Classical Simulations of Quantum Prepare and Measurement Communication Processes

by

Chai Fillerup

Student Number: 5137349
Date: October 2022

Supervisors:	Dr. D. de Laat
	Dr. J. Borregaard
Other members assessment committee:	Prof. Dr. Y.M. Blanter
	Prof. Dr. F.H.J Redig

TU Delft

Abstract

Quantum communication has been shown to be vastly superior to classical communication in many problems. However no general statements exist which tells us how much better quantum communication is to its classical counterpart. In this thesis it was studied the minimum amount of classical bits required to exactly simulate a quantum communication process. The quantum communication process specifically studied was a quantum prepare and measurement communication problem. It has been shown that the calculation of the amount of classical bits of communication required for simulation reduces to a minimization-maximization optimization problem. Several results have been presented for for solving this optimization problem and in addition a link was made between classical simulations of quantum communication and a recent debate on the reality of the quantum state.

Contents

1. Introduction

Quantum computing has been shown to be better than classical computing at many different tasks. But there are theoretical limits to how much more efficient quantum computers can be in certain tasks. For example it has been shown that quantum computers are unable to compress data more efficiently than a classical computer. With quantum internet becoming more and more advanced this then raises the question of how more efficient a quantum computer is at communication tasks than a classical computer. Efficient communication between different devices is important for tasks where data is divided over several devices which need to communicate to solve a given task. That is why in this thesis the results from a series of different papers on the subject of comparing classical and quantum communication will be presented.

The most general communication problem known to us is the communication complexity problem. This problem as depicted in Figure 1 has two parties, Alice and Bob, which both receive a set of data in the form of bit strings X and Y. Their job is then for Bob to calculate the value of function dependent on both data sets $f(x, y)$ while communicating as little as possible. The exact form of $f(x, y)$ depends on which communication problem one is studying. How much communication takes place is quantified by counting how many bits both parties exchange during the entire process.

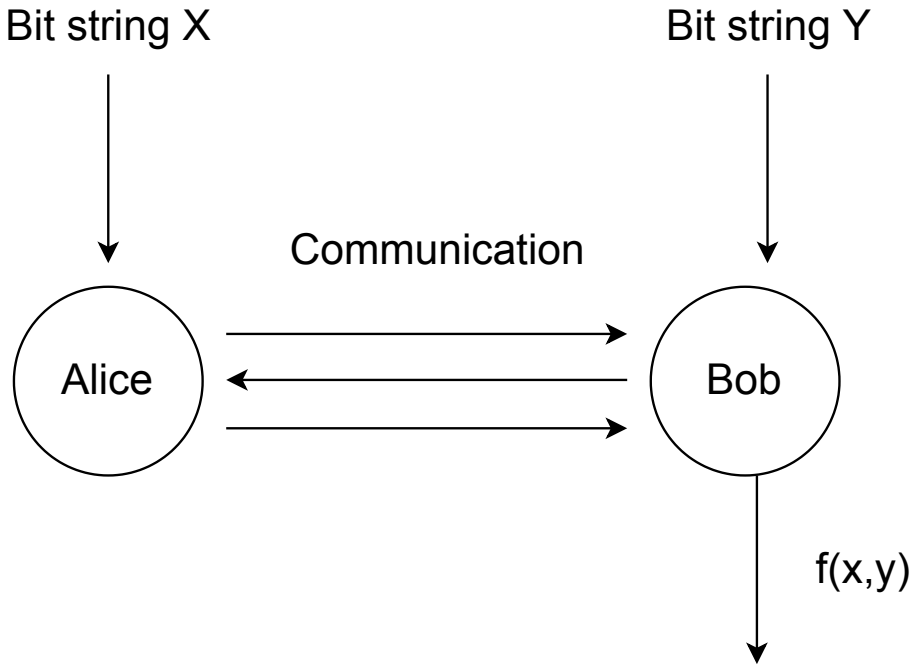


Figure 1: The general communication complexity problem where Alice and Bob receive a bit string X and Y and their job is to calculate $f(x, y)$ while communicating as little as possible.

Above the classical case for the communication complexity problem has been presented

where Bob and Alice are constrained to communicating using bits. But with the advent of quantum computing multiple methods have been found where Alice and Bob can use quantum effects to make their job easier. While plenty of models of quantum communication exist today the one we will be focusing on today is the case Alice and Bob no longer communicate with classical bits but instead start using quantum bits or rather qubits. While it has been shown that qubits are not better than regular bits at data compression, as 1 bit of classical data still requires 1 qubit to store, using clever algorithms one can make an assortment of communication tasks easier.

The main problem with directly comparing quantum communication and classical communication is that both utilise different resources. In classical communication problems the amount of classical bits communicated should be minimized while in quantum computation one tries to minimize the amount of qubits used. The difficulty being that there is no one way to say that a qubit is equal to a certain amount of classical bits. However it has been shown that there are communication problems where one uses exponentially less qubits than classical bits to solve certain communication problems so we know qubits are more efficient than regular bits. So in this thesis it will be explored how many classical bits would be required to exactly simulate a generic quantum communication process which then tells us how much more efficient quantum communication is than classical communication.

In this thesis all the necessary mathematics will be presented to evaluate the amount of classical bits needed to simulate quantum communication processes. In section 2 the key mathematical concepts required for classical communication are presented and these are expanded upon in section 3 by allowing a new kind of communication problem. In section 4 necessary concepts from quantum mechanics are introduced and this all culminates in section 5 where all mathematical theory touched upon in sections 2 and 3 are applied to quantum communication. Lastly in section 6 theoretical applications of this problem are applied to quantum ontological theories and in section 7 a discussion of the key points presented in this thesis is presented.

This research has been done as a bachelor thesis to complete the bachelors of science Applied Physics and Applied Mathematics at the Delft University of Technology.

2. Classical Information Theory

Before we can examine how to simulate qubits using classical bits, we must first examine how communication works via classical bits. To understand this we will explore the concepts of entropy, encodings and communication channels. This is all in buildup towards the main theorem of communication theory: the noisy channel coding theorem, published by Claude E. Shannon in 1948 [1]. With this theorem we have all we need to describe all classical communication processes and thus how to simulate quantum communication.

2.1 What is Information?

A big question we will start of with is what exactly is meant by communication? It is a vague term and can have plenty of interpretations. In the broadest sense communication is about distributing information, one party has information about a certain event and wants to share it with another party. In general communication theory the event that is being communicated about is the outcome of a stochastic variable. What we would like to know is: "How much communication is necessary to inform a second party about the outcomes of a stochastic variable?". As communication is the sharing of information we would first like to define how much 'information' is stored in a stochastic variable.

2.1.1 Information & Entropy

To define communication processes we must define a mathematical concept of information, as communication entails the sharing of information between two parties. We view gaining information as learning more about the possible outcomes of a random process. Say Bob rolls a six sided die in secret and tells Alice that the outcome is even, then Alice gains information on the die as 3 outcomes are discarded (1, 3 and 5). We would like to quantify exactly how much information Alice has gained by knowing the result is even.

Gaining information is thus about reducing the amount of possible outcomes a stochastic variable can have. Similarly we can also ask how much information we gain once we know what specific outcome a stochastic variable takes, as taking a specific value is simply reducing the space of possible outcomes to 1 outcome. So we define the information content of a single outcome as being how much information is contained in the removal of every other outcome. Using this definition we expect our measure of information content to satisfy the following 3 properties:

1. An event with probability 100% yields no information.
2. The less probable an event is the more information it yields.
3. If two independent events are measured separately, the total amount of information is the sum of the surprise of the individual events.

The first property is easy to grasp, if an event has 100% probability then all other events have 0%, so we do not learn anything new about the system by removing events which could not occur. To explain the second property we will provide an example. Imagine a twenty sided die and we learn that the die has taken a value less than 5 which has probability 20%,. This event reduces our space of possible outcomes to just 4 values down from 20. However

if we had only been told the die had taken a value less than 11, which has probability of 50%, we would reduce our space of possible outcomes to 10 elements thus we have learned less.

The third property ensures that we do not learn more or less about a stochastic variable X by viewing the outcome of an unrelated stochastic variable Y, one does not learn about the outcome of a die by rolling another die. Thus we require that the information contents simply sum and if we decide we no longer care about event y occuring we subtract the information we got from it and have the same information we have from event x as if event y never occured.

Definition 2.1 (Information content)**.** *Given a stochastic variable X with a set of possible outcomes $\{x_1, x_2, \ldots, x_n\}$, we denote the information content of the realision x_i as:*

$$I(x_i) = -\log_2(P(x_i)). \tag{2.1}$$

The above measure for information content satisfies all our axioms and as it turns out is the only function which does so, up to multiplication by a constant. This constant is determined by what value we take as base for the logarithm. Different applications of information theory use different bases for the logarithm which also changes in which unit information is measured. In base 2 we measure information in bits as mentioned in the definition, but in base 3 it is measured in trits and in base e in nats. From now on every logarithm is taken in base 2 thus all information is measured in bits.

The term bits can be confusing as they already have a seemingly different meaning in computer science: the binary digits which computers use for calculations. The bits in which information is measured seem to have nothing to do with the binary digits used by computers but there is a connection. This will be explored in later sections, so for now think of bits as an abstract form of representing the information of a variable and completely distinct from the binary digits used by computers.

Definition 2.2 (Entropy)**.** *The entropy, also referred to as Shannon's entropy, of a stochastic variable is given by:*

$$\mathrm{H}(X) = \mathbb{E}(I(X)) = -\sum_i P(x_i)\log(P(x_i)). \tag{2.2}$$

Shannon entropy is thus the expected amount of information contained in a stochastic variable. The name entropy was chosen for this quantity as there is a direct link between Boltzmann entropy and Shannon entropy, we will not go in depth proving this but for those familiar with the former quantity: Shannon entropy reduces to Boltzmann entropy (differing only by a constant) when the microstates are all equiprobable. For those interested further exploration of this relation can be found in [2].

An interesting question is then what probability distribution maximizes the entropy. This question can be answered with two facts, the first of which is that the function $-x\log(x)$ is a concave function and the second is the following theorem about concave and convex functions[3]:

Theorem 2.1 (Jensen's inequality)**.** *Given a convex function $\phi : [a, b] \to \mathbb{R}$ with domain $[a, b] \subseteq \mathbb{R}$. Let $y_i \in [a, b]$ be a sequence of points, for any positive weights $a_i \in \mathbb{R}$ we have that*

$$\phi\left(\frac{\sum a_i x_i}{\sum a_i}\right) \leq \frac{\sum a_i \phi(x_i)}{\sum a_i}. \tag{2.3}$$

And for concave functions we have

$$\phi\left(\frac{\sum a_i x_i}{\sum a_i}\right) \geq \frac{\sum a_i \phi(x_i)}{\sum a_i}. \tag{2.4}$$

We will now apply Jensen's inequality to the entropy of a stochastic variable X with N possible outcomes. Let the concave function be $x \mapsto -x\log(x)$ with the sequence of points being given by $y_i = P(x_i)$ and weights $a_i = 1/N$ with $N \in \mathbb{N}$, putting this all into Eq. 2.4 we get that

$$-\left(\frac{\sum \frac{1}{N} P(x_i)}{\sum \frac{1}{N}}\right) \log\left(\frac{\sum \frac{1}{N} P(x_i)}{\sum \frac{1}{N}}\right) \geq \frac{\sum \frac{1}{N} P(x_i)\log(P(x_i))}{\sum \frac{1}{N}}. \tag{2.5}$$

The above inequality can then be simplified into the following form:

$$\log(N) \geq H(X). \tag{2.6}$$

This then holds true for any discrete probability distribution as it holds true for any N and any sequence of probabilities $P(x_i)$. In addition one can easily check that equality is achieved for uniform distributions. So the distribution with the highest entropy is the uniform distribution. In fact it is the only distribution which has maximum entropy $\log(N)$ [4, p. 29].

2.1.2 Multivariable Entropies

As not all probability distributions are functions of one variable we will need to define entropy for multivariate distributions. There are two more types of entropy for different multivariate distributions of which the definition is similar to that single variable entropy but they are still important enough to give special attention to.

Definition 2.3 (Joint entropy)**.** *The joint entropy of two stochastic variables X and Y with possible outcomes given by the sets $\mathcal{X}$ and $\mathcal{Y}$ respectively is given by:*

$$\mathrm{H}(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(x,y). \tag{2.7}$$

The joint entropy describes the informational content of the two variables together. Two important properties of the joint entropy are:

Property 2.3.1. *The joint entropy of two stochastic variables X and Y is lower or equal to the sum of the individual entropies, i.e.* $\mathrm{H}(X,Y) \leq \mathrm{H}(X) + \mathrm{H}(Y)$.

Property 2.3.2. *The joint entropy of two stochastic variables X and Y is larger or equal to the maximum of the individual entropies of both variables, i.e.* $\max\{\mathrm{H}(X), \mathrm{H}(Y)\} \leq \mathrm{H}(X,Y)$.

Both properties lend themselves to nice intuitive explanations. the first property is true as the 2 variables can be correlated. So if variable Y contains some information about X we want the joint entropy to be lower than the sum of individual entropies. The second property states that we do not expect the information content of a stochastic variable to decrease as we add another variable, even if they are correlated adding more variables will never tell us more about the original variable X, similarly for Y.

Definition 2.4 (Conditional entropy)**.** *The conditional entropy of two stochastic variables X and Y with images $\mathcal{X}$ and $\mathcal{Y}$ is given by:*

$$\mathrm{H}(Y|X) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log\left(\frac{p(x,y)}{p(x)}\right). \tag{2.8}$$

A more intuitive form for the conditional entropy is:

$$\mathrm{H}(Y|X) = \mathrm{H}(X,Y) - \mathrm{H}(X). \tag{2.9}$$

Now we can see that the conditional entropy $H(Y|X)$ is the surprise left in Y after we remove the surprise in X by seeing the outcome of X.Property 2.3.1 shows us then that it must be true that $\mathrm{H}(Y|X) \leq \mathrm{H}(Y)$.

2.1.3 Mutual Information

To finish off this section we will look at one last quantity which plays a major role in communication processes.

Definition 2.5 (Mutual information)**.** *The mutual information of two stochastic variables X and Y is*

$$\mathcal{I}(X,Y) = \mathrm{H}(X) - \mathrm{H}(X|Y) = \mathrm{H}(X,Y) - \mathrm{H}(X|Y) - \mathrm{H}(Y|X) \tag{2.10}$$

The mutual information gives us a measure of how much information we gain about one variable by measuring the result of the other variable. Some important properties of the mutual information which we want to pay extra attention to are:

Property 2.5.1. *The mutual information is convex in P(Y|X).*

Property 2.5.2. *The mutual information is symmetric, that is $\mathcal{I}(X,Y) = \mathcal{I}(Y,X)$.*

In addition to these properties the mutual information also has an important inequality attached to it named the data-processing inequality [4, p. 34-35]. Before we introduce the inequality we have to introduce the concept of a Markov chain.

Definition 2.6 (Markov chain)**.** *Three stochastic variables X, Y, Z are said to form a Markov chain $X \to Y \to Z$ if the conditional probability of Z given Y is conditionally independent of X.*

A common example of a Markov chain is $Z = f(Y)$, as the values for the variable Z are entirely defined by Y where Y is a variable correlated with X. The data processing inequality states

Theorem 2.2 (Data-processing inequality)**.** *Given a Markov chain $X \to Y \to Z$ the mutual information satisfies*

$$\mathcal{I}(X,Y) \geq \mathcal{I}(Y,Z). \tag{2.11}$$

Applying this to our previous example of a Markov chain one finds that performing any transformation $f(Y)$ on the data of a stochastic variable Y will not provide any additional information on the stochastic variable X.

Now with a solid grasp on what information means we can now move on to the more complicated task of communicating information about stochastic variables to external parties. The quantities introduced in this chapter will play an important part in communication theory.

2.2 Communicating Information

Now that we know how much information is contained in a stochastic variable we would like to know how one goes about the process of communicating events of the variable to the outside world. In general we want to know how to mathematically describe the process of one party, called Alice, measuring the outcome of a stochastic variable and sending this to a second party, called Bob, such that Bob knows with 100% certainty what the outcome of the variable was.

2.2.1 Encoders & Decoders

Before Alice and Bob can start communicating over long distances they need to get together and agree on a language to translate the outcomes of the stochastic variable into. This is done via encoders and decoders which map outcomes to a set of agreed upon symbols which Alice and Bob both know beforehand. We also want to define a measure which tells us how efficient a given encoding and decoding scheme is.

In a given communication problem there is a finite set of possible messages $\mathcal{M}$ Alice can send to Bob. Bob knows the entire content of $\mathcal{M}$ but does not know which specific message Alice has received. Beforehand Alice and Bob agree on a finite alphabet $\mathcal{X}$ which defines what kind of symbols Alice is allowed to use to describe the outcome of a stochastic variable, this is called the **source code alphabet**. Strings of these symbols are called **source words**. So to effectively communicate with Bob Alice has to encode her message in terms of her source code alphabet.

Definition 2.7 (Encoder)**.** *An encoder is a map $E^n : \mathcal{M} \to \mathcal{X}^n$ that translates messages into codewords.*

Similarly there is a finite alphabet $\mathcal{Y}$ of symbols which defines what Bob receives from Alice. It is important that we do not assume that Bob and Alice use the same alphabet as it could happen that whatever medium Alice and Bob use changes the message through noise, which then changes what kind of symbols Bob could receive. This will be elaborated upon further later when dealing with noisy communication. For now all Bob has to do the reverse of Alice and decode a message from $\mathcal{Y}$ to the original set $\mathcal{M}$.

Definition 2.8 (Decoder)**.** *A decoder is a map $D^m : \mathcal{Y}^n \to \mathcal{M}$ that maps source words back to messages.*

We define a combination of a certain set of messages with a given encoder and decoder as a **coding scheme**. We define the efficiency of a coding scheme as the amount of information being transmitted divided by the amount of symbols used by Alice to encode the variable.

Definition 2.9 (Coding rate)**.** *The rate R of a coding scheme $(\mathcal{M}, E^n, D^n)$ is defined as*

$$R = \frac{\log(|\mathcal{M}|)}{n} \tag{2.12}$$

The rate as viewed from the definition is equal to the amount of information transmitted divided by the amount of symbols Alice has to use to encode this information. We take the amount of information transmitted to be equal to $\log(|\mathcal{M}|)$ as we assume a worst case scenario, which is when the input distribution of the variable taking values in $\mathcal{M}$ is the uniform distribution. Rate is often given in bits/second, as it is often known in practical applications how many symbols Alice can send to Bob per second.

2.2.2 Shannon's coding theorem for symbol codes

A very important question in data compression is what the minimal amount of bits is to encode all the events of a stochastic variable? For that we need to know what properties an encoding scheme must satisfy to be uniquely decodable. A common type of coding scheme is called a prefix coding scheme and satisfies the following properties:

1. Every codeword corresponds to a single outcome.
2. No codeword is a prefix of any other codeword.

This second property is required purely for the encoding of multiple outcomes one after the other. For instance, when you have a binary encoder using the codewords $0, 1, 10$ for a 3 outcome variable, then it is not clear what is meant by the sequence of symbols 1010.

These types of codes are referred to as **prefix-free codes**. For all symbol codes, including prefix-free codes, we have the following theorem which tells us the expected length of the code and finally gives us an operational meaning of Shannon entropy. Now the main type of codes we are interested in are known as **uniquely decodable codes**. These codes satisfy that after Alice has encoded a message and send it to Bob that Bob has a 0% chance that Bob has a different message after encoding.

Theorem 2.3 (Source coding theorem for symbol codes)**.** *Given two alphabets $\mathcal{X}$ and $\mathcal{Y}$, let $\mathcal{X}^*$ and $\mathcal{Y}^*$ be the set of all finite words composed out of symbols from their respective alphabets. Let X be a stochastic variable taking values in $\mathcal{X}$ and let $f : \mathcal{X}^* \to \mathcal{Y}^*$ be a uniquely decodable code. If f is an optimal encoding, then the expected length S of the encoding satisfies*

$$\frac{H(X)}{\log(|\mathcal{Y}|)} \leq S \leq \frac{H(X)}{\log(|\mathcal{Y}|)} + 1 \tag{2.13}$$

The simplest example of applying this theorem is that of a fair coin, which has entropy of 1 bit and can take two values we want to encode 2 possible messages. Thus if we want to encode the coin into an alphabet with two symbols the lower bound specifies that the expected message length is at least 1 symbol. Of course the optimal method for encoding the results of a fair coin is labelling one result 0 and the other 1, which has an expected message length of 1 symbol.

The +1 in the upper bound is necessary as there are situations where it is impossible to get the expected code length exactly equal to its lower bound but we still want to know something about how many symbols are necessary. For example, given a stochastic variable X taking values in $\{x_1, x_2, x_3, x_4\}$ with respective probabilities $\{0.7, 0.26, 0.02, 0.02\}$. X has an entropy of 1.09 bits so the expected code length S satisfies $0.55 < S < 1.55$. The most efficient prefix free code one can use is $\{0, 10, 110, 111\}$ for the events $\{x_1, x_2, x_3, x_4\}$, here we expect to send 1.34 binary digits on average.

The presented theorem is a weaker version of Shannon's source coding theorem which applies to all possible encodings, not just symbol codes. Shannons source coding theorem essentially states that the lower bound in Theorem 2.3 is true for every possible encoding one can think of, not just the prefix-free codes presented here.

2.3 Communication Channels

The only thing missing now for Alice and Bob to communicate is the method to go from Alice's alphabet $\mathcal{X}$ to Bob's alphabet $\mathcal{Y}$. This mapping from one alphabet to the other is

referred to as the **communication channel** used by Alice and Bob. A communication channel can be viewed as a conditional probability distribution where we give probabilities of all possible outputs given a certain input. A probability distribution is used as we want to allow for the possibility that the channel is noisy, that is that the message can get disturbed in the process of sending it. The question now is, given a communication channel, is there an intrinsic limit to how much information can be transmitted in a single use of said channel?

2.3.1 Channel Capacity

To make studying channels a bit easier we will limit ourselves to a specific type of channel. The limitation will be that we the probability distribution defining the communication channel only depends on current input and not on previous or future inputs, when a channel satisfies this requirement it is called a **memoryless channel**. From here on out it is assumed that every channel is memoryless. This means that we can define a channel by a probability distribution $P(y|x)$ where y represents the output and x any input. Often we are required to input strings of the input variable x into our channel but using that we only allow the output to depend on the input we can write

$$P(x_1, \ldots, x_n|y_1, \ldots, y_n) = P(x_1|y_1) \ldots P(x_n|y_n) \tag{2.14}$$

A simple example of a discrete memoryless channel is known as the noiseless binary channel (NBC) as shown in Figure 2. Here the input and output alphabet contain just two symbols $\{0, 1\}$ and is defined by the probability distribution $P(y|x) = \delta_{xy}$. This is what is known as a noiseless channel, there is 0% chance that Bob decodes the wrong message from Alice as long as the coding scheme is unambiguous.

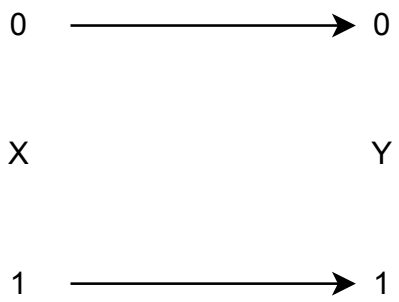


Figure 2: The simplest example of a communication channel: the noiseless binary channel. The source code alphabet for sender and receiver are both $\{0, 1\}$ and the output of the channel is equal to the input.

When Alice and Bob use a communication channel we imagine that Alice will use her encoder to encode her message in n symbols and then proceed to send these symbols one by one to Bob. Bob then reassembles the message and uses his decoder to interpret Alice's original message. However in the example of the binary channel we know that not every coding rate is possible for a given coding scheme, it will be impossible for Alice to send Bob 10 bits of data using the channel only once without introducing the chance that error occurs. Put differently, a communication channel has an inherent limit to which communication rates do not introduce error. For the binary channel this limit is obvious, it should be

equal to the limit given by Theorem 2.3. However we lack a way of evaluating this limit of information rates for a noisy channel.

To calculate this maximum rate of a channel we will first define a new, at first seemingly unrelated quantity, called the channel capacity.

Definition 2.10 (Channel capacity)**.** *Given a channel $P(y|x)$. For a random variable X with distribution $P_X(x)$ let Y be the variable with distribution*

$$P_Y(y) = \sum_X P_X(x)P(y|x). \tag{2.15}$$

The channel capacity of this communication channel is then equal to

$$C = \max_{P_X(x)} \mathcal{I}(X, Y) \tag{2.16}$$

where the maximization is over the space of possible distributions for X.

The channel capacity turns out to be equal to maximum coding rate we wanted to find previously. This was shown by Shannon in 1948 in Shannon's noisy channel coding theorem or sometimes referred to as simply Shannon's theorem.

Theorem 2.4 (Shannon's noisy channel coding theorem)**.** *Given a communication channel with channel capacity C. Then there exists a sequence of coding schemes, encoding a set of messages $\mathcal{M}$ with encoder E^n and decoder D^n, with rates $R < C$ which satisfy that the maximum probability of faulty decoding, denoted $\lambda^{(n)}$, satisfies that $\lambda^{(n)} \to 0$ as $n \to \infty$.*

Conversely any such sequence of codes with rates $R > C$ must have that the probability of faulty decoding cannot go to 0 as $n \to \infty$.

Shannon's noisy channel coding theorem gives us a way to let channels with noise simulate channels without noise. While the proof of the theorem is constructive, as in it gives a method to construct an algorithm to achieve such rates for any channel, the given algorithm is extremely computationally heavy so the discovery of more efficient algorithms for noisy channels remains an important area of research. To check if the theorem matches up with expectations, the channel capacity of the channel in Figure 2 is equal to 1 bit, where the maximization was achieved by using a uniform distribution. So it is not possible to use coding schemes with rate higher than 1 bit/second with the noiseless binary channel without introducing error.

An important fact to note is that when someone is using a noisy channel a person is not limited to sending less symbols through the channel. We are only limited in how much information we can send through the channel. To actually send a message through a noisy channel Alice has to add additional symbols to her message to ensure that Bob can properly decode the message without chance of error. The amount of symbols needed to send a message of length k through a noisy channel is equal to k divided by the channel capacity. Several examples of noisy channels and their capacities are given below but these examples will not be of major importance for the rest of this thesis.

2.3.2 Binary Symmetric Channel

An easy channel which is not noiseless is the binary symmetric channel (BSC), as represented in Figure 3. This channel uses the same alphabets as the NBC but the key difference is with

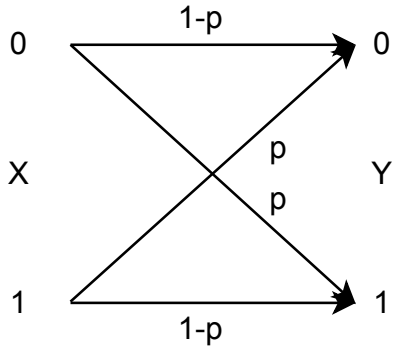


Figure 3: Binary symmetric channel. It uses the binary alphabet for input and output but has with every transmission there is a chance p that a bit gets 'flipped', a 0 becomes a 1 and a 1 becomes a 0.

every transmission there is a chance p that a bit flip occurs, that is a 0 turns into a 1 and vice versa. The channel capacity of this quantity can be derived as follows:

$$\begin{aligned} C &= \max_{P(x)} \mathcal{I}(X,Y) \\ &= \max_{P(x)} \mathrm{H}(X) - \mathrm{H}(X|Y) \\ &= 1 - \mathrm{H}(X|Y). \end{aligned}$$

The entropy in X is maximized with the uniform distribution and the conditional entropy does not depend on the input distribution as we will show now. For this we write that Y takes the value 0 with chance q, and we introduce the notation $\mathrm{H}(p)$ to be the entropy of a 2 outcome stochastic variable with respective chances of occuring p and $(1-p)$.

$$\begin{aligned} &1 - \max_{P(X)} \sum p(y)\mathrm{H}(X|Y=y) \\ &= 1 - \max_{P(X)} (q\mathrm{H}(X|Y=0) + (1-q)\mathrm{H}(X|Y=1) \\ &= 1 - \max_{P(X)} (q\mathrm{H}(p) + (1-q)\mathrm{H}(p)) \\ &= 1 - \mathrm{H}(p). \end{aligned}$$

So the maximum achievable rate for the binary symmetric channel is $1 - \mathrm{H}(p)$. This means that the binary symmetric channel with $p = 1/2$ is completely unusable.

2.3.3 Binary Deletion Channel

In the binary deletion channel, as seen in Figure 4, a fraction p of the bits will be deleted. The receiver however does know which bits are deleted so the channel can be seen as having

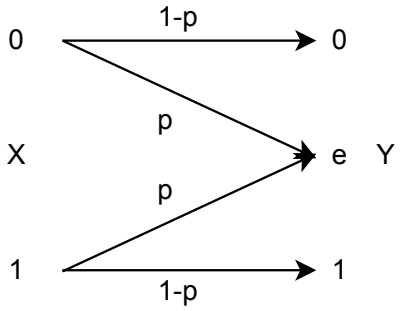


Figure 4: The binary deletion channel (BDC), every bit send through has a chance p to get erased. The receiver knows when a bit is erased so it can be represented by a symbol 'e' in the output.

three outputs: 0, 1 and e for erased. We calculate the channel capacity as follows:

$$\begin{aligned} C &= \max_{p(x)} \mathcal{I}(X;Y) \\ &= \max_{p(x)} (\mathrm{H}(Y) - \mathrm{H}(Y|X)) \\ &= \max_{p(x)} (\mathrm{H}(Y) - \mathrm{H}(p)). \end{aligned}$$

Calculating the entropy of Y is easiest by defining new variables E and Y^* with respective events $\{Y = e, Y \neq e\}$ and $\{Y = 0, Y = 1\}$. Then we can write the entropy of Y into the following form

$$\mathrm{H}(Y) = \mathrm{H}(Y^*, E) = \mathrm{H}(E) + \mathrm{H}(Y^*|E). \tag{2.17}$$

In the next step we will use that the conditional entropy can be rewritten into the form

$$\mathrm{H}(Y|X) = \sum_{x \in X} P(x) H(Y|X = x). \tag{2.18}$$

So we can now calculate both entropies on the right hand side of Eq. 2.17 and we can see that:

$$\mathrm{H}(E) = \mathrm{H}(p)$$
$$\mathrm{H}(Y^*|E) = P(Y = e)\mathrm{H}(Y^*|Y = e) + P(Y \neq e)\mathrm{H}(Y^*|Y \neq e) = p \cdot 0 + (1 - p)\mathrm{H}(q).$$

Where q is the chance that X=0. Subsituting these expressions for the initial expression for C we see that:

$$\begin{aligned} C &= \max_{q} \mathrm{H}(Y) - \mathrm{H}(p) \\ &= \max_{q} (1 - p)\mathrm{H}(q) + \mathrm{H}(p) - \mathrm{H}(p) \\ &= 1 - p. \end{aligned}$$

Where maximization is achieved by q = $\frac{1}{2}$.

In the situation where there is feedback from the receiver from the sender this capacity makes sense. Every time a bit has been corrupted resend the bit, since there is a chance 1-p that a bit gets through this translates to an effective capacity of 1-p. However Shannon's noisy channel coding theorem states that even without feedback an effective rate of 1-p should be possible.

2.3.4 Additivity of Channel Capacities

A small result we will use in later chapters about channel capacities is that when we have 2 independent channels that when they are used together that the channel capacities add.

Theorem 2.5 (Additivity of channel capacities)**.** *Given 2 channels $\mathcal{N}_1$ and $\mathcal{N}_2$. $\mathcal{N}_1$ has input alphabet $\mathcal{X}_1$, output alphabet $\mathcal{Y}_1$, $\mathcal{N}_2$ is defined similarly. We define the product channel $\mathcal{N}$ as having input and output alphabets $(\mathcal{X}_1, \mathcal{X}_2)$ $(\mathcal{Y}_1, \mathcal{Y}_2)$ respectively. The transfer probability of $\mathcal{N}$ is $p_1(y_1|x_1)p_2(y_2|x_2)$ with both distributiosn being the transfer probabilities of their respective channels. Then*

$$C(\mathcal{N}) = C(\mathcal{N}_1) + C(\mathcal{N}_2) \tag{2.19}$$

Intuitively this can be seen as Alice and Bob using two separate channels at once and then it makes sense that the channel capacities add, as when combining two noiseless binary channels we expect to be able to send 2 bits of information at a time.

2.4 Reverse Shannon Theorem

Shannon's noisy channel coding theorem gives us a way for a noisy channel to simulate a noiseless channel, now we would like to know if it is possible for a noiseless channel to simulate a channel with noise. This might seem useless as there is no reason why one might want to make a noiseless channel noisy, however it has useful applications in quantum communication theory. More importantly to some it gives the intuitive result that there is no real difference between noiseless and noisy channels, as with clever coding schemes both types of channels can simulate the other [5].

Theorem 2.6 (Reverse Shannon theorem)**.** *Let $\mathcal{N}$ be a discrete memoryless channel with channel capacity C and let $\epsilon > 0$. Then for each n there exists an exactly faithfull simulation protocol S_n simulating n uses of $\mathcal{N}$ while using a noiseless classical channel and prior random information shared between Alice and Bob. The number of bits of forward communication used by S_n on channel input $x \in \{1 \cdots d_I\}^n$ is a random variable, denoted $m_n(x)$. The simulation is exactly faithful in the sense that*

$$\forall_{nxy} P_{S^n}(y^n|x^n) = P_{N^n}(y^n|x^n). \tag{2.20}$$

In addition the protocol is asymptotically efficient in the sense that the probability that the protocol uses more than $n(C + \epsilon)$ bits of forward communication approaches zero in the limit of large n,

$$\lim_{n\to\infty} \max_{x\in\{1\cdots d_I\}^n} P(m_n(x) > n(C + \epsilon)) = 0. \tag{2.21}$$

We see that this simulation has to satisfy two properties. The first is that the conditional probabilities of the outputs given a certain input should be equal to that of the noisy channel. The second condition is a bit more complex and will be explained using the BSC as example. When sending n bits through the BSC, we know that a message of n bits was encoded. But if we want to simulate the BSC with a NBC then we can't send n bits through the NBC,

because then we encode a message of n bits as the NBC has a channel capacity of 1. So to simulate the BSC with the NBC we can send nC bits through it at most. The ϵ term in the proof denotes an error term and the theorem is true for every $\epsilon > 0$.

3. Two Input Communication Problem

To simulate a quantum communication problem we will see that both Alice and Bob both get an input, Alice a quantum state and Bob a quantum measurement. This is an example of a **2 input communication problem** where both parties get an input. We will study such communication problems without using properties specific to quantum mechanics in this section. We will then try to find how how much communication is required to classically simulate 2 input communication problems. In later sections we will start using quantum effects more to try and solve the problem for quantum prepare and measurement problems.

3.1 The Black Box

The problem is as follows: Alice receives an input a, sends this to Bob and he proceeds to alter the input a dependent on his received input b and outputs a value s. Note that s does not have to be determined deterministically so for any a and b there can be multiple possible outputs s. What we would like to know is how many bits are required to classically simulate this process. As previously stated we will not assume that we know what a, b and s actually are. They are general inputs and can be anything from quantum states to encoding and decoding schemes. This means that we cannot assume that it even is possible to use a classical channel to fully encode all information about a (we will later see that this is the case with qubits). To work around this we avoid dealing with a, b and s by pulling a 'black box' over the problem and represent it as a relatively simple probability distribution $P(s|a, b)$.

Before we start with simulating a black box we will simplify the problem by making several assumptions. It is assumed that a and b are elements of finite sets A and B, which are both known to Alice and Bob beforehand, in addition we assume that for any given a and b the amount of possible outcomes s is also finite. In later sections we will analyze the problem where we allow uncountably many a and b but the assumption that there are finitely many possible outcomes given two inputs will always be there. We introduce the notation where we describe a black box game as a 3-tuple $\mathbf{G}$=(P, A, B) , defined by the probability distribution $P(s|a, b)$ and the sets of possible inputs A and B.

To simulate the black box we will follow the process as described in [6], which is also represented in Figure 5. For this process Alice and Bob have the option to make use of a shared random variable Y which is generated in advance and is uncorrelated with both inputs. With possible use of this random variable Alice can map every input a to a message k with probability $\rho(k|y, a)$ which she proceeds to send over a noiseless channel to Bob. Bob then can decode this message k with use of his input b and finally generates an output s with probability $\rho(s|k, y, b)$.

This protocol exactly simulates the black box $P(s|a, b)$ if the probability of outputting s via this simulation is the same of generating it via the distribution $P(s|a, b)$ that is, if

$$\sum_k \int \rho(s|k, y, b)\rho(k|y, a)\rho(y)dy = P(s|a, b) \qquad \forall a, b, s. \tag{3.1}$$

To calculate the cost of our classical simulation we would like to know the average amount of information over all inputs a send through our classical channel. There are several distinct ways to measure the average length of the message but the one we will look

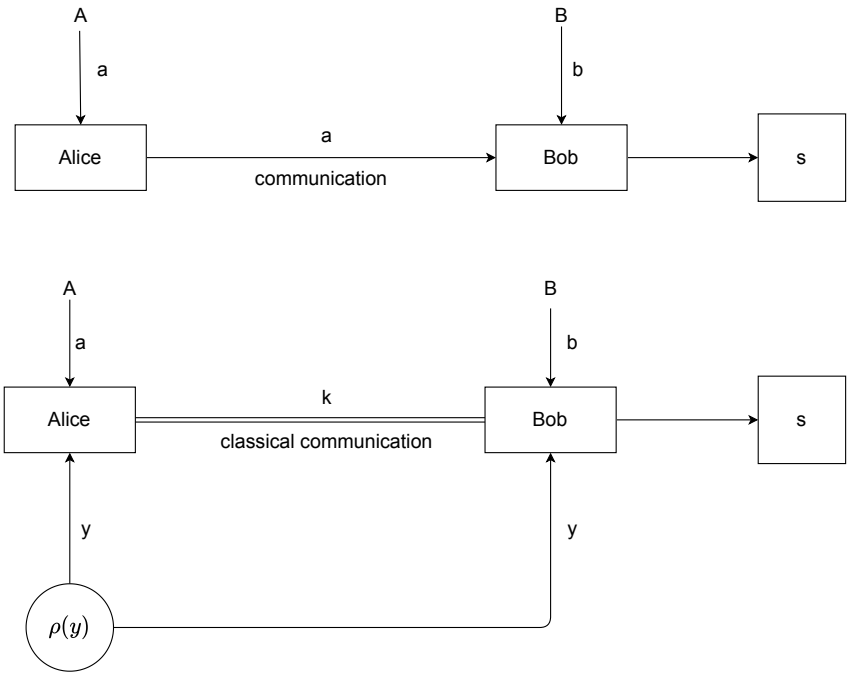


Figure 5: (a) The standard 2 player communication game, Alice receives input a and Bob receives input b. Alice sends a to Bob and he proceeds to output s. (b) The classical simulation of (a), instead of sending a Alice encodes A into a codeword k which can be send over a classical channel to Bob. For generating a and the final output s both parties can make use of a shared random variable y.

at is the average expected message length before Alice gets her input a but after she and Bob got the value for Y. Then we can calculate the average communication cost as

$$C = \max_{\rho(a)} \mathrm{H}(K|Y). \tag{3.2}$$

Shannon's source coding theorem then states that C then represents how many bits are required on average to encode k. We would like that our algorithm for determining which k to send is independent of the input distribution $\rho(a)$ so we maximize over that distribution. This way should our algorithm depend on $\rho(a)$ the communication cost will always represent the worst case scenario.

What we are interested in is the minimum value of C over all possible distributions of K and Y. We call this value C_{min} the **communication complexity** of the black box $P(s|a,b)$.

3.2 Asymptotic Simulations

Instead of trying to simulate a single shot of the communication game we will simulate many instances at once. This simplifies the problem as when one is encoding infinitely many messages the average cost per message is exactly equal to the entropy of the message, instead of an approximation.

The asymptotic simulation goes as follows, Alice gets as input $a_1, a_2, \ldots, a_N$ and similarly Bob gets as input $b_1, b_2, \ldots, b_N$. We still assume that all inputs come from finite sets A and B. Next Alice encodes all data into a single variable k which she can send over a classical channel to Bob. Bob then uses his inputs and Alice's message to output the values $s_1, s_2, \ldots$ for each game being simulated.

In the asymptotic simulation we then see that distribution $\rho(k|a, y)$ is replaced with the distribution $\rho(k|a_1, \cdots, a_N, y)$. We cannot assume that the cost of k is finite so we define the asymptotic communication cost as $\lim_{N\to\infty} C^{par}/N$ where C^{par} is the total length of k required for N simulations. The calculation C^{par} is similar to calculating C in Eq. 3.2 but with the maximization made over the space of distributions $\rho(a_1, \cdots, a_N)$.

As with the one shot communication complexity, we refer to the minimum asymptotic communication cost C_{min}^{asym} as the **asymptotic communication complexity**. An important fact is that the asymptotic communication complexity is always less than or equal to the single shot communication complexity, as in the worst case scenario the protocol for asymptotic communication simply comes down to repeating the protocol for the single shot case N times. Which in turn leads to a total communication cost of NC^{par}

3.2.1 Simulation through Classical Channels

The power of asymptotic simulations is that if we should apply Theorem 2.6 in some way we would get that the amount of bits of forward communication is exactly NC with N being the amount of simulations and C the channel capacity of the channel being used. For this theorem to then be useful we want a classical channel for which a single use exactly simulates our two input communication game. Once we have found such a channel we can calculate its channel capacity and thus calculate the amount of bits needed for k by the reverse Shannon theorem.

The naive way to construct this channel is simply creating the probability distribution $\rho(k|a)$ by averaging over the probabilities of $\rho(y)$, however this does not get us very far as we do not simply have any knowledge over k unless we construct an actual algorithm to simulate $P(s|a, b)$. Thus calculating the channel capacity will be impossible. What we would like is channel which uses the variables s, a and b. This may be unintuitive to some, as we started this section by stating that we do not posses any knowledge about s, a and b. The catch is that in actual applications of this theory we do know what s, a and b are and their properties, while k is an unknown bit string. And in addition we know the sets A and B so in fact we know a lot more about a, b than we do about k.

In the simulation of the black box game (P, A, B) , Alice sends Bob a variable k with probability $\rho(k|y, a)$. Bob then outputs the outcome s according to the probability $P(s|y, a, b)$ which satisfies Eq. 3.1. But as Alice does not know what input b Bob has gotten surely k must contain enough information for Bob to output a value s according to $P(s|a, b)$ for every b. So upon receiving k Bob should be able to construct a list $\mathbf{s} = \{s_1, \ldots, s_{|\mathcal{B}|}\}$, where s_i represents one of the possible outputs if Bob chooses $b = b_i$. Simulating $\mathbf{G}$ now comes down to Bob outputting the value $s_i \in \mathbf{s}$ which has the same index as his received input b_i. Given a k the probability that Bob should construct a certain list $\mathbf{s}$ is

$$\rho(\mathbf{s}|k, y) := \rho(s_1, \ldots, s_{|B|}|k, y) = \prod_{i=1}^{|B|} \rho(s_i|k, y, b_i). \tag{3.3}$$

As we know the probability distribution $\rho(k|y, a)$ we can remove the above dependency

on k and y

$$\rho(\mathbf{s}|a) = \sum_k \int \rho(\mathbf{s}|k,y)\rho(k|y,a)\rho(y)dy. \tag{3.4}$$

The distribution $\rho(\mathbf{s}|a)$ is exactly what we were looking for when we wanted to find a channel to classically simulate **G**. Now to confirm that a channel of this type simulates **G** we have to check that the marginal distributions of the mth variable $\rho(\mathbf{s}|a)$ is equal to the distribution $\rho(s|a,b_m)$. To do this we introduce the notation

$$\sum_{s_1} \tag{3.5}$$

to denote summing over every possible outcome Bob could output if he receives input b_1. Taking the marginal distribution of the ith variable of $\rho(\mathbf{s}|a)$ then gives us

$$\begin{aligned}
\sum_{\substack{s_1,\ldots,s_{i-1},\\ s_{i+1},\ldots,s_{|B|}}} \rho(\mathbf{s}|a) &= \sum_{\substack{s_1,\ldots,s_{i-1},\\ s_{i+1},\ldots,s_{|B|}}} \rho(s_1,\ldots,s_{i-1},s,s_{i+1},\ldots,s_{|B|}|a) \\
&= \sum_{\substack{s_1,\ldots,s_{i-1},\\ s_{i+1},\ldots,s_{|B|}}} \sum_k \int \rho(s_1,\ldots,s_{i-1},s,s_{i+1},\ldots,s_{|B|}|k,y)\rho(k|y,a)\rho(y)dy \\
&= \sum_k \int \sum_{\substack{s_1,\ldots,s_{i-1},\\ s_{i+1},\ldots,s_{|B|}}} \rho(s_1,\ldots,s_{i-1},s,s_{i+1},\ldots,s_{|B|}|k,y)\rho(k|y,a)\rho(y)dy
\end{aligned}$$

Now using Eq. 3.1 we can see that we must have that the marginal distribution of $\rho(\mathbf{s}|k,y)$ must be equal to $\rho(s|k,y,b_m)$ for exact simulation. Evaluating this marginal distribution gives

$$\begin{aligned}
&\sum_{\substack{s_1,\ldots,s_{i-1},\\ s_{i+1},\ldots,s_{|B|}}} \rho(s_1,\ldots,s_{i-1},s,s_{i+1},\ldots,s_{|B|}|k,y) = \sum_{\substack{s_1,\ldots,s_{i-1},\\ s_{i+1},\ldots,s_{|B|}}} \prod_{i=1}^{|B|} \rho(s_i|k,y,b_i) \\
&= \sum_{s_1} \sum_{\substack{s_2,\ldots,s_{i-1},\\ s_{i+1},\ldots,s_{|B|}}} \prod_{i=1}^{|B|} \rho(s_i|k,y,b_i) \\
&= \left(\sum_{s_1} \rho(s_1|k,y,b_1)\right)\left(\sum_{\substack{s_2,\ldots,s_{i-1},\\ s_{i+1},\ldots,s_{|B|}}} \prod_{i=1}^{|B|} \rho(s_i|k,y,b_i)\right) \\
&= \sum_{\substack{s_2,\ldots,s_{i-1},\\ s_{i+1},\ldots,s_{|B|}}} \rho(s_2,\ldots,s_{i-1},s,s_{i+1},\ldots,s_{|B|}|k,y) \\
&\vdots \\
&= \rho(s|k,y,b_m).
\end{aligned}$$

Here the last step denotes that it is possible to repeat the steps seen in the first four lines to remove the summation over every variable except over the ith variable which remains fixed. We have now achieved both our original goals, we wanted to remove to create a classical channel which simulates **G** which has a single input and output neither of which are the communicated variable k.

Definition 3.1. *Given a black box game* $\boldsymbol{G}$*=*(P, A, B) *the set* $\mathcal{V}(\boldsymbol{G})$ *contains any conditional probability distribution* $\rho(\mathbf{s}|a)$ *over the vector* $\mathbf{s} = (s_1, \cdots, s_{|\mathcal{B}|})^T$ *where the marginal distributions* $\rho(s|a)$ *over each variable* $s \in \mathbf{s}$ *satisfy*

$$\rho(s|a) = P(s|a, b) \qquad \forall s, a, b. \tag{3.6}$$

$\mathcal{V}(\mathbf{G})$ contains every classical channel capable of simulating **G** exactly. Now the natural question to ask is what the channel capacity of a given channel in $\mathcal{V}(\mathbf{G})$ is and to give a formal proof that this channel capacity is exactly equal to the amount of bits required for k.

Lemma 3.1. *Given a conditional probability* $\rho(\mathbf{s}|a) \in \mathcal{V}(\mathbf{G})$ *there is a classical protocol simulating N parallel black box games* $\boldsymbol{G}$*, whose asymptotic communication cost per game is equal to the capacity of the channel* $\rho(\mathbf{s}|a)$ *as N goes to infinity.*

Proof. Let $C(\mathbf{S}|A)$ be the channel capacity of the communication channel $\rho(\mathbf{s}|a)$. According to the reverse Shannon theorem there is a protocol S_N for a noiseless channel which use the channel $NC(\mathbf{S}|A)$ times and is exactly faithful to the probability distribution $\rho(\mathbf{s}|a)$ as $N \to \infty$. We are allowed to apply the reverse Shannon theorem as we allow for use of a shared random variable Y. The communication cost of this protocol is $NC(\mathbf{S}|A)$, so the asymptotic cost of this protocol is $C(\mathbf{S}|A)$ and the lemma is proven. □

So every channel in $\mathcal{V}(\mathbf{G})$ corresponds to a protocol for sending a variable k over with communication cost equal to the channel capacity. So to calculate the asymptotic communication complexity we simply have to find the channel in $\mathcal{V}(\mathbf{G})$ with the smallest capacity, for this we will now define the following quantity

$$\mathcal{D}(\mathbf{G}) = \min_{\rho(\mathbf{s}|a) \in \mathcal{V}(\mathbf{G})} \max_{\rho(a)} \mathcal{I}(\mathbf{S}, A). \tag{3.7}$$

Now we sadly cannot state that the asymptotic communication complexity is equal to this quantity $\mathcal{D}(\mathbf{G})$ and be done with it. Lemma 1 only states that every $\rho(\mathbf{s}|a)$ induces a protocol with asymptotic communication cost equal to the $C(\mathbf{S}|A)$ but not that every protocol to generate k induces a channel $\rho(\mathbf{s}|a)$.

Theorem 3.2. *The asymptotic communication complexity of the game* (P, A, B) *is the minimum of the capacity of the classical channels* $\rho(\mathbf{s}|a) \in \mathcal{V}(\mathbf{G})$*.*

Proof. Lemma 1 directly implies that $C_{min}^{asym} \leq \mathcal{D}(\mathbf{G})$. To proof the theorem we will show that $C_{min}^{asym} \geq \mathcal{D}(\mathbf{G})$. Let N denote the amount of instances of **G** being simulated. Let the distributions $\rho(k|y, a^1, \ldots, a^N)$ and $\rho(s^1, \ldots, s^N|k, y, b^1, \ldots, b^N)$ satisfy Eq. 3.1, where the superscript denote which game is simulated. These probability distributions thus effectively simulate N parallel instances of the black box game **G**. Let C_0 be the asymptotic communication cost of this simulation. The marginal distribution $\rho^i(s^i|k, y, m^1, \ldots, m^N) =$

$\rho^i(s^i|k, y, m^i)$ can then be used to generate a probability distribution $\rho^i(\mathbf{s}|a)$ as described in the previous section. This distribution can then be used to create the distribution

$$\rho(\mathbf{s^1}, \ldots, \mathbf{s^N}|a) = \prod_i^N \rho^i(\mathbf{s}^i|a). \tag{3.8}$$

Distributions build in this way satisfy Eq. 3.6 for every $s \in \bigcup_i \mathbf{s^i}$.

We now have a Markov chain $\mathbf{a} \to k \to (\mathbf{s^1}, \ldots, \mathbf{s^N})$ on which we will apply the data-processing inequality. The mutual information $\mathcal{I}(\mathbf{a}, (\mathbf{s^1}, \ldots, \mathbf{s^N}))$ can be taken to be equal to the channel capacity $C(\mathbf{s^1}, \ldots, \mathbf{s^N}|\mathbf{a})$ as the protocols defined are independent of the input distribution and the data-processing inequality is true for every input distribution. The mutual information $\mathcal{I}(k, \mathbf{a})$ is equal to $NC_0 + o(N)$[1], as we should have that $\lim_{N\to\infty} \mathcal{I}(k, \mathbf{a})/N = C_0$. The data-processing inequality now reads:

$$C(\mathbf{s^1}, \ldots, \mathbf{s^N}|\mathbf{a}) \leq NC_0 + o(N) \tag{3.9}$$

Let $\rho_0(\mathbf{s}|a) \in \mathcal{V}(\mathbf{G})$ be the probability distribution with minimal capacity $\mathcal{D}(\mathbf{G})$. Then the probability distribution

$$\rho_{min}(\mathbf{s^1}, \ldots, \mathbf{s^N}|a) = \prod_i^N \rho_0(\mathbf{s}^i|a). \tag{3.10}$$

is the channel with minimum capacity $N\mathcal{D}(\mathbf{G})$ due to the additivity of channel capacities. As this is the minimum capacity we have that

$$N\mathcal{D}(\mathbf{G}) \leq C(\mathbf{s^1}, \ldots, \mathbf{s^N}|\mathbf{a}) \tag{3.11}$$

and thus

$$N\mathcal{D}(\mathbf{G}) \leq NC_0 + o(N) \tag{3.12}$$

dividing both sides by N and taking the limit N to infinity finally gives us

$$\mathcal{D}(\mathbf{G}) \leq C_0 \tag{3.13}$$

and the theorem is proven. □

3.2.2 Tight Bounds for Communication Complexity

Now that we have a way to calculate the asymptotic communication complexity as per the optimization problem Eq. 3.7 we can start trying to calculate the one shot communication complexity C_{min}. It is currently unknown how to easily calculate this quantity as of right now but there are tight upper and lower bounds given by the asymptotic communication complexity. We first start off with the following lemma

Lemma 3.3. *Given a conditional probability $\rho_0(\mathbf{s}|a) \in \mathcal{V}(\mathbf{G})$ there is a protocol simulating a the black box game $\mathbf{G}$ with communication cost C such that*

$$C(\mathbf{S}|A) \leq C \leq C(\mathbf{S}|A) + 2\log(C(\mathbf{S}|A) + 1) + 2\log(e) \tag{3.14}$$

The proof of this lemma works similar to the proof of Theorem 3.1 and makes use of the one-shot reverse shannon theorem proved in [7].

[1] $o(N)$ is used to describe any function f satisfying $\lim_{N\to\infty} f(N)/N = 0$.

Theorem 3.4. *The communication complexity of black box game* $\mathbf{G}$ *satisfies*

$$\mathcal{D}(\mathbf{G}) \leq C_{min} \leq \mathcal{D}(\mathbf{G}) + 2\log(\mathcal{D}(\mathbf{G}) + 1) + 2\log(e) \tag{3.15}$$

Proof. The second inequality is given by Theorem 3.3, as we know that the protocol $\rho(\mathbf{s}|a)$ with minimal channel capacity $\mathcal{D}(\mathbf{G})$ induces a protocol with cost C such that

$$C \leq \mathcal{D}(\mathbf{G}) + 2\log(\mathcal{D}(\mathbf{G}) + 1) + 2\log(e) \tag{3.16}$$

and the definition of communication complexity states that $C_{min} \leq C$. The first inequality is due to that it is impossible for the asymptotic communication complexity cannot be larger than the communication complexity. □

With Eq. 3.7 and Eq. 3.16 we now have the tools to calculate the communication complexity. This is also as far as the black box will get us, as to continue we have to start doing calculations and for this we need to know the properties of our inputs and outputs. For this we will start applying our theory to a new technology which has revolutionized computer science to its core: quantum computing.

4. Quantum Communication

In this section we will apply the general theory we have developed in Section 3 to quantum communication. Quantum communication is a broad subject roughly defined by usage of quantum processes to communicate information between parties. There are many ways to exploit the postulates of quantum mechanics to achieve these goals but we will be looking at the most basic model of quantum communication where Alice and Bob simply replace their classical bits with qubits. It is assumed that the reader is comfortable with quantum mechanics but for those who are not, or need a quick refresher, can find a brief review of the basics of Quantum Mechanics in Appendix A.

The problem we will be examining here goes as follows: Alice gets as input a quantum state ψ which she proceeds to send to Bob through the quantum channel $\mathcal{L}$. Bob then performs a possible measurement $\mathcal{M}$ which is his received input. Then Bob proceeds to output the measured s. Afterwards we will see how one can construct the conditional distribution $P(s|a,b)$ for a two input communication problem from this quantum prepare and measurement communication problem. This application of a two input communication problem is also referred to as a **quantum game**.

4.1 Mixed States

Just like how classical channels can be represented by a probability distribution $\rho(y|x)$ quantum channels will also be represented by a conditional probability distribution for the output given a certain input. Only in the quantum case the input and output are of course quantum states. Due to this the math for quantum communication can get tedious as a single input corresponds to several outputs and every output corresponds to different outputs upon measurement. Luckily there is a way to represent quantum states which makes the calculations a lot easier, this representation is called the density operator.

Suppose Bob receives a quantum state $|\psi_i\rangle$ out of multiple possible states $|\psi_1\rangle, \ldots, |\psi_n\rangle$. Just like with classical channels Bob also knows the probability p_i with which he received $|\psi_i\rangle$. We then construct the **density operator** as

$$\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i| . \tag{4.1}$$

Density operators allow us to describe entire statistical ensembles of quantum states with a singular matrix. Such ensembles are also called **mixed states**, in contrast to **pure states** which are states not in ensemble. This density operator allows us to represent the output of a quantum channel as a density operator which then simplifies the math. Using only the definition of density operators we can derive the following useful properties:

1. (Trace condition) A density operator ρ has trace equal to 1.
2. (Positivity condition) ρ is a positive operator.
3. (Hermitian condition) ρ is a Hermitian operator.
4. (Idempotency condition) If the state is pure then the density matrix is idempotent, that is $\rho^2 = \rho$.

While these conditions are easy to derive using Eq. 4.1 it is worth it to examine what they physically represent. The trace condition implies that the probabilities of the different measurement outcomes sum to 1 as one would expect from any quantum state. Even if we have that the quantum state is represented by a statistical ensemble we will always get a single outcome.

Another way of viewing the density operator is, as the name implies, an operator composed out of several projection operators $|\psi\rangle\langle\psi|$ with weights p_i. As we can use projection operators to create measurement operators we then must have that for any density operator that $\langle\psi|\rho|\psi\rangle$ must be greater than 0 for every ρ and ψ. This is in fact the definition of a completely positive operator so ρ must be completely positive.

The idempotency condition also can be explained in a similar way. When we are dealing with a pure state the density operator is a projection operator. So if we apply the density operator to itself we expect no change, we are simply projecting a quantum state onto itself.

Using these properties we can rewrite all postulates of quantum mechanics in terms of density operators. The proofs that the proceeding postulates are equivalent to the regular versions of the postulates of quantum mechanics are omitted but can be found in [8].

Postulate 1′: *Associated with any isolated physical system is a complex vector space associated with an inner product, also known as a* ***Hilbert space****, which is known as the* ***state space*** *of the system. The system is completely described by a positive density operator with trace one. The density operator of a mixed state is given by*

$$\rho = \sum_i p_i \rho_i. \tag{4.2}$$

Postulate 2′: *The time evolution of a closed quantum system ρ_1 at time t_1 to the quantum state ρ_2 at time t_2 is given by a unitary transformation U dependent on $t_2 - t_1$. That is,*

$$\rho_2 = U\rho_1 U^\dagger \tag{4.3}$$

Postulate 3′: *A quantum measurement is described by a set $\{M_m\}$ of* ***measurement operators****. These are operators acting on the state space of the quantum system. The index m refers to the different outcomes of the measurement being performed. The probability that result m occurs is given by*

$$p(m) = \mathrm{Tr}(M_m^\dagger M_m \rho) \tag{4.4}$$

and the post measurement quantum state is given by

$$\frac{M_m^\dagger M_m \rho}{\mathrm{Tr}(M_m^\dagger M_m \rho)}. \tag{4.5}$$

The measurement operators must satisfy the ***completeness relation:***

$$\sum_m M_m^\dagger M_m = 1. \tag{4.6}$$

Now we can view every quantum channel, even those with some form of noise, as mappings from density operators to density operators.

4.2 The Bloch Sphere

Even the simplest form of quantum state, the qubit, requires 4 coordinates to describe so cannot be nicely visualized. However by exploiting some of the properties of qubits we can find a way to represent qubits in 3d space, which in turn will give us a nice way to visualize the effects of quantum channels.

4.2.1 Pure States on the Bloch Sphere

As a standard qubit is described by 2 complex numbers it has 4 degrees of freedom and thus requires 4 values to represent. However the requirement that the qubit is normalized removes 1 degree of freedom as the magnitude of 1 of the variables is fixed by the other complex variable. There is a second trick we can use to remove 1 more degree of a pure state qubit using only two coordinates. To do this we first rewrite the qubit into the following form

$$|\psi\rangle = e^{i\gamma}\left(\cos(\frac{\theta}{2})\,|0\rangle + e^{i\phi}\sin(\frac{\theta}{2})\,|1\rangle\right) \qquad \theta, \gamma, \phi \in \mathbb{R}. \tag{4.7}$$

We can rewrite the qubit into this form as we require that the qubit must be normalised. The trick we will use is that the factor $e^{i\gamma}$, also known as the **global phase**, turns out to be physically irrelevant and can be ignored. This can be seen by examining the probabilities for acquiring an outcome m, with associated POVM operator E_m, for the quantum state $\left|e^{i\gamma}\,|\psi\rangle\right\rangle$. One can then see that $\langle\psi|e^{-i\gamma}E_m e^{i\gamma}|\psi\rangle = \langle\psi|E_m|\psi\rangle$, so the global phase does not change the statistics of the measurement outcomes. As we are only interested in the statistics of measurement results and not the actual state itself we can then assume that the $\gamma = 0$ and write every qubit as

$$|\psi\rangle = \cos(\frac{\theta}{2})\,|0\rangle + e^{i\phi}\sin(\frac{\theta}{2})\,|1\rangle\,. \tag{4.8}$$

The values θ and ϕ now represent points on a unit sphere. Such a sphere is called the Bloch sphere and is represent in Figure 6. The Bloch sphere allows for easy visualizations of quantum operations mapping qubits to different qubits.

4.2.2 Mixed States on the Bloch Sphere

We know how to represent pure states on the Bloch sphere but we would in addition like a way to represent mixed states for qubits on the Bloch sphere. This should be possible as mixed states are described by density operators, which have 3 degrees of freedom due to the trace and hermitian condition. However it should not be allowed to map mixed states to points on the surface of the Bloch sphere, as those points are reserved for pure states. A good starting point is describing density operators in the basis of all hermitian matrices, which is given by

$$\left\{\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}, \begin{bmatrix}0 & -i\\ i & 0\end{bmatrix}, \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\right\}. \tag{4.9}$$

The latter three of these are called the Pauli matrices, denoted respectively σ_x, σ_y and σ_z. As density operators are hermitian they can be represented by a vector $\mathbf{a} \in \mathbb{R}^4$ with respect to

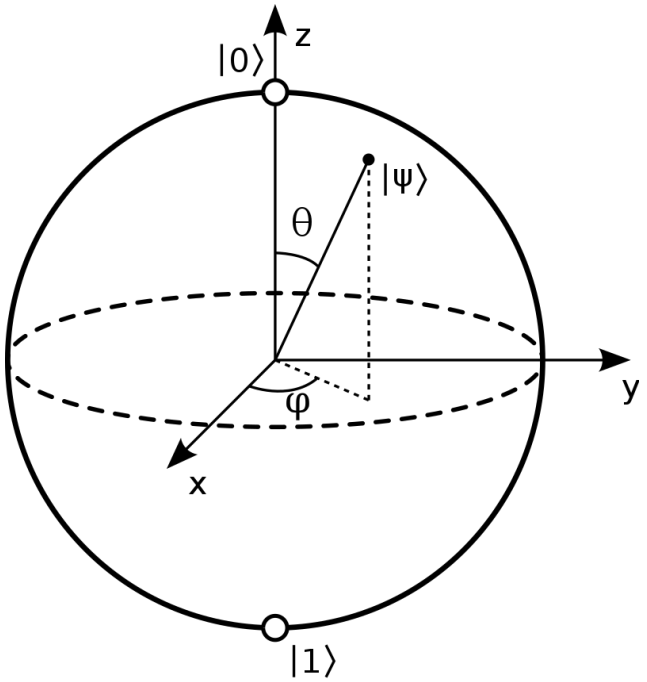


Figure 6: The Bloch sphere representation of a qubit $|\psi\rangle$. All qubits can be represented by a point on the surface of the Bloch sphere. The angles θ and ϕ represent the two variables required to describe a qubit and the poles of the Bloch sphere correspond to the $|0\rangle$ and $|1\rangle$ states.

the basis. We rewrite every density matrix to the form

$$\rho = \begin{bmatrix} a_i + a_z & a_x - ia_y \\ a_x + ia_y & a_i - a_z \end{bmatrix}. \tag{4.10}$$

Where the subscripts of a refer to their corresponding basis matrices. Thus a_i is associated with the identity matrix and a_x, a_y and a_z with their corresponding Pauli matrices. By using the fact that the density operators have trace one we can see that we must have that $a_i = 1/2$. So we will refer to the vector $\mathbf{a}$ as a 3 dimensional real vector with values a_x, a_y and a_z. With this notation we can rewrite every density operator into the following form:

$$\begin{gathered} \rho = \frac{1}{2}(I + \mathbf{a} \cdot \sigma) \\ (\mathbf{a} \cdot \sigma) := a_x \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + a_y \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} + a_z \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \end{gathered} \tag{4.11}$$

Now the 3 remaining coordinates in the vector $\mathbf{a}$ correspond to points in 3d space and what we would like to show is that these points are coordinates in the Bloch sphere. To formally prove this we have to show that a pure state corresponds to a vector satisfying $|\mathbf{a}| = 1$. To do this we will use the idempotency condition for pure states. We write

$$\frac{1}{4}(I + \mathbf{a} \cdot \sigma)(I + \mathbf{a} \cdot \sigma) = \frac{1}{2}(I + \mathbf{a} \cdot \sigma) \tag{4.12}$$

which upon expanding the brackets on the left hand side turns into

$$\frac{1}{4}(I + 2(\mathbf{a}\cdot\sigma) + (\mathbf{a}\cdot\sigma)^2) = \frac{1}{2}(I + \mathbf{a}\cdot\sigma). \tag{4.13}$$

To continue we will use the property of Pauli matrices that $(\mathbf{a}\cdot\sigma)^2 = |\mathbf{a}|^2 I$ [9]. This allows us to write

$$\frac{1}{4}(I(1 + |\mathbf{a}|^2) + 2(\mathbf{a}\cdot\sigma)) = \frac{1}{2}(I + \mathbf{a}\cdot\sigma).$$

Where the last expression is only true if and only if $|\mathbf{a}|^2 = 1$.

Now have a way to represent density operators on the Bloch sphere. But even in the case that $|\mathbf{a}|^2 < 1$ the vector $\mathbf{a}$ still points to a point in a unit sphere and still represents a valid density operator through Eq. 4.11. So what this implies is that interior points on the Bloch sphere represent mixed states and the surface represents all pure states. This will allow us visually represent the effects of quantum channels via the Bloch sphere, even if the output of a quantum channel is a mixed state.

4.3 Quantum Channels

Using the theory of the Bloch sphere and density operators we have the full mathematical toolkit to start talking about quantum channels and visualizing them as well. First we introduce what kind of mapping between density operators fully describes a quantum channel and then some examples are worked out to build up intuition.

4.3.1 Superoperators

Finally we have all the tools we need to fully present quantum channels and start applying our black box theory. At the start of this chapter we mentioned that Alice is able to send a quantum state through a quantum channel $\mathcal{L}$. With all the theory we have developed we will work through the quantum communication problem one more time and rigorously define everything that happens.

First Alice receives a quantum state, either mixed or pure. This state can be represented by a density operator and will afterwards be send through a quantum channel. This quantum channel can be noisy which then means that whatever qubit Bob receives is not necessarily the same as Alice send. It can also be the case for a channel that there is a chance that the quantum state Alice sends has a chance to be send through unchanged, or a chance to be changed in some way just like in the examples of classical channels we saw in Section 2. So a single input can be mapped to a statistical ensemble or rather, the mixed states we already encountered. A quantum channel must map density operators to density operators, this then means that a quantum channel must preserve all properties of density operators.

Definition 4.1 (Superoperator)**.** *A superoperator $\mathcal{L}$ is a linear operator acting on a vector space of linear operators which preserves trace and is completely positive.*

Superoperators are exactly what we are looking for to depict quantum channels. It has to be a linear mapping, as we expect that upon giving it a mixed state as given in Eq. 4.1 we want the same results if we used the channel multiple times with input state $|\psi_i\rangle$ with probabilites p_i. The properties of density operators have to be preserved thus the

superoperator has to preserve trace and it must be completely positive, which means that if the input is a positive operator the output must also be.

Now we are finally able to give a method to construct the probability distribution $P(s|a,b)$ for our quantum game. Alice receives a quantum state ρ and sends this to Bob through a quantum channel with superoperator $\mathcal{L}$, such that Bob receives the quantum state $\mathcal{L}[\rho]$. Bob then performs a POVM measurement where the probability of giving output s is given by

$$P(s|a,b) = \mathrm{Tr}(E_s \mathcal{L}[\rho]). \tag{4.14}$$

Which is what we originally set out to construct. With Eq. 4.14 we have a probability distribution on which we can apply all derived results in Section 3, in particular we can now start working on solving Eq. 3.7 for particular quantum channels.

4.3.2 Quantum Teleportation and the Depolarizing Channel

An important channel often seen is the quantum depolarizing channel. This channel has a chance p that any input is replaced by the maximally mixed state $I/2$ and a probability $1-p$ that the qubit is left untouched. This means that the channel corresponds to the mapping

$$\mathcal{L}[\rho] = p\frac{I}{2} + (1-p)\rho. \tag{4.15}$$

A quick calculation shows that the effect of this mapping on the Bloch sphere corresponds to contracting the Bloch Sphere by a factor p as depicted in Figure 7. When $p = 1$ for example we expect the Bloch sphere to contract entirely into the point $(0,0,0)$ which represents the maximally mixed state.

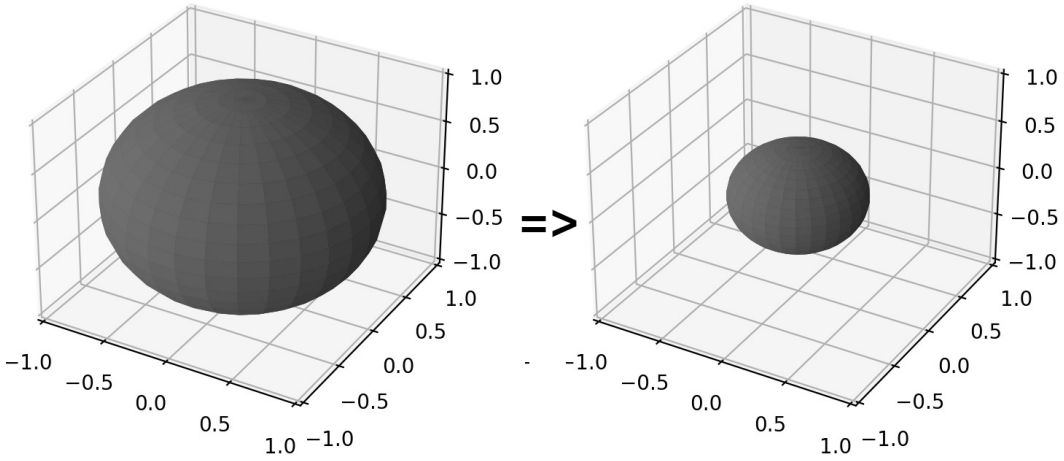


Figure 7: Visualization of the effect of the quantum depolarizing channel on the pure states, where the quantum depolarizing channel has a 50% chance to flip a state to the maximally mixed state.

This particular quantum channel arises physically in the case of quantum teleportation. Quantum teleportation entails the process of Alice wanting to send a qubit $|\psi\rangle$ to another party Bob, often over very large distances.

The protocol for quantum teleportation is as follows: say Alice has a quantum state $|\psi\rangle_C$ and wants to send this over to Bob. What they have at their disposal is a classical channel they can communicate across and a pair of entangled quantum states, where we denote Alice's entangled particle with subscript A and Bob's with B. Now we will require that the entangled particle is in one of four **Bell states**, the Bell states are a set of maximally entangled quantum states forming a basis for a 4 dimensional quantum state, thus form a basis for all systems composed of 2 qubits. The four Bell states are denoted as follows:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B \tag{4.16}$$

$$|\Phi^-\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B - |1\rangle_A \otimes |1\rangle_B) \tag{4.17}$$

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |1\rangle_B + |1\rangle_A \otimes |0\rangle_B) \tag{4.18}$$

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |1\rangle_B - |1\rangle_A \otimes |0\rangle_B). \tag{4.19}$$

Now Alice has access to particles A and C. As the Bell states form a basis Alice can perform a measurement on the combined states of these particles thus collapsing the combined state AC to one of the four Bell states. This causes Bob's particle B to collapse to a state closely related to or equal to the original state of $|\psi\rangle_C$. Once Alice has send her measurement result to Bob which only requires 2 bits of communication as there are 4 possible results, Bob can perform a unitary operation dependent on Alice's measurement result on his particle B and finally have a particle in the same state as $|\psi\rangle_C$ at the start.

Crucially for this protocol to work Alice and Bob need to know exactly what Bell state they have. Any four of the Bell states work but the unitary operation Bob performs is dependent on the initial Bell state. However as with any practical implementation it could happen that due to noise the Bell states can be turned into another state without the knowledge of Alice or Bob. But if a model is known for the quantum noise it should be possible to rewrite the entangled state as a mixed state of the original Bell state and the states which have been altered. This result is important as it is possible to model a quantum teleportation process which makes use of an arbitrary mixed state as a quantum depolarizing channel[10].

As quantum teleportation plays an important role in the development of a potential quantum internet it it is important to study how much classical information is needed to simulate a quantum depolarizing channel. As this then shows us how much better a potential quantum internet is than the current classical internet.

5. Convex Optimization

To compute the asymptotic communication complexity we now have to find a solution for the following optimization problem

$$C_{min}^{asym} = \min_{\rho(\mathbf{s}|a)\in\mathcal{V}(\mathbf{G})} \max_{\rho(a)} \mathcal{I}(\mathbf{S}, A) \tag{5.1}$$

together with the constraints

$$\sum_{\substack{s_1,\ldots,s_{i-1},\\ s_{i+1},\ldots,s_{|B|}}} \rho(\mathbf{s}|a) = P(s|a,b) \qquad \forall s, a, b \tag{5.2}$$

$$\rho(\mathbf{s}|a) \geq 0 \qquad \forall \mathbf{s}, a \tag{5.3}$$

$$\rho(a) \geq 0 \qquad \forall a. \tag{5.4}$$

$$\sum_a \rho(a) = 1 \tag{5.5}$$

An important feature of this problem is that the problem is convex in $\rho(\mathbf{s}|a)$ and concave in $\rho(a)$. Optimization of convex and concave functions is a well studied area and thus there are multiple important results we can utilize to solve our problem. In this section several of these results will be presented in addition to some results unique to the two input communication problem.

The results presented in this section can in general be applied to all two input communication problems but from here on out we will solely refer to the quantum game as the two input communication problem being studied.

5.1 The Minimax theorem

In solving our problem Eq. 5.1 we have to both minimize and maximize the objective function in regards to $\rho(\mathbf{s}|a)$ and $\rho(a)$ respectively. The current form of Eq. 5.1 implies that we are required to know the optimal distribution $\rho_{max}(a)$ before we can start minimizing over $\rho(\mathbf{s}|a)$. This makes it quite difficult to derive algorithms for solving the minimization problem, as we are not minimizing over a simple objective function but over a pointwise maximum of objective functions. Ideally we would want to swap the order in which the minimzation and maximization are done, as in that case we can search for a function $\rho(\mathbf{s}|a)$ maximizing the objective function without caring about $\rho(a)$ and vice versa. As it turns out there is a theorem which allows us to do exactly this swapping of minimization and maximization, the minimax theorem proven by John von Neuman in 1928 [11].

Theorem 5.1 (Minimax theorem)**.** *Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ be compact convex sets. If $f : X \times Y \to \mathbb{R}$ is continuous and is concave in x and convex in y then we have that*

$$\max_{y\in Y} \min_{x\in X} f(x,y) = \min_{x\in X} \max_{y\in Y} f(x,y) \tag{5.6}$$

We already know that mutual information is convex in $\rho(\mathbf{s}|a)$ and it is concave in $\rho(a)$ so we can apply Theorem 5.1 to Eq. 5.1. In proceeding sections we can then look for distributions $\rho(\mathbf{s}|a)$ minimizing the $\mathcal{I}(\mathbf{S}, A)$ without having to care about $\rho(a)$. This greatly simplifies the search for the distribution $\rho(\mathbf{s}|a)$ minimizing $\mathcal{I}(\mathbf{S}, A)$.

5.2 Maximization over Input Distributions

The first problem we then turn our attention to is the optimization of the input distributions with the restrictions given by

Problem 1.

$$\max_{\rho(a)} \quad \mathcal{I}(\mathbf{S}, A) \tag{5.7}$$

$$\text{subject to:} \quad \rho(a) > 0 \quad \forall a \tag{5.8}$$

The first problem one might notice is that our objective function is not convex in $\rho(a)$ but concave. This could be a problem as most optimization algorithms are specifically designed for convex optimization problems. However an easy fix one can do is try to solve the optimization problem

$$\min_{\rho(a)} -\mathcal{I}(\mathbf{S}, \mathbf{A}). \tag{5.9}$$

The objective function in Eq. 5.9 is convex in $\rho(a)$ and minimzing $-\mathcal{I}(\mathbf{S}, \mathbf{A})$ is equivalent to maximizing $\mathcal{I}(\mathbf{S}, \mathbf{A})$. So our solution to Eq. 5.9 is the same as for Eq. 5.1, as long as we still require that $\rho(a) \geq 0$.

The presented minimization problem is then a convex optimization over $|\mathcal{A}|$ variables. There are several algorithms available to solve such a problem. However in this section we will present a condition on $P(s|a, b)$ which is sufficient for the optimal input distribution to be a uniform distribution. We will start off by presenting two lemmas which can then be combined to proof a condition sufficient for the optimal input distribution to be uniform.

5.2.1 Conditions for Uniform Input Distribution

We will now present two lemmas which will be used to derive a result on the amount of variables defining the input distribution.

Lemma 5.2. *Given a quantum game $\boldsymbol{G} = (P, A, B)$ and the associated optimization problem with objective function $\mathcal{I}(\mathbf{S}, A)$. Without loss of generality let S be the set of all possible outcomes for every possible measurement $b \in B$. Let $\rho_{max}(a)$ be the input probability distribution maximizing $\mathcal{I}(\mathbf{S}, A)$. If there exists bijective transformations $f : A \to A$ and $g : B \times S \to B \times S$ satisfying that*

$$P(s'|f(a), b') = P(s|a, b) \qquad \forall a \in A, b \in B, s \in S \tag{5.10}$$

where $(b', s') = g(b, s)$, then $\rho_{max}(f(a))$ also maximizes the objective function $\mathcal{I}(\mathbf{S}, A)$.

Proof. Given the quantum game **G**=(P, A, B) , let S be the set of all possible measurement outcomes for every measurement. If two measurements have different labels for their respective measurement outcomes, but the same amount of outcomes, then one can change one of the measurements by relabeling the outcomes of one of the measurements to the same labels as the other measurement. This process does not alter the quantum game in any meaningful way. Similarly if two measurements have a different amount of measurement outcomes one can alter the measurement with the least amount of outcomes by adding outcomes with 0% probability of occurring, this process also does not change the quantum game. So we can assume that every measurement takes on outcomes given by the same set S.

Let $f : A \to A$ and $g : B \times S \to B \times S$ be bijective transformations such that

$$P(s'|f(a),b') = P(s|a,b) \qquad \forall a \in A, b \in B, s \in S \tag{5.11}$$

where $(b', s') = g(b, s)$. Let $\rho_{max}(a)$ be the optimal input distribution for the associated optimization problem with **G**. We will now construct a separate quantum game by applying the relevant transformations to all elements defining **G**, except the set of measurements, like so

$$\mathbf{G'} = (P(s'|f(a),b'), f(A), B) \tag{5.12}$$

As f is bijective we have that $f(A) = A$ and due to g being bijective the domain on which $P(s'|f(a),b')$ is defined remains unchanged. As $P(s|a,b)$ satisfies Eq. 5.11 we have that **G**=**G′** as all 3 elements defining the quantum game are the same. We will now proof that $\rho_{max}(f(a))$ is a solution to **G′** and then also a solution to **G** due to the games being identical, this will be done via a proof by contradiction.

Let $\rho_{max}(f(a))$ be a non-optimal solution to **G′**. We will now apply the inverse transformations used to construct **G′** to construct the new game

$$\mathbf{G''} = (P(s''|f^{-1}(f(a)),b''), f^{-1}(f(A)), B) \tag{5.13}$$

where again $(b'', s'') = g^{-1}(g(b, s))$. Trivially we then have that $\mathbf{G''} = \mathbf{G}$, so we must then have that $\rho_{max}(f^{-1}(f(a)) = \rho_{max}(a)$. By the same assumption that $\rho_{max}(f(a))$ is non-optimal we have that $\rho_{max}(f^{-1}(f(a))$, and thus $\rho_{max}(a)$, is a non-optimal solution to **G**. Hence we have a contradiction and then we must have that $\rho_{max}(f(a))$ is an optimal solution of **G′**. □

Lemma 5.3. *Given a concave maximization problem with objective function $f(\mathbf{x})$ with solutions $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Any sum $\sum_i a_i \mathbf{x}_i$ satisfying $\sum a_i = 1$ is also a solution, as long as $\sum_i a_i \mathbf{x}_i$ does not violate any constraints.*

Proof. We will apply Jennsen's inequality to the function f on points $\mathbf{x_1}, \ldots, \mathbf{x_n}$ and with any weights satisfying $\sum_i a_i = 1$ and that $\sum_i a_i x_i = 1$ not violating any possible constraints of the optimization problem. We then have

$$f\left(\frac{\sum_i a_i \mathbf{x}_i}{\sum_i a_i}\right) \geq \frac{\sum_i a_i f(\mathbf{x}_i)}{\sum_i a_i}$$

$$f\left(\sum_i a_i \mathbf{x}_i\right) \geq \sum_i a_i f(\mathbf{x}_i)$$

$$f\left(\sum_i a_i \mathbf{x}_i\right) \geq f(\mathbf{x}_1)$$

where in the last line it was used that $f(\mathbf{x_i}) = f(\mathbf{x}_1)$ for any i. As $\mathbf{x}_1$ is a solution to a maximization problem we must have that $f\left(\sum_i a_i \mathbf{x}_i\right) = f(\mathbf{x}_1)$ which implies that $\sum_i a_i \mathbf{x}_i$ is a solution. □

Finally we can combine the results of both Theorem 5.2 and Theorem 5.3 to derive the following theorem.

Theorem 5.4. *If the probability distribution $P(s|a,b)$ of a quantum game $\boldsymbol{G} = (P, A, B)$ is invariant under a transformation $a \to a + l \mod |\mathcal{A}|$ together with bijective transformations on s and b we have that the amount of variables over which must be maximized can be reduced to* $\gcd(l, |\mathcal{A}|) - 1$.

Proof. Let $\rho_{max}(a)$ be the input probability distribution solving the optimization problem associated with $\mathbf{G}$. As per Theorem 5.2 we have that $\rho_{max}(a+l)$ is also a solution to our optimization problem. Then by Theorem 5.3 we also have that

$$\rho(a) = \frac{1}{|\mathcal{A}|}\sum_{i=1}^{|\mathcal{A}|}\rho_{max}(a+il) \tag{5.14}$$

is also a solution if $\rho(a)$ satisfies all relevant conditions. We will now proof that $\rho(a)$ does satisfy the conditions and also show that the distribution has $\gcd(l,|\mathcal{A}|)-1$ degrees of freedom.

The first condition requires that $\rho(a) \geq 0$. It is trivial to show that this is true as $\rho_{max}(a) \geq 0$. The second condition requires that $\sum_a \rho(a) = 1$. To show this see that a we have that $a+il$ can take on $k := \frac{|\mathcal{A}|}{\gcd(l,|\mathcal{A}|)}$ distinct values. This means that we can rewrite Eq. 5.14 as

$$\rho(a) = \frac{1}{|\mathcal{A}|}\gcd(l,|\mathcal{A}|)\sum_{i=1}^{k}\rho_{max}(a+il) = \frac{1}{k}\sum_{i=1}^{k}\rho_{max}(a+il). \tag{5.15}$$

This can be done as we know that $\rho(a+il)$ can take on k distinct values and each of those distinct values appear $\gcd(l,|\mathcal{A}|)$ times in the sum. With this new notation we can easily write

$$\begin{aligned}\sum_a \rho(a) &= \frac{1}{k}\sum_a\sum_{i=1}^{k}\rho_{max}(a+il)\\ &= \frac{1}{k}\sum_{i=1}^{k}\sum_a\rho_{max}(a+il)\\ &= \frac{1}{k}\sum_{i=1}^{k}1\\ &= 1\end{aligned}$$

And we see that the condition holds. As $\rho(a)$ can only take on k values and one of these values is fixed due to the condition that all values summed together are equal to 1 the probability distribution $\rho(a)$ has $k-1$ degrees of freedom. □

An important corollary from this theorem is then if $\gcd(l,|\mathcal{A}|) = 1$, then we have that our input distribution is the uniform distribution.

5.2.2 Measurements for Uniform Input Distribution

An easy way to start constructing sets of states and measurements so that the input distribution is uniform is by limiting the possible measurements to only Projective valued measurements. A way to construct such sets is for example by creating

$$\mathbf{v}_x = \begin{bmatrix}\cos(\frac{\pi x}{|\mathcal{B}|})\\ \sin(\frac{\pi x}{|\mathcal{B}|})\\ 0\end{bmatrix} \tag{5.16}$$

Where of course $|\mathcal{B}|$ is the total amount of measurements. Now if the eigenvectors of the mth measurement correspond to $\pm\mathbf{v}_m$ with eigenvalues $s_m = \pm 1$ and if the allowed states $\mathcal{A}$ consist out of all $2M$ eigenvectors we have that $P(s|a,b)$ is invariant under the transformations $a \to a+1$ and $b \to b+1$.

The fact that projective measurements means we project onto vectors in our Hilbert space makes it easier to find and visualize these transformations. The probability of measuring an outcome is now only dependent on the distance between a quantum state and the eigenvectors of the measurement. While the proposed set of states and measurements restrict us to states on the xy plane of the Bloch sphere it still is useful as it can work for any amount of measurements. This is especially useful once on starts considering infinite states and measurements as we will do at the end of this chapter.

5.3 Minimization over Classical Simulation Protocols

By either applying the minimax problem or knowing the optimal input distribution one can remove the maximization performed in Eq. 5.1 and construct the following problem:

Problem 2.

$$\min_{\rho(\mathbf{s}|a)\in\mathcal{V}(\mathbf{G})} \quad \mathcal{I}(\mathbf{S}, A) \tag{5.17}$$

$$\text{subject to:} \quad \rho_i(s|a) = P(s|a,b) \quad \forall s, a, b \tag{5.18}$$

$$\rho(\mathbf{s}|a) \geq 0 \quad \forall \mathbf{s}, a \tag{5.19}$$

This is a convex minimization problem and to solve it we will apply a technique called Lagrangian Duality. Lagrangian duality relies on the idea of creating a seperate 'dual' problem to this problem and then use the solutions to that problem to study the original problem. First we will explain the main ideas and theory behind the Lagrangian dual problem and afterwards we will apply it to Problem 2.

5.3.1 Lagrangian Dual Problems

To start off we must first state a general form for convex optimization problems to which we can then apply our theory for lagrangian duality. We say that a convex optimization problem is a **primal problem** if it is in the following form

Primal problem.

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to:} \quad & f_i(x) \leq 0 \quad i = 1, \ldots, m \\ & h_i(x) = 0 \quad i = 1, \ldots, p \end{aligned}$$

Any finite optimization problem can be rewritten into such a form, hence we call it the primal form. When trying to solve such a problem one might encounter that some of the constraints have difficult forms and make the search for an analytic solution very difficult. To solve this we can relax some of the constraints and instead of making the constraints requirements we merely start viewing them as suggestions and penalizing solutions which violate the constraints instead of not permitting them.

Definition 5.1 (Lagrangian)**.** *Given an optimization problem in primal form we define the* ***Lagrangian*** *as*

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \tag{5.20}$$

where λ and ν are referred to as the ***Lagrangian multipliers****.*

This Lagrangian formalizes what we wanted to do when we wanted to relax our constraints. If one now tries to minimize this function $L(x, \lambda, \nu)$ with regards to x, as long as λ_i is positive for every i we get that positive $f_i(x)$ increase the value of the Lagrangian thus punishing us for using solutions violating our previous constraints. It then becomes important to choose optimal Lagrangian multipliers to ensure that our the minimum of our Lagrangian satisfies all our constraints. To find these optimal Lagrangian multipliers we introduce the **dual function**:

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu). \tag{5.21}$$

The dual function has several interesting properties which can help us find a solution to the primal problem. First of all, as shown in [12] the dual function is always concave even when the primal problem is not. The most important however is that the Lagrangian dual gives lower bounds on the optimal value p^* of our primal problem when $\lambda > 0$. This is easily shown by showing that if $\tilde{x}$ is a solution of our primal problem then we have that

$$g(\lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) = f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq f_0(\tilde{x}). \tag{5.22}$$

As this is true for any feasible point $\tilde{x}$ it is also true for the solution $\tilde{x}_{min}$ which minimizes $f_0(x)$. The fact that the dual problem gives lower bounds for the optimal value p^* of our primal problem is called **weak duality**. Now we would like a lower bound as big as possible to get as close to our optimal value as possible. To do this we can construct the following optimization problem called the **dual problem**

Dual problem.

$$\begin{aligned} &\text{maximize} && g(\lambda, \nu) \\ &\text{subject to:} && \lambda_i \geq 0 \quad i = 1, \ldots, m \end{aligned}$$

It is often the case that $g(\lambda, \nu) = -\infty$ which gives a nontrivial lower bound. This is why the first step in studying the Lagrangian duality is finding out which conditions are required such that $g(\lambda, \nu) > -\infty$. These conditions are then added to our dual problem as additional constraints.

As the name weak duality implies there also is a kind of stronger duality between our dual and primal problems. Whenever we have that the optimal value for the dual problem d^* is equal to p^* we speak of **strong duality**. There are many conditions which can be used to proof strong duality for a given optimization problem but the one we will be focusing on is called **Slater's condition**. Before we can introduce Slaters condition we must first define a certain type of constraint.

Definition 5.2 (Affine Function)**.** *A real function is affine when it can be expressed in the following form*

$$f(x_1, \ldots, x_n) = A_1 x_1 + \cdots + A_n x_n + b \tag{5.23}$$

where $A_i, b \in \mathbb{R}$ for all i.

There is a generalisation of affine functions for non real functions but for our purposes this definition is enough. Slaters condition for any primal problem then reads

Theorem 5.5 (Slaters condition)**.** *For a given convex primal problem where without loss of generality the first k constraints are affine we have that strong duality holds if there exists an x^* such that*

$$\begin{aligned} f_i(x^*) &\leq 0 \qquad i = 1, \ldots, k \\ f_i(x^*) &< 0 \qquad i = k+1, \ldots, m \\ Ax^* &= b. \end{aligned}$$

Thus we want that there is a feasible point which satisfies strict inequality for non-affine inequality constraints.

5.3.2 Conditions for Optimality

Our current goal is then finding the Lagrangian dual to the following problem

Problem 2a.

$$\min_{\rho(\mathbf{s}|a) \in \mathcal{V}(\mathbf{G})} \quad \mathcal{I}(\mathbf{S}, A) = \sum_{\mathbf{s},a} \rho(\mathbf{s}|a)\rho(a) \ln\left(\frac{\rho(\mathbf{s}|a)}{\rho(a)}\right) \tag{5.24}$$

$$\text{subject to:} \quad \rho_i(s|a) = P(s|a,b) \qquad \forall s, a, b \tag{5.25}$$

$$\rho(\mathbf{s}|a) \geq 0 \qquad \forall \mathbf{s}, a \tag{5.26}$$

Where we refer to the problem as problem 2a instead of simply as problem 2 like we did earlier. Important to note is that every one of our inequality constraints in this problem are affine so we automatically have strong duality.

Before we state the Lagrangian of our problem we will note that the are allowed to ignore the second constraint as the positivity of $\rho(\mathbf{s}|a)$ is already defined by the domain of the objective function. That is, $\mathcal{I}(\mathbf{S}, A)$ is not even defined if there is a negative value of $\rho(\mathbf{s}|a)$ so no need to keep it as a constraint. However we would still denote it as a constraint to remind us that in general it has to hold true and that it is a property of the solutions to problem 2a. With this in mind we write the Lagrangian as

$$L = \mathcal{I} - \sum_{s,a,b} \nu(s,a,b) \left[\sum_{\mathbf{s}/s_b} \rho(\mathbf{s}|a) - P(s|a,b) \right]. \tag{5.27}$$

Which we rewrite into the following form

$$L = L_0 + \sum_{s,a,b} \nu(s,a,b)\rho(a)P(s|a,b). \tag{5.28}$$

where

$$L_0 = \sum_{\mathbf{s},a} \rho(\mathbf{s}|a)\rho(a) \left[\ln \frac{\rho(\mathbf{s}|a)}{\rho(\mathbf{s})} - \sum_b \nu(s_b, a, b) \right]. \tag{5.29}$$

Now we will search for the constraints necessary so that the infinum of our Lagrangian is a finite value. Note that upon the relaxation of the first constraint of problem 2a we no longer strictly require that $\rho(\mathbf{s}|a)$ is normalized. This allows $\rho(\mathbf{s}|a)$ to go to infinity as we take the infinum.

The second term in Eq. 5.28 is independent of $\rho(\mathbf{s}|a)$ so we only have to focus on L_0. Specifically, if we can pick our Lagrangian multiplier ν such that the term in square brackets is negative $\rho(\mathbf{s}|a)$ our infinium is $-\infty$ as $\rho(\mathbf{s}|a)$ and $\rho(a)$ are non-negative. The term in square brackets is less than 0 when

$$\rho(\mathbf{s}|a) < \rho(\mathbf{s}) e^{\sum_b \nu(s_b,a,b)}. \tag{5.30}$$

So if any distribution $\rho(\mathbf{s}|a)$ satisfies this inequality we have that the infinum of our Lagrangian is $-\infty$. Now if we rewrite the conditional distribution $\rho(\mathbf{s}|a)$ into the following form

$$\rho(\mathbf{s}|a) = \alpha \frac{e^{\sum_b \nu(s_b,a,b)}}{\sum_{\bar{a}} \rho(\bar{a}) e^{\sum_b \nu(s_b,\bar{a},b)}}. \tag{5.31}$$

we get that

$$\rho(\mathbf{s}) = \alpha. \tag{5.32}$$

Plugging this into Eq. 5.31 one can rewrite the inequality to

$$\sum_a \rho(a) e^{\sum_b \nu(s_b,a,b)} > 1. \tag{5.33}$$

Which is exactly the kind of condition we are looking for to construct the dual problem with. As we are summing over $\mathbf{s}$ in Eq. 5.29 if there is a single $\mathbf{s}$ such that Eq. 5.33 holds then one can see that the infinum of L_0, and thus L, is equal to $-\infty$.

Now we would like to show that when Eq. 5.33 does not hold that the infinum is positive. A way to prove is to use Jensen's inequality. The function $\ln(x)$ is concave function however so to apply Jensen's inequality to prove positivity we rewrite L_0 to

$$L_0 = -\sum_{\mathbf{s},a} \rho(\mathbf{s}|a)\rho(a) \left[\ln \frac{\rho(\mathbf{s})}{\rho(\mathbf{s}|a)} - \sum_b \nu(s_b, a, b) \right] \tag{5.34}$$

Now we have the convex function $-\ln(x)$ to which we apply Jensen's inequality with the weights

$$a_{\mathbf{s}} = \sum_a \rho(\mathbf{s}|a)\rho(a) = \rho(\mathbf{s}) \tag{5.35}$$

Note that the before applying Jensen's inequality we will multiply both sides of our inequality by $\sum_{\mathbf{s}} \rho(s)$ so that the right hand side of Eq. 2.3 is equal to L_0. We then get

$$L_0 \geq -\sum_{\mathbf{s}} \rho(\mathbf{s}) \ln \left(\frac{\sum_{\mathbf{s}'} \rho(\mathbf{s}') \sum_a \rho(a) e^{\nu(s_b,a,b)}}{\sum_{\mathbf{s}'} \rho(\mathbf{s}')} \right) \geq 0. \tag{5.36}$$

The above is only true when

$$\sum_a \rho(a) e^{\sum_b \nu(s_b,a,b)} \leq 1 \tag{5.37}$$

which is the exact negation of Eq. 5.33 so we have that when the above condition is satisfied that the infinum is of our Lagrangian is non-negative. By differentiating L with respect to $\rho(\mathbf{s}|a)$ and setting the derivative equal to 0 one finds that L is equal to 0 when

$$\rho(\mathbf{s}|a) = \rho(\mathbf{s}) e^{\sum_b \nu(s_b,a,b)}. \tag{5.38}$$

Which is then a necessary condition for the minimality of the Lagrangian. By multiplying both sides of this condition by $\rho(a)$ and summing over a one finds

$$\rho(\mathbf{s}) = \rho(\mathbf{s}) \sum_a \rho(a) e^{\sum_b \nu(s_b,a,b)}. \tag{5.39}$$

This constraint then tells us that for a given $\mathbf{s}$ we must have that $\rho(\mathbf{s})$ and thus $\rho(\mathbf{s}|a)$ must be equal to 0 for every a. Now we then find our dual objective function by plugging Eq. 5.38 into Eq. 5.28 and we formulate the dual problem as follows

problem 2b.

$$\begin{aligned} \max_{\nu} \quad & \mathcal{I}_{dual} = \sum_{s,a,b} \nu(s,a,b)\rho(a)P(s|a,b) & \\ \text{subject to:} \quad & \sum_a \rho(a) e^{\sum_b \nu(s_b,a,b)} \leq 1 & \forall \mathbf{s}, b \\ & \rho(\mathbf{s}) = \rho(\mathbf{s}) \sum_a \rho(a) e^{\sum_b \nu(s_b,a,b)} & \forall \mathbf{s}, b. \end{aligned}$$

5.3.3 Reduction of Variables

We now have that the solution $\rho(\mathbf{s}|a)$ to problem 2a is only the optimal solution if and only if there exists variables $\nu(s,a,b)$ such that

$$\sum_{s_1,\ldots,s_{i-1},s_{i+1},\ldots,s_A} \rho(\mathbf{s}|a) = P(s|a,b) \qquad \forall s,a,b \tag{5.40}$$

$$\rho(\mathbf{s}|a) \geq 0 \qquad \forall \mathbf{s}, a \tag{5.41}$$

$$\rho(\mathbf{s}|a) = \rho(\mathbf{s}) e^{\sum_b \nu(s_b,a,b)} \qquad \forall \mathbf{s}, a \tag{5.42}$$

$$\sum_a \rho(a) e^{\sum_b \nu(s_b,a,b)} \leq 1 \qquad \forall \mathbf{s}, b. \tag{5.43}$$

Using additional theorems we can replace some of these constraints with weaker constraints to make the problem easier. To make sure that the new constraint is sufficient to replace the old one, we have to make sure that the new constraint is satisfied if and only if the old one is.

We will start with the easiest constraint to replace which is the second one which we will replace with $\rho(\mathbf{s}) \geq 0$. If $\rho(\mathbf{s}) \geq 0$ we have that $\rho(\mathbf{s}|a) \geq 0$ due to the third constraint. Conversely if $\rho(\mathbf{s}|a) \geq 0$ then $\rho(\mathbf{s}) \geq 0$ due to the definition of $\rho(\mathbf{s})$.

The third constraint tells us the values of $\rho(\mathbf{s}|a)$ if we know the values for $\rho(\mathbf{s})$ and $\nu(s, a, b)$. So we can reduce the amount of variables over which to optimize by replacing the first constraint by

$$\sum_{s_1,\ldots,s_{i-1},s_{i+1},\ldots,s_A} \rho(\mathbf{s}) e^{\sum_b \nu(s_b,a,b)} = P(s|a,b). \tag{5.44}$$

And we can replace the third constraint by Eq. 5.39 to finally obtain the new and equivalent constraints

$$\sum_{s_1,\ldots,s_{i-1},s_{i+1},\ldots,s_{|B|}} \rho(\mathbf{s}) e^{\sum_b \nu(s_b,a,b)} = P(s|a,b) \qquad \forall s, a, b \tag{5.45}$$

$$\rho(\mathbf{s}) \geq 0 \qquad \forall \mathbf{s} \tag{5.46}$$

$$\rho(\mathbf{s}) = \rho(\mathbf{s}) \sum_a \rho(a) e^{\sum_b \nu(s_b,a,b)} \qquad \forall \mathbf{s}, b \tag{5.47}$$

$$\sum_a \rho(a) e^{\sum_b \nu(s_b,a,b)} \leq 1 \qquad \forall \mathbf{s}, b. \tag{5.48}$$

These constraints allow for easier optimization as we have replaced $\rho(\mathbf{s}|a)$ with $\rho(\mathbf{s})$, the latter of which is only defined by $|\mathbf{S}|$ variables and the former by $|\mathbf{S}||A|$ variables, where $|\mathbf{S}|$ is the amount of possible vectors $\mathbf{s}$.

5.4 Infinite States and Measurements

So far we have limited ourselves to the case where Alice and Bob can choose from a finite set of quantum states and measurements. However one might be interested into how much classical communication a quantum channel can replace when Alice and Bob can pick from an uncountably infinite set of possible quantum states or measurements. Such quantum games will be referred to as infinite quantum games, in contrast to the finite quantum games we already know. In this section we will discuss how to calculate the asymptotic communication complexity in the case of uncountably infinite states and measurements. For this we assume the reader is acquainted with measure theory and in particular the construction of the Lebesque integral.

5.4.1 Interpreting the limit and real analysis

One method of interpreting an infinite quantum game is to view it as a limit of the asymptotic communication complexity as we let the amount of possible measurements and states approach infinity. That is, imagine we have two sequences of sets A_n and B_n which have that $A_n \uparrow A$ and $B_n \uparrow B$. Now if A and B are dense in the unaccountably infinite sets $\mathcal{A}$ and $\mathcal{B}$ we can view the quantum game $\mathcal{L} = (P(s|\psi, \mathcal{M}), \mathcal{A}, \mathcal{B})$ as a limit object of the finite quantum games $\mathbf{G}_n = (P(s|a,b), A_n, B_n)$. What we mean with being the limit object is that we expect the solutions of every game $\mathbf{G}_n$ approach the solutions to the game $\mathcal{L}$.

Now with an idea of what we want to show we will now start defining everything we need to view the infinite quantum games. To signify that the problem we are now studying

is continuous we will change some of the notation. While we still want our states and measurements to come from $\mathcal{A}$ and $\mathcal{B}$ we now denote elements from these sets as ψ and $\mathcal{M}$ respectively. To properly define our problem we need to choose integration measures on both states and measurements. We choose these measures such that

$$\int_{\mathcal{B}} d\mathcal{M} = \int_{\mathcal{A}} d\psi = 1. \tag{5.49}$$

Then if we have that the input distribution $\rho(\psi)$ is uniform we get

$$\int_{\mathcal{A}} \rho(\psi) d\psi = \rho(\psi) = 1. \tag{5.50}$$

In general we will assume for the next sections that the optimal input distribution is known and only focus on the harder problem of minimizing over classical channel. In the case of infinite states and measurements it becomes easier to find sets of measurements and states which obey the requirements in Theorem 5.4 especially if only the only measurements being considered are PVM.

5.4.2 Partitions of the Measurement Manifold

While we no longer assume that our space of states and measurements the two assumptions we made about our problem still pose true. Those are that the Hilbert space dimension is still finite and that every measurement has finitely many outcomes.

In a finite quantum game our method of classically simulating the game involved sending a variable k so that Bob can construct a vector $\mathbf{s}$ from which he can pick a value corresponding to his input b to simulate the game. This vector $\mathbf{s}$ can be summarised as a function $S : \mathcal{M} \mapsto s$ and in equations Eq. 5.40 to Eq. 5.43 we then have that every constraint which has to be true for every $\mathbf{s}$ has to be true for every function S. Analyzing such a constraint analytically is difficult so we would like to replace this function S with something else.

To be able to replace the function S this we have to make an additional assumption about the measurements Bob can use. We will require that the amount of measurement outcomes any measurement can have is bounded. Put differently is that there is a finite number M which bounds the total amount of outcomes every measurement can have. This is necessary as even though every measurement has finitely many outcomes, as we have infinite measurements it is still possible that there is no upper bound to the amount of measurement outcomes.

With this assumption we can say without loss of generality that we can label the set of possible outcomes $\{1, \ldots, M\}$. Now Alice can construct a disjoint partition of the space of all measurements dubbed $\{\Omega_1, \ldots, \Omega_M\} := \mathbf{\Omega}$ which has to satisfy that $S(\mathcal{M}) = i$ for every $\mathcal{M} \in \Omega_i$. If Alice sends Bob such a partition Bob can assign a measurement to a certain outcome by simply looking in which part of the partition his measurement is. Then the partition $\mathbf{\Omega}$ replaces the role of $\mathbf{s}$ in our infinite quantum game.

Together with the measures presented we now have enough to construct the infinite versions of both problem 2a and problem 2b.

Problem 3a.

$$\min_{\rho(\mathbf{\Omega}|a)} \quad \mathcal{I} = \int_{\mathcal{P}} \int_{\mathcal{A}} \rho(\mathbf{\Omega}|\psi)\rho(\psi) \ln\left(\frac{\rho(\mathbf{\Omega}|\psi)}{\rho(\mathbf{\Omega})}\right) d\psi d\Omega \tag{5.51}$$

$$\text{subject to:} \quad \int_{\mathcal{P}_s(\mathcal{M})} \rho(\mathbf{\Omega}|\psi) d\Omega = P(s|\psi, \mathcal{M}) \qquad \forall\, s, \psi, \mathcal{M} \tag{5.52}$$

$$\rho(\mathbf{\Omega}|\psi) \geq 0 \qquad \forall\, \mathbf{\Omega}, \psi \tag{5.53}$$

problem 3b.

$$\max_{\nu} \quad \mathcal{I}_{dual} = \sum_{s} \int_{\mathcal{B}} \int_{\mathcal{A}} \nu(s, \psi, \mathcal{M}) \rho(\psi) P(s|\psi, \mathcal{M}) d\mathcal{M} d\psi$$

$$\text{subject to:} \quad \int_{\mathcal{A}} \rho(\psi) e^{\int_{\Omega_s} \nu(s,\psi,\mathcal{M}) d\mathcal{M}} d\psi \leq 1 \qquad \forall\, \mathbf{\Omega}, \mathcal{M}$$

$$\rho(\mathbf{\Omega}) = \rho(\mathbf{\Omega}) \int_{\mathcal{A}} \rho(\psi) e^{\int_{\Omega_s} \nu(s,\psi,\mathcal{M}) d\mathcal{M}} d\psi \qquad \forall\, \mathbf{\Omega}, \mathcal{M}.$$

The strength of using $\mathbf{\Omega}$ instead of the function S is that when we restrict ourselves to the case where Bob is only allowed to use PVM. Only allowing PVM makes our problem at least intuitively easier as it allows us to represent our measurements in the Hilbert space of quantum states and there are useful mathematical results about partitioning vector spaces into different subsets. For example in [13] in the calculation of a lower bound for the asymptotic communication complexity in the case of a noiseless quantum channel and Bob using only rank-1 PVM measurements an application was found for the double cap conjecture. The conjecture states the maximum volume two subsets on a unit ball can have without containing orthogonal vectors.

While weak duality still holds for problem 3a and problem 3b it remains an open question whether strong duality still holds as Slater's conditions is only applicable to optimization problems with finite variables and constraints.

5.4.3 Proving strong duality

In [13] the case for generalizing the quantum game to uncountably many states and measurements has been presented. However it remains unclear that strong duality holds for this case. Slater's condition cannot be used for the infinite case as it only holds for optimization problems with finitely many variables and constraints. Here we present a possible method for proving strong duality but whether this method could work remains an open question.

The general approach to the proof is by using sequences of quantum games. The idea is that given two sequences of sets A_n and B_n which approach sets that are dense in the space of all quantum states and measurements respectively. Of course it is required that for every n that the sets A_n and B_n are finite. In that case one can construct a sequence of quantum games $\mathbf{G}_n = (P(s|a, b), A_n, B_n)$ for each of which strong duality holds for the associated

optimization problem. This method does however require a definition of a distance metric in the set of measurements.

The first step is proving that problems 3a and 3b are feasible. This will require the definition of a measure on the space of all partitions $\mathbf{\Omega}$ which is rather difficult. The second step is proving that the limit object of the sequence of functions $\rho_n^{min}(\mathbf{s}|a)$ and $\nu_n^{min}(s,a,b)$ which are the solutions of the nth quantum game respectively. It is hypothesized that these functions approach functions $\rho_{min}(\mathbf{s}|a)$ and $\nu_{min}(s,a,b)$ respectively which solve problems 3a and 3b. By the fact that the functions $\rho_n^{min}(\mathbf{s}|a)$ and $\nu_n^{min}(s,a,b)$ are simple functions we have that the objective functions and constraints of the games optimization problems associated with $\mathbf{G}_n$ approach those of the infinite quantum game. Then the final step is showing that as strong duality holds for every finite quantum game then that it must also for the infinite quantum game.

6. Ontological Theory

Ever since the advent of quantum mechanics the reality of the wave function has been a debated subject. The main question is whether the wave function really represents a part of reality or does it just represent our statistical knowledge of reality? The field of research dealing with this question is called ontological theory. The main results in ontological theory are the famous EPR and Kochen-Specker no-go theorems, which place limits on the properties an underlying theory of quantum mechanics can have [14][15]. In this section we will present a link between the classical simulations of the quantum games we saw in preceding sections and ontological theory. This link would allow for cross correlation between the fields of quantum communication and ontological theory to derive further results for both fields.

6.1 Reality of quantum state

To formalize the question on whether the wave function is real we will refer to a certain state of reality as the **ontic state** denoted as λ. This λ is simply a set of variables which uniquely describe reality. Reality in this case is described as an observation-independent state of the universe. Thus here it is assumed that there exists a description of reality without having to take the observer into account in said description of reality. This was the view of Einstein while on the other Niels Bohr believed that it was only possible to describe the properties of an object in regards to another object as detailed in [16].

Another assumption made in this section is that as an ontic state describes reality there must be a wave function associated with this ontic state. This assumption would turn out to be wrong if it were to turn out that quantum physics is simply an approximation of reality, in the same way that Newtonian mechanics is an approximation of relativistic mechanics. The question is now to which degree the variables defining the ontic state also define the corresponding wave function.

6.1.1 ψ-Ontic Theories

As every ontic state is a state of reality we know that there is at least one quantum state associated with said ontic state. Conversely when one has a quantum system given by ψ there must be an ontic state corresponding to the quantum state. Note that this ontic state does not have to be unique, as if it is the case that the wave function represents our statistical knowledge of reality then multiple ontic states correspond with the wave function. In any case we can conclude that there exists a mapping

$$|\psi\rangle \rightarrow \rho(\lambda|\psi) \tag{6.1}$$

which assigns to each wave function a probability distribution of ontological space. A requirement is that this mapping is injective, as every distinct wave function needs to correspond to a different state of reality.

The support of the distribution $\rho(\lambda|\psi)$ is a very important concept, as the support of this function now represents all ontic states possible for a given wave function, thus what states of reality associate with a wave function. Using the ontic state one should be able to calculate the probability of measurement outcomes for a wave function. This is why it is also possible to set up the following mapping for every measurement $\mathcal{M}$

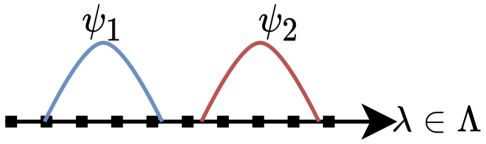
$$\mathcal{M} \to P(E_i|\lambda, \mathcal{M}) \tag{6.2}$$

where E_i are the measurement outcomes of $\mathcal{M}$.

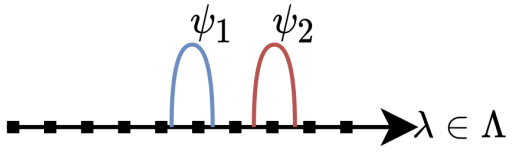
Any ontological theory can then be described by the mappings Eq. 6.1 and Eq. 6.2 when coupled with a method for the time evolution of the ontic state. A trivial ontic theory is of course the one stating that the wave function does depict reality and thus that the mappings come down to basically being the identity mappings. It is trivial that that theory matches up with quantum theory but for every other ontic theory it should be checked that the predictions made by the theory match up with reality, the predictions match up when the following equation is satisfied:

$$\int P(E_i|\lambda, \mathcal{M})\rho(\lambda|\psi)d\lambda = P(E_i|\psi, \mathcal{M}) = \langle\psi| E_i |\psi\rangle . \tag{6.3}$$

Theories which claim that the wave function represent a part of reality are called **ψ-ontic**. For a theory to be ψ-ontic one must have that the for 2 distinct quantum states that their probability distributions as induced by Eq. 6.1 have disjoint supports, as depicted in Figure 8a. Whether or not this implies that this means that the wave function has a physical interpretation has been a subject of recent debate [17]. However an important subclass of ψ-ontic theories called **complete ψ-ontic**[2] require that every wave function corresponds to a single unique ontic state, as depicted in Figure 8b.



(a) Representation of a ψ-ontic theory where the wave functions map to distinct ontic states.



(b) Representation of a complete ψ-ontic theory where every wave function corresponds to a single unique ontic state.

For every ψ-ontic theory the amount of information required to specify the ontic state is infinite [18]. As every ontic state corresponds to a wave function then knowing the ontic state is equal to knowing the wave function. As a wave function requires infinite information to encode we automatically have that an ontic state also requires this amount of information.

6.1.2 ψ-Epistemic Theories

Theories which are not ψ-ontic are called **ψ-epistemic** theories. These theories claim that the wave function represent information about reality and not reality directly, a graphical depiction of such theories can be seen in Figure 9.

While epistemic theories face heavy constrictions on predicting all of quantum mechanics it very much is possible to construct models for simple quantum mechanical situations. An example of this is the Kochen-Specker model [14] which is a ψ-epistemic model capable of simulating single qubits. The Kochen-Specker model models the ontic state as a three dimensional vector denoted **x**, which when given the Bloch vector **v** of a quantum state , is

[2]Sometimes referred to as 'ontic in the strong sense'

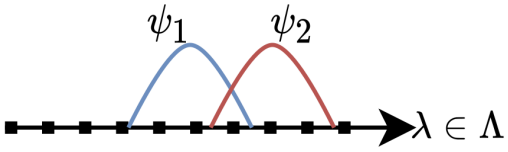


Figure 9: Graphical representation of a ψ-epistemic theory. A theory where the wave function only depicts our information of reality and thus not reality itself

set according to the following distribution

$$\rho(\mathbf{x}|\mathbf{v}) = \frac{1}{\pi}\mathbf{v}\cdot\mathbf{x}u(\mathbf{v}\cdot\mathbf{x}) \tag{6.4}$$

where u is the unit step function.

Many constraints exist on epistemic theories like the famous Bell-Kochen-Specker theorem and the more recent Pusey-Barret-Rudolph theorem. Both of which place major constraints on various types of epistemic theories. However in recent times interest in epistemic theories have been renewed due to applications of these theories in quantum communication and as thus still remain relevant field of research.

An important subclass of epistemic theories are where the information contained in the ontic state about the wave function is finite. This subclass of epistemic theories we are interested in are called **completely ψ-epistemic** and a theory has two requirements to fulfill to be classified as such. The first is that the entropy $H(\Lambda|\Psi)$ has to be finite, where Λ is a stochastic variable with outcomes given by the set of all possible ontic states and similarly for Ψ and all possible wave functions. The second is that the entropy of the maximally mixed state $\int \rho(\lambda|\psi)\rho(\psi)d\psi = \rho(\lambda)$ where $\rho(\psi)$ is the uniform distribution is finite. These conditions combined imply that the mutual information $I(\Lambda, \Psi)$ between the ontic state and wave function is finite. The Kochen-Specker model shown in Eq. 6.4 is a completely ψ-epistemic theory with a mutual information of approximately 1.28 bits.

The reason that we care about completely ψ-epistemic theorems is that there is a correspondence between simulations of quantum games and these kind of theories as shown in [18]. It is easy to show that any classical simulation induces a ψ-epistemic theory, simply let the role of the ontic state be replaced by the communicated variable k and the shared noise y and one can then see that equations Eq. 6.3 and Eq. 3.1 are the same.

To proof that a completely ψ-epistemic theory induces a classical simulation protocol requires a bit more work. That is as we cannot assume that the ontic state λ itself can be described by finitely many bits. Instead what happens in the induced protocol Alice, after receiving her input ψ, sends Bob enough information for Bob to set an ontic state according to the distribution $\rho(\lambda|\psi)$. The amount of information that Alice requires to do this is essentially equal to $I(\Lambda, \Psi)$ and since this quantity is finite we have a classical protocol simulating a quantum game.

6.2 Applications to Communication Complexity

With a strong link between completely ψ-epistemic theories and quantum communication we can now present several results from one field of research which could have impact on

the other. These highlight two more recent results but there are many more examples which could have impact.

6.2.1 Lower Bound on Required Variables

One recent result published by A. Montina which could have impact on quantum communication concerns so called Markovian ontic theories [19]. A Markovian ontic theory is a theory in which the time evolution of the ontic variables depends only on a single past state of the variables and not their entire history. An easy example of the Markovian property outside of ontological theory is the time evolution of a closed quantum system. For a closed quantum system we have that the time evolution of a quantum state is given by a unitary operator U, so we can write

$$|\psi(t_2)\rangle = U(t_2, t_1)\,|\psi(t_1)\rangle\,. \tag{6.5}$$

This system is thus Markovian as if one knows the quantum state at time t_1 and the unitary operator $U(t_2, t_1)$, then we do not require more information to determine the quantum state at time t_2. While if the time evolution where non Markovian one would also be required to know the quantum state before time t_1 to determine the state at time t_2.

A Markovian ontic theory thus requires that the variables describing the ontic state λ_{t_1} are enough knowledge about the system to determine λ_{t_2}. The result in [19] states that any Markovian theory requires at least $2N - 2$ continuous variables to describe the ontic state λ. This then gives a lower bound for the amount of variables Alice and Bob need to communicate with one another to simulate a quantum game if their communicated variable k and random shared variable y turn out to obey Markovian like transformation rules. The exact form these transformations entail is not given but due to quantum states evolving in a Markovian like manner it seems likely that almost all classical protocols would as well.

6.2.2 Possibility of Epistemic Theories

In [13] two hypotheses are presented to derive the lower bound of $\frac{1}{2}N\log(N)$, with N being the Hilbert space dimension, for the asymptotic communication complexity for noiseless channels with 2 outcome projective measurements. The hypotheses are very technical as is the derivation of the lower bound given the hypothesis so they will not be presented in this thesis. However there is an important consequence of this lower bound but to explain this consequence first some further context is required.

As has been stated the Bell-Kochen-Specker is a theorem which prohibits a large class of ψ-epistemic theories. However a recent result known as the Pusey-Barrett-Rudolph (PBR) theorem is a stronger theorem in that it disallows a larger class of ψ-epistemic theories that are impossible to integrate with quantum mechanics. Their theorem relies on the **preparation independence** axiom, which states that if two quantum states are prepared independently then that the ontic state of their product state is an independent combination of their individual ontic states [15].

Should the shown lower bound of $\frac{1}{2}N\log(N)$ hold true it turns out that the preparation independence property can be replaced with the weaker **equipartition property** [20]. The formal definition of this property requires more mathematics to describe but it suffices to know the approximate idea of this property. For an ontic theory to satisfy this theory one will require that given ψ there is a set of ontic states for which $\rho(\lambda|\psi)$ is approximately constant. The probabilities are largely required to be in the same order of magnitude so

this property largely discards any ontic theories which produce large narrow fluctuations in $\rho(\lambda|\psi)$.

Any theory satisfying the equipartition property also has another useful property stemming from quantum communication. This additional property states that should an epistemic theory satisfy the equipartition property that the theory collapses to a ψ-ontic theory in the limit of infinite qubits, as long as the asymptotic communication complexity of the theory grows faster than 2^n, where n is the amount of qubits. As of right now the consequences of these results have yet to be fully explored but it could be the case that they have consequences in a recent debate on the reality of the quantum state [13]. This debate was started by [15] which renewed interest in the subject and further results elaborating on this debate can be found in [21] and [22].

7. Conclusion

In this thesis we examined how to classically simulate a preparation and measurement quantum communication problem exactly. For this we introduced the basic concepts of information science and also elaborated on several concepts in quantum mechanics. Finally a minimization maximization problem was presented to solve our problem and this problem was then analyzed using several results from optimization theory. At the end a connection between ontological theory and quantum communication was via the classical simulations of quantum communication presented in this thesis. This connection will allow for cross correlation between the fields which could lead to interesting results of both fields. In this thesis 2 examples where given of where results in one field could have impact on the other.

An important assumption made in this thesis is that the Hilbert space is finite and thus that no measurement has infinitely many possible outcomes. It seems unlikely that this assumption can be dropped without making the amount of required communication becoming infinite but this is not proven. However any practical implementation of quantum communication will require that the Hilbert space is finite so this assumption does not limit the theory in practical applications. On the other hand it could be present a problem in the applications of the communication complexity of various quantum games to ontological theories.

An open question presented in Section 5 was whether or not strong duality holds when we consider infinite qubits and measurements being usable by Bob and Alice. A strategy to proof this was presented but not worked out further. This then immediately presents a good subject for further research which could have implications beyond just classical simulations of quantum games and have impact on other optimization problems concerning infinitely many variables.

As of this moment when the thesis was written C_{min}^{asym} has only been evaluated for a handful of situations in [6][13][18]. The only channels which have been evaluated thus far are the noiseless and the binary quantum depolarizing channel so research could focus on finding C_{min}^{asym} of other channels. However for the noiseless quantum channel only upper and lower bounds have been foundbounds for C_{min}^{asym} and for the depolarizing C_{min}^{asym} has only been found in the case where quantum states are restricted to the XY plane on the Bloch sphere. Especially finding C_{min}^{asym} for the depolarizing channel is of interest as this channel is used in quantum teleportation which plays an important in a potential quantum internet.

Lastly it was presented that there is a strong link between ψ-epistemic theories and communication complexity. A hypothesized lower bound for C_{min}^{asym} of a noiseless quantum channel could replace an assumption for the PRB no go theorem with a weaker assumption thus making it even more unlikely that a ψ-epistemic theory could ever represent quantum mechanics.

All mathematical methods presented in this thesis are also applicable to contexts out of quantum mechanics. Any communication problem represented by 2 parties receiving an input and communicating can be simulated classically and the amount of bits required for such a process is evaluated the exact same way as for the C_{min}^{asym}. Thus solving the optimization problems presented in this thesis could have an effect on a variety of communication problems.

References

[1] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x.

[2] Francisco Balibrea. On clausius, boltzmann and shannon notions of entropy. *Journal of Modern Physics*, 07:219–227, 01 2016. doi: 10.4236/jmp.2016.72022.

[3] J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(none):175 – 193, 1906. doi: 10.1007/BF02418571. URL `https://doi.org/10.1007/BF02418571`.

[4] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA, 2006. ISBN 0471241954.

[5] C.H. Bennett, P.W. Shor, J.A. Smolin, and A.V. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse shannon theorem. *IEEE Transactions on Information Theory*, 48(10):2637–2655, 2002. doi: 10.1109/TIT.2002.802612.

[6] A. Montina, M. Pfaffhauser, and S. Wolf. Communication complexity of channels in general probabilistic theories. *Phys. Rev. Lett.*, 111:160502, Oct 2013. doi: 10.1103/PhysRevLett.111.160502. URL `https://link.aps.org/doi/10.1103/PhysRevLett.111.160502`.

[7] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. *IEEE Transactions on Information Theory*, 56 (1):438–449, 2010. doi: 10.1109/TIT.2009.2034824.

[8] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[9] John Lindsay Orr. The pauli marices and the bloch sphere, 2018. URL `http://www.johnorr.us/math/pauli_matrices.html`.

[10] Garry Bowen and Sougato Bose. Teleportation as a depolarizing quantum channel, relative entropy, and classical capacity. *Phys. Rev. Lett.*, 87:267901, Dec 2001. doi: 10.1103/PhysRevLett.87.267901. URL `https://link.aps.org/doi/10.1103/PhysRevLett.87.267901`.

[11] John von Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100: 295–320, 1928.

[12] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[13] Alberto Montina and Stefan Wolf. Necessary and sufficient optimality conditions for classical simulations of quantum communication processes. *Phys. Rev. A*, 90:012309, Jul 2014. doi: 10.1103/PhysRevA.90.012309. URL `https://link.aps.org/doi/10.1103/PhysRevA.90.012309`.

[14] E. Specker Simon Kochen. The problem of hidden variables in quantum mechanics. *Indiana Univ. Math. J.*, 17:59–87, 1968. ISSN 0022-2518.

[15] Matthew F. Pusey, Jonathan Barrett, and Terry Rudolph. On the reality of the quantum state. *Nature Physics*, 8(6):475–478, may 2012. doi: 10.1038/nphys2309. URL https://doi.org/10.1038%2Fnphys2309.

[16] Harald Atmanspacher and Hans Primas. *Epistemic and Ontic Quantum Realities*, pages 301–321. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003. ISBN 978-3-662-10557-3. doi: 10.1007/978-3-662-10557-3_20. URL https://doi.org/10.1007/978-3-662-10557-3_20.

[17] R. Hermens. How Real are Quantum States in ψ -Ontic Models? *Foundations of Physics*, 51(2):38, April 2021. doi: 10.1007/s10701-021-00448-7.

[18] Alberto Montina. Epistemic view of quantum states and communication complexity of quantum channels. *Phys. Rev. Lett.*, 109:110501, Sep 2012. doi: 10.1103/PhysRevLett.109.110501. URL https://link.aps.org/doi/10.1103/PhysRevLett.109.110501.

[19] A. Montina. Exponential complexity and ontological theories of quantum mechanics. *Phys. Rev. A*, 77:022104, Feb 2008. doi: 10.1103/PhysRevA.77.022104. URL https://link.aps.org/doi/10.1103/PhysRevA.77.022104.

[20] Alberto Montina. Communication complexity and the reality of the wave function. *Modern Physics Letters A*, 30(01):1530001, jan 2015. doi: 10.1142/s0217732315300013. URL https://doi.org/10.1142%2Fs0217732315300013.

[21] Roger Colbeck and Renato Renner. Is a system's wave function in one-to-one correspondence with its elements of reality? *Phys. Rev. Lett.*, 108:150402, Apr 2012. doi: 10.1103/PhysRevLett.108.150402. URL https://link.aps.org/doi/10.1103/PhysRevLett.108.150402.

[22] Maximilian Schlosshauer and Arthur Fine. Implications of the pusey-barrett-rudolph quantum no-go theorem. *Phys. Rev. Lett.*, 108:260404, Jun 2012. doi: 10.1103/PhysRevLett.108.260404. URL https://link.aps.org/doi/10.1103/PhysRevLett.108.260404.

A. Basics of Quantum Mechanics

The basis of quantum mechanics relies upon several postulates. Three of these postulates are relevant for this thesis and are thus presented in the following sections. In addition POVM formalism gets introduced as an alternative method of formulating the second postulate.

A.1 Quantum States

To start off we introduce the first postulate of quantum mechanics, which describes the way we mathematically represent the state of a particle.

Postulate 1*: Associated with any isolated physical system is a complex vector space associated with an inner product, also known as a* ***Hilbert space****, which is known as the* ***state space*** *of the system. The system is completely described by the unit vectors of the state space which are known as the* ***state vectors*** *of the system.*

To clarify the above we will introduce the quantum equivalent of the binary digit: the qubit. The qubit is associated with a 2 dimensional Hilbert space with the regular dot product. Thus we can denote any qubit as

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle \qquad \alpha, \beta \in \mathbb{C}. \tag{A.1}$$

Where we utilise the orthonormal basis $\{|0\rangle, |1\rangle\}$ with $|0\rangle = (1,0)^T$ and $|1\rangle = (0,1)^T$. As we will see later the value $|\alpha|^2$ represents the chance that an observer measures that the quantum state is in the state $|0\rangle$ and similarly for β and $|1\rangle$. So the state vectors of this space are thus all vectors obeying the equation $|\alpha|^2 + |\beta|^2 = 1$.

A.2 Quantum Measurement

Now with a way to describe a quantum particle in isolation we want to know how the particle changes once an observer interacts with it, or put differently how it changes when we want to measure one of its properties.

Postulate 2*: A quantum measurement is described by a set* $\{M_m\}$ *of* ***measurement operators****. These are operators acting on the state space of the quantum system. The index m refers to the different outcomes of the measurement being performed. The probability that result m occurs is given by*

$$p(m) = \left\langle \psi \middle| M_m^\dagger M_m \middle| \psi \right\rangle \tag{A.2}$$

and the post measurement quantum state is given by

$$\frac{M_m |\psi\rangle}{\sqrt{p(m)}}. \tag{A.3}$$

The measurement operators must satisfy the ***completeness relation:***

$$\sum_m M_m^\dagger M_m = 1. \tag{A.4}$$

The completeness relation represents that the probabilities of the outcomes of the measurement must sum to 1. Measurements are important to quantum systems as they are the only way to physically extract values from a quantum system.

An easy but important measurement for qubits is described by the measurement operators $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ with outcomes 0 and 1. The probability of obtaining 0 is

$$p(0) = \left\langle \psi \middle| M_0^\dagger M_0 \middle| \psi \right\rangle = |\alpha|^2 \tag{A.5}$$

and similarly $p(1) = |\beta|^2$. The respective post measurement states of the measurements are $|0\rangle$ and $|1\rangle$.

Measurements explain us why quantum systems cannot be used to convey infinite information, as Eq. A.1 implies that there are uncountably many normalized qubits thus a single qubit can encode infinite information. But as we cannot directly extract the values of α and β and only approximate these values by successive measurements they carries infinite information.

A.2.1 POVM formalism

It is often the case in quantum computing that one is only performing a measurement to learn something about the current state of the quantum system and is not concerned with the state after measurement. This is the case in our quantum version of the black box game as Bob outputs the outcome of the measurement and does nothing with the post measurement state. In this case it helps to helps to define the quantity

$$E_m = M_m^\dagger M_m \tag{A.6}$$

where M_m is a measurement operator. As one can see this allows us to rewrite Eq. A.2 as

$$p(m) = \langle \psi | E_m | \psi \rangle . \tag{A.7}$$

Two important properties of E_m are

1. $\sum_m E_m = I$.
2. E_m is a positive semi definite operator.

The second property comes from that $p(m) \geq 0$ for every quantum state ψ. Thus the measurement described by $\{E_m\}$ is called a POVM (positive operator-valued measure) with the elements E_m being called POVM elements. One may suspect that given a POVM that one cannot know what the post-measurement state is but this turns out to be wrong, it still is possible to create a general measurement from a POVM and vice versa. These POVM's are simply an easier way to examine the statistical properties of a qubit without worrying about its post-measurement state.

We now understand everything we need to know about the input Alice and Bob receive and in turn what the output of the problem is. But now we will turn our attention to the how Alice and Bob communicatie: the quantum channel.

A.3 Quantum Time Evolution

Measurement only describes half of all the changes a quantum state can undergo. While measurement describes a way for a open quantum state to evolve we also want to know how a quantum state changes if there is no outside world to interact with it, or rather the system is isolated. How then do quantum states evolve when there is no external interaction such as a measurement?

Postulate 3: *The time evolution of a closed quantum system $|\psi_1\rangle$ at time t_1 to the quantum state $|\psi_2\rangle$ at time t_2 is given by a unitary transformation U dependent on $t_2 - t_1$. That is,*

$$|\psi_2\rangle = U\,|\psi_1\rangle\,. \tag{A.8}$$

The isolated time evolution play an important roll in quantum noise and thus in quantum channels. As when a particle is being transmitted through a quantum channel, thus no longer isolated, we can still view the combined system of the entire environment and sent particle as being isolated and thus the 3rd postulate applies on this new total system. We now possess two possible ways for a quantum system to evolve, either through measurement or via time evolution.