begin{aligned}
\log_2|\mathcal{L}||\mathcal{K}| =& \tilde{I}^{\varepsilon-\delta}_{\mathrm{H}}(Y;B|X)_{\sigma_{XYB}}\\
&+\log_2\left(\frac{\varepsilon-(1+c)(\varepsilon-\delta)}{2+c+c^{-1}}\right).
\end{aligned}$$

The right-hand side of the expression above should be maximized with respect to c. Since it is a down-side parabola when it comes to maximization, we pick its global maximum which occurs at $c=\frac{\delta}{2\varepsilon-\delta}$. By plugging it back into the expression we end up having:

$$\log_2|\mathcal{L}||\mathcal{K}| \leq \tilde{I}^{\varepsilon-\delta}_{\mathrm{H}}(Y;B|X)_{\sigma_{XYB}} - \log_2(\frac{4\varepsilon}{\delta^2}).$$

The derivation above ensures that in the privacy error in (4.11), Bob's error in detecting private message is satisfied (note that each separate criterion comes about by tracing out the other one).

We now turn our attention to Eve's state and security criterion which is merged into (4.11). We also assume that Eve has detected the public

$$\sigma^{m,\ell}_{X'Y'\otimes|\mathcal{L}||\mathcal{K}|_E} \coloneqq \frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma^{m,(\ell,k)}_{X'Y'\otimes|\mathcal{L}||\mathcal{K}|_E} \coloneqq \sum_x p_X(x)|x\rangle\langle x|_{X'}\otimes\sigma^{x,m,(1,1)}_{Y'}\otimes\ldots$$
$$\otimes\left[\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma^{x,m,(\ell,1)}_{Y'}\otimes\ldots\otimes\sigma^{x,m,(\ell,k)}_{Y'E}\otimes\ldots\otimes\sigma^{x,m,(\ell,|\mathcal{K}|)}_{Y'}\right]\otimes\ldots\otimes\sigma^{x,m,(|\mathcal{L}|,|\mathcal{K}|)}_{Y'}. \tag{4.25}$$

---

message. From (4.19), for a fixed (ℓ,k), Eve's state is[7]

$$\sigma^{m,(\ell,k)}_{X'Y'\otimes|\mathcal{L}||\mathcal{K}|_E} = \sum_x p_X(x)|x\rangle\langle x|_{X'}\otimes\sigma^{x,m,(1,1)}_{Y'}\otimes\ldots$$
$$\otimes\sigma^{x,m,(\ell,k)}_{Y'E}\otimes\ldots\otimes\sigma^{x,m,(|\mathcal{L}|,|\mathcal{K}|)}_{Y'}.$$

As we discussed before, k is a local key exclusively in possession of Alice and for a given private message ℓ, it is chosen uniformly at random; Hence, for a given message ℓ, the state of Eve can be written as equation (4.25). We would like her to learn almost nothing about the sent private message. In other words, her state becomes independent from the chosen index ℓ:

$$\forall m,\ell: \quad \frac{1}{2}\|\sigma^{m,\ell}_{X'Y'\otimes|\mathcal{L}||\mathcal{K}|_E}-\hat{\sigma}_{X'Y'\otimes|\mathcal{L}||\mathcal{K}|_E}\|_1\leq\sqrt{\varepsilon'}, \tag{4.26}$$

where

$$\hat{\sigma}_{X'Y'\otimes|\mathcal{L}||\mathcal{K}|_E}\coloneqq\sum_x p_X(x)|x\rangle\langle x|_{X'}\otimes\sigma^{x,m}_{Y'\otimes|\mathcal{L}||\mathcal{K}|}\otimes\tilde{\sigma}^{x,m}_E \tag{4.27}$$

for $\varepsilon'\in(0,1)$ and some state $\tilde{\sigma}^{m,x}_E$ that is the marginal of $\tilde{\sigma}^{m,x}_{Y'E}$ and $P(\sigma^{m,x}_{Y'E},\tilde{\sigma}^{m,x}_{Y'E})\leq\sqrt{\varepsilon'}-\delta'$ in which $\delta'\in(0,\sqrt{\varepsilon'})$. From the invariance of trace distance with respect to tensor-product states, we can expand the security constraint (4.26) as given by (4.28).

From the convex-split lemma and the definition of the conditional smooth max-mutual information (see Definition 10), if the following condition holds[8],

$$\log_2|\mathcal{K}| = \tilde{I}^{\sqrt{\varepsilon'}-\delta'}_{\max}(E;Y|X)_\sigma + 2\log_2(\frac{1}{\delta'}), \tag{4.29}$$

[7]Note that the state in (4.19) denotes the disturbed state after Bob finds the public message, without loss of generality, we also assume Eve affects the initial state in the same way.

[8]To maintain consistency, in the following expression, we show Eve's X' and Y' systems with X and Y, respectively.

$$
\begin{aligned}
&\frac{1}{2}\left\|\sigma^{m,\ell}_{X'Y'\otimes|\mathcal{L}\|\mathcal{K}|E} - \sum_x p_X(x)|x\rangle\langle x|_{X'} \otimes \sigma^{x,m}_{Y'\otimes|\mathcal{L}\|\mathcal{K}|} \otimes \tilde{\sigma}^{x,m}_E\right\|_1 \\
&= \frac{1}{2}\left\|\sum_x p_X(x)|x\rangle\langle x|_{X'} \otimes \left(\sigma^{x,m,\ell}_{Y'\otimes|\mathcal{L}\|\mathcal{K}|E} - \sigma^{x,m}_{Y'\otimes|\mathcal{L}\|\mathcal{K}|} \otimes \tilde{\sigma}^{x,m}_E\right)\right\|_1 \\
&= \sum_x p_X(x)\frac{1}{2}\left\|\sigma^{x,m,\ell}_{Y'\otimes|\mathcal{L}\|\mathcal{K}|E} - \sigma^{x,m}_{Y'\otimes|\mathcal{L}\|\mathcal{K}|} \otimes \tilde{\sigma}^{x,m}_E\right\|_1 \\
&= \frac{1}{2}\sum_x p_X(x)\left\|\frac{1}{K}\sum_{k=1}^{|\mathcal{K}|}\sigma^{x,m,(\ell,1)}_{Y'} \otimes \ldots \otimes \sigma^{x,m,(\ell,k-1)}_{Y'} \otimes \right. \\
&\qquad \left.\left\{\left(\sigma^{x,m,(\ell,k)}_{Y'E} - \sigma^{x,m}_{Y'} \otimes \tilde{\sigma}^{x,m}_E\right) \otimes \sigma^{x,m,(\ell,k+1)}_{Y'} \otimes \ldots \otimes \sigma^{x,m,(|\mathcal{L}|,|\mathcal{K}|)}_{Y'}\right\|_1\right\}. \quad (4.28)
\end{aligned}
$$

---

then

$$P(\sigma^{m,\ell}_{X'Y'\otimes|\mathcal{L}\|\mathcal{K}|E}, \hat{\sigma}_{X'Y'\otimes|\mathcal{L}\|\mathcal{K}|E}) \leq \sqrt{\varepsilon'}$$

is satisfied with $\hat{\sigma}_{X'Y'|\mathcal{L}\|\mathcal{K}|E}$ defined in (4.27) and from the relation between purified distance and trace distance correctness of (4.26) is guaranteed. Note also that $P(\sigma^{x,m}_E, \tilde{\sigma}^{x,m}_E) \leq P(\sigma^{x,m}_{YE}, \tilde{\sigma}^{x,m}_{YE}) \leq \sqrt{\varepsilon'} - \delta'$. So far, we have shown the correctness of two separate criteria for the assisted code. For our purposes here we would like to have a single condition for the private message encompassing both conditions discussed lately and so in the following, by sticking to the reciepe set out by [100], we try to merge two conditions and deal with a single *privacy error*. We see that the single criterion will be beneficial once we derandomize the code and upon derandomization, the requirements set out in the definition of the unassisted code will be fulfilled.

We saw that the average error probability is equal to the individual error probabilities:

$$
\begin{aligned}
&\mathrm{Tr}\{(\mathbb{1}_{XY'|\mathcal{L}\|\mathcal{K}|B} - P^{(\ell,k)}_{XY'|\mathcal{L}\|\mathcal{K}|B})\sigma^{m,(\ell,k)}_{XY\otimes|\mathcal{L}\|\mathcal{K}|B}\} = \\
&\frac{1}{|\mathcal{L}\|\mathcal{K}|}\sum_{l=1}^{|\mathcal{L}|}\sum_{k=1}^{|\mathcal{K}|}\mathrm{Tr}\{(\mathbb{1}_{Y'|\mathcal{L}\|\mathcal{K}|B} - P^{(\ell,k)}_{XY'|\mathcal{L}\|\mathcal{K}|B})\sigma^{m,(\ell,k)}_{XY\otimes|\mathcal{L}\|\mathcal{K}|B}\} \leq \varepsilon \quad (4.30)
\end{aligned}
$$

We continue by expanding $\sigma^{m,(\ell,k)}_{XY\otimes|\mathcal{L}\|\mathcal{K}|B} = \sum_x p_X(x)|x\rangle\langle x|_X \otimes \sigma^{x,m,(\ell,k)}_{Y\otimes|\mathcal{L}\|\mathcal{K}|B}$ as in equation (4.31).

Reconsider the optimal test operator Θ_{YB}, we can write the following equation:

$$
\begin{aligned}
&\sigma^{x,m,(\ell,k)}_{Y^{\otimes|\mathcal{L}||\mathcal{K}|}B} \\
&= \sigma^{x,m,(1,1)}_{Y} \otimes ... \otimes \sigma^{x,m,(1,|\mathcal{K}|)}_{Y} \otimes ... \otimes \sigma^{x,m,(\ell,k-1)}_{Y} \otimes \sigma^{x,m,(\ell,k)}_{YB} \otimes \sigma^{x,m,(\ell,k+1)}_{Y} ... \otimes \sigma^{x,m,(|\mathcal{L}|,|\mathcal{K}|)}_{Y} \\
&= \sum_{y_{11}} p(y_{11}|x)|y_{11}\rangle\langle y_{11}| \otimes ... \otimes \sum_{y_{1|\mathcal{K}|}} p(y_{1|\mathcal{K}|}|x)|y_{1|\mathcal{K}|}\rangle\langle y_{1|\mathcal{K}|}| \otimes ... \otimes \sum_{y_{\ell k-1}} p(y_{\ell k-1})|y_{\ell k-1}\rangle\langle y_{\ell k-1}| \\
&\otimes ... \otimes \sum_{y_{\ell k}} p(y_{\ell k}|x)|y_{\ell k}\rangle\langle y_{\ell k}| \otimes \\
&\qquad \left\{ \sigma^{x,m,y_{\ell k}}_{B} \otimes \sum_{y_{\ell k+1}} p(y_{\ell k+1}|x)|y_{\ell k+1}\rangle\langle y_{\ell k+1}| ... \otimes \sum_{y_{|\mathcal{L}||\mathcal{K}|}} p(y_{|\mathcal{L}||\mathcal{K}|}|x)|y_{|\mathcal{L}||\mathcal{K}|}\rangle\langle y_{|\mathcal{L}||\mathcal{K}|}| \right\} \\
&= \sum_{y_{11}y_{12}...y_{lk}...y_{|\mathcal{L}||\mathcal{K}|}} p(y_{11}|x)p(y_{12}|x)...p(y_{\ell k}|x)...p(y_{|\mathcal{L}||\mathcal{K}|}|x) \\
&\qquad\qquad |y_{11}...y_{\ell k}...y_{|\mathcal{L}||\mathcal{K}|}\rangle\langle y_{11}...y_{\ell k}...y_{|\mathcal{L}||\mathcal{K}|}| \otimes \sigma^{x,m,y_{\ell k}}_{B},
\end{aligned}
$$

hence

$$
\begin{aligned}
\sigma^{m,(\ell,k)}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}B} = &\sum_{x,y_{11}y_{12}...y_{lk}...y_{|\mathcal{L}||\mathcal{K}|}} p_{XY^{|\mathcal{L}||\mathcal{K}|}}(x, y_{11}...y_{lk}...y_{|\mathcal{L}||\mathcal{K}|}) \\
&|x\rangle\langle x|_X \otimes |y_{11}...y_{\ell k}...y_{|\mathcal{L}||\mathcal{K}|}\rangle\langle y_{11}...y_{\ell k}...y_{|\mathcal{L}||\mathcal{K}|}| \otimes \sigma^{x,m,y_{\ell k}}_{B}. \qquad (4.31)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Tr}\{\Theta_{YB}\sigma^{x}_{YB}\} &= \mathrm{Tr}\{\Theta_{YB}(\sum_{y} p(y|x)|y\rangle\langle y| \otimes \sigma^{x,y}_{B})\} \\
&= \sum_{y} p(y|x)\mathrm{Tr}\{\langle y|\Theta^{x}_{YB}|y\rangle\sigma^{x,y}_{B}\} \\
&= \sum_{y} p(y|x)\mathrm{Tr}\{G^{x,y}_{B}\sigma^{x,y}_{B}\},
\end{aligned}
$$

where $G^{x,y}_{B} \coloneqq \langle y|\Theta^{x}_{YB}|y\rangle$. In an analogous way:

$$
\mathrm{Tr}\{\Theta_{YB}(\sigma^{x}_{Y} \otimes \sigma^{x}_{B})\} = \sum_{y} p(y|x)\mathrm{Tr}\{G^{x,y}_{B}\sigma^{x}_{B}\}.
$$

The above derivations lead the test operator to be considered as $\Theta_{YB} = \sum_y |y\rangle\langle y|_Y \otimes G^{x,y}_{B}$, i.e., the operator classical on Y achieves the same optimal values as any general operator. Next we try to embed the test operator in the $N^{(\ell,k)}_{XY^{|\mathcal{L}||\mathcal{K}|}B}$ as given in (4.22) and expanded as in (4.32). Observe the structure of the POVM given in (4.33). And finally our POVM has the form given in equation (4.34). To build a POVM on the full space, we add $\Omega^{0}_{B} = \mathbb{1}_B - \sum_\ell \sum_k \Omega^{x,(\ell,k)}_{B}$ to the set $\{\Omega^{x,(\ell,k)}_{B}\}^{|\mathcal{L}|,|\mathcal{K}|}_{\ell=1,k=1}$. By combining (4.31) and

$$N^{(\ell,k)}_{XY^{|\mathcal{L}||\mathcal{K}|}B} = |x\rangle\langle x|_X \otimes \mathbb{1}_Y^{(1,1)} \otimes ... \otimes \Theta^{(\ell,k)}_{YB} \otimes ... \otimes \mathbb{1}_Y^{(|\mathcal{L}|,|\mathcal{K}|)}$$
$$= \sum_{y_{11}...y_{l\mathrm{k}}...y_{|\mathcal{L}||\mathcal{K}|}} |x\rangle\langle x|_X \otimes |y_{11}...y_{\ell k}...y_{|\mathcal{L}||\mathcal{K}|}\rangle\langle y_{11}...y_{\ell k}...y_{|\mathcal{L}||\mathcal{K}|}| \otimes G_B^{x,y_{\ell k}}. \tag{4.32}$$

$$\left(\sum_{\ell'}\sum_{k'} N^{(\ell',k')}_{XY^{|\mathcal{L}||\mathcal{K}|}B}\right)^{-\frac{1}{2}}$$
$$= \sum_{y_{11}...y_{|\mathcal{L}||\mathcal{K}|}} |x\rangle\langle x|_X \otimes |y_{11}...y_{|\mathcal{L}||\mathcal{K}|}\rangle\langle y_{11}...y_{|\mathcal{L}||\mathcal{K}|}| \otimes \left(\sum_{\ell'}\sum_{k'} G_B^{x,y_{\ell'k'}}\right)^{-\frac{1}{2}}, \tag{4.33}$$

$$P^{(\ell,k)}_{XY^{|\mathcal{L}||\mathcal{K}|}B}$$
$$= \left(\sum_{\ell'}\sum_{k'} N^{(\ell',k')}_{XY^{|\mathcal{L}||\mathcal{K}|}B}\right)^{-\frac{1}{2}} N^{(\ell,k)}_{XY^{|\mathcal{L}||\mathcal{K}|}B} \left(\sum_{l'}\sum_{k'} N^{(\ell',k')}_{XY^{|\mathcal{L}||\mathcal{K}|}B}\right)^{-\frac{1}{2}}$$
$$= \sum_{y_{11}...y_{|\mathcal{L}||\mathcal{K}|}} |x\rangle\langle x|_X \otimes |y_{11}...y_{|\mathcal{L}||\mathcal{K}|}\rangle\langle y_{11}...y_{|\mathcal{L}||\mathcal{K}|}| \otimes \Omega_B^{x,y_{\ell k}}, \tag{4.34}$$

where

$$\Omega_B^{x,(\ell,k)} \coloneqq \left(\sum_{\ell'}\sum_{k'} G_B^{x,y_{\ell'k'}}\right)^{-\frac{1}{2}} G_B^{x,y_{\ell k}} \left(\sum_{\ell'}\sum_{k'} G_B^{x,y_{\ell'k'}}\right)^{-\frac{1}{2}}.$$

(4.34), we find that

$$\mathrm{Tr}\{(\mathbb{1}_{Y^{|\mathcal{L}||\mathcal{K}|}B} - P^{(\ell,k)}_{XY^{|\mathcal{L}||\mathcal{K}|}B})\sigma^{m,(\ell,k)}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}B}\}$$
$$= \sum_{x,y_{11}...y_{|\mathcal{L}||\mathcal{K}|}} p_X(x)p(y_{11}|x)...p(y_{|\mathcal{L}||\mathcal{K}|}|x) \times \mathrm{Tr}\{(\mathbb{1}_B - \Omega_B^{x,y_{\ell k}})\sigma_B^{x,m,y_{\ell k}}\}$$

and from (4.30), the equality of the average and the individual error probabilities, yields equation (4.35).

By taking advantage of the POVMs $\{\Omega_B^{x,(\ell,k)}\}_{\ell=1,k=1}^{|\mathcal{L}|,|\mathcal{K}|}$, the following measurement channels are defined

$$\mathcal{D}'^2_{B\to\hat{L}}(\omega_B) \coloneqq \sum_{\ell=1}^{|\mathcal{L}|}\sum_{k=1}^{|\mathcal{K}|} \mathrm{Tr}\{\Omega_B^{x,y_{\ell,k}}\omega_B\}|\ell\rangle\langle\ell|_{\hat{L}}, \tag{4.36}$$

$$\frac{1}{|\mathcal{L}|}\frac{1}{|\mathcal{K}|}\sum_{\ell=1}^{|\mathcal{L}|}\sum_{k=1}^{|\mathcal{K}|}\mathrm{Tr}\{(\mathbb{1}_{Y^{|\mathcal{L}||\mathcal{K}|}B}-P^{(\ell,k)}_{XY^{|\mathcal{L}||\mathcal{K}|}B})\sigma^{m,(\ell,k)}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}B}\}$$
$$=\frac{1}{|\mathcal{L}|}\frac{1}{|\mathcal{K}|}\sum_{\ell=1}^{|\mathcal{L}|}\sum_{k=1}^{|\mathcal{K}|}\sum_{x,y_{11}\ldots y_{|\mathcal{L}||\mathcal{K}|}}p_X(x)p(y_{11}|x)\ldots p(y_{|\mathcal{L}||\mathcal{K}|}|x)\mathrm{Tr}\{(\mathbb{1}_B-\Omega_B^{x,y_{\ell k}})\sigma_B^{x,m,y_{\ell k}}\}$$
$$=\sum_{x,y_{11}\ldots y_{|\mathcal{L}||\mathcal{K}|}}p_X(x)p(y_{11}|x)\ldots p(y_{|\mathcal{L}||\mathcal{K}|}|x)\left(\frac{1}{|\mathcal{L}|}\frac{1}{|\mathcal{K}|}\sum_{\ell=1}^{|\mathcal{L}|}\sum_{k=1}^{|\mathcal{K}|}\mathrm{Tr}\{(\mathbb{1}_B-\Omega_B^{x,y_{\ell,k}})\sigma_B^{x,m,y_{\ell k}}\}\right)\leq\varepsilon. \tag{4.35}$$

$$\sum_{x,y_{11}\ldots y_{|\mathcal{L}||\mathcal{K}|}}p_X(x)p(y_{11}|x)\ldots p(y_{|\mathcal{L}||\mathcal{K}|}|x)$$
$$\left[\frac{1}{|\mathcal{L}|}\frac{1}{|\mathcal{K}|}\sum_{\ell}\sum_{k}\frac{1}{2}\left\|\mathcal{D}'^2_{B\to\hat{L}\hat{K}}(\sigma_B^{x,m,y_{\ell k}})-|\ell\rangle\langle\ell|_{\hat{L}}\otimes|k\rangle\langle k|_{\hat{K}}\right\|_1\right]\leq\varepsilon. \tag{4.38}$$

$$\mathcal{D}'^2_{B\to\hat{L}\hat{K}}(\omega_B)\coloneqq\sum_{\ell=1}^{|\mathcal{L}|}\sum_{k=1}^{|\mathcal{K}|}\mathrm{Tr}\{\Omega_B^{x,y_{\ell,k}}\omega_B\}|\ell\rangle\langle\ell|_{\hat{L}}\otimes|k\rangle\langle k|_{\hat{K}}, \tag{4.37}$$

where ω_B is a general quantum state and $\mathrm{Tr}_{\hat{K}}\circ\mathcal{D}'^2_{B\to\hat{L}\hat{K}}=\mathcal{D}'^2_{B\to\hat{L}}$. Note that in (4.36) the probability of getting a particular ℓ equals $\sum_{k=1}^{|\mathcal{K}|}\mathrm{Tr}\{\Omega_B^{x,y_{\ell k}}\omega_B\}$. By direct calculations, it can be seen that:

$$\mathrm{Tr}\{(\mathbb{1}_B-\Omega_B^{x,y_{\ell,k}})\sigma_B^{x,m,y_{\ell,k}}\}$$
$$=\frac{1}{2}\left\|\mathcal{D}'^2_{B\to\hat{L}\hat{K}}(\sigma_B^{x,m,y_{\ell,k}})-|\ell\rangle\langle\ell|_{\hat{L}}\otimes|k\rangle\langle k|_{\hat{K}}\right\|_1,$$

averaging over ℓ, k and $(x,y_{1,1}\ldots y_{|\mathcal{L}|,|\mathcal{K}|})$ and using (4.35), we get equation (4.38). In equation (4.38), if we take the average over k inside the trace distance and trace out $\hat{K}$ system, by the convexity and monotonicity of the trace distance, we obtain the equations in (4.39).

Considering the POVM $\{\Omega_B^{x,y_{\ell k}}\}_{\ell=1,k=1}^{|\mathcal{L}|,|\mathcal{K}|}$, the probability of getting ℓ' conditioned on the fact that (ℓ,k) was sent is equal to $\sum_{k'=1}^{K}\mathrm{Tr}\{\Omega_B^{x,y_{\ell'k'}}\sigma_B^{x,y_{\ell k}}\}$ and it is clear from the uniformity of the local key that the probability of getting l' given that l was sent, equals $\Pr(l'|l)=\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sum_{k'=1}^{|\mathcal{K}|}\mathrm{Tr}\{\Omega_B^{x,y_{\ell'k'}}\sigma_{BE}^{x,m,y_{\ell k}}\}=\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sum_{k'=1}^{|\mathcal{K}|}\mathrm{Tr}\{\Omega_B^{x,y_{\ell'k'}}\sigma_B^{x,m,y_{\ell k}}\}$. Note that evidently $\sum_{\ell'=1}^{|\mathcal{L}|}\Pr(\ell'|\ell)=1$. If the

$$\sum_{x,y_{11},\ldots,y_{|\mathcal{L}||\mathcal{K}|}} p_X(x)p(y_{11}|x)\ldots p(y_{|\mathcal{L}||\mathcal{K}|}|x)$$
$$\left[\frac{1}{|\mathcal{L}|}\sum_{\ell=1}^{|\mathcal{L}|}\frac{1}{2}\left\|(\mathrm{Tr}_{\hat{K}}\circ\mathcal{D}'^{2}_{B\to\hat{L}\hat{K}})\left(\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma_B^{x,m,y_{\ell k}}\right)-|\ell\rangle\langle\ell|_{\hat{L}}\right\|_1\right]$$
$$=\sum_{x,y_{11},\ldots,y_{|\mathcal{L}||\mathcal{K}|}} p_X(x)p(y_{11}|x)\ldots p(y_{|\mathcal{L}||\mathcal{K}|}|x)$$
$$\left[\frac{1}{|\mathcal{L}|}\sum_{\ell=1}^{|\mathcal{L}|}\frac{1}{2}\left\|\mathcal{D}'^{2}_{B\to\hat{L}}\left(\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma_B^{x,m,y_{\ell k}}\right)-|\ell\rangle\langle\ell|_{\hat{L}}\right\|_1\right]\leq\varepsilon. \quad (4.39)$$

---

trace above was only applied to the B system, we would have :

$$\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sum_{k'=1}^{|\mathcal{K}|}\mathrm{Tr}_B\{\Omega_B^{x,y_{\ell',k'}}\sigma_{BE}^{x,m,y_{\ell,k}}\}=\Pr(\ell'|\ell)u_E^{\ell',\ell},$$

where

$$u_E^{\ell',\ell}:=\frac{\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sum_{k'=1}^{|\mathcal{K}|}\mathrm{Tr}_B\{\Omega_B^{x,y_{\ell',k'}}\sigma_{BE}^{x,m,y_{\ell,k}}\}}{\Pr(\ell'|\ell)}.$$

And by summing up over all ℓ' we get: (see that $\sum_{k'=1}^{|\mathcal{K}|}\sum_{\ell'=1}^{|\mathcal{L}|}\mathrm{Tr}_B\{\Omega_B^{x,y_{\ell',k'}}\sigma_{BE}^{x,m,y_{\ell,k}}\}=\sigma_E^{x,m,y_{\ell,k}}$ for a given pair (ℓ,k))

$$\sum_{\ell'=1}^{|\mathcal{L}|}\Pr(\ell'|\ell)u_E^{\ell',\ell}=\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma_E^{x,m,y_{\ell,k}}.$$

Hence, the following equation follows:

$$\mathcal{D}'^{2}_{B\to\hat{L}}\left(\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma_{BE}^{x,m,y_{\ell,k}}\right)=\sum_{\ell'=1}^{|\mathcal{L}|}\Pr(\ell'|\ell)|\ell'\rangle\langle\ell'|_{\hat{L}}\otimes u_E^{\ell',\ell},$$

and by tracing out Eve's system:

$$\mathcal{D}'^{2}_{B\to\hat{L}}\left(\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma_B^{x,m,y_{\ell,k}}\right)=\sum_{\ell'=1}^{|\mathcal{L}|}\Pr(\ell'|\ell)|\ell'\rangle\langle\ell'|_{\hat{L}}.$$

We move forward with the chain of inequalities ending up in (4.40), where the first inequality follows from the convexity of the trace distance and the second equality emerges because of the invariance of the trace distance with

$$
\begin{aligned}
&\frac{1}{|\mathcal{L}|}\sum_{l=1}^{|\mathcal{L}|}\frac{1}{2}\left\|\mathcal{D}'^{2}_{B\to\hat{L}}\left(\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma_{BE}^{x,m,y_{\ell,k}}\right)-|\ell\rangle\langle\ell|_{\hat{L}}\otimes\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma_{E}^{x,m,y_{\ell,k}}\right\|_1\\
&=\frac{1}{|\mathcal{L}|}\sum_{\ell=1}^{|\mathcal{L}|}\frac{1}{2}\left\|\sum_{\ell'=1}^{|\mathcal{L}|}\Pr(\ell'|\ell)|\ell'\rangle\langle\ell'|_{\hat{L}}\otimes u_E^{\ell',\ell}-|\ell\rangle\langle\ell|_{\hat{L}}\otimes\sum_{\ell'=1}^{|\mathcal{L}|}\Pr(\ell'|\ell)u_E^{\ell',\ell}\right\|_1\\
&\leq\frac{1}{|\mathcal{L}|}\sum_{l=1}^{|\mathcal{L}|}\sum_{\ell'=1}^{|\mathcal{L}|}\Pr(\ell'|\ell)\left[\frac{1}{2}\||\ell'\rangle\langle\ell'|_{\hat{L}}\otimes u_E^{\ell',\ell}-|\ell\rangle\langle\ell|_{\hat{L}}\otimes u_E^{\ell',\ell}\|_1\right]\\
&=\frac{1}{|\mathcal{L}|}\sum_{\ell=1}^{|\mathcal{L}|}\sum_{\ell'=1}^{|\mathcal{L}|}\Pr(\ell'|\ell)\left[\frac{1}{2}\||\ell'\rangle\langle\ell'|_{\hat{L}}-|\ell\rangle\langle\ell|_{\hat{L}}\|_1\right]\\
&=\frac{1}{|\mathcal{L}|}\sum_{\ell=1}^{|\mathcal{L}|}\sum_{\ell'\neq\ell}\Pr(\ell'|\ell)=\frac{1}{|\mathcal{L}|}\sum_{\ell=1}^{|\mathcal{L}|}\frac{1}{2}\left\|\mathcal{D}'^{2}_{B\to\hat{L}}\left(\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma_{B}^{x,m,y_{\ell,k}}\right)-|\ell\rangle\langle\ell|_{\hat{L}}\right\|_1, \quad (4.40)
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{x,y_{1,1},\ldots,y_{|\mathcal{L}|,|\mathcal{K}|}}p_X(x)p(y_{1,1}|x)\ldots p(y_{|\mathcal{L}|,|\mathcal{K}|}|x)\\
&\left[\frac{1}{|\mathcal{L}|}\sum_{\ell=1}^{|\mathcal{L}|}\frac{1}{2}\left\|\mathcal{D}'^{2}_{B\to\hat{L}}\left(\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma_{BE}^{x,m,y_{\ell,k}}\right)-|\ell\rangle\langle\ell|_{\hat{L}}\otimes\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma_{E}^{x,m,y_{\ell,k}}\right\|_1\right]\leq\varepsilon. \quad (4.41)
\end{aligned}
$$

respect to tensor-product states. This result together with (4.39) leads to the inequality given by (4.41). This is equivalent to the criterion dealing with Bob's error in detecting the private message. We continue by expanding Eve's security condition as given in (4.42), where the last equality comes about by using the invariance of trace distance with respect to tensor-product states.

We deal with two important expressions in (4.41) and (4.42), the former is Bob's error in detecting the private message and the later is the security of Eve. Now it is time to unify two criteria into the so-called privacy error. To this end, let's consider (4.41) and (4.42) together with their imposed bounds on the cardinalities of $|\mathcal{L}|$ and $|\mathcal{K}|$. We employ triangle inequality for the trace distance to merge them into the privacy error as given in (4.43) (remember that in the assisted code, there is no difference between average and individual error probabilities). This immediately implies the privacy criterion given in (4.11) in the sense that if this holds, the single criterion in (4.11) also holds.

We are now done with the assisted code. As we proceed to derandomize

$$
\begin{aligned}
&\frac{1}{2}\left\|\sigma^{m,\ell}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}E} - \sum_x p_X(x)|x\rangle\langle x|_X \otimes \sigma^{x,m}_{Y^{\otimes|\mathcal{L}||\mathcal{K}|}} \otimes \tilde{\sigma}^{x,m}_E\right\|_1 \\
&= \frac{1}{2}\left\|\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sum_{x,y_{11}\ldots y_{|\mathcal{L}||\mathcal{K}|}} p_X(x)p_{Y|X}(y_{11}|x)\ldots p_{Y|X}(y_{|\mathcal{L}||\mathcal{K}|}|x)\right. \\
&\qquad\left. |x\rangle\langle x|_X \otimes |y_{11}\ldots y_{|\mathcal{L}||\mathcal{K}|}\rangle\langle y_{11}\ldots y_{|\mathcal{L}||\mathcal{K}|}|_{Y^{|\mathcal{L}||\mathcal{K}|}} \otimes (\sigma^{x,m,y_{\ell k}}_E - \tilde{\sigma}^{x,m}_E)\right\|_1 \\
&= \frac{1}{2}\left\|\sum_{x,y_{11}\ldots y_{\ell k}} p_X(x)p_{Y|X}(y_{11}|x)\ldots p_{Y|X}(y_{|\mathcal{L}|,|\mathcal{K}|}|x)\right. \\
&\qquad\left. |x\rangle\langle x|_X \otimes |y_{11}\ldots y_{|\mathcal{L}||\mathcal{K}|}\rangle\langle y_{11}\ldots y_{|\mathcal{L}|,|\mathcal{K}|}|_{Y^{|\mathcal{L}||\mathcal{K}|}} \otimes \left(\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma^{x,m,y_{\ell,k}}_E - \tilde{\sigma}^{x,m}_E\right)\right\|_1 \\
&= \sum_{x,y_{11}\ldots y_{|\mathcal{L}||\mathcal{K}|}} p_X(x)p_{Y|X}(y_{11}|x)\ldots p_{Y|X}(y_{|\mathcal{L}||\mathcal{K}|}|x)\left[\frac{1}{2}\left\|\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma^{x,m,y_{\ell k}}_E - \tilde{\sigma}^{x,m}_E\right\|_1\right] \\
&= \sum_{x,y_{11}\ldots y_{|\mathcal{L}||\mathcal{K}|}} p_X(x)p_{Y|X}(y_{11}|x)\ldots p_{Y|X}(y_{|\mathcal{L}||\mathcal{K}|}|x) \\
&\qquad\left[\frac{1}{2}\left\||\ell\rangle\langle\ell|_{L'} \otimes \frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma^{x,m,y_{\ell k}}_E - |\ell\rangle\langle\ell|_{L'} \otimes \tilde{\sigma}^{x,m}_E\right\|_1\right] \leq \sqrt{\varepsilon'}, \qquad (4.42)
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{x,y_{1,1},\ldots,y_{|\mathcal{L}|,|\mathcal{K}|}} p_X(x)p_{Y|X}(y_{1,1}|x)\ldots p_{Y|X}(y_{|\mathcal{L}|,|\mathcal{K}|}|x) \\
&\qquad\left[\frac{1}{2}\left\|\mathcal{D}'^2_{B\to\hat{L}}\left(\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma^{x,m,y_{\ell,k}}_{BE}\right) - |\ell\rangle\langle\ell|_{\hat{L}} \otimes \tilde{\sigma}^{x,m}_E\right\|_1\right] \leq \varepsilon + \sqrt{\varepsilon'}, \quad (4.43)
\end{aligned}
$$

---

the code, it will be clear that the procedure employed to unify two error criteria is helpful. Before we proceed to derandomize the code, we would like to consider two extra error terms. The error probability of the second decoder depends on the error probability of the first decoder in two directions, first, the second decoder is fed a state close to the actual received state and second, the second decoder applies a quantum instrument depending on the estimate of the transmitted message m. This can, without losing the generality, be written as follows:

$$
\mathcal{D}^2_{\hat{M}BY\to\hat{L}} \coloneqq \sum_m |m\rangle\langle m|_{M'} \otimes \mathcal{D}^{2,m}_{BY\to\hat{L}}.
$$

$$\left\|(\sum_m |m\rangle\langle m|_{M'} \otimes \mathcal{D}^{2,m}_{BY\to\hat{L}})(\mathcal{D}^1_{BX\to\hat{M}B}(\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B})) - (|m\rangle\langle m|_{M'} \otimes \mathcal{D}^{2,m}_{BY\to\hat{L}})(\mathcal{D}^1_{BX\to\hat{M}B}(\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}))\right\|_1$$

$$\begin{aligned}
&= \left\| \sum_{m'\neq m} |m'\rangle\langle m'|_{M'} \otimes \mathcal{D}^{2,m'}_{BY\to\hat{L}}(\mathcal{D}^{1,m'}_{BX\to B}(\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}))\right\|_1 \\
&\leq \sum_{m'\neq m} \left\| |m'\rangle\langle m'|_{M'} \otimes \mathcal{D}^{2,m'}_{BY\to\hat{L}}(\mathcal{D}^{1,m'}_{BX\to B}(\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B}))\right\|_1 \\
&\leq \sum_{m'\neq m} \left\| \mathcal{D}^{1,m'}_{BX\to B}(\rho^{m,(\ell,k)}_{X^{\otimes|\mathcal{M}|}B})\right\|_1 \leq \varepsilon. \qquad (4.44)
\end{aligned}$$

In the following we show how this fact contributes to the error probability. First one is the difference between the received state and the disturbed state being fed into the second decoder. Since the probability of error of the first decoder is at most ε, we know from gentle measurement lemma that:

$$\left\|\sigma^{m,(\ell,k)}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}B} - \rho^{m,(\ell,k)}_{XY^{\otimes|\mathcal{L}||\mathcal{K}|}B}\right\|_1 \leq 2\sqrt{\varepsilon},$$

and for the second term we have the chain of inequalities given by (4.44); where the equality follows from the the observation in (4.8), the first and second inequalities follow the convexity and monotonicity of trace distance, respectively. Adding these two terms to (4.43) will result in $P_{priv} \leq 2(\varepsilon + \sqrt{\varepsilon}) + \sqrt{\varepsilon'}$.

Derandomization

We can now fix the classical registers and obtain a protocol without shared randomness, i.e., derandomize the code. The derandomization is a standard technique and its mathematical details are given in the appendix.

4.5 Converse

In this section we give upper bounds for the capacity region $\mathcal{C}^{\varepsilon,\varepsilon'}(\mathcal{N})$.

Proof of theorem (4.3). Two messages $m \in \mathcal{M}$ and $\ell \in \mathcal{L}$ are sent through the channel $\mathcal{N}_{A\to BE}$ and their estimates are $\hat{M}$ and $\hat{L}$, respectively. From definition (11), an $(\varepsilon,\varepsilon')$-code satisfies $\Pr(M \neq \hat{M}) \leq \varepsilon$. A hypothesis testing problem can be associated to the problem of detecting m leading to an expression for the error probability of the public message. To see how it

works out, consider a binary hypothesis testing problem in which null and alternative hypothesis are

$$\rho_{MM'} = \frac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|} |m\rangle\langle m|_M \otimes |m\rangle\langle m|_{M'} \quad \text{and}$$
$$\rho_M \otimes \rho_{M'} = \frac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|} |m\rangle\langle m|_M \otimes \frac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|} |m\rangle\langle m|_{M'},$$

respectively. It is easily seen that type I error, i.e., deciding in favor of $\rho_M \otimes \rho_{M'}$ while the true state was $\rho_{MM'}$, is exactly equal to the error probability $\Pr(M \neq \hat{M})$ which is less than or equal to ε by assumption. On the other hand, type II error, deciding $\rho_{MM'}$ on $\rho_M \otimes \rho_{M'}$, equals $\frac{1}{|\mathcal{M}|}$ (the distribution over message set is uniform). Then from the definition of the hypothesis testing mutual information, we have the following:

$$r \leq I_{\mathrm{H}}^{\varepsilon}(M;M')_{\rho}.$$

where $r = \log|\mathcal{M}|$ is the rate of the public message. Furthermore, from the quantum DPI, we have:

$$I_{\mathrm{H}}^{\varepsilon}(M;M')_{\rho} \leq I_{\mathrm{H}}^{\varepsilon}(M;B)_{\rho}$$

Finally, using the injectivity of the encoder, we define a random variable X whose distribution is built by projecting the distribution of M on its image on X and zero otherwise. Setting $X = M$, we get the following:

$$r \leq I_{\mathrm{H}}^{\varepsilon}(X;B)_{\rho}.$$

In regards to the private rate $R = \log|\mathcal{L}|$, consider the following chain of inequalities:

$$\begin{aligned}
\varepsilon &\geq Pr\{(M,L) \neq (\hat{M},\hat{L})\} \\
&= \sum_{m,\ell} p(m)p(\ell) \sum_{(\hat{m},\hat{\ell})\neq(m,\ell)} p(\hat{m},\hat{\ell}|m,\ell) \\
&\geq \sum_{m,\ell} p(m)p(\ell) \sum_{\hat{\ell}\neq\ell} p(\hat{\ell}|m,\ell) \\
&= \sum_{m} p(m)\Pr(\hat{L} \neq L|M=m),
\end{aligned}$$

where the first line is due to the assumption. From Markov's inequality, we know that with probability at least $1-\sqrt{\varepsilon}$, the following holds for a randomly generated $m \in \mathcal{M}$:

$$\Pr(\hat{L} \neq L|M=m) \leq \sqrt{\varepsilon}.$$

Then following the same strategy as for the public rate, we consider a binary hypothesis testing problem distinguishing between $\rho^m_{L\hat{L}}$ and $\rho^m_L \otimes \rho^m_L$ conditioned on previously specified m, we will have:

$$R \leq I_{\mathrm{H}}^{\sqrt{\varepsilon}}(L;\hat{L}|M=m)_\rho.$$

Then, to get $\tilde{I}_{\mathrm{H}}^{\sqrt{\varepsilon}}(L;\hat{L}|M)_\rho$, according to Definition 9, we can optimize the expression with respect to ρ'_M where $P(\rho_M,\rho'_M) \leq \sqrt{\varepsilon}$. Then from the monotonicity of the hypothesis-testing relative entropy applied to $\hat{L}$ system, we have:

$$\tilde{I}_{\mathrm{H}}^{\sqrt{\varepsilon}}(L;\hat{L}|M)_\rho \leq \tilde{I}_{\mathrm{H}}^{\sqrt{\varepsilon}}(L;B|M)_\rho.$$

By the same argument that we defined $X \coloneqq M$, we also define $Y \coloneqq L$ and so the following results:

$$R \leq \tilde{I}_{\mathrm{H}}^{\sqrt{\varepsilon}}(Y;B|X)_\rho. \tag{4.45}$$

On the other hand, from the secrecy condition (4.3), we know that for every m, the following is true:

$$\frac{1}{2}\|\rho^m_{LE} - \rho_L \otimes \tilde{\rho}^m_E\|_1 \leq \varepsilon',$$

and from the relation between the purified distance and the trace distance it holds that:

$$P(\rho^m_{LE}, \rho_L \otimes \tilde{\rho}^m_E) \leq \sqrt{2\varepsilon'}.$$

From the definition of the smooth max-relative entropy we see that

$$D_{\max}^{\sqrt{2\varepsilon'}}(\rho^m_{LE}, \rho_L \otimes \tilde{\rho}^m_E) = 0$$

And by considering the optimization in Definition 10 over ρ'_M such that $P(\rho'_M,\rho_M) \leq \sqrt{\varepsilon'}$, we have $\tilde{I}_{\max}^{\sqrt{2\varepsilon'}}(L;E|M) = 0$. Setting $M \coloneqq X$ and $L \coloneqq Y$ as before and plugging into (4.45), the following bound on the private rate holds:

$$R \leq \tilde{I}_{\mathrm{H}}^{\sqrt{\varepsilon}}(Y;B|X) - \tilde{I}_{\max}^{\sqrt{2\varepsilon'}}(Y;E|X). \tag{4.46}$$

■

4.6 Asymptotic Analysis

We evaluate our rate region in the asymptotic limit of many uses of a memoryless channel. The capacity theorem for simultaneous transmission of classical and quantum information was proved by Devetak and Shor [20]. In this section, we recover their result from our theorems. We define the rate region of the simultaneous transmission of the classical and quantum information as follows:

$$\mathcal{C}_\infty(\mathcal{N}) \coloneqq \lim_{\varepsilon,\varepsilon'\to 0} \lim_{n\to\infty} \frac{1}{n} \mathcal{C}^{\varepsilon,\varepsilon'}(\mathcal{N}^{\otimes n}).$$

Let $\mathcal{C}(\mathcal{N})$ be the set of rate pairs (r', R') such that

$$\begin{aligned} r' \leq & I(X;B)_\rho, \\ R' \leq & I\left(Y;B|X\right)_\rho - I\left(Y;E|X\right)_\rho \end{aligned}$$

where all the entropic quantities are computed over all $\rho_{XYBE} \coloneqq \sum_{x,y} p(x,y)\,|x\rangle\langle x| \otimes |y\rangle\langle y| \otimes \mathcal{N}_{A\to BE}(\rho_A^{x,y})$ arising from the channel. Then the capacity region $\mathcal{C}_\infty(\mathcal{N})$ is the (normalized) union over ℓ uses of the channel $\mathcal{N}$ as below:

$$\mathcal{C}_\infty(\mathcal{N}) = \bigcup_{\ell=1}^{\infty} \frac{1}{\ell} \mathcal{C}(\mathcal{N}^{\otimes \ell}). \tag{4.47}$$

In the rest of this section, our aim is to prove the capacity region above. Before doing so, we slightly modify the expression for the private rate in Theorem 4.2 by using Fact 2. Note that Fact 2 deals with unconditional expressions, however, conditional expressions are trivial noting their definitions. Therefore, the following holds:

$$\begin{aligned} \tilde{I}_{\max}^{\sqrt{\varepsilon'}-\delta'}(Y;E|X) \leq \tilde{I}_{\max}^{\sqrt{\varepsilon'}-\delta'-\gamma}(Y;E|X) \\ + \log_2\left(\frac{3}{\gamma^2}\right), \end{aligned}$$

where $\gamma \in (0, \sqrt{\varepsilon'} - \delta')$. And so the achievability of the private rate appears as follows:

$$\begin{aligned} R \geq \tilde{I}_{\mathrm{H}}^{\varepsilon-\delta}\left(Y;B|X\right) - \tilde{I}_{\max}^{\sqrt{\varepsilon'}-\delta'-\gamma}(Y;E|X) \\ - \log_2(\frac{4\varepsilon}{\delta^2}) - 2\log_2(\frac{1}{\delta'}) - \log_2\left(\frac{3}{\gamma^2}\right). \end{aligned}$$

Like all capacity theorems, the proof of the aforementioned capacity region is accomplished in two steps, direct part that we show all such rates

are achievable, i.e., the right-hand side of the equation (4.47) is contained ($\subseteq$) inside $\mathcal{C}_\infty(\mathcal{N})$ and the converse part that goes in the opposite direction saying that those rates cannot be exceeded, i.e., $\mathcal{C}_\infty(\mathcal{N})$ is contained inside the union on the right-hand side of (4.47).

For the direct part, we use our one-shot lower bounds on the capacity region and apply quantum AEP for the (conditional) smooth hypothesis testing- and max-mutual information. From Theorem 4.2, for m uses of the channel $\mathcal{N}$ (or as one may like to think of it, one use of the superchannel $\mathcal{N}^{\otimes m}$), the following lower bound on the capacity region $\mathcal{C}^{\varepsilon,\varepsilon'}(\mathcal{N}^{\otimes m})$ can be seen:

$$\bigcup_{\ell=1}^{m} \mathcal{C}_a(\mathcal{N}^{\otimes \ell}) \subseteq \mathcal{C}^{3\varepsilon+2\sqrt{\varepsilon}+\sqrt{\varepsilon'},2(\varepsilon+\sqrt{\varepsilon})+\sqrt{\varepsilon'}}(\mathcal{N}^{\otimes m}),$$

where $\mathcal{C}_a(\mathcal{N}^{\otimes \ell})$ is the set of all twins (r', R') satisfying:

$$\begin{aligned} r' &\leq I_{\mathrm{H}}^{\varepsilon-\delta}\left(X^\ell; B^{\otimes \ell}\right) - \log_2\left(\frac{4\varepsilon}{\delta^2}\right), \\ R' &\leq \tilde{I}_{\mathrm{H}}^{\varepsilon-\delta}\left(Y^\ell; B^{\otimes \ell}|X^\ell\right) - \tilde{I}_{\max}^{\sqrt{\varepsilon'}-\delta'-\gamma}(Y^\ell; E^{\otimes \ell}|X^\ell) \\ &\qquad - \log_2\left(\frac{4\varepsilon}{\delta^2}\right) - 2\log_2\left(\frac{1}{\delta'}\right) - \log_2\left(\frac{3}{\gamma^2}\right). \end{aligned}$$

Since the region above is basically a lower bound on the capacity region, we are free to assume that the sequences of the random variables are generated in an *i.i.d.* fashion according to the corresponding distributions. This empowers us to make use of quantum AEP as described below. From Fact 1 we have

$$\lim_{\varepsilon\to 0}\lim_{m\to\infty}\frac{1}{m} I_{\mathrm{H}}^{\varepsilon-\delta}\left(X^m; B^{\otimes m}\right) = I(X;B).$$

Likewise, applying Lemma 4.1 and Lemma 4.2 give rise respectively to the following identities:

$$\begin{aligned} \lim_{\varepsilon\to 0}\lim_{m\to\infty}\frac{1}{m}\tilde{I}_{\mathrm{H}}^{\varepsilon-\delta}\left(Y^m; B^{\otimes m}|X^m\right) &= I(Y;B|X), \\ \lim_{\varepsilon'\to 0}\lim_{m\to\infty}\frac{1}{m}\tilde{I}_{\max}^{\sqrt{\varepsilon'}-\delta'-\gamma}(Y^m; E^{\otimes m}|X^m) &= I(Y;E|X). \end{aligned}$$

Plugging back into the respective equations, we obtain

$$\mathcal{C}(\mathcal{N}) \subseteq \lim_{\varepsilon,\varepsilon'\to 0}\lim_{m\to\infty}\frac{1}{m}\mathcal{C}^{\varepsilon,\varepsilon'}(\mathcal{N}^{\otimes m}),$$

where $\mathcal{C}(\mathcal{N})$, as defined before, consists of rate pairs (r', R') satisfying

$$\begin{aligned} r' \leq & I(X;B)_\rho, \\ R' \leq & I(Y;B|X)_\rho - I(Y;E|X)_\rho. \end{aligned}$$

Last step of the direct part is to consider a superchannel $\mathcal{N}^{\otimes \ell}$ (ℓ independent uses of the channel $\mathcal{N}$) and let $n = m\ell$ and repeat the above argument, i.e., use the superchannel m times. Finally by letting $n \to \infty$ and evaluating the union of the regions, we obtain the desired result.

To prove the converse, we consider our upper bounds given in Theorem 4.3 in the case of n uses of the channel $\mathcal{N}$ and we have:

$$\mathcal{C}^{\varepsilon,\varepsilon'}(\mathcal{N}^{\otimes n}) \subseteq \bigcup_{n=1}^{\infty} \mathcal{C}_c(\mathcal{N}^{\otimes n})$$

where $\mathcal{C}_c(\mathcal{N}^{\otimes n})$ includes all ordered twins (r', R') satisfying

$$r' \leq I_{\mathrm{H}}^{\varepsilon}\left(X^n; B^{\otimes n}\right), \tag{4.48}$$

$$R' \leq \tilde{I}_{\mathrm{H}}^{\varepsilon}(Y^n; B^{\otimes n}|X^n) - \tilde{I}_{\max}^{\sqrt{2\varepsilon'}}\left(E^{\otimes n}; Y^n|X^n\right). \tag{4.49}$$

To upper bound right-hand side of (4.48) we apply Fact 1.6. The first term on the right-hand side of (4.49) can be upper bounded by making use of lemma (4.3) and for the second term, we use Lemma 4.4 replacing $|A|$ with $|\mathcal{Y}|^n$. The inequalities are as follows:

$$\begin{aligned} r' \leq & \frac{1}{1-\varepsilon}\left(I(X^n; B^{\otimes n}) + h_b(\varepsilon)\right), \\ R' \leq & \frac{1}{1-\varepsilon}\left(I(Y^n; B^{\otimes n}|X^n) + h_b(\varepsilon)\right) - I\left(E^{\otimes n}; Y^n|X^n\right) \\ & + 2n\sqrt{2\varepsilon'}\log|\mathcal{Y}| + 2(1+\sqrt{2\varepsilon'})h_b\left(\frac{\sqrt{2\varepsilon'}}{1+\sqrt{2\varepsilon'}}\right). \end{aligned}$$

Multiplying by $\frac{1}{n}$ and taking the limits $n \to \infty$ and $\varepsilon, \varepsilon' \to 0$, (changing n with ℓ) the desired result is achieved.

4.6.1 Private information to coherent information

Here we argue that the private rate that has been given in terms of the difference between two mutual-information like quantities, is in principle, the coherent information appearing in [20]. To see how this plays out, consider an ensemble of quantum states $\mathcal{E} = \{p_X(x), |\phi^x\rangle_{RA}\}_{x\in\mathcal{X}}$ where X is a random variable with alphabet $\mathcal{X}$ and distribution $p_X(x)$ and A and R are quantum

systems such that R plays the role of a reference system. Assuming an auxiliary classical system $\sigma_X = \sum_x p_X(x)|x\rangle\langle x|_X$, the following state can be associated to the ensemble:

$$\sigma_{XRA} = \sum_x p_X(x)|x\rangle\langle x|_X \otimes |\phi^x\rangle\langle\phi^x|_{RA}. \tag{4.50}$$

If channel $\mathcal{N}_{A\to BE}$ acts on this state, we get the following *coherent* state:

$$\mathcal{N}_{A\to BE}(\sigma_{XRA}) = \sum_x p_X(x)|x\rangle\langle x|_X \otimes |\phi^x\rangle\langle\phi^x|_{RBE},$$

and the conditional coherent information on it, is evaluated as follows:

$$\begin{aligned} I(R\rangle BX) &:= -H(R|BX) = H(B|X) - H(RB|X) \\ &\overset{(a)}{=} H(B|X) - H(E|X), \end{aligned}$$

where (a) follows from the fact that the state $|\phi^x\rangle\langle\phi^x|_{RBE}$ is a pure state (conditioned on X).

We proceed with applying the Schmidt decomposition to the pure states $\{|\phi^x\rangle_{RBE}\}_{x\in\mathcal{X}}$ with respect to the cut $R|BE$. Let $\{|y^x\rangle_R\}$ and $|\psi^{x,y}\rangle_{BE}$ be orthonormal bases for R and BE systems. Then from Schmidt decomposition we have that

$$|\phi^x\rangle_{RBE} = \sum_y \sqrt{p_{Y|X}(y|x)}|y^x\rangle_R \otimes |\psi^{x,y}\rangle_{BE}.$$

We want to get a decoherent version of the state $|\phi^x\rangle_{RBE}$ by measuring the R system in the basis $\{|y^x\rangle_R\}$. Since after the measurement, R system becomes a classical system, hereafter we show it by Y. Let $|\bar{\phi}^x\rangle_{YBE}$ denote the decoherent state resulting from the measurement, then

$$\bar{\phi}^x_{YBE} = \sum_y p_{Y|X}(y|x)|y^x\rangle\langle y^x|_Y \otimes |\psi^{x,y}\rangle\langle\psi^{x,y}|_{BE},$$

and let the decoherent state $\bar{\sigma}_{XRBE}$ be as follow:

$$\begin{aligned} \bar{\sigma}_{XYBE} = &\sum_x p_X(x)|x\rangle\langle x|_X \\ &\otimes \sum_y p_{Y|X}(y|x)|y^x\rangle\langle y^x|_Y \otimes |\psi^{x,y}\rangle\langle\psi^{x,y}|_{BE}. \end{aligned}$$

This state is the same as was held by Bob and Eve after decoding for the public message. If the correctness of the following equality can be proven,

which turns out to be straightforward, we can argue about the correctness of our claim,

$$I(R\rangle BX)_\sigma = I(Y;B|X)_{\bar{\sigma}} - I(Y;E|X)_{\bar{\sigma}}. \tag{4.51}$$

The right-hand side of (4.51) can be expanded as follow:

$$\begin{aligned} & I(Y;B|X)_{\bar{\sigma}} - I(Y;E|X)_{\bar{\sigma}} \\ & \overset{(a)}{=} H(B|X) - H(B|X,Y) - H(E|X) + H(E|X,Y) \\ & \overset{(b)}{=} H(B|X) - H(E|X), \end{aligned}$$

where (a) follows by the definition of the conditional mutual information and (b) is due to the fact that conditioned on X and Y, the state on BE is a pure state. Observe the last expression is a function solely of the density operator given in (4.50). It is evident that for the regularized formula, we consider n-fold states in our proof instead. This proves our claim.

4.7 Conclusion

We studied the one-shot capacity of a quantum channel for simultaneous transmission of classical and quantum information. Our main tools are position-based decoding and convex-split lemma. We first consider the problem of simultaneous transmission of public and private classical information and then we discussed that the private rate can be translated into quantum capacity. We also provided converse bounds. By evaluating our achievability and converse bounds in asymptotic i.i.d. regime, we recovered the well-known results in the literature.

Appendix

Derandomization of the code

We aim to derandomize the assisted code. As mentioned in the introductory section, this development follows the procedure used in [100] and [125]. We start with the public message. We saw that the optimal operator Π_{XB} is such that $\mathrm{Tr}\{\Pi_{XB}\rho_{XB}\} \geq 1-(\varepsilon-\delta)$ and $\mathrm{Tr}\{\Pi_{XB}(\rho_X \otimes \rho_B)\} = 2^{-I_{\mathrm{H}}^{\varepsilon-\delta}(X;B)_\rho}$, we rewrite the two error types with slightly different notations as follows :

$$\mathrm{Tr}\{\Pi_{XB}\rho_{XB}\} = \mathrm{Tr}\left\{\Pi_{XB}\sum_x p_X(x)|x\rangle\langle x|_X \otimes \rho_B^x\right\}$$

$$\rho^{m,(l,k)}_{X^{\otimes|\mathcal{M}|}B} \coloneqq \rho^{1}_{X} ... \otimes \rho^{m,(l,k)}_{XB} \otimes ... \otimes \rho^{|\mathcal{M}|}_{X} = \sum_{x_1,...,x_{|\mathcal{M}|}} p_X(x_1)...p_X(x_{|\mathcal{M}|}) |x_1....x_{|\mathcal{M}|}\rangle\langle x_1...x_{|\mathcal{M}|}|_{X_1...X_{|\mathcal{M}|}} \otimes \rho^{m,(l,k)}_{B}. \tag{4.52}$$

$$\begin{aligned}\Gamma^{m}_{X^{|\mathcal{M}|}B} &= \mathbb{1}^{1}_{X} \otimes ... \otimes T^{m}_{XB} \otimes ... \otimes \mathbb{1}^{|\mathcal{M}|}_{X} \\ &= \sum_{x_1} |x_1\rangle\langle x_1|_X \otimes ... \otimes \left(\sum_{x_m} |x_m\rangle\langle x_m|_X \otimes W^{x_m}_{B} \right) \otimes ... \otimes \sum_{x_{|\mathcal{M}|}} |x_{|\mathcal{M}|}\rangle\langle x_{|\mathcal{M}|}|_X \\ &= \sum_{x_1...x_{|\mathcal{M}|}} |x_1...x_{|\mathcal{M}|}\rangle\langle x_1...x_{|\mathcal{M}|}|_X \otimes W^{x_m}_{B}. \end{aligned} \tag{4.53}$$

$$\begin{aligned} &= \sum_x p_X(x) \mathrm{Tr}\{\langle x|\Pi_{XB}|x\rangle_X \rho^x_B\} \\ &= \sum_x p_X(x) \mathrm{Tr}\{W^x_B \rho^x_B\}, \end{aligned}$$

in which the operator W^x_B is defined as $W^x_B \coloneqq \langle x|\Pi_{XB}|x\rangle_X$. In an analogous way, we have that

$$\begin{aligned} \mathrm{Tr}\{\Pi_{XB}(\rho_B \otimes \rho_X)\} &= \mathrm{Tr}\left\{\Pi_{XB} \sum_x p_X(x)|x\rangle\langle x|_X \otimes \rho_B\right\} \\ &= \sum_x p_X(x) \mathrm{Tr}\{\langle x|\Pi_{XB}|x\rangle_X \rho_B\} \\ &= \sum_x p_X(x) \mathrm{Tr}\{W^x_B \rho_B\}. \end{aligned}$$

These expressions imply that it is sufficient to take the optimal test to be $\Pi_{XB} = \sum_x |x\rangle\langle x|_X \otimes W^x_B$ with aforementioned W^x_B; In other words, the test Π_{XB} can achieve the same error probability as any other Π_{XB} would do. We proceed with dissecting each term involved in $\mathrm{Tr}\{(\mathbb{1}_{X^{|\mathcal{M}|}B} - \Lambda^m_{X^{|\mathcal{M}|}B})\rho^{m,(l,k)}_{X^{\otimes|\mathcal{M}|}B}\}$ where $\rho^{m,(l,k)}_{X^{\otimes|\mathcal{M}|}B}$ is given in (4.52).

By assuming the particular structure for the optimal test operator that we just introduced, the operator $\Gamma^m_{X^{|\mathcal{M}|}B}$ appears as given in (4.53). And

$$
\begin{aligned}
&\left(\sum_{m'=1}^{|\mathcal{M}|}\Gamma^{m'}_{X^{|\mathcal{M}|}B}\right)^{-\frac{1}{2}}\\
&=\left(\sum_{m'=1}^{|\mathcal{M}|}\sum_{x_1\ldots x_{|\mathcal{M}|}}|x_1\ldots x_{|\mathcal{M}|}\rangle\langle x_1\ldots x_{|\mathcal{M}|}|_X\otimes W_B^{x_{m'}}\right)^{-\frac{1}{2}}\\
&=\left(\sum_{x_1\ldots x_{|\mathcal{M}|}}|x_1\ldots x_{|\mathcal{M}|}\rangle\langle x_1\ldots x_{|\mathcal{M}|}|_X\otimes\sum_{m'=1}^{|\mathcal{M}|}W_B^{x_{m'}}\right)^{-\frac{1}{2}}\\
&=\sum_{x_1\ldots x_{|\mathcal{M}|}}|x_1\ldots x_{|\mathcal{M}|}\rangle\langle x_1\ldots x_{|\mathcal{M}|}|_X\otimes\left(\sum_{m'=1}^{|\mathcal{M}|}W_B^{x_{m'}}\right)^{-\frac{1}{2}},
\end{aligned}
$$

and finally

$$
\begin{aligned}
\Lambda^m_{X^{|\mathcal{M}|}B}&=\left(\sum_{m'=1}^{|\mathcal{M}|}\Gamma^m_{X^{|\mathcal{M}|}B}\right)^{-\frac{1}{2}}\Gamma^{m'}_{X^{|\mathcal{M}|}B}\left(\sum_{m'=1}^{|\mathcal{M}|}\Gamma^{m'}_{X^{|\mathcal{M}|}B}\right)^{-\frac{1}{2}}\\
&=\sum_{x_1\ldots x_{|\mathcal{M}|}}|x_1\ldots x_{|\mathcal{M}|}\rangle\langle x_1\ldots x_{|\mathcal{M}|}|_X\otimes\Delta^m_B,
\end{aligned}
$$

where

$$
\Delta^m_B\coloneqq\left(\sum_{m'=1}^{|\mathcal{M}|}W_B^{x_{m'}}\right)^{-\frac{1}{2}}W_B^{x_m}\left(\sum_{m'=1}^{|\mathcal{M}|}W_B^{x_{m'}}\right)^{-\frac{1}{2}}.
$$

Note that the obtained POVM, $\{\Delta^m_B\}_{m=1}^{|\mathcal{M}|}$, can be completed by adding $\Delta^0_B=\mathbb{1}-\sum_{m'=1}^{|\mathcal{M}|}\Delta^{m'}_B$. By putting everything that has derived so far into the error term, we will have:

$$
\begin{aligned}
&\mathrm{Tr}\{(\mathbb{1}_{X^{|\mathcal{M}|}B}-\Lambda^m_{X^{|\mathcal{M}|}B})\rho^{m,(l,k)}_{X^{\otimes|\mathcal{M}|}B}\}\\
&\quad=\sum_{x_1,\ldots,x_{|\mathcal{M}|}}p_X(x_1)\ldots p_{x_{|\mathcal{M}|}}\mathrm{Tr}\{(\mathbb{1}_B-\Delta^m_B)\rho^{x_m,(l,k)}_B\}.
\end{aligned}
$$

By assuming a uniform distribution on the message set, averaging it over all

messages results in

$$\begin{aligned}
&\frac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|}\mathrm{Tr}\{(\mathbb{1}_{X^{|\mathcal{M}|}B}-\Lambda^{m}_{X^{|\mathcal{M}|}B})\rho^{m,(l,k)}_{X^{\otimes|\mathcal{M}|}B}\}\\
&=\frac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|}\sum_{x_1,\ldots,x_{|\mathcal{M}|}}p_X(x_1)\ldots p_X(x_{|\mathcal{M}|})\\
&\qquad\qquad\times\mathrm{Tr}\{(\mathbb{1}_B-\Delta^m_B)\rho_B^{x_m,(l,k)}\}\\
&=\sum_{x_1,\ldots,x_{|\mathcal{M}|}}p_X(x_1)\ldots p_X(x_{|\mathcal{M}|})\\
&\qquad\qquad\times\left[\frac{1}{|\mathcal{M}|}\mathrm{Tr}\{(I_B-\Delta^m_B)\rho_B^{x_m,(l,k)}\}\right],
\end{aligned}$$

the last expression above shows averaging over all codebooks and we know that

$$\begin{aligned}
&\sum_{x_1,\ldots,x_{|\mathcal{M}|}}p_X(x_1)\ldots p_X(x_{|\mathcal{M}|})\\
&\qquad\qquad\times\left[\frac{1}{|\mathcal{M}|}\mathrm{Tr}\{(\mathbb{1}_B-\Delta^m_B)\rho_B^{x_m,(l,k)}\}\right]\leq\varepsilon,
\end{aligned}$$

which in turn, says that there exists at least one particular set of values of $\{x_1,\ldots x_{|\mathcal{M}|}\}$ such that

$$\frac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|}\mathrm{Tr}\{(I_B-\Delta^m_B)\rho_B^{x_m,(l,k)}\}\leq\varepsilon. \tag{4.54}$$

This conclusion is known as the *Shannon trick*. The sequence $\{x_1\ldots x_{|\mathcal{M}|}\}$ serves as the codebook used to transmit the public message.

As for the second part, we take (4.43) and average over all private messages as given in (4.55). And we again employ Shannon trick to conclude that there exists at least one sequence of values $(y_{1,1}\ldots y_{|\mathcal{L}|,|\mathcal{K}|}|x)$ such that equation (4.57) holds.

We can now argue that there exist values $(x_1\ldots x_{|\mathcal{M}|})$ serving as *public codebook* for the transmission of the public message and conditioned on a particular codeword of the public codebook, there exist values $(y_{1,1}\ldots y_{|\mathcal{L}|,|\mathcal{K}|})$ serving as *private codebook* ensuring that the privacy criterion holds. Now we have a codebook of size $|\mathcal{M}||\mathcal{L}||\mathcal{K}|$, $\{x_1,\ldots,x_{|\mathcal{M}|},y_1,\ldots,y_{|\mathcal{L}||\mathcal{K}|}\}$, that is publicly available serving as our deterministic codebook for simultaneous transmission of public and private messages.

$$\varepsilon + \sqrt{\varepsilon'} \geq \frac{1}{|\mathcal{L}|}\sum_{\ell=1}^{|\mathcal{L}|} \sum_{x,y_{11},\ldots,y_{|\mathcal{L}||\mathcal{K}|}} p_X(x)p_{Y|X}(y_{11}|x)\ldots p_{Y|X}(y_{|\mathcal{L}||\mathcal{K}|}|x)$$
$$\left[\frac{1}{2}\left\| \mathcal{D}'^2_{B\to\hat{L}}\left(\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma_{BE}^{x,m,y_{\ell k}}\right) - |\ell\rangle\langle\ell|_{\hat{L}}\otimes\tilde{\sigma}_E^{x,m}\right\|_1\right] \tag{4.55}$$
$$= \sum_{x,y_{11},\ldots,y_{|\mathcal{L}||\mathcal{K}|}} p_X(x)p_{Y|X}(y_{11}|x)\ldots p_{Y|X}(y_{|\mathcal{L}||\mathcal{K}|}|x)\left(\frac{1}{|\mathcal{L}|}\sum_{l=1}^{|\mathcal{L}|}\right.$$
$$\left.\left[\frac{1}{2}\left\| \mathcal{D}'^2_{B\to\hat{L}}\left(\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma_{BE}^{x,m,y_{\ell,k}}\right) - |\ell\rangle\langle\ell|_{\hat{L}}\otimes\tilde{\sigma}_E^{x,m}\right\|_1\right]\right). \tag{4.56}$$

$$\frac{1}{|\mathcal{L}|}\sum_{l=1}^{|\mathcal{L}|}\left[\frac{1}{2}\left\| \mathcal{D}'^2_{B\to\hat{L}}\left(\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}|}\sigma_{BE}^{x,m,y_{\ell,k}}\right) - |\ell\rangle\langle\ell|_{\hat{L}}\otimes\tilde{\sigma}_E^{x,m}\right\|_1\right] \leq \varepsilon + \sqrt{\varepsilon'}. \tag{4.57}$$

One-shot Quantum Capacity: Imitating Devetak's Asymptotic Proof

As we mentioned in the introduction, a one-shot version of Devetak's asymptotic proof of quantum capacity follows along the same lines [19] . Here we briefly outline the proof and the general idea. We shall freely use the notation introduced so far. We have now seen that there exists a good codebook $\{x(\ell,k)\}_{\ell\in\mathcal{L},k\in\mathcal{K}}$ selected from a distribution $p(x)$ and a corresponding POVM $\{\Omega_B^{\ell,k}\}$ such that once Alice transmits a state corresponding to $x(\ell,k)$ over the channel $\mathcal{N}_{A\to BE}$, Bob is able to reliably work out both Alice's message ℓ and the local key k and at the same time, Eve happens to learned very little about Alice's message ℓ. This holds true for the rates of the private message $\log_2|\mathcal{L}|$ and the local randomness $\log_2|\mathcal{L}|$ as are specified. Now a quantum code can be obtained by a "making coherent" of this code (see [2] for coherifying general protocols).

The first idea of making protocols coherent is that classical words/letters x become basis states $|x\rangle$ of the Hilbert space. Functions $f : x \to f(x)$ thus induce linear operators on Hilbert space, but only permutations (one-to-one functions) are really interesting, since they give rise to unitaries (isometries, resp.). The second idea is thus to make classical computations first reversible, by extending them to one-to-one functions. The last step is to use the local

decodings, which exist by the classical theorem. In summary, "making coherent" means we can take a classical protocol working on letters and turn it into a bunch of unitaries acting as permutations on the basis states, and that we can run perfectly well on superpositions.

From the recipe outlined above, Alice's messages $\ell \in \mathcal{L}$ become a basis $\{|\ell\rangle_{A_1}\}_{\ell\in\mathcal{L}}$ of the Hilbert space. Suppose that Alice shares a state $|\varphi\rangle_{RA_1}$ with a reference system R:

$$|\varphi\rangle_{RA_1} \coloneqq \sum_{i,\ell\in\mathcal{L}} \alpha_{i,\ell} |i\rangle_R |\ell\rangle_{A_1},$$

where $|i\rangle_R$ and $|\ell\rangle_{A_1}$ are some orthonormal bases for R and A_1, respectively. A number of different information-processing tasks can be considered as quantum communications. The strongest definition of quantum capacity, however, corresponds to a task known as *entanglement transmission*. According to this task, Alice aims to transfer her share of entanglement with a reference system to Bob with Alice no longer entangled with the reference, i.e., the following state:

$$|\varphi\rangle_{RB} \coloneqq \sum_{i,\ell\in\mathcal{L}} \alpha_{i,\ell} |i\rangle_R |\ell\rangle_B,$$

where now Bob holds the B system. Suppose there is an ensemble of quantum states $\{p(x), |\psi^x\rangle_A\}$. Alice uses her classical code to create a quantum codebook whose codewords are as follows:

$$\left|\phi^\ell\right\rangle_A \coloneqq \frac{1}{\sqrt{|\mathcal{K}|}} \sum_{k\in\mathcal{K}} e^{\gamma_{\ell,k}} \left|\psi^{x(\ell,k)}\right\rangle_A,$$

where the states $|\psi^{x(\ell,k)}\rangle_A$ are from the aforementioned ensemble and $x(\ell,k)$ belong to the (classical) private codebook. Alice's action would be to coherently copy the value of ℓ in register A_1 to another register A_2. She then applies some isometric encoding from A_2 register to A. These two steps are performed with the following map:

$$\left(\sum_\ell |\ell\rangle\langle\ell|_{A_2} \otimes \left|\phi^\ell\right\rangle_A\right)\left(\sum_\ell |\ell\rangle\langle\ell|_{A_1} \otimes |\ell\rangle_{A_2}\right),$$

Alice then transmits the codeword over the channel giving rise to the following state:

$$\sum_{i,\ell\in\mathcal{L}} \alpha_{i,\ell} |i\rangle_R |\ell\rangle_{A_1} \left|\phi^\ell\right\rangle_{BE}.$$

From the classical protocol we know that Bob can detect both the message ℓ and the local key k with high probability. Bob constructs a coherent version of his POVM as follows:

$$\sum_{\ell,k} \sqrt{\Omega_B^{\ell,k}} \otimes |\ell\rangle_{B_1} |k\rangle_{B_2} .$$

From gentle measurement lemma, the state after Bob's decoding will be close to the following state:

$$\sum_{i,\ell} \sum_{k} \frac{1}{\sqrt{|\mathcal{K}|}} \alpha_{i,\ell} |i\rangle_R |\ell\rangle_{A_1} e^{\delta_{\ell,k}} \left|\phi^{x(\ell,k)}\right\rangle_{BE} |\ell\rangle_{B_1} |k\rangle_{B_2} .$$

On the other hand, from secrecy requirement, it can be seen that there exists some isometry on Bob's B and B_2 systems such that after its application, Eve's system will be decoupled from the rest and the following state will result:

$$\sum_{i,\ell} \alpha_{i,\ell} |i\rangle_R |\ell\rangle_{A_1} |\ell\rangle_{B_1} .$$

So far they have successfully implemented an approximate coherent channel from systems A_1 to A_1B_1. Alice is allowed to use a forward classical channel to communicate with Bob in order to turn the above coherent channel to a quantum channel. Alice's strategy is to first perform a Fourier transform on the register A_1 then measure the register in the computational basis and communicate the classical output to Bob. Bob will perform a controlled unitary based on the classical letter he received and the desired state will be achieved. Note than it can be shown that there exists a scheme that does not require the use of this forward classical channel.

Chapter 5

Single-Serving Quantum Broadcast Channel with Common, Individualized and Confidential Messages

The two-receiver broadcast channel with primary and third party receivers is studied. The sender wishes to reliably communicate a common (or public) message to both receivers as well as individualized and confidential messages to the primary receiver only. The third party receiver must be kept completely ignorant of the confidential message but there are no secrecy requirements associated to the individualized message. A trade-off arises between the rates of the three messages: when one of the rates is high, the other rates may need to back off to guarantee the reliable transmission of all three messages. In addition, the confidentiality requirement implies availability of local randomness at the transmitter in order to implement a stochastic encoding. This chapter studies the trade-off between the rates of the common, individualized and confidential messages as well as that of the local randomness in the one-shot regime of a quantum broadcast channel. We provide an achievability region, by proving a conditional version of the convex-spit lemma combined with the position-based decoding, as well as a (weak) converse region. We study the asymptotic behavior of our bounds and recover several well-known asymptotic results in the literature, including simultaneous transmission of classical and quantum information.

5.1 Background

When a single information source transmits to multiple users, the broadcast channel is concerned. Perhaps the most important application of the broadcast channel is the wiretap channel due to Wyner [13]: a transmitter encodes a message into a codeword which is sent over the noisy channel to the legitimate receiver. Wyner's model, however, ceased to have unrestricted impact since it limits the eavesdropper by assuming that she suffers from more noise than is experienced by the legitimate receiver. It was also not clear whether the specific stochastic encoding of Wyner's model is still optimal in the non-degraded scenario. In addition, the tradeoff between reliability and security was not clear, i.e. whether one could transmit reliable messages to the eavesdropper and conceal other messages simultaneously. These issues are resolved by analyzing a more general model than Wyner's degraded wiretap channel model.

This general model introduced by Csiszár and Körner [14] concerns a broadcast channel with two receivers for which a sender attempts to transmit two messages simultaneously: a common message, which is intended for both receivers, and an individual secret message, which is intended for only one receiver, treating the other receiver as an eavesdropper. Csiszár and Körner determined the optimal information rate tuples of the secret message and the common message, and the information leakage rate of the secret message to the eavesdropper, which is measured by the conditional entropy of the secret message given eavesdropper's received signal. They called their generalized problem the broadcast channel with confidential messages (BCC). The secrecy of messages over the wiretap channel and the BCC is realized by including meaningless random variable, which is called the dummy message, into sender's transmitted signal. This obviously decreases the information rate, however, it is essential to achieve secrecy. To compensate in part for the rate loss in the wiretap channel, it is possible to replace the dummy message with some non-secret or "individualized" message [126]; it is a message intended for the legitimate receiver with no secrecy requirement imposed on it.

An important issue that was ignored in the aforementioned works was the availability of dummy randomness. In case of the degraded wiretap channel, when the randomness is constrained and not necessarily uniform, [127] showed the rate loss due to lack of unlimited randomness; more precisely, the latter paper showed how the limited randomness restricts the distributions that the optimal secrecy rate should be optimized over. Subsequently, the general BCC model with rate-constrained randomness was studied in [128] where the optimal rate region of the common, individualized, confidential and

dummy randomness was determined for classical channels[1]. The achievability is based on superposition coding and building a deterministic codebook to replace the prefixing stochastic map. The idea of using a deterministic encoder in the wiretap channel was originally proposed in [129] in the contex of three-receiver broadcast channel . It was shown in [128] that the randomness needed to select a codeword following the scheme of [129] is smaller than the randomness rate needed to simulate the prefixing map. This shows that the direct concatenation of the ordinary random encoding and channel prefixing with channel simulation is in general suboptimal.

Unlike classical wiretap channel, the capacity of the classical-quantum wiretap channel is not know in general [18] and [19]. For general trade-off capacity theorems for three or more resources see [107, 108, 130–137].

The unavailability of unlimited resources such as many instances of channels or many copies of certain states in nature, triggered a new area of research known as the information theory with finite resources. This area has drawn significant attention over the past years; see [138] for a survey. The extreme scenario where only one instance of a certain resource such as a channel use or a source state is available, is generally called the *one-shot* regime and such a channel (res. source) is called a *single-serving* channel (res. source). The one-shot channel model is the most general model and its capacity to accomplish several information-processing tasks has been studied. The question of the number of bits that can be transmitted with an error of at most $\varepsilon > 0$ by a single use of a classical channel was answered in [112] where the capacity was characterized in terms of smooth min- and max-entropies. The same question for the quantum channels was studied in [113] following a hypothesis-testing approach and the capacity was characterized in terms of the general Rényi entropies. In this context, a reformulation of a novel positive operator-valued measurement (POVM) originally introduced in [139] (see also [140]), was employed in [25] yielding an achievability bound for the capacity of the classical-quantum channels. The POVM construction as well as the converse proof followed a hypothesis-testing procedure and the result was governed by a smooth relative entropy quantity. This result was rederived in [100] by deploying a coding scheme known as position-based decoding [31]. While the position-based decoding ensures the reliability of the transmitted messages, the so-called convex-split lemma [115] also employed in [100] guarantees confidentiality resulting in a capacity theorem for the one-shot wiretap quantum channel.

[1]In [128], the non-secret or individualized message is referred to as private message. We find the nomenclature “private message" to be rather inconsistent for a message that need not be kept private and we instead use the word “individualized”.

Position-based decoding and the convex-split lemma are governed by the quantities known as the smooth relative entropies and can be regarded as generalizations of the packing and covering lemmas, respectively. Another result on the one-shot capacity of the quantum wiretap channel was given by [116] where the reliability of the messages is ensured by employing the POVMs introduced in [25] and the confidentiality of the messages is established by proving a novel one-shot covering lemma analogous in approach to [117].

From a different perspective, [114] showed that two primitive information-theoretic protocols, namely information reconciliation and privacy amplification, can be used to directly construct optimal two-terminal protocols for noisy channels without being concerned about the internal workings of the primitives. This approach yields achievability bounds for the public and confidential capacities of classical-quantum channels and their tightness also established by proving corresponding converse bounds. The quantum capacity of a quantum channel for one or a finite number of uses is studied in [26]. The authors of the content of the work of the current chapter with their colleagues in a former work [56], unified the problems of one-shot transmission of public and confidential information over quantum channels and proposed a protocol for simultaneously achieviable public and confidential rates as well as tight converse bounds. Later, following the proof of the quantum capacity in [19], they proved a one-shot result for the simultaneous transmission of classical and quantum information [55], contributing to the literature of the one-shot trade-off capacities [141–144]. Another coding scheme known as Marton coding is known to yield tight achievability and converse regions for the broadcast channel. However, since our scheme is based on superposition coding, we do not use the ideas from Marton coding. The interested reader may refer to [145–147].

This work grew out of an interest to understand the amount of dummy randomness that is needed to accomplish the task of secret message transmission in the most general channel model. As mentioned earlier, in the asymptotic limit of a memoryless classical channel, it is shown that non-secret messages may compensate for the lack of enough dummy randomness to secure certain confidential messages. This was our main motivation: to understand the price of the dummy randomness and how (much) it can be traded off for a non-secret message. In this work, we consider a broader question featuring our main goal, we study the problem of the transmission of common, individualized and confidential messages with randomness constrained encoder over a single use of a two-receiver quantum broadcast channel. This problem in the asymptotic setting of a memoryless classical channel was studied in [128]. One additional contribution of [128] is the study of the channel

resolvability problem via superposition of classical codewords. The quantum channel resolvability via superpositions in the one-shot regime was studied in [148] in the context of the Gelfand-Pinsker quantum wiretap channel. We leverage these results to derive achievability bounds based on position-based decoding and the convex-split lemma. The setup of our problem, however, requires an extension of the position-based decoding and convex-split lemma, which we refer to as the conditional position-based decoding and conditional convex-split lemma. The former leads to an operational interpretation of a recently-defined mutual information-like quantity [149] whereas the latter gives rise to another novel mutual-information like quantity and its operational interpretation. The broad scope of the rate region developed in this work enables us to recover not only the classical result of [128], but also the case of simultaneous transmission of public and private information [56], the simultaneous transmission of classical and quantum information [20], [55] and the capacity region of the quantum broadcast channel derived in [150].

The rest of the chapter is organized as follows. We start with definitions in Section II. Section III is devoted to the description of the information-processing task, the definition of the code for the task and our main results. We prove an achievability region in Section IV and a converse region in Section V. We give asymptotic analysis in Section VI. We finally conclude the chapter in Section VII. The proof of the conditional convex-split lemma as well as several other lemmas are given in the appendix.

5.2 Miscellaneous Definitions

A quantum broadcast channel $\mathcal{N}^{A\to BC}$, refers to a quantum channel with a single input and two outputs such that when the transmitter inputs a quantum state in $\mathcal{S}^A$, one receiver obtains a state in B while the other receiver obtains system. Throughout we assume the receiver obtaining B system is the primary receiver and the receiver obtaining C is a third party. It is also useful to personify the users of the channel such that Alice is the user controlling the input and Bob and Charlie are the recipients of the systems B and C, respectively. According to the Stinespring dilation of the cptp map $\mathcal{N}^{A\to BC}$ (see for example [101]), there exists an *inaccessible environment* F and a unitary operator U acting on A, C and F systems such that

$$\mathcal{N}^{A\to BC}(\rho^A) = \mathrm{Tr}_F\{U(\rho^A \otimes \sigma^C \otimes \omega^F)U^\dagger\}, \tag{5.1}$$

where ρ^A is the input state and σ^C and ω^F are some constant states on systems C and F, respectively[2]. An additional trace over C system gives the quantum channel from Alice to Bob $\mathcal{N}^{A\to B}$ implying that the composite system $E \coloneqq CF$ plays the role of an inaccessible environment for $\mathcal{N}^{A\to B}$.

Definition 12 (Hypothesis testing conditional mutual information [149])**.** *Let*

$$\rho^{XAB} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^{AB}$$

and

$$\rho^{A-X-B} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \rho_x^B$$

be two tripartite states classical on X system. Let $\varepsilon \in (0,1)$. Then, the hypothesis testing conditional mutual information is defined as:

$$I_{\mathrm{H}}^{\varepsilon}(A;B|X)_\rho \coloneqq D_{\mathrm{H}}^{\varepsilon}(\rho^{XAB}\|\rho^{A-X-B}).$$

From Eq. (1.6), the following relation can be obtained:

$$I_{\mathrm{H}}^{\varepsilon}(A;B|X)_\rho \leq \frac{1}{1-\varepsilon}(I(A;B|X)_\rho + h_b(\varepsilon)), \tag{5.2}$$

and from Eq. (1.7), we can obtain

$$\lim_{n\to\infty}\frac{1}{n}I_{\mathrm{H}}^{\varepsilon}(A^n;B^n|X^n)_{\rho^{\otimes n}} = I(A;B|X)_\rho. \tag{5.3}$$

Notice that for states $\rho^{XAB} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^{AB}$ and $\rho^{A-X-B} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \rho_x^B$, we have $D(\rho^{XAB}\|\rho^{A-X-B}) = I(A;B|X)_\rho$.

Lemma 5.1. *For quantum states ρ and σ and a parameter $\varepsilon \in (0,1)$, the following indicates the relation between the smooth max-relative entropy and quantum relative entropy.*

$$D_{\max}^{\sqrt{2\varepsilon}}(\rho\|\sigma) \leq \frac{1}{1-\varepsilon}(D(\rho\|\sigma) + h_b(\varepsilon)),$$

where $h_b(\varepsilon) \coloneqq -\varepsilon\log\varepsilon - (1-\varepsilon)\log(1-\varepsilon)$ is the binary entropy function.

[2]This can be equivalently shown via an isometric extension of the channel $V_{\mathcal{N}}^{A\to BCF}$ defined as $\mathcal{N}^{A\to BC}(\rho^A) = \mathrm{Tr}_F\{V\rho^A V^\dagger\}$ with $V^\dagger V = \mathbb{1}^A$, $VV^\dagger = \Pi_{BCF}$ where Π_{BCF} is a projection on the product Hilbert space $B\otimes C\otimes F$.

Proof. The first inequality, the upper bound on the smooth max-relative entropy, follows by a straightforward manipulation of Proposition 4.1 in [151] and relation given by Eq. (1.6.) ∎

Definition 13. *Let* $\rho^{XAB} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^{AB}$ *and* $\rho^{A-X-B} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \rho_x^B$ *be quantum states classical on* X *and* $\varepsilon \in (0,1)$*, then*

$$D_{\max}^{\varepsilon}(A;B|X)_\rho \coloneqq D_{\max}^{\varepsilon}(\rho^{XAB} \| \rho^{A-X-B})_\rho.$$

From Lemma 5.1, the following relations can be seen:

$$D_{\max}^{\sqrt{2\varepsilon}}(A;B|X)_\rho \leq \frac{1}{1-\varepsilon}(I(A;B|X) + h_b(\varepsilon)), \tag{5.4}$$

and

$$\lim_{n\to\infty} \frac{1}{n} D_{\max}^{\varepsilon}(A^n;B^n|X^n)_{\rho^{\otimes n}} = D(\rho^{XAB} \| \rho^{A-X-B}) = I(A;B|X)_\rho. \tag{5.5}$$

Lemma 5.2. *For a bipartite state* ρ^{AB} *and a parameter* $\varepsilon \in (0,1)$*, the following relation holds:*

$$D_{\max}^{\varepsilon}(A;B)_\rho \leq \widetilde{I}_{\max}^{\varepsilon}(A;B)_\rho.$$

Proof. The proof is given in the appendix. ∎

We define another mutual information-like quantity similar to the one given by Definition 13.

Definition 14. *Let* $\rho^{XAB} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^{AB}$ *and* $\rho^{A-X-B} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \rho_x^B$ *be quantum states classical on* X *and* $\varepsilon \in (0,1)$*, then*

$$\widetilde{I}_{\max}^{\varepsilon}(A;B|X)_\rho \coloneqq \min_{\rho' \in \mathcal{B}^{\varepsilon}(\rho)} D_{\max}(\rho'^{XAB} \| \sum_x p'(x)|x\rangle\langle x|^X \otimes \rho_x'^A \otimes \rho_x^B),$$

where $\mathrm{Tr}_B\, \rho'^{XAB} = \sum_x p'(x)|x\rangle\langle x|^X \otimes \rho_x'^A$.

Remark 5.1. *In the definition above, it is implied that the minimization is in fact being performed over states which are classical on subsystem* X*, leading to the conclusion that the optimal state attaining the minimum is classical on* X*. Lemma 6.6 in [138] studied two important entropic quantities, namely smooth conditional min- and max-entropies, and concluded that smoothing respects the structure of the state* ρ^{XAB}*, meaning that the optimal state* $\rho'^{XAB} \in \mathcal{B}^{\varepsilon}(\rho^{XAB})$ *will be classical on subsystem* X*. Here we make an argument showing that our definition is indeed a legitimate definition. Let* $\bar{\rho}^{XAB} \in \mathcal{B}^{\varepsilon}(\rho^{XAB})$ *be the state attaining the minimum in*

the quantity $\widetilde{I}^{\varepsilon}_{\max}(A;B|X)_{\rho}$ *if we do not restrict* X *system to be classical. Let the* pinching map *be defined as* $\mathscr{P}^{X}(.) = \sum_x |x\rangle\langle x|\,(.)\,|x\rangle\langle x|$ *and define* $\rho'^{XAB} = \mathscr{P}^{X}(\bar{\rho}^{XAB})$*. Note that the pinching map does not affect* ρ^{XAB}*, and since such maps are cptp and unital, from the monotonicity of the purified distance and also smooth max-relative entropy, we will have* $\rho'^{XAB} \in \mathcal{B}^{\varepsilon}(\rho^{XAB})$ *and* $\widetilde{I}^{\varepsilon}_{\max}(A;B|X)_{\rho'^{XAB}} \leq \widetilde{I}^{\varepsilon}_{\max}(A;B|X)_{\bar{\rho}^{XAB}}$*, respectively.*

Lemma 5.3. *For quantum states* $\rho^{XAB} = \sum_x p(x)\,|x\rangle\langle x| \otimes \rho^{AB}_x$ *and* $\rho^{A-X-B} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \rho^A_x \otimes \rho^B_x$ *and a parameter* $\varepsilon \in (0,1)$*, we have:*

$$\widetilde{I}^{2\varepsilon}_{\max}(A;B|X)_{\rho} \leq D^{\varepsilon}_{\max}(A;B|X)_{\rho} + \log\left(\frac{1}{1-\sqrt{1-\varepsilon^2}} + 1\right).$$

Proof. The proof is relegated to the appendix. ∎

Lemma 5.4. [3] *For quantum states* $\rho^{XAB} = \sum_x p(x)\,|x\rangle\langle x| \otimes \rho^{AB}_x$ *and* $\rho^{A-X-B} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \rho^A_x \otimes \rho^B_x$ *and a parameter* $\varepsilon \in (0,1)$*, the following stands:*

$$D^{\varepsilon}_{\max}(A;B|X)_{\rho} \leq \widetilde{I}^{\varepsilon}_{\max}(A;B|X)_{\rho}.$$

Proof. The proof is provided in the appendix. ∎

Lemma 5.5. *For quantum states* $\rho^{XAB} = \sum_x p(x)\,|x\rangle\langle x| \otimes \rho^{AB}_x$ *and* $\rho^{A-X-B} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \rho^A_x \otimes \rho^B_x$ *and a parameter* $\varepsilon \in (0,1)$*, it holds that:*

$$\lim_{n\to\infty} \frac{1}{n} \widetilde{I}^{\varepsilon}_{\max}(A^n;B^n|X^n)_{\rho^{\otimes n}} = I(A;B|X)_{\rho}.$$

Proof. The proof follows from Lemma 5.3 and Lemma 5.4 as well as the fact given by (5.5). ∎

The following lemma comes in handy in the proof of the conditional convex-split lemma.

Lemma 5.6. *For an ensemble of classical-quantum states* $\{\rho^{XA}_1, ..., \rho^{XA}_n\}$ *and a probability mass function* $\{p(i)\}_{i=1}^n$*, let* $\rho^{XA} = \sum_i p(i)\rho^{XA}_i$ *be the average state. Then for a state* θ^{XA} *we have the following equality:*

$$D(\rho^{XA}\|\theta^{XA}) = \sum_{i=1}^{n} p(i)\left(D(\rho^{XA}_i\|\theta^{XA}) - D(\rho^{XA}_i\|\rho^{XA})\right).$$

Proof. Proof is presented in the appendix. ∎

[3] Note that for our purposes in this chapter, the upper bound given by Lemma 5.3 is enough; we prove this lemma further for sake of completeness of our study.

Lemma 5.7 (Conditional convex-split lemma)**.** *Consider the classical-quantum state* $\rho^{XAB} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^{AB}$*, define* $\sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^B$ *such that* $\operatorname{supp}(\rho_x^B) \subseteq \operatorname{supp}(\sigma_x^B)$ *for all* x*. Let* $k \coloneqq D_{\max}(\rho^{XAB} \| \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^B)$*. Define the following state:*

$$\tau^{XAB_1\ldots B_n} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \left(\frac{1}{n}\sum_{j=1}^{n} \rho_x^{AB_j} \otimes \sigma_x^{B_1} \otimes \ldots \otimes \sigma_x^{B_{j-1}} \otimes \sigma_x^{B_{j+1}} \otimes \ldots \otimes \sigma_x^{B_n}\right),$$

on $n+2$ *systems* $X, A, B_1, \ldots, B_n$*, where for* $\forall j \in [1:n]$ *and* $x \in \operatorname{supp}(p(x))$: $\rho_x^{AB_j} = \rho_x^{AB}$ *and* $\sigma_x^{B_j} = \sigma_x^B$*. We have the following:*

$$D\big(\tau^{XAB_1\ldots B_n} \big\| \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^{B_1} \otimes \ldots \otimes \sigma_x^{B_n}\big) \leq \log\left(1 + \frac{2^k}{n}\right).$$

In particular, for some $\delta \in (0,1)$ *and* $n = \lceil \frac{2^k}{\delta^2} \rceil$ *the following holds:*

$$P(\tau^{XAB_1\ldots B_n}, \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^{B_1} \otimes \ldots \otimes \sigma_x^{B_n}) \leq \delta.$$

Proof. The proof is presented in the appendix. ∎

In the following, we present a variation of the conditional convex-split lemma which involves smooth conditional max-relative entropy.

Corollary 5.1. *Fix a* $\varepsilon > 0$*. Let* $\rho^{XAB} = \sum_x p(x)\,|x\rangle\langle x|^X \otimes \rho_x^{AB}$ *and* $\sum_x p(x)\,|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^B$ *be quantum states such that* $\operatorname{supp}(\rho_x^B) \subseteq \operatorname{supp}(\sigma_x^B)$ *for all* x*. Define* $k \coloneqq \min_{\rho' \in \mathcal{B}^\varepsilon(\rho)} D_{\max}(\rho'^{XAB} \| \sum_x p'(x)|x\rangle\langle x|^X \otimes \rho_x'^A \otimes \sigma_x^B)$ *where the optimization takes place over states classical on* X*. Further define the following state*

$$\tau^{XAB_1\ldots B_n} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \left(\frac{1}{n}\sum_{j=1}^{n} \rho_x^{AB_j} \otimes \sigma_x^{B_1} \otimes \ldots \otimes \sigma_x^{B_{j-1}} \otimes \sigma_x^{B_{j+1}} \otimes \ldots \otimes \sigma_x^{B_n}\right),$$

on $n+2$ *systems* $X, A, B_1, \ldots, B_n$*, where* $\forall j \in [1:n]$ *and* $x \in \operatorname{supp}(p(x))$: $\rho_x^{AB_j} = \rho_x^{AB}$ *and* $\sigma_x^{B_j} = \sigma_x^B$*. For* $\delta \in (0,1)$ *and* $n = \lceil \frac{2^k}{\delta^2} \rceil$*, the following holds true:*

$$P(\tau^{XAB_1\ldots B_n}, \sum_x p(x)\,|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^{B_1} \otimes \ldots \otimes \sigma_x^{B_n}) \leq 2\varepsilon + \delta.$$

Proof. Proof is presented in the appendix. ∎

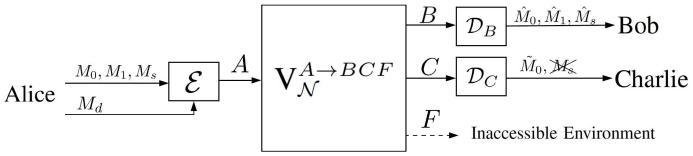

Figure 5.1: Single-serving quantum broadcast channel with isometric extension $V_{\mathcal{N}}^{A\to BCF}$. Alice attempts to transmit a common message M_0 to Bob and Charlie and a private message M_1 and a confidential message M_s to Bob only such that the confidential message must be kept secret from Charlie. The dummy randomness used by Alice for encryption is modeled by a message M_d.

5.3 Information-Processing Task, Code Definition and Main Results

Consider the quantum broadcast communication system model depicted in Fig. 5.1. A quantum broadcast channel $\mathcal{N}^{A\to BC}$ with isometric extension $\mathrm{V}_{\mathcal{N}}^{A\to BCF}$ connects a sender in possession of A system (Alice) to two receivers, a primary receiver (Bob) in possession of B and a third-party receiver (Charlie) possessing C system and the communication is surrounded by an inaccessible environment modeled as F system. Alice attempts to send three messages simultaneously: a common message M_0 that is to be decoded by Bob and Charlie, a private message M_1 addressed to Bob with no secrecy requirement and a confidential message M_s to Bob that must not be leaked to Charlie. The obfuscation of the confidential message is done by virtue of stochastic encoding, i.e., introducing local randomness into codewords in the encoding process. It is convenient to represent this local randomness as a *dummy message* M_d taking its values according to some distribution.

The encoder encodes the message triple (M_0, M_1, M_s) as well as the dummy message M_d into a quantum codeword ρ^A and transmits it over the quantum channel. Upon receiving ρ^B and ρ^C, Bob finds the estimates $\hat{M}_0, \hat{M}_1, \hat{M}_s$ of the common, individualized and confidential messages, respectively, while Charlie finds the estimate $\tilde{M}_0$ of the common message. To ensure reliability and security, a tradeoff arises between the rates of the messages. We study the one-shot limit on this tradeoff.

Definition 15. *A* $(2^{R_0}, 2^{R_1}, 2^{R_s})$ *one-shot code* $\mathcal{C}$ *for the quantum broadcast channel* $\mathcal{N}^{A\to BC}$ *consists of*

- *Three message sets* $[1 : 2^{R_0}]$, $[1 : 2^{R_1}]$ *and* $[1 : 2^{R_s}]$ *(common, individ-*

ualized and confidential, respectively),

- *A source of local randomness* $[1 : 2^{R_d}]$*,*
- *An encoding operator* $\mathcal{E} : M_0 \times M_1 \times M_s \times M_d \to A$*, which maps a message triple* $(m_0, m_1, m_s) \in [1 : 2^{R_0}] \times [1 : 2^{R_1}] \times [1 : 2^{R_s}]$ *and a realization of the local source of randomness* $m_d \in [1 : 2^{R_d}]$ *to a codeword* $\rho^A_{m_0,m_1,m_s}$*,*
- *A decoding POVM* $\mathcal{D}_B : B \to (M_0 \times M_1 \times M_s) \cup \{?\}$*, which assigns an estimate* $(\hat{m}_0, \hat{m}_1, \hat{m}_s) \in [1 : 2^{R_0}] \times [1 : 2^{R_1}] \times [1 : 2^{R_s}]$ *or an error message* $\{?\}$ *to each received state* $\rho^B_{m_0,m_1,m_s}$*,*
- *A decoding POVM* $\mathcal{D}_C : C \to M_0 \cup \{?\}$ *that assigns an estimate* $\tilde{m}_0 \in [1 : 2^{R_0}]$ *or an error message* $\{?\}$ *to each received state* $\rho^C_{m_0,m_1,m_s}$*.*

The $(2^{R_0}, 2^{R_1}, 2^{R_s})$ one-shot code is assumed to be known by all parties beforehand; likewise, the distribution of the local randomness is assumed known to all parties, however, its realization m_d is only accessible to Alice. Note that we have included the source of local randomness in the definition of the code to imply that it can be optimized over as part of the code design. Nevertheless, we do not consider the effect of non-uniform randomness in our analysis [127] and throughout we assume that the dummy message M_d is uniformly distributed over $[1 : 2^{R_d}]$. We further assume that the message triple (M_0, M_1, M_s) is uniformly distributed over $[1 : 2^{R_0}] \times [1 : 2^{R_1}] \times [1 : 2^{R_s}]$ so that the rates of the common, individualized and confidential messages are $H(M_0) = R_0, H(M_1) = R_1$ and $H(M_s) = R_s$, respectively. The reliability performance of the code $\mathcal{C}$ is measured by its average probability of error defined as follows:

$$P^1_{\text{error}} \coloneqq \Pr\{(\hat{M}_0, \hat{M}_1, \hat{M}_s) \neq (M_0, M_1, M_s) \text{ or } \tilde{M}_0 \neq M_0\}, \tag{5.6}$$

while its secrecy level, i.e., an indication of Charlie's ignorance about the confidential message, is measured in terms of the trace distance between Charlie's received state and some constant state as follows:

$$P^1_{\text{secrecy}}(m_0) \coloneqq \frac{1}{2^{R_s}} \sum_{m_s} \frac{1}{2} \|\rho^C_{m_0,m_s} - \sigma^C_{m_0}\|_1, \tag{5.7}$$

where $\rho^C_{m_0,m_s}$ and $\sigma^C_{m_0}$ are the average states at C when m_0 and m_s are transmitted. Note that the secrecy requirement indicates Charlie's ignorance about the confidential message m_s on average conditioned on the fact that he has decoded the common message m_0 correctly.

A rate quadruple (R_0, R_1, R_s, R_d) is said to be ε-achievable if there exist a one-shot code $\mathcal{C}$ satisfying the following conditions:

$$P^1_{\text{error}} \leq \varepsilon, \tag{5.8}$$

$$\forall m_0 : \quad P^1_{\text{secrecy}}(m_0) \leq \varepsilon, \tag{5.9}$$

where $\varepsilon \in (0,1)$ characterizes both the reliability and secrecy of the code. Then the ε-achievable rate region $\mathcal{R}^{\varepsilon}(\mathcal{N})$ is defined to consist of the closure of the set of all ε-achievable rate quadruples. In this chapter, our main goal is to bound the optimal rate region $\mathcal{R}^{\varepsilon}(\mathcal{N})$ by establishing achievability and converse regions.

The following theorem presents our achievability region on $\mathcal{R}^{\epsilon}(\mathcal{N})$.

Theorem 5.1 (Achievability Region)**.** *Fix $\varepsilon', \varepsilon'', \delta_1, \delta_2, \delta_3$ and η such that $0 < 3\varepsilon' + 2\sqrt{\varepsilon'} < 1, 0 < \delta_1, \delta_2, \delta_3 < \varepsilon'$, $0 < \varepsilon'' < \sqrt{2} - 1$, $0 < \eta < \varepsilon''^2$. Consider a quantum broadcast channel $\mathcal{N}^{A\to BC}$. Let the random variables U,V and X be distributed such that $U \to V \to X$ forms a Markov chain and define classical-quantum state $\rho^{UVXA} = \sum_{u,v,x} p(u,v,x) \left|u\right\rangle\!\left\langle u\right|^U \otimes \left|v\right\rangle\!\left\langle v\right|^V \otimes \left|x\right\rangle\!\left\langle x\right|^X \otimes \rho^A_x$. Let $\mathcal{R}^{(in)}(\rho)$ be the set of those quadruples (R_0, R_1, R_s, R_d) satisfying the following conditions on $\rho^{UVXBC} = \mathcal{N}^{A\to BC}(\rho^{UVXA})$:*

$$R_0 \leq \min\Big[I_{\text{H}}^{\varepsilon'-\delta_1}(U;B)_\rho - \log_2\left(\frac{4\varepsilon'}{\delta_1^2}\right), I_{\text{H}}^{\varepsilon'-\delta_2}(U;C)_\rho - \log_2\left(\frac{4\varepsilon'}{\delta_2^2}\right)\Big], \tag{5.10}$$

$$\begin{aligned} R_0 + R_1 + R_s \leq\ & I_{\text{H}}^{\varepsilon'-\delta_3}(V;B|U)_\rho - \log_2\left(\frac{4\varepsilon'}{\delta_3^2}\right) \\ & + \min\Big[I_{\text{H}}^{\varepsilon'-\delta_1}(U;B)_\rho - \log_2\left(\frac{4\varepsilon'}{\delta_1^2}\right), I_{\text{H}}^{\varepsilon'-\delta_2}(U;C)_\rho - \log_2\left(\frac{4\varepsilon'}{\delta_2^2}\right)\Big], \end{aligned} \tag{5.11}$$

$$R_s \leq I_{\text{H}}^{\varepsilon'-\delta_3}(V;B|U)_\rho - \widetilde{I}^{\varepsilon''}_{\max}(V;C|U)_\rho - \log_2\left(\frac{4\varepsilon'}{\delta_3^2}\right) - 2\log_2\left(\frac{1}{\eta}\right), \tag{5.12}$$

$$R_1 + R_d \geq \widetilde{I}^{\varepsilon''}_{\max}(V;C|U)_\rho + \widetilde{I}^{\varepsilon''}_{\max}(X;C|V)_\rho + 4\log_2\left(\frac{1}{\eta}\right), \tag{5.13}$$

$$R_d \geq \widetilde{I}^{\varepsilon''}_{\max}(X;C|V)_\rho + 2\log_2\left(\frac{1}{\eta}\right). \tag{5.14}$$

Let $\varepsilon := \max\{\sqrt[4]{\varepsilon'}, \sqrt[4]{\varepsilon''}\}$. Then $\bigcup \mathcal{R}^{(in)}(\rho) \subseteq \mathcal{R}^{\varepsilon}(\mathcal{N})$ where the union is over all ρ^{UVXBC} arising from the channel.

Proof. See Section IV. ∎

Theorem 5.2 (Converse Region)**.** *Fix $\varepsilon \in (0,1)$. Let the random variables U,V and X be distributed such that $U \to V \to X$ forms a Markov chain and define classical-quantum state $\rho^{UVXA} = \sum_{u,v,x} p(u,v,x) \left|u\right\rangle\!\left\langle u\right|^U \otimes \left|v\right\rangle\!\left\langle v\right|^V \otimes \left|x\right\rangle\!\left\langle x\right|^X \otimes \rho_x^A$. Let the state ρ^{UVXBC} be the result of the action of the quantum broadcast channel $\mathcal{N}^{A\to BC}$ on the state ρ^{UVXA}. Let $\mathcal{R}^{(co)}(\rho)$ be the set of those quadruples (R_0, R_1, R_s, R_d) satisfying the following conditions:*

$$R_0 \leq \min\left[I_{\mathrm{H}}^{\varepsilon}(U;B)_\rho, I_{\mathrm{H}}^{\varepsilon}(U;C)_\rho\right], \tag{5.15}$$

$$R_0 + R_1 + R_s \leq I_{\mathrm{H}}^{\varepsilon}(V;B|U)_\rho + \min\left[I_{\mathrm{H}}^{\varepsilon}(U;B)_\rho, I_{\mathrm{H}}^{\varepsilon}(U;C)_\rho\right], \tag{5.16}$$

$$R_s \leq I_{\mathrm{H}}^{\varepsilon}(V;B|U)_\rho - D_{\max}^{\sqrt{2\varepsilon}}(V;C|U)_\rho, \tag{5.17}$$

$$R_1 + R_d \geq D_{\max}^{\sqrt{2\varepsilon}}(V;C|U)_\rho + D_{\max}^{\sqrt{2\varepsilon}}(X;C|V)_\rho, \tag{5.18}$$

$$R_d \geq D_{\max}^{\sqrt{2\varepsilon}}(X;C|V)_\rho. \tag{5.19}$$

Then $\mathcal{R}^{\varepsilon}(\mathcal{N}) \subseteq \bigcup \mathcal{R}^{(co)}(\rho)$ and the union is over all ρ^{UVXBC} arising from the channel.

Proof. See Section V. ∎

From the theorems above, the recent result by the same authors on the simultaneous transmission of classical and quantum information can be obtained. The slight difference between the results stems from the fact that in [55], there is a single criterion for the error probability and secrecy while in this work separate criteria are considered.

Corollary 5.1 ([55])**.** *Fix $\varepsilon', \varepsilon'', \delta_1, \delta_3$ and η such that $0 < 3\varepsilon' + 2\sqrt{\varepsilon'} < 1, 0 < \delta_1, \delta_3 < \varepsilon'$, $0 < \varepsilon'' < \sqrt{2} - 1$, $0 < \eta < \varepsilon''^2$. Define $\varepsilon := \max\{\sqrt[4]{\varepsilon'}, \sqrt[4]{\varepsilon''}\}$. Let C^{ε} denote the one-shot capacity region of the channel $\mathcal{N}^{A\to BE}$ for simultaneous transmission of classical and quantum information. For a classical-quantum state $\rho^{UVA} = \sum_{u,v} p(u,v) \left|u\right\rangle\!\left\langle u\right| \otimes \left|v\right\rangle\!\left\langle v\right| \otimes \left|\psi_v\right\rangle\!\left\langle \psi_v\right|^A$, the following achievability bound holds:*

$$C^{(in)} \subseteq C^{\varepsilon},$$

where, denoting the one-shot rates of the classical and quantum information by R_c^1 and R_q^1, respectively, $C^{(in)}$ is the union over all states ρ^{UVBE} arising from the channel, of rate pairs (R_c^1, R_q^1) obeying:

$$R_c^1 \leq I_{\mathrm{H}}^{\varepsilon'-\delta_1}(U;B)_\rho - \log_2\left(\frac{4\varepsilon'}{\delta_1^2}\right),$$

$$R_q^1 \leq I_{\mathrm{H}}^{\varepsilon'-\delta_3}(V;B|U)_\rho - \widetilde{I}_{\max}^{\varepsilon''}(V;E|U)_\rho - \log_2\left(\frac{4\varepsilon'}{\delta_3^2}\right) - 2\log_2\left(\frac{1}{\eta}\right).$$

Let $\varepsilon \in (0,1)$. Then the following converse holds:

$$C^{\varepsilon} \subseteq C^{(co)},$$

where $C^{(co)}$ is the union over all states ρ^{UVBE} arising from the channel, of rate pairs (R_c^1, R_q^1) obeying

$$\begin{aligned} R_c^1 &\leq I_{\mathrm{H}}^{\varepsilon}(U;B)_{\rho}, \\ R_q^1 &\leq I_{\mathrm{H}}^{\varepsilon}(V;B|U)_{\rho} - D_{\max}^{\sqrt{2\varepsilon}}(V;E|U)_{\rho}. \end{aligned}$$

Proof. The approach for the simultaneous transmission of classical and quantum information is through finding the limits on the simultaneous transmission of common and confidential messages. From [19] it is well-known that the rate of the confidential message can be translated into the rate of quantum information. As hinted in the introductory part, when it comes to transmission of quantum information, there is zero-tolerance condition of copying quantum information; therefore the confidential messages must be kept secret from the entire universe but Bob, meaning that the output of the channel consists of a system received by Bob and another inaccessible environment E (which includes Charlie's system). From Theorem 5.1 and Theorem 5.2 onward, since there is no concern regarding the rate of the dummy randomness, the last two inequalities in both regions will be trivial. And the achievability part can be seen from (5.10) and (5.12) and the converse part from (5.15) and (5.17). ∎

5.4 Achievability

The achievability proof of the reliability condition (5.8) is based on a combination of classical superposition coding and position-based decoding whereas the secrecy condition (5.9) is handled by a version of the convex-split lemma which relies on superposition of codewords. This result is reminiscent of the channel resolvability problem via superposition studied for classical [128] and quantum [148] channels.

The proof of Theorem 5.1 follows from Lemma 5.8 and Lemma 5.9 where the former proves an alternative rate region and the latter shows the equivalence of the two regions.

Lemma 5.8. *Fix $\varepsilon', \varepsilon'', \delta_1, \delta_2, \delta_3$ and η such that $0 < 3\varepsilon' + 2\sqrt{\varepsilon'} < 1, 0 < \delta_1, \delta_2, \delta_3 < \varepsilon'$, $0 < \varepsilon'' < \sqrt{2} - 1$, $0 < \eta < \varepsilon''^2$ and define $\varepsilon := \max\{\sqrt[4]{\varepsilon'}, \sqrt[4]{\varepsilon''}\}$. Let the random variables U,V and X be distributed such that $U \to V \to X$ forms a Markov chain. We further define classical-quantum state $\rho^{UVXA} =$*

$\sum_{u,v,x} p(u,v,x)|u\rangle\langle u|^U \otimes |v\rangle\langle v|^V \otimes |x\rangle\langle x|^X \otimes \rho_x^A$. *Let* $\mathcal{R}^*(\rho)$ *be the set of those quadruples* (R_0, R_1, R_s, R_d) *satisfying the following conditions on* $\rho^{UVXBC} = \mathcal{N}^{A\to BC}(\rho^{UVXA})$*:*

$$R_0 \leq \min[I_{\mathrm{H}}^{\varepsilon'-\delta_1}(U;B)_\rho - \log_2\left(\frac{4\varepsilon'}{\delta_1^2}\right), I_{\mathrm{H}}^{\varepsilon'-\delta_2}(U;C)_\rho - \log_2\left(\frac{4\varepsilon'}{\delta_2^2}\right)], \quad (5.20)$$

$$R_1 + R_s \leq I_{\mathrm{H}}^{\varepsilon'-\delta_3}(V;B|U)_\rho - \log_2\left(\frac{4\varepsilon'}{\delta_3^2}\right), \quad (5.21)$$

$$R_1 \geq \widetilde{I}_{\max}^{\varepsilon''}(V;C|U)_\rho + 2\log_2\left(\frac{1}{\eta}\right), \quad (5.22)$$

$$R_d \geq \widetilde{I}_{\max}^{\varepsilon''}(X;C|V)_\rho + 2\log_2\left(\frac{1}{\eta}\right), \quad (5.23)$$

Then $\bigcup \mathcal{R}^*(\rho) \subseteq \mathcal{R}^\varepsilon(\mathcal{N})$ *and the union is over all* ρ^{UVXBC} *arising from the channel.*

Proof. Let $\varepsilon', \varepsilon'', \delta_1, \delta_2, \delta_3$ and η be such that $0 < 3\varepsilon' + 2\sqrt{\varepsilon'} < 1, 0 < \delta_1, \delta_2, \delta_3 < \varepsilon'$, $0 < \varepsilon'' < \sqrt{2} - 1$, $0 < \eta < \varepsilon''^2$ and $\varepsilon := \max\{\sqrt[4]{\varepsilon'}, \sqrt[4]{\varepsilon''}\}$.

Codebook generation: Fix a distribution $p(u,v,x)$ such that $U \to V \to X$. Alice, Bob and Charlie share randomness in the form of 2^{R_0} copies of the classical state $\rho^{U^A U^B U^C} := \sum_u p(u)|u\rangle\langle u|^{U^A} \otimes |u\rangle\langle u|^{U^B} \otimes |u\rangle\langle u|^{U^C} = \sum_u p(u)|uuu\rangle\langle uuu|^{U^A U^B U^C}$ as follows:

$$(\rho^{U^A U^B U^C})^{\otimes 2^{R_0}} = \rho^{U_1^A U_1^B U_1^C} \otimes \ldots \otimes \rho^{U_{2^{R_0}}^A U_{2^{R_0}}^B U_{2^{R_0}}^C},$$

where Alice possesses U^A systems, Bob U^B systems and Charlie has U^C systems (the superscripts should not be confused with the input A or output systems B and C of the channel, here they indicate the party to whom the underlying state belongs). We consider the shared state above to construct the first layer of our code. Conditioned on each of the 2^{R_0} states above, the parties are assumed to share $2^{R_s+R_1}$ copies of the state $\rho_u^{V^A V^B V^C} = \sum_v p(v|u)|vvv\rangle\langle vvv|^{V^A V^B V^C}$, as given below for the i-th state $\rho^{U_i^A U_i^B U_i^C}$:

$$\sum_u p(u)|uuu\rangle\langle uuu|^{U_i^A U_i^B U_i^C} \otimes (\rho_u^{V^A V^B V^C})^{\otimes 2^{R_s+R_1}},$$

where Alice, Bob and Charlie are in possession of the V^A, V^B and V^C systems, respectively. The set $[1 : 2^{R_s+R_1}]$ is partitioned into 2^{R_s} equal size bins. This constitutes the second layer of the code. Finally, conditioned on each of the $2^{R_s+R_1}$ states above the parties will share 2^{R_d} copies of the

state $\rho_v^{X^A X^B X^C} \coloneqq \sum_x p(x|v)|xxx\rangle\langle xxx|^{X^A X^B X^C}$, as illustrated below for the i-th state $\rho^{V_i^A V_i^B V_i^C}$:

$$\sum_{u,v} p(u,v)|vvv\rangle\langle vvv|^{V_i^A V_i^B V_i^C} \otimes (\rho_v^{X^A X^B X^C})^{\otimes 2^{R_d}},$$

where X^A, X^B and X^C systems are owned by Alice, Bob and Charlie, respectively. These states build the third layer of the code. All states above are assumed to be available to all parties before communication begins. In the following, to avoid inefficient notation we may drop the superscripts if it does not lead to ambiguity; for instance when we analyze Bob's error probability, it is obvious that we are dealing with Bob's systems or in the secrecy analysis those of Charlie are dealt with.

Encoding: To send a message triple (m_0, m_1, m_s), the encoder first chooses a dummy message $m_d \in [1 : 2^{R_d}]$. In the first layer, the encoder finds the m_0-th state, i.e. $\rho^{U_{m_0}^A}$, then it looks for the m_s-th bin, inside which it selects the state associated to the individualized message m_1; finally, the encoder picks the m_d-th state $\rho^{X_{m_d}^A}$ among those tied to the state found in the preceding step. The encoder sends the selected classical system through a modulator (a linear operator $\mathcal{V} : \mathcal{X} \to \mathcal{S}^A$ which maps the classical control variable $x \in \mathcal{X}$ to a quantum state in the input Hilbert space) resulting in a quantum codeword ρ_x^A which will be then transmitted over the channel[4].

Decoding: Bob performs a two-phase decoding strategy such that he finds the common message in the first phase and then confidential and individualized messages in the subsequent phase. The transmission of the m_0-th common message induces the following state on Bob's side:

$$\rho^{U_1} \otimes ... \otimes \rho^{U_{m_0} B} \otimes ... \otimes \rho^{U_{2^{R_0}}}, \tag{5.24}$$

where $\rho^{U_{m_0} B} = \sum_u p(u)\,|u\rangle\langle u|^{U_{m_0}} \otimes \rho_u^B$. Apparently Bob has to be able to spot the location where the received system B is tied to his U system. In other words, he should be able to distinguish between states induced for different values of the common message. Bob employs a position-based decoding to solve the raised 2^{R_0}-ary hypothesis testing problem. Moreover, for the common message m_0, the selection of the pair (m_s, m_1) will induce the following state on Bob's side:

$$\rho^{V_{(m_0,1,1)}} \otimes ... \otimes \rho^{V_{(m_0,m_s,m_1)} B} \otimes ... \otimes \rho^{V_{(m_0,2^{R_s},2^{R_1})}}, \tag{5.25}$$

[4]Note that we have included the modulator in the definition of the code meaning that it needs to be optimized over to get our capacity results.

where $\rho^{V_{(m_0,m_s,m_1)}B} = \sum_v p(v) |v\rangle\langle v|^{V_{(m_0,m_s,m_1)}} \otimes \rho_v^B$. Bob runs the second position-based POVM to solve the $2^{R_s+R_1}$-ary hypothesis testing problem. Charlie also runs the position-based decoding POVM to find out the transmitted common message. The state induced at Charlie side comes about by replacing B with C in (5.24).

Analysis of the probability of error: We first analyze the error probability of the common message by studying Bob's first decoder and the error analysis of Charlie can be carried out along the same lines. It is worth pointing out that although the messages encoded in the second layer might include dummy randomness, Bob will still decode them. The dummy messages in the third layer will not be decoded.

Reconsider the state in (5.24). To find out the transmitted common message, Bob has to distinguish between 2^{R_0} different states. As hinted before, this puts forward a 2^{R_0}-ary hypothesis testing problem. Let $\{T^{UB}, I - T^{UB}\}$ be the elements of a POVM that is chosen for discriminating between two states ρ^{UB} and $\rho^U \otimes \rho^B$. Further, we assume that the test operator T^{UB} decides correctly in favor of ρ^{UB} with probability at least[5] $1 - (\varepsilon' - \delta_1)$. Bob will use the following square-root measurement to detect the common message:

$$\Omega_{m_0} \coloneqq \left(\sum_{m_0'=1}^{2^{R_0}} \Pi_{m_0'} \right)^{-\frac{1}{2}} \Pi_{m_0} \left(\sum_{m_0'=1}^{2^{R_0}} \Pi_{m_0'} \right)^{-\frac{1}{2}},$$

where $\Pi_{m_0} \coloneqq \mathbb{1}^{U_1} \otimes \ldots \otimes T^{U_{m_0}B} \otimes \ldots \otimes \mathbb{1}^{U_{2^{R_0}}}$ and $T^{U_{m_0}B}$ is the test operator. It can be easily checked that the set $\{\Omega_{m_0}\}_{m_0}$ constitutes a valid POVM, i.e. $\sum_{m_0} \Omega_{m_0} = \mathbb{1}$. Besides, direct calculation shows that $\mathrm{Tr}\{\Pi_{m_0}(\rho^{U_1} \otimes \ldots \otimes \rho^{U_{m_0}B} \otimes \ldots \otimes \rho^{U_{2^{R_0}}})\} = \mathrm{Tr}\{T^{U_{m_0}B}\rho^{U_{m_0}B}\}$ and for any $m_0' \neq m_0$, $\mathrm{Tr}\{\Pi_{m_0}(\rho^{U_1} \otimes \ldots \otimes \rho^{U_{m_0'}B} \otimes \ldots \otimes \rho^{U_{2^{R_0}}})\} = \mathrm{Tr}\{\Pi_{m_0}(\rho^{U_{m_0}} \otimes \rho^B)\}$.

Observe that the symmetric structure of the codebook generation and decoding leads to an average error probability that is equal to the individual error probabilities. Therefore, we might assume $m_0 = 1$ was transmitted.

[5]For the sake of clarity, we choose to specify the error probability of the test operator to be $\varepsilon' - \delta_1$ to ensure that the error probability of the code will be larger than this and at most ε'.

Hence,

$$\begin{aligned}\Pr(\hat{M}_0 \neq 1|M_0 = 1) &= \mathrm{Tr}\{(\mathbb{1} - \Omega_1)(\rho^{U_1 B} \otimes \ldots \otimes \rho^{U_{2^{R_0}}})\} \\ &\leq (1+c)\,\mathrm{Tr}\{(\mathbb{1} - \Pi_1)(\rho^{U_1 B} \otimes \ldots \otimes \rho^{U_{2^{R_0}}})\} \\ &\qquad + (2 + c + c^{-1}) \sum_{m_0 \neq 1} \mathrm{Tr}\{\Pi_{m_0}(\rho^{U_1 B} \otimes \ldots \otimes \rho^{U_{2^{R_0}}})\} \\ &\leq (1+c)(\varepsilon' - \delta_1) + (2 + c + c^{-1}) 2^{R_0 - I_H^{\varepsilon' - \delta_1}(U;B)_\rho},\end{aligned}$$

where the first inequality follows from Lemma 1.1 and in the second inequality, the first term is based on the assumption and the second term follows from the definition of the hypothesis testing mutual information (see Definition 1.5). The last expression is set equal to ε' and the optimal value of c is derived as $c = \frac{\delta_1}{2\varepsilon' - \delta_1}$. Then, we will have

$$R_0 = I_H^{\varepsilon' - \delta_1}(U;B)_\rho - \log_2\left(\frac{4\varepsilon'}{\delta_1^2}\right).$$

In the same manner, it can be shown that the achievable rate of the common message to Charlie equals $R_0 = I_H^{\varepsilon' - \delta_2}(U;C)_{\rho^{UC}} - \log_2(\frac{4\epsilon'}{\delta_2^2})$.

In an analogous way, the reliability analysis of the confidential and the individualized messages goes as follows. Before we delve into the error analysis of the confidential and individualized messages, note from the gentle measurement lemma [152] that the disturbed state fed into the second decoder of Bob is impaired by at most $2\sqrt{\varepsilon'}$; this should be taken into account in final assessment of the error probability. Consider a binary POVM with elements $\{Q^{UVB}, \mathbb{1} - Q^{UVB}\}$. The POVM is to discriminate the states $\rho^{UVB} = \sum_u p(u)|u\rangle\langle u|^U \otimes \rho_u^{VB}$ and $\rho^{V-U-B} \coloneqq \sum_u p(u)|u\rangle\langle u|^U \otimes \rho_u^V \otimes \rho_u^B$ such that the value of Q^{UVB} estimates the state to be ρ^{UVB}. Assume the probability of failure to make a correct decision on ρ^{UVB} is at most $\varepsilon' - \delta_3$, i.e., $\mathrm{Tr}\{(\mathbb{1} - Q)\rho^{UVB}\} \leq \varepsilon' - \delta_3$. Bob will take the following square-root measurement POVM :

$$\Theta_{m_s,m_1} \coloneqq \left(\sum_{m_s'=1}^{2^{R_s}} \sum_{m_1'=1}^{2^{R_1}} \Gamma_{m_s',m_1'}\right)^{-\frac{1}{2}} \Gamma_{m_s,m_1} \left(\sum_{m_s'=1}^{2^{R_s}} \sum_{m_1'=1}^{2^{R_0}} \Gamma_{m_s',m_1'}\right)^{-\frac{1}{2}},$$

where $\Gamma_{m_s,m_1} \coloneqq \mathbb{1}^{V_{1,1}} \otimes \ldots \otimes Q^{UV_{m_s,m_1}B} \otimes \ldots \otimes \mathbb{1}^{V_{2^{R_s},2^{R_1}}}$ and $Q^{UV_{m_s,m_1}B}$ is the binary test operator. Observe that $\sum_{m_s,m_1} \Theta_{m_s,m_1} = \mathbb{1}$. It is easy to show that for all m_s, m_1, we have $\mathrm{Tr}\{\Gamma_{m_s,m_1}(\sum_u p(u)|u\rangle\langle u|^U \otimes \rho_u^{V_{1,1}} \otimes \ldots \otimes \rho_u^{V_{m_s,m_1}B} \otimes \ldots \otimes \rho_u^{V^B_{2^{R_s},2^{R_1}}})\} = \mathrm{Tr}\{Q\rho^{UVB}\}$. On the other hand, for any $m_s' \neq m_s$ or $m_1' \neq m_1$, $\mathrm{Tr}\{\Gamma_{m_s,m_1}(\sum_u p(u)|u\rangle\langle u|^U \otimes \rho_u^{V_{1,1}} \otimes \ldots \otimes \rho_u^{V_{m_s',m_1'}B} \otimes \ldots \otimes \rho_u^{V_{2^{R_s},2^{R_1}}})\} =$

$\mathrm{Tr}\{Q\rho^{V-U-B}\}$. By the symmetry of the random codebook construction, the average error probability is the same as the error probability of any pair (m_s, m_1), hence it suffices to find the error probability if $(m_s = 1, m_1 = 1)$ was sent. The analysis continues as follows:

$$
\begin{aligned}
&\Pr((\hat{M}_s, \hat{M}_1) \neq (1,1) | (M_s, M_1) = (1,1)) \\
&= \mathrm{Tr}\{(\mathbb{1} - \Theta_{1,1})(\sum_u p(u)|u\rangle\langle u|^U \otimes \rho_u^{V_{1,1}B} \otimes ... \otimes \rho_u^{V_{2^{R_s},2^{R_1}}})\} \\
&\leq (1+c)\,\mathrm{Tr}\{(\mathbb{1} - \Gamma_{1,1})(\sum_u p(u)|u\rangle\langle u|^U \otimes \rho_u^{V_{1,1}B} \otimes ... \otimes \rho_u^{V_{2^{R_s},2^{R_1}}})\} \\
&\quad + (2 + c + c^{-1}) \sum_{(m_s,m_1)\neq(1,1)} \mathrm{Tr}\{\Gamma_{m_s,m_1}(\sum_u p(u)|u\rangle\langle u|^U \otimes \rho_u^{V_{1,1}B} \otimes ... \otimes \rho_u^{V_{2^{R_s},2^{R_1}}})\} \\
&\leq (1+c)(\varepsilon' - \delta_3) + (2 + c + c^{-1}) 2^{R_1 + R_s - I_H^{\varepsilon'-\delta_3}(V;B|U)_{\rho^{UVB}}},
\end{aligned}
$$

where the first inequality is due to Lemma 1.1 and in the second inequality, the first term comes from the assumption about the accuracy of the test operator Q and the second term uses the definition of the hypothesis testing conditional mutual information, Definition 12. We choose the error probability be less that or equal to ε', so the optimal value of the constant is set to $c = \frac{\delta_3}{2\varepsilon' - \delta_3}$ and eventually we will get the following sum rate:

$$
R_s + R_1 = I_H^{\varepsilon'-\delta_3}(V;B|U)_\rho - \log_2\left(\frac{4\varepsilon'}{\delta_3^2}\right).
$$

Analysis of the secrecy: Our tool to study secrecy is the conditional convex-split lemma. The dummy message and perhaps the individualized message which take care of confidentiality are encoded in the second and third layers as superposition of shared states. The quantum channel resolvability via superposition coding was studied in [148]. Given the setup of our problem, here we should try to prove the resolvability problem using convex-split lemma. We gave the analysis for Charlie's successful detection of the common message; in the secrecy analysis we assume Charlie knows the common message and the correct copy of the ρ^U used in the first layer. The idea for secrecy is that Charlie's systems have to remain close to some constant state, no matter which confidential message was transmitted.

For a given confidential message, the choice of the individualized message will induce an average state on Charlie's V systems in the second layer where the dummy message induces an average state on his X systems in the third layer. Since the states in the second layer are superposed to those in the third layer, both the individualized message and the dummy message will help to

induce a state at Charlie's side that should be close enough to a target state. We first sketch the state induced by the chosen individualized and dummy messages. For a choice of the confidential message $m_s \in [1 : 2^{R_s}]$, the induced state is as follows (the subscripts of variable V should be understood as the ordered pairs $(1,1),...,(1,2^{R_1}),(2,1),...,(2,2^{R_1}),...,(m_s,m_1),...,(2^{R_s},2^{R_1})$ where the first component belongs to $[1 : 2^{R_s}]$ and the second component is in $[1 : 2^{R_1}]$):

$$\sum_u p(u)|u\rangle\langle u|^U$$
$$\otimes \Bigg(\big[\sum_v p(v|u)|v\rangle\langle v|^{V_{1,1}} \otimes (\rho_v^X)^{\otimes 2^{R_d}} \big] \otimes ... \otimes \big[\sum_v p(v|u)|v\rangle\langle v|^{V_{m_s-1,2^{R_1}}} \otimes (\rho_v^X)^{\otimes 2^{R_d}} \big] \otimes$$
$$\frac{1}{2^{R_1}} \sum_{j=1}^{2^{R_1}} \Upsilon_u^{C,j} \otimes \big[\sum_v p(v|u)|v\rangle\langle v|^{V_{m_s+1,1}} \otimes (\rho_v^X)^{\otimes 2^{R_d}} \big]$$
$$\otimes ... \otimes \big[\sum_v p(v|u)|v\rangle\langle v|^{V_{2^{R_s},2^{R_1}}} \otimes (\rho_v^X)^{\otimes 2^{R_d}} \big] \Bigg). \tag{5.26}$$

where

$$\Upsilon_u^{C,j} \coloneqq \sum_v p(v|u)|v\rangle\langle v|^{V_{m_s,1}} \otimes (\rho_v^X)^{\otimes 2^{R_d}} \otimes ... \otimes$$
$$\sum_v p(v|u)|v\rangle\langle v|^{V_{m_s,j}} \otimes \Psi_v^C \otimes ... \otimes \sum_v p(v|u)|v\rangle\langle v|^{V_{m_s,2^{R_1}}} \otimes (\rho_v^X)^{\otimes 2^{R_d}}, \tag{5.27}$$

and

$$\Psi_v^C \coloneqq \frac{1}{2^{R_d}} \sum_{i=1}^{2^{R_d}} \rho_v^{X_1^C} \otimes ... \otimes \rho_v^{X_i^C C} \otimes ... \otimes \rho_v^{X_{2^{R_d}}^C}. \tag{5.28}$$

In order to get an intuitive understanding of the equations above, first we should think of the sources of the randomness available in the protocol. Since the common message is already known, the remaining sources of the randomness with respect to the confidential message are the individualized and dummy messages. Second, we should note where each source of the randomness is being consumed. Equation (5.26) indicates that the chosen confidential message is m_s, equation (5.27) represents the uniform randomness imposed by the individualized message (see the range of the variable j) and finally, equation (5.28) reflects the randomness introduced by the dummy message (see the range of the variable i). All the messages available to the encoder are potential sources of randomness that can be used for secrecy purposes, i.e., to confuse a receiver about other messages. In our setting, the common message is decoded by Charlie and to hide the confidential

message, there is randomness coming from the individualized and dummy messages. Note that with regard to the individualized message, neither Alice's encoding nor Bob's decoding influence the role it plays as a source of randomness. Moreover, as discussed before, the individualized message may or may not contain useful information for Bob; however, Bob will decode the individualized message and then he can throw away its content. In case the individualized message contains useful information for Bob, by definition, we do not care if information about the individualized message is leaked to Charlie.

Charlie not being able to detect the confidential message amounts to his state being sufficiently close to the following state:

$$\begin{aligned}\sum_u p(u)|u\rangle\langle u|^U \otimes \Big(\big(\sum_v p(v|u)|v\rangle\langle v|^{V_{1,1}} \otimes (\rho_v^X)^{\otimes 2^{R_d}} \big) \\ \otimes \ldots \otimes \big(\sum_v p(v|u)|v\rangle\langle v|^{V_{m_s-1,2^{R_1}}} \otimes (\rho_v^X)^{\otimes 2^{R_d}} \big) \\ \otimes \Big[\big(\sum_v p(v|u)|v\rangle\langle v|^{V_{m_s}} \otimes (\rho_v^{X^C})^{\otimes 2^{R_d}} \big)^{\otimes 2^{R_1}} \otimes \rho_u^C \Big] \\ \otimes \ldots \otimes \big(\sum_v p(v|u)|v\rangle\langle v|^{V_{R_s,R_1}} \otimes (\rho_v^X)^{2^{R_d}} \big) \Big). \end{aligned} \tag{5.29}$$

where $\rho_u^C = \sum_{v,x} p(v,x|u)\rho_x^C$ is considered the constant state independent of the chosen confidential message. Concerning the trace distance between the aforementioned states, since the trace distance is invariant with respect to tensor product states, we can remove the same terms from both states. Eventually the following is the distance required to be small enough:

$$\begin{aligned}\frac{1}{2} \Big\| \sum_u p(u)|u\rangle\langle u|^U \otimes \frac{1}{2^{R_1}} \sum_{j=1}^{2^{R_1}} \Upsilon_u^{C,j} - \sum_u p(u)|u\rangle\langle u|^U \\ \otimes \big(\sum_v p(v|u)|v\rangle\langle v|^V \otimes (\rho_v^X)^{\otimes 2^{R_d}} \big)^{\otimes 2^{R_1}} \otimes \rho_u^C \Big\|_1, \end{aligned} \tag{5.30}$$

where the expression being subtracted refers to the state associated to the chosen confidential message given inside the brackets in (5.29). We proceed to bound equation (5.30) from above by envisioning an intermediate state which is, intuitively, closer to either of the states involved in (5.30) than the two states themselves. We define such an intermediate state as $\sum_u p(u)|u\rangle\langle u|^U \otimes$

Ξ_u^C where

$$\Xi_u^C \coloneqq \frac{1}{2^{R_1}} \sum_{j=1}^{2^{R_1}} \Bigg(\big[\sum_v p(v|u)|v\rangle\langle v|^{V_{m_s,1}} \otimes (\rho_v^X)^{\otimes 2^{R_d}} \big] \otimes \ldots \\ \otimes \big[\sum_v p(v|u)|v\rangle\langle v|^{V_{m_s,j}} \otimes (\rho_v^{X_1} \\ \otimes \ldots \otimes \rho_v^{X_{R_d}} \otimes \rho_v^C) \big] \otimes \ldots \otimes \big[\sum_v p(v|u)|v\rangle\langle v|^{V_{m_s,2^{R_1}}} \otimes (\rho_v^X)^{\otimes 2^{R_d}} \big] \Bigg).$$

Next, we have to bring in the intermediate state. We do so by the triangle inequality as follows

$$\begin{aligned} &\frac{1}{2} \Big\| \sum_u p(u)|u\rangle\langle u|^U \otimes \frac{1}{2^{R_1}} \sum_{j=1}^{2^{R_1}} \Upsilon_u^{C,j} \\ &\qquad - \sum_u p(u)|u\rangle\langle u|^U \otimes \Big(\sum_v p(v|u)|v\rangle\langle v|^V \otimes (\rho_v^X)^{\otimes 2^{R_d}} \Big)^{\otimes 2^{R_1}} \otimes \rho_u^C \Big\|_1 \\ &\leq \frac{1}{2} \Big\| \sum_u p(u)|u\rangle\langle u|^U \otimes \frac{1}{2^{R_1}} \sum_{j=1}^{2^{R_1}} \Upsilon_u^{C,j} - \sum_u p(u)|u\rangle\langle u|^U \otimes \Xi_u^C \Big\|_1 \\ &\qquad + \frac{1}{2} \Big\| \sum_u p(u)|u\rangle\langle u|^U \otimes \Xi_u^C - \sum_u p(u)|u\rangle\langle u|^U \otimes \Big(\rho_u^{V^C} \otimes (\rho_v^{X^C})^{\otimes 2^{R_d}} \Big)^{\otimes 2^{R_1}} \otimes \rho_u^C \Big\|_1. \end{aligned}$$

We now try to upper bound each term appearing on the right-hand side. For the first term, simply by expanding the summation and subtracting equal terms from both side, it can be seen that:

$$\begin{aligned} &\frac{1}{2} \Big\| \sum_u p(u)|u\rangle\langle u|^U \otimes \big(\frac{1}{2^{R_1}} \sum_{j=1}^{2^{R_1}} \Upsilon_u^{C,j} - \Xi_u^C \big) \Big\|_1 \\ &= \frac{1}{2} \sum_u p(u) \Big\| \sum_v p(v|u)|v\rangle\langle v|^V \otimes \\ &\qquad \Big(\frac{1}{2^{R_d}} \sum_{i=1}^{2^{R_d}} \rho_v^X \otimes \ldots \otimes \rho_v^{X_i C} \otimes \ldots \otimes \rho_v^{X_{2^{R_d}}} - \rho_v^{X_1} \otimes \ldots \otimes \rho_v^{X_{2^{R_d}}} \otimes \rho_v^C \Big) \Big\|_1. \end{aligned}$$

Then immediately by noting the Markov chain, the conditional convex-split lemma asserts that if $R_d = \widetilde{I}_{\max}^{\varepsilon''}(X;C|V)_\rho + 2\log_2(\frac{1}{\eta})$, then

$$P\big(\sum_u p(u)|u\rangle\langle u|^U \otimes \frac{1}{2^{R_1}} \sum_{j=1}^{2^{R_1}} \Upsilon_u^{C,j}, \sum_u p(u)|u\rangle\langle u|^U \otimes \Xi_u^C \big) \leq 2\varepsilon'' + \eta.$$

and from the relation between the purified distance and the trace distance, we have

$$\frac{1}{2}\Big\| \sum_u p(u)|u\rangle\langle u|^U \otimes \big(\frac{1}{2^{R_1}} \sum_{j=1}^{2^{R_1}} \Upsilon_u^{C,j} - \Xi_u^C\big)\Big\|_1 \leq 2\varepsilon'' + \eta,$$

For the second term, from the invariance of the trace distance with respect to tensor product states, we can trace out X systems from both expressions leading to the following:

$$\frac{1}{2}\Bigg\| \sum_u p(u)|u\rangle\langle u|^U \otimes \left(\frac{1}{2^{R_1}} \sum_{j=1}^{R_1} (\rho_u^{V_1^C} \otimes ... \otimes \rho_u^{V_i^C C} \otimes ... \otimes \rho_u^{V_{R_1}^C}) - \rho_u^{V_1^C} \otimes ... \otimes \rho_u^{V_{2^{R_1}}^C} \otimes \rho_u^C\right)\Bigg\|_1,$$

then the conditional convex-split lemma guarantees the purified distance between states to be less than or equal to $(2\varepsilon'' + \eta)$ if we choose $R_1 = \widetilde{I}_{\max}^{\varepsilon''}(V;C|U)_\rho + 2\log_2(\frac{1}{\eta})$, which in turn, implies that the trace distance between the states is also less than or equal to $(2\varepsilon'' + \eta)$.

Derandomization: The proposed protocol relies upon shared randomness among parties. In order to show that the results also hold without assistance of shared randomness, the code needs to be derandomized. Derandomization is a standard procedure which can be done by expanding the states and corresponding POVM's and using a property of the trace distance given by the equality in (1.3) (see [55], [100], [125]). The only point that might be needed to be made here is the structure of the test operators in Bob's decoders (as well as that of Charlie). Note than the test operators were described generally as T^{UB} and Q^{UVB} without specifying the nature of the subsystems, i.e., whether each of U, V or B systems are classical or quantum. For our purposes, it is sufficient to consider the test operators as $T^{UB} \coloneqq \sum_u |u\rangle\langle u|^U \otimes \overline{T}_u^B$ where $\overline{T}_u^B \coloneqq \langle u|T^{UB}|u\rangle$. Likewise, we only need to have $Q^{UVB} \coloneqq \sum_{u,v} |u\rangle\langle u|^U \otimes |v\rangle\langle v|^V \otimes \overline{Q}_{u,v}^B$ where $\overline{Q}_{u,v}^B \coloneqq \langle u, v|Q^{UVB}|v, u\rangle$.

Expurgation: So far we have come to know that there exists at least one code that satisfies the reliability criterion in (5.8) and at least one codebook that satisfies the secrecy requirement (5.9). We should use Markov inequality to find a good code that satisfies both the reliability (5.8) and secrecy (5.9) simultaneously. We have the average error probability over all codes $P_{\text{error}}^1 \leq 3\varepsilon' + 2\sqrt{\varepsilon'}$ (with $2\sqrt{\varepsilon'}$ coming from the gentle measurement lemma) and the secrecy over all code $P_{\text{secrecy}}^1 \leq 4\varepsilon'' + 2\eta$. From Markov inequality we know that $\Pr(P_{\text{error}}^1 \geq \sqrt[4]{\varepsilon'}) \leq 3(\varepsilon')^{3/4} + 2\sqrt[4]{\varepsilon'}$ and $\Pr(P_{\text{secrecy}}^1 \geq \sqrt[4]{\varepsilon''}) \leq 4(\varepsilon'')^{3/4} + 2(\varepsilon'')^{7/4}$.

Then there is a good code for which, with high probability neither statement is true:

$$\begin{aligned}\Pr(P^1_{\text{error}} \leq \sqrt[4]{\varepsilon'}, P^1_{\text{secrecy}} \leq \sqrt[4]{\varepsilon''}) \\ \geq 1 - (3(\varepsilon')^{3/4} + 2\sqrt[4]{\varepsilon'}) - (4(\varepsilon'')^{3/4} + 2(\varepsilon'')^{7/4}).\end{aligned}$$

Let $\varepsilon \coloneqq \max\{\sqrt[4]{\varepsilon'}, \sqrt[4]{\varepsilon''}\}$. This parameter works for both requirements and the results is concluded.

■

Lemma 5.9. *We have* $\bigcup \mathcal{R}^{(in)}(\rho) \subseteq \mathcal{R}^{\varepsilon}(\mathcal{N})$ *and the union is over all* ρ^{UVXBC} *arising from the channel.*

Proof. To prove the lemma we need to show that for all ρ arising from the channel $\mathcal{R}^{\text{in}}(\rho) \subseteq \mathcal{R}^{*}(\rho)$. While this can be proven in a standard way by Fourier-Motzkin elimination (see for example appendix D of [153]), we follow the approach of [128] to show the lemma. Note that from the definition of $\mathcal{R}^{\varepsilon}(\mathcal{N})$, if a quadruple $(R_0 + r_0, R_1 - r_0 - r_s + r_d, R_s + r_s, R_d - r_d) \in \mathcal{R}^{\varepsilon}(\mathcal{N})$ for some $r_0, r_s, r_d \geq 0$, then $(R_0, R_1, R_s, R_d) \in \mathcal{R}^{\varepsilon}(\mathcal{N})$ as well. Then one can find explicit values of (r_0, r_s, r_d) such that for any given $(R_0, R_1, R_s, R_d) \in \mathcal{R}^{(in)}(\rho)$ we have $(R_0 + r_0, R_1 - r_0 - r_s + r_d, R_s + r_s, R_d - r_d) \in \mathcal{R}^{*}(\rho)$. In fact for $R_1 < \widetilde{I}^{\varepsilon''}_{\max}(V;C|U)_\rho + 2\log_2(\frac{1}{\eta})$, the values $(r_0, r_s, r_d) \coloneqq (0, 0, \widetilde{I}^{\varepsilon''}_{\max}(V;C|U)_\rho + 2\log_2(\frac{1}{\eta}) - R_1)$ can be seen to satisfy equations (5.20) to (5.23) and thus imply $(R_0 + r_0, R_1 - r_0 - r_s + r_d, R_s + r_s, R_d - r_d) \in \mathcal{R}^{*}(\rho)$. This combined with Lemma 5.8 proves $\bigcup \mathcal{R}^{(\text{in})}(\rho) \subseteq \mathcal{R}^{\varepsilon}(\mathcal{N})$. ■

5.5 Converse

Consider the common message rate R_0 bound from (5.15). This bound was proved in [25] by relating the communication problem to a problem in binary hypothesis testing. We briefly explain the approach here. From the definition of the reliability given in (5.8) both $\Pr\{\tilde{M}_0 \neq M_0\} \leq \varepsilon$ and $\Pr\{(\hat{M}_0, \hat{M}_1, \hat{M}_s) \neq (M_0, M_1, M_s)\} \leq \varepsilon$ must be satisfied. We concentrate now on the fulfillment of the first condition, i.e., $\Pr\{\tilde{M}_0 \neq M_0\} \leq \varepsilon$. Consider the task of distinguishing between two quantum states $\rho^{\tilde{M}_0 M_0} = \frac{1}{2^{R_0}} \sum_{m_0} |m_0\rangle\langle m_0|^{\tilde{M}_0} \otimes |m_0\rangle\langle m_0|^{M_0}$ and $\rho^{\tilde{M}_0} \otimes \rho^{M_0}$ where the former is the null hypothesis and the latter the alternative hypothesis. It can be easily verified that $\Pr\{\tilde{M}_0 \neq M_0\} \leq \varepsilon$ implies that the type I error is less that or equal to ε and the type II error equals 2^{-R_0}. Then from the definition of the hypothesis-testing mutual information and the monotonicty of the hypothesis testing relative entropy with cptp maps,

we have $R_0 \leq I^{\varepsilon}_{\mathrm{H}}(M_0;B)_\rho$. Let $U \coloneqq M_0$, then the converse follows. The proof of $R_0 \leq I^{\varepsilon}_{\mathrm{H}}(M_0;C)_\rho$ follows the same argument.

We now analyze the condition $\Pr\{(\hat{M}_0,\hat{M}_1,\hat{M}_s) \neq (M_0,M_1,M_s)\} \leq \varepsilon$. Following a procedure similar to [55], we expand this expression as follows:

$$\begin{aligned}
\varepsilon &\geq \Pr\{(\hat{M}_0,\hat{M}_1,\hat{M}_s) \neq (M_0,M_1,M_s)\} \\
&= \sum_{m_0,m_1,m_s} p(m_0)p(m_1)p(m_s)\Pr\{(\hat{M}_0,\hat{M}_1,\hat{M}_s) \neq (m_0,m_1,m_s)|m_0,m_1,m_s\} \\
&= \sum_{m_0,m_1,m_s} p(m_0)p(m_1)p(m_s) \sum_{(m_0',m_1',m_s')\neq(m_0,m_1,m_s)} p(m_0',m_1',m_s'|m_0,m_1,m_s) \\
&\geq \sum_{m_0,m_1,m_s} p(m_0)p(m_1)p(m_s) \sum_{\substack{m_0',\\(m_1',m_s')\neq(m_1,m_s)}} p(m_0',m_1',m_s'|m_0,m_1,m_s) \\
&= \sum_{m_0,m_1,m_s} p(m_0)p(m_1)p(m_s) \sum_{(m_1',m_s')\neq(m_1,m_s)} p(m_1',m_s'|m_0,m_1,m_s) \\
&= \sum_{m_0} p(m_0)\Pr\{(\hat{M}_1,\hat{M}_s) \neq (M_1,M_s)|M_0 = m_0\}.
\end{aligned}$$

Notice that the final expression indicates the probability of erroneous detection of (M_s,M_1) when M_0 is transmitted. We find an upper bound on the sum rate of (M_s,M_1) by considering a binary hypothesis testing problem with null and alternative hypotheses given respectively as follows:

$$\begin{aligned}
\rho^{M_0\hat{M}_s\hat{M}_1M_sM_1} &\coloneqq \frac{1}{2^{R_0}}\sum_{m_0}|m_0\rangle\langle m_0|^{M_0}\otimes\rho_{m_0}^{\hat{M}_s\hat{M}_1M_sM_1}, \\
\rho^{\hat{M}_s\hat{M}_1-M_0-M_sM_1} &\coloneqq \frac{1}{2^{R_0}}\sum_{m_0}|m_0\rangle\langle m_0|^{M_0}\otimes\rho_{m_0}^{\hat{M}_s\hat{M}_1}\otimes\rho_{m_0}^{M_sM_1},
\end{aligned}$$

where $\rho_{m_0}^{\hat{M}_s\hat{M}_1M_sM_1} = \frac{1}{2^{R_s+R_1}}\sum_{m_sm_1}|m_sm_1\rangle\langle m_sm_1|^{\hat{M}_s\hat{M}_1}\otimes|m_sm_1\rangle\langle m_sm_1|^{M_sM_1}$. It can be easily verified that type I error is equivalent to $\sum_{m_0}p(m_0)\Pr\{(\hat{M}_1,\hat{M}_s)\neq(M_1,M_s)|M_0=m_0\}$ which is assumed to be less that or equal to ε. On the other hand, the type II error can be written as follows:

$$\begin{aligned}
&\sum_{m_0,m_s,m_1} p_{M_0}(m_0)p_{M_sM_1}(m_s,m_1)p_{\hat{M}_s\hat{M}_1}(m_s,m_1) \\
&\qquad = \frac{1}{2^{R_s+R_1}}\sum_{m_0,m_s,m_1} p_{M_0}(m_0)p_{\hat{M}_s\hat{M}_1}(m_s,m_1) \\
&\qquad = \frac{1}{2^{R_s+R_1}}.
\end{aligned}$$

Then we have the following:

$$R_s+R_1 \leq I^{\varepsilon}_{\mathrm{H}}(M_s,M_1;\hat{M}_s,\hat{M}_1|M_0)_\rho \leq I^{\varepsilon}_{\mathrm{H}}(M_s,M_1;B|M_0)_\rho,$$

where the first inequality stems from the definition of the conditional hypothesis testing mutual information and the second inequality is from monotonicity under cptp maps. Identifying the random variables $V \coloneqq (M_s, M_1)$ and $U \coloneqq M_0$ concludes the intended bound. So far we have dealt with the reliability condition and have derived (5.15) and (5.16).

Next we turn our attention to the secrecy criterion. The secrecy condition (5.7) requires that the state of the Charlie and the confidential message become close to a product state for every transmitted common message. In converse proof, we consider a less strict criterion such that we demand the aforementioned states to be close on average over the common messages. i.e.,

$$\frac{1}{2}\Big\|\rho^{CM_0M_s} - \frac{1}{2^{R_0}}\sum_{m_0}|m_0\rangle\langle m_0|^{M_0}\otimes\rho_{m_0}^{M_s}\otimes\sigma_{m_0}^{C}\Big\|_1 = \frac{1}{2^{R_0+R_s}}\sum_{m_0,m_s}\frac{1}{2}\|\rho_{m_0,m_s}^{C}-\sigma_{m_0}^{C}\|_1 \leq \varepsilon.$$

From the relation between the purified distance and the trace distance, the purified distance between the above-mentioned states is less that or equal to $\sqrt{2\varepsilon}$. Then from the definition of the smooth conditional relative entropy, it is easily checked that the following holds:

$$D_{\max}^{\sqrt{2\varepsilon}}(M_s;C|M_0)_\rho \coloneqq D_{\max}^{\sqrt{2\varepsilon}}\Big(\rho^{M_0M_sC}\Big\|\frac{1}{2^{R_0}}\sum_{m_0}|m_0\rangle\langle m_0|^{M_0}\otimes\rho_{m_0}^{M_s}\otimes\sigma_{m_0}^{C}\Big)_\rho = 0.$$

Therefore, in the quantity $D_{\max}^{\sqrt{2\varepsilon}}(M_s;C|M_0)_\rho = 0$, we define $U \coloneqq M_0$ and $V \coloneqq M_s$ to get $D_{\max}^{\sqrt{2\varepsilon}}(V;C|U)_\rho = 0$. Similarly, we let $V \coloneqq M_0$ and $X \coloneqq M_s$ to get $D_{\max}^{\sqrt{2\varepsilon}}(X;C|V)_\rho = 0$. Finally the bound on the rate of the confidential message (5.17) can be seen from the bound derived on $R_s + R_1$ and the preceding discussion.

5.6 Asymptotic Analysis

So far we have studied the scenario in which a quantum channel is available only once and the transmission was subject to some non-zero error and secrecy parameters. In the asymptotic regime, however, a memoryless channel is considered to be available for an unlimited number of uses; if we denote the uses of the channel by n, the one-shot scenario corresponds to $n = 1$ where in the asymptotic regime $n \to \infty$. Moreover, in the asymptotic regime, as long as the achievability bounds and weak converses are concerned, the error and secrecy parameters are assumed to be vanishing in the limit of many channel uses, i.e., $\varepsilon \to 0$ as $n \to \infty$. The following formally defines the rate region in the asymptotic regime from the one-shot rate region defined before:

$$\mathcal{R}_\infty(\mathcal{N}) \coloneqq \lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\mathcal{R}^\varepsilon(\mathcal{N}^{\otimes n}), \tag{5.31}$$

where the tensor power channel $\mathcal{N}^{\otimes n}$ indicates the n independent uses of the channel $\mathcal{N}$. In the following we first prove a theorem then we will recover several well-known results as corollaries.

Theorem 5.3. *The asymptotic rate region $\mathcal{R}_\infty(\mathcal{N})$ of the broadcast channel $\mathcal{N}^{A\to BC}$ for simultaneous transmission of common, individualized and confidential messages with a rate-limited randomness encoder is given as follows:*

$$\mathcal{R}_\infty(\mathcal{N}) = \bigcup_{\ell=1}^{\infty} \frac{1}{\ell} \mathcal{R}_\infty^{(1)}(\mathcal{N}^{\otimes \ell}), \tag{5.32}$$

where $\mathcal{R}_\infty^{(1)}(\mathcal{N}) \coloneqq \bigcup_{\rho^{UVXBC}} \mathcal{R}_\infty^{(2)}(\mathcal{N})$, in which $\mathcal{R}_\infty^{(2)}(\mathcal{N})$ is the set of quadruples (R_0, R_1, R_s, R_d) satisfying the following conditions:

$$R_0 \le \min\left[I(U;B)_\rho, I(U;C)_\rho\right], \tag{5.33}$$

$$R_0 + R_1 + R_s \le I(V;B|U)_\rho + \min\left[I(U;B)_\rho, I(U;C)_\rho\right], \tag{5.34}$$

$$R_s \le I(V;B|U)_\rho - I(V;C|U)_\rho, \tag{5.35}$$

$$R_1 + R_d \ge I(V;C|U)_\rho + I(X;C|V)_\rho, \tag{5.36}$$

$$R_d \ge I(X;C|V)_\rho, \tag{5.37}$$

where (R_0, R_1, R_s, R_d) denotes the rates of the common, individualized, confidential and dummy messages, respectively and

$$\rho^{UVXBC} = \sum_{u,v,x} p(u)p(v|u)p(x|v)\, |u\rangle\langle u|^U \otimes |v\rangle\langle v|^V \otimes |x\rangle\langle x|^X \otimes \mathcal{N}(\rho_x^A),$$

is the state arising from the channel.

Proof of Theorem 5.3. We need to show the direct part and the converse. To establish the direct part, we appeal to our one-shot achievability region and seek to show that the right-hand side of equation (5.32) is contained inside the left-hand side, i.e., the following:

$$\bigcup_{\ell=1}^{\infty} \frac{1}{\ell} \mathcal{R}_\infty^{(1)}(\mathcal{N}^{\otimes \ell}) \subseteq \mathcal{R}_\infty(\mathcal{N}).$$

From our achievability result, Theorem 5.1, if we use the channel m times independently (memoryless channel), or equivalently if we consider one use of the tensor power channel $\mathcal{N}^{\otimes m}$, we will have:

$$\bigcup_{\rho^m} \mathcal{R}^{(\mathrm{in})}(\rho^m) \subseteq \mathcal{R}^{\varepsilon}(\mathcal{N}^{\otimes m}), \tag{5.38}$$

where $\mathcal{R}^{(\text{in})}(\rho^m)$ is the convex closure over all states ρ^m arising from m uses of the channel, of the rate quadruples (R_0, R_1, R_s, R_d) obeying the following:

$$R_0 \leq \min\left[I_{\text{H}}^{\varepsilon'-\delta_1}(U^m; B^{\otimes m})_{\rho^m} - \log_2\left(\frac{4\varepsilon'}{\delta_1^2}\right), I_{\text{H}}^{\varepsilon'-\delta_2}(U^m; C^{\otimes m})_{\rho^m} - \log_2\left(\frac{4\varepsilon'}{\delta_2^2}\right)\right],$$

$$R_0 + R_1 + R_s \leq I_{\text{H}}^{\varepsilon'-\delta_3}(V^m; B^{\otimes m}|U^m)_{\rho^m} - \log_2\left(\frac{4\varepsilon'}{\delta_3^2}\right)$$

$$+ \min\left[I_{\text{H}}^{\varepsilon'-\delta_1}(U^m; B^{\otimes m})_{\rho^m} - \log_2\left(\frac{4\varepsilon'}{\delta_1^2}\right), I_{\text{H}}^{\varepsilon'-\delta_2}(U^m; C^{\otimes m})_{\rho^m} - \log_2\left(\frac{4\varepsilon'}{\delta_2^2}\right)\right],$$

$$R_s \leq I_{\text{H}}^{\varepsilon'-\delta_3}(V^m; B^{\otimes m}|U^m)_{\rho^m} - \widetilde{I}_{\max}^{\varepsilon''}(V^m; C^{\otimes m}|U^m)_{\rho^m} - \log_2\left(\frac{4\varepsilon'}{\delta_1^2}\right) - 2\log_2\left(\frac{1}{\eta}\right),$$

$$R_1 + R_d \geq \widetilde{I}_{\max}^{\varepsilon''}(V^m; C^{\otimes m}|U^m)_{\rho^m} + \widetilde{I}_{\max}^{\varepsilon''}(X^m; C^{\otimes m}|V^m)_{\rho^m} + 4\log_2\left(\frac{1}{\eta}\right),$$

$$R_d \geq \widetilde{I}_{\max}^{\varepsilon''}(X^m; C^{\otimes m}|V^m)_{\rho^m} + 2\log_2\left(\frac{1}{\eta}\right),$$

where U^m, V^m and X^m refer to the random variables drawn from the joint distributions $p(u_1, ..., u_m), p(v_1, ..., v_m)$ and $p(x_1, ..., x_m)$, respectively and $B^{\otimes m}$ and $C^{\otimes m}$ refer to the m-fold tensor product of the Hilbert spaces B and C, respectively. Since we want to prove an achievability theorem, we can assume that each sequence of random variables is drawn from the corresponding distributions in an i.i.d. fashion, i.e., $p(u_1, ..., u_m) = \prod_{i=1}^m p(u_i)$, $p(v_1, ..., v_m) = \prod_{i=1}^m p(v_i)$ and $p(x_1, ..., x_m) = \prod_{i=1}^m p(x_i)$. Therefore, the state over which the above quantities are assessed, is $\rho^{\otimes m} = \rho \otimes ... \otimes \rho$.

The i.i.d. encoding assumption implies $\rho^m = \rho^{\otimes m}$ and enables us to simplify the entropic quantities in the asymptotic limit of many channel uses. To see this, we divide both sides of (5.38) by m and let $m \to \infty$:

$$\lim_{m\to\infty} \frac{1}{m} \bigcup_{\rho^{\otimes m}} \mathcal{R}^{(\text{in})}(\rho^{\otimes m}) \subseteq \lim_{\varepsilon\to 0} \lim_{m\to\infty} \frac{1}{m} \mathcal{R}^{\varepsilon}(\mathcal{N}^{\otimes m}), \tag{5.39}$$

This results in dividing the entropic quantities comprising $\mathcal{R}^{(\text{in})}(\rho^{\otimes m})$ by m and evaluate limits as $m \to \infty$. All the constant terms will vanish as $m \to \infty$ and from the asymptotic i.i.d. behavior of the quantities given in (1.9), (5.3) and Lemma 5.5, we get the region $\mathcal{R}_\infty^1(\mathcal{N})$. So far we have shown the following:

$$\mathcal{R}_\infty^1(\mathcal{N}) \subseteq \lim_{\varepsilon\to 0} \lim_{m\to\infty} \frac{1}{m} \mathcal{R}^{\varepsilon}(\mathcal{N}^{\otimes m}).$$

Finally we consider m uses of the tensor power channel $\mathcal{N}^{\otimes \ell}$ and let $n = m\ell$. Taking the limits as $n \to \infty$ concludes the direct part.

For the converse part, from Theorem 5.2 onward, if the channel $\mathcal{N}$ gets used m independent times, we will have

$$\mathcal{R}^{\varepsilon}(\mathcal{N}^{\otimes m}) \subseteq \bigcup_{\ell=1}^{m} \bigcup_{\rho^{\ell}} \mathcal{R}^{(\mathrm{co})}(\mathcal{N}^{\otimes \ell}), \tag{5.40}$$

where $\mathcal{R}^{(\mathrm{co})}(\mathcal{N}^{\otimes \ell})$ consists of the rate quadruples (R_0, R_1, R_s, R_d) obeying the following:

$$\begin{aligned}
R_0 &\leq \min\left[I_{\mathrm{H}}^{\varepsilon}(U^{\ell}; B^{\otimes \ell})_{\rho^{\ell}}, I_{\mathrm{H}}^{\varepsilon}(U^{\ell}; C^{\otimes \ell})_{\rho^{\ell}}\right], \\
R_0 + R_1 + R_s &\leq I_{\mathrm{H}}^{\varepsilon}(V^{\ell}; B^{\otimes \ell}|U^{\ell})_{\rho^{\ell}} + \min\left[I_{\mathrm{H}}^{\varepsilon}(U^{\ell}; B^{\otimes \ell})_{\rho^{\ell}}, I_{\mathrm{H}}^{\varepsilon}(U^{\ell}; C^{\otimes \ell})_{\rho^{\ell}}\right], \\
R_s &\leq I_{\mathrm{H}}^{\varepsilon}(V^{\ell}; B^{\otimes \ell}|U^{\ell})_{\rho^{\ell}} - D_{\max}^{\sqrt{2\varepsilon}}(V^{\ell}; C^{\otimes \ell}|U^{\ell})_{\rho^{\ell}}, \\
R_1 + R_d &\geq D_{\max}^{\sqrt{2\varepsilon}}(V^{\ell}; C^{\otimes \ell}|U^{\ell})_{\rho^{\ell}} + D_{\max}^{\sqrt{2\varepsilon}}(X^{\ell}; C^{\otimes \ell}|V^{\ell})_{\rho^{\ell}}, \\
R_d &\geq D_{\max}^{\sqrt{2\varepsilon}}(X^{\ell}; C^{\otimes \ell}|V^{\ell})_{\rho^{\ell}},
\end{aligned}$$

where $(\rho^{UVXBC})^{\ell}$ is the state inducing by ℓ independent uses of the channel such that its classical systems, U^{ℓ}, V^{ℓ} and X^{ℓ} correspond to the random variables drawn from the joint distributions $p(u_1, ..., u_{\ell}), p(v_1, ..., v_{\ell})$ and $p(x_1, ..., x_{\ell})$, respectively and quantum systems $B^{\otimes \ell}$ and $C^{\otimes \ell}$ refer to the ℓ-fold tensor product of the Hilbert spaces B and C, respectively. We can now consider t i.i.d. uses of the superchannel $\mathcal{N}^{\otimes m}$ for large t. This means we evaluate the region over "tensor product" states $(\rho^{m})^{\otimes t}$ and divide both sides of (5.40) by t and let $t \to \infty$. By invoking the asymptotic results from (1.9),(5.3) and (5.5), $\mathcal{R}^{(\mathrm{co})}(\mathcal{N}^{\otimes \ell})$ can be seen to be included in the following region:

$$\begin{aligned}
R_0 &\leq \min\left[I(U^{\ell}; B^{\otimes \ell})_{\rho^{\ell}}, I(U^{\ell}; C^{\otimes \ell})_{\rho^{\ell}}\right], \\
R_0 + R_1 + R_s &\leq I(V^{\ell}; B^{\otimes \ell}|U^{\ell})_{\rho^{\ell}} + \min\left[I(U^{\ell}; B^{\otimes \ell})_{\rho^{\ell}}, I(U^{\ell}; C^{\otimes \ell})_{\rho^{\ell}}\right], \\
R_s &\leq I(V^{\ell}; B^{\otimes \ell}|U^{\ell})_{\rho^{\ell}} - I(V^{\ell}; C^{\otimes \ell}|U^{\ell})_{\rho^{\ell}}, \\
R_1 + R_d &\geq I(V^{\ell}; C^{\otimes \ell}|U^{\ell})_{\rho^{\ell}} + I(X^{\ell}; C^{\otimes \ell}|V^{\ell})_{\rho^{\ell}}, \\
R_d &\geq I(X^{\ell}; C^{\otimes \ell}|V^{\ell})_{\rho^{\ell}}.
\end{aligned}$$

The proof will be completed by dividing both sides of (5.40) by m and letting $m \to \infty$ as well as $\varepsilon \to 0$. ∎

Corollary 5.1 (Theorem 1 in [20])**.** *Consider the quantum channel $\mathcal{N}^{A \to B}$ with an isometric extension $V^{A \to BE}$ and let $\rho^{URA} = \sum_u p(u) |u\rangle\langle u| \otimes |\phi_u\rangle \langle \phi_u|^{RA}$ be a classical-quantum state in which R is a reference system. The capacity*

region of simultaneous transmission of classical and quantum information for the channel is given by

$$S^{\infty}(\mathcal{N}) = \bigcup_{\ell=1}^{\infty} \frac{1}{\ell} S_1^{\infty}(\mathcal{N}^{\otimes \ell}),$$

where $S_1^{\infty}(\mathcal{N})$ *is the union, over all states of the form* $\rho^{URB} = \sum_u p(u) |u\rangle\langle u| \otimes \mathcal{N}^{A\to B}(|\phi_u\rangle \langle\phi_u|^{RA})$ *arising from the channel, of the rate pairs* $(R_c^{\infty}, R_q^{\infty})$ *obeying:*

$$R_c^{\infty} \leq I(U;B)_{\rho},$$
$$R_q^{\infty} \leq I(R\rangle BU)_{\rho},$$

where R_c^{∞} *and* R_q^{∞} *denote respectively the rates of the classical and quantum information and* $I(R\rangle BU)_{\rho} \coloneqq -H(R|BU)_{\rho}$ *is the coherent information.*

Proof. Following the discussion of Corollary 5.1 and Theorem 5.3, we only need to argue that the coherent information of the ensemble $\{p(u), |\phi_u\rangle \langle\phi_u|^{RBE}\}$ is equal to the rate of the confidential message in Theorem 3, i.e., the following:

$$I(R\rangle BU)_{\rho} = I(V;B|U)_{\rho} - I(V;E|U)_{\rho}.$$

We apply the Schmidt decomposition to the pure states $\{|\phi_u\rangle^{RBE}\}_u$ with respect to the cut $R|BE$ and then measure the R system in a suitable orthonormal basis. This measurement decoherifies the states such that the R system can be shown by a classcial system, say V. Then the equality of the coherent information and the confidential message rate can be easily checked (see for example exercise 11.6.7 in [101]). ∎

Corollary 5.2 (Theorem 3 of [128])**.** *Let* $\mathcal{N}_C^{X\to(Y,Z)}$ *be a classical channel taking inputs to outputs according to some distribution* $p(y,z|x)$*. Moreover, let* $\mathcal{R}^{\infty}(\mathcal{N}_C)$ *be the capacity region of* $\mathcal{N}_C^{X\to(Y,Z)}$ *for simultaneous transmission of the common, individualized and confidential messages with a rate-limited randomness encoder, defined similar to (5.31). Then there exist random variables* U *and* V *satisfying* $p(u,v,x,y,z) = \sum p(u,v)p(x|v)p(y,z|x)$ *such that* $\mathcal{R}^{\infty}(\mathcal{N}_C)$ *equals the union over all distributions of rate quadruples* (R_0, R_1, R_s, R_d) *obeying:*

$$\begin{aligned}
R_0 &\leq \min\big[I(U;Y)_p, I(U;Z)_p\big],\\
R_0 + R_1 + R_s &\leq I(V;Y|U)_p + \min\big[I(U;Y)_p, I(U;Z)_p\big],\\
R_s &\leq I(V;Y|U)_p - I(V;Z|U)_p,\\
R_1 + R_d &\geq I(V;Z|U)_p + I(X;Z|V)_p,\\
R_d &\geq I(X;Z|V)_p,
\end{aligned}$$

where (R_0, R_1, R_s, R_d) *denotes the rates of the common, individualized, confidential and dummy messages, respectively.*

Proof. This is a simple corollary of Theorem 5.3. If we assume the channel outputs B and C are classical, then we know that all systems will be simultaneously diagonalizable and the regularization is not needed. Letting $Y \coloneqq B$ and $Z \coloneqq C$ finishes the proof. ∎

In the following corollary we recover a result for quantum broadcast channel without any secrecy requirement [150].

Corollary 5.3 (Theorem in [150])**.** *Consider the quantum broadcast channel* $\mathcal{N}^{A\to BC}$*. The capacity region for the transmission of common and individualized messages of* $\mathcal{N}$*, denoted by* $C^{\infty}(\mathcal{N})$*, is given as follows*[6]*:*

$$C^{\infty}(\mathcal{N}) = \bigcup_{\ell=1}^{\infty} \frac{1}{\ell} C_1^{\infty}(\mathcal{N}^{\otimes \ell}),$$

where $C_1^{\infty}(\mathcal{N})$ *is the union over all states* ρ^{UVBC} *arising from the channel, of the rate pairs* (R_0, R_1) *obeying*

$$\begin{aligned} R_0 &\leq \min\left[I(U;B)_\rho, I(U;C)_\rho\right], \\ R_0 + R_1 &\leq I(V;B|U)_\rho + \min\left[I(U;B)_\rho, I(U;C)_\rho\right]. \end{aligned}$$

Proof. By dropping the secrecy requirement, the rate of the confidential message in Theorem 5.3 will add up to that of the individualized message. Note that this region is slightly different in appearance compared to the Theorem 1 in [150]. However, the discussion leading to the equations (17) and (18) in that paper indicates their equivalence: part (or whole) of the common message may contain information intended for Charlie such that Bob does not have any interest in learning those information; this leads to a slightly different region but the scenario and the rate region are essentially the same in that in superposition coding Bob is supposed to decode the common message in whole and maybe ignore its content afterwards. ∎

5.7 Conclusion

We have studied the interplay between common, individualized and confidential messages with rate-limited randomness in the one-shot regime of a quantum broadcast channel. To establish our achievability results, we have proved

[6] This is defined similar to (5.31).

a conditional version of the convex-split lemma whereby we have shown the channel resolvability problem in the one-shot regime via superpositions. To assess the tightness of our achievability region, we have also derived a (weak) converse region. By evaluating our rate regions in the asymptotic i.i.d setting, we recovered several well-known results in the literature.

Appendix

Proof of lemmas

We need the following auxiliary lemmas.

Lemma 5.10 (Lemma 17 in [154]). *Let $\rho \in \mathcal{S}_{\leq}^{A}$ and Π a projector on A, then*

$$\mathrm{P}(\rho, \Pi\rho\Pi) \leq \sqrt{2\,\mathrm{Tr}\,\rho\Pi_{\perp} - (\mathrm{Tr}\,\rho\Pi_{\perp})^2},$$

where $\Pi_{\perp} = \mathbb{1} - \Pi$.

Lemma 5.11 (corollary 16 in [154]). *Let $\rho^{AB} = |\varphi\rangle\langle\varphi|^{AB} \in \mathcal{S}^{AB}$ be a pure state, $\rho^{A} = \mathrm{Tr}_{B}\,\rho^{AB}$, $\rho^{B} = \mathrm{Tr}_{A}\,\rho^{AB}$ and let $\Pi^{A} \in \mathcal{P}^{A}$ be a projector in $\mathrm{supp}(\rho^{A})$. Then, there exists a dual projector Π^{B} on B such that*

$$(\Pi^{A} \otimes (\rho^{B})^{-\frac{1}{2}})\,|\varphi\rangle^{AB} = ((\rho^{A})^{-\frac{1}{2}} \otimes \Pi^{B})\,|\varphi\rangle^{AB}.$$

Lemma 5.12 ([24]). *Let $\rho, \sigma \in \mathcal{P}$, then*

- *For any $\omega \geq \rho$,*

$$\left\|\sqrt{\omega}\sqrt{\sigma}\right\|_{1} \geq \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_{1}.$$

- *For any projector $\Pi \in \mathcal{P}$,*

$$\begin{aligned}\left\|\sqrt{\Pi\rho\Pi}\sqrt{\sigma}\right\|_{1} &= \left\|\sqrt{\rho}\sqrt{\Pi\sigma\Pi}\right\|_{1} \\ &= \left\|\sqrt{\Pi\rho\Pi}\sqrt{\Pi\sigma\Pi}\right\|_{1}.\end{aligned}$$

To prove Lemma 5.2, we need the following lemma.

Lemma 5.13. *For quantum states ρ^{AB} and σ^{B}, there exists a state $\rho'^{A} \in \mathcal{B}^{\varepsilon}(\rho^{A})$ such that:*

$$D_{\max}(\rho^{AB} \| \rho'^{A} \otimes \sigma^{B}) \leq D_{\max}(\rho^{AB} \| \rho^{A} \otimes \sigma^{B}).$$

Proof. Trivial. ∎

Proof of lemma 5.2. In the result of Lemma 5.13, let ρ^{*AB} be the optimizer in the definition of $\widetilde{I}^{\varepsilon}_{\max}(A;B)_{\rho}$, by substituting this state we will have,

$$D_{\max}(\rho^{*AB}\|\rho'^{A}\otimes\sigma^{B}) \leq D_{\max}(\rho^{*AB}\|\rho^{*A}\otimes\sigma^{B}).$$

Let $\sigma^{B} \coloneqq \rho^{B}$ and choose $\rho'^{A} = \rho^{A}$ (this is possible since $\mathrm{P}(\rho^{A},\rho^{*A}) \leq \varepsilon$) and then

$$D_{\max}(\rho^{*AB}\|\rho^{A}\otimes\rho^{B}) \leq D_{\max}(\rho^{*AB}\|\rho^{*A}\otimes\rho^{B}).$$

Then the result follows by definitions of the quantities. ∎

We need the following lemma to prove Lemma 5.3.

Lemma 5.14. *For quantum states $\rho^{XAB} = \sum_x p(x)\,|x\rangle\langle x|\otimes\rho_x^{AB}$ and $\sigma^{XAB} = \sum_x q(x)\,|x\rangle\langle x|\otimes\sigma_x^{A}\otimes\sigma_x^{B}$, there exists a state $\rho'^{XAB}\in\mathcal{B}^{\varepsilon}(\rho^{XAB})$ classical on X such that:*

$$D_{\max}\Big(\rho'^{XAB}\Big\|\sum_x p'(x)\,|x\rangle\langle x|\otimes\rho_x'^{A}\otimes\sigma_B^{x}\Big) \leq D_{\max}\Big(\rho^{XAB}\Big\|\sum_x q(x)\,|x\rangle\langle x|\otimes\sigma_x^{A}\otimes\sigma_x^{B}\Big) + \log\left(\frac{1}{1-\sqrt{1-\varepsilon^2}}+1\right).$$

Proof. The proof is inspired by [31] and [120]. Let ρ^{XABC} be a purification of ρ^{XAB} and $\varepsilon > 0$. Further let $\Pi^{BC}\in BC$ be a projector that is defined as the dual projector of the minimum rank projector Π^{XA} with $\mathrm{supp}(\Pi^{XA}) \subseteq \mathrm{supp}(\rho^{XA})$. The projector Π^{XA} is set to minimize $\left\|\Pi^{XA}\Gamma^{XA}\Pi^{XA}\right\|_{\infty}$ while fulfilling $P(\rho^{XABC},\tilde{\rho}^{XABC}) \leq \varepsilon$ in which $\Gamma^{XA} \coloneqq (\rho^{XA})^{-\frac{1}{2}}\sigma^{XA}(\rho^{XA})^{-\frac{1}{2}}$ and $\tilde{\rho}^{XABC} \coloneqq \Pi^{BC}\rho^{XABC}\Pi^{BC}$. From Lemma 5.10, we know the following

$$\begin{aligned}\mathrm{P}(\rho^{XABC},\Pi^{BC}\rho^{XABC}\Pi^{BC}) &\leq \sqrt{2\,\mathrm{Tr}\,\Pi_{\perp}^{BC}\rho - (\mathrm{Tr}\,\Pi_{\perp}^{BC}\rho)^2}\\ &= \sqrt{2\,\mathrm{Tr}\,\Pi_{\perp}^{XA}\rho - (\mathrm{Tr}\,\Pi_{\perp}^{XA}\rho)^2}.\end{aligned}$$

If we let $\mathrm{Tr}\,\Pi_{\perp}^{XA}\rho \leq 1-\sqrt{1-\varepsilon^2}$, then we will have $P(\rho^{XABC},\tilde{\rho}^{XABC}) \leq \varepsilon$ since $t \mapsto \sqrt{2t-t^2}$ is monotonically increasing over $[0,1]$. Now we choose Π^{XA} to be the projector onto the smallest eigenvalues of Γ^{XA} such that the aforementioned restriction holds, which in turn, results in the minimization of $\left\|\Pi^{XA}\Gamma^{XA}\Pi^{XA}\right\|_{\infty}$. Let Π'^{XA} denote the projector onto the largest remaining eigenvalue of $\Pi^{XA}\Gamma^{XA}\Pi^{XA}$. Notice that Π^{XA} and Π'^{XA} commute with Γ^{XA}. Then we have the following:

$$\left\|\Pi^{XA}\Gamma^{XA}\Pi^{XA}\right\|_{\infty} = \mathrm{Tr}(\Pi'^{XA}\Gamma^{XA}) = \min_{\mu^{XA}}\frac{\mathrm{Tr}(\mu^{XA}\Gamma^{XA})}{\mathrm{Tr}\,\mu^{XA}},$$

where the minimization is over all operators in the support of $\Pi'^{XA} + \Pi_\perp^{XA}$. Choosing $\mu^{XA} = (\Pi'^{XA} + \Pi_\perp^{XA})\rho^{XA}(\Pi'^{XA} + \Pi_\perp^{XA})$, we will have:

$$\left\|\Pi^{XA}\Gamma^{XA}\Pi^{XA}\right\|_\infty \leq \frac{\mathrm{Tr}\{(\Pi'^{XA} + \Pi_\perp^{XA})\rho^{XA}(\Pi'^{XA} + \Pi_\perp^{XA})\Gamma^{XA}\}}{\mathrm{Tr}\{(\Pi'^{XA} + \Pi_\perp^{XA})\rho^{XA}(\Pi'^{XA} + \Pi_\perp^{XA})\}} \leq \frac{1}{1-\sqrt{1-\varepsilon^2}},$$

where from the fact that Π'^{XA} and $\Pi_\perp^{XA}$ commute with Γ^{XA}, we have $\mathrm{Tr}\{(\Pi'^{XA} + \Pi_\perp^{XA})\rho^{XA}(\Pi'^{XA} + \Pi_\perp^{XA})\Gamma^{XA}\} = \mathrm{Tr}\{(\Pi'^{XA} + \Pi_\perp^{XA})(\rho^{XA})^{1/2}\Gamma^{XA}(\rho^{XA})^{1/2}\} \leq \mathrm{Tr}\{(\rho^{XA})^{1/2}\Gamma^{XA}(\rho^{XA})^{1/2}\} = \mathrm{Tr}\,\sigma^{XA} = 1$. Moreover, the definition of Π^{XA} implies that $\mathrm{Tr}\{(\Pi'^{XA} + \Pi_\perp^{XA})\rho^{XA}\} \geq 1 - \sqrt{1-\varepsilon^2}$. Let $\gamma \coloneqq D_{\max}\big(\rho^{XAB}\big\|\sum_x q(x)\,|x\rangle\langle x| \otimes \sigma_x^A \otimes \sigma_x^B\big)$ and $\sigma^{X-B} \coloneqq \sum_x |x\rangle\langle x| \otimes \sigma_x^B$. For state $\tilde{\rho}^{XABC}$ introduced above, we can write:

$$\begin{aligned}
&D_{\max}(\tilde{\rho}^{XAB}\|\sum_x p(x)\,|x\rangle\langle x| \otimes \rho_x^A \otimes \sigma_x^B)\\
&= \log\Big\|\Big(\sum_x p(x)\,|x\rangle\langle x| \otimes \rho_x^A \otimes \sigma_x^B\Big)^{-\frac{1}{2}}\tilde{\rho}^{XAB}\Big(\sum_x p(x)\,|x\rangle\langle x| \otimes \rho_x^A \otimes \sigma_x^B\Big)^{-\frac{1}{2}}\Big\|_\infty\\
&= \log\Big\|\Big(\sum_x p(x)\,|x\rangle\langle x| \otimes \rho_x^A \otimes \sigma_x^B\Big)^{-\frac{1}{2}}\,\mathrm{Tr}_C\{\Pi^{BC}\rho^{XABC}\Pi^{BC}\}\Big(\sum_x p(x)\,|x\rangle\langle x| \otimes \rho_x^A \otimes \sigma_x^B\Big)^{-\frac{1}{2}}\Big\|_\infty\\
&= \log\Big\|(\sigma^{X-B})^{-\frac{1}{2}}\,\mathrm{Tr}_C\{(\rho^{XA})^{-\frac{1}{2}} \otimes \Pi^{BC}\rho^{XABC}(\rho^{XA})^{-\frac{1}{2}} \otimes \Pi^{BC}\}(\sigma^{X-B})^{-\frac{1}{2}}\Big\|_\infty\\
&= \log\Big\|(\sigma^{X-B})^{-\frac{1}{2}}(\rho^{XA})^{-\frac{1}{2}}\Pi^{XA}\rho^{XAB}(\rho^{XA})^{-\frac{1}{2}}\Pi^{XA}(\sigma^{X-B})^{-\frac{1}{2}}\Big\|_\infty\\
&\leq \log 2^\gamma\Big\|(\sigma^{X-B})^{-\frac{1}{2}}(\rho^{XA})^{-\frac{1}{2}}\Pi^{XA}(\sum_x q(x)\,|x\rangle\langle x| \otimes \sigma_x^A \otimes \sigma_x^B)(\rho^{XA})^{-\frac{1}{2}}\Pi^{XA}(\sigma^{X-B})^{-\frac{1}{2}}\Big\|_\infty\\
&= \log 2^\gamma\Big\|(\rho^{XA})^{-\frac{1}{2}}\Pi^{XA}\sum_x q(x)\,|x\rangle\langle x| \otimes \sigma_x^A \otimes (\sigma_x^B)^{-\frac{1}{2}}\sigma_x^B(\sigma_x^B)^{-\frac{1}{2}}(\rho^{XA})^{-\frac{1}{2}}\Pi^{XA}\Big\|_\infty\\
&= \log 2^\gamma\Big\|(\rho^{XA})^{-\frac{1}{2}}\Pi^{XA}\sum_x q(x)\,|x\rangle\langle x| \otimes \sigma_x^A \otimes \mathbb{1}^B(\rho^{XA})^{-\frac{1}{2}}\Pi^{XA}\Big\|_\infty\\
&= \gamma + \log\left\|\Pi^{XA}\Gamma^{XA}\Pi^{XA}\right\|_\infty\\
&\leq D_{\max}\big(\rho^{XAB}\big\|\sum_x q(x)\,|x\rangle\langle x| \otimes \sigma_x^A \otimes \sigma_x^B\big) + \log\frac{1}{1-\sqrt{1-\varepsilon^2}}.
\end{aligned}$$

Define the positive semi-definite operator $\kappa^{XA} \coloneqq \rho^{XA} - \tilde{\rho}^{XA}$ and Let $\bar{\rho}^{XAB} \coloneqq \tilde{\rho}^{XAB} + \kappa^{XA} \otimes \sigma^{X-B}$. It can be easily checked that $\bar{\rho}^{XA} = \rho^{XA}$. Moreover, in the following we show that $P(\bar{\rho}^{XAB}, \rho^{XAB}) \leq \varepsilon$:

$$\begin{aligned}
F(\bar{\rho}^{XAB}, \rho^{XAB}) &\geq \left\|\sqrt{\tilde{\rho}^{XAB}}\sqrt{\rho^{XAB}}\right\|_1 + 1 - \mathrm{Tr}\,\rho^{XAB}\\
&\geq \left\|\sqrt{\tilde{\rho}^{XABC}}\sqrt{\rho^{XABC}}\right\|_1 + 1 - \mathrm{Tr}\,\rho^{XAB}\\
&= 1 - \mathrm{Tr}\,\Pi_\perp^{BC}\rho^{BC}\\
&\geq \sqrt{1-\varepsilon^2}.
\end{aligned}$$

The first inequality follows from Lemma 5.12 and the fact that by construction $\tilde{\rho}^{XAB} \leq \bar{\rho}^{XAB}$, therefore $\left\|\sqrt{\tilde{\rho}^{XAB}}\sqrt{\rho^{XAB}}\right\|_1 \leq \left\|\sqrt{\bar{\rho}^{XAB}}\sqrt{\rho^{XAB}}\right\|_1$. The second inequality follows from the fact that fidelity is monotonically non-decreasing with respect to cptp maps. The equality stems from Lemma 5.12 and the last inequality is the assumption. And finally from the relation between the purified distance and the fidelity the desired inequality follows. We continue as follows:

$$
\begin{aligned}
&D_{\max}\big(\bar{\rho}^{XAB}\big\|\bar{\rho}^{XA}\otimes\sigma^{X-B}\big)\\
&=\log\left\|(\bar{\rho}^{XA})^{-\frac{1}{2}}\otimes(\sigma^{X-B})^{-\frac{1}{2}}\bar{\rho}^{XAB}(\bar{\rho}^{XA})^{-\frac{1}{2}}\otimes(\sigma^{X-B})^{-\frac{1}{2}}\right\|_\infty\\
&=\log\left\|(\rho^{XA})^{-\frac{1}{2}}\otimes(\sigma^{X-B})^{-\frac{1}{2}}\bar{\rho}^{XAB}(\rho^{XA})^{-\frac{1}{2}}\otimes(\sigma^{X-B})^{-\frac{1}{2}}\right\|_\infty\\
&\leq\log\left(\left\|(\rho^{XA})^{-\frac{1}{2}}\otimes(\sigma^{X-B})^{-\frac{1}{2}}\tilde{\rho}^{XAB}(\rho^{XA})^{-\frac{1}{2}}\otimes(\sigma^{X-B})^{-\frac{1}{2}}\right\|_\infty+1\right)\\
&\leq\log\left(2^{\gamma}\frac{1}{1-\sqrt{1-\varepsilon^2}}+1\right)\\
&\leq D_{\max}\big(\rho^{XAB}\big\|\sum_x q(x)\,|x\rangle\langle x|\otimes\sigma_x^A\otimes\sigma_x^B\big)+\log\left(\frac{1}{1-\sqrt{1-\varepsilon^2}}+1\right),
\end{aligned}
$$

where in the first inequality we have used $\bar{\rho}^{XAB} \leq \tilde{\rho}^{XAB} + \rho^{XA}\otimes\sigma^B$ and in the final inequality we have used the fact that $2^\gamma \geq \operatorname{Tr}\rho^{XAB} = 1$. Now similar to Remark 5.1, a pinching map is applied to the left hand-hand side to conclude from the monotonicity of the max-relative entropy that X system is classical. ∎

Proof of Lemma 5.3. From the result given in Lemma 5.14 onward, let ρ^{*XAB} be the optimizer for $D^{\varepsilon}_{\max}\big(\rho^{XAB}\big\|\sum_x q(x)\,|x\rangle\langle x|\otimes\sigma_x^A\otimes\sigma_x^B\big)$. We argued that this state will be classical on X. Then there exists a state $\bar{\rho}^{XAB}\in\mathcal{B}^{\varepsilon}(\rho^{*XAB})$ classical on X such that

$$
\begin{aligned}
&D_{\max}\big(\bar{\rho}^{XAB}\big\|\sum_x \bar{p}(x)\,|x\rangle\langle x|\otimes\bar{\rho}_x^A\otimes\sigma_B^x\big)\\
&\leq D_{\max}\big(\rho^{*XAB}\big\|\sum_x q(x)\,|x\rangle\langle x|\otimes\sigma_x^A\otimes\sigma_x^B\big)+\log\left(\frac{1}{1-\sqrt{1-\varepsilon^2}}+1\right).
\end{aligned}
$$

From the triangle inequality for the purified distance it is seen that $\bar{\rho}^{XAB}\in\mathcal{B}^{2\varepsilon}(\rho^{XAB})$. Choosing $q(x)=p(x), \sigma_x^A=\rho_x^A, \sigma_x^B=\rho_x^B$ for all x, finishes the job. ∎

To prove Lemma 5.4, we need to following lemma.

Lemma 5.15. *Let ρ^{XAB} and σ^B be a quantum states. There exists a state $\rho'^{XAB}\in\mathcal{B}^{\varepsilon}(\rho)$ classical on X such that:*

$$
D_{\max}(\rho^{XAB}\|\sum_x p'(x)\,|x\rangle\langle x|\otimes\rho_x'^A\otimes\sigma_x^B)\leq D_{\max}(\rho^{XAB}\|\sum_x p(x)\,|x\rangle\langle x|\otimes\rho_x^A\otimes\sigma_x^B).
$$

Proof. Trivial. ∎

proof of Lemma 5.4. Let ρ^{*XAB} be the optimizer in the definition of the PSCMMI. By substituting it in Lemma 5.15, we will have:

$$D_{\max}(\rho^{*XAB}\|\sum_x p'(x)\,|x\rangle\langle x|\otimes\rho_x'^A\otimes\sigma_x^B)\leq D_{\max}(\rho^{*XAB}\|\sum_x p^*(x)\,|x\rangle\langle x|\otimes\rho_x^{*A}\otimes\sigma_x^B).$$

Let $\rho'^{XA}=\rho^{XA}$ and $\sigma^B=\rho^B$. Then the result follows from the definition of the quantities. ∎

Proof of Lemma 5.6. Similar to Lemma 11 in [115], the proof follows by straightforward calculation as shown below:

$$\begin{aligned}
&\sum_i p(i)\left(D(\rho_i^{XA}\|\theta^{XA})-D(\rho_i^{XA}\|\rho^{XA})\right)\\
&=\sum_i p(i)\left(\mathrm{Tr}\{\rho_i^{XA}\log\rho_i^{XA}\}-\mathrm{tr}\{\rho_i^{XA}\log\theta^{XA}\}-\mathrm{Tr}\{\rho_i^{XA}\log\rho_i^{XA}\}+\mathrm{Tr}\{\rho_i^{XA}\log\rho^{XA}\}\right)\\
&=\mathrm{Tr}\{\sum_i p(i)\rho_i^{XA}\log\rho^{XA}\}-\mathrm{Tr}\{\sum_i p(i)\rho_i^{XA}\log\theta^{XA}\}\\
&=\mathrm{Tr}\{\rho^{XA}\log\rho^{XA}\}-\mathrm{Tr}\{\rho^{XA}\log\theta^{XA}\}\\
&=D(\rho^{XA}\|\theta^{XA}).
\end{aligned}$$

∎

Proof of Lemma 5.7. The proof is similar to the proof of its uncontional version [115]. For the convenience sake, we let $\sigma_x^{B_{-j}}\coloneqq\sigma_x^{B_1}\otimes...\sigma_x^{B_{j-1}}\otimes\sigma_x^{B_{j+1}}\otimes...\otimes\sigma_x^{B_n}$ and $\sigma_x^{B_{+j}}\coloneqq\sigma_x^{B_1}\otimes...\otimes\sigma_x^{B_n}$. By adopting this notation, we can see that $\tau^{XAB_1...B_n}=\frac{1}{n}\sum_{j=1}^n\sum_x p(x)|x\rangle\langle x|^X\otimes\rho_x^{AB}\otimes\sigma_x^{B_{-j}}$. We use Lemma 5.6 to write the following:

$$\begin{aligned}
&D\big(\tau^{XAB_1...B_n}\|\sum_x p(x)|x\rangle\langle x|^X\otimes\rho_x^A\otimes\sigma_x^{B_{+j}}\big)\\
&=\frac{1}{n}\sum_j D\big(\sum_x p(x)|x\rangle\langle x|^X\otimes\rho_x^{AB_j}\otimes\sigma_x^{B_{-j}}\|\sum_x p(x)|x\rangle\langle x|^X\otimes\rho_x^A\otimes\sigma_x^{B_{+j}}\big) \quad (5.41)\\
&-\frac{1}{n}\sum_j D\big(\sum_x p(x)|x\rangle\langle x|^X\otimes\rho_x^{AB_j}\otimes\sigma_x^{B_{-j}}\|\tau^{XAB_1...B_n}\big). \quad (5.42)
\end{aligned}$$

From the invariance of the relative entropy with respect to tensor product states, the term inside the summation in (5.41) equals $D\big(\sum_x p(x)|x\rangle\langle x|^X\otimes\rho_x^{AB_j}\|\sum_x p(x)|x\rangle\langle x|^X\otimes\rho_x^A\otimes\sigma_x^{B_j}\big)$. Besides, from the monotonicity of the quantum relative entropy, by applying $\mathrm{Tr}_{B_1,...B_{j-1},B_{j+1},...,B_n}\{.\}$ to the term inside summation in (5.42), it is lower bounded by $D\big(\sum_x p(x)|x\rangle\langle x|^X\otimes\rho_x^{AB_j}\|\tau^{XAB_j}\big)$

where $\tau^{XAB_j} \coloneqq \sum_x p(x)|x\rangle\langle x|^X \otimes \big(\frac{1}{n}\rho_x^{AB_j} + (1-\frac{1}{n})(\rho_x^A \otimes \sigma_x^{B_j})\big)$. Let k be such that $\rho^{XAB_j} \leq 2^k \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^{B_j}$. Therefore, we will have $\rho^{XAB_j} \leq (1+\frac{2^k-1}{n}) \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^{B_j}$. Consider the following chain:

$$
\begin{aligned}
&D\big(\rho^{XAB_j}\big\|\tau^{XAB_j}\big) = \mathrm{Tr}\left\{\rho^{XAB_j}\log\rho^{XAB_j}\right\} - \mathrm{Tr}\left\{\rho^{XAB_j}\log\tau^{XAB_j}\right\} \\
&\geq \mathrm{Tr}\left\{\rho^{XAB_j}\log\rho^{XAB_j}\right\} - \mathrm{Tr}\left\{\rho^{XAB_j}\log\left(\sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^{B_j}\right)\right\} - \log\left(1+\frac{2^k-1}{n}\right) \\
&= D\big(\rho^{XAB_j}\big\|\sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^{B_j}\big) - \log\left(1+\frac{2^k-1}{n}\right),
\end{aligned}
$$

where the inequality comes from the fact that if A and B are positive semidefinite operators and $A \leq B$, then $\log A \leq \log B$. Plugging the findings above into (5.41) and (5.42) yields:

$$
\begin{aligned}
&D\big(\tau^{XAB_1\ldots B_n}\big\|\sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^{B_{+j}}\big) \\
&\leq \frac{1}{n}\sum_j D\big(\rho^{XAB_j}\big\|\sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^{B_j}\big) \\
&- \frac{1}{n}\sum_j D\big(\rho^{XAB_j}\big\|\sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^{B_j}\big) + \log\left(1+\frac{2^k-1}{n}\right) \\
&\leq \log\left(1+\frac{2^k}{n}\right).
\end{aligned}
$$

By choosing $n = \lceil\frac{2^k}{\delta^2}\rceil$, it follows that $D\big(\tau^{XAB_1\ldots B_n}\big\|\sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^{B_{+j}}\big) \leq \log(1+\delta^2)$. From Pinsker's inequality (1.2), we also can see that $F^2\big(\tau^{XAB_1\ldots B_n}, \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^{B_{+j}}\big) \geq \frac{1}{1+\delta^2} \geq 1-\delta^2$. From definition of the purified distance, it can be easily seen that $P\big(\tau^{XAB_1\ldots B_n}, \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^{B_{+j}}\big) \leq \delta$. ∎

Proof of Corollary 5.1. Let $\tilde{\rho}^{XAB}$ be the optimal state achieving the minimum for k. Then from the conditional convex-split lemma we know that:

$$
P(\tilde{\tau}^{XAB_1\ldots B_n}, \sum_x \tilde{p}(x)|x\rangle\langle x|^X \otimes \tilde{\rho}_x^A \otimes \sigma_x^{B_1} \otimes \ldots \otimes \sigma_x^{B_n}) \leq \delta, \tag{5.43}
$$

where

$$
\tilde{\tau}^{XAB_1\ldots B_n} \coloneqq \sum_x \tilde{p}(x)|x\rangle\langle x|^X \otimes \big(\frac{1}{n}\sum_{j=1}^n \tilde{\rho}_x^{AB_j} \otimes \sigma_x^{B_1} \otimes \ldots \otimes \sigma_x^{B_{j-1}} \otimes \sigma_x^{B_{j+1}} \otimes \sigma_x^{B_n}\big).
$$

From the concavity of the fidelity as well as its invariance with respect to tensor product states, the following can be seen:

$$P(\tilde{\tau}^{XAB_1\ldots B_n}, \tau^{XAB_1\ldots B_n}) \leq P(\tilde{\rho}^{XAB}, \rho^{XAB}) \leq \varepsilon. \tag{5.44}$$

Analogously, we have

$$P(\sum_x \tilde{p}(x)|x\rangle\langle x|^X \otimes \tilde{\rho}_x^A \otimes \sigma_x^{B_1} \otimes \ldots \otimes \sigma_x^{B_n}, \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^A \otimes \sigma_x^{B_1} \otimes \ldots \otimes \sigma_x^{B_n}) \leq P(\tilde{\rho}^{XA}, \rho^{XA}) \leq \epsilon. \tag{5.45}$$

Then the desired result is inferred by applying the triangle inequality to (5.43), (5.44) and (5.45). ■

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[2] I. Devetak, A. W. Harrow, and A. Winter, "A Resource Framework for Quantum Shannon Theory," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4587–4618, Oct 2008.

[3] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, "Concentrating partial entanglement by local operations," *Phys. Rev. A*, vol. 53, pp. 2046–2052, Apr 1996.

[4] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum entanglement," *Rev. Mod. Phys.*, vol. 81, no. 2, pp. 865–942, 2009.

[5] C. H. Bennett, S. Popescu, D. Rohrlich, J. A. Smolin, and A. V. Thapliyal, "Exact and asymptotic measures of multipartite pure-state entanglement," *Phys. Rev. A*, vol. 63, p. 012307, 2000.

[6] N. Linden, S. Popescu, B. Schumacher, and M. Westmoreland, "Reversibility of local transformations of multiparticle entanglement," *Quantum Inf. Proc.*, vol. 4, no. 3, pp. 241–250, 2005.

[7] M. Walter, D. Gross, and J. Eisert, "Multipartite Entanglement," in *Quantum Information: From Foundations to Quantum Technology Applications*, D. Bruss and G. Leuchs, Eds., 2016, ch. 14, pp. 293–330.

[8] M. Horodecki, J. Oppenheim, and A. Winter, "Partial quantum information," *Nature*, vol. 436, pp. 673–676, 2005.

[9] ——, "Quantum State Merging and Negative Information," *Commun. Math. Phys.*, vol. 269, no. 1, pp. 107–136, 2007.

[10] D. Yang and J. Eisert, "Entanglement Combing," *Phys. Rev. Lett.*, vol. 103, p. 220501, Nov 2009.

[11] H. J. Kimble, "The quantum internet," *Nature*, vol. 453, pp. 1023–1030, 2008.

[12] M. Skotiniotis and A. Winter, "Quantum Godwin's Law," 2020, arXiv[quant-ph]:2003.13715.

[13] A. D. Wyner, "The wire-tap channel," *The Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, Oct 1975.

[14] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[16] Y. Steinberg and S. Verdu, "Channel simulation and coding with side information," *IEEE Transactions on Information Theory*, vol. 40, no. 3, pp. 634–646, May 1994.

[17] I. CsiszÃąr and J. KÃűrner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[18] N. Cai, A. Winter, and R. W. Yeung, "Quantum privacy and quantum wiretap channels," *Problems of Information Transmission*, vol. 40, no. 4, pp. 318–336, Oct 2004. [Online]. Available: https://doi.org/10.1007/s11122-005-0002-x

[19] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 44–55, Jan 2005.

[20] I. Devetak and P. W. Shor, "The capacity of a quantum channel for simultaneous transmission of classical and quantum information," *Communications in Mathematical Physics*, vol. 256, no. 2, pp. 287–303, Jun 2005. [Online]. Available: https://doi.org/10.1007/s00220-005-1317-6

[21] W. F. Stinespring, "Positive functions on c*-algebras," *Proceedings of the American Mathematical Society*, vol. 6, no. 2, pp. 211–216, 1955.

[22] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On quantum Rényi entropies: A new generalization and some properties," *Journal of Mathematical Physics*, vol. 54, no. 12, pp. 122 203–122 203, Dec. 2013.

[23] M. Tomamichel, R. Colbeck, and R. Renner, "Duality between smooth min- and max-entropies," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4674–4681, 2010.

[24] R. Wilmink, "Quantum Broadcast Channels and Cryptographic Applications for Separable States," Ph.D. dissertation, Universität Bielefeld, Department of Mathematics, 2003. [Online]. Available: https://pub.uni-bielefeld.de/download/2303480/2303483/dissertation_wilmink_publikation.pdf

[25] L. Wang and R. Renner, "One-shot classical-quantum capacity and hypothesis testing," *Phys. Rev. Lett.*, vol. 108, p. 200501, May 2012. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.108.200501

[26] F. Buscemi and N. Datta, "The quantum capacity of channels with arbitrarily correlated noise," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1447–1460, March 2010.

[27] T. Ogawa and H. Nagaoka, "Strong converse and Stein's lemma in quantum hypothesis testing," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2428–2433, Nov 2000.

[28] F. Hiai and D. Petz, "The proper formula for relative entropy and its asymptotics in quantum probability," *Comm. Math. Phys.*, vol. 143, no. 1, pp. 99–114, 1991. [Online]. Available: https://projecteuclid.org:443/euclid.cmp/1104248844

[29] N. Datta, "Min- and max-relative entropies and a new entanglement monotone," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2816–2826, June 2009.

[30] M. Tomamichel and M. Hayashi, "A hierarchy of information quantities for finite block length analysis of quantum tasks," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7693–7710, Nov 2013.

[31] A. Anshu, R. Jain, and N. A. Warsi, "Building blocks for communication over noisy quantum networks," *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 1287–1306, Feb 2019.

[32] K. Li, "Second-order asymptotics for quantum hypothesis testing," *Ann. Statist.*, vol. 42, no. 1, pp. 171–189, 02 2014. [Online]. Available: https://doi.org/10.1214/13-AOS1185

[33] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1753–1768, 2003.

[34] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2481–2485, Nov 1999.

[35] M. Hayashi, "Discrimination of Two Channels by Adaptive Methods and Its Application to Quantum System," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3807–3820, Aug 2009.

[36] A. W. Harrow, A. Hassidim, D. W. Leung, and J. Watrous, "Adaptive versus nonadaptive strategies for quantum channel discrimination," *Physical Review A*, vol. 81, p. 032339, Mar 2010. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.81.032339

[37] M. M. Wilde, M. Berta, C. Hirche, and E. Kaur, "Amortized channel divergence for asymptotic quantum channel discrimination," *Letters in Mathematical Physics*, vol. 110, no. 8, pp. 2277–2336, 2020. [Online]. Available: https://doi.org/10.1007/s11005-020-01297-7

[38] R. Ahlswede and I. Csiszar, "Common randomness in information theory and cryptography. II. CR capacity," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 225–240, Jan 1998.

[39] U. M. Maurer, "Secret Key Agreement by Public Discussion from Common Information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, 1993.

[40] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. I. Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, July 1993.

[41] I. Devetak and A. Winter, "Distillation of secret key and entanglement from quantum states," *Proc. Roy. Soc. London Ser. A*, vol. 461, no. 2053, pp. 207–235, Jan 2005, arXiv:quant-ph/0306078.

[42] ——, "Distilling common randomness from bipartite quantum states," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3183–3196, Dec 2004.

[43] L. Henderson and V. Vedral, "Classical, quantum and total correlations," *J. Phys. A: Math. Gen.*, vol. 34, no. 35, pp. 6899–6905, 2001.

[44] G. Murta, F. Grasselli, H. Kampermann, and D. Bruss, "Quantum Conference Key Agreement: A Review," 2020, arXiv[quant-ph]:2003.10186.

[45] I. Csiszár and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3047–3061, Dec 2004.

[46] R. García-Patrón, W. Matthews, and A. Winter, "Quantum Enhancement of Randomness Distribution," *IEEE Trans. Inf. Theory*, vol. 64, no. 6, pp. 4664–4673, June 2018.

[47] M. Hillery, V. Bužek, and A. Berthiaume, "Quantum secret sharing," *Phys. Rev. A*, vol. 59, pp. 1829–1834, Mar 1999.

[48] J. A. Smolin, F. Verstraete, and A. Winter, "Entanglement of assistance and multipartite state distillation," *Phys. Rev. A*, vol. 72, p. 052317, Nov 2005.

[49] B. Fortescue and H.-K. Lo, "Random Bipartite Entanglement from W and W-Like States," *Phys. Rev. Lett.*, vol. 98, p. 260501, 2007.

[50] A. Streltsov, C. Meignant, and J. Eisert, "Rates of multi-partite entanglement transformations and applications in quantum networks," 2019, arXiv[quant-ph]:1709.09693v2.

[51] S. Bravyi, D. Fattal, and D. Gottesman, "GHZ extraction yield for multipartite stabilizer states," *J. Math. Phys.*, vol. 47, p. 062106, 2006.

[52] P. Vrana and M. Christandl, "Distillation of Greenberger-Horne-Zeilinger States by Combinatorial Methods," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5945–5958, Sep 2019.

[53] F. Salek and A. Winter, "Multi-user distillation of common randomness and entanglement from quantum states," 2020.

[54] ——, "Multi-User Distillation of Common Randomness and Entanglement from Quantum States," in *Proc. 2020 IEEE International Symposium on Information Theory (ISIT), 21-26 June 2020, Los Angeles, CA*. IEEE, 2020, pp. 1967–1972.

[55] F. Salek, A. Anshu, M. Hsieh, R. Jain, and J. R. Fonollosa, "One-shot capacity bounds on the simultaneous transmission of classical and quantum information," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2141–2164, 2020.

[56] F. Salek, A. Anshu, M. Hsieh, R. Jain, and J. R. Fonollosa, "One-shot capacity bounds on the simultaneous transmission of public and private information over quantum channels," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 296–300.

[57] F. Salek, M. Hsieh, and J. R. Fonollosa, "Single-serving quantum broadcast channel with common, individualized and confidential messages," *IEEE Transactions on Information Theory*, pp. 1–1, 2020.

[58] ——, "Publicness, privacy and confidentiality in the single-serving quantum broadcast channel," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1712–1716.

[59] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *The Annals of Mathematical Statistics*, vol. 23, no. 4, pp. 493–507, Dec 1952. [Online]. Available: https://doi.org/10.1214/aoms/1177729330

[60] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *The Annals of Mathematical Statistics*, vol. 36, no. 2, pp. 369–401, Apr 1965. [Online]. Available: https://doi.org/10.1214/aoms/1177700150

[61] T. Han and K. Kobayashi, "The strong converse theorem for hypothesis testing," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 178–180, Jan 1989.

[62] F. Hiai and D. Petz, "The proper formula for relative entropy and its asymptotics in quantum probability," *Communications in Mathematical Physics*, vol. 143, no. 1, pp. 99–114, 1991. [Online]. Available: https://doi.org/10.1007/BF02100287

[63] M. Nussbaum and A. Szkoła, "The Chernoff lower bound for symmetric quantum hypothesis testing," *The Annals of Statistics*, vol. 37, no. 2, pp. 1040–1057, Apr 2009. [Online]. Available: https://doi.org/10.1214/08-AOS593

[64] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, "Discriminating States:

The Quantum Chernoff Bound," *Physical Review Letters*, vol. 98, no. 16, Apr 2007. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.98.160501

[65] M. Hayashi, *Quantum Information Theory: Mathematical Foundation*, 2nd ed. Springer Publishing Company, Incorporated, 2016.

[66] T. Ogawa and M. Hayashi, "On error exponents in quantum hypothesis testing," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1368–1372, June 2004.

[67] M. Hayashi, "Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding," *Physical Review A*, vol. 76, no. 6, Dec 2007. [Online]. Available: http://dx.doi.org/10.1103/PhysRevA.76.062301

[68] H. Nagaoka, "The Converse Part of The Theorem for Quantum Hoeffding Bound," Nov 2006, arXiv:quant-ph/0611289.

[69] M. Berta, C. Hirche, E. Kaur, and M. M. Wilde, "Amortized Channel Divergence for Asymptotic Quantum Channel Discrimination," Aug 2018, arXiv[quant-ph]:1808.01498.

[70] R. Duan, Y. Feng, and M. Ying, "Perfect Distinguishability of Quantum Operations," *Physical Review Letters*, vol. 103, p. 210501, Nov 2009. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.103.210501

[71] R. Duan, C. Guo, C. Li, and Y. Li, "Parallel distinguishability of quantum operations," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 2259–2263.

[72] D. Puzzuoli and J. Watrous, "Ancilla Dimension in Quantum Channel Discrimination," *Annales Henri Poincaré*, vol. 18, no. 4, pp. 1153–1184, 2017.

[73] X. Wang and M. M. Wilde, "Resource theory of asymmetric distinguishability for quantum channels," *Physical Review Research*, vol. 1, p. 033169, Dec 2019. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevResearch.1.033169

[74] K. Fang, O. Fawzi, R. Renner, and D. Sutter, "A chain rule for the quantum relative entropy," Sep. 2019, arXiv[quant-ph]:1909.05826.

[75] T. Cooney, M. Mosonyi, and M. M. Wilde, "Strong Converse Exponents for a Quantum Channel Discrimination Problem and Quantum-Feedback-Assisted Communication," *Communications in Mathematical Physics*, vol. 344, no. 3, pp. 797–829, June 2016. [Online]. Available: https://doi.org/10.1007/s00220-016-2645-4

[76] C. Hirche, M. Hayashi, E. Bagan, and J. Calsamiglia, "Discrimination power of a quantum detector," *Physical Review Letters*, vol. 118, no. 16, p. 160502, 2017.

[77] N. Yu and L. Zhou, "Chernoff Bound for Quantum Operations is Faithful," May 2017, arXiv[quant-ph]:1705.01642.

[78] C. A. Fuchs and J. van de Graaf, "Cryptographic distinguishability measures for quantum-mechanical states," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1216–1227, 1999.

[79] M. A. Naimark (Neumark), "On a representation of additive operator set functions," *Doklady Akademii Nauk SSSR - Comptes Rendus de l'Acad 'e mie des Sciences de l'URSS (NS)*, vol. 41, no. 9, pp. 359–361, 1943.

[80] T. Ogawa and M. Hayashi, "On Error Exponents in Quantum Hypothesis Testing," Jun. 2002, arXiv:quant-ph/0206151.

[81] C. P. Niculescu and L. Persson, *Convex Functions and Their Applications*, 2nd ed. Springer-Verlag, 2018.

[82] C. King and M.-B. Ruskai, "Minimal entropy of states emerging from noisy quantum channels," *IEEE Transactions on Information Theory*, vol. 47, no. 1, pp. 192–209, 2001.

[83] A. Fujiwara and P. Algoet, "One-to-one parametrization of quantum channels," *Physical Review A*, vol. 59, no. 5, pp. 3290–3294, May 1999. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.59.3290

[84] I. Devetak and A. Winter, "Relating Quantum Privacy and Quantum Coherence: An Operational Approach," *Phys. Rev. Lett.*, vol. 93, p. 080501, Aug 2004.

[85] I. Devetak, A. W. Harrow, and A. Winter, "A Family of Quantum Protocols," *Phys. Rev. Lett.*, vol. 93, p. 230504, Dec 2004.

[86] A. W. Harrow, "Coherent Communication of Classical Messages," *Phys. Rev. Lett.*, vol. 92, p. 097902, Mar 2004.

[87] T. M. Cover and J. A. Thomas, *Elements of Information Theory (2nd ed.)*. Wiley Interscience, 2006.

[88] A. Winter, "Coding Theorems of Quantum Information Theory," Ph.D. dissertation, Universität Bielefeld, Department of Mathematics, July 1999, arXiv:quant-ph/9907077.

[89] I. Devetak and A. Winter, "Classical data compression with quantum side information," *Phys. Rev. A*, vol. 68, p. 042301, Oct 2003.

[90] P. Sen, "Simultaneous decoding, unions, intersections and a one-shot quantum joint typicality lemma," 2018, arXiv[quant-ph]:1806.07278v2.

[91] W. F. Stinespring, "Positive Functions on C^*-Algebras," *Proc. Amer. Math. Soc.*, vol. 6, no. 2, pp. 211–216, 1955.

[92] M.-H. Hsieh, I. Devetak, and A. Winter, "Entanglement-Assisted Capacity of Quantum Multiple-Access Channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3078–3090, July 2008.

[93] K. Horodecki, M. Horodecki, P. Horodecki, and J. Oppenheim, "Locking Entanglement with a Single Qubit," *Phys. Rev. Lett.*, vol. 94, p. 200501, 2005.

[94] H. Maassen and J. B. M. Uffink, "Generalized Entropic Uncertainty Relations," *Phys. Rev. Lett.*, vol. 60, no. 12, pp. 1103–1106, 1988.

[95] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized Privacy Amplification," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1915–1923, 1995.

[96] H. Tyagi and P. Narayan, "How Many Queries Will Resolve Common Randomness?" *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5363–5378, 2013.

[97] A. A. Gohari and V. Anantharam, "Information-Theoretic Key Agreement of Multiple Terminals: Part I," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3973–3996, 2010.

[98] P. Sen, "Inner bounds via simultaneous decoding in quantum network information theory," 2018, arXiv[quant-ph]:1806.07276.

[99] R. Ahlswede, "The Capacity Region of a Channel with Two Senders and Two Receivers," *Ann Probab.*, vol. 2, no. 5, pp. 805–814, 1974.

[100] M. M. Wilde, "Position-based coding and convex splitting for private communication over quantum channels," *Quantum Information Processing*, vol. 16, no. 10, p. 264, Sep 2017.

[101] ——, *Quantum Information Theory*, 1st ed. New York, NY, USA: Cambridge University Press, 2013.

[102] A. S. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 269–273, Jan 1998.

[103] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A*, vol. 56, pp. 131–138, Jul 1997. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.56.131

[104] A. Holevo, *Statistical problems in quantum physics*, 11 2006, vol. 1, pp. 104–119.

[105] S. Lloyd, "Capacity of the noisy quantum channel," *Phys. Rev. A*, vol. 55, pp. 1613–1622, Mar 1997. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.55.1613

[106] P. Shor, "The quantum channel capacity and coherent information„" in *MSRI Seminar, unpublished.*

[107] M.-H. Hsieh and M. Wilde, "Public and private communication with a quantum channel and a secret key," *Phys. Rev. A*, vol. 80, p. 022306, Aug 2009. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.80.022306

[108] M. M. Wilde and M.-H. Hsieh, "The quantum dynamic capacity formula of a quantum channel," *Quantum Information Processing*, vol. 11, no. 6, pp. 1431–1463, 2012. [Online]. Available: http://dx.doi.org/10.1007/s11128-011-0310-6

[109] T. S. Han and S. Verdu, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, 1993.

[110] N. Datta and T. C. Dorlas, "The coding theorem for a class of quantum channels with long-term memory," *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 28, pp. 8147–8164, jun 2007.

[111] T. Dorlas and C. Morgan, "Invalidity of a strong capacity for a quantum channel with memory," *Phys. Rev. A*, vol. 84, p. 042318, Oct 2011.

[112] R. Renner, S. Wolf, and J. Wullschleger, "The single-serving channel capacity," in *2006 IEEE International Symposium on Information Theory*, July 2006, pp. 1424–1427.

[113] M. Mosonyi and N. Datta, "Generalized relative entropies and the capacity of classical-quantum channels," *Journal of Mathematical Physics*, vol. 50, no. 7, p. 072104, 2009. [Online]. Available: https://doi.org/10.1063/1.3167288

[114] J. M. Renes and R. Renner, "Noisy channel coding via privacy amplification and information reconciliation," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7377–7385, Nov 2011.

[115] A. Anshu, V. K. Devabathini, and R. Jain, "Quantum communication using coherent rejection sampling," *Phys. Rev. Lett.*, vol. 119, p. 120506, Sep 2017. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.119.120506

[116] J. Radhakrishnan, P. Sen, and N. A. Warsi, "One-Shot Private Classical Capacity of Quantum Wiretap Channel: Based on one-shot quantum covering lemma," *ArXiv e-prints*, p. arXiv:1703.01932, Mar. 2017.

[117] R. Ahlswede and A. Winter, "Strong converse for identification via quantum channels," *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp. 569–579, March 2002.

[118] M. Tomamichel, R. Colbeck, and R. Renner, "A fully quantum asymptotic equipartition property," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5840–5847, 2009.

[119] M. Berta, M. Christandl, and R. Renner, "The quantum reverse shannon theorem based on one-shot information theory," *Communications in Mathematical Physics*, vol. 306, no. 3, p. 579, Aug 2011. [Online]. Available: https://doi.org/10.1007/s00220-011-1309-7

[120] N. CiganoviÄĞ, N. J. Beaudry, and R. Renner, "Smooth max-information as one-shot generalization for mutual information," *IEEE*

Transactions on Information Theory, vol. 60, no. 3, pp. 1573–1581, March 2014.

[121] M. E. Shirokov, "Tight continuity bounds for the quantum conditional mutual information, for the Holevo quantity and for capacities of quantum channels," *arXiv e-prints*, p. arXiv:1512.09047, Dec. 2015.

[122] A. Winter, "Tight uniform continuity bounds for quantum entropies: Conditional entropy, relative entropy distance and energy constraints," *Communications in Mathematical Physics*, vol. 347, no. 1, pp. 291–313, 2016. [Online]. Available: https://doi.org/10.1007/s00220-016-2609-8

[123] K. Horodecki, M. Horodecki, P. Horodecki, and J. Oppenheim, "General paradigm for distilling classical key from quantum states," *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1898–1929, 2009.

[124] M. M. Wilde, M. Tomamichel, and M. Berta, "Converse bounds for private communication over quantum channels," *IEEE Transactions on Information Theory*, vol. 63, no. 3, pp. 1792–1817, 2017.

[125] H. Qi, Q. Wang, and M. M. Wilde, "Applications of position-based coding to classical communication over quantum channels," *Journal of Physics A: Mathematical and Theoretical*, vol. 51, no. 44, p. 444002, 2018.

[126] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[127] M. R. Bloch and J. Kliewer, "On secure communication with constrained randomization," in *2012 IEEE International Symposium on Information Theory Proceedings*, July 2012, pp. 1172–1176.

[128] S. Watanabe and Y. Oohama, "The optimal use of rate-limited randomness in broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 61, no. 2, pp. 983–995, Feb 2015.

[129] Y. K. Chia and A. E. Gamal, "Three-receiver broadcast channels with common and confidential messages," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 2748–2765, May 2012.

[130] M. Hsieh and M. M. Wilde, "Trading classical communication, quantum communication, and entanglement in quantum shannon theory,"

IEEE Transactions on Information Theory, vol. 56, no. 9, pp. 4705–4730, 2010.

[131] ——, "Entanglement-assisted communication of classical and quantum information," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4682–4704, 2010.

[132] E. Y. Zhu, Q. Zhuang, M.-H. Hsieh, and P. W. Shor, "Superadditivity in trade-off capacities of quantum channels," *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3973–3989, 2019.

[133] E. Chitambar, M.-H. Hsieh, and A. Winter, "The private and public correlation cost of three random variables with collaboration," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 2034–2043, April 2016.

[134] M. M. Wilde and M.-H. Hsieh, "Public and private resource trade-offs for a quantum channel," *Quantum Information Processing*, vol. 11, no. 6, pp. 1465–1501, 2012. [Online]. Available: http://dx.doi.org/10.1007/s11128-011-0317-z

[135] M.-H. Hsieh, T. Brun, and I. Devetak, "Entanglement-assisted quantum quasicyclic low-density parity-check codes," *Phys. Rev. A*, vol. 79, p. 032340, Mar 2009. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.79.032340

[136] M.-H. Hsieh, Z. Luo, and T. Brun, "Secret-key-assisted private classical communication capacity over quantum channels," *Phys. Rev. A*, vol. 78, p. 042306, Oct 2008. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.78.042306

[137] M.-H. Hsieh, I. Devetak, and A. Winter, "Entanglement-assisted capacity of quantum multiple-access channels," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3078–3090, July 2008.

[138] M. Tomamichel, *Quantum information processing with finite resources : mathematical foundations*, ser. SpringerBriefs in mathematical physics ; v. 5, 2016.

[139] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1753–1768, July 2003.

[140] M. Hayashi, "Role of hypothesis testing in quantum information," 2017.

[141] N. Datta and M.-H. Hsieh, "One-shot entanglement-assisted quantum and classical communication," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1929–1939, March 2013.

[142] N. Datta, M. Mosonyi, M.-H. Hsieh, and F. G. Brandao, "A smooth entropy approach to quantum hypothesis testing and the classical capacity of quantum channels," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8014–8026, Dec 2013.

[143] N. Datta and M.-H. Hsieh, "The apex of the family tree of protocols: optimal rates and resource inequalities," *New Journal of Physics*, vol. 13, no. 9, p. 093042, 2011. [Online]. Available: http://stacks.iop.org/1367-2630/13/i=9/a=093042

[144] N. Datta, M.-H. Hsieh, and J. Oppenheim, "An upper bound on the second order asymptotic expansion for the quantum communication cost of state redistribution," *Journal of Mathematical Physics*, vol. 57, no. 5, 2016. [Online]. Available: http://scitation.aip.org/content/aip/journal/jmp/57/5/10.1063/1.4949571

[145] I. Savov and M. M. Wilde, "Classical codes for quantum broadcast channels," *IEEE Transactions on Information Theory*, vol. 61, no. 12, pp. 7017–7028, Dec 2015.

[146] J. Radhakrishnan, P. Sen, and N. Warsi, "One-shot marton inner bound for classical-quantum broadcast channel," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2836–2848, May 2016.

[147] C. Hirche and C. Morgan, "An improved rate region for the classical-quantum broadcast channel," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 2782–2786.

[148] A. Anshu, M. Hayashi, and N. A. Warsi, "Secure communication over fully quantum Gel'fand-Pinsker wiretap channel," *arXiv e-prints*, p. arXiv:1801.00940, Jan. 2018.

[149] P. Sen, "Inner bounds via simultaneous decoding in quantum network information theory," *arXiv e-prints*, p. arXiv:1806.07276, Jun 2018.

[150] J. Yard, P. Hayden, and I. Devetak, "Quantum broadcast channels," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 7147–7162, Oct 2011.

[151] F. Dupuis, L. Kraemer, P. Faist, J. M. Renes, and R. Renner, "Generalized Entropies," *arXiv e-prints*, p. arXiv:1211.3141, Nov 2012.

[152] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2481–2485, Nov 1999.

[153] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. New York, NY, USA: Cambridge University Press, 2012.

[154] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner, "Leftover hashing against quantum side information," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5524–5535, Aug 2011.