Technische Universität München
Fakultät für Mathematik

Simultaneous Identification Capacity of the Classical-Quantum Multiple Access Channel

Masterarbeit von Stephen Diadamo

Aufgabensteller: Prof. Dr. H. Boche
Betreuer: Prof. Dr. R. König

Abgabedatum: September 17, 2018

Ich erkläre hiermit, dass ich diese Masterarbeit selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe.

(Garching bei München, 17.09.2018, Stephen Diadamo)

Abstract

The problem of transmitting information securely and effectively over various communication channels is a highly important topic in information theory. A different type of communication problem was introduced by Rudolf Ahlswede and Gunter Dueck in 1989 in which they labeled "identification". In contrast to transmission problems, the receiver of information is not interested in what exactly the information contains, rather they are interested in whether it contains specific information or not. It was shown by Ahlswede et al that the size of randomized identification codes for discrete memoryless classical channels grow doubly exponentially in blocklength, where transmission codes sizes grow singly exponentially. For the discrete memoryless classical-quantum channel, Peter Löber introduced the idea of a simultaneous identification code and showed that the simultaneous identification capacity, a double logarithmic scale, for the this channel is equal to its transmission capacity, a single logarithmic scale. For the classical memoryless multiple access channel, Yosef Steinberg showed that the identification capacity region is equal to the transmission capacity region. We extend these ideas in this thesis and prove that simultaneous identification capacity region for the discrete memoryless classical-quantum multiple access channel is equal to its transmission capacity region.

Acknowledgments

In completing this thesis, I give thanks to those who have helped and supported me throughout the process. I firstly give thanks to my supervisors professors Holger Boche and Robert König for their time and their support with which I have been able to study a truly fascinating topic. Moreover, I give thanks to Gisbert Janßen who has helped me most in the completion of this thesis, whose guidance and suggestions I have benefited from greatly. I also thank Alihan Kaplan and Sajad Saeedinaeeni especially for the time they spent with me in discussion. Finally, I give thanks to my parents, who have always been supportive of me in my endeavors, for which I am truly grateful.

Contents

Chapter 1

Introduction

1.1 Transmission

In 1948, Claude Shannon contributed ground breaking theory regarding the transmission of information over a communication channel. He was the first to formalize what is called a "transmission problem", a single sender sending messages over a noisy channel to be recovered by a single receiver. In his theory, he modeled a noisy channel W via a stochastic matrix,

$$W := \{W(y|x) \mid x \in \mathcal{X}, y \in \mathcal{Y}\}, \tag{1.1.1}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are finite alphabets. At a high level, what this represents is that if there are two parties, call them Alice and Bob, and Alice sends the letter $x \in \mathcal{X}$ to Bob, the probability that Bob then receives $y \in \mathcal{Y}$ is $W(y|x)$. For a perfect channel, Bob would always be able to recover x, but when noise is considered, the likelihood that x is recovered perfectly is not certain.

For longer sequences of characters, consecutive uses of the channel will be needed. There could be correlations between the uses of a channel, but a channel that is unchanged between its uses is called a discrete memoryless channel. With a noisy channel W, a discrete memoryless channel can be generated as follows: for k-length sequences, with $k \in \mathbb{N}$, of characters, $x^k := (x_1, x_2, ..., x_k) \in \mathcal{X}^k$ and $y^k := (y_1, y_2, ..., y_k) \in \mathcal{Y}^k$ being the sequence of characters received, the transmission probability is

$$W^k(y^k|x^k) := \prod_{i=1}^{k} W(y_i|x_i). \tag{1.1.2}$$

The discrete memoryless property is explicitly clear in this model, seen by using a product of single uses of W. The theory we aim to introduce in this chapter relies on the discrete memoryless channel. There are, naturally, models for general channels that are not memoryless; we will explore this distinction further in later chapters.

Another important concept that Shannon made use of in his theory is channel coding. The idea of channel coding is that there is a set of M messages translated deterministically into a set of codewords that are called a codebook, indexed by $[M] := \{1, ..., M\}$. Alice then chooses, with uniform probability, which message to send to Bob. Uniformity here implies it is not particularly important which message Alice chooses to send; the goal is just to be able to transmit it reliably. At Bob's side, he would like to recover the codeword sent by Alice and the underlying message despite the noise of the channel. Shannon's idea was to incorporate redundancy into codewords and have decoding sets such that the probability that Bob can recover Alice's message tends to certainty.

When Bob receives a message from Alice, because they are generally communicating over a noisy channel, it is possible that the message becomes disturbed along the way. The goal for Bob then is to try to recover the message as best as possible. One way to do this is to add redundancy based on the channel in a decoding codebook such that multiple codewords represent the same message. We clarify this idea with an example.

Example 1.1.1. *Let $\mathcal{X} = \mathcal{Y} \coloneqq \{0,1\}$. With $p \in (0,1)$, assume that channel W Alice and Bob are communicating with is defined as,*

$$W(0|0) = p, \;\; W(1|1) = p, \;\; W(0|1) = 1-p, \;\; W(1|0) = 1-p. \tag{1.1.3}$$

Assume Alice intends to send the two messages {"hi", "bye"} to Bob. Let Alice's codebook be

$$(0 : \text{"}hi\text{"}, 1 : \text{"}bye\text{"}), \tag{1.1.4}$$

and Bob's decoding the same set,

$$(0 : \text{"}hi\text{"}, 1 : \text{"}bye\text{"}). \tag{1.1.5}$$

Then with one use of W, the probability that Bob receives the message Alice intends to send is p and the probability of error is $1-p$.

Now modify Alice's codebook to be,

$$(000 : \text{"}hi\text{"}, 111 : \text{"}bye\text{"}), \tag{1.1.6}$$

and Bob's decodings to

$$(\{000, 001, 010, 100\} : \text{"}hi\text{"}, \{111, 110, 101, 011\} : \text{"}bye\text{"}), \tag{1.1.7}$$

where two bits of redundancy have been added. The probability that Bob receives "hi" when Alice sends it is then,

$$\begin{aligned} W^3(\{000, 001, 010, 100\}|000) &= W(000|000) + W(001|000) + W(010|000) + W(100|000) && (1.1.8)\\ &= W(0|0)^3 + 3(W(0|0) \cdot W(0|0) \cdot W(1|0)) && (1.1.9)\\ &= p^3 + 3p^2(1-p), && (1.1.10) \end{aligned}$$

which is the same if Alice sends "bye". When $p \geq 0.5$, $p^3 + 3p^2(1-p) \geq p$, and so the probability of a successful transmission for a sufficiently reliable channel is larger with two bits of redundancy added than with none. With even more redundant bits, the success probability can become much higher.

There are of course other more efficient coding schemes, but with this rough intuition of how a decoding set could work, the definition for codes that we use in various forms throughout this thesis will be sufficiently understandable.

Definition 1.1.2 ($(k, M) - code$)**.** *With a channel W, for $k, M \in \mathbb{N}$ a (k, M)-code $\mathcal{C}$ is a family $(x_i, D_i)_{i=1}^M$ such that for all $i \in [M]$, $x_i \in \mathcal{X}^k$, $D_i \subset \mathcal{Y}^k$, and for all $j \in [M], i \neq j$, $D_i \cap D_j = \emptyset$. For a figure of merit for the quality of $\mathcal{C}$, an average transmission error is calculated via*

$$\overline{e}(\mathcal{C}, W^k) \coloneqq 1 - \frac{1}{M} \sum_{i=1}^{M} W^k(D_i|x_i). \tag{1.1.11}$$

For error threshold $\lambda \geq 0$, we define the maximum code size for k uses of W with error bounded by λ as,

$$M(k, \lambda) \coloneqq \max_M \{M \in \mathbb{N} \mid \exists \; a \; (k, M) - code \; \mathcal{C} \; s.t. \; \overline{e}(\mathcal{C}, W^k) \leq \lambda\}. \tag{1.1.12}$$

Remark 1.1.3. *In the definition above, we use an average transmission error for a figure of merit. We could have also used a maximal error rate which we define as follows, and will make use of later,*

$$e(\mathcal{C}, W^k) \coloneqq \max_{i \in [M]} 1 - W^k(D_i|x_i). \tag{1.1.13}$$

It can be shown that a code with a bounded average error, with a negligible loss to the number of codewords, also has a maximal error bounded on the order of the average error bound. That is, for a (k, M)-code $\mathcal{C}$, $\overline{e}(\mathcal{C}, W^k) \leq \epsilon$ implies $e(\mathcal{C}, W^k) \leq O(\epsilon)$.

The aim of a channel capacity problem is to determine, given a channel, how fast the maximum code size $M(k, \lambda)$ can grow per channel use with many uses of the channel, that is, in the limit $k \to \infty$. The rate of growth per channel use for $M(k, \lambda)$ is defined,

$$R := \frac{1}{k} \log M(k, \lambda) \tag{1.1.14}$$

where $\log M(k, \lambda)$ is used because it represents the number of bits needed to represent all of the messages in a codebook. In this thesis, log will always be assumed to be base 2.

With C representing the maximal capacity of a channel, Shannon gave the following theorem, which was proved partially by Shannon himself and partially by Jacob Wolfowitz.

Theorem 1.1.4 (Shannon et al)**.** *For a discrete memoryless channel W, for all $\lambda \in (0, 1)$,*

$$\lim_{k \to \infty} \frac{1}{k} \log M(k, \lambda) = C \tag{1.1.15}$$

where $C = \max_A I(A; B)$.

What this theorem tells us is that regardless of error probability, other than the extreme cases, with unlimited uses of the channel, the maximal rate for a codebook is exactly equal to the channel capacity. Further, the channel capacity is exactly the maximum possible mutual information between the sender Alice represented with random variable A and receiver Bob, represented by random variable B. In this thesis we do not go into depth regarding information quantities such as mutual information, but the necessary definitions are provided in Appendix B. For an adequate understanding of the theorem, the mutual information here can be thought of as the amount of information contained in B about A. Fixing the receiver B and maximizing over all A gives the largest amount of shared information between Alice and Bob, and thus the capacity of the channel has an intuitive operational meaning. Due to the symmetric property of mutual information, maximizing over A is equivalent to maximizing over B, and so either can be performed.

1.2 Identification

To introduce identification, we contrast Shannon's transmission theory with Ahlswede and Dueck's theory of identification. A message transmission problem as we have seen an example, is when one or more parties are attempting to transmit messages over a noisy communication channel to other receiving parties such that the receivers can recover the message contents. An identification problem is similar in the sense that there are still one or more senders using a noisy communication channel to send messages, but what changes is the objective of the receiver. The receiver is no longer interested in what exactly the message is, only if the message they received is a specific message they are interested in. This is what is referred to as an identification problem. We give a simple example of when a problem can be considered an identification problem.

Example 1.2.1. *Assume there is a radio broadcast transmitting weather forecasts using codewords to represent different types of weather patterns. For example, 1 for rain, 2 for sun, and so on. Assume Bob is a farmer and cares only for when he can expect rain and all information about other weather systems is useless to him. If he is listening to the radio broadcast, since he only cares to receive information about rain, instead of decoding every message received from the broadcast, he can instead identify when a message signaling rain is received using potentially less resources.*

The above example is indeed different from a transmission problem. Because Bob is only interested in a single message, intuitively he should not have to use as many resources to determine if the message he received is that message. With channel uses as the resource, Ahlswede and Dueck found, surprisingly, that Bob would need exponentially fewer uses of the channel than if he was to decode every received message.

To model an identification problem, a similar approach is taken as for transmission problems. Identification also uses a coding structure, but a key difference is that the codes use a randomized encoder for codeword generation. In the introduction to transmission codes, Alice selects a message to send to Bob with uniform probability and it is mapped deterministically to the same codeword every time she sends that particular message, and therefore each codeword has the same likelihood of appearing. With identification codes, as will be seen in the definition, the messages are not deterministically mapped to a particular codeword, and therefore each codeword has an arbitrary likelihood of being sent. For transmission codes, using a randomized encoding has no consequence regarding Shannon's capacity theorem. Ahlswede and Dueck, however, found that the result from using deterministic codes for identification codes was rather unsatisfactory [1], and hence used the following coding scheme.

Definition 1.2.2 ((k, M)-ID-code). *With a channel W, for $k, M \in \mathbb{N}$, a randomized $(k, M)-$ID-code $\mathcal{C}_{id}$ is a family $(P_i, D_i)_{i=1}^{M}$ where for all $i \in [M]$, $P_i \in \mathcal{P}(\mathcal{X}^k)$ and $D_i \subset \mathcal{Y}^k$. For a figure of merit, there are two error types.*

$$e_1(\mathcal{C}_{id}, W^k) := \max_{i\in[M]} 1 - \sum_{x^k\in\mathcal{X}^k} P_i(x^k)W^k(D_i|x^k) \tag{1.2.1}$$

$$e_2(\mathcal{C}_{id}, W^k) := \max_{\substack{i,j\in[M]\\ i\neq j}} \sum_{x^k\in\mathcal{X}^k} P_i(x^k)W^k(D_j|x^k). \tag{1.2.2}$$

We define,

$$\begin{aligned} M_{id}(k, \lambda) := \max_{M}\{M \in \mathbb{N} \mid \exists \text{ a } (k, M)\text{-ID-code } \mathcal{C}_{id} \text{ s.t. } \lambda_1, \lambda_2 \leq \lambda, \\ e_1(\mathcal{C}_{id}, W^k) \leq \lambda_1 \text{ and } e_2(\mathcal{C}_{id}, W^k) \leq \lambda_2\}. \end{aligned} \tag{1.2.3}$$

Note that in this definition there is no condition that the D_is of the code must be mutually disjoint. The first type of error can be thought of as the probability of a message being identified erroneously. Here, the effect of randomized encoding is not seen because a small first kind error implies a high probability of the encoding of x^k existing in the set D_i, as with transmission. The figure of merit to control the effect of randomization is therefore seen in the second kind of error. Because in identification the condition of mutually disjoint D_is is not enforced, the D_is can have a lot of overlap and the randomized encoders P_i a lot of support, resulting in a potentially bad identification code, and so the second type of error provides a figure of merit to analyze this.

As with transmission, to find the identification capacity of a channel, one is interested in how fast the maximum ID-code size $M_{\text{id}}(k, \lambda)$ grows with many uses of the channel. The rate of growth per channel use for $M_{\text{id}}(k, \lambda)$ is given by,

$$R := \frac{1}{k} \log \log M_{\text{id}}(k, \lambda). \tag{1.2.4}$$

The double logarithm here is not particularly intuitive but can be motivated by a consequence of the following theorem.

Theorem 1.2.3 (Alhswede, Dueck). *[2, Theorem 2] For a discrete memoryless channel W, for all $\lambda \in (0, 1/2)$,*

$$\lim_{k\to\infty} \frac{1}{k} \log \log M_{id}(k, \lambda) = C \tag{1.2.5}$$

with $C = \max_A I(A; B)$, the transmission capacity of the channel.

For the remainder of this thesis, the goal is to extend the theory of identification to the quantum multiple access channel. In Chapter 2, we review the proof for achievability of Theorem 1.2.3 by reviewing the Transformator lemma. The structure of the proof will be useful when proving achievability for the quantum case, and so we give the complete proof in the classical case for a better overall understanding when the more complicated quantum scenario arises. In Chapter 3, we review the needed quantum theory and review the transmission capacity theorem for the quantum multiple access channel from Andreas Winter. During this review, we introduce useful lemmas which will be needed in proving equivalence of capacity regions in Chapter 4. Lastly, in Chapter 4, we prove the capacity region for the discrete memoryless classical-quantum multiple access channel by introducing two multi-letter regions and showing their equivalence to both the identification and transmission capacity region.

Chapter 2

The Transformator Lemma

The focus of this chapter is to review a lemma used by Ahlswede and Dueck to prove the direct part of their theorem, that is, showing that there exist ID-codes that can achieve achievable transmission rates. The idea of the lemma, called the "Transformator" lemma, is to join two transmission codes, one with large blocklength and one with, in the limit, negligible blocklength to show that with random generations of identification codes, there exist ID-codes that achieve the desired rate. An important thing to note is that the two transmission codes will use different figures of merit. One will use an average error criterion and the other a maximum. In the single sender-single receiver case, both types of codes can achieve the same rates and making the distinction is therefore not beneficial. In latter chapters, when dealing with the double sender-single receiver multiple access channel, this distinction is important since the two types of codes do not achieve the same rates, as was shown with a counter example given by Dueck in [12].

2.1 Transformator Lemma for Classical Channels

In the previous chapter, we introduced the concept of the rate at which codebooks increase in size per channel use with the number of uses of the channel, or blocklength, and reviewed Shannon's theorem for the rate in which the maximal codes achieves. An alternative interpretation of a rate is how much information is being transmitted per channel use and so although optimally one would always like to transmit information at the channel capacity, it is also practical to use a rate less than capacity. We therefore formalize what it means for a code to achieve a particular rate.

Definition 2.1.1 (Achievable rate)**.** *With channel W, for $R \in \mathbb{R}, R \geq 0$, we say R is an achievable rate if for all $\epsilon, \delta > 0$, there exists a k_0 such that for all $k \geq k_0$, there is a (k, M)-code $\mathcal{C}$ such that,*

$$\frac{1}{k} \log M \geq R - \delta, \text{ and } \quad \overline{e}(\mathcal{C}, W^k) \leq \epsilon. \tag{2.1.1}$$

*We say a (k, M)-code $\mathcal{C}$ achieves rate R if (*2.1.1*) holds for all $\delta, \epsilon > 0$.*

As we will see in many instances in this thesis, there are often analogous definitions for identification, and hence we have the following.

Definition 2.1.2 (Achievable ID-rate)**.** *With channel W, for $R \in \mathbb{R}$, $R \geq 0$, we say R is an achievable ID-rate if for all $\epsilon_1, \epsilon_2, \delta > 0$, there exists a k_0 such that for all $k \geq k_0$, there is a (k, M)-ID-code $\mathcal{C}_{id}$ with*

$$\frac{1}{k} \log\log M \geq R - \delta, \quad e_1(\mathcal{C}_{id}, W^k) \leq \epsilon_1, \text{ and } e_2(\mathcal{C}_{id}, W^k) \leq \epsilon_2. \tag{2.1.2}$$

*We say a (k, M)-ID-code $\mathcal{C}_{id}$ achieves ID-rate R if (*2.1.2*) holds for all $\delta, \epsilon_1, \epsilon_2 > 0$.*

The goal of the Transformator lemma is then to show that any achievable rate is also an achievable ID-rate. In the proof, we will make use of the following lemma which we state without proof, as it is a well known result.

Lemma 2.1.3 (Chernov-Hoeffding Bound)**.** *Let $M \in \mathbb{N}$, and let a sequence of random variables $(\psi_i)_{i\in[M]}$ be such that for each i, $\psi_i \in \{0,1\}$. Assume for each i and $\mu, \lambda \in (0,1)$ the expectation value $\mathbb{E}(\psi_i) \le \mu < \lambda$. Then it holds,*

$$\Pr\left(\sum_{j=1}^{M} \psi_j > \lambda M\right) \le 2^{-M\cdot D(\lambda\|\mu)},$$

where $D(\lambda \parallel \mu)$ is the relative entropy between distributions $(\lambda, 1-\lambda)$ and $(\mu, 1-\mu)$. We refer the reader to Appendix B for the definition of relative entropy.

Lemma 2.1.4 (Transformator lemma)**.** *[2, Lecture 2, Section 3] For a discrete memoryless channel W and an achievable rate R, there is an ID-code that achieves the ID-rate R.*

Proof. Let W be a discrete memoryless channel. For an achievable rate R, there exists by definition, for any $\eta > 0$, a k_0 such that for all $k \ge k_0$, there is a (k, M')-code that achieves R. Let ϵ also be an achievable rate. Then again by definition (and Remark 1.1.3) there exists an k'_0 dependent on η such that for any $k' \ge k'_0$ there is a (k', M'')-code that achieves ϵ. Let k be such that $k \ge k_0$ and $k'_0 \le \lceil\sqrt{k}\rceil$, then there exists two codes, a (k, M')-code $\mathcal{C}' := (u'_j, D'_j)_{j=1}^{M'}$ such that $M' = \lceil 2^{k(R-\eta)}\rceil$ and a $(\lceil\sqrt{k}\rceil, M'')$-code $\mathcal{C}'' := (u''_l, D''_l)_{l=1}^{M''}$ with $M'' = \lceil 2^{\sqrt{k}\epsilon}\rceil$, where for notational simplicity, we let $\epsilon = \epsilon - \eta$. Further, there exist sequence that with $k \to \infty$, $\lambda(k) \to 0$ and $\lambda(\sqrt{k}) \to 0$, $\overline{e}(\mathcal{C}', W^k) \le \lambda(k)$ and $e(\mathcal{C}'', W^{\lceil\sqrt{k}\rceil}) \le \lambda(\sqrt{k})$. Now, define the family of maps $\mathcal{T}$ as,

$$\mathcal{A} := (T_i : [M'] \to [M''])_{i=1}^{M}. \tag{2.1.3}$$

We aim to show that M is doubly exponential in k. With $m := k + \lceil\sqrt{k}\rceil$ we construct an (m, M)-ID-code $\mathcal{C}_{\text{id}} := (P_i, D_i)_{i=1}^{M}$, where each P_i has the structure,

$$P_i(x^m) := \begin{cases} \frac{1}{M'} & \text{if } \exists j \in [M'] : x^m = u'_j \cdot u''_{A_{i(j)}} \\ 0 & \text{otherwise,} \end{cases}$$

and each D_i,

$$D_i := \bigcup_{j=1}^{M'} D'_j \times D''_{A_i(j)}. \tag{2.1.4}$$

We show that for a random construction of $\mathcal{T}$ there exists an (m, M)-ID-code with the P_is and D_is structured as above with errors decreasing sufficiently quickly.

Define for $j \in [M']$, independent random variables U_j such that $\Pr\big(U_j = u'_j \cdot u''_l\big) = \frac{1}{M''}$ for $l \in [M'']$. For $i \in \mathbb{N}$, we denote

$$\overline{\mathcal{U}}_i := \{U_1, ..., U_{M'}\} \tag{2.1.5}$$

as a collection of random variables U_j and we define a random distribution $\overline{P}_i$ as the uniform distribution on $\overline{\mathcal{U}}_i$. The random decoding set is then defined as

$$\mathcal{D}(\overline{\mathcal{U}_i}) := \bigcup_{j=1}^{M'} D(U_j) \tag{2.1.6}$$

with $D(U_j) := D'_j \times D''_l$ when $U_j = u'_j \cdot u''_l$. We analyze the error performance of the random code $\mathcal{C}_{\text{id}} := (\overline{P}_i, \mathcal{D}(\overline{\mathcal{U}}_i))_{i=1}^{M}$. Fix a realization $\mathcal{U}_i$ of $\overline{\mathcal{U}}_i$ and define $r := \lceil\sqrt{k}\,\rceil$. For the error of the first kind,

$$1 - \sum_{x^m\in\mathcal{X}^m} P_i(x^m) W^m(\mathcal{D}(\mathcal{U}_i)|x^m) = 1 - \frac{1}{M'} \sum_{u\in\mathcal{U}_i} W^m(\mathcal{D}(\mathcal{U}_i)|x^m) \tag{2.1.7}$$

$$
\begin{aligned}
&= 1 - \frac{1}{M'} \sum_{u'_j \cdot u''_l \in \mathcal{U}_i} W^m \left(\bigcup_{j'=1}^{M'} D'_{j'} \times D''_{l_{j'}} | u'_j \cdot u''_l \right) && \text{(2.1.8)} \\
&= 1 - \frac{1}{M'} \sum_{u'_j \cdot u''_l \in \mathcal{U}_i} \sum_{j'=1}^{M'} W^m \left(D'_{j'} \times D''_{l_{j'}} | u'_j \cdot u''_l \right) && \text{(2.1.9)} \\
&\leq 1 - \frac{1}{M'} \sum_{u'_j \cdot u''_l \in \mathcal{U}_i} W^m \left(D'_j \times D''_l | u'_j \cdot u''_l \right) && \text{(2.1.10)} \\
&= \underbrace{1 - \frac{1}{M'} \sum_{u'_j \cdot u''_l \in \mathcal{U}_i} W^k \left(D'_j | u'_j \right)}_{>1-\lambda(k)} \cdot \underbrace{W^r \left(D''_l | u''_l \right)}_{>1-\lambda(\sqrt{k})} && \text{(2.1.11)} \\
&< \lambda(k) + \lambda(\sqrt{k}). && \text{(2.1.12)}
\end{aligned}
$$

Since i is chosen arbitrarily as is the realization $\mathcal{U}_i$, $e_1(\mathcal{C}_{\text{id}}, W^m) \leq \lambda(k) + \lambda(\sqrt{k})$.

For the second type of error, we consider a realization $\mathcal{U}_1$ of $\overline{\mathcal{U}}_1$ and random set $\overline{\mathcal{U}}_2$. We denote the elements of $\overline{\mathcal{U}}_2$ as U_j^2 and define a function, for $j \in [M']$,

$$
\psi_j(\overline{\mathcal{U}}_2) := \begin{cases} 1, & \text{if } U_j^2 \in \mathcal{U}_1 \\ 0, & \text{otherwise.} \end{cases}
$$

Note that because each U_j^2 is independent of the other elements of $\overline{\mathcal{U}}_2$, each $\psi_j(\overline{\mathcal{U}}_2)$ is also independent of $\psi_i(\overline{\mathcal{U}}_2), i \neq j$. It is easy to see that,

$$
\mathbb{E}(\psi_j(\overline{\mathcal{U}}_2)) = \frac{1}{M''} \quad \forall j \in [M'] \tag{2.1.13}
$$

since there is a $1/M''$ chance that $U_j^2 \in \mathcal{U}_1$, seen by considering the structure of the random variables. Further, for $\lambda \in (0, 1)$

$$
\begin{aligned}
D\left(\lambda \parallel \frac{1}{M''}\right) &= \lambda \log(\lambda 2^{r\epsilon}) + (1-\lambda) \log\left(\frac{1-\lambda}{1-2^{-r\epsilon}}\right) && \text{(2.1.14)} \\
&= \lambda \log(\lambda) + \lambda \log(2^{r\epsilon}) + (1-\lambda)\log(1-\lambda) \\
&\quad - (1-\lambda)\log\big(1-2^{-r\epsilon}\big) && \text{(2.1.15)} \\
&\geq \lambda \log(2^{r\epsilon}) + \log(0.5) && \text{(2.1.16)} \\
&\geq \lambda\sqrt{k}\epsilon - 1, && \text{(2.1.17)}
\end{aligned}
$$

where we use the fact that $\lambda = 0.5$ minimizes $\lambda\log(\lambda) + (1-\lambda)\log(1-\lambda)$. By Lemma 2.1.3 with $\lambda \in (0, 1)$ and k large enough such that $1/M'' < \lambda$, it holds that

$$
\Pr\left(\sum_{j=1}^{M'} \psi_j(\overline{\mathcal{U}}_2) > \lambda M'\right) \leq 2^{-M'(\lambda\sqrt{k}\epsilon - 1)}. \tag{2.1.18}
$$

Moreover, for any realization $\mathcal{U}_2$ of $\overline{\mathcal{U}}_2$, it holds for $u'_j \cdot u''_l \in \mathcal{U}_1 \setminus \mathcal{U}_2$, $\mathcal{D}(\mathcal{U}_2) \cap D'_j \times D''_l = \emptyset$ and so $D(\mathcal{U}_2) \subseteq (D'_j \times D''_l)^c$. Therefore it holds that

$$
\begin{aligned}
\frac{1}{M'} \sum_{u \in \mathcal{U}_1 \setminus \mathcal{U}_2} W^m(\mathcal{D}(\mathcal{U}_2) \mid u) &\leq \frac{1}{M'} \sum_{u'_j \cdot u''_l \in \mathcal{U}_1 \setminus \mathcal{U}_2} W^m((D'_j \times D''_l)^c \mid u'_j \cdot u''_l) && \text{(2.1.19)} \\
&\leq \frac{1}{M'} \sum_{u'_j \cdot u''_l \in \mathcal{U}_1} W^m((D'_j \times D''_l)^c \mid u'_j \cdot u''_l) && \text{(2.1.20)} \\
&= \frac{1}{M'} \sum_{u'_j \cdot u''_l \in \mathcal{U}_1} W^m(D'^c_j \times D''_l \cup D'_j \times D''^c_l \cup D'^c_j \times D''^c_l \mid u'_j \cdot u''_l) && \text{(2.1.21)}
\end{aligned}
$$

$$\begin{aligned}
&= \frac{1}{M'} \sum_{u'_j \cdot u''_l \in \mathcal{U}_1} W^m(D'^c_j \times D''_l \mid u'_j \cdot u''_l) + W^m(D'_j \times D''^c_l \mid u'_j \cdot u''_l) \\
&\qquad\qquad\qquad\qquad + W^m(D'^c_j \times D''^c_l \mid u'_j \cdot u''_l)
\end{aligned} \tag{2.1.22}$$

$$\begin{aligned}
&= \underbrace{\frac{1}{M'} \sum_{u'_j \cdot u''_l \in \mathcal{U}_1} W^k(D'^c_j \mid u'_j)}_{\leq \lambda(k)} \cdot \underbrace{W^r(D''_l \mid u''_l)}_{\leq 1} \\
&\quad + \frac{1}{M'} \sum_{u'_j \cdot u''_l \in \mathcal{U}_1} \underbrace{W^k(D'_j \mid u'_j)}_{\leq 1} \cdot \underbrace{W^r(D''^c_l \mid u''_l)}_{\leq \lambda(\sqrt{k})} \\
&\quad + \underbrace{\frac{1}{M'} \sum_{u'_j \cdot u''_l \in \mathcal{U}_1} W^k(D'^c_j \mid u'_j)}_{\leq \lambda(k)} \cdot \underbrace{W^r(D''^c_l \mid u''_l)}_{\leq \lambda(\sqrt{k})}
\end{aligned} \tag{2.1.23}$$

$$\leq \lambda(k) + \lambda(\sqrt{k}) + \lambda(k)\lambda(\sqrt{k}). \tag{2.1.24}$$

Define $\lambda_k := \lambda(k) + \lambda(\sqrt{k}) + \lambda(k)\lambda(\sqrt{k})$. Using this information we can estimate,

$$\sum_{x^m \in \mathcal{X}^m} P_1(x^m) W^m(\mathcal{D}(\overline{\mathcal{U}}_2) \mid x^m) = \frac{1}{M'} \sum_{u \in \mathcal{U}_1} W^m(\mathcal{D}(\overline{\mathcal{U}}_2) \mid u) \tag{2.1.25}$$

$$= \frac{1}{M'} \sum_{u \in \mathcal{U}_1 \cap \overline{\mathcal{U}}_2} \underbrace{W^m(\mathcal{D}(\overline{\mathcal{U}}_2) \mid u)}_{\leq 1} + \frac{1}{M'} \sum_{u \in \mathcal{U}_1 \setminus \overline{\mathcal{U}}_2} W^m(\mathcal{D}(\overline{\mathcal{U}}_2) \mid u) \tag{2.1.26}$$

$$\leq \frac{1}{M'} \mid \mathcal{U}_1 \cap \overline{\mathcal{U}}_2 \mid + \lambda_k \tag{2.1.27}$$

$$= \frac{1}{M'} \sum_{j=1}^{M'} \psi_j(\overline{\mathcal{U}}_2) + \lambda_k \tag{2.1.28}$$

By Lemma 2.1.3 we have that for $\lambda \in (0,1)$ and large enough k such that $1/M'' < \lambda$, with non-zero probability it holds that

$$\frac{1}{M'} \sum_{j=1}^{M'} \psi_j(\overline{\mathcal{U}}_2) + \lambda_k < \lambda + \lambda_k, \tag{2.1.29}$$

and therefore with non-zero probability,

$$\sum_{x^m \in \mathcal{X}^m} P_1(x^m) W^m(\mathcal{D}(\overline{\mathcal{U}}_2) | x^m) < \lambda + \lambda_k. \tag{2.1.30}$$

Similar arguments can be made to justify,

$$\sum_{x^m \in \mathcal{X}^m} \overline{P}_2(x^m) W^m(\mathcal{D}(\mathcal{U}_1) | x^m) < \lambda + \lambda_k. \tag{2.1.31}$$

Hence there exists a realization of $\overline{\mathcal{U}}_2, \mathcal{U}_2$, such that (2.1.30) and (2.1.31) hold. We follow the argumentation of [1]. Using $\mathcal{U}_1$ and $\mathcal{U}_2$, we have then that with positive probability $|\mathcal{U}_1 \cap \mathcal{U}_2| \leq \lambda M'$. We would now like to add a third element, $\mathcal{U}_3$ to the set such that with positive probability, $|\mathcal{U}_1 \cap \mathcal{U}_3| \leq \lambda M'$ and $|\mathcal{U}_2 \cap \mathcal{U}_3| \leq \lambda M'$. This holds when,

$$2 \cdot \Pr\left(\sum_{j=1}^{M'} \psi_j(\overline{\mathcal{U}}_3) > \lambda M' \right) < 1, \tag{2.1.32}$$

where we define the analogous of ψ_j for $\mathcal{U}_2$. Repeating this argument for $i = 4, ..., M$ we have that for each new $\mathcal{U}_i$, it should hold that

$$(M-1) \cdot \Pr\left(\sum_{j=1}^{M'} \psi_j(\overline{\mathcal{U}}_i) > \lambda M' \right) < 1, \tag{2.1.33}$$

Thus, it should hold, for each $i = 2, ..., M$,

$$\Pr\left(\sum_{j=1}^{M'} \psi_j(\overline{\mathcal{U}}_i) > \lambda M'\right) \leq 2^{-M'(\lambda\sqrt{k}\epsilon - 1)} < \frac{1}{M-1}, \tag{2.1.34}$$

and so, considering that we need to fulfill the condition $M - 1$ times, it must hold that

$$2^{-M'(\lambda\sqrt{k}\epsilon-1)} < \frac{1}{(M-1)^2} \Rightarrow (M-1)^2 < 2^{M'(\lambda\sqrt{k}\epsilon-1)} \tag{2.1.35}$$

$$\Rightarrow M - 1 < 2^{M'(\lambda\sqrt{k}\epsilon-1)/2} \tag{2.1.36}$$

$$\Rightarrow M \leq 2^{M'(\lambda\sqrt{k}\epsilon-1)/2} = 2^{2^{k(R-\eta)}(\lambda\sqrt{k}\epsilon-1)/2} \tag{2.1.37}$$

for there to exist a collection $(\mathcal{U}_1, ..., \mathcal{U}_M)$ such that the error of second kind is bounded. Therefore, with non-zero probability and M chosen with respect to the above conditions, since the errors are bounded for all $\lambda \in (0, 1)$, there is an (m, M)-ID-code $\mathcal{C}_{\text{id}}$ that achieves ID-rate R as $k \to \infty$. Further, in the limit $k \to \infty$, $m = k$ and so $\mathcal{C}_{\text{id}}$ is a (k, M) ID-code. □

A consequence of the Transformator lemma is the achievability part of Theorem 1.2.3, that is, finding the existence of an ID-code that can achieve the capacity of the channel. What we can learn from the lemma is that transmission codes can be "transformed" such that by appending small amounts of randomness, an ID-code can be formed that achieves ID rates equal to the transmission rates. We will see that this strategy works in the quantum case as well.

Chapter 3

The Quantum Multiple Access Channel

The study of information transfer over quantum channels is known as quantum Shannon theory as it extends much of the theory invented by Shannon into the quantum regime. In this chapter we review the definitions of quantum states and channels and some properties of them. We will introduce the quantum multiple access channel, the channel in which the main result of this thesis relies on. We will then review the transmission capacity theorem for the multi-sender-multi-receiver channel proved by Andreas Winter in [6] in the less general case of the multiple access channel.

3.1 Quantum Shannon Theory

In classical computing, the most fundamental object is the bit, up or down, usually represented by 0 or 1. The analogous fundamental system in terms of quantum computing is called a qubit, short for quantum bit. A qubit is a two level quantum system, the two levels sometimes denoted with bra-ket notation as $|0\rangle$ and $|1\rangle$, where the key difference between a bit and qubit is that a qubit can be in a superposition of the two levels, for example, $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, where $\alpha, \beta \in \mathbb{C}$ such that $|\alpha|^2 + |\beta|^2 = 1$. This difference is a main reason why quantum theory is embedded into a linear algebraic setting, treating qubits as vectors in a complex Hilbert space.

When a quantum system becomes more complicated, involving things like entanglement and coherence, it is beneficial to model it via density matrices. The following definition is how quantum states will be regarded in this thesis.

Definition 3.1.1 (Quantum State)**.** *For a Hilbert space $\mathcal{H}$, a quantum state $\rho \in \mathcal{L}(\mathcal{H})$ is such that $\rho \geq 0$ and $\mathrm{Tr}(\rho) = 1$. The set of quantum states in $\mathcal{H}$ is denoted $\mathcal{S}(\mathcal{H})$.*

Composing many quantum states into one system allows us to ask more interesting questions regarding quantum Shannon theory, and therefore a method describing the composition is needed. In this Hilbert space setting, a natural method is to use the tensor product. For a family of quantum states $\{\rho_i\}_{i=1}^{M} \subseteq \mathcal{S}(\mathcal{H})$, $M \in \mathbb{N}$, the tensor product of the states will be represented as,

$$\bigotimes_{i=1}^{M} \rho_i := \rho_1 \otimes \rho_2 \otimes \ldots \otimes \rho_M. \tag{3.1.1}$$

This tensor product state is then a quantum state in the product space $\mathcal{H}^{\otimes M}$, or in other terms an element of $\mathcal{S}(\mathcal{H}^{\otimes M})$, where

$$\mathcal{H}^{\otimes M} := \underbrace{\mathcal{H} \otimes \ldots \otimes \mathcal{H}}_{M}. \tag{3.1.2}$$

In Shannon theory, both classical and quantum, the task is to transmit information in some way and, to accomplish this, a type of communication channel is used. In the classical setting as we have seen already, a stochastic matrix is used which can model the noise when transmitting classical information over the channel. In the quantum setting, because quantum information is encoded in a Hilbert space, the model must be generalized such that quantum states are mapped to quantum states. We therefore define a quantum channel as follows.

Definition 3.1.2 (Quantum Channel)**.** *For Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, a linear map $W : \mathcal{S}(\mathcal{H}_A) \to \mathcal{S}(\mathcal{H}_B)$ is called a quantum channel if it is completely positive and trace preserving (CPTP). We refer the reader to Appendix A for the definition of CPTP.*

Classical information can be transmitted via traditional means, that is, without the notions of quantum mechanics. It is possible though, to transmit classical information via a quantum channel by preparing quantum systems in a particular way to encode the classical information and to also recover it on the receiving end of the transmission. We denote that a quantum channel is sending classical information by setting the domain of channels as finite alphabet instead of the general set of states in a Hilbert space, with the implication that there is an underlying encoding map that converts codewords into quantum states to be transmitted.

As with the classical channel, we are interested in multiple uses of a quantum channel to send composite quantum systems encoded with classical codewords. To encapsulate this idea, we define what we call a classical-quantum (CQ) channel.

Definition 3.1.3 (CQ channel)**.** *A CQ channel $\mathbf{W}$ is a family of quantum channels*

$$\{W^k : \mathcal{X}^k \to \mathcal{S}(\mathcal{H}^{\otimes k})\}_{k\in\mathbb{N}}. \tag{3.1.3}$$

When a quantum channel is memoryless, that is, each use of the channel is independent of the last, it a special case of a CQ channel and so we formulate a distinct definition to contain the concept of a discrete memoryless classical-quantum (DM-CQ) channel.

Definition 3.1.4 (DM-CQ channel)**.** *Let $\mathbf{W}$ be a CQ channel. We say that $\mathbf{W}$ is a DM-CQ channel generated by W if for the family $\{W^k : \mathcal{X}^k \to \mathcal{S}(\mathcal{H}^{\otimes k})\}_{k\in\mathbb{N}}$, where, with $x^k = (x_1, ..., x_k) \in \mathcal{X}^k$, for each k, it holds*

$$W^k(x^k) = \bigotimes_{i=1}^{k} W(x_i). \tag{3.1.4}$$

In this thesis, we focus on the multiple access channel. The multiple access channel is such that there are two senders transmitting information to one receiver. The multiple access channel we investigate here is one which has two senders transmitting classical information to a receiver who receives a single quantum state. This channel is what we refer to as a classical-classical-quantum (CCQ) channel. Again, we would like to allow for multiple uses of this channel and so we define the following.

Definition 3.1.5 (CCQ channel)**.** *A CCQ channel $\mathbf{W}$ is a family of quantum channels*

$$\{W^k : \mathcal{X}^k \times \mathcal{Y}^k \to \mathcal{S}(\mathcal{H}^{\otimes k})\}_{k\in\mathbb{N}}. \tag{3.1.5}$$

For the analogous memoryless channel, we define,

Definition 3.1.6 (DM-CCQ channel)**.** *Let $\mathbf{W}$ be a CCQ channel. We say that $\mathbf{W}$ is a DM-CCQ channel generated by W if for the family $\{W^k : \mathcal{X}^k \times \mathcal{Y}^k \to \mathcal{S}(\mathcal{H}^{\otimes k})\}_{k\in\mathbb{N}}$, where with $x^k = (x_1, ..., x_k) \in \mathcal{X}^k$ and $y^k = (y_1, ..., y_k) \in \mathcal{Y}^k$, for each k,*

$$W^k(x^k, y^k) = \bigotimes_{i=1}^{k} W(x_i, y_i). \tag{3.1.6}$$

Without going deeply into the physics, classical information is encoded quantumly using natural properties of a quantum system, such as its energy level or polarization. Natural properties of quantum states can be measured in a laboratory, and therefore make reasonable resources to use for encoding. We can abstract these properties by embedding them into a Hilbert space and using a basis to represent the various energy levels, for example. When sending a quantumly encoded message over a quantum channel, because in general there is noise, it cannot be guaranteed that the state maintains its exact encoding, possibly due to decoherence, and so the receiver can generally learn less from the incoming message. One way to overcome this loss of information is for the sender to "record" the message they send by adding an additional quantum system with the classical information encoded within it. To represent the overall state for possible messages, we have the following.

Definition 3.1.7 (Channel state)**.** *For CCQ channel* $\mathbf{W}$*, probability distributions* $p_1 \in \mathcal{P}(\mathcal{X}^k)$ *and* $p_2 \in \mathcal{P}(\mathcal{Y}^k)$*, and Hilbert spaces* $\mathcal{H}_A$ *and* $\mathcal{H}_B$ *with respective orthonormal bases* $\{|x^k\rangle\}_{x^k \in \mathcal{X}^k}$ *and* $\{|y^k\rangle\}_{y^k \in \mathcal{Y}^k}$*, the channel state is defined as*

$$\gamma_2^k(p_1, p_2) := \sum_{\substack{x^k \in \mathcal{X}^k \\ y^k \in \mathcal{Y}^n}} p_1(x^k) p_2(y^k) \, |x^k\rangle\!\langle x^k| \otimes |y^k\rangle\!\langle y^k| \otimes W^k(x^k, y^k). \tag{3.1.7}$$

For $p_1 \in \mathcal{P}(\mathcal{X})$ *and* $p_2 \in \mathcal{P}(\mathcal{Y})$*, we write for notational simplicity* $\gamma_2(p_1, p_2) := \gamma_2^1(p_1, p_2)$*. We will also use CQ channel states defined similarly, where we denote* γ_1 *and* γ_1^k *for single sender channel states in a similar way, with a single distribution parameter over one orthonormal basis.*

One way to extract information from a quantum state is to perform a measurement on it. Measurements on quantum systems, as with measurements on classical systems, are actions performed on a system to determine information about it. For example, in the classical transmission problem, when the receiver receives an encoded message, he performs a decoding operation on it to extract the message that it represents. This process can be thought of as a measurement. With a classical message encoded quantumly, the receiver is still interested in decoding what classical information the system contains, but is faced with the additional complexity that comes with measuring quantum systems, therefore a different model is needed for decoding.

Measurable natural properties of a quantum system are known as observables. When observing a quantum system, the "quantumness" of it is generally lost and the state falls into a classical state. In communication problems that transmit classical information encoded quantumly, this is not a problem, but the goal. Measuring the quantum state that represents the message, which is then outputted by the measurement, is the decoding process. What is important then is to minimize the probability that the measurement performed by the receiver outputs the incorrect result. When an observable has a finite set of outcomes we can model the measurement mathematically using a positive operator-value measure.

Definition 3.1.8 (Positive Operator-Valued Measure)**.** *For a Hilbert space* $\mathcal{H}$*, a positive operator valued measure (POVM) is a family of linear operators* $(D_i)_{i \in \mathcal{I}} \subset \mathcal{L}(\mathcal{H})$*,* $\mathcal{I}$ *a finite index set, such that for all* $i \in \mathcal{I}$*,* $0 \leq D_i \leq \mathbb{1}_{\mathcal{H}}$ *and* $\sum_{i \in \mathcal{I}} D_i = \mathbb{1}_{\mathcal{H}}$*, where* $\mathbb{1}_{\mathcal{H}}$ *is the identity operator on* $\mathcal{H}$*.*

With a quantum state $\rho \in \mathcal{S}(\mathcal{H})$, a random variable is induced over the finite set of outputs for the measurement. For a POVM $(D_i)_{i=1}^M$, $M \in \mathbb{N}$ finite, the random variable A with realizations in $[M]$ has the distribution

$$\Pr(A = i) = \operatorname{tr}(\rho D_i). \tag{3.1.8}$$

This provides a straight forward way to have a figure of merit when modeling codes for quantum channels, as the aim would be to construct a random variable with the channel output state and POVM with elements representing the finite set of messages such that there is a high likelihood that the realization is the encoded message.

As we are now in the quantum setting, as we will be for the remainder of the thesis, we recycle the naming conventions for the terminology of the classical setting and formalize the quantum analogies for the definitions seen in the previous chapters for classical transmission and identification. With this, we can formalize the coding structure for a CQ channel.

Definition 3.1.9 ((k, M)-code)**.** *For a CQ channel* $\mathbf{W}$*, a* (k, M)*-code for classical message transmission is the family* $\mathcal{C} := (x_m, D_m)_{m=1}^M$*, where* $x_1, ..., x_M \in \mathcal{X}^k$ *and* $(D_m)_{m=1}^M \subset \mathcal{L}(\mathcal{H}^{\otimes k})$ *forms a*

POVM.

For a (k, M)-code $\mathcal{C}$, the average error of transmission is defined as

$$\overline{e}(\mathcal{C}, W^k) := 1 - \frac{1}{M}\sum_{m=1}^{M} \operatorname{tr}\big(D_m W^k(x_m)\big) \tag{3.1.9}$$

and the maximal,

$$e(\mathcal{C}, W^k) := \max_{m} \;\; 1 - \operatorname{tr}\big(D_m W^k(x_m)\big). \tag{3.1.10}$$

Further, we define,

$$M(k, \lambda) := \max_{M} \{M \in \mathbb{N} \mid \exists \textit{ a } (k, M) - \textit{code } \mathcal{C} \textit{ s.t. } \overline{e}(\mathcal{C}, W^k) \leq \lambda\}. \tag{3.1.11}$$

For a CCQ channel, there are two senders and one receiver and so naturally the code will have two sets of codewords, one for each sender, and a single decoding POVM for the receiver. The decoding measure will contain an event for each possible pair of incoming messages. We formalize this in the following definition.

Definition 3.1.10 ((k, M, N)-code)**.** *For a CCQ channel* $\mathbf{W}$*, a (k, M, N)-code for classical message transmission is the family $\mathcal{C} := (x_m, y_n, D_{mn})_{m=1,n=1}^{M,N}$ where $x_1, ..., x_M \in \mathcal{X}^k$, $y_1, ..., y_N \in \mathcal{Y}^k$, and $(D_{mn})_{m=1,n=1}^{M,N} \subset \mathcal{L}(\mathcal{H}^{\otimes k})$ forms a POVM.*

For a (k, M, N)-code $\mathcal{C}$, the average error of transmission is defined as

$$\overline{e}(\mathcal{C}, W^k) := 1 - \frac{1}{MN}\sum_{\substack{m=1\\n=1}}^{M,N} \operatorname{tr}\big(D_{mn} W^k(x_m, y_n)\big), \tag{3.1.12}$$

and the maximal,

$$e(\mathcal{C}, W^k) := \max_{m,n} \;\; 1 - \operatorname{tr}\big(D_{mn} W^k(x_m, y_n)\big). \tag{3.1.13}$$

As seen with the classical codes, we would like to determine the rates at which information can be transmitted per channel use, and so we define the following.

Definition 3.1.11 (Achievable rate)**.** *For a CQ channel* $\mathbf{W}$*, we say $R \in \mathbb{R}, R \geq 0$, is an achievable rate if for all $\epsilon, \delta > 0$, there exists a k_0 such that for all $k \geq k_0$, there is a (k, M)-code $\mathcal{C}$ such that,*

$$\frac{1}{k}\log M \geq R - \delta, \;\; \textit{and} \quad e(\mathcal{C}, W^k) \leq \epsilon. \tag{3.1.14}$$

Considering that the CCQ channel has two senders, determining the set of achievable rate pairs becomes a two dimensional problem, a rate for each sender, and it forms a region which we call its capacity region. To formalize this, we define the following.

Definition 3.1.12 (Achievable rate pair)**.** *For a CCQ channel* $\mathbf{W}$*, we say $(R_1, R_2) \in \mathbb{R}^2$, $R_1, R_2 \geq 0$, is an achievable rate pair if for all $\epsilon, \delta > 0$, there exists a k_0 such that for all $k \geq k_0$ there exists a (k, M, N)-code $\mathcal{C}$ such that,*

$$\frac{1}{k}\log M \geq R_1 - \delta, \quad \frac{1}{k}\log N \geq R_2 - \delta, \quad \overline{e}(\mathcal{C}, W^k) \leq \epsilon. \tag{3.1.15}$$

The capacity region for $\mathbf{W}$ *is defined as*

$$C(\mathbf{W}) := \{(R_1, R_2) \mid (R_1, R_2) \textit{ is an achievable rate pair}\}. \tag{3.1.16}$$

We say a (k, M, N)-code $\mathcal{C}$ achieves the rate pair (R_1, R_2) if (3.1.15) holds for all $\epsilon, \delta > 0$.

Notation 3.1.13. *If* $\mathbf{W}$ *is a discrete memoryless channel generated by a quantum channel* W*, we refer to its capacity as* $C(W)$*, that is, without the bold face character.*

In many cases, it is useful to use a CCQ channel as a CQ channel. We show with the next lemma that is possible to do this without any loss of precision or rate.

Lemma 3.1.14. *For a CCQ channel* $\mathbf{W}$*, a* (k, M, N)*-code* $\mathcal{C}$ *with* $\overline{e}(\mathcal{C}, W^k) \leq \epsilon$ *can be a* (k, M) *or* (k, N)*-code* $\mathcal{C}_M$ *or* $\mathcal{C}_N$ *over channels* W_M^k *and* W_N^k *respectively with* $\overline{e}(\mathcal{C}_M, W_M^k) = \overline{e}(\mathcal{C}_N, W_N^k) \leq \epsilon$*, where* W_M^k *and* W_N^k *are the CQ channels generated by averaging out one sender of the CCQ channel* W^k*.*

Proof. Let $\mathcal{C} \coloneqq (x_m, y_n, D_{mn})_{m=1,n=1}^{M,N}$ be a (k, M, N)-code with $\overline{e}(\mathcal{C}, W^k) \leq \epsilon$. Consider the channel

$$W_M^k : \mathcal{X}^k \to \mathcal{S}(\mathcal{H}^{\otimes k}) \tag{3.1.17}$$

$$: x^k \mapsto \frac{1}{N} \sum_{n=1}^{N} W^k(x^k, y_n). \tag{3.1.18}$$

Define $s_m(D_{mn}) \coloneqq \sum_{n=1}^{N} D_{mn}$ and the code $\mathcal{C}_M \coloneqq (x_m, s_m(D_{mn}))_{m=1}^{M}$.

$$\overline{e}(W_M^k, \mathcal{C}_M) = 1 - \frac{1}{M} \sum_{m=1}^{M} \operatorname{tr}\big(s_m(D_{mn}) W_M^k(x_m)\big) \tag{3.1.19}$$

$$= 1 - \frac{1}{MN} \sum_{n=1}^{N} \sum_{m=1}^{M} \operatorname{tr}\left(\left(\sum_{n'=1}^{N} D_{mn'}\right) W^k(x_m, y_n)\right) \tag{3.1.20}$$

$$\leq 1 - \frac{1}{MN} \sum_{n=1}^{N} \sum_{m=1}^{M} \operatorname{tr}\big(D_{mn} W^k(x_m, y_n)\big) \tag{3.1.21}$$

$$\leq \epsilon. \tag{3.1.22}$$

Thus $\mathcal{C}_M$ is a (k, M)-code that has average error bounded by ϵ over the channel W_M^k. Analogous arguments can be made to construct a (k, N)-code with bounded average error. □

In the definition for an achievable rate pair, we used an average error figure of merit. In latter parts of this thesis, we will make use of (k, M, N)-codes with a bounded maximal error, and so it is useful to distinguish between the rate pairs that are achievable with a maximal error criterion.

Definition 3.1.15 (Maximal-error achievable rate pair)**.** *For a CCQ channel* $\mathbf{W}$*, we say* $(R_1, R_2) \in \mathbb{R}^2$*,* $R_1, R_2 \geq 0$*, is a max-error achievable rate pair if for all* $\epsilon, \delta > 0$*, there exists a* k_0 *such that for all* $k \geq k_0$ *there exists a* (k, M, N)*-code* $\mathcal{C}$ *such that,*

$$\frac{1}{k} \log M \geq R_1 - \delta, \quad \frac{1}{k} \log N \geq R_2 - \delta, \quad e(\mathcal{C}, W^k) \leq \epsilon. \tag{3.1.23}$$

The max-error capacity region for $\mathbf{W}$ *is defined as*

$$C_{\max}(\mathbf{W}) \coloneqq \{(R_1, R_2) \mid (R_1, R_2) \textit{ is a max-error achievable rate pair}\}. \tag{3.1.24}$$

We say a (k, M, N)*-code* $\mathcal{C}$ *maximally achieves the rate pair* (R_1, R_2) *if (3.1.23) holds for all* $\epsilon, \delta > 0$*.*

For the remainder of this chapter we use what was introduced here to review the DM-CCQ capacity theorem proved in a more general sense by Andreas Winter.

Theorem 3.1.16 (Winter, [6])**.** *For a DM-CCQ generate by* $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{S}(\mathcal{H}_C)$*, it holds that*

$$C(W) = \bigcup_{\substack{p_1 \in \mathcal{P}(\mathcal{X}) \\ p_2 \in \mathcal{P}(\mathcal{Y})}} \big\{(R_1, R_2) \mid 0 \leq R_1 \leq I(A, C|B)_{\gamma_2(p_1,p_2)}, 0 \leq R_2 \leq I(B, C|A)_{\gamma_2(p_1,p_2)}, \tag{3.1.25}$$

$$R_1 + R_2 \leq I(A, B; C)_{\gamma_2(p_1,p_2)}\big\},$$

where A *and* B *refer to the respective Hilbert spaces for the two senders and* C *refers to the output Hilbert space of the channel.*

3.2 Review of the Achievability Theorem

The strategy to prove the achievability part of the theorem is to analyze a (k, M, N)-code constructed using random codebooks and show that there exist realizations of the codebooks that will achieve the desired rate. By the time sharing principle, if we show all extremal points of the capacity region are achievable then we have also shown the closed convex hull of the extremal points is also achievable. With the random codebooks, we construct two single sender-single receiver codes for the DM-CQ channel generated by removing one sender channel by assuming they always send their average message. By the Holveo-Schumacher-Westmoreland theorem, we will have the existence of such codes and we can then merge them in such a way that a POVM can be constructed that recovers the combined messages with arbitrarily high probability. We introduce the following theory used to prove the achievability theorem.

Lemma 3.2.1 (Gentle Measurement, Average Version)**.** *[6, Lemma 8] For $M, N \in \mathbb{N}$ finite, a sequence of states $(\rho_i)_{i\in[M]} \subseteq \mathcal{S}(\mathcal{H})$, a POVM $(D_j)_{j\in[N]} \subset \mathcal{L}(\mathcal{H})$, probability distribution $p \in \mathcal{P}([M])$, and any map $\psi : [M] \to [N]$ such that for any $\epsilon > 0$ it holds,*

$$1 - \sum_{i\in[M]} p(i)\,\mathrm{tr}\big(\rho_i D_{\psi(i)}\big) \leq \epsilon, \tag{3.2.1}$$

with the operator defined as, with orthonormal basis $(|x_j\rangle)_{j\in[N]}$ for an additional space $\mathcal{H}'$,

$$\Delta : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H}' \otimes \mathcal{H}) \tag{3.2.2}$$

$$: \rho \mapsto \frac{1}{N}\sum_{j=1}^{N} |x_j\rangle\!\langle x_j| \otimes \sqrt{D_j}\rho\sqrt{D_j}, \tag{3.2.3}$$

it holds that,

$$\sum_{i=1}^{M} p(i)\,\big\| |x_{\psi(i)}\rangle\!\langle x_{\psi(i)}| \otimes \rho_i - \Delta(\rho_i)\big\|_1 \leq \sqrt{8\epsilon} - \epsilon. \tag{3.2.4}$$

Theorem 3.2.2. *[5, Theorem 11.2.2] For channel state $\gamma \coloneqq \gamma_1(p_A)$, where p_A is the distribution that governs a random variable A with realizations in $\mathcal{X}$, it holds,*

$$H(A, C)_\gamma = H(A) + \sum_{x\in\mathcal{X}} p_A(x) H(C)_\gamma. \tag{3.2.5}$$

with $H(A)$ the Shannon entropy of A where we refer the reader to Appendix B for the definition of Shannon entropy.

Theorem 3.2.3 (Holevo-Schumacher-Westmoreland)**.** *For a DM-CQ channel generated by W,*

$$\lim_{k\to\infty} \frac{1}{k}\log M(\lambda, k) = \sup_{p\in\mathcal{P}(\mathcal{X})} \chi(p, W) \tag{3.2.6}$$

where $\chi(p, W)$ is the Holveo quantity for the pair (p, W) where we refer the reader to Appendix B for the definition of the Holevo quantity.

Lemma 3.2.4. *[6, Appendix] For a fixed channel state γ for a DM-CCQ The upper extremal points of the region*

$$\left\{(R_1, R_2) \mid 0 \leq R_1 \leq I(A, C|B)_\gamma, 0 \leq R_2 \leq I(B, C|A)_\gamma, R_1 + R_2 \leq I(A, B; C)_\gamma\right\}. \tag{3.2.7}$$

are

$$(I(A; C)_\gamma, I(B; A, C)_\gamma) \quad \textit{and} \quad (I(A; B, C)_\gamma, I(B; C)_\gamma). \tag{3.2.8}$$

Theorem 3.2.5. *[6, Theorem 9] For the DM-CCQ channel generated by $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{S}(\mathcal{H}_C)$, let R_1, R_2 be non-negative real numbers such that with $p_1 \in \mathcal{P}(\mathcal{X})$ and $p_2 \in \mathcal{P}(\mathcal{Y})$, the constraints*

$$R_1 \leq I(A; C \mid B)_\gamma \tag{3.2.9}$$

$$R_2 \leq I(B; C \mid A)_\gamma \tag{3.2.10}$$

$$R_1 + R_2 \leq I(A, B; C)_\gamma, \tag{3.2.11}$$

evaluated with respect to the channel state $\gamma \coloneqq \gamma_2(p_1, p_2)$ hold, then $(R_1, R_2) \in C(W)$.

Proof. By the time-sharing principle, we need only prove that the upper extremal points of the capacity region defined by the above constraints are achievable. Since the proof runs the same way regardless of the selection of extremal point, we prove, without loss of generality, the achievability of one upper extremal point. By Lemma 3.2.4, the extremal points are

$$(I(A; C)_\gamma, I(B; A, C)_\gamma) \quad \text{and} \quad (I(A; B, C)_\gamma, I(B; C)_\gamma), \tag{3.2.12}$$

and so we show $(R_1, R_2) \coloneqq (I(A; C)_\gamma, I(B; A, C)_\gamma)$ is an achievable rate pair. Let $\epsilon, \delta > 0$ and $k \in \mathbb{N}$. Consider two random codebooks $C_1 \coloneqq (x_1^k, ..., x_M^k) \subset \mathcal{X}^k$ and $C_2 \coloneqq (y_1^k, ..., y_N^k) \subset \mathcal{Y}^k$ with $M \coloneqq \lceil 2^{k(R_1 - \delta)} \rceil$ and $N \coloneqq \lceil 2^{k(R_2 - \delta)} \rceil$ chosen independently and identically distributed according to

$$p_1^k : x^k \in \mathcal{X}^k \mapsto \prod_{i=1}^k p_1(x_i) \quad \text{and} \quad p_2^k : y^k \in \mathcal{Y}^k \mapsto \prod_{i=1}^k p_2(y_i), \tag{3.2.13}$$

respectively, where p_1 and p_2 govern random variables A and B, respectively.

Consider the two channels,

$$V_1^k : \mathcal{X}^k \to \mathcal{S}(\mathcal{H}_C^{\otimes k}) \tag{3.2.14}$$

$$: x^k \mapsto \frac{1}{N} \sum_{n=1}^N W^k(x^k, y_n^k), \tag{3.2.15}$$

and

$$V_2^k : \mathcal{Y}^k \to \mathcal{S}(\mathcal{H}_A^{\otimes k} \otimes \mathcal{H}_C^{\otimes k}) \tag{3.2.16}$$

$$: y^k \mapsto \frac{1}{M} \sum_{m=1}^M |x_m^k\rangle\langle x_m^k| \otimes W^k(x_m^k, y^k). \tag{3.2.17}$$

V_1^k is a random CQ channel depending on the codebook C_2, as is V_2^k with C_1. Now, define the two quantum channels as following.

$$S_1 : \mathcal{X} \to \mathcal{S}(\mathcal{H}_C) \tag{3.2.18}$$

$$: x \mapsto \sum_{y \in \mathcal{Y}} p_2(y) W(x, y), \tag{3.2.19}$$

and

$$S_2 : \mathcal{Y} \to \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_C) \tag{3.2.20}$$

$$: y \mapsto \sum_{x \in \mathcal{X}} p_1(x) \, |x\rangle\langle x| \otimes W(x, y). \tag{3.2.21}$$

Taking the expected channel output over the choice of C_2 we compute,

$$\langle V_1^k(x^k) \rangle_{C_2} = \sum_{y^k \in \mathcal{Y}^k} p_2^k(y^k) W^k(x^k, y^k) \tag{3.2.22}$$

$$= \sum_{y^k \in \mathcal{Y}^k} \bigotimes_{i=1}^k p_2(y_i) W(x_i, y_i) \tag{3.2.23}$$

$$= \bigotimes_{i=1}^{k} \sum_{y\in\mathcal{Y}} p_2(y)W(x_i, y) \tag{3.2.24}$$

$$= S_1^k(x^k), \tag{3.2.25}$$

where the use of the discrete memorylessness of W and p_2 is used in the second equality. Hence $\langle V_1^k(x^k)\rangle_{C_2}$ is a DM-CQ generated by S_1. Similarly,

$$\langle V_2^k(y^k)\rangle_{C_1} = \sum_{x^k\in\mathcal{X}^k} p_1^k(x^k)\,|x^k\rangle\langle x^k| \otimes W^k(x^k, y^k) \tag{3.2.26}$$

$$= \sum_{x^k\in\mathcal{X}^k} \bigotimes_{i=1}^{k} p_1(x_i)\,|x_i\rangle\langle x_i| \otimes W(x_i, y_i) \tag{3.2.27}$$

$$= \bigotimes_{i=1}^{k} \sum_{x\in\mathcal{X}} p_1(x)\,|x\rangle\langle x| \otimes W(x, y_i) \tag{3.2.28}$$

$$= S_2^k(y^k), \tag{3.2.29}$$

and therefore $\langle V_2^k(y^k)\rangle_{C_1}$ is a DM-CQ channel generated by S_2. Now,

$$I(A;C)_\gamma = H(A)_\gamma + H(C)_\gamma - H(A,C)_\gamma \tag{3.2.30}$$

$$= H(A)_\gamma + H\left(\sum_{x,y} p_1(x)p_2(y)W(x,y)\right) - H(A)_\gamma - \sum_x p_1(x)H\left(\sum_y p_2(y)W(x,y)\right) \tag{3.2.31}$$

$$= H\left(\sum_x p_1(x)S_1(x)\right) - \sum_x p_1(x)H\left(S_1(x)\right) \tag{3.2.32}$$

$$= \chi(p_1, S_1), \tag{3.2.33}$$

where Lemma 3.2.2 has been applied from the first to second equality with $H(A,C)_\gamma$. Further,

$$I(B;A,C)_\gamma = H(B)_\gamma + H(A,C)_\gamma - H(B,A,C)_\gamma \tag{3.2.34}$$

$$= H(B)_\gamma + H(A,C)_\gamma - H(B)_\gamma - \sum_y p_2(y)H\left(\sum_x p_1(x)\,|x\rangle\langle x| \otimes W(x,y)\right) \tag{3.2.35}$$

$$= H\left(\sum_{x,y} p_2(y)p_1(x)\,|x\rangle\langle x| \otimes W(x,y)\right) - \sum_y p_2(y)H\left(\sum_x p_1(x)\,|x\rangle\langle x| \otimes W(x,y)\right) \tag{3.2.36}$$

$$= H\left(\sum_y p_2(y)S_2(y)\right) - \sum_y p_2(y)H\left(S_2(y)\right) \tag{3.2.37}$$

$$= \chi(p_2, S_2), \tag{3.2.38}$$

and therefore, by Theorem 3.2.3, since $R_1 = \chi(p_1, S_1)$ and $R_2 = \chi(p_2, S_2)$ are achievable, there are, for large enough $k \in \mathbb{N}$, (k, M)-codes and (k, N)-codes

$$\mathcal{C}_1 := (x_m^k, D_m^1)_{m=1}^M \quad \text{and} \quad \mathcal{C}_2 := (y_n^k, D_n^2)_{n=1}^N, \tag{3.2.39}$$

with $(D_m^1)_{m=1}^M$ a POVM on $\mathcal{H}_C^{\otimes k}$ and $(D_n^2)_{n=1}^M$ a POVM on $\mathcal{H}_A^{\otimes k} \otimes \mathcal{H}_C^{\otimes k}$. Further, with the given ϵ, it holds,

$$\overline{e}(\mathcal{C}_1, S_1^k) = 1 - \frac{1}{M}\sum_{m=1}^{M} \operatorname{tr}\bigl(S_1^k(x_m^k)D_m^1\bigr) \leq \epsilon/2 \tag{3.2.40}$$

$$\overline{e}(\mathcal{C}_2, S_2^k) = 1 - \frac{1}{N}\sum_{n=1}^{N} \operatorname{tr}\bigl(S_2^k(y_n^k)D_n^2\bigr) \leq \epsilon/2. \tag{3.2.41}$$

We consider the codes $\mathcal{C}_1$ and $\mathcal{C}_2$ for random channels V_1 and V_2. The random average errors are then,

$$\overline{e}_1 \coloneqq \overline{e}(\mathcal{C}_1, V_1^k) = 1 - \frac{1}{M}\sum_{m=1}^{M} \operatorname{tr}\left(V_1^k(x_m^k)D_m^1\right) \tag{3.2.42}$$

$$\overline{e}_2 \coloneqq \overline{e}(\mathcal{C}_2, V_2^k) = 1 - \frac{1}{N}\sum_{n=1}^{N} \operatorname{tr}\left(V_2^k(y_n^k)D_n^2\right). \tag{3.2.43}$$

Because the only random part of each random average error is the channel, averaging $\overline{e}_1$ and $\overline{e}_2$ over codebooks gives

$$\langle \overline{e}_1\rangle_{C_2} = \overline{e}(\mathcal{C}_1, S_1^k) \leq \epsilon/2 \tag{3.2.44}$$

$$\langle \overline{e}_2\rangle_{C_1} = \overline{e}(\mathcal{C}_2, S_2^k) \leq \epsilon/2 \tag{3.2.45}$$

and therefore $\langle \overline{e}_1 + \overline{e}_2\rangle_{C_1,C_2} = \langle \overline{e}_1\rangle_{C_2} + \langle \overline{e}_2\rangle_{C_1} \leq \epsilon$ which implies that there are indeed codebooks C_1 and C_2 such that $\overline{e}_1 \leq \epsilon$ and $\overline{e}_2 \leq \epsilon$. Fixing codebooks C_1 and C_2 that obtain the desired error bounds, we redefine for V_1^k and V_2^k, $\mathcal{C}_1$ to use $C_1 \coloneqq (x_m^k)_{m=1}^M$ as its codebook and $\mathcal{C}_2$ to use $C_2 \coloneqq (y_n^k)_{n=1}^N$ as its codebook, with the decoding POVMs unchanged.

We now show that $\mathcal{C}_1$ and $\mathcal{C}_2$ can be merged to form a (k, M, N)-code with bounded average error on the order of ϵ. Define the two operators

$$\Delta_1 : \mathcal{S}(\mathcal{H}_C^{\otimes k}) \to \mathcal{S}(\mathcal{H}_A^{\otimes k} \otimes \mathcal{H}_C^{\otimes k}) \tag{3.2.46}$$

$$: \rho \mapsto \frac{1}{M}\sum_{m=1}^{M} |x_m^k\rangle\langle x_m^k| \otimes \sqrt{D_m^1}\rho\sqrt{D_m^1}, \tag{3.2.47}$$

and

$$\Delta_2 : \mathcal{S}(\mathcal{H}_A^{\otimes k} \otimes \mathcal{H}_C^{\otimes k}) \to \mathcal{S}(\mathcal{H}_B^{\otimes k} \otimes \mathcal{H}_A^{\otimes k} \otimes \mathcal{H}_C^{\otimes k}) \tag{3.2.48}$$

$$: \rho \mapsto \frac{1}{N}\sum_{n=1}^{N} |y_n^k\rangle\langle y_n^k| \otimes \sqrt{D_n^2}\rho\sqrt{D_n^2}, \tag{3.2.49}$$

By Lemma 3.2.1, with the uniform distribution on $[M] \times [N]$, it holds

$$\sqrt{8\epsilon} + \epsilon \geq \frac{1}{M}\sum_{m=1}^{M} \left\| |x_m^k\rangle\langle x_m^k| \otimes V_1^k(x_m^k) - \Delta_1(V_1^k(x_m^k))\right\|_1 \tag{3.2.50}$$

$$= \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} \left\| |x_m^k\rangle\langle x_m^k| \otimes W^k(x_m^k, y_n^k) - \Delta_1\left(W^k(x_m^k, y_n^k)\right)\right\|_1 \tag{3.2.51}$$

where linearity of Δ_1 and the definition of V_1^k is used. Similarly

$$\sqrt{8\epsilon} + \epsilon \geq \frac{1}{N}\sum_{n=1}^{N} \left\| |y_n^k\rangle\langle y_n^k| \otimes V_2^k(y_n^k) - \Delta_2(V_2^k(y_n^k))\right\|_1 \tag{3.2.52}$$

$$= \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} \left\| |y_n^k\rangle\langle y_n^k| \otimes |x_m^k\rangle\langle x_m^k| \otimes W^k(x_m^k, y_n^k) - \Delta_2\left(|x_m^k\rangle\langle x_m^k| \otimes W^k(x_m^k, y_n^k)\right)\right\|_1 \tag{3.2.53}$$

Then,

$$\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} \left\| |y_n^k\rangle\langle y_n^k| \otimes |x_m^k\rangle\langle x_m^k| \otimes W^k(x_m^k, y_n^k) - \Delta_2 \circ \Delta_1\left(W^k(x_m^k, y_n^k)\right)\right\|_1 \tag{3.2.54}$$

$$
\begin{aligned}
&\leq \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} \| \, |y_n^k\rangle\langle y_n^k| \otimes |x_m^k\rangle\langle x_m^k| \otimes W^k(x_m^k, y_n^k) - \Delta_2(|x_m^k\rangle\langle x_m^k| \otimes W^k(x_m^k, y_n^k))\| \\
&\quad + \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} \|\Delta_2(|x_m^k\rangle\langle x_m^k| \otimes W^k(x_m^k, y_n^k)) - \Delta_2 \circ \Delta_1(W^k(x_m^k, y_n^k))\|_1 \qquad (3.2.55)\\
&\leq \sqrt{8\epsilon} + \epsilon + \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} \left\|\Delta_2\left(|x_m^k\rangle\langle x_m^k| \otimes W^k(x_m^k, y_n^k)) - \Delta_1(W^k(x_m^k, y_n^k))\right)\right\|_1 \qquad (3.2.56)\\
&\leq \sqrt{8\epsilon} + \epsilon + \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} \underbrace{\|\Delta_2\|_1}_{\leq 1} \left\||x_m^k\rangle\langle x_m^k| \otimes W^k(x_m^k, y_n^k)) - \Delta_1(W^k(x_m^k, y_n^k))\right\|_1 \qquad (3.2.57)\\
&\leq 2(\sqrt{8\epsilon} + \epsilon). \qquad (3.2.58)
\end{aligned}
$$

The first inequality is by adding and subtracting a term and using a triangle inequality. The Second inequality holds by using the bound from (3.2.53) and linearity of Δ_2. The third inequality holds by a Cauchy-Schwartz inequality, and the last inequality holds by 3.2.51.

Taking $\Delta := \Delta_2 \circ \Delta_1$, it therefore holds

$$
\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} \left\||y_n^k\rangle\langle y_n^k| \otimes |x_m^k\rangle\langle x_m^k| \otimes W^k(x_m^k, y_n^k) - \Delta\left(W^k(x_m^k, y_n^k)\right)\right\|_1 \leq 2(\sqrt{8\epsilon} + \epsilon), \qquad (3.2.59)
$$

Since the partial trace does not increase the trace norm, we construct $\tilde{\Delta} := \operatorname{tr}_{\mathcal{H}_C^{\otimes k}} \circ \Delta$ and maintain that,

$$
\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} \left\||y_n^k\rangle\langle y_n^k| \otimes |x_m^k\rangle\langle x_m^k| - \tilde{\Delta}\left(W^k(x_m^k, y_n^k)\right)\right\|_1 \leq 2(\sqrt{8\epsilon} + \epsilon). \qquad (3.2.60)
$$

Therefore $\tilde{\Delta}$ is a quantum channel that, on average, accurately recovers classical outcomes $(x_m^k, y_n^k) \in C_1 \times C_2$ from $W^k(x_m^k, y_n^k)$ for all m and n. Because of this, it is possible to construct a POVM $(D_{mn})_{m=1,n=1}^{M,N}$ for the observables in which the basis $\left(|y_n^k\rangle \otimes |x_m^k\rangle\right)_{m,n}$ represents. With this, we can construct a (k, M, N)-code $\mathcal{C} := (x_m^k, y_n^k, D_{mn})_{m=1,n=1}^{M,N}$ such that,

$$
\operatorname{tr}\big(D_{mn} W^k(x_m^k, y_n^k)\big) \geq 1 - 2(\sqrt{8\epsilon} + \epsilon). \qquad (3.2.61)
$$

The average error of transmission is then,

$$
\begin{aligned}
\overline{e}(\mathcal{C}, W^k) &= 1 - \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} \operatorname{tr}\big(D_{mn} W^k(x_m^k, y_n^k)\big) \qquad (3.2.62)\\
&\leq 1 - (1 - 2(\sqrt{8\epsilon} + \epsilon)) \qquad (3.2.63)\\
&= 2(\sqrt{8\epsilon} + \epsilon), \qquad (3.2.64)
\end{aligned}
$$

and therefore with the selected codebook, the code achieves the rate (R_1, R_2) as was to show. □

3.3 Review of the Converse Theorem

In this section we review the proof for the converse of Theorem 3.2.5, that is, the containment of $C(W)$ in the region defined in 3.2.5. To show this, we introduce the following lemmas, which will also prove useful in latter sections on this thesis.

Lemma 3.3.1 (Fano's Inequality)**.** *For a quantum channel $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$, let A be a random variable for a set of messages $[M] \subseteq \mathcal{X}$ and let C be the quantum output of W. Let the POVM $\mathbf{D} := \{D_i\}_{i=1}^{M} \subseteq \mathcal{L}(\mathcal{H})$ have one element for each realization of A. Suppose $\hat{A}$ is a random variable for the result of measuring $\mathbf{D}$ on C, then $A \to C \to \hat{A}$ forms a Markov chain. Define the probability of error as $p_e := [A \neq \hat{A}]$. Then, it holds,*

$$H(A|C)_\gamma \leq H(A|\hat{A}) \leq 1 + p_e \log(M), \tag{3.3.1}$$

where γ is the channel state constructed with A. We refer the reader to Appendix B for the definition of condition entropy.

Lemma 3.3.2 (Subadditivity of mutual information)**.** *[6, Lemma 1] For a quantum channel $W : \mathcal{S}(\mathcal{H}_A) \to \mathcal{S}(\mathcal{H}_C)$, for all $k \in \mathbb{N}, k > 0$, and any channel state $\gamma^k \in \mathcal{S}(\mathcal{H}_A^{\otimes k} \otimes \mathcal{H}_C^{\otimes k})$, it holds*

$$I(A^k; C^k)_{\gamma^k} \leq k \cdot I(A; C)_\gamma. \tag{3.3.2}$$

where γ is γ^k restricted to one instance of A and C. We refer the reader to Appendix B for the definition of mutual information.

Theorem 3.3.3. *For a DM-CCQ channel generated by W, let $(R_1, R_2) \in C(W)$, then there exists probability distributions $p_1 \in \mathcal{P}(\mathcal{X})$ and $p_2 \in \mathcal{P}(\mathcal{Y})$ for random variables A and B such that with channel state $\gamma_2 := \gamma_2(p_1, p_2)$,*

$$\begin{aligned} R_1 &\leq I(A; C|B)_{\gamma_2}, && (3.3.3)\\ R_2 &\leq I(B; C|A)_{\gamma_2}, && (3.3.4)\\ R_1 + R_2 &\leq I(A, B; C)_{\gamma_2}. && (3.3.5) \end{aligned}$$

Proof. Let $(R_1, R_2) \in C(W)$ and $\delta, \epsilon > 0$, then there exists a k_0 such that for $k \geq k_0$, $\mathcal{C}$ is a (k, M, N)-code that achieves (R_1, R_2) with $M = \lceil 2^{k(R_1 - \delta)} \rceil$ and $N = \lceil 2^{k(R_2 - \delta)} \rceil$ and $\overline{e}(C, W^k) \leq \epsilon$. Let p_1 be the uniform distribution on the codewords in $[M]$ and p_2 uniform on the codewords in $[N]$. Then, with $k \geq k_0$ and channel states $\gamma_2^k := \gamma_2^k(p_1, p_2)$, by Lemma 3.3.1

$$\begin{aligned} H(A^k \mid C^k, B^k)_{\gamma_2^k} &\leq 1 + \epsilon \cdot \lceil k(R_1 - \delta) \rceil && (3.3.6)\\ H(B^k \mid C^k, A^k)_{\gamma_2^k} &\leq 1 + \epsilon \cdot \lceil k(R_2 - \delta) \rceil && (3.3.7)\\ H(A^k, B^k \mid C^k)_{\gamma_2^k} &\leq 1 + \epsilon \cdot \lceil k(R_1 + R_2 - 2\delta) \rceil. && (3.3.8) \end{aligned}$$

where for the first two inequalities, we use Lemma 3.1.14 to justify that a (k, M, N)-code can be modified to be used as a (k, M) and (k, N) over a DM-CQ channel using receiver Hilbert spaces of dimension $\lceil \log M \rceil$ and $\lceil \log N \rceil$ respectively with average error bound maintained. Then it holds,

$$\begin{aligned} H(A^k \mid C^k, B^k)_{\gamma_2^k} &= H(A^k)_{\gamma_2^k} - I(A^k; B^k, C^k)_{\gamma_2^k} && (3.3.9)\\ &= H(A^k) - I(A^k; B^k, C^k)_{\gamma_2^k} && (3.3.10)\\ &= \lceil k(R_1 - \delta) \rceil - I(A^k; B^k, C^k)_{\gamma_2^k} && (3.3.11) \end{aligned}$$

where the first equality is by definition of the mutual information, the second since γ_2^k is a channel state, and the third since p_A is uniform over $[M]$. Therefore,

$$\begin{aligned} (1 - \epsilon)(R_1 - \delta) &\leq (1 - \epsilon)\lceil (R_1 - \delta) \rceil && (3.3.12)\\ &\leq \frac{1}{k} + \frac{1}{k} I(A^k; B^k, C^k)_{\gamma_2^k} && (3.3.13) \end{aligned}$$

$$\leq \frac{1}{k} + I(A;B,C)_{\gamma_2} \tag{3.3.14}$$

$$= \frac{1}{k} + I(A;C|B)_{\gamma_2} \tag{3.3.15}$$

where we start by combining (3.3.6) and (3.3.11). The third inequality holds by Lemma 3.3.2. The equality holds because it is straight forward to see by the chain rule of mutual information (see Appendix B) $I(A;B,C)_{\gamma_2} = I(A;C|B)_{\gamma_2}$ when A and B are independent and using the definition of conditional entropy. Taking k large enough, and since ϵ and δ are taken arbitrarily, it holds that $R_1 \leq I(A;C|B)_{\gamma_2}$. It can be similarly shown that $R_2 \leq I(B;C|A)_{\gamma_2}$ and $R_1 + R_2 \leq I(A,B;C)_{\gamma_2}$, and therefore the statement holds. □

3.4 Maximal Error Capacity Region

Because we will use codes for a CCQ channels which use a maximal error figure of merit, we will prove some facts about the maximal error capacity region. We aim to show that when the average error capacity region for a CCQ has a non-empty interior, the interior of maximal error capacity region is also non-empty.

Lemma 3.4.1 (Maximal error capacity region is closed)**.** *For CCQ channel* $\mathbf{W}$*,* $C_{max}(\mathbf{W})$ *is closed.*

Proof. Let $(R_{1i}, R_{2i})_{i\in\mathbb{N}}$ be a sequence in $C_{\max}(\mathbf{W})$ that converge to some (R_1, R_2). By the convergence of $(R_{1i}, R_{2i})_{i\in\mathbb{N}}$, there exists an i_0 large enough such that for $\delta > 0$,

$$R_{1i_0} \geq R_1 - \delta/2 \quad \text{and} \quad R_{2i_0} \geq R_2 - \delta/2. \tag{3.4.1}$$

Since (R_{1i_0}, R_{2i_0}) is a maximally achievable rate pair, there exists a k_0 such that for all $k \geq k_0$, there is a (k, M, N)-code $\mathcal{C}$ such that for all $\epsilon > 0$, $e(\mathcal{C}, W^k) \leq \epsilon$ and

$$\frac{1}{k}\log M \geq R_{1i_0} - \delta/2 \geq R_1 - \delta \quad \text{and} \quad \frac{1}{k}\log N \geq R_{2i_0} - \delta/2 \geq R_2 - \delta. \tag{3.4.2}$$

Since δ was chosen arbitrarily, $\mathcal{C}$ also achieves (R_1, R_2) under maximum error criterion, hence $(R_1, R_2) \in C_{\max}(\mathbf{W})$ and therefore $C_{\max}(\mathbf{W})$ is closed. □

Lemma 3.4.2 (Maximal error capacity region is convex)**.** *For CCQ channel* $\mathbf{W}$*,* $C_{max}(\mathbf{W})$ *is convex.*

Proof. Let (R_1, R_2) and (R_1', R_2') be two pairs in $C_{\max}(\mathbf{W})$ and $\epsilon, \delta > 0$, then there exists a k_0 such that for all $k \geq k_0$, (k, M, N)-code $\mathcal{C} := (x_m, y_n, D_{mn})_{m=1,n=1}^{M,N}$ satisfies $e(\mathcal{C}, W^k) \leq \epsilon^{1/2}$, $M \geq 2^{k(R_1-\delta_1)}$, $N \geq 2^{k(R_2-\delta_1)}$, and a k_0' such that for all $k' \geq k_0'$, (k', M', N')-code $\mathcal{C}' := (x_m', y_n', D_{mn}')_{m=1,n=1}^{M',N'}$ satisfies $e(\mathcal{C}', W^{k'}) \leq \epsilon^{1/2}$, $M' \geq 2^{k'(R_1'-\delta_2)}$, $N' \geq 2^{k'(R_2'-\delta_2)}$. We make the choice δ_1 and δ_2 such that the rate calculations below are satisfied. We show that for all $\alpha \in [0,1]$ that $(\alpha R_1 + (1-\alpha)R_1', \alpha R_2 + (1-\alpha)R_2')$ is an achievable rate pair. The strategy is to construct a new code that sends first a message (x_m, y_n) from code $\mathcal{C}$ with k uses of the channel and then a message $(x_{m'}', y_{n'}')$ from $\mathcal{C}'$ with k' uses of the channel. The maximum error of such a code is bounded by,

$$\begin{aligned}\operatorname{tr}\Big(D_{mn} \otimes D_{m'n'}' W^k(x_m, y_n) \otimes W^{k'}(x_{m'}, y_{n'})\Big) = \\ \operatorname{tr}\big(D_{mn} W^k(x_m, y_n)\big) \operatorname{tr}\Big(D_{m'n'}' W^{k'}(x_{m'}, y_{n'})\Big) \leq \epsilon.\end{aligned} \tag{3.4.3}$$

Let $\alpha \in [0,1]$, then

$$\frac{1}{k+k'}\log(M \cdot M') \geq \frac{1}{k+k'}\log\Big(2^{k(R_1-\delta_1)}2^{k'(R_1'-\delta_2)}\Big) \tag{3.4.4}$$

$$= \frac{k}{k+k'}(R_1 - \delta_1) + \frac{k'}{k+k'}(R_1' - \delta_2) \tag{3.4.5}$$

$$= \alpha R_1 + (1-\alpha)R_1' - \alpha(\delta_1 + \delta_2) + \epsilon_e \tag{3.4.6}$$

$$= \alpha R_1 + (1-\alpha)R_1' - \delta \tag{3.4.7}$$

where k and k' can be chosen such that $\frac{k}{k+k'}$ is arbitrarily close to α with ϵ_e the inaccuracy. We can also choose δ_1 and δ_2 such that $\delta = \alpha(\delta_1 + \delta_2) - \epsilon_e$. We can make the same arguments to show that $\frac{1}{k+k'}\log(N \cdot N') \geq \alpha R_2 + (1-\alpha)R_2' - \delta$. Therefore $(\alpha R_1 + (1-\alpha)R_1', \alpha R_2 + (1-\alpha)R_2')$ is achievable and therefore $C_{\max}(\mathbf{W})$ is convex. □

Lemma 3.4.3 (Non-empty maximal error capacity region)**.** *For CCQ channel* $\mathbf{W}$*, if int*$(C(\mathbf{W}))$ *is non-empty, then int*$(C_{\max}(\mathbf{W}))$ *is also non-empty.*

Proof. Assume $\mathrm{int}(C(\mathbf{W})) \neq \emptyset$ and $(R_1, R_2) \in \mathrm{int}(C(\mathbf{W}))$ and so $R_1 \neq 0$ and $R_2 \neq 0$. From an average error code that achieves the rate pairs (R_1, R_2), we construct two codes that maximally achieve $(R_1, 0)$ and $(0, R_2)$ respectively. Since $C_{\max}(\mathbf{W})$ is convex and closed, by a time sharing argument, the interior will thus be non-empty.

By definition, since $(R_1, R_2) \in \mathrm{int}(C(\mathbf{W}))$, there exists a k_0 such that for all $k \geq k_0$, the (k, M, N) code $\mathcal{C} := (x_m, y_n, D_{mn})_{m=1,n=1}^{M,N}$ satisfies, for all $\delta, \epsilon > 0$, $\frac{1}{k}\log M \geq R_1 - \delta$, $\frac{1}{k}\log N \geq R_2 - \delta$, and $\overline{e}(\mathcal{C}, W^k) \leq \epsilon$. Using this code, we construct a code in the following way. Since $\overline{e}(\mathcal{C}, W^k) \leq \epsilon$, we can write,

$$\epsilon \geq 1 - \frac{1}{MN}\sum_{\substack{m=1\\n=1}}^{M,N} \mathrm{tr}\big(D_{mn}W^k(x_m, y_n)\big) = \frac{1}{N}\sum_{n=1}^{N}\left(1 - \frac{1}{M}\sum_{m=1}^{M}\mathrm{tr}\big(D_{mn}W^k(x_m, y_n)\big)\right). \tag{3.4.8}$$

Therefore, there exists at least one index $n_0 \in [N]$ such that

$$1 - \frac{1}{M}\sum_{m}\mathrm{tr}\big(D_{mn_0}W^k(x_m, y_{n_0})\big) \leq \epsilon. \tag{3.4.9}$$

With this n_0, we construct the code $(x_m, y_{n_0}, D_{mn_0})_{m=1}^{M}$ which achieves (under average error) $(R_1, 0)$. Note, the decoders no longer form a POVM, but can be modified to form a POVM via an expurgation as in [5, Ch. 16.5] with negligible effects for large k. We transform this code to a maximally achieve $(R_1, 0)$. Assume without loss of generality that the codewords x_m are ordered such that $\big(1 - \mathrm{tr}\big(D_{mn_0}W^k(x_m, y_{n_0})\big)\big)_{m\in[M]}$ is non-decreasing. For $M' := \lceil 2^{(k/2)(R_1-\delta)}\rceil$, let $\lambda := 1 - \mathrm{tr}\big(D_{M'n_0}W^k(x_{M'}, y_{n_0})\big)$. It holds,

$$\epsilon \geq 1 - \frac{1}{M}\sum_{m=1}^{M}\mathrm{tr}\big(D_{mn_0}W^k(x_m, y_{n_0})\big) \tag{3.4.10}$$

$$\geq \frac{1}{\lceil 2^{k(R_1-\delta)}\rceil}\sum_{m=1}^{M'-1}\underbrace{1 - \mathrm{tr}\big(D_{mn_0}W^k(x_m, y_{n_0})\big)}_{\geq 0} + \frac{1}{\lceil 2^{k(R_1-\delta)}\rceil}\sum_{m=M'}^{\lceil 2^{k(R_1-\delta)}\rceil}\underbrace{1 - \mathrm{tr}\big(D_{mn_0}W^k(x_m, y_{n_0})\big)}_{\geq \lambda} \tag{3.4.11}$$

$$\geq \frac{1}{\lceil 2^{k(R_1-\delta)}\rceil}(\lceil 2^{k/2(R_1-\delta)}\rceil)\lambda \tag{3.4.12}$$

$$= \lambda/2. \tag{3.4.13}$$

Thus $\lambda \leq 2\epsilon$, $M' \to M$ as $k \to \infty$, and $\mathcal{C}' := (x_m, y_{n_0}, D_{mn_0})_{m=1}^{M'}$ achieves maximal error rate $(R_1, 0)$. We can make analogous steps to find another code $\mathcal{C}''$ that achieves maximal error rate $(0, R_2)$. By the convexity of $C_{\max}(\mathbf{W})$, we have that, for all $\alpha \in [0,1], (\alpha R_1, (1-\alpha)R_2)$ is an achievable maximal error rate and thus the interior of $C_{\max}(\mathbf{W})$ is not empty. □

Chapter 4

Identification Capacity of the Quantum Multiple Access Channel

Many ideas of classical Shannon theory can be adapted to the quantum setting. Here we investigate how we can model an identification problem via classical-quantum channels using analogies from the classical setting, motivated by the results of Peter Löber. After we introduce the necessary theory, we prove the main result of the thesis, namely, we determine the simultaneous identification capacity of the classical-quantum multiple access channel.

4.1 Classical-Quantum Identification

In this section we will build the necessary models for identification over a CCQ channel. We begin with the single sender-single receiver case and extend to the multiuser case.

Definition 4.1.1 ((k, M)-ID-code)**.** *For a CQ channel* $\mathbf{W}$*, a randomized* (k, M)*-ID-code is a family* $\mathcal{C}_{id} := (P_m, D_m)_{m=1}^{M}$*, where* $P_1, ..., P_M \in \mathcal{P}(\mathcal{X}^k)$ *are probability distributions, and for each* $m \in [M]$*,* $D_m \in \mathcal{L}(\mathcal{H}^{\otimes k})$*,* $0 \leq D_m \leq \mathbb{1}_{\mathcal{H}^{\otimes k}}$*.*

For an (k, M)*-ID-code* $\mathcal{C}_{id}$ *and CQ channel* $\mathbf{W}$*, we define two types of errors,*

$$e_1(\mathcal{C}_{id}, W^k) := \max_{m \in [M]} 1 - \sum_{x^k \in \mathcal{X}^k} P_m(x^k) \operatorname{tr}\left(D_m W^k(x^k)\right), \tag{4.1.1}$$

$$e_2(\mathcal{C}_{id}, W^k) := \max_{\substack{m,n \in [M] \\ m \neq n}} \sum_{x^k \in \mathcal{X}^k} P_m(x^k) \operatorname{tr}\big(D_n W^k(x^k)\big). \tag{4.1.2}$$

In this definition, we, similar to the classical definition, use two types of errors with analogous interpretations. The first type of error provides a figure of merit for how accurate the encoding and identifying procedure is. For the second, an important thing to notice is that the decoding operators need not form a POVM. The implications of this are analogous to that of classical ID codes not requiring mutual disjointness amongst the identifying sets, that is, multiple events can be associated to a single identifying operator. Therefore, because the code is randomized, the second type of error provides a figure of merit for the precision of the identifying operators and encoders.

Another difference is that, unlike classical measurements, performing a measurement on quantum data will destroy the quantumness. What this implies is that for the identification task, the receiver cannot use the same state to identify two different messages since a measurement must be made. Further, if the first receiver is not the final receiver in the communication chain, an identifying measurement would ruin the state for the next receiver. To overcome these problems, Löber introduced the idea of a simultaneous ID-code. This more restrictive model has the receivers identify all possible messages simultaneously with a single measurement.

Definition 4.1.2 (Simultaneous (k, M)-ID-code)**.** *A* (k, M)*-ID-code* $\mathcal{C}_{id} = (P_m, D_m)_{m=1}^{M}$ *is called simultaneous if for* $R \in \mathbb{N}$ *there exists a POVM* $(E_r)_{r=1}^{R}$ *and subsets* $A_1, ..., A_M \subset [R]$ *such that* *for each $m \in [M]$*

$$D_m = \sum_{r \in A_m} E_r. \tag{4.1.3}$$

We can extend these ideas to the multiple access channel.

Definition 4.1.3 ((k, M, N)-ID-code)**.** *For a a CCQ channel $\mathbf{W}$, a (k, M, N)-ID-code for classical message identification is the family $\mathcal{C}_{id} := (P_m, Q_n, D_{mn})_{m=1,n=1}^{M,N}$ where $P_1, ..., P_M \in \mathcal{P}(\mathcal{X}^k)$, $Q_1, ..., Q_N \in \mathcal{P}(\mathcal{Y}^k)$, and $(D_{mn})_{m=1,n=1}^{M,N} \subseteq \mathcal{L}(\mathcal{H}^{\otimes k})$ such that $0 \leq D_{mn} \leq \mathbb{1}_{\mathcal{H}^{\otimes k}}$ for all $m \in [M]$ and $n \in [N]$.*

For a (k, M, N)-ID-code $\mathcal{C}_{id}$, we define two types of errors,

$$e_1(C_{id}, W^k) := \max_{m \in [M], n \in [N]} \quad 1 - \sum_{x^k \in \mathcal{X}^k} \sum_{y^k \in \mathcal{Y}^k} P_m(x^k) Q_n(y^k) \operatorname{tr}\left(D_{mn} W^k(x^k, y^k)\right), \tag{4.1.4}$$

$$e_2(C_{id}, W^k) := \max_{\substack{m,m' \in [M], n,n' \in [N] \\ (m,n) \neq (m',n')}} \quad \sum_{x^k \in \mathcal{X}^k} \sum_{y^k \in \mathcal{Y}^k} P_m(x^k) Q_n(y^k) \operatorname{tr}\left(D_{m'n'} W^k(x^k, y^k)\right). \tag{4.1.5}$$

Definition 4.1.4 (Simultaneous (k, M, N)-ID-code)**.** *A (k, M, N)-ID-code $\mathcal{C}_{id} := (P_m, Q_n, D_{mn})_{m=1,n=1}^{M,N}$ is called simultaneous if for $R, S \in \mathbb{N}$ there exists a POVM $(E_{rs})_{r=1,s=1}^{R,S}$ with subsets $A_1, ..., A_M \subset [R]$ and $B_1, ..., B_N \subset [S]$ such that for each $m \in [M]$ and $n \in [N]$,*

$$D_{mn} = \sum_{i \in A_m} \sum_{j \in B_n} E_{ij}. \tag{4.1.6}$$

As with the transmission capacity problem, we are interested in the region of achievable ID-rates. We define the analogous definitions for identification over a CCQ channel.

Definition 4.1.5 (Achievable ID-rate pair)**.** *For a CCQ channel $\mathbf{W}$, we say $(R_1, R_2) \in \mathbb{R}^2$, $R_1, R_2 \geq 0$, is an achievable ID-rate pair if for all $\epsilon_1, \epsilon_2, \delta > 0$, there exists a k_0 such that for all $k \geq k_0$, there is a (k, M, N)-ID-code $\mathcal{C}_{id}$ with*

$$\frac{1}{k} \log\log M \geq R_1 - \delta, \quad \frac{1}{k} \log\log N \geq R_2 - \delta, \quad e_1(\mathcal{C}_{id}, W^k) \leq \epsilon_1, \quad e_2(\mathcal{C}_{id}, W^k) \leq \epsilon_2. \tag{4.1.7}$$

The ID capacity region of $\mathbf{W}$ is defined as

$$C_{id}(\mathbf{W}) := \{(R_1, R_2) \mid (R_1, R_2) \text{ is an achievable ID-rate pair}\}. \tag{4.1.8}$$

Definition 4.1.6 (Achievable simultaneous ID-rate pair)**.** *For a CCQ channel $\mathbf{W}$, we say $(R_1, R_2) \in \mathbb{R}^2$, $R_1, R_2 \geq 0$, we say (R_1, R_2) is an achievable simultaneous ID-rate pair if for $\epsilon_1, \epsilon_2, \delta > 0$, there exists a k_0 such that for all $k \geq k_0$ there is a simultaneous (k, M, N)-ID-code $\mathcal{C}_{id}^{sim}$ with*

$$\frac{1}{k} \log\log M \geq R_1 - \delta, \quad \frac{1}{k} \log\log N \geq R_2 - \delta, \quad e_1(\mathcal{C}_{id}^{sim}, W^k) \leq \epsilon_1, \quad e_2(\mathcal{C}_{id}^{sim}, W^k) \leq \epsilon_2. \tag{4.1.9}$$

The simultaneous ID capacity region for a CCQ channel $\mathbf{W}$ is defined as

$$C_{id}^{sim}(\mathbf{W}) := \{(R_1, R_2) \mid (R_1, R_2) \text{ is an achievable simultaneous ID-rate pair}\}. \tag{4.1.10}$$

We say a simultaneous ID-code $\mathcal{C}_{id}^{sim}$ simultaneously achieves (R_1, R_2) if there exists a k_0 such that for all $k \geq k_0$, conditions (4.1.9) hold.

Remark 4.1.7. *Since the simultaneous case is more restrictive, it is clear that*

$$C_{id}^{sim}(\mathbf{W}) \subseteq C_{id}(\mathbf{W}). \tag{4.1.11}$$

For the remainder of this chapter, we aim to prove the following capacity theorem.

Theorem 4.1.8. *For a DM-CCQ channel generated by $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{S}(\mathcal{H})$,*

$$C(W) = C_{id}^{sim}(W) \tag{4.1.12}$$

4.2 Achievability Theorem

Here we prove the achievability direction of the theorem, that is, we show that there exists a simultaneous ID-code that simultaneously ID-achieve an arbitrary achievable rate from $C(W)$. To do this, we use the strategy from the classical case and derive a Transformator-like lemma for DM-CCQ channels.

Theorem 4.2.1. *For a DM-CCQ channel generated by $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{S}(\mathcal{H})$,*

$$C(W) \subseteq C_{id}^{sim}(W). \tag{4.2.1}$$

4.2.1 Transformator Lemma for DM-CCQ Channels

Lemma 4.2.2 (Transformator lemma for a DM-CCQ channel)**.** *For the DM-CCQ channel generated by $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{S}(\mathcal{H})$ with an achievable rate $(R_1, R_2) \in C(W)$, there is a simultaneous (n, M', N')-ID-code $\mathcal{C}_{id}^{sim}$ that achieves the simultaneous ID-rate (R_1, R_2).*

Proof. Assume for DM-CCQ generated by W that $\operatorname{int}(C(W)) \neq \emptyset$, otherwise the problem is reduced to the single sender case, a problem solved in [4]. For a rate pair $(R_1, R_2) \in C(W)$, by definition, there exists a k_0 such that for all $k \geq k_0$, there is a (k, M', N')-code $\mathcal{C}' := (u_i', v_j', D_{ij}')_{i=1,j=1}^{M',N'}$ that achieves (R_1, R_2) with $\overline{e}(\mathcal{C}', W^k) \leq \lambda(k)$, where $\lambda(k) \to 0$ as $k \to \infty$. Further, since $\operatorname{int}(C(W))$ is non-empty, by Lemma 3.4.3, there is a non-trivial achievable maximal error rate pair $(\epsilon_1, \epsilon_2) \in C_{\max}(W)$ and thus a k_0' such that for $k' \geq k_0'$, there is a $(\lceil\sqrt{k}\rceil, M'', N'')$-code $\mathcal{C}'' := (u_i'', v_j'', D_{ij}'')_{i=1,j=1}^{M'',N''}$, $M'' = 2^{\lceil\sqrt{k}\rceil\epsilon_1}$ and $N'' = 2^{\lceil\sqrt{k}\rceil\epsilon_2}$, with $e(\mathcal{C}'', W^{\lceil\sqrt{k}\rceil}) \leq \lambda(\sqrt{k})$, with $\lambda(\sqrt{k}) \to 0$ as $k \to \infty$, where the assumption that $\lceil\sqrt{k}\rceil \geq k_0'$ is made. We define $m := k + \lceil\sqrt{k}\rceil$ and two families of maps

$$\mathcal{A} := (A_i : [M'] \to [M''])_{i=1}^{M} \tag{4.2.2}$$

$$\mathcal{B} := (B_j : [N'] \to [N''])_{j=1}^{N}. \tag{4.2.3}$$

With these families of maps, we define an (m, M, N)-ID-code $(P_i, Q_j, D_{ij})_{i=1,j=1}^{M,N}$ where

$$P_i(x^m) := \begin{cases} \frac{1}{M'} & \text{if } \exists a \in [M'] : x^m = u_a' \cdot u''_{A_i(a)} \\ 0 & \text{otherwise} \end{cases}, \tag{4.2.4}$$

and

$$Q_j(y^m) := \begin{cases} \frac{1}{N'} & \text{if } \exists b \in [N'] : y^m = v_b' \cdot v''_{B_j(b)} \\ 0 & \text{otherwise} \end{cases}. \tag{4.2.5}$$

The decoders are defined as

$$D_{ij} := \sum_{a=1}^{M'} \sum_{b=1}^{N'} D'_{ab} \otimes D''_{A_i(a)B_j(b)}. \tag{4.2.6}$$

We show that with this structure, there exists a random construction of $\mathcal{A}$ and $\mathcal{B}$ such that there is a simultaneous ID-code which achieves the simultaneous ID-rate pair (R_1, R_2).

For $a \in [M']$ and $b \in [N']$ define the random variables U_a such that

$$\Pr(U_a = u_a' \cdot u_c'') = \frac{1}{M''} \tag{4.2.7}$$

with $c \in [M'']$, and V_b such that

$$\Pr(V_b = v_b' \cdot v_d'') = \frac{1}{N''} \tag{4.2.8}$$

with $d \in [N'']$. For $(i,j) \in [M] \times [N]$, define

$$\overline{\mathcal{U}}_i := \{U_a\}_{a=1}^{M'} \tag{4.2.9}$$

$$\overline{\mathcal{V}}_j := \{V_b\}_{b=1}^{N'}, \tag{4.2.10}$$

Let $\overline{P}_i$ and $\overline{Q}_j$ be the uniform distributions on $\overline{\mathcal{U}}_i$ and $\overline{\mathcal{V}}_j$ respectively. Define the random decoder

$$\mathcal{D}(\overline{\mathcal{U}}_i, \overline{\mathcal{V}}_j) := \sum_{a=1}^{M'} \sum_{b=1}^{N'} D(U_a, V_b) \tag{4.2.11}$$

where $D(U_a, V_b) := D'_{ab} \otimes D''_{p_a q_b}$ when $U_a = u'_a \cdot u''_{p_a}$ and $V_b = v'_b \cdot v''_{q_b}$. With this, we can construct a random ID-code $(\overline{P}_i, \overline{Q}_j, \mathcal{D}(\overline{\mathcal{U}}_i, \overline{\mathcal{V}}_j)_{i=1,j=1}^{M,N}$.

We analyze the errors of realizations of the random code. Let $r := \lceil \sqrt{k} \rceil$ and fix any two realizations $\mathcal{U}_i$ of $\overline{\mathcal{U}}_i$ and $\mathcal{V}_j$ of $\overline{\mathcal{V}}_j$.

$$1 - \sum_{\substack{x^m \in \mathcal{X}^m \\ y^m \in \mathcal{Y}^m}} P_i(x^m) Q_j(y^m) \operatorname{tr}(\mathcal{D}(\mathcal{U}_i, \mathcal{V}_j) W^m(x^m, y^m)) \tag{4.2.12}$$

$$= 1 - \frac{1}{M'N'} \sum_{\substack{u \in \mathcal{U}_i \\ v \in \mathcal{V}_j}} \operatorname{tr}(\mathcal{D}(\mathcal{U}_i, \mathcal{V}_j) W^m(u, v)) \tag{4.2.13}$$

$$= 1 - \frac{1}{M'N'} \sum_{\substack{u'_i \cdot u''_{p_i} \in \mathcal{U}_i \\ v'_j \cdot v''_{q_j} \in \mathcal{V}_j}} \operatorname{tr}\left(\sum_{\substack{i'=1 \\ j'=1}}^{M',N'} (D'_{i'j'} \otimes D''_{p_{i'} q_{j'}}) W^m(u'_i \cdot u''_{p_i}, v'_j \cdot v''_{q_j}) \right) \tag{4.2.14}$$

$$\leq 1 - \frac{1}{M'N'} \sum_{\substack{u'_i \cdot u''_{p_i} \in \mathcal{U}_i \\ v'_j \cdot v''_{q_j} \in \mathcal{V}_j}} \operatorname{tr}\Big((D'_{ij} \otimes D''_{p_i q_i}) W^m(u'_i \cdot u''_{p_i}, v'_j \cdot v''_{q_j}) \Big) \tag{4.2.15}$$

$$= 1 - \underbrace{\frac{1}{M'N'} \sum_{\substack{u'_i \cdot u''_{p_i} \in \mathcal{U}_i \\ v'_j \cdot v''_{q_j} \in \mathcal{V}_j}} \operatorname{tr}\big(D'_{ij} W^k(u'_i, v'_j) \big)}_{> 1 - \lambda(k)} \cdot \underbrace{\operatorname{tr}\Big(D''_{p_i q_j} W^r(u''_{p_i}, v''_{q_j}) \Big)}_{> 1 - \lambda(\sqrt{k})} \tag{4.2.16}$$

$$< \lambda(k) + \lambda(\sqrt{k}). \tag{4.2.17}$$

We start with the definition of the first kind error. The first equality is by replacing P_i and Q_j with their definitions, that is, P_i is the uniform distribution on $\mathcal{U}_i$ and $\mathcal{V}_j$ is the uniform distribution on $\mathcal{V}_j$. The second equality is from replacing $\mathcal{D}(\mathcal{U}_i, \mathcal{V}_j)$ with its definition. The inequality holds since each operator in the trace is positive, if we were to expand using linearity of the trace, we would remove fewer positive values from 1 giving the upper bound. The third equality holds from using the memorylessness of the CCQ channel. By splitting the channel into two components and use properties of the trace and tensor products we get the equality. Expanding the product $(1 - \lambda(\sqrt{k}))(1 - \lambda(\sqrt{k}))$, we get the final inequality.

For the second type of error, we can start by considering the error between realizations $\mathcal{U}_1$ of $\overline{\mathcal{U}}_1$ and $\mathcal{V}_1$ of $\overline{\mathcal{V}}_1$ and random sets $\overline{\mathcal{U}}_2$ and $\overline{\mathcal{V}}_2$. We define two random functions, with the ith element of a realization of $\overline{\mathcal{U}}_2$ denoted U_i^2 and similarly the jth element of $\overline{\mathcal{V}}_2$ as V_j^2,

$$\psi_i(\overline{\mathcal{U}}_2) := \begin{cases} 1, & \text{if } U_i^2 \in \mathcal{U}_1 \\ 0, & \text{otherwise} \end{cases}, \quad \text{and} \quad \phi_j(\overline{\mathcal{V}}_2) := \begin{cases} 1, & \text{if } V_j^2 \in \mathcal{V}_1 \\ 0, & \text{otherwise} \end{cases}. \tag{4.2.18}$$

Note that because each U_i^2 is independent of the other elements of $\overline{\mathcal{U}}_2$ and each V_j^2 is independent of the other elements of $\overline{\mathcal{V}}_2$, for $i \neq j$, $\psi_i(\overline{\mathcal{U}}_2)$ is independent of $\psi_j(\overline{\mathcal{U}}_2)$ and $\phi_j(\overline{\mathcal{V}}_2)$ is independent of $\phi_i(\overline{\mathcal{V}}_2)$. Further, it is easy to see

$$\mathbb{E}(\psi_i(\overline{\mathcal{U}}_2)) = \frac{1}{M''} \quad \text{and} \quad \mathbb{E}(\phi_j(\overline{\mathcal{V}}_2)) = \frac{1}{N''}, \quad \forall i \in [M'], j \in [N'], \tag{4.2.19}$$

since the $\psi_i(\overline{\mathcal{U}}_2) = 1$ when the ending of U_i^2 is equal to the ending of the ith element of $\mathcal{U}_1$ which occurs with $1/M''$ chance, and the analogous for $\phi_j(\overline{\mathcal{V}}_2)$. For $\lambda \in (0,1)$ and that $\epsilon_1 = \log(M'')/r$, and $\epsilon_2 = \log(N'')/r$,

$$
\begin{aligned}
D\left(\lambda \parallel \frac{1}{M''}\right) &= \lambda \log(\lambda 2^{r\epsilon_1}) + (1-\lambda)\log\left(\frac{1-\lambda}{1-2^{-r\epsilon_1}}\right) && (4.2.20)\\
&= \lambda\log(\lambda) + \lambda\log(2^{r\epsilon_1}) + (1-\lambda)\log(1-\lambda) \\
&\quad - (1-\lambda)\log\bigl(1-2^{-r\epsilon_1}\bigr) && (4.2.21)\\
&\geq \lambda\log(2^{r\epsilon_1}) + \log(0.5) && (4.2.22)\\
&\geq \lambda\sqrt{k}\epsilon_1 - 1, && (4.2.23)
\end{aligned}
$$

where we use the fact that $\lambda = 0.5$ minimizes $\lambda\log(\lambda) + (1-\lambda)\log(1-\lambda)$. Similarly,

$$
D\left(\lambda \parallel \frac{1}{N''}\right) \geq \lambda\sqrt{k}\epsilon_2 - 1. \tag{4.2.24}
$$

Moreover, for any two realizations $\mathcal{U}_2$ and $\mathcal{V}_2$ when

$$
(u_i \cdot u_{p_i} \notin \mathcal{U}_2 \text{ or } v_j \cdot v_{q_j} \notin \mathcal{V}_2) \text{ and } (u_i \cdot u_{p_i} \in \mathcal{U}_1 \text{ and } v_j \cdot v_{q_j} \in \mathcal{V}_1) \tag{4.2.25}
$$

is true, it holds that,

$$
\begin{aligned}
\mathcal{D}(\mathcal{U}_2, \mathcal{V}_2) &= \sum_{\substack{i'=1\\ j'=1}}^{M',N'} D'_{i'j'} \otimes D''_{p_{i'}q_{j'}} && (4.2.26)\\
&\leq \sum_{\substack{i'=1\\ j'=1}}^{M',N'} D'_{i'j'} \otimes \left(\mathbb{1}_{\mathcal{H}^{\otimes r}} - D''_{p_iq_j}\right) && (4.2.27)\\
&= D'_{ij} \otimes \left(\mathbb{1}_{\mathcal{H}^{\otimes r}} - D''_{p_iq_j}\right) + \sum_{\substack{i'=1, i'\neq i\\ j'=1, j'\neq j}} D'_{i'j'} \otimes \left(\mathbb{1}_{\mathcal{H}^{\otimes r}} - D''_{p_iq_j}\right) && (4.2.28)\\
&= D'_{ij} \otimes \left(\mathbb{1}_{\mathcal{H}^{\otimes r}} - D''_{p_iq_j}\right) + \bigl(\mathbb{1}_{\mathcal{H}^{\otimes k}} - D'_{ij}\bigr) \otimes \left(\mathbb{1}_{\mathcal{H}^{\otimes r}} - D''_{p_iq_j}\right), && (4.2.29)
\end{aligned}
$$

where the first inequality is true because when condition (4.2.25) holds, $D''_{p_{i'}q_{j'}}$ will never equal $D''_{p_iq_j}$. Since the $D''_{p_iq_j}$s form a POVM, the first inequality holds. The last equality holds by using the fact that the D'_{ij}s also forms a POVM. Therefore,

$$
\begin{aligned}
&\frac{1}{M'N'} \sum_{\substack{u \in \mathcal{U}_1 \setminus \mathcal{U}_2\\ v \in \mathcal{V}_1 \setminus \mathcal{V}_2}} \operatorname{tr}(\mathcal{D}(\mathcal{U}_2, \mathcal{V}_2) W^m(u,v)) && (4.2.30)\\
&\leq \frac{1}{M'N'} \sum_{\substack{u \in \mathcal{U}_1\\ v \in \mathcal{V}_1}} \operatorname{tr}(\mathcal{D}(\mathcal{U}_2, \mathcal{V}_2) W^m(u,v)) && (4.2.31)\\
&\leq \frac{1}{M'N'} \sum_{\substack{u'_i \cdot u''_{p_i} \in \mathcal{U}_1\\ v'_j \cdot v''_{q_j} \in \mathcal{V}_1}} \operatorname{tr}\left[\left(D'_{ij} \otimes \left(\mathbb{1}_{\mathcal{H}^{\otimes r}} - D''_{p_iq_j}\right)\right) W^m(u'_i \cdot u''_{p_i}, v'_j \cdot v''_{q_j})\right. \\
&\qquad \left. + \left(\bigl(\mathbb{1}_{\mathcal{H}^{\otimes k}} - D'_{ij}\bigr) \otimes \left(\mathbb{1}_{\mathcal{H}^{\otimes r}} - D''_{p_iq_j}\right)\right) W^m(u'_i \cdot u''_{p_i}, v'_j \cdot v''_{q_j})\right] && (4.2.32)
\end{aligned}
$$

$$\begin{aligned} &= \frac{1}{M'N'} \sum_{\substack{u'_i \cdot u''_{p_i} \in \mathcal{U}_1 \\ v'_j \cdot v''_{q_j} \in \mathcal{V}_1}} \underbrace{\operatorname{tr}\big(D'_{ij} W^k(u'_i, v'_j)\big)}_{\leq 1} \underbrace{\operatorname{tr}\Big((\mathbb{1}_{\mathcal{H}^{\otimes r}} - D''_{p_i q_j}) W^r(u''_{p_i}, v''_{q_j})\Big)}_{\leq \lambda(\sqrt{k})} \\ &\quad + \underbrace{\frac{1}{M'N'} \sum_{\substack{u'_i \cdot u''_{p_i} \in \mathcal{U}_1 \\ v'_j \cdot v''_{q_j} \in \mathcal{V}_1}} \operatorname{tr}\big((\mathbb{1}_{\mathcal{H}^{\otimes k}} - D'_{ij}) W^k(u'_i, v'_j)\big)}_{\leq \lambda(k)} \underbrace{\operatorname{tr}\Big((\mathbb{1}_{\mathcal{H}^{\otimes r}} - D''_{p_i q_j}) W^r(u''_{p_i}, v''_{q_j})\Big)}_{\lambda(\sqrt{k})} \end{aligned} \tag{4.2.33}$$

$$\leq \lambda(\sqrt{k}) + \lambda(k)\lambda(\sqrt{k}) \tag{4.2.34}$$

$$=: \lambda_k, \tag{4.2.35}$$

The first inequality holds by including more positive elements in the sum, that is, the set subtraction is disregarded. The second inequality follows because condition (4.2.25) holds and the operator inequality above also holds. The equality holds by using the memorylessness of W and then using properties of the the tensor product and trace operator. For the two other cases when (4.2.25) holds, the same logical steps apply and the inequality holds. With this, we have,

$$\sum_{\substack{x^m \in \mathcal{X}^m \\ y^m \in \mathcal{Y}^m}} P_1(x^m) Q_1(Y^m) \operatorname{tr}\big(\mathcal{D}(\overline{\mathcal{U}}_2, \overline{\mathcal{V}}_2) W^m(x^m, y^m)\big) \tag{4.2.36}$$

$$= \frac{1}{M'N'} \sum_{\substack{u \in \mathcal{U}_1 \\ v \in \mathcal{V}_1}} \operatorname{tr}\big(\mathcal{D}(\overline{\mathcal{U}}_2, \overline{\mathcal{V}}_2) W^m(u, v)\big) \tag{4.2.37}$$

$$\begin{aligned} &= \frac{1}{M'N'} \sum_{\substack{u \in \mathcal{U}_1 \cap \overline{\mathcal{U}}_2 \\ v \in \mathcal{V}_1 \cap \overline{\mathcal{V}}_2}} \underbrace{\operatorname{tr}\big(\mathcal{D}(\overline{\mathcal{U}}_2, \overline{\mathcal{V}}_2) W^m(u, v)\big)}_{\leq 1} + \frac{1}{M'N'} \sum_{\substack{u \in \mathcal{U}_1 \setminus \overline{\mathcal{U}}_2 \\ v \in \mathcal{V}_1 \setminus \overline{\mathcal{V}}_2}} \operatorname{tr}\big(\mathcal{D}(\overline{\mathcal{U}}_2, \overline{\mathcal{V}}_2) W^m(u, v)\big) \\ &\quad + \frac{1}{M'N'} \sum_{\substack{u \in \mathcal{U}_1 \setminus \overline{\mathcal{U}}_2 \\ v \in \mathcal{V}_1 \cap \overline{\mathcal{V}}_2}} \operatorname{tr}\big(\mathcal{D}(\overline{\mathcal{U}}_2, \overline{\mathcal{V}}_2) W^m(u, v)\big) + \frac{1}{M'N'} \sum_{\substack{u \in \mathcal{U}_1 \cap \overline{\mathcal{U}}_2 \\ v \in \mathcal{V}_1 \setminus \overline{\mathcal{V}}_2}} \operatorname{tr}\big(\mathcal{D}(\overline{\mathcal{U}}_2, \overline{\mathcal{V}}_2) W^m(u, v)\big) \end{aligned} \tag{4.2.38}$$

$$\leq \frac{1}{M'N'} (|\mathcal{U}_1 \cap \overline{\mathcal{U}}_2| \cdot |\mathcal{V}_1 \cap \overline{\mathcal{V}}_2|) + 3\lambda_k \tag{4.2.39}$$

$$= \frac{1}{M'N'} \left(\sum_{i=1}^{M'} \psi_i(\overline{\mathcal{U}}_2) \cdot \sum_{j=1}^{N'} \phi_j(\overline{\mathcal{V}}_2) \right) + 3\lambda_k. \tag{4.2.40}$$

The second equality holds by splitting the sum into four components. The inequality holds by using the bound from above for each piece of the sum which satisfy (4.2.25) and using that the number of elements in the first sum is bounded by the magnitude of the set in which the sum is taken over. Now, using Lemma 2.1.3 twice, we have that for $\lambda \in (0, 1)$ and k large enough such that both $1/M'' < \lambda$ and $1/N'' < \lambda$, with non-zero probability, it holds that

$$\frac{1}{M'} \sum_{i=1}^{M'} \psi_i(\overline{\mathcal{U}}_2) < \lambda \quad \text{and} \quad \frac{1}{N'} \sum_{j=1}^{N'} \phi_j(\overline{\mathcal{V}}_2) < \lambda. \tag{4.2.41}$$

Therefore with non-zero probability,

$$\frac{1}{M'N'} \left(\sum_{i=1}^{M'} \psi_i(\overline{\mathcal{U}}_2) \cdot \sum_{j=1}^{N'} \phi_j(\overline{\mathcal{V}}_2) \right) + 3\lambda_k \leq \lambda^2 + 3\lambda_k, \tag{4.2.42}$$

which implies

$$\sum_{\substack{x^m \in \mathcal{X}^m \\ y^m \in \mathcal{Y}^m}} P_1(x^m) Q_1(y^m) \operatorname{tr}\big(\mathcal{D}(\overline{\mathcal{U}}_2, \overline{\mathcal{V}}_2) W^m(x^m, y^m)\big) \leq \lambda^2 + 3\lambda_k. \tag{4.2.43}$$

Similar arguments can be made to show that,

$$\sum_{\substack{x^m\in\mathcal{X}^m\\ y^m\in\mathcal{Y}^m}} \overline{P}_2(x^m)\overline{Q}_2(y^m)\operatorname{tr}(\mathcal{D}(\mathcal{U}_1,\mathcal{V}_1)W^m(x^m,y^m)) \le \lambda^2+3\lambda_k \tag{4.2.44}$$

$$\sum_{\substack{x^m\in\mathcal{X}^m\\ y^m\in\mathcal{Y}^m}} \overline{P}_2(x^m)Q_1(y^m)\operatorname{tr}\big(\mathcal{D}(\mathcal{U}_1,\overline{\mathcal{V}}_2)W^m(x^m,y^m)\big) \le \lambda^2+3\lambda_k \tag{4.2.45}$$

$$\sum_{\substack{x^m\in\mathcal{X}^m\\ y^m\in\mathcal{Y}^m}} P_1(x^m)\overline{Q}_2(y^m)\operatorname{tr}\big(\mathcal{D}(\overline{\mathcal{U}}_2,\mathcal{V}_1)W^m(x^m,y^m)\big) \le \lambda^2+3\lambda_k. \tag{4.2.46}$$

Hence there exists a realizations of $\overline{\mathcal{U}}_2$, $\mathcal{U}_2$ and $\overline{\mathcal{V}}_2$, $\mathcal{V}_2$ such that (4.2.43), (4.2.44), (4.2.45), and (4.2.46) are satisfied. We follow the argumentation of [1]. With $\mathcal{U}_1$, $\mathcal{U}_2$, $\mathcal{V}_1$ and $\mathcal{V}_2$, with positive probability, it holds that $|\mathcal{U}_1\cap\mathcal{U}_2|\le\lambda M'$ and $|\mathcal{V}_1\cap\mathcal{V}_2|\le\lambda N'$. we would like to add two elements $\mathcal{U}_3$ and $\mathcal{V}_3$ such that $|\mathcal{U}_1\cap\mathcal{U}_3|\le\lambda M'$, $|\mathcal{U}_2\cap\mathcal{U}_3|\le\lambda M'$, $|\mathcal{V}_1\cap\mathcal{V}_3|\le\lambda N'$, and $|\mathcal{V}_2\cap\mathcal{V}_3|\le\lambda N'$, which implies that the second kind errors will hold for all elements in the code. We bound probability that such a $\mathcal{U}_3$ and $\mathcal{V}_3$ do not simultaneously exist with,

$$2\cdot\Pr\left(\sum_{k=1}^{M'}\psi_k(\overline{\mathcal{U}}_3)>\lambda M'\right)+2\cdot\Pr\left(\sum_{l=1}^{N'}\phi_l(\overline{\mathcal{V}}_3)>\lambda N'\right)<1, \tag{4.2.47}$$

where the analogous of ψ_k and ϕ_k are defined. We repeat the argument for $i=4,...,M$ and $j=4,...,N$ and it should hold for existence that,

$$(M-1)\cdot\Pr\left(\sum_{k=1}^{M'}\psi_k(\overline{\mathcal{U}}_i)>\lambda M'\right)+(N-1)\cdot\Pr\left(\sum_{l=1}^{N'}\phi_l(\overline{\mathcal{V}}_j)>\lambda N'\right)<1, \tag{4.2.48}$$

We can therefore enforce that for all $i=2,...,M$,

$$(M-1)\cdot\Pr\left(\sum_{k=1}^{M'}\psi_k(\overline{\mathcal{U}}_i)>\lambda M'\right)<\frac{1}{2}, \tag{4.2.49}$$

and all $j=2,...,N$ that,

$$(N-1)\cdot\Pr\left(\sum_{l=1}^{N'}\phi_l(\overline{\mathcal{V}}_j)>\lambda N'\right)<\frac{1}{2}. \tag{4.2.50}$$

Thus it should hold that, for all $i=2,...,M$ and $j=2,...,N$,

$$\Pr\left(\sum_{k=1}^{M'}\psi_k(\overline{\mathcal{U}}_i)>\lambda M'\right)\le 2^{-M'(\lambda\sqrt{k}\epsilon_1-1)}<\frac{1}{2(M-1)}, \tag{4.2.51}$$

and

$$\Pr\left(\sum_{l=1}^{N'}\phi_l(\overline{\mathcal{V}}_j)>\lambda N'\right)\le 2^{-N'(\lambda\sqrt{k}\epsilon_2-1)}<\frac{1}{2(N-1)}. \tag{4.2.52}$$

With the choices

$$M\le 2^{(2^{k(R_1-\delta)}(\lambda\sqrt{k}\epsilon_1-1)-1)/2} \quad\text{and}\quad N\le 2^{(2^{k(R_2-\delta)}(\lambda\sqrt{k}\epsilon_2-1)-1)/2}, \tag{4.2.53}$$

there exists realizations of $(\overline{\mathcal{U}}_1,...,\overline{\mathcal{U}}_M)$, and $(\overline{\mathcal{V}}_1,...,\overline{\mathcal{V}}_N)$ such that with the family of maps $\mathcal{A}$ and $\mathcal{B}$ constructed from them, there is a simultaneous (m,M,N)-ID-code $\mathcal{C}_{\text{id}}^{\text{sim}}$ that achieves the simultaneous ID-rate (R_1,R_2) as $k\to\infty$. In the limit as $k\to\infty$, $m=k$ and thus the $\mathcal{C}_{\text{id}}^{\text{sim}}$ is a simultaneous (k,M,N)-ID-code. □

4.2.2 Proof of the Achievability Theorem

With the Transformator lemma for DM-CCQ channels, it is straight forward to prove the theorem.

Proof of Theorem 4.2.1. Let $(R_1, R_2) \in C(W)$. By definition, there exists a code $\mathcal{C}$ that achieves it. By the transformator lemma, Lemma 4.2.2, there exists a simultaneous ID code $\mathcal{C}_{\mathrm{id}}$ that simultaneous ID-achieves (R_1, R_2). Therefore, $(R_1, R_2) \in C_{\mathrm{id}}^{\mathrm{sim}}(W)$. □

Corollary 4.2.3. *For a DM-CCQ channel generated by $W : \mathcal{X} \times \mathcal{Y} \to \mathcal{S}(\mathcal{H})$,*

$$C(W) \subseteq C_{id}(W). \tag{4.2.54}$$

Proof. For a DM-CCQ generated by W, by Theorem 4.2.1, $C(W) \subseteq C_{\mathrm{id}}^{\mathrm{sim}}(W)$ By Remark 4.1.7, we have that $C_{\mathrm{id}}^{\mathrm{sim}}(W) \subseteq C_{\mathrm{id}}(W)$. Combining these results, it holds that $C(W) \subseteq C_{\mathrm{id}}(W)$. □

4.3 Converse Theorem

In this section we prove the converse theorem. After reviewing the necessary theory, we introduce a multi-letter capacity region for the CCQ channel and show that any achievable simultaneous ID-rate is contained in this set. In the next section, we will prove that this multi-letter region is equal to the transmission capacity region for DM-CCQ channels.

4.3.1 Proof of the Converse Theorem

When proving converse theorems for transmission problems, it is common to use a "Fano"-like inequality which provides a bound for the achievable rates in terms of a mutual information. For identification problems, because of the double-logarithmic nature of ID-rates, this method is no longer applicable and generally proving converse theorems is, for now, difficult. Introduced by Te Sun Han and Sergio Verdú in [14] is a theory of resolvability which can be used to prove converse theorems for identification. In [7], Steinberg extends this theory and finds the capacity region for classical multiple access channels with memory and frame synchronization and in the process provides a useful lemma which we introduce in this section and use in our proof of the converse.

Notation 4.3.1. *For probability distribution $P^k \in \mathcal{P}(\mathcal{X}^k)$ and CQ channel $\mathbf{W}$, we write*

$$P^k W^k := \sum_{x^k \in \mathcal{X}^k} P^k(x^k) W^k(x^k), \tag{4.3.1}$$

or with an additional distribution $Q^k \in \mathcal{P}(\mathcal{Y}^k)$, for a CCQ channel $\mathbf{W}$,

$$P^k Q^k W^k := \sum_{\substack{x^k \in \mathcal{X}^k \\ y^k \in \mathcal{Y}^k}} P^k(x^k) Q^k(y^k) W^k(x^k, y^k). \tag{4.3.2}$$

When dealing with classical-classical channels $\mathbf{W} := \{W^k(y^k|x^k) : x^k \in \mathcal{X}^k, y^k \in \mathcal{Y}^k\}_{k \in \mathbb{N}}$, we write for $y^k \in \mathcal{Y}^k$ the output of the channel,

$$P^k W^k(y^k) := \sum_{x^k \in \mathcal{X}^k} P^k(x^k) W^k(y^k|x^k). \tag{4.3.3}$$

Notation 4.3.2. *In some cases, for notational simplicity, we refer to a random variable by its distribution. For example, for a random variable A with distribution p, we may refer to A by p.*

Definition 4.3.3. *Let $\rho \in \mathcal{S}(\mathcal{H})$ and $D := \{D_i\}_{i \in [M]}$ a POVM for $M \in \mathbb{N}$. Then a probability distribution $\rho(D)$ is induced on $[M]$ such that $\rho(D)(i) = \mathrm{tr}(\rho D_i)$ for $i \in [M]$. For a second state $\sigma \in \mathcal{S}(\mathcal{H})$, we define, with D,*

$$d_D(\rho, \sigma) := d_1(\rho(D), \sigma(D)), \tag{4.3.4}$$

with d_1 the total variational distance, that is, for a set A and two distributions $p, q \in \mathcal{P}(A)$

$$d_1(p, q) := \sum_{a \in A} |p(a) - q(a)| = 2 \sup_{A' \subseteq A} \{p(A') - q(A')\}. \tag{4.3.5}$$

Definition 4.3.4. *For $D \in \mathcal{L}(\mathcal{H})$ such that $0 \leq D \leq \mathbb{1}$, we define POVM $P(D) \coloneqq \{D, \mathbb{1} - D\}$.*

Lemma 4.3.5. *Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ and $E = (E_i)_{i \in [M]}$ be a POVM on Hilbert space $\mathcal{H}$ with $M \in \mathbb{N}$ events. Let $A_1, A_2 \subset [M]$ such that $A_1 \cap A_2 = \emptyset$ and $A_1 \cup A_2 = [M]$. Then, for $D \coloneqq \sum_{i \in A_1} E_i$,*

$$d_1(\rho(E), \sigma(E)) \geq d_1(\rho(P(D)), \sigma(P(D))), \tag{4.3.6}$$

with $P(D) \coloneqq (D, \mathbb{1} - D)$, as defined in Definition 4.3.4.

Proof. Let ρ, σ, E, A_1 and A_2 be defined as in the lemma statement. Then,

$$\begin{aligned}
d_1(\rho(E), \sigma(E)) &= \sum_{m \in [M]} |\rho(E)(m) - \sigma(E)(m)| && (4.3.7)\\
&= \sum_{m \in A_1} |\rho(E)(m) - \sigma(E)(m)| + \sum_{m \in A_2} |\rho(E)(m) - \sigma(E)(m)| && (4.3.8)\\
&\geq |\sum_{m \in A_1} \rho(E)(m) - \sigma(E)(m)| + |\sum_{m \in A_2} \rho(E)(m) - \sigma(E)(m)| && (4.3.9)\\
&= |\sum_{m \in A_1} \operatorname{tr}(\rho E_m) - \operatorname{tr}(\sigma E_m)| + |\sum_{m \in A_2} \operatorname{tr}(\rho E_m) - \operatorname{tr}(\sigma E_m)| && (4.3.10)\\
&= |\operatorname{tr}(\rho D) - \operatorname{tr}(\sigma D)| + |\operatorname{tr}(\rho(\mathbb{1} - D)) - \operatorname{tr}(\sigma(\mathbb{1} - D))| && (4.3.11)\\
&= |\rho(P(D))(1) - \sigma(P(D))(1)| + |\rho(P(D))(2) - \sigma(P(D))(2)| && (4.3.12)\\
&= d_1(\rho(P(D)), \sigma(P(D))). && (4.3.13)
\end{aligned}$$

□

Lemma 4.3.6 (Steinberg [7], Lemma 6)**.** *With a classical-classical channel $\mathbf{W} \coloneqq \{W^k(y^k|x^k) : x^k \in \mathcal{X}^k, y^k \in \mathcal{Y}^k\}_{k \in \mathbb{N}}$, for a fixed $R > 0$, $\rho > 0$, $k_0 \geq 1$, assume that for every $k > k_0$ there exists a collection of distributions $\mathbf{P} \coloneqq \{P_i^k\}_{i=1}^N \subseteq \mathcal{P}(\mathcal{X}^k)$ such that,*

$$\frac{1}{k} \log \log N \geq R, \tag{4.3.14}$$

and,

$$\min_{i \neq j} d_1(P_i^k W^k, P_j^k W^k) > 2(1 - \rho). \tag{4.3.15}$$

Then for every $\gamma < (\rho/4) \cdot \min(1, R)$, there exists a subset $\tilde{\mathbf{P}} \subseteq \mathbf{P}$ such that for every $k > k_0(m, R, \rho, \gamma)$ independent of $\mathbf{W}$ and $\mathbf{P}$ it holds,

$$|\tilde{\mathbf{P}}| \geq \exp\exp(kR) - \exp\exp(k(R - \gamma)) \tag{4.3.16}$$

and for every $\tilde{P}^k \in \tilde{\mathbf{P}}$,

$$R(1 - 4\rho) \leq \frac{1}{k} I(\tilde{P}^k; \tilde{P}^k W^k). \tag{4.3.17}$$

Definition 4.3.7. *For a CCQ channel $\mathbf{W}$,*

$$C'(\mathbf{W}) \coloneqq cl\left(\liminf_{k \to \infty} C_k(\mathbf{W})\right), \tag{4.3.18}$$

with

$$C_k(\mathbf{W}) \coloneqq \bigcup_{\substack{p_1 \in \mathcal{P}(\mathcal{X}^k) \\ p_2 \in \mathcal{P}(\mathcal{Y}^k)}} \left\{(R_1, R_2) \mid 0 \leq R_1 \leq \frac{1}{k} I(A^k; C^k)_{\gamma_2^k(p_1, p_2)}, 0 \leq R_2 \leq \frac{1}{k} I(B^k; C^k)_{\gamma_2^k(p_1, p_2)}\right\}, \tag{4.3.19}$$

where A^k and B^k refer to the respective output Hilbert spaces of the two senders and C^k the output Hilbert space of the channel. For the definition of the limit for sets, we refer the reader to Appendix A.

Theorem 4.3.8. *For CCQ channel* $\mathbf{W}$*,*

$$C_{id}^{sim}(\mathbf{W}) \subseteq C'(\mathbf{W}). \tag{4.3.20}$$

Proof. Given a CCQ channel $\mathbf{W} \coloneqq \{W^k : \mathcal{X}^k \times \mathcal{Y}^k \to \mathcal{S}(\mathcal{H}_C^{\otimes k})\}$, let $(R_1, R_2) \in C_{\mathrm{id}}^{\mathrm{sim}}(\mathbf{W})$, then for all $\lambda_1, \lambda_2, \delta > 0$ there exists a k_0 such that for all $k \geq k_0$, there is a simultaneous $(k, M, N)-$ID-code $\mathcal{C}_{\mathrm{id}}^{\mathrm{sim}}$ with

$$\frac{1}{k}\log\log M \geq R_1 - \delta, \;\; \frac{1}{k}\log\log N \geq R_2 - \delta, \;\; e_1(\mathcal{C}_{\mathrm{id}}^{\mathrm{sim}}, W^k) \leq \lambda_1, \;\; e_2(\mathcal{C}_{\mathrm{id}}^{\mathrm{sim}}, W^k) \leq \lambda_2, \tag{4.3.21}$$

and specifically there are codes with $\lambda_1 + \lambda_2 < 1$. Without loss of generality, assume for $\mu > 0$,

$$R_2 = R_1 - \mu. \tag{4.3.22}$$

Let $\lambda_1 + \lambda_2 < 1$ and define $\epsilon \coloneqq 1 - \lambda_1 - \lambda_2$. For $k \geq k_0$, let $\mathcal{C}_{\mathrm{id}}^{\mathrm{sim}} \coloneqq (P_i^k, Q_j^k, D_{ij}^k)_{i=1,j=1}^{M,N}$ be the simultaneous (k, M, N)-ID-code, where with $\delta > 0$,

$$M \coloneqq \lceil 2^{2^{(k(R_1-\delta))}} \rceil \quad \text{and} \quad N \coloneqq \lceil 2^{2^{(k(R_2-\delta))}} \rceil, \tag{4.3.23}$$

with errors as in (4.3.21). Define POVM E^k indexed by $z^k \in \mathcal{Z}^k$ as the the common refinement of $\{D_{ij}^k\}_{i=1,j=1}^{M,N}$. For a fixed $j \in [N]$ define the channel,

$$W_j^k : \mathcal{X}^k \to \mathcal{S}(\mathcal{H}_C^{\otimes k}), \;\; x^k \mapsto \sum_{y^k \in \mathcal{Y}^k} Q_j^k(y^k) W^k(x^k, y^k). \tag{4.3.24}$$

Then it is easy to see that with W_j^k, $(P_i^k, D_{ij}^k)_{i=1}^M$ is an simultaneous (k, M)-ID-code, since rate and error bounds hold. Further, for all $1 \leq a < b \leq M$,

$$\begin{aligned}
d_{E^k}(P_a^k W_j^k, P_b^k W_j^k) &= d_1(P_a^k W_j^k(E^k), P_b^k W_j^k(E^k)) && (4.3.25)\\
&\geq d_1(P_a^k W_j^k(P(D_{aj}^k)), P_b^k W_j^k(P(D_{aj}^k))) && (4.3.26)\\
&\geq 2\left(\mathrm{tr}\big(P_a^k W_j^k D_{aj}^k\big) - \mathrm{tr}\big(P_b^k W_j^k D_{aj}^k\big)\right) && (4.3.27)\\
&= 2\left(\mathrm{tr}\big(P_a^k Q_j^k W^k D_{aj}^k\big) - \mathrm{tr}\big(P_b^k Q_j^k W^k D_{aj}^k\big)\right) && (4.3.28)\\
&> 2\left(1 - e_1(\mathcal{C}_{\mathrm{id}}^{\mathrm{sim}}, W^k) - e_2(\mathcal{C}_{\mathrm{id}}^{\mathrm{sim}}, W^k)\right) && (4.3.29)\\
&\geq 2(1 - \lambda_1 - \lambda_2) && (4.3.30)\\
&= 2\epsilon. && (4.3.31)
\end{aligned}$$

The first inequality holds by Lemma 4.3.5. The second inequality holds by using Definition 4.3.3 with a fixed subset for the variational distance. The second equality comes from simply replacing W_j^k with its definition. The third inequality holds by definition of the error types. The fourth inequality holds by the error bound restrictions imposed on $\mathcal{C}_{\mathrm{id}}^{\mathrm{sim}}$. With this, we see if we construct the classical channel

$$\tilde{W}_j^k(z^k|x^k) = \mathrm{tr}\big(W_j^k(x^k) E_{z^k}^k\big) \tag{4.3.32}$$

for some finite output alphabet $\mathcal{Z}^k$ and $z^k \in \mathcal{Z}^k$, then for $1 \leq a \leq M$,

$$\begin{aligned}
P_a^k \tilde{W}_j^k(z^k) &= \sum_{x^k \in \mathcal{X}^k} P_a^k(x^k) \tilde{W}_j^k(z^k|x^k) && (4.3.33)\\
&= \sum_{x^k \in \mathcal{X}^k} P_a^k(x^k)\, \mathrm{tr}\big(W_j^k(x^k) E_{z^k}^k\big) && (4.3.34)\\
&= \mathrm{tr}\left(\sum_{x^k \in \mathcal{X}^k} P_a^k(x^k) W_j^k(x^k) E_{z^k}^k\right) && (4.3.35)\\
&= P_a^k W_j^k(E^k)(z^k), && (4.3.36)
\end{aligned}$$

where the first equality is by notational choice in Notation 4.3.1, the second by replacing $\tilde{W}_j^k(z^k|x^k)$ with its definition, the third by the linearity of the trace and the last again by notational choice. Then it holds for all $1 \leq a < b \leq M$,

$$d_1(P_a^k\tilde{W}_j^k, P_b\tilde{W}_j^k) = d_1(P_a^kW_j^k(E^k), P_b^kW_j^k(E^k)) > 2\epsilon, \tag{4.3.37}$$

where the inequality holds from above. Letting $\rho := 1 - \epsilon$ and choosing

$$\gamma < \min\left(\mu, \rho/4, (\rho/4)(R_1 - \delta)\right), \tag{4.3.38}$$

by Lemma 4.3.6, there exists a subset $\tilde{\mathbf{P}}_j \subset (P_i^k)_{i=1}^M$ such that

$$|\tilde{\mathbf{P}}_j| \geq \exp\exp(k(R_1 - \delta)) - \exp\exp(k(R_1 - \delta - \gamma)) \tag{4.3.39}$$

and for all $\tilde{P}_{i^*}^k \in \tilde{\mathbf{P}}_j$,

$$(R_1 - \delta)(1 - 4\rho) \leq \frac{1}{k}I(\tilde{P}_{i^*}^k; \tilde{P}_{i^*}^kW_j^k(E^k)), \tag{4.3.40}$$

for all k sufficiently large, depending only on $k_0, R_1 - \delta, \gamma$, and ρ. We show that

$$\bigcap_{j=1}^N \tilde{\mathbf{P}}_j \neq \emptyset, \tag{4.3.41}$$

which implies that regardless of j, we can always find such a $\tilde{P}_{i^*}^k$. In reference to the proof of Lemma 6 in [7], a set $\mathcal{Z}_j^k$, for a fixed j, is defined as,

$$\mathcal{Z}_j^k := \left\{P^k \in (P_i^k)_{i=1}^M \mid \frac{1}{k}I(P^k; P^kW_j^k(E^k)) < R_1 - \delta\right\}, \tag{4.3.42}$$

and it is shown that $|\mathcal{Z}_j^k| \leq \exp\exp(k(R_1 - \delta - \gamma))$ and $\tilde{\mathbf{P}}_j := (P_i^k)_{i=1}^M \setminus \mathcal{Z}_j^k$. It clear then that

$$\left|\bigcup_{j=1}^N \mathcal{Z}_j^k\right| \leq \sum_{j=1}^N |\mathcal{Z}_j^k| \leq N\exp\exp(k(R_1 - \delta - \gamma)). \tag{4.3.43}$$

Therefore, it holds that,

$$\begin{aligned}
\left|\bigcap_{j=1}^N \tilde{\mathbf{P}}_j\right| &= M - \left|\bigcup_{j=1}^N \mathcal{Z}_j^k\right| && (4.3.44)\\
&\geq M - N\exp\exp(k(R_1 - \delta - \gamma)) && (4.3.45)\\
&\geq \lceil\exp\exp(k(R_1 - \delta))\rceil - \lceil\exp\exp(k(R_2 - \delta))\rceil\exp\exp(k(R_1 - \delta - \gamma)) && (4.3.46)\\
&= \lceil\exp\exp(k(R_1 - \delta))\rceil - \lceil\exp\exp(k(R_1 - \delta - \mu))\rceil\exp\exp(k(R_1 - \delta - \gamma)) && (4.3.47)\\
&> 0 && (4.3.48)
\end{aligned}$$

for k large enough, where the second inequality is by how M and N are defined, and the second equality is by the assumption that $R_2 = R_1 - \mu$. Therefore, there exists at least one distribution $\tilde{P}_{i^*}^k$ for all $j \in [N]$ with

$$(R_1 - \delta)(1 - 4\rho) \leq \frac{1}{k}I(\tilde{P}_{i^*}^k; \tilde{P}_{i^*}^kW_j^k(E^k)). \tag{4.3.49}$$

With such a $\tilde{P}_{i^*}^k$, construct the channel,

$$V_{\tilde{P}}^k : \mathcal{Y}^k \to \mathcal{S}(\mathcal{H}_C^{\otimes k}),\ \ y^k \mapsto \sum_{x^k \in \mathcal{X}^k} \tilde{P}_{i^*}^k(x^k)W^k(x^k, y^k). \tag{4.3.50}$$

Repeating the same argumentation, we can conclude that there is a subset $\tilde{\mathbf{Q}} \subset (Q_j^k)_{j=1}^N$ with

$$|\tilde{\mathbf{Q}}| \geq \exp\exp(k(R_2 - \delta)) - \exp\exp(k(R_2 - \delta - \gamma)) > 0 \tag{4.3.51}$$

and for any element $\tilde{Q}_{j^*}^k \in \tilde{\mathbf{Q}}$, where j^* indicates the index of such a distribution,

$$(R_2 - \delta)(1 - 4\rho) \leq \frac{1}{k} I(\tilde{Q}_{j^*}^k; \tilde{Q}_{j^*}^k V_{\tilde{P}}^k(E^k)). \tag{4.3.52}$$

Now, with channel state $\alpha \in \mathcal{S}(\mathcal{H}_A^{\otimes k} \otimes \mathcal{H}_C^{\otimes k})$,

$$\alpha := \sum_{x^k \in \mathcal{X}^k} \tilde{P}_{i^*}^k(x^k) \left|x^k\middle\rangle\!\middle\langle x^k\right| \otimes W_{j^*}^k(x^k) \tag{4.3.53}$$

it holds,

$$\begin{aligned} I(A^k, C^k)_\alpha &= \chi(\tilde{P}_{i^*}^k, W_{j^*}^k) && (4.3.54)\\ &\geq I_{\text{acc}}(\tilde{P}_{i^*}^k, W_{j^*}^k) && (4.3.55)\\ &= \max_{\text{POVM } \tilde{E}^k} I(\tilde{P}_{i^*}^k, \tilde{P}_{i^*}^k W_{j^*}^k(\tilde{E}^k)) && (4.3.56)\\ &\geq I(\tilde{P}_{i^*}^k, \tilde{P}_{i^*}^k W_{j^*}^k(E^k)) && (4.3.57) \end{aligned}$$

where I_{acc} is the accessible information defined in Appendix B. The first equality holds since α is a channel state. The first inequality is the Holveo bound. The second equality is the definition of accessible information which we refer the reader to Appendix B for. Now, defining channel state $\beta \in \mathcal{S}(\mathcal{H}_B^{\otimes k} \otimes \mathcal{H}_C^{\otimes k})$

$$\beta := \sum_{y^k \in \mathcal{Y}^k} \tilde{Q}_{j^*}^k(y^k) \left|y^k\middle\rangle\!\middle\langle y^k\right| \otimes V_{\tilde{P}}^k(y^k), \tag{4.3.58}$$

by the same reasoning,

$$I(B^k, C^k)_\beta \geq I(\tilde{Q}_{j^*}^k, \tilde{Q}_{j^*}^k V_{\tilde{P}}^k(E^k)). \tag{4.3.59}$$

It is clear that for the channel state $\gamma_2^k \in \mathcal{S}(\mathcal{H}_A^{\otimes k} \otimes \mathcal{H}_B^{\otimes k} \otimes \mathcal{H}_C^{\otimes k})$,

$$\gamma_2^k := \gamma_2^k(\tilde{P}_{i^*}^k, \tilde{Q}_{j^*}^k) = \sum_{x^k \in \mathcal{X}^k} \sum_{y^k \in \mathcal{Y}^k} \tilde{P}_{i^*}^k(x^k) \left|x^k\middle\rangle\!\middle\langle x^k\right| \otimes \tilde{Q}_{j^*}^k(y^k) \left|y^k\middle\rangle\!\middle\langle y^k\right| \otimes W^k(x^k, y^k), \tag{4.3.60}$$

it holds,

$$I(A^k, C^k)_\alpha = I(A^k, C^k)_{\gamma_2^k} \quad \text{and} \quad I(B^k, C^k)_\beta = I(B^k, C^k)_{\gamma_2^k}. \tag{4.3.61}$$

Therefore,

$$(R_1 - \delta)(1 - 4\rho) \leq \frac{1}{k} I(A^k; C^k)_{\gamma_2^k} \tag{4.3.62}$$

and

$$(R_2 - \delta)(1 - 4\rho) \leq \frac{1}{k} I(B^k, C^k)_{\gamma_2^k}. \tag{4.3.63}$$

Since δ and ρ can be arbitrarily small, it implies that,

$$(R_1, R_2) \in C'(\mathbf{W}). \tag{4.3.64}$$

□

4.4 Equivalence of Capacity Regions

In this section, we show that for a DM-CCQ channel generated by W, the multi-letter capacity region $C'(W)$ in the previous section is indeed equal to the transmission capacity $C(W)$. To do so, we make use of the strategy used by Yosef Steinberg in [7, Chapter 3] and Sergio Verdú in [8] where the combined results provide a method to show the equivalence. We define an intermediate region which will be used to prove the result.

Definition 4.4.1. *For a CCQ channel* $\mathbf{W}$ *and* $k \in \mathbb{N}$*,*

$$R_k(\mathbf{W}) := \bigcup_{\substack{p_1 \in \mathcal{P}(\mathcal{X}^k) \\ p_2 \in \mathcal{P}(\mathcal{Y}^k)}} \Big\{ (R_1, R_2) \mid R_1 \leq \frac{1}{k} I(A^k; C^k | B^k)_{\gamma_2^k(p_1,p_2)},\ R_2 \leq \frac{1}{k} I(B^k; C^k | A^k)_{\gamma_2^k(p_1,p_2)}, \\ R_1 + R_2 \leq \frac{1}{k} I(A^k, B^k; C^k)_{\gamma_2^k(p_1,p_2)} \Big\} \tag{4.4.1}$$

where A^k *and* B^k *refer to the respective output Hilbert spaces of the two senders and* C^k *the output Hilbert space of the channel.*

Lemma 4.4.2. *For a CCQ channel* $\mathbf{W}$*, for all* $k \in \mathbb{N}$*,*

$$C_k(\mathbf{W}) \subseteq R_k(\mathbf{W}). \tag{4.4.2}$$

Proof. We show for any two of random variables A and B such that A and B are mutually independent, C the channel output, and γ_2 any channel state, that it holds,

$$I(A;C)_{\gamma_2} \leq I(A;C|B)_{\gamma_2}, \tag{4.4.3}$$
$$I(B;C)_{\gamma_2} \leq I(B;C|A)_{\gamma_2}, \tag{4.4.4}$$
$$I(A,C)_{\gamma_2} + I(B,C)_{\gamma_2} \leq I(A,B;C)_{\gamma_2}. \tag{4.4.5}$$

With A and B independent, $I(A;B)_{\gamma_2} = 0$. Moreover, $I(A;B|C)_{\gamma_2} \geq 0$. So

$$\begin{aligned}
I(A;C|B)_{\gamma_2} &\geq I(A;C|B)_{\gamma_2} + I(A;B)_{\gamma_2} - I(A;B|C)_{\gamma_2} && (4.4.6)\\
&= H(A|B)_{\gamma_2} + H(C|B)_{\gamma_2} - H(A,C|B)_{\gamma_2} \\
&\quad + H(A)_{\gamma_2} + H(B)_{\gamma_2} - H(A,B)_{\gamma_2} && (4.4.7)\\
&\quad - H(A|C)_{\gamma_2} - H(B|C)_{\gamma_2} + H(A,B|C)_{\gamma_2} \\
&= H(A,B)_{\gamma_2} - H(B)_{\gamma_2} + H(B,C)_{\gamma_2} - H(B)_{\gamma_2} - H(A,B,C)_{\gamma_2} + H(B)_{\gamma_2} \\
&\quad - H(A)_{\gamma_2} + H(B)_{\gamma_2} - H(A,B)_{\gamma_2} && (4.4.8)\\
&\quad - H(A,C)_{\gamma_2} + H(C)_{\gamma_2} - H(B,C)_{\gamma_2} + H(C)_{\gamma_2} + H(A,B,C)_{\gamma_2} - H(C)_{\gamma_2} \\
&= H(A)_{\gamma_2} + H(C)_{\gamma_2} - H(A,C)_{\gamma_2} && (4.4.9)\\
&= I(A;C)_{\gamma_2}, && (4.4.10)
\end{aligned}$$

where all equalities hold simply by definition of entropy and mutual information. It can be shown in a similar way that similarly $I(B;C|A)_{\gamma_2} \geq I(B;C)_{\gamma_2}$. Further,

$$\begin{aligned}
I(A;C)_{\gamma_2} + I(B;C)_{\gamma_2} &= H(A) - H(A|C)_{\gamma_2} + H(B) - H(B|C)_{\gamma_2} && (4.4.11)\\
&= H(A,B)_{\gamma_2} - H(A|C)_{\gamma_2} - H(B|C)_{\gamma_2} && (4.4.12)\\
&= H(A,B)_{\gamma_2} - H(A,B|C)_{\gamma_2} + H(B|A,C)_{\gamma_2} - H(B|C)_{\gamma_2} && (4.4.13)\\
&= I(A,B;C)_{\gamma_2} - I(A;B|C)_{\gamma_2} && (4.4.14)\\
&\leq I(A,B;C)_{\gamma_2}, && (4.4.15)
\end{aligned}$$

where the first equality is by definition of quantum mutual information and also the fact that for classical variables the Shannon entropy is equal to the von Neumann entropy. The second equality follows by the independence of A and B and again that the Shannon and von Neumann entropies are equal for classical variables. The third equality follows because, with some manipulation $H(A|C)_{\gamma_2} = H(A,B|C)_{\gamma_2} - H(B|A,C)_{\gamma_2}$, seen as follows. By definition,

$$H(A,B|C)_{\gamma_2} = H(A,B,C)_{\gamma_2} - H(C) \tag{4.4.16}$$

and,

$$H(B|A,C)_{\gamma_2} = H(A,B,C)_{\gamma_2} - H(AC), \tag{4.4.17}$$

and so,

$$\begin{align} H(A,B|C)_{\gamma_2} - H(B|A,C)_{\gamma_2} &= H(A,B,C)_{\gamma_2} - H(C)_{\gamma_2} - H(A,B,C)_{\gamma_2} + H(AC)_{\gamma_2} \tag{4.4.18}\\ &= H(AC)_{\gamma_2} - H(C)_{\gamma_2} \tag{4.4.19}\\ &= H(A|C)_{\gamma_2}. \tag{4.4.20}\end{align}$$

The fourth equality, (4.4.14), holds because $I(A;B|C)_{\gamma_2} = H(B|C)_{\gamma_2} - H(B|A,C)_{\gamma_2}$. To see this we just need that $H(B|A,C)_{\gamma_2} = H(A,B|C)_{\gamma_2} - H(A|C)_{\gamma_2}$ and the rest follows by definition. Explicitly,

$$\begin{align} H(A|C)_{\gamma_2} &= H(AC)_{\gamma_2} - H(C)_{\gamma_2} \tag{4.4.21}\\ H(A,B|C)_{\gamma_2} &= H(A,B,C)_{\gamma_2} - H(C)_{\gamma_2}, \tag{4.4.22}\end{align}$$

so,

$$\begin{align} H(A,B|C)_{\gamma_2} - H(A|C)_{\gamma_2} &= H(A,B,C)_{\gamma_2} - H(C)_{\gamma_2} - H(AC)_{\gamma_2} + H(C)_{\gamma_2} \tag{4.4.23}\\ &= H(A,B,C)_{\gamma_2} - H(AC)_{\gamma_2} \tag{4.4.24}\\ &= H(B|A,C)_{\gamma_2}. \tag{4.4.25}\end{align}$$

The inequality follows from positivity of mutual information. Since all random variables are chosen arbitrarily and this holds for any channel state, the statement holds for all k as was to show. □

Lemma 4.4.3. *For a CCQ channel* $\mathbf{W}$*,*

$$C(\mathbf{W}) \subseteq cl\left(\liminf_{k\to\infty} R_k(\mathbf{W})\right) \tag{4.4.26}$$

Proof. The proof follows the structure of [8, Theorem 1]. Let $(R_1, R_2) \in C(\mathbf{W})$, and let $\epsilon \in (0,1)$, $\delta > 0$. By definition, there exists a k_0 such that for all $k \geq k_0$, there is a (k, M, N)-code $\mathcal{C} := (x_m, y_n, D_{mn})_{m=1,n=1}^{M,N}$ such that,

$$\frac{1}{k}\log M \geq R_1 - \delta, \quad \frac{1}{k}\log N \geq R_2 - \delta, \quad \overline{e}(\mathcal{C}, W^k) \leq \epsilon. \tag{4.4.27}$$

Let $\mathcal{C}$ be such a code with k fixed. Let A^k and B^k be independent random variables uniformly distributed on the codewords represented by $[M]$ and $[N]$ respectively. Let C^k be the output of W^k when A^k and B^k are sent, and let $(\hat{A}, \hat{B})$ be random variables for the decoding of C^k. Then, the Markov chain $(A^k, B^k) \to C^k \to (\hat{A}, \hat{B})$ is formed. By Fano's inequality, Lemma 3.3.1, it holds,

$$\begin{align} H(A^k, B^k|C^k)_{\gamma_2^k} &\leq 1 + \epsilon\log(MN) \tag{4.4.28}\\ H(A^k|C^k)_{\gamma_2^k} &\leq 1 + \epsilon\log(M) \tag{4.4.29}\\ H(B^k|C^k)_{\gamma_2^k} &\leq 1 + \epsilon\log(N). \tag{4.4.30}\end{align}$$

where γ_2^k is the channel state constructed with A^k and B^k. We use Lemma 3.1.14 for existence of codes for the single sender inequalities (4.4.29) and (4.4.30) with no loss of rate or accuracy. Now, since A^k and B^k are uniformly distributed,

$$I(A^k, B^k; C^k)_{\gamma_2^k} \geq (1-\epsilon)\log(MN) - 1$$

$$\begin{aligned} I(A^k; C^k)_{\gamma_2^k} &\geq (1-\epsilon)\log(M) - 1 \\ I(B^k; C^k)_{\gamma_2^k} &\geq (1-\epsilon)\log(N) - 1. \end{aligned}$$

Since A^k and B^k are independent, it holds,

$$\begin{aligned} I(A^k; C^k|B^k)_{\gamma_2^k} &= I(A^k; C^k|B^k)_{\gamma_2^k} + \underbrace{I(A^k; B^k)_{\gamma_2^k}}_{=0} && (4.4.31) \\ &= I(A^k; C^k, B^k)_{\gamma_2^k} && (4.4.32) \\ &\geq I(A^k; C^k)_{\gamma_2^k} && (4.4.33) \end{aligned}$$

The second equality is by the chain rule property of quantum mutual information (see Appendix B). The inequality holds since for any quantum state ρ,

$$\begin{aligned} I(A^k; C^k, B^k)_\rho &= H(A^k)_\rho - H(A^k|C^k, B^k)_\rho && (4.4.34) \\ &\geq H(A^k)_\rho - H(A^k|C^k)_\rho && (4.4.35) \\ &= I(A^k; C^k)_\rho, && (4.4.36) \end{aligned}$$

where we use that conditioning does not increase quantum entropy [5, Theorem 11.4.1]. Similarly it can be shown that,

$$I(B^k; C^k|A^k)_{\gamma_2^k} \geq I(B^k; C^k)_{\gamma_2^k} \tag{4.4.37}$$

Combining these results, it holds,

$$\begin{aligned} (1-\epsilon)(R_1 - \delta) - \frac{1}{k} &\leq \frac{1}{k} I(A^k; C^k|B^k)_{\gamma_2^k} && (4.4.38) \\ (1-\epsilon)(R_2 - \delta) - \frac{1}{k} &\leq \frac{1}{k} I(B^k; C^k|A^k)_{\gamma_2^k} && (4.4.39) \\ (1-\epsilon)(R_1 + R_2 - 2\delta) - \frac{1}{k} &\leq \frac{1}{k} I(A^k, B^k; C^k)_{\gamma_2^k} && (4.4.40) \end{aligned}$$

and therefore,

$$(1-\epsilon)(R_1 - \delta, R_2 - \delta) - \left(\frac{1}{k}, \frac{1}{k}\right) \in R_k. \tag{4.4.41}$$

for k large enough which further implies,

$$(1-\epsilon)(R_1 - 2\delta, R_2 - 2\delta) \in \liminf_{k\to\infty} R_k. \tag{4.4.42}$$

Since ϵ and δ are chosen arbitrarily, (4.4.42) gives us that (R_1, R_2) is a limit of a sequence of points in $\liminf_{k\to\infty} R_k$ and is therefore in the closure of $\liminf_{k\to\infty} R_k$ with channel state constructed with the distributions on A^k and B^k. □

Lemma 4.4.4. *For a DM-CCQ channel generated by W, for all $k_0 \in \mathbb{N}$ it holds,*

$$cl\left(\bigcup_{k\geq k_0} R_k(W)\right) \subseteq C(W). \tag{4.4.43}$$

Proof. Let $k_0 \in \mathbb{N}$ and $k \geq k_0$. Further, let $(R_1, R_2) \in R_k$. With this, it holds,

$$\begin{aligned} R_1 &\leq \frac{1}{k} I(A^k; C^k|B^k)_{\gamma_2^k} && (4.4.44) \\ &= \frac{1}{k} I(A^k; C^k, B^k)_{\gamma_2^k} && (4.4.45) \\ &\leq \frac{1}{k} \cdot k I(A; C, B)_{\gamma_2} && (4.4.46) \end{aligned}$$

$$= I(A;C|B)_{\gamma_2}, \tag{4.4.47}$$

with γ_2^k the channel state constructed from A^k and B^k. The first inequality is by definition of $R_k(W)$. For mutually independent random variables A^k and B^k, it is easy with the chain rule for mutual information that, for C^k the channel output space, $I(A^k;C^k|B^k)_\rho = I(A^k;B^k,C^k)_\rho$, for any state ρ and so the first equality holds. The second inequality is due to Lemma 3.3.2. It can be similarly shown that $R_2 \le I(B;C|A)_{\gamma_2}$. Further, Lemma 3.3.2 gives $R_1 + R_2 \le I(A,B;C)_{\gamma_2}$. Thus, with the channel state γ_2, $(R_1, R_2) \in C(W)$. □

Using these results, it is now straight forward to prove the following theorem.

Theorem 4.4.5. *For a DM-CCQ channel generated by W, it holds,*

$$C(W) = cl\left(\liminf_{k\to\infty} R_k(W)\right).$$

Proof. Combining the results of the previous lemmas, we have,

$$C(W) \subseteq \mathrm{cl}\left(\liminf_{k\to\infty} R_k(W)\right) \subseteq \mathrm{cl}\left(\limsup_{k\to\infty} R_k(W)\right) \subseteq \mathrm{cl}\left(\bigcup_{k\geq k_0} R_k(W)\right) \subseteq C(W), \tag{4.4.48}$$

where the first containment is from Lemma 4.4.3, the second and third by the structure of the limits for sets with any $k_0 > 0$, and the last by Lemma 4.4.4. □

Using these results, we can prove the main theorem of the thesis.

Proof of Theorem 4.1.8. It holds,

$$C(W) \subseteq C_{\mathrm{id}}^{\mathrm{sim}}(W) \subseteq \mathrm{cl}\left(\liminf_{k\to\infty} C_k(W)\right) \subseteq \mathrm{cl}\left(\liminf_{k\to\infty} R_k(W)\right) = C(W), \tag{4.4.49}$$

where the first containment is by 4.2.1, the second from 4.3.8, the third by Lemma 4.4.2 and the equality by Lemma 4.4.5. □

Chapter 5

Conclusion

In this thesis, we have reviewed a highly interesting communication problem proposed by Rudolf Ahlswede and Gunter Dueck labeled "identification". As the label suggests, the problem of identification is not to determine what a received message is exactly, but just determining if a received message is one of interest. We reviewed the doubly exponential capacity theorem of Ahlswede, Dueck, et al for the classical single sender-single receiver channel and gave the full proof of achievability using the Transformator lemma.

From here, we introduced concepts of quantum information theory and most importantly to this thesis, the classical-quantum multiple access channel. We reviewed the transmission capacity theorem given by Andreas Winter for the quantum multiple access channel, reviewing the achievability and converse proof. We also introduced useful lemmas such as the Fano's inequality and subadditivity of quantum mutual information which we make use of in the latter sections.

In the following chapters, we extended the models for ID-codes introduced by Ahlswede and Peter Löber for single sender-single receiver classical-quantum (CQ) channels to the classical-classical-quantum (CCQ) channel. Using Löber's simultaneous ID-code for the CQ channel, we define a simultaneous ID-code for the CCQ channel and proved that the transmission capacity region of the discrete memoryless CCQ channel is equal to its simultaneous identification capacity region.

Appendices

Appendix A

Definitions and Notations

In this section, we give specific definitions and explain notations that will be used throughout the thesis.

Definition A.1 (Finite Alphabet)**.** *A finite alphabet is simply a finite set of symbols, for example* $\{0,1\}$*. Finite alphabets will be referred to in this thesis with symbols such as* $\mathcal{X}$*,* $\mathcal{Y}$ *or* $\mathcal{Z}$*. Strings are finite sequences of elements from a finite alphabet and the set of* k*-length,* $k \in \mathbb{N}$*, strings composed of symbols in* $\mathcal{X}$ *is notated as* $\mathcal{X}^k$*.*

Notation A.2 (Probability Distribution)**.** *The set of probability distributions on a set* $\mathcal{X}$ *is denoted* $\mathcal{P}(\mathcal{X})$*.*

Notation A.3 (Linear Operators)**.** *The set of linear operators on a Hilbert space* $\mathcal{H}$ *is denoted* $\mathcal{L}(\mathcal{H})$*.*

Notation A.4. *For* $M \in \mathbb{N}$ *we define* $[M] := \{1, ..., M\}$*.*

Definition A.5 (Hermitian Operator)**.** *For a Hilbert space* $\mathcal{H}$*, an linear operator* $D \in \mathcal{L}(\mathcal{H})$ *is called Hermitian or self-adjoint if it is equal to its conjugate transpose, which we denote* $D^\dagger$*.*

Definition A.6 (Unitary Operator)**.** *For a Hilbert space* $\mathcal{H}$*, an linear operator* $D \in \mathcal{L}(\mathcal{H})$ *is called unitary if* $D^\dagger D = DD^\dagger = \mathbb{1}_{\mathcal{H}}$*.*

Definition A.7 (Positive Semi-Definite Operator)**.** *For a Hilbert space* $\mathcal{H}$*, a linear operator* $D \in \mathcal{L}(\mathcal{H})$ *is said to be a positive operator if,*

$$D = D^\dagger, \text{ and } \quad \forall \, |x\rangle \in \mathcal{H}, \quad \langle x|D|x\rangle \geq 0.$$

We write $D \geq 0$ *to indicate that* D *is positive semi-definite. We further introduce a partial ordering such that for a second operator* $C \in \mathcal{L}(\mathcal{H})$*,* $D \leq C$ *if* $C - D \geq 0$*.*

Definition A.8 (Completely Positive)**.** *For Hilbert spaces* $\mathcal{H}_A$ *and* $\mathcal{H}_B$*, a linear map* $D : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ *is called completely positive if for any reference Hilbert space* R*,* $\mathbb{1}_R \otimes D$ *is positive.*

Definition A.9 (Common refinement)**.** *Let* $P = \{[x_0, x_1], [x_1, x_2], ..., [x_{n-1}, x_n]\}$ *be a partition of the interval* $[a, b]$*. Then the partition* P' *of* $[a, b]$ *is a refinement of* P *if the partition points in* P' *include the partition points of* P*. A third partition* Q *of* $[a, b]$ *is a common refinement of two partitions* P *and* P' *of* $[a, b]$ *if* Q *comprises of the partition points of* P *together with* P'*.*

Definition A.10 (Set Subtraction)**.** *For two finite sets* A *and* B *we define set subtraction with the implication that* $A \setminus B = A \setminus (B \cap A)$*.*

Definition A.11 (Limit Superior and Limit Inferior for Sets)**.** *For a sequence of sets* $(C_k)_{k\in\mathbb{N}}$*,*

$$\liminf_{k\to\infty} C_k := \bigcup_{n\geq 1} \bigcap_{k\geq n} C_k \quad \text{and} \quad \limsup_{k\to\infty} C_k := \bigcap_{n\geq 1} \bigcup_{k\geq n} C_k. \tag{A.1}$$

Appendix B

Information Quantities

Here we define the information quantities used in the thesis and some properties of them.

Definition B.1 (Shannon Entropy)**.** *The Shannon entropy of a discrete random variable X with a probability distribution $p \in \mathcal{P}(X)$ is*

$$H(X) \coloneqq -\sum_{x \in X} p(x) \log(p(x)).$$

Definition B.2 (von Neumann Entropy)**.** *For a quantum state $\rho_A \in \mathcal{S}(\mathcal{H}_A)$, the von Neumann entropy of ρ_A is defined as*

$$H(A)_\rho \coloneqq -\operatorname{tr}\left(\rho_A \log \rho_A\right).$$

We may also write the entropy in the form $H(\rho_A)$ with implication that it should be evaluated as defined above.

Definition B.3 (Joint von Neumann Entropy)**.** *For a quantum state $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, the joint von Neumann entropy of ρ_{AB} is defined as*

$$H(A, B)_\rho \coloneqq -\operatorname{tr}\left(\rho_{AB} \log \rho_{AB}\right).$$

Definition B.4 (Conditional von Neumann Entropy)**.** *For a quantum state $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, the conditional von Neumann entropy of ρ_{AB} is defined as*

$$H(A|B)_\rho \coloneqq H(A, B)_\rho - H(B)_\rho.$$

Definition B.5 (Classical Mutual Information)**.** *For discrete random variables A and B with joint probability distribution $p \in \mathcal{P}(A \times B)$, the mutual information is*

$$I(A; B) \coloneqq H(A) - H(A \mid B).$$

Definition B.6 (Accessible Information)**.** *For a probability distribution $p \in \mathcal{P}(\mathcal{X})$, and a CQ channel $\mathbf{W}$, let the channel state $\gamma \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be defined as*

$$\gamma \coloneqq \sum_{x \in \mathcal{X}} p(x) \left|x\middle\rangle\middle\langle x\right| \otimes W(x). \tag{B.1}$$

For a POVM $\mathbf{E} \coloneqq (E_y)_{y=1}^M$ on $\mathcal{H}_B$ a random variable B can be constructed via probability distribution $q(y) = \operatorname{tr}(E_y \gamma) \in \mathcal{P}(\mathcal{Y})$. With A a random variable governed by p, the accessible information I_{acc} is then defined as

$$I_{acc} \coloneqq \max_{POVM\ \mathbf{E}} I(A, B). \tag{B.2}$$

Definition B.7 (Quantum Mutual Information)**.** *For a bipartite state $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, the quantum mutual information is defined as*

$$I(A; B)_\rho \coloneqq H(A)_\rho + H(B)_\rho - H(A, B)_\rho.$$

Definition B.8 (Conditional Quantum Mutual Information). *For parties A, B, and C the conditional quantum mutual information for tripartite state $\rho_{ABC} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ is defined as*

$$I(A; C \mid B)_\rho \coloneqq H(A \mid B)_\rho + H(C \mid B)_\rho - H(A, C \mid B)_\rho,$$

where conditional von Neumann entropies are taken similarly as with the above definition.

Property B.9 (Chain Rule for Quantum Mutual Information). *For a quantum state $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$, quantum mutual information has the following property*

$$I(A; B, C)_\rho = I(A; B)_\rho + I(A; C|B)_\rho \tag{B.3}$$

Definition B.10 (Relative Entropy). *For two distributions p and q on a set A, the relative entropy is defined as*

$$D(p \parallel q) \coloneqq \sum_{a \in A} p(a) \log\left(\frac{p(a)}{q(a)}\right).$$

Definition B.11 (Holevo Quantity). *For a quantum channel $W : \mathcal{X} \to \mathcal{S}(\mathcal{H})$ and $p \in \mathcal{P}(\mathcal{X})$, the Holveo quantity is defined as*

$$\chi(p, W) \coloneqq H\left(\sum_{x \in \mathcal{X}} p(x) W(x)\right) - \sum_{x \in \mathcal{X}} p(x) H\left(W(x)\right).$$

Bibliography

[1] R. Ahlswede and G. Dueck - *Identification via Channels* - IEEE Trans. Inf. Theory, vol. 35, no. 1, pp. 1529, 1989

[2] R. Ahlswede, I. Althöfer, C. Deppe, U. Tamm - *Rudolf Ahlswede's Lectures on Information Theory* - Preprint

[3] R. Ahlswede, G. Dueck - *Identification in the Presence of Feedback - A Discovery of New Capacity Formulas* - IEEE Transactions on Information Theory, vol. 35, no. 1, pp. 30-36, Jan 1989. doi: 10.1109/18.42173

[4] P. Löber - *Quantum Channels and Simultaneous ID Coding* - arXiv:quant-ph/9907019v1

[5] M. Wilde - *From Classical to Quantum Shannon Theory* - arXiv:quant-ph/1106.1445v7

[6] A. Winter - *The Capacity of the Quantum Multiple-Access Channel* - arXiv:quant-ph/9807019v3

[7] Y. Steinberg - *New Converses in the Theory of Identification via Channels* - IEEE Transactions on Information Theory, vol. 44, no. 3, pp. 984-998, May 1998.

[8] S. Verdú - *Multiple-Access Channels with Memory with and without Frame Synchronism* - IEEE Transactions on Information Theory, vol. 35, no. 3, pp. 605-620, May 1989

[9] S. Moser - *Advanced Topics in Information Theory (Lecture Notes) v. 3.0* - http://moser-isi.ethz.ch/cgi-bin/request_script.cgi?script=atit_script_v30.pdf&version=3.0

[10] I.Csiszár, J. Körner - *Information Theory - Coding Theorems for Discrete Memoryless Systems*

[11] A.Winter - *Identification Via Quantum Channels* - arXiv:1212.0494v1

[12] G. Dueck - *Maximal error capacity regions are smaller than average error capacity regions for multi-user channels*

[13] D. Brannan - *A First Course in Mathematical Analysis*

[14] T.S. Han, S. Verdú - *Approximation Theory of Output Statistics* - IEEE Transactions on Information Theory, vol. 39, no. 3, pp. 752-772

[15] H. Boche, C. Deppe, A. Winter - *Secure and Robust Identification via Classical-Quantum Channels* arXiv:1801.09967v1