Quantum Entanglement in Time

by

Del Rajan

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of
Doctor of Philosophy

Victoria University of Wellington
2020

Abstract

This thesis is in the field of quantum information science, which is an area that reconceptualizes quantum physics in terms of information. Central to this area is the quantum effect of entanglement in space. It is an interdependence among two or more spatially separated quantum systems that would be impossible to replicate by classical systems. Alternatively, an entanglement in space can also be viewed as a resource in quantum information in that it allows the ability to perform information tasks that would be impossible or very difficult to do with only classical information. Two such astonishing applications are quantum communications which can be harnessed for teleportation, and quantum computers which can drastically outperform the best classical supercomputers.

In this thesis our focus is on the theoretical aspect of the field, and we provide one of the first expositions on an analogous quantum effect known as entanglement in time. It can be viewed as an interdependence of quantum systems across time, which is stronger than could ever exist between classical systems. We explore this temporal effect within the study of quantum information and its foundations as well as through relativistic quantum information.

An original contribution of this thesis is the design of one of the first quantum information applications of entanglement in time, namely a quantum blockchain. We describe how the entanglement in time provides the quantum advantage over a classical blockchain. Furthermore, the information encoding procedure of this quantum blockchain can be interpreted as non-classically influencing the past, and hence the system can be viewed as a 'quantum time machine.'

Acknowledgements

Somtimes a few words carry more weight: There have been an eclectic set of mentors and books in my life that were pivotal to my positive development. However there is truly one mentor who I owe so much to, and who has impacted my life the most in a positive way. And that is my supervisor, Professor Matt Visser. Matt gave me an opportunity when I needed it the most in my life; I am not really sure where I would have been without it. He taught me so much about theoretical physics, supervised me with care, and his work especially on wormholes inspired me. I am extraordinarily grateful to be trained under this great scientist.

Thank you Matt.

Dedicated to Albert Einstein[1]

[1]It was a temporal narrative (the twin paradox) from his theories of relativity that shocked a less than ordinary boy to first open the door to the magical world of theoretical physics.

Contents

One of the fourteen untitled pieces from Francisco Goya's *Black Paintings* series.

"Quantum mechanics: Real Black Magic Calculus" - Albert Einstein

1

Introduction

"Anyone who is not shocked by quantum theory has not understood it."

– Niels Bohr, co-inventor of quantum theory

THIS THESIS explores the most shocking temporal effects in quantum physics. These recently discovered phenomena violently overthrow the classical picture of the world [1]. Each of these effects has been described as an 'entanglement in time' and yet arise from different contexts within quantum physics. This thesis collects these results to provide one of the first systematic expositions on entanglement in time, which also entails a comparison with the extensively researched entanglement in space. Furthermore, this project is carried out using the theoretical framework of quantum information science [2].

Quantum information science reconceptualizes quantum physics in terms of information. It is the theoretical and experimental study of quantum information and its applications. It can be regarded as a fundamental subject in that it distils questions on the nature of quantum physics to distinctions between quantum information and classical information. One very powerful advantage of quantum information is that it can contain a resource known as entanglement in space. As a result, quantum information has the ability to perform information tasks that would be impossible or very difficult to do with only classical information. One prominent example of such tasks is quantum communication which can be used to teleport quantum information. Another remarkable example are quan-

tum computers which can be shown to efficiently solve problems that would be infeasible to perform on any classical computer that could ever be built. Both of these applications of entanglement in space are major research programs, and hence emphasizes the crucial role of this resource in quantum information.

Apart from its importance, entanglement in space is a perplexing phenomenon when expressed in terms of the physical systems that instantiate the quantum information. It can be described as an interdependence among two or more spatially separated quantum information systems in which any one system can instantaneously affect the other systems that can in principle be arbitrarily far.

Entanglement in time, of which we bring to light in this thesis, corresponds to an analogous effect. It can be viewed as an interdependence of quantum information systems across time, which is stronger than could ever exist between classical information systems; it can even arise for the case of a single system (across multiple times). Moreover, the interpretations associated with each manifestation of the effect are far more bizarre; for example, a newly created photon can *affect the physical description of a photon in the past* that has long since been destroyed; in another scenario, a quantum detector that is switched on and off at say quarter to 12:00 can form a non-classical interdependence with another detector at the same spatial location in the future, but *only if the future detector waits* to be switched on and off at precisely quarter past 12:00.

To gain a deeper understanding of these effects, a natural question that arises is if there is a common trait that marks these various entanglement phenomena as shocking? This is a challenging question given that these effects are theoretically expressed using different mathematical areas; we can conceive of a large number of factors that contribute to its radical departure from classical properties. This project provides an insight towards an answer, and captures this in the form of the following overarching theme of this thesis: *The interdependence in any entanglement in space is shocking due to the absense of a time interval involved. The interdependence in any entanglement in time is shocking due to the existence of a time interval involved.* To elaborate on this observation, in an entanglement in space the ability of a system to instantaneously affect a distant system signifies a lack of a time interval. Introducing a time interval in this scenario will only

make the effect clash less harshly with our classical intuition (as it allows for an explanation involving hidden causal signals of some form or the other); such an insight regarding this spatial case was first expressed in the concluding remarks in [3]. However in an entanglement in time, it is precisely the introduction of a time interval that makes the effect completely unpalatable to the mind of a classical physicist. This is non-trivial as it was already noted that a time interval allows for a classical dependence among systems across time in the form of a causal relationship. We aim to provide a compelling case for this theme.

In quantum information science, both quantum communications and quantum computers are established applications of entanglement in space. A central aim of the field is the creation of new quantum information applications. In this thesis, we make an original contribution by designing one of the first novel applications of entanglement in time, namely a quantum blockchain. In our mathematical design, we show that the entanglement in time (as opposed to an entanglement in space) provides the quantum advantage over a classical blockchain. Furthermore, the information encoding procedure of this quantum blockchain can be interpreted as non-classically influencing the past, and hence the system can be viewed as a 'quantum time machine.' This advancement forms one piece of the various original works and insights presented in this thesis.

Rather than provide a chronological presentation, our thesis will place the entanglements in quantum information science as the conceptual core and coherently organize the diverse topics as backgrounds or extensions of this core. We believe this approach achieves the most clarity. Hence the structure of this thesis is as follows: In Chapter 2, we provide mathematical descriptions of classical information and highlight three of its applications. These fall under the respective sections of classical communications, classical computers and classical blockchains. In Chapter 3, we introduce quantum information and contrast this with the properties of classical information. As a result, it contains a description of the mathematical tools of quantum information science, and this will closely follow the material in [2] along with recent developments; from the perspective of a theoretical physicist, this subject can be viewed as an information-theoretic reformulation of non-relativistic quantum mechanics. Chapter 4 is the core chap-

ter of this thesis which starts by describing entanglement in space using the tools obtained in the previous chapter. Furthermore, it introduces quantum communications and quantum computers which are the applications of entanglement in space. We proceed to describe the central topic of entanglement in time, and its various non-classical properties. We conclude this chapter by presenting a mathematical design of the quantum blockchain which is an application of entanglement in time. All three quantum applications will be contrasted with the classical case. We proceed to Chapter 5, where the notions of both entanglement in space and entanglement in time are extended to the subject of quantum foundations. These are respectively described in sections titled nonlocality in space and nonlocality in time. The emphasis will be placed on the aspects of quantum foundations which shares an interface with quantum information science. In Chapter 6, we see entanglement in space and entanglement in time manifesting itself in the relativistic regime. These effects are respectively termed spacelike entanglement and timelike entanglement in the subject of relativistic quantum information. One can think of this new subject as placing quantum information science within the broader framework of quantum field theory (in flat and curved spacetimes). Finally in Chapter 7, we provide a conclusion which includes a discussion on future research projects concerning entanglement in time.

Though entanglement in time may resemble some of the concepts of closed timelike curves, we exclude a detailed study of the latter for two reasons. The first reason is that closed timelike curves have already been extensively reviewed within relativity [4] as well as in quantum theory [5]; whereas this thesis is one of the first to systematically compile the diverse literature on entanglement in time. The second reason is that unlike closed timelike curves, the effects of entanglement in time have been experimentally verified for a number of cases, thereby warranting itself as a distinct phenomenon.

Although we refer to various literature in experimental physics, information theory, and computer science, this thesis falls under the theoretical physics aspect of quantum information science. Hence, the focus is solely on the mathematics with an emphasis on the physical theories.

2

Classical Information

"A [classical] computer on every desk and in every home."

– Microsoft's founding vision statement

THE MODERN CONCEPT of classical information will be articulated against the backdrop of three technological applications. The associated mathematical models in each of the three cases provide an abstraction for how this information is represented and transformed. For the sole study of information, this abstraction provides the necessary framework to ignore the details of the various physical systems used which store that information.

Of particular importance to this thesis is the notion of systems exhibiting *interdependence* with each other. In the realm of classical information, we will find this to be naturally captured through the constructs of probability theory.

2.1 Review of Probability Theory

In this thesis, probability theory will prove to be essential in two primary ways:

i) To assist in mathematically defining classical information.
ii) To understand the probabilities derived from quantum information.

2.1.1 Single random variable

The fundamental object of probability theory is the random variable, which we denote as X. The random variable can take one of a number of values, x, with respective probabilities $p(X = x)$; we limit ourselves to the case where the values form a finite set; we also use the convention that $p(x)$ can represent $p(X = x)$.

The expectation value of X is defined as

$$\mathbb{E}(X) \equiv \sum_x p(x)\, x. \tag{2.1}$$

It is a particular type of mean for all the values the random variable can take. Furthermore if a and b are constants, then it can be shown that $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.

The variance and standard deviation are respectively defined as

$$\mathrm{Var}(X) \equiv \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - \mathbb{E}(X)^2, \tag{2.2}$$

$$\Delta(X) \equiv \sqrt{\mathrm{Var}(X)}. \tag{2.3}$$

Both are statistical measures of the 'spread' of values about the average. One advantage of using the standard deviation, as opposed to the variance, is that it has the same units as the expectation value.

2.1.2 Multiple random variables

When considering the case of more than one random variable, several new constructions can be introduced. Suppose X and Y are random variables. Then the probability that $X = x$ *and* $Y = y$ is known as the joint probability,

$$p(X = x, Y = y). \tag{2.4}$$

An equivalent notation is simply $p(x, y)$.

The conditional probability that $X = x$ given that $Y = y$ is defined as

$$p(X = x \mid Y = y) \equiv \frac{p(X = x, Y = y)}{p(Y = y)}. \tag{2.5}$$

Bayes' rule allows one to 'invert' conditional probabilities

$$p(X = x \mid Y = y) = p(Y = y \mid X = x) \frac{p(X = x)}{p(Y = y)}. \tag{2.6}$$

Given two random variables, the law of total probability provides an alternative way to calculate probabilities of one of the variables,

$$p(Y = y) = \sum_x p(Y = y \mid X = x)\, p(X = x). \tag{2.7}$$

The sum is over all values that the other random variable can take.

The expectation value for two random variables is the rather simple result

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y). \tag{2.8}$$

2.1.3 Independent random variables

The pertinent question, from the view of this thesis, is how to describe random variables where the realization of one has no effect on the other? This can be enunciated by the following mathematical definition: Random variables X and Y are said to be *independent* if

$$\boxed{p(X = x, Y = y) = p(X = x)\, p(Y = y), \quad \forall x, y.} \tag{2.9}$$

When viewed through the concepts from the previous subsection, it is not difficult to see that if X and Y are independent random variables, then:

$$p(Y = y \,|\, X = x) = p(Y = y) \quad \forall x, y \tag{2.10}$$

$$\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y), \tag{2.11}$$

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y). \tag{2.12}$$

A further consequence from considering independent random variables is the following theorem.

Theorem 2.1. ***(Law of large numbers)*** *Suppose $X_1, X_2, \ldots$ are independent random variables that all have identical probability distributions as X, where $|\mathbb{E}(X)| < \infty$ and $|\mathbb{E}(X^2)| < \infty$. Then for any $\epsilon > 0$, and where $S_n \equiv \sum_{i=1}^{n} X_i/n$, we have that $p(|S_n - \mathbb{E}(X)| > \epsilon) \to 0$ as $n \to \infty$. (See e.g. [2].)* □

The utility of this result arises in many applications such as in games involving chance. Hence, it is of no surprise that origins of probability theory can be traced to systemic study of dice games [6].

2.1.4 Application: Monty Hall game

We devote this subsection to applying some of the reviewed concepts to the well known Monty Hall game [7, 8, 9, 10]. This game is perhaps the most bizarre application of classical probabilities.

a) Classic Monty Hall game: A character named Monty hosts a game show. There are three doors respectively labelled $\{1, 2, 3\}$. There is a car prize behind one door, and goats behind the remaining two. We let a random variable A represent the prize door which can take value i from the set of door labels. We assume in the game that when a random choice needs to be made, all options are chosen with equal probability. Note that this implies that the choice for prize door has probabilities $p(A = i) = 1/3$ for each value i.

The contestant on the show, who doesn't know which door the prize is behind, is given a choice to pick a door; we represent the chosen door using a random

variable denoted B which can take value j from the set of door labels. Provided this is a random choice, we have $p(B = j \mid A = i) = 1/3$ for all values i, j.

Next, Monty who knows where the prize is, has to open a goat door, which we represent using random variable C; in a similar manner, the variable takes value k from the set of door labels. But unlike the previous choices, Monty's decision is constrained through the game rule that he is not allowed to open the door chosen by the contestant. From this, we derive the following conditional probabilities:

$$p(C = k \mid B = j, A = i) = \begin{cases} \frac{1}{2}, & \text{if } i = j \neq k \\ 1, & \text{if } i \neq j \neq k \\ 0, & \text{otherwise} \end{cases} \tag{2.13}$$

Once a goat door is opened, Monty offers the contestant the option to stick with the original choice, or alternatively switch to the other unopened door. By sticking, the contestant's probability of opening the prize door is $1/3$. Counter-intuitively, by switching doors, the probability of winning increases to $2/3$.

This can be seen by proceeding to compute the non-zero joint probabilities

$$p(A = i, B = j, C = k) = p(C = k \mid B = j, A = i)\, p(B = j \mid A = i)\, p(A = i). \tag{2.14}$$

Then we sum the joint probabilities corresponding to the combination of door labels where the contestant would win by switching. This leads to the desired result

$$p(\text{win if switch}) = \sum_{i \neq j \neq k} p(A = i, B = j, C = k) = \frac{2}{3}. \tag{2.15}$$

b) Ignorant Monty Hall game: Let us consider the case where Monty does not know what lies behind any of the doors. Nonetheless, we still have $p(A = i) = 1/3$, and also $p(B = j \mid A = i) = 1/3$ for all values i, j. The only constraint as in the Classic game is that Monty cannot open the door chosen by the contestant.

This means that (2.13) is modified to

$$p(C = k \mid B = j, A = i) = \begin{cases} 0, & \text{if } j = k \\ \frac{1}{2}, & \text{otherwise} \end{cases} \tag{2.16}$$

Unlike the previous case, there is a probability in this scenario that Monty opens the prize door by accident; this be seen as the set of cases where $i = k$:

$$p(\text{opens prize door}) = \sum_{i=k\neq j} p(A = i, B = j, C = k) = \frac{1}{3}. \tag{2.17}$$

By respecting that probabilities sum to unity, we derive from (2.17) that the probability Monty opens a goat door is $2/3$. The joint probability that Monty opens a goat door and the contestant wins by switching doors can be computed to be $1/3$. Substituting the last two values into the conditional probability formula (2.5), we obtain

$$p(\text{win if switch} \mid \text{opens goat door}) = \frac{1/3}{2/3} = \frac{1}{2}. \tag{2.18}$$

Thus, in this modified game, the contestant essentially acquires the same probability of winning whether a choice to switch is made or not.

2.2 Classical Communication

Classical information theory [11] is a powerful application of probability theory. It arose from considering engineering problems associated with classical communication systems. The central mathematical object of the subject is the Shannon entropy. It turns out that there are two rather separate ways to interpret this quantity; the first is derived on an intuitive notion of what properties information should have; the second is based on an operational definition in terms of data compression. For an extensive treatment on the subject, refer to [2, 12, 13].

2.2.1 Information content

Consider a random variable X which can take one of the values x with respective probabilities $p(x)$. The information content of x is defined as

$$I(x) \equiv -\log_2(p(x)). \tag{2.19}$$

This mathematical definition captures the intuition that the occurrence of a value associated with a lower probability provide a greater 'information' gain than the occurrence of a value associated with a larger probability.

a) Single random variable: Generalizing (2.19) to the case of the random variable gives $I(X) = -\log_2(p(X))$. The Shannon entropy of X is defined as the expectation value of $I(X)$:

$$H(X) \equiv \mathbb{E}(I(X)) = -\sum_x p(x) \log_2 p(x). \tag{2.20}$$

This quantity is a function of only the probability distribution. We take the convention that $0 \log_2 0 \equiv 0$, which is supported through $\lim_{x\to 0} x \log_2 x = 0$. Within this intuitive definition, there are three ways to view the Shannon entropy:

i) It represents the information content of random variable X.
ii) It quantifies the information gained after we learn the value of X.
iii) It measures the uncertainty before we know the value of X.

It can be shown that the entropy has the bounds $0 \leq H(X) \leq \log_2 d$, where d is the number of values X can take.

b) Multiple random variables: By extracting the notions developed in probability theory, one can develop various information-theoretic constructions for multiple random variables. An example of this is the joint entropy of random variables X and Y, which is defined as

$$H(X, Y) \equiv -\sum_{x,y} p(x, y) \log_2 p(x, y). \tag{2.21}$$

The joint entropy corresponds to the total uncertainty of both the variables con-

sidered. Using (2.21), we define the conditional entropy, of X conditioned on Y, as

$$H(X|Y) \equiv H(X, Y) - H(Y). \tag{2.22}$$

It can be interpreted as the remaining uncertainty of X once the value of Y is known. A quantity of great importance is the mutual information as it provides a way to measure how much information X and Y have in common:

$$H(X : Y) \equiv H(X) + H(Y) - H(X, Y). \tag{2.23}$$

Next, consider the case where $p(x)$ and $q(x)$ are probability distributions over the same index set, x. The relative entropy provides measure of 'distance' between these distributions; it is defined (from $p(x)$ to $q(x)$) as

$$H(p(x)||q(x)) \equiv \sum_x p(x) \log_2 \frac{p(x)}{q(x)}. \tag{2.24}$$

Given our emphasis on temporal phenomena in this thesis, we want to consider the relationships between random variables across time. This can be exemplified by a Markov chain, which is a sequence of random variables $X_1 \to X_2 \to \cdots$ such that

$$p(X_{n+1} = x_{n+1} \,|\, X_n = x_n, \ldots, X_1 = x_1) = p(X_{n+1} = x_{n+1} \,|\, X_n = x_n). \tag{2.25}$$

c) Properties: We list out some elementary properties including how the con-

sidered quantities relate to one another:

$$H(X:Y) = H(X) - H(X|Y), \tag{2.26}$$
$$H(X,Y) = H(Y,X), \tag{2.27}$$
$$H(X:Y) = H(Y:X), \tag{2.28}$$
$$H(Y|X) \geq 0, \tag{2.29}$$
$$H(X) \leq H(X,Y), \tag{2.30}$$
$$H(X,Y) \leq H(X) + H(Y), \tag{2.31}$$
$$H(Y|X) \leq H(Y), \tag{2.32}$$
$$H(X:Y) = H(p(x,y)||p(x)p(y)), \tag{2.33}$$
$$H(X,Y,Z) + H(Y) \leq H(X,Y) + H(X,Z) \tag{2.34}$$
$$H(X|Y,Z) \leq H(X|Y). \tag{2.35}$$

Along with that, the chaining rule for conditional entropies is a result that relates random variable Y to a set of random variables $X_1, \ldots, X_n$ in the following way

$$H(X_1, \ldots, X_n|Y) = \sum_{i=1}^{n} H(X_i|Y, X_1, \ldots, X_{i-1}). \tag{2.36}$$

With respect to temporal relationships, we expect that once information is lost over time, it is gone forever. This idea is mathematically captured by the data processing inequality: If $X \to Y \to Z$ is a Markov chain, then

$$H(X) \geq H(X:Y) \geq H(X:Z). \tag{2.37}$$

d) Independent random variables: For the special case of random variables that are independent, each of the following are a biconditional property:

$$H(X,Y) = H(X) + H(Y), \tag{2.38}$$
$$H(Y|X) = H(Y), \tag{2.39}$$
$$H(X:Y) = 0. \tag{2.40}$$

2.2.2 Data compression

The fundamental results of classical information theory are the noiseless channel coding theorem and the noisy channel coding theorem; the former is concerned with the problem of compressing a message in a communication channel; the latter quantifies the reliability of transmitting that message over a noisy channel.

However, our focus is solely on the noiseless coding theorem as it provides an operational definition of the Shannon entropy. Instead of viewing $H(X)$ as the information content of X, it will be seen as the minimal physical resource necessary and sufficient to reliably store the output of a classical information source.

a) Defining an information source: In order to derive the noiseless coding theorem, we define a classical information source as a sequence of random variables $(X_1, X_2, \dots)$. The output of the source are the values the variables take. Furthermore, we assume the variables are independent and have identical distributions, which we abbreviate as *i.i.d.* Hence we have $H(X) \equiv H(X_1) = H(X_2) = \dots$ Developing on this model, we define a compression scheme of rate R as mapping output $x = (x_1, \dots, x_n)$ to a string of length nR, which we represent by $C^n(x) = C^n(x_1, \dots, x_n)$. Conversely, the corresponding decompression scheme, $D^n(C^n(x))$, maps the string of length nR to a string of length n. The compression-decompression scheme is defined to be reliable if the probability that $D^n(C^n(x)) = x$ goes to one as n goes to ∞.

b) Defining typical sequences: The possible outputs of the information source can divided into two sets, namely typical sequences and its complement, atypical sequences. More precisely, given $\epsilon > 0$, a sequence $x_1, \dots, x_n$ is ϵ-typical if it satisfies

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, \dots, x_n) \leq 2^{-n(H(X)-\epsilon)}. \tag{2.41}$$

We can reformulate (2.41) as

$$\left| \frac{1}{n} \log \frac{1}{p(x_1, \dots, x_n)} - H(X) \right| \leq \epsilon. \tag{2.42}$$

We also denote $T(n, \epsilon)$ as the set of of all ϵ-typical sequences of length n.

c) Application of the law of large numbers: In the case of large n, it can be

observed that most sequences are typical. This hypothesis is rigorously proved in the following theorem using the law of large numbers.

Theorem 2.2. ***(Theorem of typical sequences)***

i) Fix $\epsilon > 0$. Then for any $\delta > 0$, for sufficiently large n, the probability that a sequence is ϵ-typical is at least $1 - \delta$.

ii) For any fixed $\epsilon > 0$ and $\delta > 0$, for sufficiently large n, the number of ϵ-typical sequences, $|T(n, \epsilon)|$, satisfies

$$(1 - \delta)\, 2^{n(H(X)-\epsilon)} \leq |T(n, \epsilon)| \leq 2^{n(H(X)+\epsilon)}. \tag{2.43}$$

iii) Suppose $R < H(X)$. Let $S(n)$ be a collection of size at most 2^{nR}, of length n sequences from the source. Then for any $\delta > 0$ and for sufficiently large n,

$$\sum_{x \in S(n)} p(x) \leq \delta. \tag{2.44}$$

□

Proof. (See e.g. [2].)

i) Given that X_i are a set of *i.i.d* random variables, this implies $-\log p(X_i)$ are also a set of *i.i.d* random variables. Using the law of large numbers (Theorem (2.1)), we have for any $\epsilon > 0$ and $\delta > 0$ for sufficiently large n that

$$p\left(\left|\sum_{i=1}^{n} \frac{-\log p(X_i)}{n} - \mathbb{E}(-\log_2 p(X))\right| \leq \epsilon\right) \geq 1 - \delta. \tag{2.45}$$

Using (2.20), we can substitute $H(X)$ for $\mathbb{E}(-\log_2 p(X))$. Furthermore using the product property of logarithms, we have that $\sum_{i=1}^{n} \log p(X_i) = \log(p(X_1, \ldots, X_n))$. This modifies (2.45) to give the desired result that the probability a sequence is ϵ-typical is at least $1 - \delta$:

$$p\left(\left|\frac{1}{n} \log \frac{1}{p(X_1, \ldots, X_n)} - H(X)\right| \leq \epsilon\right) \geq 1 - \delta. \tag{2.46}$$

ii) The sum of the probabilities of the typical sequences cannot be greater than

one. Along with (2.41), we see that

$$1 \geq \sum_{x \in T(n,\epsilon)} p(x) \tag{2.47}$$

$$\geq \sum_{x \in T(n,\epsilon)} 2^{-n(H(X)+\epsilon)} \tag{2.48}$$

$$= |T(n,\epsilon)| 2^{-n(H(X)+\epsilon)}. \tag{2.49}$$

Therefore, we obtain that $|T(n,\epsilon)| \leq 2^{n(H(X)+\epsilon)}$. Conversely, from (2.46), we can also deduce that the sum of the probabilities of typical sequences must be at least $1-\delta$. Under this requirement, along with (2.41), we can write

$$1 - \delta \leq \sum_{x \in T(n,\epsilon)} p(x) \tag{2.50}$$

$$\leq \sum_{x \in T(n,\epsilon)} 2^{-n(H(X)-\epsilon)} \tag{2.51}$$

$$= |T(n,\epsilon)| 2^{-n(H(X)-\epsilon)}. \tag{2.52}$$

Hence, we can compute that $|T(n,\epsilon)| \geq (1-\delta) 2^{n(H(X)-\epsilon)}$.

iii) Fix an ϵ such that $R < H(X) - \epsilon$, and $0 < \epsilon < \delta/2$. The total probability for ϵ-atypical sequences in $S(n)$ can be made small, ie less than $\delta/2$, for large enough n. The total number of ϵ-typical sequences is at most 2^{nR} since that is the upper bound for the total number of sequences in $S(n)$. Furthermore, each ϵ-typical sequence has probability at most $2^{-n(H(X)-\epsilon)}$. Therefore, the total probability of ϵ-typical sequences in $S(n)$ is $2^{-n(H(X)-\epsilon-R)}$. Given $R < H(X) - \epsilon$, we can see that $2^{-n(H(X)-\epsilon-R)} \to 0$ as $n \to \infty$. Hence the total probability of sequences in set $S(n)$ is less than δ for sufficiently large n.

■

d) Application of theorem of typical sequences: The usefulness of Theorem (2.2) becomes apparent when proving the main result:

Theorem 2.3. ***(Shannon's noiseless channel coding theorem)*** *Consider an i.i.d. information source represented by $\{X_i\}$, with entropy rate $H(X)$:*

i) If $R > H(X)$, then there exists a reliable compression scheme of rate R for the

information source.

ii) Conversely, if $R < H(X)$, then any compression scheme will not be reliable.

□

Proof. (See e.g. [2].)

i) Consider the case $R > H(X)$. We choose an ϵ such that $H(X) + \epsilon < R$. From Theorem (2.2), we have that for any $\delta > 0$ and for sufficiently large n, there are at most $2^{n(H(x)+\epsilon)} < 2^{nR}$ ϵ-typical sequences produced by the information source. Given that there are most 2^{nR} of such sequences, it only requires nR bits to uniquely identify a particular ϵ-typical output. Hence we can compress the ϵ-typical output, using some scheme, to a string of nR bits which can be decompressed later. Furthermore, using Theorem (2.2), we have that the probability of producing such an ϵ-typical sequences is at least $1 - \delta$. If on the other hand, we have an ϵ-atypical sequence, we declare an error and give up on compression.

ii) Consider the case $R < H(X)$. There are at most 2^{nR} outputs for the combined compression-decompression scheme. Using Theorem (2.2), the probability, for sufficiently large n, of the information output belonging to a subset of the 2^{nR} sequences tends to zero. Hence any compression scheme for this case will not be reliable.

■

e) Comments:

i) The entropy can be operationally defined as the minimum physical resource required to reliably store the output of a classical information source.

ii) The idea is that we only need to compress typical sequences, as they are the outputs that are overwhelmingly likely to occur in the asymptotic limit.

2.3 Classical Computing

The wide proliferation of digital computers across the globe has led to a period in human history known as the 'Information Age.' However, the conception of

these physical devices stemmed from abstract work in the foundations of mathematics [14]. This investigation brought about a mathematical model of computation known as Turing machine, which has since had a profound influence across different spheres of thought[15].

Surprisingly, there are a number of different models of computation which are equivalent to the Turing machine. One such example is the circuit model which we briefly cover in this section. We also look at how one can probe at the resources required for a model to solve a computational problem; this can quantitatively captured by a framework known as the asymptotic notation. For a broader survey on the theory of computation, we refer the reader to [16].

2.3.1 Circuit model

An enormous range of computations can be performed by using a combination of circuits. Circuits are abstractions which can be physically instantiated, most commonly through classical electrical systems. They are composed of three primary elements. The first is that they encode the information in a bit, whose state is either a 0 or a 1. The second element is that circuits are made up of 'wires' which carry that information through space or time. The final piece is that circuits contain logic gates which are a particular application of Boolean logic; more precisely a logic gate is a function $f : \{0,1\}^k \rightarrow \{0,1\}^l$ where k and l respectively denote the number of input and output bits.

We briefly describe various elementary logic gates as follows:

a) NOT: The NOT gate inverts the input value

$$f(a) = 1 \oplus a \tag{2.53}$$

where $\oplus$ represents modulo 2 addition.

b) AND: The AND gate outputs bit 1 if both input values are 1.

c) OR: The OR gate produces output 1 if at least one of the input values are 1.

d) XOR: The XOR gate outputs bit 1 if only one of the input values are 1.

e) NAND : The NAND gate produces the negation of an AND gate.

f) NOR : The NOR gate produces the negation of an OR gate.

Using a combination of these gates, one can construct integrated circuits to solve computational problems with sophisticated mathematical structures. The particular step by step procedure to do so are collectively known as an algorithm for that problem.

However, a related issue to consider are what are the minimal number of gates required to solve a particular problem of interest? More broadly speaking, how does one quantify the resources required by a specific algorithm? Furthermore, is there a limit to the computational capabilities provided by classical resources?

2.3.2 Asymptotic notation

Computational resources can be measured in a multitude of forms depending on the nature of the problem in question. Common examples include the number of evaluations of a function, space requirements (say in the form of memory), time requirements (in terms of run time of an algorithm) or even energy.

For an appropriate framework to analyze specific algorithms, an important consideration is that one cares only about how the resource consumed scales with the 'size' of the corresponding problem. Roughly speaking, each problem has a quantity of interest that can be used to describe the problem, and the magnitude of that quantity represents the size of the problem. As an example, n could be the number of input bits for an algorithm which takes $30n + \log_2 n$ gates to execute. The only term that dominates for large sizes is $30n$ hence we say that the number of operations required scales like n. The asymptotic notation captures this idea.

Suppose $f(n)$ and $g(n)$ are two functions where n is a non-negative integer. With this in mind, one can define the three tools provided by the asymptotic notation.

a) The 'big O': The first tool in the asymptotic notation is the O notation. It quantifies the upper bound on the behaviour of a function. A function $f(n)$ is $O(g(n))$ if there are constants c and n_0 such that for all values of n greater than n_0, $f(n) \leq cg(n)$.

b) The 'big Omega': Conversely, the Ω notation provides a lower bound. A function $f(n)$ is in $\Omega(g(n))$ if there are constants c and n_0 such that for all values of n greater than n_0, $cg(n) \leq f(n)$.

c) The 'big Theta': The final tool is the Θ notation which corresponds to the notion that $f(n)$ and $g(n)$ are similar in the asymptotic regime. More precisely, $f(n)$ is in $\Theta(g(n))$ if it is both $O(g(n))$ and $\Omega(g(n))$.

The asymptotic notation provides a way to quantify the resources used by an algorithm for a specific problem. By harnessing this framework, it allows superior algorithms to be quantitatively expressed in that they use fewer resources than previous ways of solving the relevant problem. The design of such powerful algorithms is one of the central aims in the field of classical computation.

2.4 Classical Blockchain

Information security systems harness concepts from both communication and computing. One prominent example of this class of technologies is the classical blockchain system which stores data securely over time. Furthermore, this task is accomplished among computer nodes in a communication network that do not necessarily trust each other.

The pioneering invention of the blockchain system was first described pseudonymously in [17]. However, many of the individual subsystems draw their inspiration from a large body of disconnected theoretical research [18]. Over recent years, countless variants have been proposed [19], but we devote this section to describing the original design, with an emphasis on the mathematical concepts.

The aim of a blockchain system is to have a single database of records about the past that every node in the network can agree on. Furthermore, it should not require a centralized management node. We start with describing the two primary elements of such a system. The first is the blockchain data structure which encodes the classical information using an algorithm. The second component involves a communication network to provide the decentralization feature. We conclude this section by conveying the essential ideas of public key cryptogra-

phy; this is used in various tasks within the blockchain system.

2.4.1 Blockchain data structure

Records about the past, which occurred at around the same time, are received and collected into a data block. These blocks are time-stamped to ensure that the data existed at the specified time. Furthermore, the blocks are linked in chronological order through mathematical functions known as cryptographic hash functions [20]. We provide a more careful treatment of the linked blocks as follows.

A cryptographic hash function, h, is a deterministic function that maps a string of arbitrary length to a string of fixed length (eg 256 bits). The output is known as the hash digest, $h(x)$. This computing task can be accomplished by various cryptographic hash algorithms (eg SHA-256).

The function, h, satisfies the following properties:

a) Preimage resistant: It is infeasible through classical computation that given output $d = h(x)$, one can derive the input string x. This gives the implication that the function is one-way. This is based on the assumption that the search space of outputs is large.

b) Second preimage resistant: It is infeasible through classical computation given that given input x, one can find $y \neq x$, such that $h(x) = h(y)$.

c) Collision resistant: It is infeasible through classical computation to find any two inputs, x and y, that produce the same digest $h(x) = h(y)$.

d) Efficient: It requires polynomial (ideally linear) computational resources to compute the digest, d, given the size of input x.

e) Pseudo-random: If one modifies any of the bits in the input, x, it has a significant unpredictable change in the output of $h(x)$.

Using these mathematical properties, each block, with its string of bits, is mapped using the hash function to a specific digest. More crucially, each block's data contains the hash digest of the previous block. This latter property provides the required notion of a 'chain' of blocks, resulting in the term blockchain.

This *interdependence of the time-stamped blocks*, through the cryptographic hash functions, provides the necessary sensitivity for the role of securing the records in a blockchain. Any party that attempts to falsify the past records in a block would need to find a way to alter the data such that it does not change the digest of that block. This task, as we have mentioned, is computationally infeasible. Hence, the resulting change in the digest of the tampered block would cause all subsequent blocks to have different digests. This is due to the design that each block's data contains the digest of the previous block. Hence, the consequence of this sensitivity is that altering the data in a block would tamper all subsequent blocks and hence invalidate them. Furthermore, given that only future blocks following the tampered block are invalidated, this implies that the the older the time stamp on the block, the more secure it is in the blockchain. In summary, the blockchain data structure provides a tamper proof system for storing records, precisely because tampering with it can easily be detected.

2.4.2 Network consensus protocol

Along with a blockchain data structure, the second part to the system is a classical communication network. Each node on the network carries a local copy of the blockchain data structure. This provides the mechanism if one local copy is destroyed, other nodes with local copies would serve to provide replication.

However, the primary objective of the network component is to add valid blocks to each local copy without a centralized management node. The challenge of the task is that it must be accomplished without the assumption that all the nodes are 'honest.' Typically, this involves invoking a node on the network to confirm the validity of records in a new block, and then communicating that block to other nodes on the network. The different nodes accept the block if the block is valid and they can successfully link it to their own local copy of the blockchain data structure through the cryptographic hash functions. For this procedure to maintain ongoing accuracy, the validating node gets chosen at random for each block; this prevents preplanned node-specific attacks. Furthermore, the validating node is also incentivised through the network for carrying out these tasks. Despite some dishonest nodes, this is all successfully accomplished through a

non-trivial consensus protocol.

In the original design, the consensus protocol is coined 'proof-of-work' or is also known as the Nakamoto consensus. In this scenario, the node that successfully validates the block has to expend a specific amount of computational resource. This resource is used to solve a tractable problem involving the hash digest associated to the new block in question. After the node verifies the validity of the block, it is rewarded by an economic incentive.

However, the consensus protocols in the blockchain systems do not fit into the traditional framework of fault-tolerant distributed computing [18, 21]. More specifically, it is not rigorously clear that 'proof-of-work' satisfies a security standard known as BFT (Byzantine Fault Tolerance) [22]. In this setting, byzantine nodes refer to computer nodes that may take arbitrary actions such as sending faulty messages, as opposed to crash failure nodes which fail by stopping. An well known example of a BFT protocol in the fault-tolerant literature is PBFT (practical Byzantine fault tolerance) [23].

2.4.3 Public key cryptography

Public key cryptography forms the security backbone of the classical information infrastructure of the modern world. In the specific case of blockchain technologies, it is most notably implemented for digitally signing the records in a block [24]. The subject of public key cryptography is infeasible to cover in a short section, and hence we refer the reader to [25] for a deeper mathematical coverage. We limit our discussion to the RSA (Rivest–Shamir–Adleman) public key cryptosystem which relies on ideas extracted from number theory. Furthermore, our aim is to articulate the essential concepts by focusing within the simplified context of two parties wishing to communicate in private.

a) Number-theoretic preliminaries: We briefly digress to results regarding prime numbers and modular arithmetic. Two integers a and b are defined as co-prime if their greatest common divisor is one. The Euler $\varphi(n)$ function is defined to be the number of positive integers less than n which are co-prime to n.

Suppose that n has prime factorization $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ where $p_1, \cdots, p_k$ represent

the distinct prime numbers, and $\alpha_1, \cdots, \alpha_k$ are positive integers. Then one can derive the formula

$$\varphi(n) = \prod_{j=1}^{k} p_j^{\alpha_j - 1}(p_j - 1). \tag{2.54}$$

Furthermore, it can be proven that if a is co-prime to n, then

$$a^{\varphi(n)} = 1 \,(\text{mod } n). \tag{2.55}$$

b) Communication problem: Suppose a party, say 'Alice', wants to transmit a message to another party, say 'Bob', over a classical communication channel. More crucially, they want to ensure that no other party can access the contents of the message. This can be accomplished, with a significant degree of confidence, by invoking the mathematical notions of a public and private key.

c) Encrypting the message: The message Alice wants to transmit is denoted m. She is said to have encrypted her message to c if she performs the computation

$$c = m^e \,(\text{mod } n), \tag{2.56}$$

where the values n and e are collectively known as the public key. These values are generated by Alice. She first selects two large prime numbers p and q. Then she computes $n = pq$. From this, Alice picks an $e \in \mathbb{N}$ such that e is co-prime to n and also satisfies $1 < e < \varphi(n)$ where

$$\varphi(n) = (p - 1)(q - 1). \tag{2.57}$$

d) Decrypting the message: Bob receives the encrypted message c over the communication channel. He is said to have decrypted the message back to m if he performs the computation

$$m = c^d (\text{mod } n), \tag{2.58}$$

where the values n and d are collectively known as the private key. The value

$d \in \mathbb{N}$ is generated by

$$d = \frac{1 \bmod \varphi(n)}{e}. \tag{2.59}$$

The decryption procedure can seen more clearly by considering the specific case that m is co-prime to n (although this can be generalized to the case when m is not co-prime to n). From (2.59), we have $ed = 1 + k\varphi(n)$ for some $k \in \mathbb{N}$. Using result (2.55), we find that $m^{k\varphi(n)} = 1 \,(\text{mod } n)$. Substituting this into the decryption procedure results in

$$\begin{aligned}
(c)^d &= (m^e)^d \,(\text{mod } n) && (2.60)\\
&= m^{ed} \,(\text{mod } n) && (2.61)\\
&= m^{1+k\varphi(n)} \,(\text{mod } n) && (2.62)\\
&= m \cdot m^{k\varphi(n)} \,(\text{mod } n) && (2.63)\\
&= m \,(\text{mod } n). && (2.64)
\end{aligned}$$

Using the symmetry property of modular arithmetic, this implies the desired result that $m = c^d (\text{mod } n)$.

e) Breaking encryption: The private key is kept in secret by the intended party. This is in contrast with the public key which is available to anyone. Despite this wide access, there is no increase in the security vulnerability as we shall describe below. The outside party that aims to eavesdrop to the transmission between Alice and Bob is commonly referred to as 'Eve'. If Eve has access to the private key, she can extract the message m from c. One way to obtain the private key would be if she could derive p and q by factoring $n = pq$. She would then be able to compute $\varphi(n) = (p-1)(q-1)$, and consequently obtain the private key (d, n).

However, the problem of prime factorization with classical computation is currently believed to require exponential resources (but this hypothesis is not formally proven). More accurately, the best known classical algorithm for this task is the NFS (Number Field Sieve) algorithm which has a performance of $\exp(\Theta\,(n^{1/3} \log^{2/3} n))$ operations for an n-bit integer. It is precisely the on-going computational difficulty of this problem that ensures durability of this information security system.

3

Quantum Information

> *"Is it, for example, information about some underlying reality, or about the effects of our intervention in it? Information universal to all observers, or personal to each? And can it be meaningful to speak of quantum information as something that flows, like liquid in a pipe, from place to place? No one knows (despite what they might tell you)."*
>
> – Philip Ball, *Quantum teleportation is even weirder than you think*

QUANTUM INFORMATION SCIENCE is the theoretical and experimental study of quantum information and its applications. The field is largely concerned with designing quantum systems to perform information tasks. This novel exploration has the following consequences that make the subject fundamental:

i) It reconceptualizes the probability amplitude of quantum theory as a quantity that can be harnessed for representing and transforming information; it is precisely this quantity that is termed 'quantum information.'
ii) Analogous to the study of classical information, a generalized framework is developed that abstracts away from the physical (quantum mechanical) systems that could be used to store the quantum information.
iii) It distils questions on the nature of quantum physics to distinctions between quantum information and classical information.

In this chapter, we look at three theoretical tools of quantum information science.

3.1 Review of Linear Algebra

Prior to examining the three main topics in this chapter, we provide a brief overview of linear algebra with an emphasis on the use of the Dirac notation.

3.1.1 Vector spaces

The vector space that is commonly used in quantum information science is $\mathbb{C}^n$. An element of the space, namely a vector, can be denoted $|\psi\rangle$ (referred to as a ket), where ψ is simply a label for the vector. The vector can have a column matrix representation of its n-tuples of complex numbers. Vector addition in $\mathbb{C}^n$ proceeds as

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \equiv \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{pmatrix}. \tag{3.1}$$

Scalar multiplication is computed as

$$\alpha \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \equiv \begin{pmatrix} \alpha a_1 \\ \vdots \\ \alpha a_n \end{pmatrix}. \tag{3.2}$$

Note that it does not make a difference if a scalar stands on the left or the right of a ket, $\alpha |\psi\rangle = |\psi\rangle \alpha$. We exclude the use of the ket notation for the zero vector and rather denote it as 0. A vector subspace of a vector space is a subset of the vector space such that the subset is also a vector space.

3.1.2 Basic definitions

A spanning set for a vector space is a set of vectors $|v_1\rangle, \ldots, |v_n\rangle$ such that any vector $|v\rangle$ in the vector space can be written as $|v\rangle = \sum_i \alpha_i |v_i\rangle$. Another core concept is that a set of non-zero vectors $|v_1\rangle, \ldots, |v_n\rangle$ is said to be linearly de-

pendent if the equation

$$\alpha_1 |v_1\rangle + \alpha_2 |v_2\rangle + \cdots + \alpha_n |v_n\rangle = 0, \tag{3.3}$$

has a solution where $\alpha_i \neq 0$ for at least one value of i. A set of vectors is linearly independent if it is not linearly dependent. A set of vectors that spans the vector space and is linearly independent is called a basis for the vector space. The dimension of the vector space is the number of elements in a basis set. With the exception of chapter 6, this thesis is only concerned with finite dimensional vector spaces.

An example of a basis for $\mathbb{C}^2$ is the computational basis set

$$|0\rangle \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{3.4}$$

Another basis for the space is

$$|+\rangle \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad |-\rangle \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{3.5}$$

3.1.3 Operators and Matrices

Suppose V and W are vector spaces. A linear operator between V and W is defined to be any function $A : V \to W$ which is linear in inputs

$$A\left(\sum_i \alpha_i |v_i\rangle\right) = \sum_i \alpha_i A(|v_i\rangle). \tag{3.6}$$

We can write $A|v_i\rangle$ to denote $A(|v_i\rangle)$. We say a linear operator A is defined on a vector space V if $A : V \to V$. The identity operator I maps all vectors to their respective self, $I|v\rangle = |v\rangle$. The zero operator 0 maps any vector to the zero vector, $0|v\rangle = 0$. The composition of two operators, say A and B, on a vector is defined as $(AB)(|v\rangle) \equiv A(B|v\rangle)$.

Operator addition is commutative, $A + B = B + A$, and associative $A + (B + C) =$

$(A + B + C)$. However operator multiplication is not commutative $AB \neq BA$ but is associative $A(BC) = (AB)C = ABC$.

Operators have an equivalent matrix representation. A m by n complex matrix A with entries A_{ij} can be thought as a linear operator that maps vectors from $\mathbb{C}^n$ to $\mathbb{C}^m$ under matrix multiplication. Conversely to view operators as matrices, suppose V and W are vector spaces with operator $A : V \rightarrow W$. More crucially, let $|v_1\rangle, \ldots, |v_m\rangle$ be a basis for V, and let $|w_1\rangle, \ldots, |w_n\rangle$ be a basis for W. Then for every j between 1 and m, there exists complex coefficients A_{1j} through A_{nj} such that

$$A|v_j\rangle = \sum_i A_{ij}|w_i\rangle . \tag{3.7}$$

The complex numbers A_{ij} form the matrix representation of the operator A.

Of critical importance are the topics of eigenvectors and eigenvalues. An eigenvector of operator A is a non-zero vector $|v\rangle$ that satisfies the equation $A|v\rangle = \lambda|v\rangle$, where λ is a complex number known as the eigenvalue corresponding to $|v\rangle$. The solution to the characteristic equation $c(\lambda) = 0$, where $c(\lambda) \equiv \det|A - \lambda I|$, are the eigenvalues of operator A. The eigenspace corresponding to eigenvalue λ, is a vector subspace on which A acts, that contains all the eigenvectors which have λ as its eigenvalue. When the dimension of the eigenspace is greater than one, we say it is degenerate.

3.1.4 Types of products

One can go beyond the basic abstraction of a vector space with its scalar multiplication; we will discuss four types of products that occur between vectors.

a) Inner product: An inner product maps two vectors, say $|v\rangle$ and $|w\rangle$, to a complex number. We denote this complex number as $\langle v|w\rangle$. The notation $\langle v|$ is referred to as the dual vector (or a bra). A vector space with an inner product is

called an inner product space. An inner product satisfies properties:

$$\langle v| \left(\sum_i \alpha_i |w_i\rangle \right) = \sum_i \alpha_i \langle v|w_i\rangle , \tag{3.8}$$

$$\langle v|w\rangle = \langle w|v\rangle^* , \tag{3.9}$$

$$\langle v|v\rangle \geq 0. \tag{3.10}$$

One can define the following inner product for $\mathbb{C}^n$: For two vectors with respective column matrix entries $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$, an inner product is given by $\sum_i a_i^* b_i$. In the case of finite dimensional complex vector spaces, an inner product space is also referred to as a Hilbert space.

Using the inner product, one can develop several useful notions. Vectors $|v\rangle$ and $|w\rangle$ are said to be orthogonal if $\langle v|w\rangle = 0$. The norm of a vector $|v\rangle$ is defined as $|||v\rangle|| = \sqrt{\langle v|v\rangle}$. A unit vector has a norm of value one; any vector with this property is said to be normalized. Furthermore, for any non-zero vector $|v\rangle$, its normalized form is given by $|v\rangle / |||v\rangle||$. A set of vectors $|v_i\rangle$ with index i is said to be orthonormal if $\langle v_i|v_j\rangle = \delta_{ij}$. The Gram-Schmidt procedure transforms an arbitrary basis of a vector space with an inner product, to an orthonormal basis; suppose $|w_1\rangle, \ldots, |w_d\rangle$ is an arbitrary basis; then an orthonormal basis $|v_1\rangle, \ldots, |v_d\rangle$ is computed first by $|v_1\rangle \equiv |w_1\rangle / |||w_1\rangle||$, and then the rest inductively obtained through formula,

$$|v_{k+1}\rangle \equiv \frac{|w_{k+1}\rangle - \sum_{i=1}^{k} \langle v_i|w_{k+1}\rangle |v_i\rangle}{|||w_{k+1}\rangle - \sum_{i=1}^{k} \langle v_i|w_{k+1}\rangle |v_i\rangle||}. \tag{3.11}$$

An orthonormal basis has the advantage of simplifying various computations. Let $|i\rangle$ be an orthonormal basis, with the following vectors, $|w\rangle = \sum_i w_i |i\rangle$ and $|v\rangle = \sum_i v_i |i\rangle$. Then the inner product is given by

$$\langle v|w\rangle = \sum_{ij} v_i^* w_j \delta_{ij} = \sum_i v_i^* w_i = \begin{pmatrix} v_1^* \ldots v_n^* \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}. \tag{3.12}$$

The dual vector can be interpreted as a row vector whose elements are complex

conjugates of the components of the column vector form of $|v\rangle$.

b) Outer product: Suppose $|v\rangle$ and $|w\rangle$ are vectors from respective inner product spaces, V and W. Then the outer product $|w\rangle\langle v|$ is a linear operator from $V \to W$ which is defined by $(|w\rangle\langle v|)|u\rangle \equiv |w\rangle\langle v|u\rangle = \langle v|u\rangle|w\rangle$. This is a valid operation as long as we are dealing with 'legal' products. This property is also referred to as the associative axiom [26] as it is an extension of the associativity of operator multiplication. More generally,

$$\left(\sum_i \alpha_i |w_i\rangle\langle v_i|\right)|u\rangle = \sum_i \alpha_i |w_i\rangle\langle v_i|u\rangle . \tag{3.13}$$

An application of the outer product is the completeness relation: If $|i\rangle$ is an orthonormal basis, then the identity operator can be written as $I = \sum_i |i\rangle\langle i|$. Using this property, one can obtain an outer product representation of operator $A : V \to W$:

$$A = I_W A I_V \tag{3.14}$$

$$= \sum_{ij} |w_j\rangle\langle w_j| A |v_i\rangle\langle v_i| \tag{3.15}$$

$$= \sum_{ij} \langle w_j|A|v_i\rangle |w_j\rangle\langle v_i| . \tag{3.16}$$

The quantity $\langle w_j|A|v_i\rangle$ is the matrix element in the jth row and ith column; the matrix representation is with respect to basis $|v_i\rangle$ and $|w_j\rangle$. The completeness relation is also used to prove the Cauchy-Schwarz inequality which states that for any two vectors in a Hilbert space, $|v\rangle$ and $|w\rangle$, we have $|\langle v|w\rangle|^2 \leq \langle v|v\rangle\langle w|w\rangle$.

Suppose $|i\rangle$ is an orthonormal set of eigenvectors for operator A with corresponding eigenvalues λ_i. Then a diagonal representation (or an orthonormal decomposition) for A is given by $A = \sum_i \lambda_i |i\rangle\langle i|$. An operator that has a diagonal representation is said to be diagonalizable.

c) Tensor product: One can construct a larger vector space from two or more different vector spaces. The mathematical machinery for such a construction is named the tensor product. To be more precise, suppose V and W are Hilbert spaces with respective dimensions m and n. Then $V \otimes W$ is a vector space with

dimension mn. The elements of $V \otimes W$ are linear combinations of $|v\rangle \otimes |w\rangle$, which is a tensor product of elements $|v\rangle$ of V, and $|w\rangle$ of W. For the case that $|i\rangle$ and $|j\rangle$ are respective orthonormal bases for V and W, $|i\rangle \otimes |j\rangle$ forms a basis for $V \otimes W$. The tensor product has the following properties:

$$z(|v\rangle \otimes |w\rangle) = (z\,|v\rangle) \otimes |w\rangle = |v\rangle \otimes (z\,|w\rangle), \tag{3.17}$$

$$(|v_1\rangle + |v_2\rangle) \otimes |w\rangle = |v_1\rangle \otimes |w\rangle + |v_2\rangle \otimes |w\rangle\,, \tag{3.18}$$

$$|v\rangle \otimes (|w_1\rangle + |w_2\rangle) = |v\rangle \otimes |w_1\rangle + |v\rangle \otimes |w_2\rangle\,, \tag{3.19}$$

where z is an arbitrary scalar, and the rest are vectors from their respective vector spaces. One can extend the tensor product to operators; suppose $|v\rangle$ and $|w\rangle$ are vectors in V and W, and A and B are linear operators respectively on V and W; then one can define a linear operator $A \otimes B$ which acts on $V \otimes W$ as

$$(A \otimes B)(|v\rangle \otimes |w\rangle) \equiv A\,|v\rangle \otimes B\,|w\rangle\,. \tag{3.20}$$

More generally, one has

$$(A \otimes B)\left(\sum_i \alpha_i\,|v_i\rangle \otimes |w_i\rangle\right) \equiv \sum_i \alpha_i A\,|v_i\rangle \otimes B\,|w_i\rangle\,. \tag{3.21}$$

The inner product on $V \otimes W$ is defined as follows; suppose we have two vectors $\sum_i \alpha_i\,|v_i\rangle \otimes |w_i\rangle$ and $\sum_j \beta_j\,|v'_j\rangle \otimes |w'_j\rangle$, then the inner product is defined as

$$\sum_{ij} \alpha_i^* \beta_j\,\langle v_i|v'_j\rangle\,\langle w_i|w'_j\rangle\,. \tag{3.22}$$

The tensor product can also be computed in terms of matrices. If A is an m by n matrix, and B is an p by q matrix, then we have

$$A \otimes B \equiv \begin{pmatrix} A_{11}B & A_{12}B & A_{13}B & \dots & A_{1n}B \\ A_{21}B & A_{22}B & A_{23}B & \dots & A_{2n}B \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{m1}B & A_{m2}B & A_{m3} & \dots & A_{mn}B \end{pmatrix}. \tag{3.23}$$

For tensor product $|v\rangle \otimes |w\rangle$, one can use equivalent notations $|v\rangle\,|w\rangle$, or $|v, w\rangle$, or simply $|vw\rangle$. Additionally, one often writes $|\psi\rangle^{\otimes n}$ to signify that $|\psi\rangle$ is tensored with itself n times.

d) Illegal products: Certain products are nonsensical in the Dirac notation [26] and should be avoided. Unlike the tensor product, if vectors $|v\rangle$ and $|w\rangle$ belong to the same vector space, then the product $|v\rangle\,|w\rangle$ is illegal; a similar condition holds for the dual vectors. Furthermore, operators always stand on the left of a ket and to the right of a bra; hence, the products $|v\rangle\,B$ and $A\,\langle v|$ are illegal.

3.1.5 Common operations

a) Hermitian conjugate: If A is a linear operator on V, then the Hermitian conjugate (or adjoint) of A is denoted $A^\dagger$ and it satisfies

$$\langle v|A^\dagger|w\rangle = \langle w|A|v\rangle^*\,, \tag{3.24}$$

for all vectors $|v\rangle$, $|w\rangle$ in V. In terms of a matrix representation of operator A, the Hermitian conjugation can be defined as $A^\dagger \equiv (A^*)^T$ where $*$ represents complex conjugation and T represents the transpose operation. For the case of a scalar, the Hermitian conjugate reduces to the complex conjugate. For the case of a vector, we have $|v\rangle^\dagger \equiv \langle v|$. We list a number of further properties:

$$(AB)^\dagger = B^\dagger A^\dagger, \tag{3.25}$$

$$(A\,|v\rangle)^\dagger = \langle v|\,A^\dagger, \tag{3.26}$$

$$(|w\rangle\,\langle v|)^\dagger = |v\rangle\,\langle w|\,, \tag{3.27}$$

$$(A^\dagger)^\dagger = A, \tag{3.28}$$

$$\left(\sum_i \alpha_i A_i\right)^\dagger = \sum_i \alpha_i^* A_i^\dagger. \tag{3.29}$$

b) Function of an operator: Suppose we have a function $f : \mathbb{C} \to \mathbb{C}$. If linear operator A has a diagonal representation $A = \sum_i \lambda_i\,|i\rangle\,\langle i|$, then the corresponding

operator function is defined as

$$f(A) \equiv \sum_i f(\lambda_i) \left|i\right\rangle \left\langle i\right| . \tag{3.30}$$

c) Trace: The trace of a matrix is the sum of its diagonal elements. Furthermore, the trace of an operator is defined as the trace of any matrix representation of the operator. Hence, for the case of an operator A, we have

$$\mathrm{tr}(A) = \sum_i A_{ii}. \tag{3.31}$$

This operation has the following properties:

$$\mathrm{tr}(AB) = \mathrm{tr}(BA), \tag{3.32}$$

$$\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B), \tag{3.33}$$

$$\mathrm{tr}(\alpha A) = \alpha \mathrm{tr}(A), \tag{3.34}$$

$$\mathrm{tr}(A \left|v\right\rangle \left\langle v\right|) = \left\langle v|A|v\right\rangle . \tag{3.35}$$

d) Commutator: The commutator of two operators, A and B is defined as

$$[A, B] \equiv AB - BA. \tag{3.36}$$

The anti-commutator for the two operators is computed as $\{A, B\} \equiv AB + BA$. The important case of $[A, B] = 0$, is expressed by saying A commutes with B.

3.1.6 Types of Operators

Using the Hermitian conjugate, operators can be classified into certain classes.

a) Hermitian: A Hermitian (or self-adjoint) operator is an operator that is equal to its Hermitian conjugate, $A^\dagger = A$. One of the most useful theorems regarding Hermitian operators is,

Theorem 3.1. ***(Simultaneous diagonalization theorem)*** *Suppose A and B are two Hermitian operators. Then $[A, B] = 0$ if and only if there exists an orthonormal*

basis such that both A and B are diagonal with respect to that basis. (See e.g. [2].)

Hence, for simultaneous diagonalizable Hermitian operators, A and B, we express them as $A = \sum_i \alpha_i |i\rangle \langle i|$ and $B = \sum_i \beta_i |i\rangle \langle i|$ for some common orthonormal set of eigenvectors $|i\rangle$.

b) Projectors: A particular subset of Hermitian operators are known as projectors (or projection operators). Suppose we have vector space V, along with a vector subspace W that has orthonormal basis $|1\rangle, \ldots, |k\rangle$. Then a projector onto W is defined as

$$P \equiv \sum_{i=1}^{k} |i\rangle \langle i| . \tag{3.37}$$

A projector satisfies the property $P^2 = P$. Furthermore, all the eigenvalues of a projector are all either 0 or 1.

c) Positive: Another class of Hermitian operators are known as positive operators. An operator A is said to be a positive if for every vector $|v\rangle$ we have $\langle v|A|v\rangle \geq 0$. An even stricter case is that an operator is said to be positive definite if $\langle v|A|v\rangle > 0$ for every non-zero vector $|v\rangle$. An interesting property is that if A is any operator, then $A^\dagger A$ is positive.

d) Unitary: A operator U is unitary if $UU^\dagger = U^\dagger U = I$. Alternatively an operator is unitary if and only if each of its matrix representations are unitary matrices. Furthermore, all the eigenvalues of a unitary matrix take the form $e^{i\theta}$ for some real θ. Of importance is the result that any unitary operator U can be formulated as

$$U = \exp(iA) \tag{3.38}$$

for some Hermitian operator A. Aside from the algebraic properties, unitary operators are geometrically significant in that they preserve the inner product between vectors; as an example the inner product between $U|v\rangle$ and $U|w\rangle$ is computed as $\langle v|U^\dagger U|w\rangle = \langle v|I|w\rangle = \langle v|w\rangle$.

e) Normal: An operator A is said to be normal if $AA^\dagger = A^\dagger A$. Both Hermitian and unitary operators are normal. One of the most important results in linear algebra is the spectral decomposition theorem:

Theorem 3.2. ***(Spectral decomposition)*** *Any normal operator M on a vector space V is diagonal with respect to some orthonormal basis for V. Conversely, any diagonalizable operator is normal (See e.g. [2].)*

□

More explicitly, this can be expressed as

$$M = \sum_i \lambda_i |i\rangle \langle i| , \tag{3.39}$$

where λ_i are the eigenvalues of M with each $|i\rangle$ signifying the corresponding eigenvector. Furthermore, the set of eigenvectors form an orthonormal basis for the vector space. One can also derive the projectors $P_i = |i\rangle \langle i|$ which results in $M = \sum_i \lambda_i P_i$. The set of projectors in this 'spectral expansion' of M satisfy both $\sum_i P_i = I$ and $P_i P_j = \delta_{ij} P_i$.

3.2 Qubits

The postulates of quantum theory [26, 27] are most commonly framed through state vectors. An information-theoretic view of these mathematical objects results in the quantum circuit model for qubits. At a coarse level, this framework can be viewed as a quantum analogue of the classical circuit model described in Chapter 2. There are four concepts to the quantum circuit model; we provide a description of each concept, their difference to the classical counterpart, and their inception from the postulates of quantum theory. We conclude this section with noting implications that portray further distinctions between quantum information and classical information.

3.2.1 Single qubit

a) Description: A bit can be physically manifested by a classical two state system. A qubit is a quantum analogue of a bit. It corresponds to an abstraction, that relates to a classical bit, which can be physically instantiated by a two-level

quantum system. More precisely, a qubit is a unit vector in a two-dimensional Hilbert space which takes the general form,

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \tag{3.40}$$

where we have used the computational basis set (3.4), and where $\alpha, \beta \in \mathbb{C}$. It is these complex numbers that are referred to as *quantum information*. Given $\langle\psi|\psi\rangle = 1$, known as the normalization condition, it can easily be shown that

$$|\alpha|^2 + |\beta|^2 = 1. \tag{3.41}$$

Depending on the values of α and β, a qubit is in one of the orthogonal computational basis vectors ($|0\rangle$ or $|1\rangle$), or in some linear combination of those vectors (3.40) which is referred to as a superposition. For the former case, a qubit would then map to the notion of a classical bit. This alludes to the idea that orthogonal vectors can be thought of as the different states of classical information.

b) Difference to classical information: The classical information of a bit, namely 0 or 1, can directly correspond to some physical feature of the classical system such as the voltage value of an electrical circuit. This is in vast contrast to quantum information, such as α and β in (3.40), which does not have a direct correspondence with the physical properties of the quantum system. The fundamental mystery [28, 29] is: What do these complex numbers physically represent? We do not exactly know what quantum information is! Nevertheless, these values do carry direct experimental consequences. From a historical view, this problem is known as the issue of the interpretation of quantum mechanics.

c) Quantum-theoretic origin: The relationship between a two-level quantum system and a qubit (3.40) stems from a postulate of quantum theory which states that: Associated to any isolated quantum system is a Hilbert space known as the state space of the system; the system is completely described by its state vector (also known as the quantum state), which is a unit vector in the system's state space. In regards to terminology, if a state vector is represented as $\sum_i \alpha_i\,|v_i\rangle$ where it is a linear combination of basis states $|v_i\rangle$, then the complex coefficients α_i are referred to as its probability amplitudes. The central tenet of quantum

theory is that to describe the state of a system, one needs to assign one amplitude for each possible configuration that you would the find the system in upon measuring it. For the case of a two-level quantum system its state vector (or its quantum state) is adapted as a qubit, and its amplitudes are referred to as its quantum information. The power of the quantum circuit framework can be seen in that a qubit can be physically instantiated by a diverse range of two-level quantum systems [2, 30]. A few examples include the spin of a spin-1/2 particle, the polarization of a photon or the energy levels of a two-state atom.

3.2.2 Multiple qubits

a) Description: A 'string' of qubits is connected by a tensor product structure. As an example, a two qubit system can be in one of the four computational basis vectors $|0\rangle \otimes |0\rangle$, $|0\rangle \otimes |1\rangle$, $|1\rangle \otimes |0\rangle$, $|1\rangle \otimes |1\rangle$, or in some linear combination of these vectors

$$|\psi\rangle = \alpha_{00} |00\rangle + \alpha_{01} |01\rangle + \alpha_{10} |10\rangle + \alpha_{11} |11\rangle . \tag{3.42}$$

The vector $|\psi\rangle$ satisfies the normalization condition $\sum_{x\in\{0,1\}^2} |\alpha_x|^2 = 1$, where $\{0,1\}^2$ refers to 'the set of strings of length two with each letter being either 0 or 1.' More generally for a system of n qubits, the associated state vector $|\psi\rangle$ is referred to as its quantum state, with the computational basis states of the form $|x_1 x_2 \dots x_n\rangle$ with $x_i \in \{0,1\}$. The number of complex coefficients involved is 2^n and it is these coefficients that are the quantum information.

b) Difference to classical information: The superposition property of a qubit provides the key distinction from a classical bit. Moreover, it has a remarkable consequence for multiple qubits; for a relatively small number of qubits such as $n = 500$, the superposition property gives 2^n values of quantum information. These are more complex numbers than can be stored on any classical computer that could ever feasibly be built. Fortunately, this exponential relationship between the number of qubits and the amount of quantum information makes quantum systems a compelling platform to design information technologies on.

c) Quantum-theoretic origin: The idea that the tensor product is the appropri-

ate mathematical machinery for multiple qubits comes from a postulate of quantum theory concerning composite systems. It assumes that the state space of a composite quantum system is the tensor product of the states spaces of the component quantum systems. Furthermore, if we have systems numbered 1 through to n, and system number i is in state $|\psi_i\rangle$, then the joint state vector of the total system is given by $|\psi_1\rangle \otimes |\psi_2\rangle \otimes \ldots |\psi_n\rangle$.

3.2.3 Transforming qubits

a) Description: Information as an abstraction is useful when it can be transformed. Classical gates transform the classical information through the algebra of boolean logic. The quantum circuit model introduces the concept of a quantum gate as a means to transform quantum information. The only constraint on the notion of a quantum gate is that it be a unitary operator, $U^\dagger U = U^\dagger U = I$. These gates are applied to qubits as operators acting on vectors. The most important single qubit gates are the Pauli operators. With respect to basis set (3.4), they represented as

$$\sigma_x \equiv \sigma_1 \equiv X \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \quad \sigma_y \equiv \sigma_2 \equiv Y \equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}; \quad \sigma_z \equiv \sigma_3 \equiv Z \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.43}$$

It is also standard to include the identity operator as part of this set which we label as σ_0. In terms of outer products, the Pauli operators are expressed as

$$\sigma_0 = |0\rangle\langle 0| + |1\rangle\langle 1|, \tag{3.44}$$

$$\sigma_x = |0\rangle\langle 1| + |1\rangle\langle 0|, \tag{3.45}$$

$$\sigma_y = -i\,|0\rangle\langle 1| + i\,|1\rangle\langle 0|, \tag{3.46}$$

$$\sigma_z = |0\rangle\langle 0| - |1\rangle\langle 1|. \tag{3.47}$$

The commutators between the different Pauli operators equate to

$$[X, Y] = 2iZ; \quad [Y, Z] = 2iX; \quad [Z, X] = 2iY. \tag{3.48}$$

The Hadamard gate, phase gate, and $\pi/8$ gate (denoted T) are respectively

$$H \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}; \quad S \equiv \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}; \quad T \equiv \begin{pmatrix} 1 & 0 \\ 0 & e^{(i\pi/4)} \end{pmatrix}. \tag{3.49}$$

The Hadamard gate turns the computational basis states into particular superposition states as follows

$$H|0\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} = |+\rangle, \quad H|1\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}} = |-\rangle. \tag{3.50}$$

The states $|+\rangle$ and $|-\rangle$ have column vector representation (3.5). The quantum gates mentioned so far satisfy the following well known identities

$$XYX = -Y, \tag{3.51}$$

$$HXH = Z, \tag{3.52}$$

$$HYH = -Y, \tag{3.53}$$

$$HZH = X, \tag{3.54}$$

$$H = \frac{(X+Z)}{\sqrt{2}}, \tag{3.55}$$

$$S = T^2. \tag{3.56}$$

Another important set of quantum gates, which are derived from the Pauli operators, are known as the rotation operators:

$$R_x(\theta) \equiv e^{-i\theta X/2} = \cos\frac{\theta}{2} I - i \sin\frac{\theta}{2} X = \begin{pmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}, \tag{3.57}$$

$$R_y(\theta) \equiv e^{-i\theta Y/2} = \cos\frac{\theta}{2} I - i \sin\frac{\theta}{2} Y = \begin{pmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}, \tag{3.58}$$

$$R_z(\theta) \equiv e^{-i\theta Z/2} = \cos\frac{\theta}{2} I - i \sin\frac{\theta}{2} Z = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix}. \tag{3.59}$$

The significance of these rotation operators is that we can express an arbitrary single qubit quantum gate U as

$$U = e^{i\alpha} R_z(\beta)\, R_y(\gamma)\, R_z(\delta), \tag{3.60}$$

for some real numbers α, β, γ, and δ.

For the case of two qubits, an important quantum gate is the controlled-NOT operator. This unitary operator has the matrix representation

$$U_{CN} \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{3.61}$$

The action of the operator on a quantum state $|a, b\rangle$ is to transform it into $|a, b \oplus a\rangle$ where $\oplus$ denotes addition modulo two. A more explicit description is this gate acts on two registers where the first qubit is known as the control qubit and the second as the target qubit; if the control qubit is in state $|0\rangle$, then the target qubit is left unchanged; however if the control qubit is in state $|1\rangle$, then an X (NOT) operator is applied to the target qubit. One can generalize the essence of the controlled-NOT operator to any other gate in that the X operator is replaced by the appropriate gate.

The importance of the controlled-NOT operator can be stated by the result that any multiple qubit quantum gate may be composed from controlled-NOT gates and single qubit gates.

b) Difference to classical information: The mathematical difference between boolean functions and unitary operators is clearly self-evident. However the non-trivial differences between classical and quantum gates are subtle. Some classical gates such as the NAND gate or the XOR gate are non-invertible; it is not possible to derive the input given the output. In contrast, all quantum gates are invertible as the inverse of a unitary matrix is also a unitary matrix, hence a valid quantum gate. In the classical case, the only non-trivial single bit gate is the NOT gate; in the quantum model, we have several important single qubit

gates. It is interesting to note that there are some subtle similarities. The Pauli X operator can be thought of as a quantum analogue of classical NOT gate since it inverts the computational basis states

$$X|0\rangle = |1\rangle\,, \quad X|1\rangle = |0\rangle\,. \tag{3.62}$$

c) Quantum-theoretic origin: In the quantum circuit model, we have seen the use of unitary operators as a means to transform qubits. It turns out that this is directly connected to a postulate of quantum theory regarding dynamics. Namely that the continuous time evolution of a state vector of a closed quantum system is governed by the Schrödinger equation

$$i\hbar\frac{\partial|\psi(t)\rangle}{\partial t} = H|\psi(t)\rangle\,. \tag{3.63}$$

The Hamiltonian H is a Hermitian operator which specifies the physics of the system. The solution to (3.63) is

$$|\psi(t_2)\rangle = \exp\left[\frac{-iH(t_2-t_1)}{\hbar}\right]|\psi(t_1)\rangle\,. \tag{3.64}$$

Using relationship (3.38), one can naturally define a unitary operator

$$U(t_1,t_2) \equiv \exp\left[\frac{-iH(t_2-t_1)}{\hbar}\right]. \tag{3.65}$$

Hence a discrete time transformation of states is provided by unitary operators.

3.2.4 Measuring qubits

a) Description: The final element of the quantum circuit model is measuring the qubits to extract their values. One way to mathematically represent the measurement of qubits is using any orthonormal bases. We have seen a qubit (3.40) represented using the computational basis states (3.4). More generally, suppose

a qubit is represented using an arbitrary orthonormal basis, $|x\rangle$ and $|y\rangle$

$$|\psi\rangle = \alpha_x |x\rangle + \beta_y |y\rangle , \tag{3.66}$$

where $\alpha_x, \beta_y \in \mathbb{C}$. Then by measuring the qubit, with respect to the $|x\rangle$, $|y\rangle$ basis, we find the qubit is in state $|x\rangle$ or $|y\rangle$; we never find it in the superposition state (3.66); hence measurement is said to instantaneously 'collapse' the state into one of the basis states. Furthermore, the probability of finding the qubit in state $|x\rangle$ is given by the modulus square of its coefficient, $|\alpha_x|^2$; similarly the probability of finding it in state $|y\rangle$ is given by $|\beta_y|^2$. Due to the normalization condition, these 'quantum' probabilities (that are derived from quantum information) sum to one. As an example, if the qubit is in state $|+\rangle = (1/\sqrt{2})|0\rangle + (1/\sqrt{2})|1\rangle$, then the probability of finding it in state $|0\rangle$ upon measurement is $1/2$, and the probability of finding it in state $|1\rangle$ is also $1/2$. One can generalize this technique to multiple qubits by using an arbitrary orthonormal basis of the respective Hilbert space.

b) Difference to classical information: Unlike classical information, quantum information such as α_x, β_y in (3.66) can be described as 'hidden.' No single measurement allows us to directly extract those values. On a related matter, measurement can generally be viewed as a process that converts quantum information into classical information in the form of an orthogonal basis state. This presents us with another fundamental mystery: Why does this 'collapse' occur? Or perhaps can we derive 'measurement' from a unitary process. Like with the first mystery, a deep answer still unknown. This issue from a quantum-theoretic perspective is referred to as the measurement problem [31].

c) Quantum-theoretic origin: In quantum theory, an alternative way of evolving a state forward in time is through the measurement of quantum states (as opposed to the Schrödinger equation). We describe the relevant postulate: Quantum measurements are denoted by a set of measurement operators $\{M_m\}$ which satisfy

$$\sum_m M_m^\dagger M_m = I. \tag{3.67}$$

This is known as the completeness relation. These operators act on the relevant state space. The index m refers to the outcome obtained from the measurement;

if the quantum system is in state $|\psi\rangle$, then probability that result m occurs upon measurement is given by

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle\,. \tag{3.68}$$

These quantum probabilities sum to one

$$\sum_m p(m) = \sum_m \langle\psi|M_m^\dagger M_m|\psi\rangle = 1 \tag{3.69}$$

Furthermore, given result m, the post-measurement state can be written as

$$\frac{M_m\,|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}. \tag{3.70}$$

The quantity $e^{i\theta}$ known as a global phase factor, where $\theta \in \mathbb{R}$, is irrelevant with respect to measurement. The state $|\psi\rangle$ and $e^{i\theta}\,|\psi\rangle$ are equivalent from the perspective of observation since

$$\langle\psi|e^{-i\theta}M_m^\dagger M_m e^{i\theta}|\psi\rangle = \langle\psi|M_m^\dagger M_m|\psi\rangle\,. \tag{3.71}$$

A special case of these measurement operators are projective measurements. Projective measurements satisfy (3.67) as well as carry the property that M_m are Hermitian and that $M_m M_{m'} = \delta_{m,m'} M_m$. A projective measurement corresponds to an observable (a physical quantity that can be measured). An observable is represented as a Hermitian operator M on the state space. Using spectral decomposition (3.39), one can state this more precisely as

$$M = \sum_m m\,P_m, \tag{3.72}$$

where P_m represents the orthogonal projector onto the eigenspace of M with eigenvalue m. The measurement outcomes are the eigenvalues of M. Furthermore, the probability of obtaining result m if system was in state $|\psi\rangle$ before measurement, is given by

$$p(m) = \langle\psi|P_m|\psi\rangle\,. \tag{3.73}$$

This is often referred to as the Born rule. The post-measurement state after obtaining outcome m becomes

$$\frac{P_m |\psi\rangle}{\sqrt{\langle\psi|P_m|\psi\rangle}}. \tag{3.74}$$

It is convention to sometimes not emphasize the observable, but rather focus on the orthogonal projectors in (3.72) or its associated kets. To be more precise, the phrase 'measure in basis $|m\rangle$' means to measure any observable that has $|m\rangle$ as its eigenbasis. The corresponding projectors are $P_m = |m\rangle\langle m|$ which satisfy $\sum_m P_m = I$ and $P_m P_{m'} = \delta_{m,m'} P_m$. The Born rule can be re-written as

$$p(m) = |\langle m|\psi\rangle|^2. \tag{3.75}$$

Hence it can be seen that every orthonormal basis of the Hilbert space corresponds to a quantum measurement, and has outcome probabilities given by the Born rule. In the quantum circuit model, we use projective measurements, and therefore equations (3.73) and (3.74) are implicit in our explanation regarding the measurement of qubits. Furthermore, it is often the case that we measure in the computational basis states (3.4).

3.2.5 Further distinctions from classical information

From our treatment on the qubits, a number of implications arise. These results further emphasize the non-trivial distinctions between quantum information and classical information.

a) No-cloning: An essential task of classical information systems is to copy bits. For an unknown quantum state, this operation is impossible to carry out. The no-cloning theorem [32, 33, 34] states that there exists no unitary operator that can clone an unknown quantum state. To see this as true, suppose such a unitary operator U did exist, and we have two unknown quantum states, $|\psi\rangle$ and $|\phi\rangle$. Furthermore, let $|s\rangle$ denote a blank state to copy in. We have mappings of

the form

$$U(|\psi\rangle \otimes |s\rangle) = |\psi\rangle \otimes |\psi\rangle\,, \tag{3.76}$$

$$U(|\phi\rangle \otimes |s\rangle) = |\phi\rangle \otimes |\phi\rangle\,. \tag{3.77}$$

We then compute the inner product of the left hand side of both equations. We carry the same task on the right hand side. By equating the quantities, we obtain

$$\langle\psi|\phi\rangle = (\langle\psi|\phi\rangle)^2. \tag{3.78}$$

This implies that either the two unknown states are orthogonal ($\langle\psi|\phi\rangle = 0$), or they are equal to each other ($|\psi\rangle = |\phi\rangle$). Hence a general cloning machine is impossible. However, it is not surprising that a set of orthogonal states can be copied since these can be viewed as different states of classical information. The no-cloning theorem is a basic result in quantum information science, and is related to other fundamental constraints in physics such as in regards to closed timelike curves [35, 36]. Having introduced the no-cloning theorem, it seems appropriate to say that other related no-go theorems exist including a no-deletion theorem [37]. Both no-cloning and no-deletion collectively allude to the conservation of quantum information.

b) Indistinguishability: In principle, one can always distinguish classical bits from each other. This is an impossible task for non-orthogonal quantum states. As an example, there is no single measurement process that can reliably distinguish states $|0\rangle$ or $|+\rangle$. We provide a rough argument. Consider the simpler case of a fixed set of orthogonal quantum states denoted $|\psi_i\rangle$. One can define general measurement operators consisting of $M_i \equiv |\psi_i\rangle\langle\psi_i|$ as well as the positive square root of $I - \sum_{i\neq 0} |\psi_i\rangle\langle\psi_i|$. In this case, if state $|\psi_i\rangle$ is prepared, then $p(i) = \langle\psi_i|M_i|\psi_i\rangle = 1$. This implies that this set of states can be reliably distinguished. If on the other hand $|\psi_i\rangle$ denotes a set of non-orthogonal states, then a crucial property is that say state $|\psi_2\rangle$ can be broken into a component parallel to say state $|\psi_1\rangle$ as well as a component orthogonal to $|\psi_1\rangle$. Due to the component of $|\psi_2\rangle$ that is parallel to $|\psi_1\rangle$, there is a non-zero probability of mistaking $|\psi_1\rangle$ as the state when in fact it was $|\psi_2\rangle$ that was prepared. Thus these non-orthogonal

states cannot be reliably distinguished. This means a measurement has a limit on its ability to exact information which conveys additional support to the notion that quantum information is hidden. Surprisingly, quantum indistinguishability is also related to the constraints regarding closed timelike curves [38].

c) Uncertainty: Unlike classical bits, qubits have a probabilitic property that is intrinsic to them (3.73). One can view these quantum probabilities in terms of an expectation value (2.1). More precisely, the expectation value of an observable, M, (with respect to quantum state $|\psi\rangle$) in (3.72), is defined as

$$\begin{aligned}\langle M\rangle &\equiv \sum_m m\, p(m) && (3.79)\\ &= \sum_m m\, \langle\psi|P_m|\psi\rangle && (3.80)\\ &= \langle\psi|\left(\sum_m m\, P_m\right)|\psi\rangle && (3.81)\\ &= \langle\psi|M|\psi\rangle\,. && (3.82)\end{aligned}$$

In the derivation we have used quantum probabilities (3.73). We can invoke further classical probabilistic concepts, and introduce the standard deviation (2.3) of an observable,

$$\Delta(M) = \sqrt{\langle M^2\rangle - \langle M\rangle^2}. \tag{3.83}$$

This quantity is also known as the uncertainty of the observable, and it represents a statistical measure of the spread of measurements about the expectation value. The Heisenberg uncertainty principle states that for observables A and B we have

$$\Delta(A)\Delta(B) \geq \frac{|\langle\psi|\,[A,B]\,|\psi\rangle|}{2}. \tag{3.84}$$

This result has a shocking implication; suppose we had two observables that do not commute and we performed a large number of these measurements on systems which are in identical states $|\psi\rangle$; then if we make the uncertainty on the results of B decrease, then the uncertainty of the results of A must increase, regardless of the sophistication of the measurement. The uncertainty principle is also related to constraints regarding closed timelike curves [39].

3.3 Density Operators

Density operators are a widely used mathematical tool in the study of open quantum systems and quantum statistical mechanics [40]. Within quantum information science, it provides a natural framework for quantifying the information concerning subsystems. In this section, we briefly describe quantum information in the language of density operators, while relaying its relationship to the quantum circuit model.

3.3.1 Single density operator

Associated to any isolated quantum system is a Hilbert space with an operator known as the density operator, ρ, which acts on the space. The density operator is a positive operator with $\mathrm{tr}(\rho) = 1$. The relationship to the quantum circuit model is as follow: If a system is known to be in state $|\psi_i\rangle$ with associated classical probability p_i then the density operator of the system is given by

$$\rho = \sum_i p_i \rho_i, \tag{3.85}$$

where $\rho_i \equiv |\psi_i\rangle \langle\psi_i|$. The set $\{|\psi_i\rangle, p_i\}$ is referred to as an ensemble. For the limited case where $|\psi\rangle$ is the only member of an ensemble (like in the circuit model), we have $\rho = |\psi\rangle \langle\psi|$ which is then called a pure state. Otherwise it is known as a mixed state, which means we do not know with certainty what quantum state it is in. In terms of computations, a pure state satisfies

$$\mathrm{tr}(\rho^2) = 1, \tag{3.86}$$

whereas a mixed state results in

$$\mathrm{tr}(\rho^2) < 1. \tag{3.87}$$

The largest statistical ignorance is expressed by the maximally mixed state

$$\rho = (1/n)\, I_n, \tag{3.88}$$

where n is the dimension of the Hilbert space. The density operator is a broader framework than the quantum circuit model. It captures both the quantum information with its quantum probabilities, as well as the classical probabilities related to our ignorance of that quantum information.

3.3.2 Multiple density operators

a) Composite system: In the quantum circuit model, we have employed the tensor product as a means to describe multiple qubits. Similarly, in the density operator language, a composite system is represented as a tensor product of the Hilbert spaces of the component systems. If system number i is in state ρ_i, then the composite system is described by density operator

$$\rho_1 \otimes \rho_2 \otimes \ldots \rho_n. \tag{3.89}$$

b) Subsystem: A central role of the density operator in quantum information science is as an information tool to describe subsystems. This particular task is carried out by the reduced density operator. Suppose ρ^{AB} describes a composite system made up of system A and system B. Then the reduced density operator for A is defined as

$$\rho^A \equiv \mathrm{tr}_B(\rho^{AB}), \tag{3.90}$$

where tr_B is known as the partial trace over system B. The partial trace is defined as

$$\mathrm{tr}_B(|a_1\rangle \langle a_2| \otimes |b_1\rangle \langle b_2|) \equiv |a_1\rangle \langle a_2| \, \mathrm{tr}(|b_1\rangle \langle b_2|) \tag{3.91}$$

where we have used the usual trace operation on the right hand side; the vectors $|a_1\rangle$ and $|a_2\rangle$ are any vectors from the Hilbert space associated to system A; similarly $|b_1\rangle$ and $|b_2\rangle$ are any vectors from the Hilbert space associated to system B. As an example, one can apply this operation to the trivial composite system $\rho^{AB} = \rho_1 \otimes \rho_2$, and obtain as expected

$$\rho^A = \mathrm{tr}_B(\rho_1 \otimes \rho_2) = \rho_1 \mathrm{tr}(\rho_2) = \rho_1. \tag{3.92}$$

3.3.3 Transforming density operators

a) Closed system: Like in the quantum circuit model, density operators associated with closed systems also transform according to unitary operators. This is computed as

$$\rho' = U\rho U^\dagger, \tag{3.93}$$

where U is a unitary operator.

b) Open system: For open quantum systems, a generalized framework for dynamics known as quantum operations is employed. The density operators transform as

$$\rho' = \Phi(\rho) = \sum_k A_k \rho A_k^\dagger, \tag{3.94}$$

where Φ is known as a quantum operation. The operators A_k are known as operation elements or as the Krauss operators. These are not necessarily unitary, but rather satisfy the condition

$$\sum_k A_k^\dagger A_k \leq I. \tag{3.95}$$

The mapping (3.93) can be regarded as a quantum operation where $\Phi(\rho) = U\rho U^\dagger$. But the utility of the framework is best captured when considering open systems such in the case of environmental noise on a qubit. Suppose a qubit flips from $|0\rangle$ to $|1\rangle$ (or vice versa) with classical probability $1-p$. The associated operation elements are $A_0 = \sqrt{p}I$, and $A_1 = \sqrt{1-p}X$. The quantum operation, known as the bit flip channel, is written as

$$\Phi(\rho) = p\rho + (1-p)X\rho X. \tag{3.96}$$

More generally, quantum operations have been used to quantify a broad range of noise-related phenomenon on qubits.

3.3.4 Measuring density operators

a) General measurement: The measurement postulates of quantum theory can be reformulated for density operators. Quantum measurements are described by a collection of measurement operators $\{M_m\}$ which satisfy the completeness relation (3.67). If the quantum system is in state ρ before measurement, then the probability of obtaining result m is given by

$$p(m) = \mathrm{tr}(M_m^\dagger M_m \rho). \tag{3.97}$$

The post-measurement operator is expressed as

$$\frac{M_m \rho M_m^\dagger}{\mathrm{tr}(M_m^\dagger M_m \rho)}. \tag{3.98}$$

b) POVM: POVM stands for positive operator valued measure and is a formalism that is usually expressed with the language of density operators. Given general measurement operators $\{M_m\}$, one can define

$$E_m \equiv M_m^\dagger M_m. \tag{3.99}$$

These positive operators, E_m, are known as the POVM elements. It can be shown that

$$\sum_m E_m = I. \tag{3.100}$$

If the density operator prior to measurement is denoted ρ, then probability of obtaining outcome m is given by

$$p(m) = \mathrm{tr}(E_m \rho). \tag{3.101}$$

One example of a POVM are projection measurements which are described by projectors, P_m, and satisfy $E_m = P_m^\dagger P_m = P_m$. For this specific case, (3.101), equates to

$$p(m) = \mathrm{tr}(P_m \rho). \tag{3.102}$$

which is just a reformulation of the Born rule (3.73).

c) Tomography: We have seen that a single measurement on a qubit does not allow us to obtain the quantum information. This means in general it is impossible to characterize an unknown state ρ if we are given a single copy. Quantum state tomography is a procedure to estimate the unknown quantum state with many measurements. Suppose we have many copies of the density operator ρ of an unknown qubit. Using the Pauli operators (3.43), one can express ρ as

$$\rho = \frac{\mathrm{tr}(\rho)I + \mathrm{tr}(X\rho)X + \mathrm{tr}(Y\rho)Y + \mathrm{tr}(Z\rho)Z}{2}. \tag{3.103}$$

For large sample sizes, one can obtain a reasonable estimation of the values of $\mathrm{tr}(X\rho)$, $\mathrm{tr}(Y\rho)$ and $\mathrm{tr}(Z\rho)$ and identify the quantum information of the qubit. Generalizing this procedure to n qubits results in the expression,

$$\rho = \sum_{\vec{v}} \frac{\mathrm{tr}(\sigma_{v_1} \otimes \sigma_{v_2} \otimes \cdots \otimes \sigma_{v_n}\rho)\sigma_{v_1} \otimes \sigma_{v_2} \otimes \cdots \otimes \sigma_{v_n}}{2^n}, \tag{3.104}$$

where $\vec{v} = (v_1, \ldots, v_n)$ with entries v_i chosen from the set 0, 1, 2, 3

3.3.5 Further properties

a) No-broadcasting: The no-broadcast theorem [41] generalizes the no-cloning theorem to the case of mixed states. It states that given state ρ_1, it is not possible to create a composite system ρ^{AB} such that $\mathrm{tr}_A(\rho^{AB}) = \rho_1$ and $\mathrm{tr}_B(\rho^{AB}) = \rho_1$.

b) Antidistinguishability: In the previous section, we looked at the general case of distinguishing non-orthogonal quantum states. In terms of density operators, distinguishability can be stated as the existence of a POVM E_j for set of states ρ_k such that

$$\mathrm{tr}(E_j\rho_k) = \delta_{ij}, \tag{3.105}$$

for all j and k. A related property is the notion of antidistinguishability [42, 43]. A set of states ρ_k is antidistinguishable if there exists a POVM E_j such that for each j,

$$\mathrm{tr}(E_j\rho_k) = 0. \tag{3.106}$$

Distinguishability lets us know that a particular state was definitely prepared.

This in contrast to antidistinguishability which lets us know that a particular state was definitely not prepared.

c) Distance measures: To quantitatively capture the idea of how 'close' two quantum states are, there are two useful tools that we proceed to describe. The first is the trace distance between two density operators ρ and σ, which is defined as

$$D(\rho, \sigma) \equiv \frac{1}{2}\text{tr}|\rho - \sigma|, \tag{3.107}$$

where

$$|A| \equiv \sqrt{A^\dagger A}. \tag{3.108}$$

The second method is known as the fidelity which is given by

$$F(\rho, \sigma) \equiv \text{tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}} \tag{3.109}$$

for density operators ρ and σ. The fidelity is invariant under unitary transformations

$$F(U\rho U^\dagger, U\sigma U^\dagger) = F(\rho, \sigma). \tag{3.110}$$

If both density operators represent pure states, $\rho = |\psi_\rho\rangle \langle\psi_\rho|$ and $\sigma = |\psi_\sigma\rangle \langle\psi_\sigma|$, the fidelity reduces to

$$F(\rho, \sigma) = |\langle\psi_\rho|\psi_\sigma\rangle|^2 \tag{3.111}$$

The quantity (3.111) measures the probability of confusing the two states if one is only able to carry out only one measurement on one system which is prepared in one of the two states. If the two states are orthogonal, then the fidelity is computed to be zero and the states can be fully distinguished.

3.4 Entropy

An alternative approach to view quantum information science is based on entropy. In chapter 2, we introduced the Shannon entropy of a random variable as a means of describing classical information. In this section, we define the von Neumann entropy of a quantum density operator. A limited perspective is

that the Shannon entropy applies only in the classical realm, whereas the von Neumann entropy strictly conveys quantum properties.

Rather in the modern setting of quantum information, we'll see that the Shannon entropy can employed with respect to classical probabilities (in a mixed state) as well quantum probabilities (derived from quantum information). Moreover, the von Neumann entropy can represent classical ignorance (in the case of a mixed state) as well as signify a reliable storage of quantum information (through the quantum analogue of data compression). All of the material in this section reformulates or builds on concepts seen in the previous sections.

3.4.1 Indistinguishability using Shannon entropy

Although we have treated quantum indistinguishability in the qubit and the density operator frameworks, a description through entropy is most insightful [44].

a) Scenario: Suppose a state is prepared from an ensemble ξ of density operators $\{\rho_j : j = 1, 2, \ldots, N\}$ with a prior classical probability distribution $\{\eta_j : j = 1, 2, \ldots, N\}$. Hence, the resulting operator can be written as

$$\rho = \sum_{j=1}^{N} \eta_j \rho_j, \tag{3.112}$$

with $\mathrm{tr}(\rho) = 1$. The task of distinguishability is to identify which state was prepared through a single measurement. We perform this measurement using POVM elements which we denote by $\Pi \equiv \{\Pi_k : k = 1, 2, \ldots, M\}$, where $M \geq N$.

b) Probabilistic quantities: To develop an entropic model of this task, we proceed to derive several quantities. The joint probability that the state ρ_j is prepared and that the outcome obtained is Π_k, is given by

$$P(\rho_j, \Pi_k) = \eta_j \,\mathrm{tr}(\Pi_k \rho_j), \tag{3.113}$$

where we have used (3.101). The total probability of obtaining outcome Π_k is

computed as

$$P_{\Pi_k} = \sum_{j=1}^{N} \eta_j \operatorname{tr}(\Pi_k \rho_j) = \operatorname{tr}(\Pi_k \rho). \tag{3.114}$$

Summing over k in (3.113) results in

$$\sum_{k=1}^{M} P(\rho_j, \Pi_k) = \eta_j. \tag{3.115}$$

b) Entropic quantities: The quantities η_j signify a probability distribution in (3.112). Therefore, we can evaluate the Shannon entropy (2.20) of this distribution,

$$H(\xi) = -\sum_{j=1}^{N} \eta_j \log_2 \eta_j. \tag{3.116}$$

Recall the mutual information (2.23) and its property $H(X:Y) = H(X) - H(X|Y)$. Using (3.116), we have the following mutual information associated to the measurement process

$$H(\xi : \Pi) = H(\xi) - \sum_{k=1}^{M} P_{\Pi_k} H(\xi|\Pi_k). \tag{3.117}$$

It quantifies how much information is gained about inferring the state that was prepared through the measurement. Moreover, the quantity $H(\xi|\Pi_k)$ signifies the conditional entropy (2.22) of the remaining ignorance after outcome Π_k is obtained. Therefore, a reasonable goal for this task is to choose a measurement that maximizes $H(\xi : \Pi)$.

c) Accessible information: Of crucial importance is the accessible information which is defined as the maximum mutual information attainable over all possible POVM measurements,

$$I_{\text{acc}} = H(\xi) - \min_{\text{all}\,\Pi} \sum_{k=1}^{M} P_{\Pi_k} H(\xi|\Pi_k). \tag{3.118}$$

The accessible information is a marker of how well a measurement can do at identifying the state prepared. Moreover, it has an upper bound known as the Holevo bound [2]. From this point of view, the accessible information quantita-

tively captures the notion that quantum information has a hidden nature.

d) Subsequent measurements: Suppose measurement outcome Π_k is obtained. After our first measurement, there may be subsequent measurements performed to extract further accessible information. To compute the relevant entropic quantity, recall (3.99); each of the POVM elements corresponds to a general measurement operator M_k, where $\Pi_k = M_k^\dagger M_k$. Then with respect to state ρ_j, the normalized postmeasurement states (3.98) are written as

$$\rho_j^{(k)} = \frac{M_k \rho_j M_k^\dagger}{\mathrm{tr}(\Pi_k \rho_j)}. \tag{3.119}$$

The respective new probabilities (3.97) are found as

$$\eta_j^{(k)} = \frac{\eta_j \mathrm{tr}(\Pi_k \rho_j)}{\mathrm{tr}(\Pi_k \rho)}. \tag{3.120}$$

Moreover, we let $\xi^{(k)}$ denote the postmeasurement ensemble consisting of states (3.119) with respective probabilities (3.120). The Shannon entropy of $\xi^{(k)}$ using (3.120) is equal to value of $H(\xi|\Pi_k)$ in (3.117).

If one performs a optimal subsequent POVM on $\xi^{(k)}$, this reduces the remaining ignorance for distinguishability in ξ to $H(\xi^{(k)}) - I_{\mathrm{acc}}(\xi^k)$. Hence the maximum mutual information between the original ensemble ξ and the outcomes of optimal subsequent measurements is given by

$$I'_{\max}(\xi, \Pi) = H(\xi) - \sum_{k=1}^{M} P_{\Pi_k}[H(\xi^{(k)}) - I_{\mathrm{acc}}(\xi^k)]. \tag{3.121}$$

e) Efficiency of a measurement: Using the computed quantities, one can characterize a quantum measurement using the following framework. The amount of extracted information from a measurement is defined as

$$\bar{E} \equiv \frac{H(\xi : \Pi)}{I_{\mathrm{acc}}(\xi)}. \tag{3.122}$$

The residual information is defined as the information that can be potentially

extracted from subsequent measurements

$$\bar{R} \equiv \frac{I'_{\max}(\xi, \Pi) - H(\xi : \Pi)}{I_{\text{acc}}(\xi)}. \tag{3.123}$$

This leaves us with a definition of the destroyed information, which quantifies the reduction of the accessible information due to measurement Π:

$$\bar{D} \equiv \frac{I_{\text{acc}}(\xi) - I'_{\max}(\xi, \Pi)}{I_{\text{acc}}(\xi)} \tag{3.124}$$

The conservation of the total accessible information can thus be expressed as

$$\bar{E} + \bar{R} + \bar{D} = 1. \tag{3.125}$$

For the task of distinguishability, these entropic quantities express the idea of the 'efficiency' of a single quantum measurement.

3.4.2 Uncertainty using Shannon entropy

In the context of a large number of measurements, an unavoidable consequence of quantum information is the Heisenberg uncertainty principle (3.84). However from an information-theoretic perspective, the entropy is a preferred quantity over the standard deviation to measure uncertainty. Indeed, it can be seen therefore, that the uncertainty principle can be reformulated in terms of the Shannon entropy [45, 46, 47, 48].

a) Entropic uncertainty relation: In the uncertainty principle, the standard deviation of observables, X and Z, must satisfy (3.84). Using the spectral expansion (3.39), one obtains the corresponding eigenvectors and their eigenvalues

$$X = \sum_x x \left|x\right\rangle \left\langle x\right|, \tag{3.126}$$

$$Z = \sum_z z \left|z\right\rangle \left\langle z\right|. \tag{3.127}$$

Suppose we measure either one of these observables on a system represented by density operator ρ. Through (3.102), one obtains a distribution for the quantum

probabilities, denoted $p(x)$, associated with the measurement of X; likewise, one obtains a probability distribution, denoted $q(z)$, associated with the measurement of Z. The Shannon entropy (2.20) is a function of only a probability distribution. Hence it is not too difficult to see that we can derive an entropy from $p(x)$, as well as entropy from $q(z)$; these are respectively labelled $H(X)$ and $H(Z)$. The entropic uncertainty relation states that

$$H(X) + H(Z) \geq \log \frac{1}{c}, \tag{3.128}$$

where c is the maximum value of the possible quantities, $c_{xz} = |\langle x|z\rangle|^2$. Moreover, for a system with an associated Hilbert space of dimension d we have the following bounds,

$$0 \leq \log \frac{1}{c} \leq \log d. \tag{3.129}$$

b) Guessing game: One can view the entropic uncertainty relation through the lens of a game. Suppose we have two players whom we name Alice and Bob. The initial role of Bob is to prepare a system in state ρ, and send it to Alice. Alice proceeds to measure either observable A or B with equal probability; the measurement choice is stored in bit Θ whereas the outcome is stored in bit K. The final step of the game is that Alice reveals the choice Θ to Bob. The aim of the game is for Bob to guess K, given the value of Θ.

It can be shown [48], that regardless of the state ρ prepared, the entropic uncertainty relation (3.128) implies that Bob will not be able to perfectly guess K if $\log(1/c) > 0$.

c) Temporal version: Recently [49], it was shown that an entropic uncertainty relation can be formulated for energy and time. The Hamiltonian, H, in (3.65) corresponds to the energy of a system. However, capturing the temporal aspect is non-trivial as there does not exist a Hermitian time operator. Hence, an entropic uncertainty relation was formulated through the construction of a 'quantum clock.' The uncertainty about time corresponds to how well one can 'read off' the time from measuring this clock.

It would be illuminating to view this in terms of a guessing game. Bob prepares a quantum clock in state ρ. He then sends this to Alice. In this modified scenario,

Alice either measures the clock's energy or randomly sets the clock's time; the choicce is made with equal probability; the latter task is accomplished by applying $\exp(-iHt)$ using a random chosen t from a set of values. Depending on what Alice chose to do, Bob's task is either to guess the clock's energy or estimate the value of t by reading the clock. The entropic energy-time uncertainty relation limits Bob's ability to win this guessing game.

3.4.3 The von Neumann entropy

We have witnessed the application of the Shannon entropy in settings involving quantum information. An alternative entropic tool is the von Neumann entropy. The usual treatment of this quantity is found in the subject of quantum statistical mechanics. The approach taken by quantum information science is to describe this quantity in relation to the concepts in classical information theory.

a) Single system: The von Neumann entropy of a quantum density operator ρ is defined as

$$S(\rho) \equiv -\mathrm{tr}(\rho \log_2 \rho). \tag{3.130}$$

One finds that this entropy can re-written as

$$S(\rho) = -\sum_x \lambda_x \log_2 \lambda_x, \tag{3.131}$$

where λ_x are the eigenvalues of ρ. With the latter form, $S(\rho)$ can be seen as a Shannon entropy (2.20) where the eigenvalues are substituted for the probabilities. We also take the convention that $0 \log_2 0 \equiv 0$.

The bounds of the von Neumann entropy are $0 \leq S(\rho) \leq \log_2 d$, where d is the dimension of the Hilbert space. Moreoever, the case of $S(\rho) = 0$ corresponds to a pure state, whereas for a completely mixed state (3.88) we have $S(\rho) = \log_2 d$. Hence a non-zero von Neumann entropy signifies an ignorance (through classical probabilities) as to what the state of the system is.

b) Multiple systems: Suppose we have composite system with two components denoted A and B. This system is collectively described by density operator ρ^{AB}. Analogous to (2.21), we define the von Neumann joint entropy of this sys-

tem as

$$S(\rho^{AB}) \equiv -\text{tr}(\rho^{AB} \log_2(\rho^{AB})). \tag{3.132}$$

Following (2.22), we can define the von Neumann conditional entropy as

$$S(A|B) \equiv S(A, B) - S(B). \tag{3.133}$$

In classical communications, the quantity $H(X|Y)$ can interpreted as the number of additional bits that need to be transmitted to have full knowledge of X, after knowing Y. In an analogous manner, it was recently [50] shown that $S(A|B)$ can be interpreted as a number of qubits that needs to be transmitted to make the task of quantum teleportation (which we'll discuss in the next chapter) possible. The von Neumann mutual information is defined as

$$S(A : B) \equiv S(A) + S(B) - S(A, B), \tag{3.134}$$

and resembles the form of (2.23). Furthermore, it can shown that

$$S(A : B) = S(A) - S(A|B) \tag{3.135}$$

$$= S(B) - S(B|A). \tag{3.136}$$

By considering (2.24), we are then led to define the von Neumann relative entropy (of ρ to σ) as

$$S(\rho||\sigma) \equiv \text{tr}(\rho \log_2 \rho) - \text{tr}(\rho \log_2 \sigma), \tag{3.137}$$

where it can be derived that $S(\rho||\sigma) = 0$ if and only if $\rho = \sigma$.

c) Transformation: For a density operator, recall that a unitary transformation is given by

$$\rho' = U\rho U^\dagger. \tag{3.138}$$

The von Neumann entropy is invariant under this unitary transformation, hence

$$S(\rho') = S(U\rho U^\dagger). \tag{3.139}$$

d) Measurement: Suppose we have a system in state ρ that we would like to

perform a projective measurement on. Let P_i be the corresponding orthogonal projectors for that measurement. If we never learn the result of the measurement, the post-measurement state can be represented as

$$\rho' = \sum_i P_i \rho P_i. \tag{3.140}$$

It can be shown that this procedure in general increases the entropy,

$$S(\rho') \geq S(\rho), \tag{3.141}$$

with equality if and only if $\rho = \rho'$.

e) Properties: The first important property regarding von Neumann entropies is the subadditivity inequality

$$S(A, B) \leq S(A) + S(B). \tag{3.142}$$

A related property is the triangle inequality which is written as

$$S(A, B) \geq |S(A) - S(B)|. \tag{3.143}$$

Of considerable importance is the strong subadditivity inequality

$$S(A, B, C) + S(B) \leq S(A, B) + S(B, C) \tag{3.144}$$

which applies for a system composed of three components denoted A, B, and C. For the conditional entropy associated to a trio of systems, we have the result

$$S(A|B, C) \leq S(A, B). \tag{3.145}$$

In regards to mutual information, one finds that

$$S(A : B) \leq S(A : B, C). \tag{3.146}$$

The monotocity of the relative entropy is a result regarding subsystems

$$S(\rho^A||\sigma^A) \leq S(\rho^{AB}||\sigma^{AB}), \tag{3.147}$$

where ρ^{AB} and σ^{AB} are any two density operators of a system AB. Another significant result is the concavity of the von Neumann entropy which is expressed as

$$S\left(\sum_i p_i\rho_i\right) \geq \sum_i p_i S(\rho_i) \tag{3.148}$$

for probabilities p_i (which sum to unity) and their corresponding density operators ρ_i.

3.4.4 Quantum data compression

Data compression plays a fundamental role in classical information theory; the noiseless channel coding theorem (Theorem 2.3) forms the basis for an operational definition of the Shannon entropy. In this subsection, we provide a brief overview of the quantum noiseless channel coding theorem [51], which provides an operational definition of the von Neumann entropy. A large part of the development towards the theorem relies on the mathematical machinery associated with classical data compression. However, the pioneering nature of the work stems from the conceptual shift of treating the states of quantum theory as information in the most genuine manner. Hence, the significance of this quantum coding theorem cannot be understated for the development of quantum information theory, which is also referred to as the quantum Shannon theory [13].

a) Defining an information source: As in the classical case, the first step is to construct a valid notion of an information source. We define a *i.i.d* quantum information source,$\{H, \rho\}$, as one that can described by a Hilbert space H, and a density operator ρ on that Hilbert space. Furthermore, we utilize the framework of quantum operations (3.94) to help us define a compression scheme of rate R. The compression operation, C^n maps states in $H^{\otimes n}$ to states in a 2^{nR}-dimensional state space. Conversely, D^n represents a decompression operation which takes states in the compressed space back to states in the original Hilbert space.

b) Defining typical states: It will be necessary to recall the definition of a typical sequence (2.42) that was described in Chapter 2. To harness this result, we note that the density operator associated with our information source has a spectral expansion

$$\rho = \sum_x p(x) \left|x\right\rangle \left\langle x\right|, \tag{3.149}$$

where $|x\rangle$ are the eigenvectors with associated eigenvalues $p(x)$. Of crucial importance is that the eigenvalues, in this case, behave like a probability distribution in that they are non-negative and sum to unity. Thus, $S(\rho)$ can be viewed as the Shannon entropy of the set of eigenvalues. Therefore, by using (2.42) we obtain the ϵ-typical sequence $x_1, x_2, \ldots, x_n$ where

$$\left|\frac{1}{n}\log\frac{1}{p(x_1)p(x_2)\ldots,p(x_n)} - S(\rho)\right| \leq \epsilon. \tag{3.150}$$

We define an ϵ-typical state $|x_1\rangle |x_2\rangle \ldots |x_n\rangle$ as one for which $x_1, x_2, \ldots, x_n$ is an ϵ-typical sequence. Related to this concept is the definition of an ϵ-typical subspace, denoted $T(n,\epsilon)$; this is a subspace spanned by all ϵ-typical states, $|x_1\rangle, \ldots, |x_n\rangle$. Moreover, to project onto the subspace $T(n,\epsilon)$, we can use the operator,

$$P(n,\epsilon) = \sum_{x\ \epsilon-\text{typical}} \left|x_1\right\rangle \left\langle x_1\right| \otimes \left|x_2\right\rangle \left\langle x_2\right| \otimes \ldots \left|x_n\right\rangle \left\langle x_n\right|. \tag{3.151}$$

c) Application of Theorem 2.2: One can use the classical theorem regarding typical sequences to prove the following quantum theorem:

Theorem 3.3. ***(Typical subspace theorem)***

i) Fix $\epsilon > 0$. Then for any $\delta > 0$, for sufficiently large n,

$$tr(P(n,\epsilon)\rho^{\otimes n}) \geq 1 - \delta. \tag{3.152}$$

ii) For any fixed $\epsilon > 0$ and $\delta > 0$, for sufficiently large n, the dimension of the subspace, $|T(n,\epsilon)| = tr(P(n,\epsilon))$, satisfies

$$(1-\delta)\, 2^{n(S(\rho)-\epsilon)} \leq |T(n,\epsilon)| \leq 2^{n(S(\rho)+\epsilon)}. \tag{3.153}$$

iii) Let $S(n)$ be a projector onto any subspace of $H^{\otimes n}$ of dimension at most 2^{nR}, where $R < S(\rho)$ is fixed. Then for any $\delta > 0$ and for sufficiently large n,

$$tr(S(n)\rho^{\otimes n}) \leq \delta. \tag{3.154}$$

(See e.g. [2].) □

d) Application of typical subspace theorem: The utility of Theorem 3.3 manifests by its use in proving the quantum analogue of noiseless channel coding theorem (Theorem 2.3). For the sake of briefness, we simply state the end result:

Theorem 3.4. ***(Schumacher's noiseless channel coding theorem)*** *Let $\{H, \rho\}$ be an i.i.d. quantum information source:*

i) If $R > S(\rho)$ then there exists a reliable compression scheme of rate R for the information source

ii) Conversely, if $R < S(\rho)$, then any compression scheme will not be reliable.

(See e.g. [2].) □

e) Comments:

i) From Theorem 3.4, the von Neumann entropy can be operationally defined as the minimum physical resource required to reliably store the output of a quantum information source. Recall that the Shannon entropy is the minimum physical resource required to reliably store the output of a classical information source. Hence in this precise manner, the von Neumann entropy can be considered a quantum generalization of the Shannon entropy. More importantly, we see that entropies in both information theories play the role of signifying optimal data compression.

ii) We have seen that most of the quantum results rely on the mathematics of classical data compression. This is part of a broader framework in which quantum information theory can be seen as a generalization of classical information theory.

4

Quantum Entanglement

> *"I cannot seriously believe in [the quantum theory] because it cannot be reconciled with the idea that physics should represent a reality in time and space, free from spooky actions at a distance."*
>
> – Albert Einstein, co-inventor of quantum theory

THE INTERDEPENDENCE among classical information systems is developed on the violation of probabilistic independence (2.9) described in Chapter 2. We portrayed this property of independence only after introducing the case of a single variable followed by the consideration of multiple variables. Our presentation of quantum information science will proceed in an analogous manner. In Chapter 3, we examined single and multiple quantum information systems through a variety of theoretical tools. Hence, in this chapter we are led to introduce the mathematical description of 'independent' quantum information systems; the notion of interdependence arises naturally in a form known as entanglement; it turns out that entanglement exists across spatial distances (entanglement in space) as well as across temporal intervals (entanglement in time). In Chapter 2, we also described three applications namely classical communication, classical computing, and classical blockchain. In this chapter, we introduce their quantum information analogues using entanglement. Both quantum communications and quantum computing rely on an entanglement in space. The quantum blockchain is one of the first novel applications of an entanglement in time.

4.1 Entanglement in Space

Entanglement, or more precisely entanglement in space, was first theoretically discovered in the Einstein-Podolsky-Rosen (EPR) paradox [52]. They attempted to dismiss the framework of quantum theory by assuming that such an effect could not reasonably exist in the physical world, due to the bizarre implications associated with it. However, the effect has been experimentally well established, most recently to spatial distances exceeding a 1000 kilometers [53]. Entanglement in space has also been historically described as *the* single property that radically distinguishes quantum physics from classical physics [3]. From a modern perspective, such a statement has manifested itself in that the property plays a central and pervasive role in quantum information science. It can be seen as an interdependence among two or more spatially separated quantum information systems that would be impossible to replicate by classical information systems.

In this thesis, we observe that *the interdependence in any entanglement in space is shocking due to the absense of a time interval involved.* Introducing a time interval in the relevant scenario will only make the effect clash less harshly with our classical intuition. This observation was first described in [3], where it was crucially noted that "*The [EPR] paradox would be shaken, though, if an observation did not relate to a definite moment.*"

Our description of entanglement in space will be introduced through the theoretical tools of qubits, density operators and entropy. Each provides a different perspective into the perplexing nature of the spatial interdependence. For detailed reviews on the subject of entanglement in space, we refer the reader to [54, 55, 56, 57] whose material we follow closely. For the rest of this section, we use the term entanglement to solely mean an entanglement in space.

4.1.1 Through qubits

a) Bipartite definition: The entanglement among pure states can easily be described using the quantum circuit model. We constrain our focus even further by considering the bipartite case which is a system composed of two quantum

information subsystems. These can be respectively labelled A and B. The Hilbert space associated to each subsystem is written as H_A with dimension d_A, and H_B with dimension d_B. Then any state vector, representing the composite system, in the Hilbert space $H = H_A \otimes H_B$ is given by

$$|\psi\rangle = \sum_{i,j=1}^{d_A,d_B} c_{ij} |a_i\rangle \otimes |b_j\rangle , \tag{4.1}$$

with a $d_A \times d_B$ matrix C consisting of complex numbers c_{ij}.

A pure state $|\psi\rangle \in H$ is known as *separable*, or as a *product state*, if we can obtain states $|\phi^A\rangle \in H_A$ and $|\phi^B\rangle \in H_B$ such that

$$\boxed{|\psi\rangle = |\phi^A\rangle \otimes |\phi^b\rangle .} \tag{4.2}$$

Otherwise the state $|\psi\rangle$ is referred to as *entangled* or as *nonseparable*.

Quantum separability can be seen to be comparable in some respects to the definition of classical independence (2.9). By looking ahead, we can generalize separability to multipartite systems which consist of multiple subsystems.

b) Multipartite definition: Consider a pure N-partite state $|\psi\rangle$. We refer to the state $|\psi\rangle$ as *fully separable* if it can be written as

$$\boxed{|\psi\rangle = \bigotimes_{i=1}^{N} |\phi_i\rangle .} \tag{4.3}$$

If a state does not satisfy the condition of fully separable, then it contains some entanglement. A pure state is called *m-separable* where $1 < m < N$, if there exists a division of the N parties into m parts $P_1, \ldots, P_m$ such that

$$\boxed{|\psi\rangle = \bigotimes_{i=1}^{m} |\phi_i\rangle_{P_i} .} \tag{4.4}$$

The m-separable state may still contain some entanglement. A state is referred to

as truly *N-partite entangled* when it is neither fully separable, nor m-separable, for any $m > 1$.

As an example, consider the case of $N = 3$ where the respective quantum information subsystems are labelled A, B, and C. The pure three-qubit state are fully separable if they can be written as

$$|\phi^{fs}\rangle_{A|B|C} = |\alpha\rangle_A \otimes |\beta\rangle_B \otimes |\gamma\rangle_C \,. \tag{4.5}$$

Let $m = 2$ to consider the associated biseparable states:

$$|\phi^{bs}\rangle_{A|BC} = |\alpha\rangle_A \otimes |\delta\rangle_{BC} \,, \tag{4.6}$$

$$|\phi^{bs}\rangle_{B|AC} = |\beta\rangle_B \otimes |\delta\rangle_{AC} \,, \tag{4.7}$$

$$|\phi^{bs}\rangle_{C|AB} = |\gamma\rangle_C \otimes |\delta\rangle_{AB} \,. \tag{4.8}$$

Note that the state $|\delta\rangle$ may contain entanglement.

c) Implications: We proceed to describe some properties of well known entangled pure states starting with the bipartite case.

The simplest entangled states are the four *Bell states* (also known as EPR states or EPR pairs)

$$|\Phi^{\pm}_{AB}\rangle = \frac{1}{\sqrt{2}}(|0_A 0_B\rangle \pm |1_A 1_B\rangle), \tag{4.9}$$

$$|\Psi^{\pm}_{AB}\rangle = \frac{1}{\sqrt{2}}(|0_A 1_B\rangle \pm |1_A 0_B\rangle). \tag{4.10}$$

We can describe the generation of these states using the quantum circuit model. Consider starting with the computational basis state $|0_A 0_B\rangle$. After applying the Hadamard gate to the first qubit, we obtain state $(|0_A\rangle + |1_A\rangle)\,|0_B\rangle\,/\sqrt{2}$. The next step of applying the CNOT gate results in the desired output $|\Phi^{+}_{AB}\rangle$. Similar procedures can produce the remaining Bell states.

An entanglement (in space) has an associated interdependence among quantum information systems across spatial distances. This can be portrayed in the fol-

lowing scenario. Suppose we have a bipartite system in Bell state

$$|\Phi^+_{AB}\rangle = \frac{1}{\sqrt{2}}(|0_A 0_B\rangle + |1_A 1_B\rangle), \tag{4.11}$$

where subsystem A can be arbitrarily far from subsystem B. The state $|\Phi^+_{AB}\rangle$ has the property that if we make a projective measurement *only* on subsystem A (in the computational basis), then the post-measurement result for the system is either $|0_A 0_B\rangle$ or $|1_A 1_B\rangle$ (each occuring with probability $1/2$). The point we want to stress is that the state of subsystem B will equate to whatever binary state that subsystem A 'collapses' to. The measurement outcomes are correlated. It also is important to emphasize that prior to the measurement on A, both subsystems are in a superposition in (4.11) and neither can be described to be in a definite state. (Note that a similar analysis occurs for the inverted case where the measurement is on subsystem B). This is remarkable in that subsystem B, who is arbitrarily far away from system A, *instantaneously* takes whatever value that subsystem A is measured to be found in. How is it that subsystem B instantaneously 'knows' the measurement outcome of subsystem A and follows accordingly? This property is what Einstein referred [58] to as "spooky action at a distance." This interdependence of quantum information systems across space is "spooky" precisely due to the instantaneous aspect of it. In other words, it is the *lack of a time interval involved that makes this spatial interdependence shocking*. However, it is important to note that the measurement outcomes $|0_A 0_B\rangle$ or $|1_A 1_B\rangle$ occur randomly. Hence such an effect cannot be used to send classical information instantaneously across vast distances.

It turns out the measurements results are always interdependent. We have witnessed the case of correlated results. Consider the Bell state

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle), \tag{4.12}$$

where the measurements are anti-correlated with respect to the computational basis states. If $\vec{v}$ is any real three-dimensional unit vector, then we can define the observable,

$$\vec{v} \cdot \vec{\sigma} \equiv v_1\sigma_x + v_2\sigma_y + v_3\sigma_z, \tag{4.13}$$

which is referred to as a measurement of spin along the $\vec{v}$ axis. Let the eigenvectors of the observable be denoted $|a\rangle$ and $|b\rangle$. Then it can be shown that

$$\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) = \frac{1}{\sqrt{2}}(|ab\rangle - |ba\rangle), \tag{4.14}$$

up to a global phase factor which we can ignore. This quantitatively shows that the measurement outcomes, for this Bell state, are always anti-correlated.

The Bell states, (4.9) and (4.10), also form an orthonormal basis for a two qubit four dimensional Hilbert space. Hence, one can perform a joint quantum measurement of two qubits that determine which of the four Bell states the two qubits are in. This is known as a Bell state measurement. On a related matter, an important class of operations are LOCC which is an acronym for local operations and classical communications. This means that operations can only be performed locally on the individual subsystems and the subsystems can communicate classically with each other. An example of this is the local application of the Pauli operators (3.43) to change between any of the Bell states

$$(\sigma_x \otimes I)\,|\Phi^{\pm}_{AB}\rangle = |\Psi^{\pm}_{AB}\rangle\,, \tag{4.15}$$

$$(\sigma_x \otimes I)\,|\Psi^{\pm}_{AB}\rangle = |\Phi^{\pm}_{AB}\rangle\,, \tag{4.16}$$

$$(\sigma_z \otimes I)\,|\Phi^{\pm}_{AB}\rangle = |\Phi^{\mp}_{AB}\rangle\,, \tag{4.17}$$

$$(\sigma_z \otimes I)\,|\Psi^{\pm}_{AB}\rangle = |\Psi^{\mp}_{AB}\rangle\,. \tag{4.18}$$

In contrast to Bell state measurements, the ability to distinguish the four Bell states using LOCC is an impossible task and its violation is related to notions of closed timelike curves [59].

Moving from the bipartite case, we proceed to briefly list some well known examples of multipartite entangled pure states. The first of these are the *GHZ (Greenberger-Horne-Zeilinger) states* which are perhaps the most well studied. The GHZ state for N qubits is defined as

$$\boxed{|GHZ_N\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes N} + |1\rangle^{\otimes N}).} \tag{4.19}$$

The second example we wish highlight are the graph states which are defined as follows. Let G be a graph with a set of N vertices and certain number of edges connecting them. For each vertex i, let neigh(i) be defined as the neighborhood of i, which is the set of vertices that are connected to i by an edge. Then for each vertex i, one can construct what is known as as a stabilizer operator,

$$g_i = X_i \bigotimes_{j \in \text{neigh}(i)} Z_j, \tag{4.20}$$

where X_i, Y_i, and Z_i represent Pauli matrices (3.43) applied to the i-th qubit. Using this notation, the *graph state* $|G\rangle$ associated with graph G is the unique common eigenvector to all stabilizing operators g_i,

$$\boxed{g_i |G\rangle = |G\rangle, \quad \text{for } i \in \{1, \ldots, N\}.} \tag{4.21}$$

Notice the important property that

$$\langle G|g_i|G\rangle = 1 \quad \text{for } i \in \{1, \ldots, N\}. \tag{4.22}$$

An important subset of graph states are *cluster states* which are based on square lattice graphs. An example of this is the four qubit cluster state

$$|CL_4\rangle = \frac{1}{2}(|0000\rangle + |0011\rangle + |1100\rangle - |1111\rangle). \tag{4.23}$$

Our third and final example of multipartite entangled pure states are the *Dicke states* which are physically associated with the light emission of a cloud of atoms in excited states. In relation to quantum information science, the most important are the *symmetric Dicke states*, which for N qubits and and k excitations is given by

$$\boxed{|D_{k,N}\rangle = \binom{N}{k}^{-\frac{1}{2}} \sum_j P_j \Big\{ |1\rangle^{\otimes k} \otimes |0\rangle^{\otimes N-k} \Big\},} \tag{4.24}$$

where $\sum_j P_j\{\ldots\}$ represents the sum over all possible permutations of the qubits. An example of such a Dicke state is the *W state* which is the symmetric state of

N particles with a single excitation,

$$|W_n\rangle = \frac{1}{\sqrt{n}}(|0\ldots01\rangle + \ldots|10\ldots0\rangle). \tag{4.25}$$

d) Detection: An important question is how do we show that a state is entangled? For bipartite systems, we consider two types of entanglement detection.

The first is known as the Schmidt decomposition. Suppose we have the pure state

$$|\psi\rangle = \sum_{i,j=1}^{d_A,d_B} c_{ij}\,|a_i b_j\rangle\,, \tag{4.26}$$

which is a state vector in the space $H_A \otimes H_B$. Moreover, we have an associated $d_A \times d_B$ matrix C consisting of the complex numbers c_{ij}. Then the Schmidt decomposition states that there exists an orthonormal basis $|\alpha_i\rangle$ of H_A and an orthonormal basis $|\beta_j\rangle$ of H_B such that

$$|\psi\rangle = \sum_{k=1}^{R} \lambda_k\,|\alpha_k\beta_k\rangle\,, \tag{4.27}$$

where λ_k are positive real coefficients. The values of λ_k are the unique square roots of the eigenvalues of the matrix $CC^\dagger$. The number $R \leq \min\{d_A, d_B\}$ is known as the Schmidt rank of $|\psi\rangle$. Pure product states correspond to states of Schmidt rank one. If it is greater than one, then the state is entangled.

The second method is known as the Bell inequality or more precisely the CHSH inequality [60]. Suppose we have a bipartite system, composed of A and B, in which each subsystem can be measured in two quantities; for system A, this is denoted by A_1 and A_2 and similarly for system B, we have B_1 and B_2; each can take either value $+1$ or -1. The CHSH inequality states that

$$\langle A_1B_1\rangle + \langle A_2B_1\rangle + \langle A_2B_2\rangle - \langle A_1B_2\rangle \leq 2. \tag{4.28}$$

We will see in Chapter 5 that the violation of this result has profound implications for fundamental physics. However from an operational perspective, the violation of this inequality (and its generalization) detects all pure entangled states. More

precisely, for any entangled pure state it is possible to find local measurements such that it violates the CHSH inequality. Furthermore, the only states that do not violate it are product states. To see an explicit example, consider the entangled state

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle), \tag{4.29}$$

and let

$$A_1 = \sigma_z, \quad A_2 = \sigma_x, \quad B_1 = \frac{-\sigma_z - \sigma_x}{\sqrt{2}}, \quad B_2 = \frac{\sigma_z - \sigma_x}{\sqrt{2}}. \tag{4.30}$$

From this we can compute the expectation values of each observable through (3.79). We find that the violation of (4.28) occurs by the left hand side of the inequality equating to $2\sqrt{2}$.

For the case of multiple subsystems, it can be shown that all pure entangled N-partite states violate a generalization of this Bell inequality [57, 61].

4.1.2 Through density operators

a) Bipartite definition: Expressing the definition of entanglement through density operators allows the property to be extended to mixed states. We begin by constraining our attention to the bipartite case, with the subsystems labelled A and B. Suppose we have the density operator

$$\sigma = \sum_i p_i \left|\phi_i\right\rangle \left\langle\phi_i\right|, \tag{4.31}$$

where the state of the the system is known to be in one of $|\phi_i\rangle \in H = H_A \otimes H_B$ with respective classical probabilities p_i. In the literature regarding entanglement, it is often the case that the probabilities which satisfy

$$p_i \geq 0, \qquad \sum_i p_i = 1, \tag{4.32}$$

are referred to as convex weights; this terminology stems from a geometric interpretation. Moreover, a convex combination of density operators σ_i refers to

the quantity

$$\sum_i p_i \sigma_i. \tag{4.33}$$

We say that σ is a *product state* if there exists state σ^A for subsystem A, and state σ^B for subsystem B, such that

$$\boxed{\sigma = \sigma^A \otimes \sigma^B.} \tag{4.34}$$

The density operator σ is called *separable* if there exists convex weights p_i and product states $\sigma_i^A \otimes \sigma_i^B$ such that

$$\boxed{\sigma = \sum_i p_i \sigma_i^A \otimes \sigma_i^B.} \tag{4.35}$$

Otherwise the density operator σ is referred to as *entangled*.

b) Multipartite definition: For an N-partite system, a density operator σ is *fully separable* if it can be written as a convex combination of pure fully separable states

$$\boxed{\sigma = \sum_i p_i \left|\phi_i^{fs}\right\rangle \left\langle\phi_i^{fs}\right|,} \tag{4.36}$$

which can also be written as

$$\boxed{\sigma = \sum_k p_k \sigma_k^{(1)} \otimes \sigma_k^{(2)} \otimes \ldots \sigma_k^{(N)}.} \tag{4.37}$$

A density operator is called *m-separable*, where $1 < m < N$, if it can be written as a convex combination of pure m-separable states. The density operator is said to be *N-partite entangled* when it is neither fully separable, nor m-separable for any $m > 1$.

c) Implications: Through the qubit framework, we witnessed some non-trivial properties regarding entanglement best exemplified through the Bell state (4.11). Other than extending the definition of entanglement to mixed states, density

operators provide a widely different perspective on the puzzling nature of entanglement. To elaborate on this point, consider once again the Bell state (4.11). This can be expressed through the density operator

$$\rho = \left(\frac{|00\rangle + |11\rangle}{\sqrt{2}}\right)\left(\frac{\langle 00| + \langle 11|}{\sqrt{2}}\right) \tag{4.38}$$

$$= \frac{|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|}{2}. \tag{4.39}$$

One can compute the reduced density operator (3.90) of the first qubit as

$$\rho^1 = \mathrm{tr}_2(\rho) \tag{4.40}$$

$$= \frac{\mathrm{tr}_2(|00\rangle\langle 00|) + \mathrm{tr}_2(|11\rangle\langle 00|) + \mathrm{tr}_2(|00\rangle\langle 11|) + \mathrm{tr}_2(|11\rangle\langle 11|)}{2} \tag{4.41}$$

$$= \frac{|0\rangle\langle 0|\langle 0|0\rangle + |1\rangle\langle 0|\langle 0|1\rangle + |0\rangle\langle 1|\langle 1|0\rangle + |1\rangle\langle 1|\langle 1|1\rangle}{2} \tag{4.42}$$

$$= \frac{|0\rangle\langle 0| + |1\rangle\langle 1|}{2} \tag{4.43}$$

$$= \frac{1}{2}I. \tag{4.44}$$

The result is that we obtain a maximally mixed state (3.88) for its subsystem. We can verify this by computing $\mathrm{tr}((I/2)^2) = 1/2 < 1$. Of more interest is the interpretation of this computation. This result is truly perplexing in that the joint state of the system is known *exactly* (ρ is a pure state), and yet at the *at the same time*, we do not have maximal knowledge about its subsystem (ρ^A is a mixed state)! If there was a time interval involved, then perhaps such a property could be explained by a loss or transfer of information among the systems during some period of time. Hence, it is precisely the *lack of a time interval involved that makes this interdependence among the system and its subsystems shocking*.

More broadly speaking, a pure bipartite state is said to be *maximally entangled* if the reduced density matrix on either system is maximally mixed.

d) Detection: Detecting entanglement in mixed states is non-trivial. One way to articulate this is that the the test of the Bell inequality or CHSH inequality (4.28) fails for some entangled mixed states; they do not violate the inequality.

An example of such mixed states are a subset of the Werner states

$$\sigma_W = F\left|\Psi^-\right\rangle\left\langle\Psi^-\right| + \frac{1-F}{3}(\left|\Psi^+\right\rangle\left\langle\Psi^+\right| + \left|\Phi^+\right\rangle\left\langle\Phi^+\right| + \left|\Phi^-\right\rangle\left\langle\Phi^-\right|), \tag{4.45}$$

where we have used Bell states (4.9) and (4.10), and where $0 \leq F \leq 1$. When $F > 0.5$, the density operator σ_w is entangled, and yet these mixed states only violate the Bell inequality when $F > 0.78$.

From such an example, it becomes readily apparent that one needs a new set of theoretical tools. However the question of whether a given density operator is separable or entangled has no known general solution. This problem is called the separability problem. The challenge in mixed states is in detecting the quantum interdependence while ignoring the classical interdependence. Nevertheless we introduce two methods that succeed for certain scenarios.

For the case of bipartite entanglement, there is a tool known as the PPT criterion which is also known as the Peres-Horodecki criterion. Suppose we have a density operator for a composite system and this is expanded in terms of a product basis such that

$$\sigma = \sum_{i,j}^{N}\sum_{k,l}^{M}\sigma_{ij,kl}\left|i\right\rangle\left\langle j\right| \otimes \left|k\right\rangle\left\langle l\right|. \tag{4.46}$$

We define the partial transposition of σ as the transposition with respect to one of its subsystems. An example is that the partial transposition with respect to subsystem A is written as

$$\sigma^{T_A} = \sum_{i,j}^{N}\sum_{k,l}^{M}\sigma_{ji,kl}\left|i\right\rangle\left\langle j\right| \otimes \left|k\right\rangle\left\langle l\right|, \tag{4.47}$$

where we have exchanged the indices i and j. In a similar manner, one can define σ^{T_B} by exchanging k and l. Moreover, a density operator σ is said to have a PPT (positive partial transpose) if its partial transposition has no negative eigenvalues. It is important to note that the spectrum of the density matrix does not depend on what product basis the density operator was expanded in.

The PPT criterion states that if σ is a bipartite separable state, then σ is PPT. Hence, this provides us with a method to detect entanglement. If for a given

density matrix, we compute the partial transpose with its spectrum and obtain negative eigenvalues, then the state is entangled. However, this method does not provide a general sufficient criteria for separability. Nevertheless, we can see its utility on detecting the entanglement such as for the case of Werner states (4.45). We have

$$\sigma_W = F\,|\Psi^-\rangle\langle\Psi^-| + \frac{1-F}{3}(|\Psi^+\rangle\langle\Psi^+| + |\Phi^+\rangle\langle\Phi^+| + |\Phi^-\rangle\langle\Phi^-|), \tag{4.48}$$

and we can compute the partial tranposition with respect to subsystem B in the following manner. For the Bell states we obtain

$$|\Phi^+\rangle\langle\Phi^+|^{T_B} = \frac{1}{2}(|00\rangle\langle 00| + |01\rangle\langle 10| + |11\rangle\langle 00| + |11\rangle\langle 11|), \tag{4.49}$$

$$|\Phi^-\rangle\langle\Phi^-|^{T_B} = \frac{1}{2}(|00\rangle\langle 00| - |01\rangle\langle 10| - |10\rangle\langle 01| + |11\rangle\langle 11|), \tag{4.50}$$

$$|\Psi^+\rangle\langle\Psi^+|^{T_B} = \frac{1}{2}(|01\rangle\langle 01| + |00\rangle\langle 11| + |11\rangle\langle 00| + |10\rangle\langle 10|), \tag{4.51}$$

$$|\Psi^-\rangle\langle\Psi^-|^{T_B} = \frac{1}{2}(|01\rangle\langle 01| - |00\rangle\langle 11| - |11\rangle\langle 00| + |10\rangle\langle 10|). \tag{4.52}$$

From this we can compute

$$\frac{1-F}{3}(|\Psi^-\rangle\langle\Psi^-| + |\Psi^+\rangle\langle\Psi^+|)^{T_B} = \frac{1-F}{3}(2\,|01\rangle\langle 01| + 2\,|10\rangle\langle 10|), \tag{4.53}$$

and

$$\left(\frac{3F}{3}\,|\Phi^-\rangle\langle\Phi^-| + \frac{1-F}{3}\,|\Psi^+\rangle\langle\Psi^+|\right)^{T_B} = \frac{1}{3}((2F+1)(|00\rangle\langle 00| + |11\rangle\langle 11|) + (4F-1)(|01\rangle\langle 01| + |10\rangle\langle 01|)). \tag{4.54}$$

Combining these quantities, the partial transpose of σ in a matrix can be obtained as

$$\sigma^{T_B} = \frac{1}{3}\begin{pmatrix} 2F+1 & 0 & 0 & 0 \\ 0 & 2-2F & 4F-1 & 0 \\ 0 & 4F-1 & 2-2F & 0 \\ 0 & 0 & 0 & 2F+1 \end{pmatrix} \tag{4.55}$$

with eigenvalues equating to $(2F+1)$ and to $(3-6F)$. Therefore, we can correctly identify that entanglement occurs when $F > 0.5$, as this results in $(3-6F)$ becoming negative. This is in contrast to the Bell inequality which is only violated when $F > 0.78$.

Another partial solution to the separability problem are through what are known as entanglement witnesses. These are widely used in experimental settings. Theoretically, these are Hermitian operators (observables) that assist in determining whether a density operator is entangled or not. More formally, an observable W is defined as an entanglement witness if

$$\text{tr}(W\sigma_s) \geq 0 \quad \text{for all separable } \sigma_s, \tag{4.56}$$

$$\text{tr}(W\sigma_e) < 0 \quad \text{for at least one entangled } \sigma_e. \tag{4.57}$$

The underlying mathematical reasoning is based on the Hahn-Banach theorem regarding Hilbert spaces. Physically what is important is that for any entangled state σ_e there always exists an entanglement witness that detects it. However, constructing an entanglement witness is a difficult problem. One construction of an entanglement witness is given by

$$W = \alpha I - |\psi\rangle \langle\psi|, \tag{4.58}$$

where $|\psi\rangle$ represents an entangled pure state, and where the value of α is specific to the case in question. As an example, in the tripartite case an entanglement witness for GHZ is given by

$$W_{GHZ_N} = \frac{3}{4} I - |GHZ_3\rangle |GHZ_3\rangle, \tag{4.59}$$

where for mixed states σ we have

$$\text{tr}(W_{GHZ_N}\sigma) < 0 \quad \rightarrow \quad \sigma \text{ is in the GHZ class,} \tag{4.60}$$

$$\text{tr}(W_{GHZ_N}\sigma) \geq 0 \quad \rightarrow \quad \sigma \text{ is not detected.} \tag{4.61}$$

4.1.3 Through entropy

a) Definition: Another interpretation of the von Neumann entropy (3.130) is in relation to entanglement. More precisely, suppose we have a bipartite system with the subsystems labelled A and B. Moreover, let $|AB\rangle$ denote a pure state of this composite system. Then $|AB\rangle$ is *entangled* if and only if

$$\boxed{S(A|B) < 0,} \tag{4.62}$$

where we have used the conditional von Neumann entropy (3.133), which we rewrite here as

$$S(A|B) \equiv S(A,B) - S(B). \tag{4.63}$$

b) Implications: We aim to examine two properties regarding entangled states from the perspective of entropy.

The first is that the inequality (4.62) implies that for entangled states

$$S(B) > S(A,B), \tag{4.64}$$

which means the uncertainty about the subsystem B is greater than the uncertainty of the composite system AB. This characteristic was expressed earlier through our analysis via density operators. However, the implications of this entropic inequality are far more interesting when we consider the strong subadditivity inequality (3.144). For a tripartite system this can be written as

$$S(A,B,C) + S(B) \leq S(A,B) + S(B,C), \tag{4.65}$$

which can be shown to be equivalent to

$$S(A) + S(B) \leq S(A,C) + S(B,C). \tag{4.66}$$

For entangled systems, it is possible to obtain counter-intuitive results such as $S(A) > S(A,C)$ or $S(B) > S(B,C)$. However we see that the strong subadditivity constrains this freedom in that both of these cases cannot be true *at the same*

time. Hence the *lack of a time interval in this tripartite scenario makes the interdependence among these three quantum systems extremely non-trivial.*

The second property we wish to consider is how entanglement may influence the entropic uncertainy relation (3.128). We refer the reader to [48] for a detailed analysis. To briefly see this, we rewrite the entropic uncertainty relation as

$$H(X) + H(Z) \geq \log \frac{1}{c}. \tag{4.67}$$

More specific to the scenario is how would the uncertainty relation be modified if one is able to have access to entangled states. These would serve as memory or side information that assists in predicting the results of the measurement of X and Z. To answer this we need to introduce what is known as a classical-quantum state which is a classical register X correlated with a quantum memory B, modelled by density operator

$$\rho_{XB} = \sum_x p(x) |x\rangle \langle x| \otimes \rho_B^x. \tag{4.68}$$

Note that $p(x)$ refers to the probability distribution associated with X, and ρ_B^x is the quantum state of the memory conditioned on the classical register taking value $X = x$. From this quantity, we can compute the classical-quantum entropy which is the von Neumann entropy of X conditioned on B,

$$S(X|B) \equiv S(\rho_{XB}) - S(\rho_B), \tag{4.69}$$

where

$$\rho_B = \mathrm{tr}_X(\rho_{XB}) = \sum_x p(x)\rho_B^x. \tag{4.70}$$

The classical-quantum entropy (4.69) is a specific form of the conditional von Neumann entropy (4.63). From these constructions, one can prove the following entropic uncertainty relation

$$S(X|B) + S(Z|B) \geq \log \frac{1}{c} + S(A|B), \tag{4.71}$$

for bipartite quantum state ρ^{AB}, for observables X and Z, and where c, as in (4.67),

is the maximum value of the possible quantities, $c_{xz} = |\langle x|z\rangle|^2$, where

$$0 \leq \log \frac{1}{c} \leq \log d. \tag{4.72}$$

Both classical-quantum conditional entropies $S(X|B)$ and $S(Z|B)$ quantify the uncertainty of X and Z given that one has access to quantum memory B. For a maximally entangled state it can be shown that $S(A|B) = -\log d$ where d is the dimensionality of the respective Hilbert space. Hence we have

$$\log \frac{1}{c} + S(A|B) = \log \frac{1}{c} - \log d \leq 0. \tag{4.73}$$

To interpret this result, recall the guessing game between Alice and Bob associated with (3.128). If we allow Bob access to a maximally entangled quantum memory, then it can be shown that Bob can win the game with probability one. This highlights how entanglement allows one to perform tasks that would be impossible to carry out with only classical resources.

Finally suppose we have a tripartite system ABC represented by density operator ρ_{ABC}. Moreover we have associated observables X and Z. Then it can be shown that

$$S(X|B) + S(Z|C) \geq \log d, \tag{4.74}$$

where d is the dimension of the Hilbert space associated with subsystem A. More generally, one can obtain

$$S(X|B) + S(Z|C) \geq \log \frac{1}{c}, \tag{4.75}$$

where c is defined as in (4.67).

c) Measures: Through the qubit and density operator framework, we were introduced to methods that detected whether a state was entangled or not. Using entropic concepts, we can develop tools to quantify the amount of entanglement in an entangled state. Such tools are known as entanglement measures. We expect for a density operator, σ, an entanglement measure, denoted $E(\sigma)$, satisfies the following properties:

i) For a separable state σ, we have $E(\sigma) = 0$.

ii) It is invariant under unitary transformation, that is

$$E(\sigma) = E(U_A \otimes U_B \sigma U_A^\dagger \otimes U_B^\dagger) \tag{4.76}$$

for a unitary transformation of the form $U_A \otimes U_B$.

iii) $E(\sigma)$ should not increase under an LOCC operations.

We briefly list four common entanglement measures discussed in the literature. Our focus is on the bipartite case (labelled AB), and how they are related to entropic concepts.

The first of these is the entanglement of formation which for density operator σ is written as

$$E_F(\sigma) \equiv \min \sum_i p_i S(\sigma_i^A), \tag{4.77}$$

where we use von Neumann entropy

$$S(\sigma^A) = -\text{tr}\, \sigma^A \log \sigma^A. \tag{4.78}$$

The minimum is over all possibilities of state

$$\sigma^{AB} = \sum_j p_j \left|\psi_j\right\rangle \left\langle\psi_j\right|, \tag{4.79}$$

where

$$\sigma_i^A = \text{tr}_B(\left|\psi_i\right\rangle \left\langle\psi_i\right|). \tag{4.80}$$

It can be interpreted as the minimum number of maximally entangled states that is required to to obtain a certain number of copies of the given state by LOCC.

The second quantity is known as the entanglement of distillation which for a pure state is given by the von Neumann entropy of the reduced state σ_A,

$$E_D(\left|\psi\right\rangle) = S(\sigma_A) = -\text{tr}(\sigma_A \log \sigma_A). \tag{4.81}$$

It can be interpreted as the number of maximally entangled states that can be derived from an initial number of non-maximally entangled states using LOCC.

Another useful measure is known as the relative entropy of entanglement which

is defined as

$$E_R(\sigma) \equiv \min_{\varrho \in D} S(\sigma||\varrho), \tag{4.82}$$

where $S(\sigma||\varrho)$ is the von Neumann relative entropy (3.137), and D is the set of all disentangled states. It quantifies the amount of entanglement through a distance measure.

Finally the concurrence for a pure state is given by

$$C(|\psi\rangle) = \sqrt{2(1 - \mathrm{tr}(\sigma_A^2))}, \tag{4.83}$$

where σ_A is the reduced subsystem of $|\psi\rangle$. For the two qubit case, the concurrence is related to the entanglement of formation

$$E_F(\sigma) = h\left(\frac{1 + \sqrt{1 - C^2(\sigma)}}{2}\right), \tag{4.84}$$

where we use binary version of the Shannon entropy, $h(p) = -p \log p - (1 - p) \log(1 - p)$.

4.2 Application: Quantum Communication

Entanglement in space can be seen as a resource in quantum information in that it allows the ability to perform information tasks that would be impossible or very difficult to do with only classical information. The three different communication protocols described in this section serve to illustrate this point. Each protocol is described in the context of two parties, named Alice and Bob, who are some arbitrary distance apart. More crucially, each share a qubit from a spatial Bell state. It is also common in these protocols to design a code that relates the classical and quantum information. These applications are instrumental for the construction of a useful quantum communications network [62, 63].

4.2.1 Superdense coding

a) Protocol: This information task requires Alice to send two bits of classical information to Bob using a single qubit [64]. The protocol starts by assuming Alice and Bob share the spatial Bell state

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle\,|0\rangle + |1\rangle\,|1\rangle). \tag{4.85}$$

Moreover, they have agreed to encode the classical information in the following way: The bit string xy, where $xy = 00, 01, 10, 11$ corresponds to Bell state

$$|\beta_{xy}\rangle = \frac{1}{\sqrt{2}}(|0\rangle\,|y\rangle + (-1)^x\,|1\rangle\,|\bar{y}\rangle), \tag{4.86}$$

where $\bar{y}$ is the negation of y.

The protocol is as follows: If Alice wants to send bit string 00, she simply sends her qubit to Bob. However if Alice wants to send string 01, she applies the X operator on her qubit before sending it to Bob

$$(X \otimes I)\,|\psi\rangle = \frac{1}{\sqrt{2}}(|1\rangle\,|0\rangle + |0\rangle\,|1\rangle) = |\beta_{01}\rangle\,. \tag{4.87}$$

For the case of sending bits 10, she applies a Z operator,

$$(Z \otimes I)\,|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle\,|0\rangle - |1\rangle\,|1\rangle) = |\beta_{10}\rangle\,. \tag{4.88}$$

And for the last case of 11, she applies the iY gate before sending her qubit to Bob

$$(iY \otimes I)\,|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle\,|1\rangle - |1\rangle\,|0\rangle) = |\beta_{11}\rangle\,. \tag{4.89}$$

Once Bob receives the qubit from Alice, he performs a projective measurement, in the Bell basis on both qubits. From that, he is able to recover bit string xy from identifying state $|\beta_{xy}\rangle$.

b) Comments:

i) This information task would be impossible to perform, in the classical case,

had Alice only transmitted a single classical bit.

ii) Superdense coding has recently been experimentally demonstrated within an optical fiber infrastructure [65].

4.2.2 Quantum teleportation

a) Protocol: The following task [66] requires Alice to send a particular set of quantum information to Bob without that information traversing the space between them. More precisely, Alice wants to send Bob a qubit $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, where the values of α and β are unknown to both parties. They both share the Bell state

$$|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \tag{4.90}$$

as well as have access to a classical communications channel which transmits bits. The initial state of this scenario can written as

$$|\psi_0\rangle = |\psi\rangle\,|\beta_{00}\rangle = \frac{1}{\sqrt{2}}[\alpha|0\rangle\,(|00\rangle + |11\rangle) + \beta|1\rangle\,(|00\rangle + |11\rangle)], \tag{4.91}$$

where the first two qubits are in Alice's possession, while the third qubit belongs to Bob. The first step is that Alice applies a CNOT gate (3.61) to both of her qubits, in which case the state transforms to

$$|\psi_1\rangle = \frac{1}{\sqrt{2}}[\alpha|0\rangle\,(|00\rangle + |11\rangle) + \beta|1\rangle\,(|10\rangle + |01\rangle)]. \tag{4.92}$$

From there, she proceeds to apply a Hadamard gate to her first qubit. This produces the overall state

$$|\psi_2\rangle = \frac{1}{2}[\alpha(|0\rangle + |1\rangle)(|00\rangle + |11\rangle) + \beta(|0\rangle - |1\rangle)(|10\rangle + |01\rangle)], \tag{4.93}$$

which can re-written as

$$\begin{aligned}|\psi_2\rangle = \frac{1}{2}\Big(&|00\rangle\,(\alpha|0\rangle + \beta|1\rangle) + |01\rangle\,(\alpha|1\rangle + \beta|0\rangle)\\ &+ |10\rangle\,(\alpha|0\rangle - \beta|1\rangle) + |11\rangle\,(\alpha|1\rangle - \beta|0\rangle)\Big).\end{aligned} \tag{4.94}$$

When Alices measures her qubits, in her computational basis states, she gets one of the results on the left in (4.95). Bob would then apply the corresponding Pauli operator (3.43) on his qubit to obtain $|\psi\rangle$:

$$\begin{aligned} 00 &\rightarrow \text{Does nothing,} \\ 01 &\rightarrow \text{Applies } \sigma_x = |0\rangle\langle 1| + |1\rangle\langle 0|, \\ 10 &\rightarrow \text{Applies } \sigma_z = |0\rangle\langle 0| - |1\rangle\langle 1|, \\ 11 &\rightarrow \text{Applies } \sigma_z\sigma_x. \end{aligned} \tag{4.95}$$

Bob receives the two bits from Alice in (4.95) through the classical channel. In this case, one view (4.95) as a code to relate the classical and quantum information.

b) Comments:

i) The notion of teleportation can be seen in that the quantum information disappears from Alice's location and re-appears in Bob's location. Of crucial necessity to perform this task is the initial Bell state (4.90). This emphasizes the point that entanglement in space can be regarded as a resource in quantum information; it allows us to carry out information tasks that would be impossible to do with only classical resources.

ii) In Chapter 3 we saw measurement as a process that converts quantum information into classical information. This protocol alludes to a more general property in that one can also convert the classical information back to the quantum information as long as the quantum measurement does not reveal any information about the state being measured.

iii) Throughout the duration of the protocol, there is always at most one copy of $|\psi\rangle$. Hence at no time is the no-cloning theorem violated.

iv) When Alice performs a measurement on her qubits in (4.94), the quantum information residing in Bob's qubit is instantaneously affected. The lack of a time interval involved in this process suggests a violation of relativity. However, such a concern can be largely alleviated in that Bob still requires the two bits from the classical channel (whose transmission is limited by the speed of light) to obtain $|\psi\rangle$. The density operator framework clearly illustrates this point: Each of the outcomes in (4.95) from measuring (4.94)

occur with probability 1/4. Hence the density operator of the system after Alice's measurement is given by

$$\begin{aligned}\rho = \; & \tfrac{1}{4}\Big(|00\rangle\langle 00|\,(\alpha|0\rangle + \beta|1\rangle)(\alpha^*\langle 0| + \beta^*\langle 1|) \\ + \; & |01\rangle\langle 01|\,(\alpha|1\rangle + \beta|0\rangle)(\alpha^*\langle 1| + \beta^*\langle 0|) \\ + \; & |10\rangle\langle 10|\,(\alpha|0\rangle - \beta|1\rangle)(\alpha^*\langle 0| - \beta^*\langle 1|) \\ + \; & |11\rangle\langle 11|\,(\alpha|1\rangle - \beta|0\rangle)(\alpha^*\langle 1| - \beta^*\langle 0|)\Big).\end{aligned} \tag{4.96}$$

By using (3.90), we obtain the reduced density operator of Bob's system which can be computed as

$$\begin{aligned}\rho^B = \; & \tfrac{1}{4}\Big((\alpha|0\rangle + \beta|1\rangle)(\alpha^*\langle 0| + \beta^*\langle 1|) \\ + \; & (\alpha|1\rangle + \beta|0\rangle)(\alpha^*\langle 1| + \beta^*\langle 0|) \\ + \; & (\alpha|0\rangle - \beta|1\rangle)(\alpha^*\langle 0| - \beta^*\langle 1|) \\ + \; & (\alpha|1\rangle - \beta|0\rangle)(\alpha^*\langle 1| - \beta^*\langle 0|)\Big).\end{aligned} \tag{4.97}$$

One can simplify this expression to

$$\rho^B = \frac{2(|\alpha|^2 + |\beta|^2)\,|0\rangle\langle 0| + 2(|\alpha|^2 + |\beta|^2)\,|1\rangle\langle 1|}{4} \tag{4.98}$$

$$= \frac{|0\rangle\langle 0| + |1\rangle\langle 1|}{2} \tag{4.99}$$

$$= \frac{I}{2}. \tag{4.100}$$

This means that prior to receiving the classical measurement results from Alice, the state appears totally random to Bob. Nevertheless, there is an instantaneous effect across space on the quantum information held by Bob when Alice makes a measurement. This remaining issue would be resolved where a time interval introduced into that process.

v) Quantum teleportation has been demonstrated experimentally, most recently from a ground station to a space-based satellite [67]

c) Monty Hall teleportation: The teleportation protocol has been extended to probabilistic scenarios [68, 69, 70]. In this section, we present a probabilistic

version [71] of quantum teleportation that is *part of the original component of this thesis* (which was done in collaboration with my supervisor). We modify the standard teleportation protocol into the Monty Hall game which was described in detail in Chapter 2. Alice can be viewed as Monty, and Bob as the contestant. The four doors are respectively labelled $(00, 01, 10, 11)$. This coincides with Alice's possible measurement results in (4.95); the prize door is Alice's actual result, whose bits we denote ab, and what Bob would need get the desired state $|\psi\rangle$. The contestant's initial choice of door would be equivalent to what Bell state was used at the start of the protocol. In this modification, the contestant is allowed to choose any of the four doors $(00, 01, 10, 11)$, which we denote xy. This event coincides with using Bell state

$$|\beta_{xy}\rangle = \frac{1}{\sqrt{2}}(|0\rangle\,|y\rangle + (-1)^x\,|1\rangle\,|\bar{y}\rangle), \tag{4.101}$$

where $\bar{y}$ is the negation of y. As an example, if the contestant chooses door 10, then a way to implement this is that Bob applies the operator $(\sigma_0 \otimes \sigma_z)\,|\beta_{00}\rangle = |\beta_{10}\rangle$, and communicates that to Alice; the last step would be analogous to Monty being aware of what door the contestant chooses. In this modified protocol, the initial state is represented as

$$\begin{aligned} |\psi\rangle\,|\beta_{xy}\rangle =& \frac{1}{\sqrt{2}}(\alpha\,|0\rangle + \beta\,|1\rangle)(|0\rangle\,|y\rangle + (-1)^x\,|1\rangle\,|\bar{y}\rangle) \\ =& \frac{\alpha(|00y\rangle + (-1)^x\,|01\bar{y}\rangle) + \beta(|10y\rangle + (-1)^x\,|11\bar{y}\rangle)}{\sqrt{2}}. \end{aligned} \tag{4.102}$$

After Alice applies a CNOT gate to her qubits, the state can be found in

$$\frac{\alpha(|00y\rangle + (-1)^x\,|01\bar{y}\rangle) + \beta(|11y\rangle + (-1)^x\,|10\bar{y}\rangle)}{\sqrt{2}}. \tag{4.103}$$

This is equivalent to

$$\frac{\alpha\,|0\rangle\,(|0y\rangle + (-1)^x\,|1\bar{y}\rangle)}{\sqrt{2}} + \frac{\beta\,|1\rangle\,(|1y\rangle + (-1)^x\,|0\bar{y}\rangle)}{\sqrt{2}}. \tag{4.104}$$

Alice proceeds to apply the relevant Hadamard gate which provides the result

$$\frac{1}{2}\Big(|00\rangle (\alpha |y\rangle + \beta(-1)^x |\bar{y}\rangle) + |01\rangle (\alpha(-1)^x |\bar{y}\rangle + \beta |y\rangle) \\ + |10\rangle (\alpha |y\rangle - \beta(-1)^x |\bar{y}\rangle) + |11\rangle (\alpha(-1)^x |\bar{y}\rangle - \beta |y\rangle)\Big).$$

At this step, Alice measures her qubits to get her result. If Alice's result is $ab = xy$, meaning it coincides with the Bell state used $|\beta_{xy}\rangle$, then Bob has to do nothing and he has the desired state $|\psi\rangle$ (the exception is if the initial Bell state used was $|\beta_{11}\rangle$ in which case Bob has to apply operator $(-\sigma_0)$ to get $|\psi\rangle$ if result is 11). This is why the contestant's initial choice relates to the Bell state used. As an example, if the initial Bell state was $|\beta_{01}\rangle$ and Alice's measurement outcome was bits 01, then Bob's state is automatically in $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$.

In this Monty Hall protocol, Alice sends Bob two bits as in (4.95) with the following modification: she chooses two bits denoted cd (ie goat door) that are not xy (ie contestant's initial choice) and are not ab (ie prize door). Should Bob do nothing, or apply one of the possible operators (which depend on what Bell state was used) to get $|\psi\rangle$ ie should the contestant stick or switch?

To answer this, let B_{xy} be the door chosen by contestant. For this example, assume we use $|\beta_{00}\rangle$, hence $P(B_{00}) = 1$. Let A_{ab} be the prize door and due to Born probabilities we have $P(A_{ab}) = 1/4$. Let C_{cd} be the goat door opened by Monty whose probabilities, from the protocol description, work out as:

$$P(C_{cd} \mid B_{00}, A_{ab}) = \begin{cases} \frac{1}{3}, & \text{if } 00 = ab \neq cd, \\ \frac{1}{2}, & \text{if } 00 \neq ab \neq cd, \\ 0, & \text{otherwise.} \end{cases} \tag{4.105}$$

If Bob always does nothing (ie, stick strategy), then

$$P(\text{win if stick}) = \sum_{ab=00\neq cd} P(A_{ab}, B_{00}, C_{cd}) = \frac{2}{8}. \tag{4.106}$$

Suppose Bob decides to always apply one of the two operators (ie, switch strategy). Then there are one of two possibilities which we denote ef and given its

a random choice, each occur with probability $1/2$. Let D_{ef} represent that door, and P(win if switch) is

$$\sum_{ab=ef\neq cd\neq 00} P(A_{ab}, B_{00}, C_{cd}, D_{ef}) = \frac{3}{8}. \tag{4.107}$$

This means Bob should apply one of the two operators (switch) rather than do nothing (stick) to get state $|\psi\rangle$.

d) Unreliable teleportation: The effect of noise has been widely analyzed for teleportation [72, 73, 74, 75, 76]. In this part, we present a second modification [71] of quantum teleportation, involving noise, that is *part of the original component of this thesis* (which was done in collaboration with my supervisor). Consider the standard teleportation protocol with the following unreliability: one of the two bits (either the first or second) Alice sends to Bob in (4.95) is received but the other is lost; each event occurs with probability $1/2$. If the initial Bell state is $|\beta_{00}\rangle$ and Alice's result is 00, then Bob can do nothing. But in this scenario, if Bob receives the single bit as 1, then the possible options are 01, 10, or 11; in this case he should apply one of the operators (switch). If Bob receives bit 0, then his options are 00, 01, 10. Should he stick (to 00) or switch (to 01 or 10)? To answer this, let us use the notation developed in the Monty Hall protocol.

We have $P(B_{00}) = 1$ and $P(A_{ab}) = 1/4$. Let d in C_d be the single bit received by Bob; based on the scenario described above, we have $P(C_0 \,|\, B_{00}\, A_{00}) = 1$, $P(C_0 \,|\, B_{00}, A_{01}) = 1/2$, and $P(C_0 \,|\, B_{00}, A_{10}) = 1/2$. We can compute the probability that Bob receives bit 0:

$$P(\text{received bit 0}) = \sum_{ab\neq 11} P(C_0, B_{00}, A_{ab}) = \frac{1}{2}.$$

If Bob decides to always do nothing then this would be like a sticking strategy. The probability that bit 0 is received and Bob wins by sticking is given by $P(A_{00}, B_{00}, C_0) = 1/4$. Hence we can compute the conditional probability:

$$P(\text{win if stick} \,|\, \text{received bit 0}) = \frac{1/4}{1/2} = \frac{1}{2}. \tag{4.108}$$

If an always switching strategy is adopted, then there are two possibilities (01 or 10) each occuring with probability 1/2. In this case probability of winning if switched and bit 0 is received is given by $P(A_{01}, B_{00}, C_0, D_{01})+P(A_{10}, B_{00}, C_0, D_{10}) = 1/8$. With that we compute,

$$P(\text{win if switch} \mid \text{received bit } 0) = \frac{1/8}{1/2} = \frac{1}{4}. \tag{4.109}$$

It is an advantage to stick ie Bob should do nothing. This strategy may serve to be useful in an error-correcting design for reliability issues in practical quantum networks [62, 67]

4.2.3 Quantum cryptography

a) Preliminaries: The secure exchange of messages in classical communications is mainly carried out using public key cryptography, which was described in Chapter 2. However, as we shall show in the next section on quantum computing, a dramatic result is that a scalable quantum computer would be able to break public key cryptography by solving prime factorization. This discovery has radically changed the landscape of cryptographic research. There are various investigations that aim to build information security systems based on mathematical problems that many believe a quantum computer would not be able to solve. These are collectively referred to as post-quantum cryptography [77, 78]. Perhaps the greatest drawback of this set of solutions is that their durability can be questioned; it may be the case that one finds a way for a quantum computer to solve such problems in the future; more precisely stated, there are no formal proofs that such solutions are secure against a quantum computing attack.

Besides public key cryptography, another classical method to encrypt and decrypt messages is through private key cryptography like the one-time pad. In this scenario, Alice and Bob each possess an identical copy of a random string of bits known as the private key. More crucially, only they are aware of the key values and keep them in secret. As long as the key is kept in secrecy, this method is is shown to be provably secure. When Alice wants to transmit a secure message to Bob, she encrypts the message using this private key by adding the random

key bits to the message. When Bob receives the encrypted message, he decrypts it by subtracting the key bits, using his own copy of the private key. Hence the problem of transmitting secure messages can be reduced to the problem of how can Alice and Bob acquire pre-established perfectly correlated random keys that an adversary would not be able to acquire. This subroutine can be accomplished using quantum information, and is known a quantum key distribution (QKD) or quantum cryptography [79]. The keys generated by this task are guaranteed to be secure through the properties of quantum information, and hence through the laws of physics; this is in vast contrast to the security of public key cryptography which is based on the difficulty of solving certain mathematical problems. We discuss two protocols that implement QKD, where the second involves an entanglement in space.

b) BB84 protocol: In this scenario [80, 34], our task is for Alice and Bob to acquire identical private keys. There are two communication channels between Alice and Bob. The first is a quantum communication channel that transmits qubits, while the second is a public classical channel for transmitting bits. In this protocol, we consider the two bases, $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$. They are related to one another in the following way:

$$|0\rangle = \frac{|+\rangle + |-\rangle}{\sqrt{2}}, \quad |1\rangle = \frac{|+\rangle - |-\rangle}{\sqrt{2}}, \tag{4.110}$$

$$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}, \quad |-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}. \tag{4.111}$$

We also use the encoding that logical 0 is represented by states $|0\rangle$ and $|+\rangle$, whereas logical 1 is represented by states $|1\rangle$ and $|-\rangle$. Alice creates a random string of classical bits. She encodes this into a corresponding string of qubits using the code. She sends these qubits through the quantum channel to Bob. From there, Bob chooses to measure each qubit in either the $\{|0\rangle, |1\rangle\}$ basis or the $\{|+\rangle, |-\rangle\}$ basis; he makes this choice, for each qubit, randomly.

All the four states, in (4.110) and (4.111), are not mutually orthogonal, and therefore there is no quantum measurement to distinguish each of them with certainty. This creates two cases. If Alice and Bob used the same bases for their respective

tasks, then their results are perfectly correlated. As an example if Alice prepared state $|0\rangle$ and Bob measures in basis $\{|0\rangle, |1\rangle\}$, then he will find state $|0\rangle$ with certainty. If on the other hand they used different bases, there is a chance of an error. As an example if Alice prepared state $|0\rangle$ and Bob measures in the $\{|+\rangle, |-\rangle\}$ basis, then there is a probability of $1/2$ of Bob obtaining the incorrect state $|1\rangle$. From these measurements and the code, Bob obtains a corresponding string of classical bits.

From there, Alice and Bob proceed to employ the classical channel to tell each other what basis was used at each position. They discard the bits in their strings where they used a different basis for their respective quantum tasks. As a consequence, both Alice and Bob end up with perfectly correlated classical private keys whose values are only known to them.

c) Security analysis: Suppose an adversary, named Eve, is attempting to obtain information about the private key. There are a number of features of quantum information that make this impossible. We have seen that Alice and Bob use the classical communication channel to share what basis was used at each position. This information can be public since it cannot be used to infer what the prepared and measured value of the qubit was at the respective position. In regards to the quantum communication channel, Eve cannot copy the qubits transmitted due to the no-cloning theorem. Even more striking is that it is impossible for Eve gain any information on non-orthogonal qubits without introducing a disturbance on the signal. This is why we have used non-orthogonal states in the protocol. More broadly speaking,

Proposition 4.1. ***(Information gain implies disturbance)*** *In any attempt to distinguish between two non-orthogonal quantum states, information gain is only possible at the expense of introducing disturbance to the signal. (See e.g. [2].)*

Hence at the end of the protocol, Alice and Bob select a subset of bits from their final strings to compare the values. If more than an acceptable number disagree, they abort the protocol and try again.

d) E91 protocol: Having introduced the need for correlations in the BB84 protocol, it seems appropriate to ask whether the *interdependence in an entanglement*

in space can be used for a QKD protocol? It turns out that such an answer was first developed in [81]. This is known as the E91 protocol. It utilizes the Bell state

$$|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \tag{4.112}$$

which can be re-written in terms of the $\{|+\rangle, |-\rangle\}$ basis as

$$|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(|++\rangle + |--\rangle). \tag{4.113}$$

Once again the task is for Alice and Bob to acquire identical private keys made up of random values. In this protocol, one qubit from state $|\beta_{00}\rangle$ is held by Alice and the other qubit by Bob. By considering both (4.112) and (4.113), it appears that if Alice and Bob measure in the same basis, then their results are perfectly correlated and yet random. Furthermore, to obtain an appropriate key length, we suppose that Alice and Bob share many copies of $|\beta_{00}\rangle$ and repeat the measurement procedure over many rounds; they use a public classical channel to randomly agree to measure in either basis $\{|0\rangle, |1\rangle\}$ or $\{|+\rangle, |-\rangle\}$.

In the BB84 protocol, the key is fundamentally produced by Alice and sent to Bob (before measurement and the removal of some results). In the E91 protocol, Alice and Bob can measure their qubits simultaneously and obtain their respective keys. *The lack of a time interval involved* suggests that the key is not fundamentally distributed, in any way, from one location to another like in BB84. Rather identical keys are generated at same time at two different locations, and whose values cannot be pre-determined by Alice nor Bob!

e) Security analysis: We provide a brief outline of a security proof [48, 82] for E91. In order to do accomplish this result, we have to show that the following two statements are mutually exclusive:

i) The measurement results between Alice and Bob agree on most rounds.
ii) An adversary, whom we can name Eve, possesses a large amount of information on the results of either Alice or Bob.

In the protocol, we assumed that Alice and Bob share state $|\beta_{00}\rangle$. However, it may be the case that Eve interfered. Hence, let ρ_{ABE} represent a density operator

where A represents Alice's qubit, B represents Bob's qubit and E signifies any quantum subsystem acquired by Eve. Let Θ be a mixed state whose role is to be a binary register which signifies whether the qubits are to be measured in basis $\{|0\rangle, |1\rangle\}$ or $\{|+\rangle, |-\rangle\}$. Furthermore, let Y denote the measurement results of Alice, and let $\bar{Y}$ denote the measurement results Bob obtains. Alice and Bob measure their system in the basis as indicated by Θ; we assume that Eve also holds state Θ. We first consider an analysis on Alice's results. Using entropic concepts, we obtain

$$S(Y|B\Theta) = \frac{1}{2} S(X|B) + \frac{1}{2} S(Z|B), \tag{4.114}$$

$$S(Y|E\Theta) = \frac{1}{2} S(X|E) + \frac{1}{2} S(Z|E). \tag{4.115}$$

Applying the tripartite entropic uncertainty principle (4.75) with quantum memory results in,

$$S(Y|B\Theta) + S(Y|E\Theta) \geq 1, \tag{4.116}$$

given $q_{MU} = 1$ for bases $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$. In Bob's case, the following result can be derived

$$S(Y|B\Theta) \leq H(Y|\bar{Y}). \tag{4.117}$$

This brings us to the final result

$$S(Y|E\Theta) \geq 1 - H(Y|\bar{Y}). \tag{4.118}$$

The interpretation of (4.118) is that the von Neumann entropy relating to Eve's uncertainty is small given that the conditional entropy between Alice and Bob $H(Y|\bar{Y})$ is large; this provides the necessary expression to show that the two statements of the proof are mutually exclusive as required. The argument can be extended to multiple rounds.

f) Comments:

i) QKD protocols are rigorously proven to be secure using the laws of physics. A formal definition for security along with a proof can be found in [2]. Hence, unlike classical cryptography, quantum cryptography provides a guaranteed level of protection against a quantum computing attack.

ii) It is of noteworthy interest that the initial idea of using quantum physics for cryptography was first formulated in a design for money bank notes that would be impossible to forge. However this work was rejected for publication. For a historic and broad review of the subfield of quantum cryptography, we refer the reader to [79].

iii) Other than Bell states, the more general GHZ states have also been employed in cryptographic settings. One notable example is in a quantum version of the classical secret sharing protocol [83].

iv) Quantum key distribution systems have moved from theory to commercial reality. There are a number of quantum cryptographic companies deploying these systems in the public and private sector.

4.3 Application: Quantum Computing

We have witnessed the use of entanglement in space in quantum information to perform communication tasks that would be classically unimaginable. In this section, we introduce the notion of a quantum computer [84, 85, 86] that harnesses quantum information, which includes an entanglement in space resource [87, 88, 56], to solve computational problems. There are a number of different models for quantum computation such as the gate model [2], the adiabatic model [89, 90], the topological model [91] and the one-way measurement model [92]. Our sole focus is on the gate model which is based on the quantum circuit model presented in Chapter 3. Moreover, we present three quantum algorithms that remarkably outperform the best known classical algorithms for the same task. However, it is not formally proven that quantum computers are more powerful than classical computers; it may very well be the case that we find classical algorithms that are equivalent in computational performance. Nevertheless, small-scale quantum computing systems have been experimentally realized and shown [93] to drastically outpeform the world's best classical supercomputers. For a broader survey of quantum algorithms, we refer the reader to [94, 95].

4.3.1 Quantum search

a) Grover's algorithm: The classical computational problem of search involves finding M solutions in a search space of N elements, where $1 \leq M \leq N$. To solve this problem using a quantum computer [96], we encode each of these N elements into a quantum state $|x\rangle$, and create the following state that is in an equal superposition

$$|\psi\rangle = \frac{1}{N^{\frac{1}{2}}} \sum_{x=0}^{N-1} |x\rangle \,. \tag{4.119}$$

Suppose a single solution is marked as x'. Then the goal of the quantum computer is to transform state $|\psi\rangle$ into the state $|x'\rangle$ using the fewest number of operations and measurements. To perform this task, we simply need to construct an operator known as the Grover operator,

$$G = (2\,|\psi\rangle\,\langle\psi| - I)O, \tag{4.120}$$

where $(2\,|\psi\rangle\,\langle\psi| - I)$ can be constructed using (4.119). The operator O is known as the oracle; the action of the oracle is given by

$$O\,|x\rangle = (-1)^{f(x)}\,|x\rangle\,, \tag{4.121}$$

where $f(x') = 1$, and otherwise the function evaluates to zero for all other x. It is important to note that the oracle can only recognize the solution to the search problem. There is a clear distinction between recognizing the solution and knowing the solution. The former does not mean the latter.

The quantum algorithm is straightforward in that it consists of repeatedly applying the Grover operator $\pi\sqrt{N}/4$ times on state $|\psi\rangle$. This transforms the quantum information in $|\psi\rangle$ such that when measured gives with high probability the result $|x'\rangle$. To see this to be the case, let $\sum_x^{'}$ represent the sum over all x which are solutions, and $\sum_x^{''}$ represent the sum over all x which are not solutions. We can

construct the following normalized quantum states

$$|\alpha\rangle = \frac{1}{\sqrt{N-M}} \sum_{x}^{''} |x\rangle \,, \tag{4.122}$$

$$|\beta\rangle = \frac{1}{\sqrt{M}} \sum_{x}^{'} |x\rangle \,. \tag{4.123}$$

We can re-express initial state $|\psi\rangle$ of the quantum computer, represented in (4.119), in terms of $|\alpha\rangle$ and $|\beta\rangle$,

$$|\psi\rangle = \sqrt{\frac{N-M}{N}} |\alpha\rangle + \sqrt{\frac{M}{N}} |\beta\rangle \,. \tag{4.124}$$

Furthermore, let

$$\cos\frac{\theta}{2} = \sqrt{\frac{N-M}{M}}, \tag{4.125}$$

so that we can write (4.124) as

$$|\psi\rangle = \cos\frac{\theta}{2} |\alpha\rangle + \sin\frac{\theta}{2} |\beta\rangle \,. \tag{4.126}$$

The effect of applying the Grover operator on the initial state results in

$$G|\psi\rangle = \cos\frac{3\theta}{2} |\alpha\rangle + \sin\frac{3\theta}{2} |\beta\rangle \,. \tag{4.127}$$

The repeated iteration of the operator on the state computes to

$$G^k |\psi\rangle = \cos\left(\frac{(2k+1)\theta}{2}\right) |\alpha\rangle + \sin\left(\frac{(2k+1)\theta}{2}\right) |\beta\rangle \,. \tag{4.128}$$

This has the requirement of transforming the state $|\psi\rangle$ to $|\beta\rangle$. More precisely, the number of iterations required is upper bounded by

$$\frac{\pi}{4}\sqrt{\frac{N}{M}}. \tag{4.129}$$

After this repeated application of G on initial state $|\psi\rangle$, a measurement in the computational basis provides the answer to the problem with a high probability.

b) Analysis of algorithm: In Chapter 2, we provided a brief overview of the asymptotic notation. Using those tools, it can be said that Grover's algorithm requires $O(\sqrt{N/M})$ operations for an N item search problem with M solutions. For the case of a single solution, this equates to $O(\sqrt{N})$. A classical computer for the same single solution case requires $O(N)$ operations. To highlight the shocking aspect of this situation, consider a search space of a million items; a classical computer would need to, at worst, go through all million of them whereas a quantum computer simply needs to search through, at worst, a thousand of them; this is remarkable as there does not seem to be any geometric structure in the problem to offer such a quadratic speed up.

4.3.2 Quantum factoring

a) Preliminaries: Perhaps the most influential result in quantum information science is Shor's algorithm [97, 98]:

i) This is a quantum algorithm that can efficiently derive the prime factorization of an integer. In Chapter 2, we saw that the reliability of public key cryptography is based on the hypothesis that prime factorization cannot be computed in any reasonable time. Hence, Shor's algorithm has dramatic consequences on the information infrastructure of the modern world.

ii) Shor's algorithm provided an concrete instantiation of the notion that quantum computer could be far more powerful than classical computers on real-world problems. The faith in this idea led to a drastic growth in the theoretical and experimental progress of the quantum computation.

iii) Shor's algorithm has to some degree directed the attention of the subject of cryptography, rooted in number theory and abstract algebra [25, 99] towards quantum physics. Hence, broader development of quantum information science has, in some part, been predicated on cryptographic aims.

Before stating the computational steps of Shor's algorithm, we aim to discuss two of its subroutines. The first is the implementation of a quantum version of a discrete Fourier transform. Suppose we have a quantum computer represented by a Hilbert space with an orthonormal basis $|0\rangle, \ldots, |N-1\rangle$. Then the action

of the quantum Fourier transform on arbitrary state of the computer is given by

$$\sum_{j=0}^{N-1} x_j |j\rangle \rightarrow \sum_{k=0}^{N-1} y_k |k\rangle , \tag{4.130}$$

where

$$y_k \equiv \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} x_j e^{2\pi ijk/N}. \tag{4.131}$$

The quantum information y_k is the discrete Fourier transform of the quantum information x_j. The quantum Fourier transform can be expressed in terms of a sequence of qubit gates and can be shown to be unitary. For the case of a single basis state and where $N = 2^n$, the action of the quantum Fourier transform can be written as

$$|j\rangle \rightarrow \frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} e^{2\pi ijk/2^n} |k\rangle . \tag{4.132}$$

We adopt the following two notations: For a state $|j\rangle$, we express it in terms of binary representation $j = j_1 j_2 \dots j_n$ meaning $j = j_1 2^{n-1} + j_2 2^{n-2} + \dots j_n 2^0$; we also use the notation $0.j_l j_{l+1} \dots j_m$ to represent $j_l/2 + j_{l+1}/4 + \dots j_m/2^{m-l+1}$.

Hence, we can expand the output in (4.132) as

$$|j\rangle \rightarrow \frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} e^{2\pi ijk/2^n} |k\rangle \tag{4.133}$$

$$= \frac{1}{2^{n/2}} \sum_{k_1=0}^{1} \cdots \sum_{k_n=0}^{1} e^{2\pi ij(\sum_{l=1}^{n} k_l 2^{-l})} |k_1 \dots k_n\rangle \tag{4.134}$$

$$= \frac{1}{2^{n/2}} \sum_{k_1=0}^{1} \cdots \sum_{k_n=0}^{1} \bigotimes_{l=1}^{n} e^{2\pi ijk_l 2^{-l}} |k_l\rangle \tag{4.135}$$

$$= \frac{1}{2^{n/2}} \bigotimes_{l=1}^{n} \left(\sum_{k_l=0}^{1} e^{2\pi ijk_l 2^{-l}} |k_l\rangle \right) \tag{4.136}$$

$$= \frac{1}{2^{n/2}} \bigotimes_{l=1}^{n} \left(|0\rangle + e^{2\pi ij2^{-l}} |1\rangle \right) \tag{4.137}$$

$$\tag{4.138}$$

This can be expanded into what is known as the product representation of the quantum Fourier transform

$$\frac{(|0\rangle + e^{2\pi i 0.j_n} |1\rangle)(|0\rangle + e^{2\pi i 0.j_{n-1}j_n} |1\rangle) \cdots (|0\rangle + e^{2\pi i 0.j_1 j_2 \ldots j_n} |1\rangle)}{2^{n/2}}. \tag{4.139}$$

The second subroutine in Shor's algorithm is known as phase estimation; the computational task is that given a unitary operator U with an eigenvector $|u\rangle$, find the unknown value φ in the corresponding eigenvalue $e^{2\pi i\varphi}$. For simplicity, assume that φ can be written in t bits as $\varphi = 0.\varphi_1 \ldots \varphi_t$. The quantum computer starts with in the state

$$|0\rangle^{\otimes t} |u\rangle , \tag{4.140}$$

where the first register contains t qubits in state $|0\rangle$ and the second register contains the eigenvector. We proceed to apply the Hadamard transform to the first register to obtain

$$\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)^{\otimes t} |u\rangle = \frac{1}{2^{t/2}} \sum_{j=0}^{2^t-1} |j\rangle |u\rangle . \tag{4.141}$$

Recall that $U |u\rangle = e^{2\pi i\varphi} |u\rangle$. This implies that when we apply a controlled-U operation on the second register with U raised to successive powers of two, the state results in

$$\frac{(|0\rangle + e^{2\pi i 0.\varphi_t} |1\rangle)(|0\rangle + e^{2\pi i 0.\varphi_{t-1}\varphi_t} |1\rangle) \cdots (|0\rangle + e^{2\pi i 0.\varphi_1 \varphi_2 \ldots \varphi_t} |1\rangle)}{2^{t/2}} |u\rangle . \tag{4.142}$$

After this step, we apply the inverse of the quantum Fourier transform (4.139) to obtain the desired output $|\varphi_1 \ldots \varphi_t\rangle$.

b) Shor's algorithm: The classical computational problem is to find the prime factorization of an integer N. This problem is equivalent to the order-finding problem which can be described as follows. Suppose x and N are positive integers with no common factors and where $x < N$. The order of x modulo N is the smallest positive integer, r, such that

$$x^r = 1 \,(\text{mod } N). \tag{4.143}$$

The order-finding problem is that given x and N, determine r. Showing the math-

ematical equivalence of these two problems is beyond the scope of this thesis. Assuming this equivalence, Shor's algorithm can be seen as a classical algorithm with a quantum subroutine for order-finding. Every step in the following algorithm can be performed efficiently on a classical computer except the quantum subroutine. Over the course of repeating the algorithm, the complete prime factorization of N can be computed.

The first step of the algorithm is check if N is even, and if so return the factor 2. The second step is to use a known classical algorithm determine whether $N = a^b$ for integers $a \geq 1$ and $b \geq 2$, and if so return the factor a. The third step is to randomly choose an x in the range 1 to $N - 1$. If $\gcd(x, N) > 1$, then return the factor $\gcd(x, N)$. The fourth step is the quantum order-finding subroutine to derive the order r of x modulo N. The last step is if r is even and $x^{r/2} \neq -1(\mathrm{mod} N)$, then compute both $\gcd(x^{r/2} - 1, N)$ and $\gcd(x^{r/2} + 1, N)$; check if any one of these is a factor and output that factor; otherwise the algorithm fails.

Shor's algorithm crucially depends on the quantum subroutine for order-finding, which we now describe. We encode the order-finding problem into the quantum computer as unitary operator,

$$U\left|y\right\rangle \equiv \left|xy(\mathrm{mod\ N})\right\rangle, \tag{4.144}$$

where $y \in \{0, 1\}^L$. The eigenvectors of U can expressed as

$$\left|u_s\right\rangle \equiv \frac{1}{\sqrt{r}} \sum_{k=0}^{r-1} \exp\left(\frac{-2\pi isk}{r}\right) \left|x^k \mathrm{mod}\, N\right\rangle, \tag{4.145}$$

for integer $0 \leq s \leq r - 1$. Then the eigenvalues can be written in the following equation as

$$U\left|u_s\right\rangle = \exp\left(\frac{2\pi is}{r}\right) \left|u_s\right\rangle. \tag{4.146}$$

After the encoding, we can use phase estimation to obtain s/r in the eigenvalues $\exp(2\pi is/r)$. After that, we can use a procedure known as the continued fractions algorithm to efficiently obtain the order r.

c) Analysis of algorithm: The best known classical algorithm for the task of

prime factorization of an n-bit integer is the number field sieve which requires $\exp(\Theta(n^{1/3}\log^{2/3} n))$ operations. Shor's algorithm is exponentially faster than this as it can be shown that it performs the same task in $O(n^2 \log n \log\log n)$ operations. The concrete output at the end of this algorithm makes the fundamental question of what is quantum information unavoidable to easily dismiss; it is natural to ask, in this case, where do the quantum computations in Shor's algorithm physically take place [100]? In this way, quantum information science can be seen to provide a resurgence on the historical inquiry [29] regarding the fundamental nature of quantum physics as a whole.

4.3.3 Quantum machine learning

a) Preliminaries: Machine learning [101] is a relatively new field of computer science with the goal to significantly advance artificial intelligence. The central tenet of the field is that computational machines can 'learn' from large data sets to perform tasks (traditionally assigned to only humans) as opposed to being explicitly programmed to do so. The area has found many real-world applications, as well as exhibiting its progress in the domain of games; recently [102, 103] a machine learning system defeated the world champion in the game of Go. Machine learning can be crudely separated into supervised and unsupervised learning. In the former, each piece of data with a corresponding labelled category, collectively known as the training set, is provided to the machine; using this, the machine is supposed to carry out the task of correctly labelling data that exists outside of the training set. This is in contrast to unsupervised learning where the training set does not contain any categories; rather the machine is supposed to find natural categories in which the data could be indexed under; furthermore the machine is then tasked with classifying data outside of that training set.

Within quantum information science, there has been an effort to investigate whether quantum computers could outperform classical computers to implement machine learning [104, 105, 106]. One prominent example of this is in relation to a supervised learning algorithm known as a support vector machine. A quantum support vector machine was designed in [107] with a drastic improvement over the classical case. Central to that work as well as many other quantum ma-

chine learning algorithms is the HHL algorithm. This is a quantum algorithm that efficiently performs matrix inversion on data such as a training set.

b) HHL algorithm: The classical computational problem is that given $N \times N$ complex matrix A and a vector $\vec{b} \in \mathbb{C}^n$, solve for $\vec{x} \in \mathbb{C}^n$ in the equation $A\vec{x} = \vec{b}$. In other words, derive A^{-1} to compute $\vec{x} = A^{-1}\vec{b}$. For our discussion, assume that A is also Hermitian. However, this can easily be generalized if we let

$$C = \begin{pmatrix} 0 & A \\ A^\dagger & 0 \end{pmatrix}, \tag{4.147}$$

where one can solve the equation

$$C\vec{y} = \begin{pmatrix} \vec{b} \\ 0 \end{pmatrix}. \tag{4.148}$$

This results in the solution

$$\vec{y} = \begin{pmatrix} 0 \\ \vec{x} \end{pmatrix}. \tag{4.149}$$

Of noteworthy importance is result that matrix A has an eigenvalue λ if and only if A^{-1} has eigenvalue λ^{-1}. Hence, if A is diagonalizable then computing the inverse of the eigenvalues allows us to construct A^{-1} in an analogous way.

We want to encode this classical problem onto a quantum computer. Let A be the corresponding Hermitian operator with eigenbasis $|u_j\rangle$ with corresponding eigenvalues λ_j. Moreover, we encode the N variable vector $\vec{b}$ into a quantum state,

$$|b\rangle = \sum_{i=1}^{N} b_i |i\rangle, \tag{4.150}$$

using $\log_2 N$ qubits. Our goal is to construct

$$|x\rangle = A^{-1} |b\rangle, \tag{4.151}$$

where $|x\rangle$ encodes the solution $\vec{x}$ over $\log_2 N$ qubits.

The first step of the algorithm is to use a version of phase estimation to decompose the state $|b\rangle$ into the eigenbasis of A and obtain the eigenvalues of A.

Roughly this amounts to applying the unitary operator e^{iAt} on state $|b\rangle$ for a superposition of different times t. After this phase estimation stage, we can represent $|b\rangle$ as

$$|b\rangle = \sum_{j=1}^{N} \beta_j |u_j\rangle , \tag{4.152}$$

and the total state can be, informally, written as

$$\sum_{j=1}^{N} \beta_j |u_j\rangle |\lambda_j\rangle . \tag{4.153}$$

For a more precise description of this step, we introduce the state

$$|\Psi_0\rangle := \sqrt{\frac{2}{T}} \sum_{\tau=0}^{T-1} \sin \frac{\pi(\tau + \frac{1}{2})}{T} |\tau\rangle , \tag{4.154}$$

for some large T; this is to minimize a quadratic loss function which we will not concern us here. Hence we can express the process discussed more accurately as

$$\sum_{\tau=0}^{T-1} |\tau\rangle \langle\tau| \otimes e^{\frac{iA\tau t_0}{T}} (|\Psi_0\rangle \otimes |b\rangle), \tag{4.155}$$

where t_0 is dependent on the condition number (ratio between A's largest and smallest eigenvalues) and the additive error achieved in output state $|x\rangle$. This is followed by a Fourier transform on the first register which gives the state

$$\sum_{j=1}^{N} \sum_{k=0}^{T-1} \alpha_{k|j} \beta_j |k\rangle |u_j\rangle , \tag{4.156}$$

where states $|k\rangle$ represent the Fourier basis states; the value $|\alpha_{k|j}|$ is large if and only if $\lambda_j \approx (2\pi k)/(t_0)$. We proceed to define $\bar{\lambda_k} \equiv (2\pi k)/(t_0)$ and re-express our $|k\rangle$ register as

$$\sum_{j=1}^{N} \sum_{k=0}^{T-1} \alpha_{k|j} \beta_j |\bar{\lambda_k}\rangle |u_j\rangle \tag{4.157}$$

The second step of the algorithm is acquire the inverse of the eigenvalues into the quantum information; this is the critical step as it allows us construct A^{-1}. This

is roughly accomplished by performing a linear map taking state $|\lambda_j\rangle$ in (4.153) to

$$C\lambda_j^{-1} |\lambda_j\rangle , \tag{4.158}$$

where C is some normalization constant. A more precise description of this step can be described by adding an extra qubit in state $|0\rangle$ to (4.157). This extra qubit will be rotated conditioned on state $|\bar{\lambda_k}\rangle$ to produce

$$\sum_{j=1}^{N} \sum_{k=0}^{T-1} \alpha_{k|j} \beta_j |\bar{\lambda_k}\rangle |u_j\rangle \left(\sqrt{1 - \frac{C^2}{\bar{\lambda_k^2}}} |0\rangle + \frac{C}{\bar{\lambda_k}} |1\rangle \right), \tag{4.159}$$

where C is chosen based on the condition number of the matrix. This procedure is not unitary so it does have a probability of failure.

The third step of the algorithm is to uncompute the $|\lambda_j\rangle$ register in (4.158) and the quantum computer outputs a state proportional to

$$\sum_{j=1}^{N} \beta_j \lambda_j^{-1} |u_j\rangle = A^{-1} |b\rangle = |x\rangle . \tag{4.160}$$

The more precise description, of this third step, following (4.159) is to undo phase estimation to uncompute $|\bar{\lambda_k}\rangle$. For the case where phase estimation is perfect, we have the value $\alpha_{k|j} = 1$ if $\bar{\lambda_k} = \lambda_j$, and 0 otherwise. Supposing this case, we can write the resulting state

$$\sum_{j=1}^{N} \beta_j |u_j\rangle \left(\sqrt{1 - \frac{C^2}{\lambda_j^2}} |0\rangle + \frac{C}{\lambda_j} |1\rangle \right), \tag{4.161}$$

and from there measure the last qubit. Conditioned on seeing the result 1, we have the final state

$$\sqrt{\frac{1}{\sum_{j=1}^{N} C^2 |\beta_j|^2 / |\lambda_j|^2}} \sum_{j=1}^{N} \beta_j \frac{C}{\lambda_j} |u_j\rangle , \tag{4.162}$$

which corresponds to the state

$$|x\rangle = \sum_{j=1}^{n} \beta_j \lambda_j^{-1} |u_j\rangle \,, \tag{4.163}$$

up to a normalization.

The final step is that we can make measurement M which provides us with expectation value $\langle x| M |x\rangle$; this could be used to estimate features of $\vec{x}$ that we may be interested in.

c) Analysis of algorithm: The best known classical algorithm for the task of finding $\vec{x}$ is $O(N \log N)$, where N is the number of variables. By contrast the HHL quantum algorithm is exponentially better in that it requires takes $O((\log N)^2)$ steps to find $|x\rangle$. However, due to the hidden nature of quantum information, we can only output some expectation value $\langle x|M|x\rangle$ for a measurement M, rather than $|x\rangle$ itself. Nevertheless, the broad applicability of quantum information to develop artificial intelligence is proving to be a promising area of research. Perhaps more exciting are the investigations on whether quantum information could form the direct basis for biological intelligence [108, 109, 110].

4.4 Entanglement in Time

To introduce an entanglement in time, it is perhaps useful to consider the properties of an entanglement in space. The latter entanglement involves an interdependence of quantum information systems across a spatial distance. This trivially implies that the systems cannot be in the same location. In an entanglement in time, the interdependence of quantum information systems is across a temporal interval. In its strictest form, this implies that the entanglement is between systems that do not coexist! Rather remarkably, such an entanglement has recently been experimentally realized [111, 112, 113].

Our aim is to describe this entanglement in time. Moreover, we will highlight examples of it through a temporal Bell state, a temporal GHZ state and a temporal graph state. Despite the experimental realization of such states, the role of this

temporal effect in quantum information science is largely unknown.

We aim to provide some insight by arguing for the following overarching theme that was derived solely on the basis of examining the collected literature. *The interdependence in any entanglement in space is shocking due to the absense of a time interval involved. The interdependence in any entanglement in time is shocking due to the existence of a time interval involved.* The latter statement is rather non-trivial as an interdependence across time exists for even classical information systems through a causal dependence. Nevertheless, we will articulate that the interdependence across time for this entanglement is stronger than could ever exist between classical information systems.

We proceed to describe the experimental realization of this temporal effect through the theoretical tools of qubits and density operators. For a general survey on the experimental procedures involving temporal quantum information systems, we refer reader to [114].

4.4.1 Preliminaries

a) Entanglement swapping: To generate an entanglement in time, between subsystems that do not coexist, one uses a modification of an entanglement in space procedure known as entanglement swapping [115]. The spatial entanglement is essentially transferred from a composite system to a different composite system. This procedure can also be viewed as a teleportation protocol for a spatially entangled state [116]. It has been generalized to multiple swappings [117], and entropic analysis of the procedure can be found in [50]. It also allows one to extend the range of quantum communication networks through repeaters [118]. Of noteworthy importance is a delayed choice version of this procedure [119], which has recently been experimentally realized [120]. In this delayed choicce experiment, the choice to transfer the entanglement to a desired composite system is made after the desired system has already been measured. This results in a portrayal of the total system exhibiting retrocausality (the influence of future actions on past events). However all the subsystems in this procedure coexisted; the entanglement in time that we will describe between subsystems that do not

coexist will prove to be far more bizarre. For a broad overview on delayed choice experiments, we refer the reader to [121].

b) Quantum optics: To mathematically describe the optical experimental generation of this entanglement in time, we provide a brief 'dictionary' that explains the experimental physics in terms of the theoretical physics:

i) A photon with horizontal (h) and vertical (v) polarization states can be seen as a physical instantiation of a qubit,

$$|\psi\rangle = \alpha\,|h\rangle + \beta\,|v\rangle\,. \tag{4.164}$$

where $|\alpha|^2 + |\beta|^2 = 1$. The left and right circularly polarized states are respectively,

$$|l\rangle \equiv \frac{1}{\sqrt{2}}(|h\rangle + i\,|v\rangle), \tag{4.165}$$

$$|r\rangle \equiv \frac{1}{\sqrt{2}}(|h\rangle - i\,|v\rangle). \tag{4.166}$$

ii) Wave plates are devices that mathematically transform (4.164) to

$$|\psi\rangle = \alpha\,|h\rangle + e^{i\phi}\beta\,|v\rangle\,, \tag{4.167}$$

where $\phi = \pi$ represents a half-wave plate (HWP) and $\phi = \pi/2$ represents a quarter-wave plate (QWP).

iii) A beam splitter (BS) transforms the two spatially separated inputs $|T\rangle$ and $|B\rangle$ (which refer to top and bottom respectively) to two spatially separated outputs

$$|T\rangle \rightarrow \frac{1}{\sqrt{2}}(|T\rangle + |B\rangle), \tag{4.168}$$

$$|B\rangle \rightarrow \frac{1}{\sqrt{2}}(|T\rangle - |B\rangle). \tag{4.169}$$

iv) A polarizing beam splitter (PBS) directs the $|h\rangle$ photons in one direction and directs the $|v\rangle$ photons in another direction. After the PBS, it is typical that there are two detectors after that measure the photons of the two

different polarizations.

v) Spontaneous parametric down-conversion (SPDC) is a method where a photon with frequence v is converted to two photons with respective frequencies v_1 and v_2 such that $v = v_1 + v_2$.

vi) Post-selection is also known as conditional detection. The process of SPDC is not certain to happen given that the photon detectors do not have perfect efficiency. Hence the event is recorded only if two photons are detected.

4.4.2 Through qubits

a) Temporal bipartite systems: The aim is to manifest the intended entanglement in time through a bipartite system. More precisely, the experiment [111] generates a temporal version of a Bell state between a pair of photons that do not coexist. Our review of the experiment begins by considering a PDC to create polarized photons in any of the four spatial Bell states

$$|\phi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|h_a h_b\rangle \pm |v_a v_b\rangle), \tag{4.170}$$

$$|\psi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|h_a v_b\rangle \pm |v_a h_b\rangle). \tag{4.171}$$

The labels h and v represent the respective polarization states, and the spatial modes are denoted a or b. These can simply be viewed as a physical instantiation of the Bell states described in (4.9) and (4.10).

In order to observe the intended effect, the experiment generates two pairs of photons (1-2 and 3-4) separated by a well-defined time interval τ. Hence, there are a total of four photons across a span of time. The quantum state of such a system can be described as

$$|\psi^-\rangle^{0,0}_{a,b} \otimes |\psi^-\rangle^{\tau,\tau}_{a,b} = \frac{1}{2}(|h^0_a v^0_b\rangle - |v^0_a h^0_b\rangle) \otimes (|h^\tau_a v^\tau_b\rangle - |v^\tau_a h^\tau_b\rangle). \tag{4.172}$$

In this case, the spatial modes are located in the subscripts and the time labels of the photons are in the superscripts.

The aim is to perform a Bell state projection on the second photon of the first

pair and the first photon of the second pair. To achieve this the former particle is delayed in a delay line. The same delay is also used on the second photon of the second pair. The resulting state can be reordered and written as

$$|\psi^-\rangle_{a,b}^{0,\tau}\,|\psi^-\rangle_{a,b}^{\tau,2\tau} = \frac{1}{2}(|\psi^+\rangle_{a,b}^{0,2\tau}\,|\psi^+\rangle_{a,b}^{\tau,\tau} - |\psi^-\rangle_{a,b}^{0,2\tau}\,|\psi^-\rangle_{a,b}^{\tau,\tau} - |\phi^+\rangle_{a,b}^{0,2\tau}\,|\phi^+\rangle_{a,b}^{\tau,\tau} + |\phi^-\rangle_{a,b}^{0,2\tau}\,|\phi^-\rangle_{a,b}^{\tau,\tau}). \tag{4.173}$$

We now describe the explicit sequence of measurements undertaken by the experimentalists. They measure the first photon of the first pair (1) immediately after it is created, while the second photon of that pair (2) is delayed by temporal interval τ in a free-space delay line. The length of the delay line is chosen so that there is adequate time for the measurement of the first photon (1) before the second pair of photons is created (3-4).

After the second pair (3-4) of photons is generated, the first photon of that pair (3) is projected onto a Bell state with the delayed photon of the first pair (2). The last photon (4), which is the second photon of the second pair, is delayed by an interval τ through the same delay line. Moreoever, the last photon (4) is measured only after that delay period.

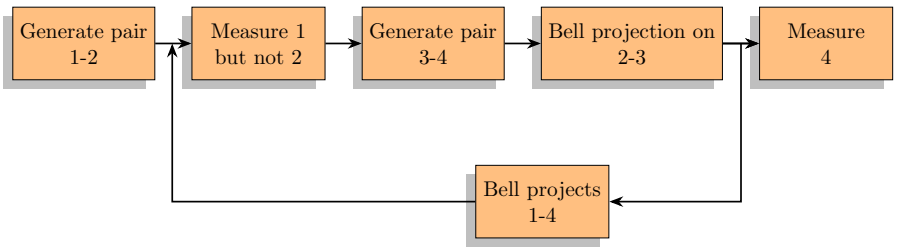

Figure 4.1: Time is increasing to the right

When the photons at time τ (2-3) are projected onto any Bell state, the first and last photons (1-4), which share no prior interdependence, also collapse into the same state and the entanglement is 'swapped.' The first photon (1) and the last photon (4) become entangled. It is important to emphasize that the first photon (1) was measured before the last photon (4) was even created. We can write these

temporal Bell states more explicity as

$$|\phi^{\pm}\rangle_{a,b}^{0,2\tau} = \frac{1}{\sqrt{2}}(|h_a^0 h_b^{2\tau}\rangle \pm |v_a^0 v_b^{2\tau}\rangle), \tag{4.174}$$

$$|\psi^{\pm}\rangle_{a,b}^{0,2\tau} = \frac{1}{\sqrt{2}}(|h_a^0 v_b^{2\tau}\rangle \pm |v_a^0 h_b^{2\tau}\rangle). \tag{4.175}$$

This is an *entanglement in time* between subsystems that do not coexist. The mathematical description as in (4.174) describe the state in terms of the particle that existed just prior to $t = 0$ and the particle that existed just prior to $t = 2\tau$; we emphasize that the particle at $t = 0$ did not coexist with the particle at $t = 2\tau$.

We see that the time at which each photon is measured has no effect on the final outcome. Hence the timing of each photon simply serves an additional label to differentiate between the various photons. From the view of this thesis, it can be alternatively described as that the quantum information (held in these photons) is labelled in both space (a and b) and time (0 and 2τ).

However, a Bell state measurement using linear optical elements can only simultaneously discriminate between two of the four Bell states. Relaying this restriction back to the experiment, we can describe the procedure in more detail as follows. The delayed photon of the first pair (2) and the first photon of the second pair (3) are projected onto a Bell state by combining them at a PBS. Moreover, postselection is carried out in that the photons must exit the PBS at different ports and that the photons must be indistinguishable. Conditioned on that requirement, the photons are rotated by HWPs to the polarization basis $|p/m\rangle = 1/\sqrt{2}(|h\rangle \pm |v\rangle)$. If the the polarization of the middle photons (2-3) were measured to be correlated (hh or vv), then they were projected onto a $|\phi^+\rangle_{a,b}^{\tau,\tau}$ state. This results in the first and last photon being entangled through the temporal Bell state $|\phi^+\rangle_{a,b}^{0,2\tau}$. However, if the polarization of the middle photons (2-3) were anti-correlated (hv or vh), then they were were projected onto the $|\phi^-\rangle_{a,b}^{\tau,\tau}$ state. In an analogous manner, this resulted in the first and last photon being entangled through the temporal Bell state $|\phi^-\rangle_{a,b}^{0,2\tau}$. In closing, this experimental procedure provided a means of generating temporal Bell states between subsystems that do not coexist.

b) Temporal multipartite systems: We will now describe the generation of temporal multipartite entangled systems whose subsystems existed at different times. This was experimentally realized in [112]. To describe the generation of such an effect, first consider two spatially entangled photons in the following Bell state

$$|\phi_{12}^{+}\rangle = \frac{1}{\sqrt{2}}(|h_1 h_2\rangle + |v_1 v_2\rangle). \tag{4.176}$$

Using two pairs of such states

$$|\phi_{12}^{+}\rangle \otimes |\phi_{34}^{+}\rangle = \frac{1}{2}(|h_1 h_2\rangle + |v_1 v_2\rangle) \otimes (|h_3 h_4\rangle + |v_3 v_4\rangle), \tag{4.177}$$

one can fuse these states, using a PBS, into a four-photon spatial GHZ state

$$|\Psi_{GHZ}^{(4)}\rangle = \frac{1}{\sqrt{2}}(|h_1 h_2 h_3 h_4\rangle + |v_1 v_2 v_3 v_4\rangle). \tag{4.178}$$

This state can be viewed as a specific physical instantiation of (4.19). One can fuse further entangled photon pairs to create a growing GHZ state.

In the experimental work in [112], they were able to realize a temporal version of the above GHZ state using the same experimental setup that used in the temporal bipartite case [111]. Once again, they considered pairs of photons in spatial Bell states generated at consecutive intervals of time. The first photon of a pair was directed to a PBS whereas the second photon enters a delay line of time τ. This second photon met the first photon of the next pair which was then fused at the PBS. Using post-selection, this projected the two entangled pairs onto a temporal four-photon GHZ state. This resulted in an entanglement between four photons that were across different spatial modes and existed at different times! This *entanglement in time* can be mathematically written as

$$\boxed{|\Psi_{GHZ}^{(4)}\rangle = \frac{1}{\sqrt{2}}(|h_1^0, h_2^\tau h_1^\tau h_{2'}^{2\tau}\rangle + |v_1^0, v_2^\tau v_1^\tau v_{2'}^{2\tau}\rangle).} \tag{4.179}$$

From this, we see that there were two spatial modes (1 and 2) after the projecting PBS and $1'$ and $2'$ before the projecting PBS) and three temporal modes (0, τ, 2τ). Contrasting this equation to the spatial case (4.178), the different spatial

modes that would exist for the different photons are replaced in the temporal case by different time slots for only two spatial modes. In principle, one can create larger GHZ states using this technique, solving many of the scalability problems encountered in alternative experimental setups. The most general *temporal GHZ state*, from n photon pairs, would take the form

$$|\Psi_{GHZ}^{(2n)}\rangle = \frac{1}{\sqrt{2}}(|h_1^0, h_2^\tau h_1^\tau \cdots h_2^{(n-1)\tau} h_1^{(n-1)\tau} h_{2'}^{n\tau}\rangle + |v_1^0, v_2^\tau v_1^\tau \cdots v_2^{(n-1)\tau} v_1^{(n-1)\tau} v_{2'}^{n\tau}\rangle). \tag{4.180}$$

The experimental procedure was also able to generate temporal graph states, using polarization rotations on the respective photons. One such example was a six-photon *temporal graph state* with an 'H-shape'

$$|\Psi_H^{(6)}\rangle = \frac{1}{\sqrt{2}}(|h_1^0, h_2^\tau h_1^\tau h_2^{2\tau} h_1^{2\tau} h_{2'}^{3\tau}\rangle + |h_1^0, h_2^\tau h_1^\tau v_2^{2\tau} v_1^{2\tau} v_{2'}^{3\tau}\rangle + |v_1^0, v_2^\tau v_1^\tau h_2^{2\tau} h_1^{2\tau} h_{2'}^{3\tau}\rangle - |v_1^0, v_2^\tau v_1^\tau v_2^{2\tau} v_1^{2\tau} v_{2'}^{3\tau}\rangle). \tag{4.181}$$

There have been other experiments that have generated an entanglement in time (given this definition); one remarkable achievement was generating a time-multiplexed cluster state containing more than 10, 000 entangled modes [122].

4.4.3 Through density operators

a) Temporal bipartite systems: We saw the generation of temporal Bell states (4.174) in the experimental setup described in [111]. An example of such a state was $|\phi^+\rangle_{a,b}^{0,2\tau}$ which is just one instantiation of the entanglement between the first and last photon of the experiment. The density operator framework allows us to articulate this entanglement in an alternative way.

The experimentalists constructed a density matrix to characterize this temporal bipartite system (4.174). More precisely, the density matrix of the first and last photons was constructed, conditioned on the measurement outcome of the pro-

jection of the two photons at time τ. This was accomplished using a modification of quantum state tomography (3.103). Moreoever, the modified tomography required projection measurements that used states such as $|hv\rangle$ as well as states such as $|lr\rangle$ described in (4.165). For the experimental details, refer to [111].

Using such a density matrix, they were able to detect and measure entanglement. This was accomplished a posteriori, meaning only after the measurement of all the photons in the experiment. We briefly list only some of the experimental results of the measured matrices:

i) The fidelity (3.109) between the measured and theoretical density operators were $(77 \pm 1)\%$. Entanglement is said to be demonstrated when the fidelity exceeds 50%.
ii) The CHSH (4.28) value was 2.04 ± 0.04 which was a marginal violation to demonstrate entanglement.
iii) The PPT criterion (4.47) was -0.28 ± 0.01. This aligns with the result as a negative value is needed to demonstrate entanglement.
iv) The concurrence (4.83) was 0.57 ± 0.03. This once again demonstrates entanglement as the number needs to be positive for such an effect.

Though the two photons in a temporal Bell state do not coexist, their quantum state is entangled. This is experimentally expressed through the measured density matrix of the two photons conditioned on the result of the Bell state projection measurement concerning (4.173).

However, if the Bell state projection is not carried out perfectly on the photons at time τ (e.g. indistinguishability is introduced into the projected photons), then it can be measured that the first and last photon do not become entangled. Rather they share classical correlations. This is an important observation as it emphasizes that prior to the Bell projection on the middle photons, the first and last photon do not somehow share any entanglement.

b) Temporal multipartite systems: Our aim is to describe the work in [113] which uses the density operator formalism to efficiently characterize temporal GHZ states (4.180). Moreover, we focus on the theoretical aspects. As described earlier, these temporal multipartite states were experimentally generated in [112]. We recall the four photon case (4.179) whose generation involved two

spatial Bell states with a delay line followed by a fusion at the PBS. This process can be expressed with alternative initial Bell states as

$$\begin{aligned} |\psi+\rangle_{a,b}^{0,0} \otimes |\psi+\rangle_{a,b}^{\tau,\tau} &\xrightarrow{delay} |\psi+\rangle_{a,b}^{0,\tau} \otimes |\psi+\rangle_{a,b}^{\tau,2\tau} \\ &= \frac{1}{2}(|h_a^0 v_b^\tau\rangle + |v_a^0 h_b^\tau\rangle) \otimes (|h_a^\tau v_b^{2\tau}\rangle + |v_a^\tau h_b^{2\tau}\rangle) \\ &\xrightarrow{PBS} \frac{1}{2}(|h_a^0 v_b^\tau v_a^\tau h_b^{2\tau}\rangle + |v_a^0 h_b^\tau h_a^\tau v_b^{2\tau}\rangle) = |GHZ\rangle_{1,2,3,4}\,. \end{aligned} \tag{4.182}$$

We adopt the labels 1, 2, 3, 4 to allow for a more compact notation. The final four photon state can be re-expressed in terms of a density operator

$$\rho_{1,2,3,4} = E(\rho_{1,2} \otimes \rho_{3,4})E^\dagger, \tag{4.183}$$

where $\rho_{i,j}$ is the density matrix of the ith and jth photons, and E is the operator that represents the four-photon entangling process. Recall that at time τ, only the photons 2 and 3 are interacting at the PBS. This implies that we can decompose E as

$$E = \sigma_0^1 F_{2,3} \sigma_0^4, \tag{4.184}$$

where

$$F_{2,3} = (|h_2 h_3\rangle \langle h_2 h_3| + |v_2 v_3\rangle \langle v_2 v_3|) \tag{4.185}$$

$$= \frac{1}{2}(\sigma_0^2 \sigma_0^3 + \sigma_3^2 \sigma_3^3). \tag{4.186}$$

Here σ_0^i and σ_3^i are the identity and Pauli z operator which are applied to the ith photon. The recursive nature of the experimental set up implies that all the entangled pairs originate from the same source and the fusion process operation is also identical. This means that by measuring $\rho_{1,2}$ and $F_{2,3}$, the density matrix of any temporal GHZ state can be computed by combining identical two-photon states with identical projections

$$\begin{aligned} \rho_{1,2,\ldots,n} = \sigma_0^1 F_{2,3} \cdots F_{n-2,n-1} \sigma_0^n \\ (\rho_{1,2} \otimes \cdots \otimes \rho_{n-1,n}) \\ (\sigma_0^1 F_{2,3} \cdots F_{n-2,n-1} \sigma_0^n)^\dagger. \end{aligned} \tag{4.187}$$

One can obtain the entire information about a GHZ state containing any number of photons without getting their full statistics or even observing them. This provides a far more efficient method to characterize the state than standard quantum state tomography (3.104). For the experimental details, refer to [113].

4.4.4 Implications

a) For the theme: We shall now consider how properties of this entanglement, between subsystems that do not coexist, relate to the formulation of the overarching theme of this thesis. Namely that *the interdependence in any entanglement in time is shocking due to the existence of a time interval involved.* For simplicity, we devote our analysis on the bipartite case [111]. Recall for spatial Bell states, a measurement on one of its subsystems instantaneously affects the other spatially distant subsystem. Similarly, an analogous effect exists for temporal Bell states, such as for $|\phi^+\rangle_{a,b}^{0,2\tau}$, where the state represents an entanglement between the first photon and last photon which do not coexist. In [111], this was stated more directly and we quote, "*In the standard entanglement [in space] case, the measurement of any one of the particles instantaneously changes the physical description of the other. This result was described by Einstein as "spooky action at a distance." In the scenario we present here, measuring the last photon affects the physical description of the first photon in the past, before it has even been measured. Thus, the "spooky action" is steering the system's past. Another point of view that one can take is that the measurement of the first photon is immediately steering the future physical description of the last photon. In this case, the action is on the future of a part of the system that has not yet been created.*" These interpretations are shocking because of a time interval between the measurement of the first photon and generation of the last photon. If one were to diminish the time interval to zero, then this entanglement in time loses its perplexing character.

Our thesis aims to provide a finer level of analysis on these matters. It needs to be articulated that the collapses of states are a mathematical description rather than a physical observable. Denying the quantum state a physical reality allows for a pragmatic approach in terms of measurement correlations. The measurement of the first photon yielded a definite outcome, and the measurement of the last

photon is affected by it in a way that is stronger than could ever exist between classical systems. The inability to replicate these correlations using classical systems is quantitatively captured using the CHSH (Bell) inequalities.

However if one assumes that the quantum state has a physical reality, then the interpretation quoted above does indeed highlight that quantum physics allows for influences to propagate into the past. Perhaps this is seen far more concretely in that prior to the Bell state measurements of the photons at time τ (photons 2 and 3), the outcome of the Bell state measurement is unknown. Using linear optical elements, the two possible outcomes are $|\phi^+\rangle_{a,b}^{\tau,\tau}$ or $|\phi^-\rangle_{a,b}^{\tau,\tau}$. This implies that two possible states that the first and last photon can collectively collapse to are either $|\phi^+\rangle_{a,b}^{0,2\tau}$ or $|\phi^-\rangle_{a,b}^{0,2\tau}$. But this is determined only much after the measurement of the first photon (see Figure 4.1)!

b) For relativity: Similar to the spatial case, the individual measurement result of a subsystem in the entanglement in time is probabilistically random. Hence this does not violate causality (in a strict sense). But there may be consequences for approaching the quantum physics of gravitation (the problem of quantum gravity), or at the very least of viewing relativity with an alternative perspective. It is often the case that spacetimes which contain closed timelike curves (CTC) are regarded as pathological. This experimentally verified effect of entanglement in time can be interpreted to exhibit some similar properties to CTCs. Hence, for the pursuit towards quantum gravity, the premature dismissal of such pathological spacetimes may prove to be erroneous.

c) For the nature of quantum physics: As mentioned in Chapter 3, a most fundamental mystery is what is quantum information? With respect to its historical origins, this problem is referred to as the interpretation issue of quantum mechanics. We shall briefly describe how the above effect of entanglement in time gives support to a subset of the proposed interpretations. One of the proposals is known as the transactional interpretation [123] which involves sending signals back in time. In this experiment of entanglement between subsystems that do not coexist, we have so far seen that it possible to some extent to non-classically influence the past. Hence, this effect certainly adds considerable weight to furthering the transactional interpretation or some modification of it.

An alternative interpretation of quantum mechanics is known as the sum over paths approach [124]. One of the insights (gained from my supervisor) is that if the the framework of quantum mechanics is taken seriously along with the sum over paths approach, then it can be seen that an event has multiple histories and that there is no single definite past. The latter statement aligns with the description presented in this entanglement in time where the past is seen to be unsettled with respect to quantum properties. The final interpretation that aligns well with the experiment is the two-state vector formalism [125, 126] where the present is affected by both the past and the future. In fact this was shown [127] to be related to the entangled histories formalism, which was then used to describe an alternative temporal version of a GHZ state.

d) For the nature of time: One of the most fundamental areas in the philosophy of time [128] is in regards to answering the question: Is there more to the world than the present moment? The classification of answers within this particular sub-branch of metaphysics can be crudely categorized in three groups. The first are the "eternalists" who believe that the past, present, and future are real. The second call themselves the "possibilists" who claim that the past and present are real, but the future is not. The remaining category are known as "presentists" who hold the position that only the present is real. From the perspective of modern physics, it can easily be seen that the theories of relativity (which we shall review in Chapter 6) are in conflict, to a large degree, with presentism. However, a defense [129] was put forth that included the dismissal of the relativistic attack since those theories are challenged by quantum physics. Hence it is surprising that in our investigation of entanglement in time, it is precisely a phenomena in quantum physics that shows from the present one can non-classically affect the past or immediately affect the future that has not yet been created. This alludes to denying presentism, and rather taking the eternalists' view in that the past and future are as real as the present.

4.5 Application: Quantum Blockchain

A central aim in the field of quantum information science is the creation of new applications. Both quantum communications and quantum computers are established applications of entanglement in space. An open question in the field was whether entanglement in time is also a resource in quantum information? More concretely, would it enhance the advantage of established applications? Or more astonishingly, would it lead to the development of novel applications and thereby open the door to new areas of research?

There have been various proposals to modify parts of quantum communications and quantum computers to adopt an entanglement in time. For example, it was noted in [111] that a memory system in a communication network would benefit from using this entanglement in time. The entanglement in time which we will review in Chapter 5 has found applications for its use in a communication protocol [130] as well as in the analysis of computing [131]. In Chapter 6, we examine yet another entanglement in time with proposals for memory systems [132], teleportation [133], and quantum key distribution[134].

In this thesis, we make an original contribution (which was done in collaboration with my supervisor) by designing one of the first quantum information applications of entanglement in time, namely a quantum blockchain [135]. Our primary innovation is in encoding the blockchain into a temporal GHZ state. It will be shown that the entanglement in time, as opposed to an entanglement in space, provides the crucial quantum advantage over a classical blockchain. More shockingly, the information encoding procedure in this quantum blockchain can be interpreted as non-classically influencing the past, and hence the system can be viewed as a 'quantum time machine.' Furthermore, all the subsystems of this design have already been shown to be experimentally realizable [111, 112, 113, 136].

However, the scope of our original research into the quantum blockchain is limited to only specifying a conceptual design. This can be seen as the major step before providing a fully detailed design. On a coarse level, there are three phases to the design of a quantum information system. The first phase is to identify and extract the most essential task that characterizes the information system

under consideration. As an example in certain classical cryptographic systems, the most fundamental aspect is to have identical private keys at two different locations. The second phase is to articulate the particular quantum system to be used to represent that information task along with an appropriate encoding method. As an example in the E91 quantum protocol, it was realized that a spatial Bell state was the desirable structure for generating private keys at two different locations. The third phase is to specify the dynamics of the quantum information system. In the E91 protocol this is a clear sequence of steps carried out by Alice and Bob in terms of quantum operations, quantum measurements, classical communication and classical processing to generate that key. For our conceptual design we will concern ourselves with the first two phases and merely provide an outline for how it could be used for developing the third phase.

4.5.1 Preliminaries

a) Quantum computing attacks: Before describing the quantum blockchain, we want to convey that one instantiation of the interplay between blockchains and quantum information arises as a security threat. Quantum computers pose a significant threat to the security features of a classical blockchain, thereby potentially invalidating it as an information security system. We refer the reader to [137] for an in-depth analysis on this topic whose crucial points we briefly outline. In Chapter 2 on classical blockchains, it was emphasized that the certain quantities associated with cryptographic hash functions would be infeasible to classically compute. The security of the system crucially depends on such properties. However, a quantum computer running Grover's search algorithm can perform quadratically fewer computations for this problem than is needed by classical computing. Therefore over time, this vulnerability will pose an imminent threat. The second risk to classical blockchains posed by quantum computers derives from Shor's factoring algorithm. The classical blockchains require public key cryptography for various operations involving digital signatures. Hence Shor's algorithm and its attack on public key cryptography poses a dramatic risk to the classical system.

b) Post-quantum cryptography: One can address the second risk of Shor's

algorithm by developing classical blockchains which replace the public key cryptography with a post-quantum cryptographic component [77, 78]. Recall that these protocols are classical cryptographic systems which utilize mathematical problems that many believe a quantum computer would not be able to solve. However, the durability of these solutions can be questioned in that there are no formal proofs supporting this hypothesis. Nevertheless, these post-quantum blockchains have been proposed [138, 139, 140].

c) Quantum cryptography: In a previous section, we described how quantum key distribution provides an alternative solution towards the risk posed by Shor's algorithm. In [141], a classical blockchain with a quantum key distribution subroutine was proposed. It was also experimentally realized among a small number of network nodes. In addition to this work, classical blockchains with various added quantum features have also been put forward in [142, 143, 144, 145, 146, 147].

d) Design methodology: Recall that a classical blockchain system stores data securely over time and in a decentralized manner. It is composed of two parts, namely the temporal blockchain data structure and a decentralized network consensus algorithm. Our aim is to redesign the classical blockchain system into a full quantum information application to not only protect it from a quantum computing attack, but highlight further superior advantages as an information security system. To do so our conceptual design focuses on creating quantum analogues of the blockchain data structure as well as the network consensus algorithm. However, a number of low level design gaps do exist, but the intention was to open up a novel area where at least the core functionalities are covered.

4.5.2 Quantum data structure

a) Description: In this section, our aim is to replace the data structure component of the classical blockchain with a quantum information system which harnesses an entanglement in time. In the classical case, records are chained in a chronological order through cryptographic hash functions. In the quantum information case, we will capture the notion of the chain through the non-

separability (entanglement) of quantum systems. For a spatially bipartite system $|\psi\rangle_{AB}$, this means that

$$|\psi\rangle_{AB} \neq |a\rangle_A |b\rangle_B , \tag{4.188}$$

for all single qubit states $|a\rangle$ and $|b\rangle$; the subscripts refer to the respective Hilbert spaces. In particular multipartite GHZ states are ones in which all subsystems contribute to the shared entangled property. This enables us to create the concept of a chain. However we need a method to encode the records into the chain, and develop a temporal structure to identify the chronological order.

To create the appropriate code to utilize this chain, it is helpful to use a concept from superdense coding [64]. In this protocol, recall that a code (4.86) converts classical information into spatially entangled Bell states; two classical bits, xy, where $xy = 00, 01, 10$ or 11, are encoded to the state

$$|\beta_{xy}\rangle = \frac{1}{\sqrt{2}}(|0\rangle |y\rangle + (-1)^x |1\rangle |\bar{y}\rangle), \tag{4.189}$$

where $\bar{y}$ is the negation of y. Given that Bell states are orthonormal, they can be distinguished by quantum measurements. This decoding process allows one to extract the classical bit string, xy, from $|\beta_{xy}\rangle$.

We still need a temporal structure to encode the chronological order. This can be accomplished using an entanglement in time rather than a spatial entanglement. For our conceptual design, we temporarily simplify the data characterizing the records in the classical block to a string of two bits. Our encoding procedure converts each block with its classical record, say $r_1 r_2$, into a temporal Bell state , generated at a particular time, say $t = 0$:

$$|\beta_{r_1 r_2}\rangle^{0,\tau} = \frac{1}{\sqrt{2}}(|0^0\rangle |r_2^\tau\rangle + (-1)^{r_1} |1^0\rangle |\bar{r_2}^\tau\rangle). \tag{4.190}$$

From the entanglement in time section, it was seen that the superscripts in the kets signify the time at which the photon is absorbed; notice that the first photon of a block is absorbed immediately. For our purposes, this provides a way to do time stamps for each block.

Recall such temporal Bell states were experimentally generated in the work by

[111] which we described in the last section. In their procedure, spatially entangled qubits were represented through polarized photons,

$$|\phi\pm\rangle = \frac{1}{\sqrt{2}}(|h_a h_b\rangle \pm |v_a v_b\rangle), \quad |\psi\pm\rangle = \frac{1}{\sqrt{2}}(|h_a v_b\rangle \pm |v_a h_b\rangle), \tag{4.191}$$

where h_a (v_a) represent the horizontal (vertical) polarization in spatial mode a (b). To create the temporally entangled states, consecutive pairs of spatially entangled pairs were generated at well-defined times separated by time interval τ:

$$|\psi-\rangle_{a,b}^{0,0} \otimes |\psi-\rangle_{a,b}^{\tau,\tau} = \frac{1}{2}(|h_a^0 v_b^0\rangle - |v_a^0 h_b^0\rangle) \otimes (|h_a^\tau v_b^\tau\rangle - |v_a^\tau h_b^\tau\rangle), \tag{4.192}$$

where the added superscripts provide the time labels for the photons. In the experiment, a delay line of time τ is introduced to one of the photons of each entangled pair. This resulting state equated to

$$\begin{aligned} |\psi-\rangle_{a,b}^{0,\tau} |\psi-\rangle_{a,b}^{\tau,2\tau} = \frac{1}{2}(&|\psi+\rangle_{a,b}^{0,2\tau} |\psi+\rangle_{a,b}^{\tau,\tau} - |\psi-\rangle_{a,b}^{0,2\tau} |\psi-\rangle_{a,b}^{\tau,\tau} \\ &- |\phi+\rangle_{a,b}^{0,2\tau} |\phi+\rangle_{a,b}^{\tau,\tau} + |\phi-\rangle_{a,b}^{0,2\tau} |\phi-\rangle_{a,b}^{\tau,\tau}). \end{aligned} \tag{4.193}$$

When Bell projection was carried out on two photons at time $t = \tau$, entanglement is created between the photon absorbed at $t = 0$ and the photon absorbed at $t = 2\tau$; this is despite the fact that the latter two photons have never coexisted.

Going back to our design, as records are generated, the system encodes them as blocks into temporal Bell states; these photons are then created and absorbed at their respective times. A specific example of such blocks would be:

$$|\beta_{00}\rangle^{0,\tau} = \frac{1}{\sqrt{2}}(|0^0\rangle |0^\tau\rangle + |1^0\rangle |1^\tau\rangle), \tag{4.194}$$

$$|\beta_{10}\rangle^{\tau,2\tau} = \frac{1}{\sqrt{2}}(|0^\tau\rangle |0^{2\tau}\rangle - |1^\tau\rangle |1^{2\tau}\rangle), \tag{4.195}$$

$$|\beta_{11}\rangle^{2\tau,3\tau} = \frac{1}{\sqrt{2}}(|0^{2\tau}\rangle |1^{3\tau}\rangle - |1^{2\tau}\rangle |0^{3\tau}\rangle), \tag{4.196}$$

and so forth. To create the desired quantum design, the system should chain the bit strings of the Bell states together in chronological order, through an entanglement in time. Such a task can be accomplished by using a fusion process [112], described in the last section, in which temporal Bell states are recursively projected into a growing temporal GHZ state. Physically, the fusion process is carried out through the entangled photon-pair source, a delay line and a polarizing beam splitter (PBS). As an example, two Bell states can be fused into the following four-photon GHZ state:

$$\begin{aligned} |\psi+\rangle_{a,b}^{0,0} \otimes |\psi+\rangle_{a,b}^{\tau,\tau} &\xrightarrow{delay} |\psi+\rangle_{a,b}^{0,\tau} \otimes |\psi+\rangle_{a,b}^{\tau,2\tau} \\ &= \frac{1}{2}(|h_a^0 v_b^\tau\rangle + |v_a^0 h_b^\tau\rangle) \otimes (|h_a^\tau v_b^{2\tau}\rangle + |v_a^\tau h_b^{2\tau}\rangle) \\ &\xrightarrow{PBS} \frac{1}{2}(|h_a^0 v_b^\tau v_a^\tau h_b^{2\tau}\rangle + |v_a^0 h_b^\tau h_a^\tau v_b^{2\tau}\rangle) = |GHZ\rangle^{0,\tau,\tau,2\tau} . \end{aligned} \tag{4.197}$$

Recall that in this GHZ state, entanglement exists between the four photons that propagate in different spatial modes and exist at different times. Implementing this procedure in our design, the state of the *quantum blockchain*, at $t = n\tau$ (from $t = 0$) is given by

$$\begin{aligned} &|GHZ_{r_1 r_2 \ldots r_{2n}}\rangle^{0,\tau,\tau,2\tau,2\tau\ldots,(n-1)\tau,(n-1)\tau,n\tau} \\ &= \frac{1}{\sqrt{2}}(|0^0 r_2^\tau r_3^\tau \ldots r_{2n}^{n\tau}\rangle + (-1)^{r_1} |1^0 \bar{r}_2^\tau \bar{r}_3^\tau \ldots \bar{r}_{2n}^{n\tau}\rangle). \end{aligned} \tag{4.198}$$

The subscripts on the LHS of (4.198) denote the concatenated string of all the blocks, while superscripts refer to the time stamps. The time stamps allow each blocks' bit string to be differentiated from the binary representation of the temporal GHZ basis state. Note that at $t = n\tau$, there is only one photon remaining.

The dynamics of this procedure can be illustrated with our example above. Out of the first two blocks, $|\beta_{00}\rangle^{0,\tau}$ and $|\beta_{10}\rangle^{\tau,2\tau}$, the system creates the (small) blockchain,

$$|\beta_{00}\rangle^{0,\tau} \otimes |\beta_{10}\rangle^{\tau,2\tau} \rightarrow |GHZ_{0010}\rangle^{0,\tau,\tau,2\tau} \tag{4.199}$$

Concatenating the third block $|\beta_{11}\rangle^{2\tau,3\tau}$ produces

$$|GHZ_{001011}\rangle^{0,\tau,\tau,2\tau,2\tau,3\tau} = \frac{1}{\sqrt{2}}(|0^0 0^\tau 1^{2\tau} 0^{2\tau} 1^{2\tau} 1^{3\tau}\rangle + |1^0 1^\tau 0^{2\tau} 1^{2\tau} 0^{2\tau} 0^{3\tau}\rangle). \quad (4.200)$$

The decoding process extracts the classical information, $r_1 r_2 \ldots r_{2n}$, from the state (4.198). As mentioned in the previous section, it was shown [113] how to characterize any such temporally generated GHZ state efficiently compared to standard tomography techniques. This can be accomplished without measuring the full photon statistics, or even detecting them.

b) Security analysis: Recall that in the classical blockchain system the relevant performance metric is nontampering for the data structure. This is accomplished by the data structure being extremely sensitive to tampering through the *interdependence of classical blocks achieved by cryptographic hash functions*. If one attempts to modify even a single block, the extreme sensitivity is such that is invalidates all future blocks following the tampered block. This provides a tamper proof system for storing records because tampering with it can easily be detected. In the quantum blockchain, the sensitivity to tampering is achieved through *the interdependence of the quantum blocks in an entanglement in time*.

To elaborate, for the quantum blockchain we have replaced the important functionality of time stamped blocks and hash functions linking them, by a temporal GHZ state with an entanglement in time. The quantum advantage is that the sensitivity towards tampering is significantly amplified, meaning that the blockchain is destroyed if one tampers with a single block (due to entanglement); on a classical blockchain only the blocks after the tampered block are destroyed (due to cryptographic hash functions) which leaves it open to vulnerabilities. For the classical case, it is often stated that the farther back the block was time stamped in, the more "secure" it is; this is precisely because of the above invalidation. Even if we had used an entanglement in space (with all the photons coexisting) that would still have provided an advantage since if an attacker tries to tamper with any photon, the full blockchain would be invalidated immediately; this already provides a benefit over the classical case where only the future blocks of the tampered block are invalidated. The temporal GHZ blockchain (4.198) adds a far greater benefit in that the attacker cannot even attempt to access the previ-

ous photons since they no longer exist. They can at best try to tamper with the last remaining photon, which would invalidate the full state. Hence in this application of quantum information, we see that the entanglement in time provides a far greater security benefit than an entanglement in space. There still needs to be a careful case by case analysis of potential tampering with the ultimately classical measurement results, but that would entail full security proofs which is left for future work.

c) Comments:

i) The temporal GHZ state, that we use in our design, involve an entanglement between photons that do not share simultaneous coexistence, yet they share non-classical interdependence. This temporal interdependence, between two entangled photons that existed at different times, was interpreted in [111] as follows: "*...measuring the last photon affects the physical description of the first photon in the past, before it has even been measured. Thus, the "spooky action" is steering the system's past*". Stated more shockingly, in our quantum blockchain, we can interpret our encoding procedure as linking the current records in a block, not to a record *of* the past, but linking it to the actual record *in* the past, a record which does not exist anymore. Hence the system can be viewed as a 'quantum time machine.'

ii) Much of the performance of the quantum blockchain data structure is simply due to the properties of a temporal GHZ state. The non-trivial aspect was in obtaining the appropriate quantum structure and finding an efficient encoding method. This phase of design is comparable to realizing that a spatial Bell state was a useful structure for key generation in E91

iii) We imagine that future designs of quantum blockchains may harness the other entanglement in time effects discussed in Chapter 5 and Chapter 6.

iv) This conceptual design presented the case that a security advantage exists given that the previously existing photons are not able to be accessed (since they no longer exist). However a full security proof of this remains to be worked out. Of great interest would be whether from an operational point of view there is a security advantage between using photons that do not coexist as opposed to using photons from a single Bell-pair in which a mea-

surement of one particle is simply delayed with respect to the other. Such an understanding may provide the necessary basis for the development of further temporal based quantum information protocols.

4.5.3 Quantum consensus protocol

a) Description: Our aim in this section is to develop a quantum analogue of the network consensus protocol. The θ-protocol [136] was originally designed to verify GHZ entanglement in a quantum network. Given that we have encoded a quantum blockchain into a temporal GHZ state we can harness the θ-protocol as a consensus algorithm for blocks.

To provide some elaboration, recall that a classical blockchain system has a number of different components. A blockchain data structure, a copy of this data structure at each node of a classical network, and a consensus network algorithm to verify the correctness of new blocks (before adding that new block to a blockchain). In our design, we replace the classical network with a quantum network and with that, digital signatures would be covered by a quantum key distribution (QKD) protocol. In fact, others have used this way of reasoning when introducing new quantum protocols. For example in the θ-protocol [136] the authors also simply assume a QKD layer before moving onto their original work. We quote their paper, "*it is assumed that the verifier and each of the parties share a secure private channel for the communication. This can be achieved by using either a one-time pad or a quantum key distribution.*" Furthermore, in this design, each node on the quantum network would host a copy of the quantum blockchain (4.198); hence if a node tampers with its own local copy, it does not affect the copies at the other nodes analogous to the classical case. New blocks (that come from a sender) need to be verified for their correctness, before being copied and added to each node's blockchain. Since correct blocks are GHZ entangled states, one needs a verification test to do it.

At this stage of the design, we assume that newly generated blocks are spatial GHZ states (converting this to the related temporal case is at this stage of the design process unnecessary, and is left for future work). As in the classical case,

the objective is to add valid blocks in a decentralized manner. The challenge is that the network can consist of dishonest nodes, and the generated blocks can come from a dishonest source. To solve this problem, the quantum network uses the θ-protocol [136], which is a consensus algorithm where a random node in the quantum network can verify that the untrusted source created a valid block (ie spatial GHZ state). More crucially, this is accomplished in a decentralized way by using other network nodes, who may also be dishonest (ie Byzantine nodes).

To start off this verification protocol, we need to pick a randomly chosen verifier node (analogous to proof-of-stake or proof-of-work); this can be accomplished through a low level sub-algorithm involving a quantum random number generator. The untrusted source shares a possible valid block, an n-qubit state, ρ. Since it knows the state, it can share as many copies of the block as is needed without running afoul of the no-cloning theorem. For verification, it distributes each of the qubits to each node, j. The verifying node generates random angles $\theta_j \in [0, \pi)$ such that $\sum_j \theta_j$ is a multiple of π. The (classical) angles are distributed to each node, including the verifier. They respectively measure their qubit in the basis,

$$|+_{\theta_j}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{i\theta_j}|1\rangle\right), \tag{4.201}$$

$$|-_{\theta_j}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - e^{i\theta_j}|1\rangle\right). \tag{4.202}$$

The results, $Y_j = \{0, 1\}$, are sent to the verifier. If the n-qubit state was a valid block, ie a spatial n-qubit GHZ state, the necessary condition

$$\oplus_j Y_j = \frac{1}{\pi}\sum_j \theta_j \pmod 2, \tag{4.203}$$

is satisfied with probability 1. The protocol links the verification test to the state that is used; the paper [136] explicitly mentions this and we quote, "*It is important to remark that our verification protocols go beyond merely detecting entanglement; they also link the outcome of the verification tests to the state that is actually used by the honest parties of the network with respect to their ideal target state. This is non-trivial and of great importance in a realistic setting where such resources*

are subsequently used by the parties in distributed computation and communication applications executed over the network." Hence the block can be copied and distributed to each node on the network to be added onto their blockchain.

b) Security analysis: We refer the reader to [136] for an in-depth security analysis whose results we briefly outline. Let $P(\rho)$ denote probability of passing the verification test. Furthermore, let the fidelity of the shared state ρ with respect to an ideal GHZ state be computed as

$$F(\rho) = \langle GHZ_n|\rho|GHZ_n\rangle \,. \tag{4.204}$$

One can obtain a lower bound on the passing the verification test. It can be proven that if the n parties are honest, then we have the relationship,

$$F(\rho) \geq 2P(\rho) - 1. \tag{4.205}$$

When the protocol is performed in the presence of dishonest nodes (byzantine nodes), then the results are modified. Suppose we have $n - k$ nodes that apply local or joint unitary operation U to their state. This encodes the various ways these nodes may attempt to cheat the system. The modified fidelity for this scenario is given by

$$F'(\rho) = \max_U F((I_k \otimes U_{n-k})\rho(I_k \otimes U_{n-k}^\dagger)), \tag{4.206}$$

and the associated lower bound for pass probability can be computed to be

$$F'(\rho) \geq 4P(\rho) - 3. \tag{4.207}$$

Compared to other quantum verification protocols, the θ-protocol can be shown to be more sensitive to detecting dishonest nodes.

c) Comments:

i) Combining the data structure component with the network consensus protocol provides us with the conceptual design of a quantum blockchain.

ii) Our work provided a conceptual design. This is the major step before providing a fully detailed protocol design. The latter is left for future

work. However there are some challenges that we foresee: Standard blockchain protocols do not easily fit into the traditional framework of distributed computing [18] and proof of their security functionalities in a rigorous manner is not well articulated. Hence developing a detailed quantum blockchain protocol with security proofs would be predicated on also undertaking many research problems from the classical case.

iii) Given the rise of classical blockchains and the development of a quantum network, we hope this conceptual design may potentially open the door to a new research frontier in quantum information science.

5

Quantum Foundations

"Quantum information is more like the information in a dream."

– Charles Bennett, co-inventor of quantum teleportation

QUANTUM INFORMATION SCIENCE is based on the framework of quantum theory. In an almost paradoxical manner, quantum theory provides an extraordinary degree of applicability, and yet its fundamental structures remain deeply mysterious. Quantum foundations is a field that is devoted to examining the nature of these structures. Perhaps the two great mysteries are: What is the nature of the quantum state? And how does the quantum state 'collapse' upon measurement? The first question stems as a generalization concerning the unknown physical representation of quantum information. Whereas the second question arises from the unarticulated notion that an undefined observer causes an instant transformation from quantum information to classical information.

In this thesis, we highlight how concepts from the entanglements in both space and time can progress us towards these two questions. To elaborate, we will focus our study on a particular *interdependence* witnessed in the entanglements, known as non-locality. We will see that it a stricter form of non-classical interdependence than entanglement. Our focus in this chapter is to describe non-locality across space as well as across time. We aim to show how these properties can shed at least a partial understanding on the two questions.

5.1 Quantum Measurements

Our focus will initially be on the question of quantum measurement. The approach will involve deriving the condition for non-locality in space, and using it to unravel issues regarding measurements. This procedure requires two points:

i) The probabilistic aspect of quantum theory.
ii) A concept known as realism.

Before moving to non-locality in space, we provide an aid by illuminating results concerning these two points. Gleason's theorem is a result concerning the first point, and the Kochen–Specker theorem elaborates on the second point. For a comprehensive overview on these results, refer to [148].

5.1.1 Gleason's theorem

The probabilistic aspect of quantum theory is conveyed through the measurement postulate such as the Born rule (3.102). Roughly speaking, Gleason's theorem [149] states that if one is given the non-probabilistic structure of quantum theory (e.g. Hilbert spaces, projection operators) and one also assumes that the theory requires a probabilistic character, then that character must be expressed in no other way than the Born rule. An alternative view is that if one requires non-Born rule quantum probabilities, then one must give up using projection operators to describe measurements. In this sense, Gleason's theorem can be interpreted as a 'derivation' of the Born rule. However, it is important to emphasize that the assumption of a probabilistic aspect is still needed and the underlying nature of it is currently unknown. Thus at present, the Born rule cannot be derived solely from the non-probabilistic postulates of quantum theory.

a) Preliminaries: Within quantum foundations as well certain areas of pure mathematics, an extensive investigation of Gleason's theorem has been carried out [150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161]. The theorem addresses the minimal, (in fact, quite surprisingly minimal), assumptions required to deduce the existence of a quantum density matrix, (a unit trace Hermitian matrix encoding the notion of quantum probability), and as mentioned underlies

the theoretical justification for adopting the Born rule. Early proofs of Gleason's theorem were implicit and non-constructive, and for some time there was controversy as to whether a constructive proof was even possible [151, 152, 154, 155]. With hindsight, disagreement on what methods are legitimately to be deemed "constructive" is the key point of the constructivist debate. Even with modern constructive (in principle) proofs, the construction is not particularly explicit, and often very little is said as to what the quantum density matrix actually looks like. Traditionally the analysis stops, and the theorem is complete, once the existence of the quantum density matrix is established.

In this subsection, we will now present work [162] this is *part of the original component of this thesis* (which was done in collaboration with my supervisor). It has very little to say about the theorem and proof themselves, focusing more on the implications: We shall say a little more about the density matrix itself — and shall provide two constructions (one implicit, one explicit) for the density matrix.

b) Gleason's theorem: An explicit statement of the theorem runs thus [149]:

Theorem 5.1. ***(Gleason's theorem)***
Suppose H is a separable[1] *Hilbert space, (either real or complex).*

A measure on H is defined to be a function $v(\cdot)$ that assigns a nonnegative real number to each closed subspace of H in such a way that: If $\{A_i\}$ is any countable collection of mutually orthogonal subspaces of H, and the closed linear span of this collection is B, then $v(B) = \sum_i v(A_i)$. Furthermore we normalize to $v(H) = 1$.

Then if the Hilbert space H has dimension at least three, (either real or complex), every measure $v(\cdot)$ can be written in the form $v(A) = \mathrm{tr}(\rho P_A)$, where ρ is a positive semidefinite trace class operator with $\mathrm{tr}(\rho) = 1$, and P_A is the orthogonal projection onto A. □

(Physicists would almost immediately focus on complex Hilbert spaces; but some of the mathematical literature also works with real Hilbert spaces.) The original theorem gives one very little idea of what the density matrix might look like, and it is this topic we shall address. Indeed, the original theorem spends many pages

[1] A Hilbert space is separable if and only if it has a countable orthonormal basis.

proving that the valuation $v(P)$ uniformly continuous; while this is certainly an extremely useful result, most physicists, (and applied mathematicians for that matter), would simply assume continuity on physical grounds.

c) Elementary observations: Our first observation is that since ρ is Hermitian we can diagonalize it and define

$$\rho = \sum_i \lambda_i \; Q_i. \tag{5.1}$$

Here the Q_i are taken to be 1-dimensional subspaces, and the λ_i are to be repeated with the appropriate multiplicity. Per Gleason's theorem,

$$v(Q_j) = \text{tr}(\rho Q_j) = \text{tr}\left(\left[\sum_i \lambda_i Q_i\right] \; Q_j\right) = \sum_i \lambda_i \; \text{tr}\left(Q_i \; Q_j\right) = \lambda_j. \tag{5.2}$$

So actually

$$\rho = \sum_i v(Q_i) \; Q_i, \tag{5.3}$$

which does not (yet) help unless you can somehow extract the Q_i in terms of the underlying valuation function $v(\cdot)$. Furthermore note that for each 1-dimensional subspace Q_i we can identify

$$Q_i \sim |\psi_i\rangle \; \langle\psi_i| \tag{5.4}$$

where $|\psi_i\rangle$ is any arbitrary vector in the 1-dimensional subspace Q_i. Then

$$v(Q_i) = \langle\psi_i|\rho|\psi_i\rangle. \tag{5.5}$$

Now let P_i be any arbitrary collection of orthogonal 1-dimensional projection operators

$$v\left(\sum_i P_i\right) = \sum_i v(P_i) = 1. \tag{5.6}$$

Using Gleason's theorem, we can calculate

$$v(P_j) = \text{tr}(\rho P_j) = \text{tr}\left(\left[\sum_i v(Q_i) Q_i\right] \; P_j\right) = \sum_i v(Q_i) \; \text{tr}\left(Q_i \; P_j\right) = \sum_i v(Q_i) \; S_{ij}, \tag{5.7}$$

with $S_{ij} = \mathrm{tr}\left(Q_i\, P_j\right)$ a bi-stochastic matrix (which is a square matrix of non-negative real numbers with each row and column summing to unity). That is, Gleason's theorem implies

$$v(P_j) = \sum_i v(Q_i)\, S_{ij}; \qquad \text{with} \qquad S_{ij} = |\langle q_i|p_j\rangle|^2 = |U_{ij}|^2. \tag{5.8}$$

So we see that the matrix S_{ij} is actually unitary-stochastic (which is a bi-stochastic matrix whose entries are the squares of the absolute values of the entries of some unitary matrix); both unitary and unitary-stochastic matrices drop out automatically.

Now pick some random basis P_i and construct

$$\rho_P = \sum_i v(P_i)\, P_i. \tag{5.9}$$

This is not ρ itself, but it is what you get from ρ by hitting it with $\$_P$, the decoherence super-scattering operator with respect to the basis P_i [163]. (At a basic level, a super-scattering operator can be viewed as a trace-preserving linear mapping from density matrices to density matrices.) To see this note

$$\$_P\ \rho = \sum_i P_i\ \mathrm{tr}(P_i\,\rho) = \sum_i P_i\ v(P_i) = \rho_P. \tag{5.10}$$

Finally consider what happens if you average over the P_i:

$$\left\langle \$_P \right\rangle\ \rho = \left\langle \sum_i P_i\ \mathrm{tr}(P_i\,\rho) \right\rangle = \left\langle \sum_i P_i\ \ v(P_i) \right\rangle = \left\langle \rho_P \right\rangle. \tag{5.11}$$

In d dimensions for a uniform average over the $(P_i)_{ab}$ we have

$$\left\langle \sum_i (P_i)_{ab}\ (P_i)_{cd} \right\rangle = \frac{\delta_{ac}\delta_{bd} + \delta_{ab}\delta_{cd}}{d+1}. \tag{5.12}$$

This arises from symmetry plus the normalization condition $\langle I_{d\times d}\rangle = I_{d\times d}$. But

then we can reconstruct

$$\rho = (d+1)\,\langle \rho_P \rangle - I_{d\times d}. \tag{5.13}$$

(Note this does have the correct trace, $\mathrm{tr}(\rho) = 1$.) So if you know all possible ways in which the density matrix decoheres $\rho \to \rho_P$, and uniformly average over all choices of decoherence basis, then one can reconstruct the full density matrix. While certainly an elegant result, this is by no means explicit.

d) Implicit construction: Let us now set up a reasonably explicit construction of the density matrix ρ directly from the valuation function $v(P)$. To construct ρ proceed as follows: First for any 1-dimensional subspace note $Q \sim |n\rangle\,\langle n|$ where n can be taken to be a unit vector in S^{d-1}. This defines a valuation $v(n)$ on S^{d-1}. Then find a n_1 such that $v(Q_{n_1}) = \max_{n\in S^{d-1}}\{v(P_n)\} = \max_{n\in S^{d-1}}\langle n|\rho|n\rangle$.

Now consider the S^{d-2} perpendicular to n_1: Proceed as follows — find a n_2 such that $v(Q_{n_2}) = \max_{n\in S^{d-2}}\{v(P_n)\}$. By construction $n_1 \perp n_2$ and $P_{n_1}P_{n_2} = 0$. Iterate this construction: Consider the S^{d-i} perpendicular to $n_1, n_2, \ldots, n_{i-1}$: Find a n_i such that $v(Q_{n_i}) = \max_{n\in S^{d-i}}\{v(P_n)\}$. By construction the n_j for $j \in \{1, 2, \cdots, i\}$ are mutually perpendicular, and $P_{n_j}P_{n_k} = 0$ for $j \neq k$ and $j, k \in \{1, 2, \cdots, i\}$. Ultimately we have $n_d = \max_{n\in S^0}\{v(P_n)\} = \min_{n\in S^{d-1}}\{v(P_n)\}$. The construction terminates after d steps with an orthonormal basis $n_1, n_2, \ldots, n_d$, and the corresponding valuations $v(Q_{n_i})$. Now construct

$$\rho = \sum_{i=1}^{d} v(Q_{n_i})\, Q_{n_i}. \tag{5.14}$$

This is the density matrix you want. □

Proof. It is clearly a density matrix; it only remains to check that it is the density matrix. But this is obvious from the construction — the n_i are the simply eigenvectors of ρ, with the corresponding projection operators Q_{n_i}, and the $v(Q_{n_i})$ are the eigenvalues. (Basically the construction above is just an application of the Rayleigh–Ritz min-max variational theorem for finding eigenvectors/eigenvalues of Hermitian matrices.) The density matrix is constructed in terms of the values, $v(Q_{n_i})$, and locations, n_i, of the maximum, minimum, and

extremal points of the valuation function $v(\cdot)$. ■

Note the construction is still rather implicit. Once Gleason's theorem guarantees the existence of the density matrix, this construction implicitly allows one to determine the density matrix. The more purist of constructivist mathematicians might not call this constructive, but most others would. On the other hand, as we shall now show, much better can be done in terms of a fully explicit construction.

e) Explicit construction: This second construction is completely explicit but considerably more subtle. We assert that within the framework of Gleason's theorem, for any arbitrary basis on complex Hilbert space we can write:

$$\rho = \sum_j |n_j\rangle\ v(n_j)\ \langle n_j| \tag{5.15}$$
$$+\frac{1}{2}\sum_{j\neq k} |n_j\rangle \left\{ v\left(\frac{n_j+n_k}{\sqrt{2}}\right) - v\left(\frac{n_j-n_k}{\sqrt{2}}\right) - i\ v\left(\frac{n_j+in_k}{\sqrt{2}}\right) + iv\left(\frac{n_j-in_k}{\sqrt{2}}\right)\right\}\ \langle n_k|.$$

That is, to reconstruct the full density matrix we need only determine the valuations $v(\cdot)$, which is a collection of real numbers, on the specific set of unit vectors

$$n_j; \qquad \left(\frac{n_j \pm n_k}{\sqrt{2}}\right); \qquad \left(\frac{n_j \pm in_k}{\sqrt{2}}\right). \tag{5.16}$$

There are a total of $d + d(d-1) + d(d-1) = 2d^2 - d$ such unit vectors to deal with. This formula for the density matrix can also be rearranged as follows

$$\begin{aligned}\rho \;=\; & \sum_j v(n_j)|n_j\rangle\ \langle n_j| \\ & +\frac{1}{2}\sum_{j<k}\left\{v\left(\frac{n_j+n_k}{\sqrt{2}}\right) - v\left(\frac{n_j-n_k}{\sqrt{2}}\right)\right\}\left(|n_j\rangle\ \langle n_k| + |n_k\rangle\ \langle n_j|\right) \\ & -\frac{i}{2}\sum_{j<k}\left\{\ v\left(\frac{n_j+in_k}{\sqrt{2}}\right) + v\left(\frac{n_j-in_k}{\sqrt{2}}\right)\right\}\left(|n_j\rangle\ \langle n_k| - |n_k\rangle\ \langle n_j|\right).\end{aligned} \tag{5.17}$$

In this form, Hermiticity of the density matrix is manifest. The situation for a

real Hilbert space is considerably simpler:

$$\rho = \sum_j |n_j\rangle\ v(n_j)\ \langle n_j| + \frac{1}{2}\sum_{j\neq k} |n_j\rangle \left\{ v\left(\frac{n_j + n_k}{\sqrt{2}}\right) - v\left(\frac{n_j - n_k}{\sqrt{2}}\right)\right\}\ \langle n_k|. \tag{5.18}$$

There are now only a total of $d + d(d-1) = d^2$ unit vectors to deal with. This formula for the (real) density matrix can also be rearranged as follows

$$\begin{aligned}\rho &= \sum_j v(n_j)|n_j\rangle\ \langle n_j| \\ &\quad + \frac{1}{2}\sum_{j<k}\left\{ v\left(\frac{n_j + n_k}{\sqrt{2}}\right) - v\left(\frac{n_j - n_k}{\sqrt{2}}\right)\right\}\left(|n_j\rangle\ \langle n_k| + |n_k\rangle\ \langle n_j|\right).\end{aligned} \tag{5.19}$$

In this form, symmetry of the (real) density matrix is manifest. To start the construction, following [154], we extend the valuation $v(P) \longleftrightarrow v(n)$ from S^{d-1} to all of H as follows:

$$f(n) = ||n||^2\ v\left(\frac{n}{||n||}\right), \tag{5.20}$$

Now, again following [154],

$$\langle x|\rho|y\rangle = \frac{f(x+y) - f(x-y)}{4} - i\ \frac{f(x+iy) - f(x-iy)}{4}, \tag{5.21}$$

which in the real case reduces to

$$\langle x|\rho|y\rangle = \frac{f(x+y) - f(x-y)}{4}. \tag{5.22}$$

In [154], it asserts the equivalence of:

- $\langle ax|\rho|by\rangle = \overline{a}\, b\, \langle x|\rho|y\rangle$.
- $\langle x|\rho|y\rangle = \overline{\langle y|\rho|x\rangle}$.
- $\langle x|\rho|y_1 + y_2\rangle = \langle x|\rho|y_1\rangle + \langle x|\rho|y_2\rangle$.

where the overline signifies the complex conjugation. This is needed to verify that $\langle x|\rho|y\rangle$ actually represents a bilinear form. Then the density matrix ρ can itself be defined by

$$\rho = \sum_j \sum_k |n_j\rangle\ \langle n_j|\rho|n_k\rangle\ \langle n_k|. \tag{5.23}$$

So

$$\rho = \sum_j \sum_k |n_j\rangle \left\{ \frac{f(n_j + n_k) - f(n_j - n_k)}{4} - i \frac{f(n_j + in_k) - f(n_j - in_k)}{4} \right\} \langle n_k|. \tag{5.24}$$

Whence, splitting the sum into diagonal and off-diagonal pieces, and noting that both $||n_j \pm n_k||^2 = 2 = ||n_j \pm in_k||^2$, while $\widehat{n_j \pm n_k} = (n_j \pm n_k)/\sqrt{2}$, and finally $\widehat{n_j \pm in_k} = (n_j \pm in_k)/\sqrt{2}$, we have:

$$\begin{aligned} \rho &= \sum_j |n_j\rangle \, v(n_j) \, \langle n_j| \\ &+ \frac{1}{2} \sum_{j \neq k} |n_j\rangle \left\{ v\left(\frac{n_j + n_k}{\sqrt{2}}\right) - v\left(\frac{n_j - n_k}{\sqrt{2}}\right) - i \, v\left(\frac{n_j + in_k}{\sqrt{2}}\right) + iv\left(\frac{n_j - in_k}{\sqrt{2}}\right) \right\} \langle n_k|. \end{aligned} \tag{5.25}$$

That is, in terms of the decohered density matrix ρ_P we have:

$$\begin{aligned} \rho &= \rho_P \\ &+ \frac{1}{2} \sum_{j \neq k} |n_j\rangle \left\{ v\left(\frac{n_j + n_k}{\sqrt{2}}\right) - v\left(\frac{n_j - n_k}{\sqrt{2}}\right) - i \, v\left(\frac{n_j + in_k}{\sqrt{2}}\right) + iv\left(\frac{n_j - in_k}{\sqrt{2}}\right) \right\} \langle n_k|. \end{aligned} \tag{5.26}$$

For a real Hilbert space this reduces to

$$\rho = \rho_P + \frac{1}{2} \sum_{j \neq k} |n_j\rangle \left\{ v\left(\frac{n_j + n_k}{\sqrt{2}}\right) - v\left(\frac{n_j - n_k}{\sqrt{2}}\right) \right\} \langle n_k|. \tag{5.27}$$

One aspect of the "miracle" of Gleason's theorem is that this construction is actually independent of the specific basis chosen. To see why this construction works, note that from Gleason's theorem, for unit vectors

$$\hat{x} \sim |\hat{x}\rangle = \frac{|x\rangle}{||x||} \sim \frac{x}{||x||}, \tag{5.28}$$

we have

$$v(\hat{x}) = \langle \hat{x}|\rho|\hat{x}\rangle = \frac{\langle x|\rho|x\rangle}{||x||^2}, \tag{5.29}$$

or more prosaically

$$\langle x|\rho|x\rangle = ||x||^2 v(\hat{x}). \tag{5.30}$$

But then

$$\langle x+y|\rho|x+y\rangle = ||x+y||^2 v(\widehat{x+y}) = \langle x|\rho|x\rangle + \langle y|\rho|y\rangle + (\langle x|\rho|y\rangle + \langle y|\rho|x\rangle), \quad (5.31)$$

and

$$\langle x-y|\rho|x-y\rangle = ||x-y||^2 v(\widehat{x-y}) = \langle x|\rho|x\rangle + \langle y|\rho|y\rangle - (\langle x|\rho|y\rangle + \langle y|\rho|x\rangle), \quad (5.32)$$

whence

$$\langle x|\rho|y\rangle + \langle y|\rho|x\rangle = \frac{1}{2}\left\{||x+y||^2 v(\widehat{x+y}) - ||x-y||^2 v(\widehat{x-y})\right\}. \quad (5.33)$$

(In a real Hilbert space we could stop here since then $\langle x|\rho|y\rangle = \langle y|\rho|x\rangle$.) Similarly, in a complex Hilbert space,

$$\langle x+iy|\rho|x+iy\rangle = ||x+iy||^2 v(\widehat{x+iy}) = \langle x|\rho|x\rangle + \langle y|\rho|y\rangle + i(\langle x|\rho|y\rangle - \langle y|\rho|x\rangle), \quad (5.34)$$

and

$$\langle x-iy|\rho|x-iy\rangle = ||x-iy||^2 v(\widehat{x-iy}) = \langle x|\rho|x\rangle + \langle y|\rho|y\rangle - i(\langle x|\rho|y\rangle - \langle y|\rho|x\rangle), \quad (5.35)$$

whence

$$\langle x|\rho|y\rangle - \langle y|\rho|x\rangle = -\frac{i}{2}\left\{||x+iy||^2 v(\widehat{x+iy}) - ||x-iy||^2 v(\widehat{x-iy})\right\}. \quad (5.36)$$

Combining these results

$$\begin{aligned}\langle x|\rho|y\rangle \quad = \quad & +\frac{1}{4}\left\{||x+y||^2 v(\widehat{x+y}) - ||x-y||^2 v(\widehat{x-y})\right\} \\ & -\frac{i}{4}\left\{||x+iy||^2 v(\widehat{x+iy}) - ||x-iy||^2 v(\widehat{x-iy})\right\}. \end{aligned} \quad (5.37)$$

This finally justifies our construction of the density matrix ρ as presented above.

f) Two dimensions: Although Gleason's theorem does not apply in two dimensions, there are improved versions of Gleason's theorem based on POVMs, see [156, 157], that do apply to 2-dimensional Hilbert space. In this case the for-

malism simplifies even further: Let $\hat{x}$ and $\hat{y}$ be any orthonormal basis for the 2-dimensional Hilbert space. Then in terms of the valuation $v(\cdot)$ the density matrix is

$$\begin{aligned}\rho &= v(\hat{x})\,|\hat{x}\rangle\,\langle\hat{x}| + v(\hat{y})\,|\hat{y}\rangle\,\langle\hat{y}| \\ &\quad + \frac{1}{2}\left\{v\left(\frac{\hat{x}+\hat{y}}{\sqrt{2}}\right) - v\left(\frac{\hat{x}-\hat{y}}{\sqrt{2}}\right)\right\}\left(|\hat{x}\rangle\,\langle\hat{y}| + |\hat{y}\rangle\,\langle\hat{x}|\right) \\ &\quad - \frac{i}{2}\left\{v\left(\frac{\hat{x}+i\hat{y}}{\sqrt{2}}\right) - v\left(\frac{\hat{x}-i\hat{y}}{\sqrt{2}}\right)\right\}\left(|\hat{x}\rangle\,\langle\hat{y}| - |\hat{y}\rangle\,\langle\hat{x}|\right). \qquad (5.38)\end{aligned}$$

If desired one can further rewrite this in terms of the Pauli σ matrices

$$\begin{aligned}\rho &= \frac{v(\hat{x})+v(\hat{y})}{2}\,I_{2\times 2} + \frac{v(\hat{x})-v(\hat{y})}{2}\,\sigma_z \\ &+ \frac{1}{2}\left\{v\left(\frac{\hat{x}+\hat{y}}{\sqrt{2}}\right) - v\left(\frac{\hat{x}-\hat{y}}{\sqrt{2}}\right)\right\}\sigma_x - \frac{i}{2}\left\{v\left(\frac{\hat{x}+i\hat{y}}{\sqrt{2}}\right) - v\left(\frac{\hat{x}-i\hat{y}}{\sqrt{2}}\right)\right\}\sigma_y. \qquad (5.39)\end{aligned}$$

For real 2-dimensional Hilbert space this further simplifies to

$$\begin{aligned}\rho &= v(\hat{x})\,|\hat{x}\rangle\,\langle\hat{x}| + v(\hat{y})\,|\hat{y}\rangle\,\langle\hat{y}| \\ &\quad + \frac{1}{2}\left\{v\left(\frac{\hat{x}+\hat{y}}{\sqrt{2}}\right) - v\left(\frac{\hat{x}-\hat{y}}{\sqrt{2}}\right)\right\}\left(|\hat{x}\rangle\,\langle\hat{y}| + |\hat{y}\rangle\,\langle\hat{x}|\right). \qquad (5.40)\end{aligned}$$

(For completeness, note that for one dimension the valuation trivializes to $v(\cdot) \equiv 1$, and so the density matrix trivializes to $\rho \equiv I_{1\times 1}$.)

g) Comments:

i) We have not attempted to provided a new proof of Gleason's theorem. We have in mind a much more modest attempt at trying to understand what the density matrix actually looks like directly in terms of the probability valuations $v(\cdot)$ on a limited number of subspaces of the Hilbert space.

ii) Gleason's theorem is profound that it shapes the probabilistic nature of quantum theory resulting in the Born rule. It places strong constraints on any attempts to modify this probabilistic formalism. However, it still requires the assumption of a probabilistic aspect for its derivation.

iii) Future work regarding this explicit construction of the density operator may involve applications to quantum information science. This may reveal

interesting links between quantum foundations, and to the fundamental quantum information results such as no-cloning or no-broadcasting.

5.1.2 Kochen–Specker theorem

In this subsection, we want to articulate a concept known as realism. Realism is the view that physical properties have definite values which exist independent of observation. (This of course seems obvious to classical intuition.) It is also known as value definiteness [164] where it is said that the properties of physical objects always have definite values even if they are not measured or accessible for any observer. In quantum theory, values of physical objects are revealed at the moment of measurement; prior to that we only have access to the quantum state and are not given a physical picture of the world. The crucial question is whether there could be a value definite structure underlying quantum theory?

The Kochen-Specker theorem (sometimes called the Bell–Kochen–Specker theorem) can be crudely stated that if a theory reproduces the results of quantum theory and also has value definiteness, then that theory must be contextual. Contextuality is the property that the result of a measurement can depend on what combination of measurements we chose to do! In other words, the outcome of a question depends on what other questions we are simultaneously trying to answer alongside it. For a classical analogy, suppose one is trying to measure a person's height. Then contextuality in this scenario implies one gets a different value for height if one measured the person's weight along with it than one would get if one measured the person's shoe size along with it! To avoid a contextual characteristic to a theory, the alternative method is to give up value definiteness. In this case the values do not exist before one does a measurement!

a) Preliminaries: Our aim is to state the Kochen-Specker theorem and provide a proof. We shall phrase our discussion in terms of real Hilbert spaces, noting that a complex Hilbert space can always be viewed as a real Hilbert space of double the dimensionality $\mathbb{C}^n \sim \mathbb{R}^{2n}$. One view of the Kochen–Specker theorem is that it demonstrates the impossibility of consistently assigning $\{0, 1\}$ truth values to quantum propositions. It was originally proved some fifty years ago by explicitly

finding a set of 117 distinct projection operators on 3-dimensional Hilbert space [165, 166], and then showing that there was no way to consistently assign values in $\{0, 1\}$ to these projection operators. (That is, these 117 "quantum questions" that one might ask could not be consistently assigned yes-no answers.) A later version of the Kochen–Specker theorem reduced the number of projection operators to 33 [167]. This was further reduced to 24 [167], to 20 [168], and then to 18 [169, 170], at the cost of slightly increasing the dimension of the Hilbert space to 4. Ultimately the number of projection operators was further reduced to 13 in an 8-dimensional Hilbert space in reference [171]. Interest in these foundational issues has continued unabated [172, 173], with at least two "geometrical" proofs that avoid explicit construction of sets of projection operators [174, 175].

In this subsection, we will now present work [176] this is *part of the original component of this thesis* (which was done in collaboration with my supervisor). We shall provide yet another even more simplified "geometrical" proof of the Kochen–Specker theorem, which, while it is still non-constructive, (proceeding by establishing an inconsistency), is utterly minimal in its technical requirements, and so hopefully instructive.

b) Kochen–Specker theorem: An explicit statement of the Kochen–Specker theorem, (based on the discussion in the Stanford encyclopaedia of philosophy), runs thus:

Theorem 5.2. ***(Kochen-Specker – mathematical version)***

Let H be a Hilbert space of quantum state vectors of real dimension $d \geq 3$. Then there is a set M of observables on H, containing n elements, such that the following two assumptions are contradictory:

KS1: *All n members of M simultaneously have values, that is, they are unambiguously mapped onto real numbers (designated, for specific observables $A, B, C, \ldots$, by values $v(A), v(B), v(C), \ldots$).*

KS2: *Values of observables conform to the following constraints:*

- **(a)** *If A, B, C are all compatible and $C = A + B$, then $v(C) = v(A) + v(B)$.*
- **(b)** *If A, B, C are all compatible and $C = AB$, then $v(C) = v(A)v(B)$.*

(c) $\exists$ *at least one observable* X *with* $v(X) \neq 0$.
(Here "compatible" means that the observables commute.)

□

The statement **KS1** essentially captures the notion of value definiteness (or realism). The assumptions **KS2a** and **KS2b** are respectively referred to as the sum rule and product rule. Both of these assumptions are based on what is known as the functional composition principle which is in turn a consequence of non-contextuality. (The explicit connection among these various statements can be found in the Stanford encyclopaedia of philosophy).

There are several technical issues with the above presentation. Without condition **KS2c** the theorem is actually false — the trivial valuation where for all observables X one sets $v(X) = 0$ provides an explicit counter-example. Without condition **KS2c**, $v(I) = v(I^2) = v(I)^2$ only implies $v(I) \in \{0, 1\}$. With condition **KS2c** we have the stronger statement that $v(X) = v(IX) = v(I)v(X)$, which since $v(X) \neq 0$ implies $v(I) = 1$.

A more subtle issue is this: Physically, we would like to have $v(zA) = z\,v(A)$, for any $z \in \mathbb{C}$. But using the conditions **KS2a** and **KS2b** we could only deduce this for rational numbers. Extending this to the complex numbers requires us to first construct the real numbers "on the fly" using Dedekind cuts, and then to formally construct the complex numbers as an algebraic extension of the field of real numbers — while this is certainly possible, in a physics context it is rather pointless — it would seem more reasonable to start with the complex numbers as being given, even if you then need slightly stronger axioms.

Improved KS2 axioms:

(a) If $[A, B] = 0$ and $a, b \in \mathbb{C}$, then $v(aA + bB) = a\,v(A) + b\,v(B)$.

(b) If $[A, B] = 0$ then $v(AB) = v(A)\,v(B)$.

(c) $\exists$ at least one observable X with $v(X) \neq 0$.

If one accepts these improved **KS2** axioms then immediately

$$v(I) = 1; \qquad v(aI) = a; \tag{5.41}$$

and for any analytic function with a non-zero radius of convergence

$$v(f(A)) = f(v(A)). \tag{5.42}$$

Note that this last condition, $v(f(A)) = f(v(A))$, is where physics discussions of the Kochen–Specker theorem often start. Indeed let us write $A = \sum_i a_i P_i$ where the a_i are real and the P_i are projection operators onto 1-dimensional subspaces; so the projection operators $P_i = |\psi_i\rangle \langle\psi_i|$ can be identified with the vectors $|\psi_i\rangle$ which form a basis for the Hilbert space. Then

$$v(A) = v\left(\sum_i a_i P_i\right) = \sum_i a_i \, v(P_i). \tag{5.43}$$

This now focusses attention on the valuations $v(P_i)$. Since $P_i^2 = P_i$, condition **KS2b** implies that $v(P_i) \in \{0, 1\}$; the valuation must be a yes-no valuation. Now consider the identity operator $I = \sum_i P_i$ and note

$$\sum_i v(P_i) = v(I) = 1. \tag{5.44}$$

It is customary to identify the projectors $P_i = |n_i\rangle \langle n_i|$ with the corresponding unit vectors n_i, (defined up to a sign), with the n_i forming a basis for Hilbert space, and in d dimensions write

$$\sum_{i=1}^{d} v(n_i) = 1; \qquad v(n_i) \in \{0, 1\}; \qquad v(-n) = v(n). \tag{5.45}$$

It is the claimed existence of this function $v(n)$, having the properties stated above for any arbitrary basis of Hilbert space, which is the central point of the **KS1** and **KS2** conditions. This discussion allows us to rephrase the Kochen–Specker theorem in terms of the non-existence of such a valuation.

Theorem 5.3. ***(Kochen-Specker — physics-based version)***

For $d \geq 3$ there is no valuation $v(n) : S^{d-1} \to \{0, 1\}$, where S^{d-1} is the unit

hypersphere, such that $v(-n) = v(n)$ *for all* n *and*

$$\sum_{i=1}^{d} v(n_i) = 1, \tag{5.46}$$

for every basis (frame, d*-bein) of orthogonal unit vectors* n_i*.* □

It is this statement about bases in Hilbert space that is often more practical to work with, rather than the formulation at the start of this section — of course without that initial formulation it would be less than clear why the basis formulation is physically interesting.

We will start by looking in a non-traditional place, by considering one-dimensional and two-dimensional Hilbert spaces, before dealing with three-dimensional Hilbert space, (which then settles things for any higher dimensionality). Since one is trying to prove an inconsistency result, there will be an infinite number of ways of doing so; the question is whether one learns anything new by coming up with a different proof. We shall do so with a modified and simplified "descent" argument, one that requires only two steps in the descent process.

c) One dimensions: There is no Kochen–Specker no-go result in one dimension, since in one dimension all operators are multiples of the identity, $A = aI$, and then

$$v(f(A)) = v(f(aI)) = v(f(a)I) = f(a)\, v(I) = f(a). \tag{5.47}$$

In particular, as long as $f(a) \not\equiv 0$, (which is implied by the **KS2c** axiom), then for the (unique) normalized basis vector n we have $v(n) = 1$. Conversely if we are considering a one-dimensional subspace of a higher-dimensional Hilbert space then the **KS2c** axiom tells us nothing; for the (unique) normalized basis vector we merely have $v(n) \in \{0, 1\}$, and we have no further constraint on the valuation.

d) Two dimensions: There is no Kochen–Specker no-go result in two dimensions, but there are still quite interesting things to say. Consider the valuation $v : S^1 \to \{0, 1\}$ (where S^1 is the unit circle) such that $v(-n) = v(n)$ for all n and

$$v(n_1) + v(n_2) = 1 \tag{5.48}$$

for every dyad (every pair of orthogonal unit vectors) n_1, n_2. Indeed in two dimensions we can construct such a valuation. Re-characterize n_1 and n_2 in terms of the angle they make with (say) the x axis; then the constraints we want to impose are

$$v(\theta) = v(\theta + \pi); \qquad v(\theta) + v(\theta \pm \tfrac{\pi}{2}) = 1. \tag{5.49}$$

But these conditions are easily solved: Let $g(\theta)$ be some arbitrary (not necessarily continuous) function mapping the interval $\left[0, \frac{\pi}{2}\right) \to \{0, 1\}$, and define

$$v(\theta) = \begin{cases} g(\theta) & \text{for } \theta \in [0, \frac{\pi}{2}); \\ 1 - g(\theta - \frac{\pi}{2}) & \text{for } \theta \in [\frac{\pi}{2}, \pi); \\ g(\theta - \pi) & \text{for } \theta \in [\pi, \frac{3\pi}{2}); \\ 1 - g(\theta - \frac{3\pi}{2}) & \text{for } \theta \in [\frac{3\pi}{2}, 2\pi). \end{cases} \tag{5.50}$$

So the existence of a Kochen–Specker valuation is easily verified in two dimensions, and because points separated by $\pi/2$ radians must be given opposite valuations, the image $v(S^1)$ is automatically 50%–50% zero-one. Note in particular that the function $v(\theta)$ cannot be everywhere continuous. (We will recycle these results repeatedly when we turn to three and higher dimensions.)

e) Three dimensions: It is in 3 dimensions that things first get interesting. We are interested in valuations $v : S^2 \to \{0, 1\}$, (where S^2 is the unit 2-sphere), such that $v(-n) = v(n)$ for all n and

$$v(n_1) + v(n_2) + v(n_3) = 1 \tag{5.51}$$

for every triad (every triplet of orthogonal unit vectors) n_1, n_2, n_3. In the argument below we shall make extensive use of the great circles S^1 in the unit 2-sphere S^2.

Lemma: On any great circle in S^2, under the conditions given above, the valuation is either 50%–50% zero-one (as in two dimensions), or is 100% zero (identically zero). □

Proof. Pick any great circle and for convenience align it with the equator.

Now look at the poles:

- If v(poles) = 1, then v(equator) ≡ 0 is identically zero.
  (Since points on the equator will be part of some triad that includes the unit vector pointing to the poles.)
- If v(poles) = 0, then any dyad lying in the equator will satisfy the conditions of the two dimensional argument given above, and so will be 50%–50% zero-one.

■

Now bootstrap this to a modified "great circle descent" argument, one that needs only two steps in the descent process. We start with a purely geometrical result. From the argument above if we arrange v(poles) = 1, then v(equator) ≡ 0, and for each line of longitude v(meridian) will be 50%–50% zero-one. (See figure 5.1.)

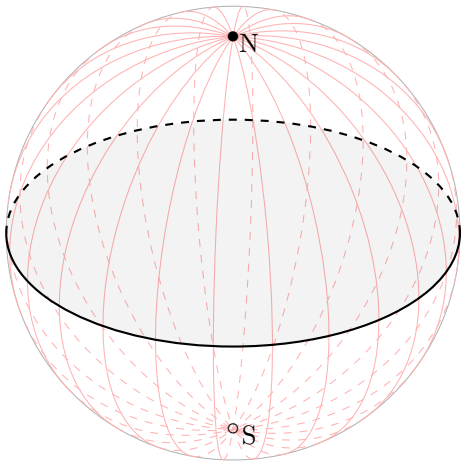

Figure 5.1: Setup with v(poles) = 1, v(equator) ≡ 0, and v(meridians) 50%–50% zero-one.

We define a "great circle descent" $C(p)$ through a point p on the sphere as a great circle that starts off at constant latitude. (So the point p is either the northernmost or southernmost point on the great circle. See figures 5.2 and 5.3.)

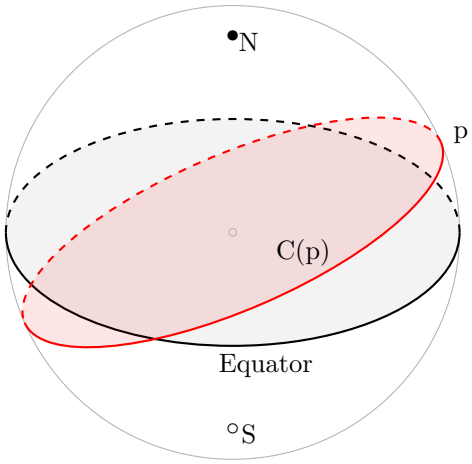

Figure 5.2: Descent great circle $C(p)$, with northernmost point at p.

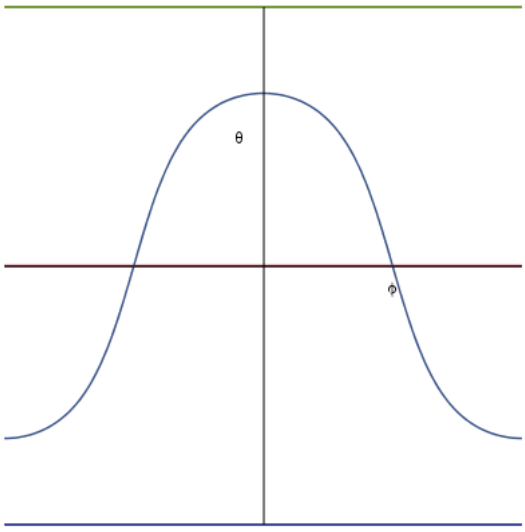

Figure 5.3: Descent great circle represented in terms of $\theta(\phi)$.

Lemma: Let q be any other point at the same longitude as p (the same meridian) that is closer to the equator than p. Then there exists a point r such that r lies on the great circle descent through p, and q lies on the great circle descent through r. □

That is $r \in C(p)$ and $q \in C(r)$, so one can always zig-zag directly to towards the equator via exactly two great circle descents. Note that this is a much easier geometric result than that used in the Gill–Keane [174] or Calude–Hertling–Svozil [175] approaches where a finite but possibly large number of great circle descents is used to get to any point closer to the equator, not necessarily at the same longitude. (See figure 5.4.)

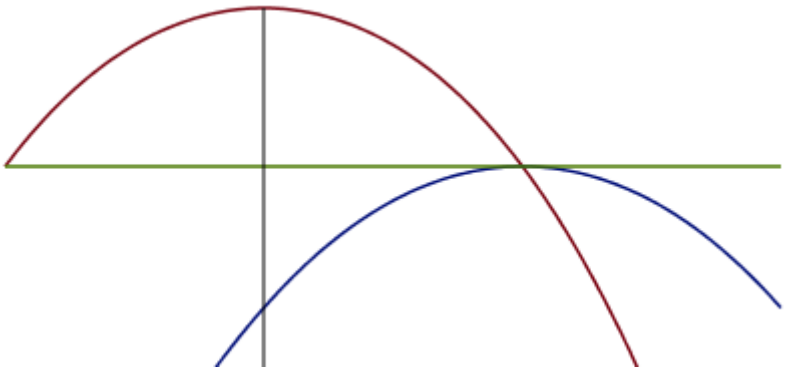

Figure 5.4: Example of a two-step descent with net motion along a meridian.

Proof. Using spherical coordinates (θ, ϕ) let the generic point x be represented by the 3-vector

$$\vec{x} = (\cos\theta\cos\phi, \cos\theta\sin\phi, \sin\theta). \tag{5.52}$$

(Somewhat unusually, we adopt conventions close to the usual latitude nomenclature: $\theta = +\pi/2$ represents the north pole, $\theta = 0$ represents the equator, while while $\theta = -\pi/2$ represents the south pole. Doing this simplifies some of the formulae below.)
Now let the specific point p of interest be represented by the 3-vector

$$\vec{p} = (\cos\theta_p\cos\phi_p, \cos\theta_p\sin\phi_p, \sin\theta_p). \tag{5.53}$$

Consider the great circle descent $C(p)$. This great circle will be orthogonal to the vector

$$\vec{p}_\perp = (\sin\theta_p \cos\phi_p, \sin\theta_p \sin\phi_p, -\cos\theta_p). \tag{5.54}$$

The entire great circle $C(p)$ will be characterized by $\vec{p}_\perp \cdot \hat{x}(\theta,\phi) = 0$, that is

$$\sin\theta_p \cos\theta(\cos\phi_p \cos\phi + \sin\phi_p \sin\phi_p) - \cos\theta_p \sin\theta = 0, \tag{5.55}$$

implying

$$\sin\theta_p \ \cos\theta \ \cos(\phi - \phi_p) = \cos\theta_p \sin\theta. \tag{5.56}$$

That is

$$\tan\theta = \tan\theta_p \cos(\phi - \phi_p), \tag{5.57}$$

or more explicitly

$$\theta(\phi) = \tan^{-1}\left(\tan\theta_p \ \cos(\phi - \phi_p)\right). \tag{5.58}$$

This explicitly yields $\theta(\phi)$ along the entire descent circle $C(p)$.

Note that this descent circle crosses the equator at $\theta = 0$, implying $(\phi - \phi_p) = \pm\pi/2$. This occurs at the points s such that $\vec{s} = \pm(-\sin\phi_p, \cos\phi_p, 0)$. In particular, for the three points p, r, q, (and using $\phi_p = \phi_q$ because we want p and q to have the same longitude), we have

$$\tan\theta_r = \tan\theta_p \ \cos(\phi_r - \phi_p); \qquad \tan\theta_q = \tan\theta_r \ \cos(\phi_r - \phi_p); \tag{5.59}$$

implying

$$\tan\theta_q = \tan\theta_p \ \cos^2(\phi_r - \phi_p). \tag{5.60}$$

That is

$$\cos^2(\phi_r - \phi_p) = \frac{\tan\theta_q}{\tan\theta_p}. \tag{5.61}$$

Alternatively

$$|\phi_r - \phi_p| = \cos^{-1}\sqrt{\frac{\tan\theta_q}{\tan\theta_p}}. \tag{5.62}$$

The azimuthal difference $|\phi_r - \phi_p|$ tells you exactly how much you have to zig-zag along the descent circles for the net motion to be directly along the line of

longitude towards the equator. Note $|\phi_r - \phi_p|$ is real only if you move towards (rather than away from) the equator. ■

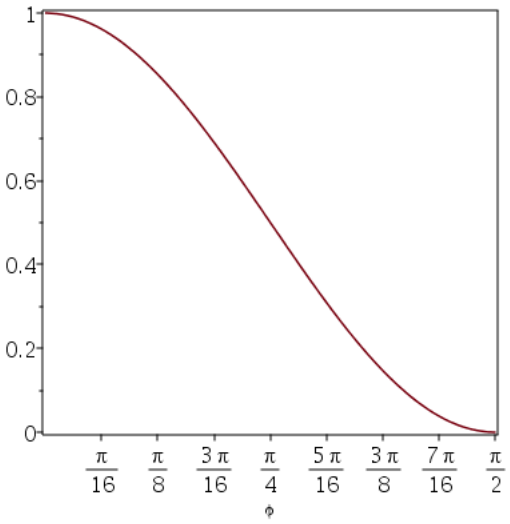

Figure 5.5: Quantifying $\delta\phi$ in terms of $\cos^{-1}\sqrt{\frac{\tan\theta_q}{\tan\theta_p}}$ for two-step descent towards the equator.

Application to the Kochen-Specker theorem:
Consider any point p_0 such that $v(p) = 1$ and rotate to put it at the north or south pole. Then by hypothesis $v(equator) = 0$ for any point on the equator. Now consider any other point p such that $v(p) = 0$ and p is not on the equator. Consider the descent circle $C(p)$; we have $v(p) = 0$ by hypothesis, and $v(s) = 0$ at the perpendicular point s with $\vec{s} = (0, -\sin\phi_0, \cos\phi_0)$ where $C(p)$ crosses the equator. Therefore $v(C(p)) \equiv 0$ everywhere on this descent circle. But in particular this implies that $v(r) = 0$. Thence $v(C(r)) \equiv 0$ everywhere on this descent circle. Thence $v(q) = 0$. This means we have proved:

Lemma: If $v(pole) = 1$ and $v(p) = 0$ then also $v(q) = 0$ for q any point on the same line of longitude (same meridian) as p that is closer to the equator than p.
□

Consequently, for any line of longitude for which $v(poles) = 1$, we see that $v^{-1}(0)$

is path connected. Specifically this implies that $\exists\ \pi/2 \leq \theta_* \leq 0$ such that either

$$v(\theta) = \begin{cases} 1 & \text{for } \frac{\pi}{2} \geq \theta \geq \theta_*; \\ 0 & \text{for } \theta_* > \theta \geq \theta_* - \frac{\pi}{2}; \\ 1 & \text{for } \theta_* - \frac{\pi}{2} > \theta \geq -\frac{\pi}{2}; \end{cases} \tag{5.63}$$

or

$$v(\theta) = \begin{cases} 1 & \text{for } \frac{\pi}{2} \geq \theta > \theta_*; \\ 0 & \text{for } \theta_* \geq \theta > \theta_* - \frac{\pi}{2}; \\ 1 & \text{for } \theta_* - \frac{\pi}{2} \geq \theta \leq -\frac{\pi}{2}. \end{cases} \tag{5.64}$$

(See figure 5.6.)

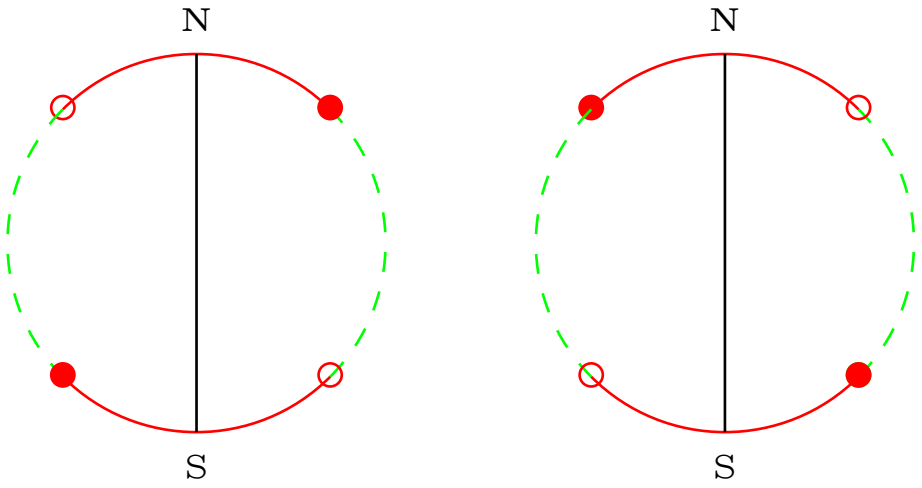

Figure 5.6: Assuming $v(poles) = 1$, as a consequence of the two-step descent argument *any* meridian can be put into one of these two forms for *some* value of θ_*.

Now pick any specific line of longitude, by interchanging the north and south

poles we can without loss of generality assert

$$v(\theta) = \begin{cases} 1 & \text{for } \frac{\pi}{2} \geq \theta \geq \theta_*; \\ 0 & \text{for } \theta_* > \theta \geq \theta_* - \frac{\pi}{2}; \\ 1 & \text{for } \theta_* - \frac{\pi}{2} > \theta \geq -\frac{\pi}{2}. \end{cases} \tag{5.65}$$

Now rotate the sphere S^2 around the polar axis so that the line of longitude we have chosen lies on the zero meridian $\phi_* = 0$ (the prime meridian). Then furthermore rotate the sphere S^2 around the axis perpendicular to the zero meridian so that point $p_* = (\sin\theta_*, 0, \cos\theta_*)$ is moved to the north pole. That is:

Lemma: Without any loss of generality we can choose the zero meridian to satisfy

$$v(\theta) = \begin{cases} 1 & \text{for } \theta = \frac{\pi}{2}; \\ 0 & \text{for } \frac{\pi}{2} > \theta \geq 0; \\ 1 & \text{for } 0 > \theta \geq -\frac{\pi}{2}. \end{cases} \tag{5.66}$$

(See figure 5.7.)

This will now quickly lead to a contradiction. □

First consider all the descent great circles $C(p)$ based on this particular choice of zero meridian. These descent great circles will (in the northern hemisphere) sweep out the entire half-hemisphere $\phi \in (-\pi/2, +\pi/2)$ and $\theta \in (\pi/2, 0)$. Similarly, in the southern hemisphere these decent circles will in turn sweep out the complementary half-hemisphere $\phi \in (+\pi/2, \pi] \cup [-\pi, -\pi/2)$ and $\theta \in (0, -\pi/2)$. But, following previous arguments, since $v(\theta) = 0$ at the apex of all these descent great circles, $v(C(p)) = 0$ for all these descent great circles. That is:

Lemma: Without loss of generality we have chosen the zero meridian such that (except possibly at the poles themselves)

$$v(\theta > 0, |\phi| < \pi/2) = 0; \qquad \text{and} \qquad v(\theta < 0, |\phi| < \pi/2) = 1; \tag{5.67}$$

$$v(\theta < 0, |\phi| > \pi/2) = 0; \qquad \text{and} \qquad v(\theta > 0, |\phi| > \pi/2) = 1. \tag{5.68}$$

Thus the valuation $v(\cdot)$ is 50%–50% zero-one over the entire 2-sphere S^2. □

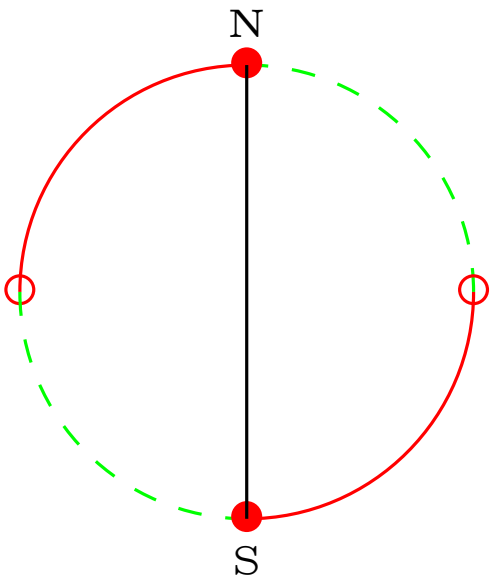

Figure 5.7: Assuming $v(poles) = 1$, after suitable rotations the prime meridian can always be put into this standardized form.

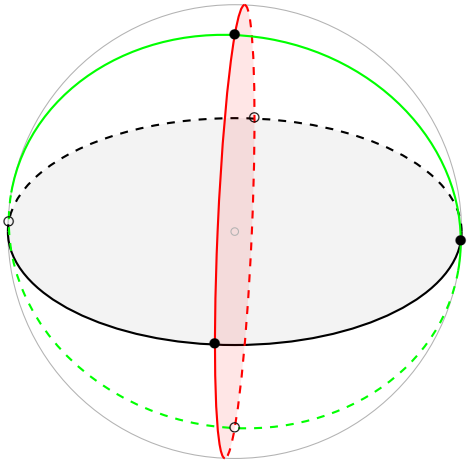

Figure 5.8: Let green denote the prime meridian at $\phi = 0$, black the equator at $\theta = 0$, and red the meridians at $\phi = \pm\pi/2$. The equator and red meridians split the sphere into four segments, with two of these segments having valuation zero, and two segments having valuation unity.

Completing the inconsistency argument can now be done in many ways (in fact, an infinite number of ways). Consider any meridian with $\phi_* \neq 0$ and $|\phi_*| < \pi/2$. On the one hand this meridian will also have the same valuation, equation (5.66), as the zero meridian. On the other hand by considering the descent great circles based on this new meridian we have

$$v(\theta > 0, |\phi - \phi_*| < \pi/2) = 0; \qquad \text{and} \qquad v(\theta > 0, |\phi - \phi_*| < \pi/2) = 1; \quad (5.69)$$

$$v(\theta < 0, |\phi - \phi_*| > \pi/2) = 0; \qquad \text{and} \qquad v(\theta < 0, |\phi - \phi_*| > \pi/2) = 1. \quad (5.70)$$

But this is incompatible with the behaviour based on the zero meridian, equations (5.67) and (5.68), so we have a contradiction. This completes the proof of Kochen–Specker in three dimensions. We feel that this is a nice simple proof of Kochen–Specker that does not rely on finding explicit bases for the Hilbert space — it also seems to us to be considerably simpler than the other geometric or colouring arguments.

f) Four dimensions and higher: What happens in a $d > 3$-dimensional Hilbert space? The 3-dimensional logic carries over with utterly minimal modifications.

- In $d = 4$ one needs to study the unit 3-sphere S^3. Pick any point n on S^3 such that $v(n) = 0$. This can always be done. Then consider the 2-sphere perpendicular to chosen point n. On that 2-sphere the 4-dimensional Kochen–Specker theorem will reduce to the 3-dimensional Kochen–Specker theorem, which we have already established. So nothing more need be done.

- In $d \geq 4$ dimensions one needs to study the unit $(d-1)$-sphere S^{d-1}. Pick any $d-3$ mutually-orthogonal points n_i on S^3 such that $v(n_i) = 0$. If this cannot be done then the existence of the claimed valuation $v(\cdot)$ already fails at this elementary level so that the d-dimensional Kochen–Specker theorem is established; so without loss of generality we can assume this can be done. Then consider the 2-sphere perpendicular to all the n_i. On that 2-sphere the d-dimensional Kochen–Specker theorem will reduce to the 3-dimensional Kochen–Specker theorem, which we have already established. So nothing more need be done.

It is interesting to note that 3-dimensions is the key part of the theorem; in 1 and 2 dimensions related results are trivial. In 4 or more dimensions the Kochen–Specker theorem follows immediately from the 3-dimensional result.

g) Comments:

i) We have presented a geometric approach where one constructs and exploits the properties of great circles on a n-sphere. This has the power to significantly simplify the argument, while maintaining the validity of the theorem for a minimum dimension of three.

ii) The Kochen–Specker theorem is more basic and fundamental than Gleason's theorem. Indeed, if one assumes Gleason's theorem then the Kochen–Specker theorem is trivial. The point is that once one asserts that the valuation $v(\cdot)$ is inherited from a density matrix $v(n) = \langle n|\rho|n\rangle$, then one knows that the valuation is continuous. But no function from the connected space S^n to the discrete set $\{0, 1\}$ (with its implied discrete topology) can possibly be continuous.

iii) The main implication of the result is that quantum theory fails to allow a underlying non-contextual model. More precisely, it states that it is impossible for the predictions of quantum mechanics to be in line with measurement outcomes which are pre-determined in a non-contextual manner. Hence this would rule out a large class of models that might otherwise seem at first sight to be intuitive representations of the physical world.

iv) With respect to quantum information science, there has been recent evidence that contextuality may be the primary reason for the speedup for quantum computation. This has been shown through 'magic' state injection [177].

5.2 Non-locality across Space

Non-locality across space is the characteristic that an action on a subsystem can instantaneously influence another subsystem at an arbitrarily far spatial location. We have seen this strange property exemplified in the previous chapter regarding the entanglement in space. In this section we will describe two different non-

localities across space. One requires an entanglement in space, and is known as Bell non-locality. The other does not require the entanglement and is known as the violation of preparation independence.

By considering both the probabilistic aspects of quantum theory and the concept of realism, one is led to a mathematical formulation of the Bell non-locality across space. This will expressed through what is known as Bell's theorem [178]. We will show that Bell non-locality across space represents a stricter form of non-classical interdependence than an entanglement in space. The particular version of Bell's theorem we will focus our attention on is the Bell-CHSH or known simply as the CHSH (Clauser-Horne-Shimony-Holt) inequality [60]. For the questions in quantum foundations, the CHSH inequality sheds a partial understanding on the nature of quantum measurement. Due to its implications, Bell's theorem has also been viewed by some as the most profound discovery of science [179]. For a thorough review of Bell non-locality across space, we refer the reader to [57].

The second non-locality across space is known as the violation of preparation independence. It does not require entanglement and applies to product states. It will be expressed through the PBR (Pusey-Barrett-Rudolph) theorem [180]. The original aim of the theorem was to shed a partial understanding on the nature of the quantum state. Due to its implications, the PBR theorem has been referred to as the most important theorem in quantum foundations since Bell's theorem [181]. For a comprehensive reviews of the PBR theorem and the violation of preparation independence, we refer the reader to [182, 1]

Our aim in this section is therefore to present the CHSH inequality and the PBR theorem. We will also articulate both of these results through the lens of a game.

5.2.1 Bell-CHSH inequality

a) CHSH inequality: The CHSH inequality will be used to demonstrate a non-locality across space. It will be derived without any reference to quantum theory. Suppose there are three parties who are each spatially apart named Alice, Bob and Charlie. Charlie prepares two particles and sends one particle to Alice and

the other one to Bob. Each particle can be measured in two quantities. For Alice's particle we denote these quantities as A_1 and A_2, and similarly for Bob's particle we have quantities B_1 and B_2. Each of these can take either value $+1$ or -1. We assume realism, and hence the values are objective properties which exist independent of observation; these values are merely revealed by measurement.

Both Alice and Bob each choose to measure their respective particles at the same time. With this constraint, we can assume that the measurement of one particle cannot effect the result of the other particle. This is known as the assumption of locality. Furthermore, we also require that each choose to measure their particle randomly using their two options. This is also known as the free will assumption.

We proceed to consider the quantity

$$A_1B_1 + A_2B_1 + A_2B_2 - A_1B_2. \tag{5.71}$$

This can be re-expressed as

$$A_1B_1 + A_2B_1 + A_2B_2 - A_1B_2 = (A_1 + A_2)B_1 + (A_2 - A_1)B_2. \tag{5.72}$$

Given that $A_1, A_2 = \pm 1$, we have that

$$(A_1 + A_2)B_1 = 0, \tag{5.73}$$

or

$$(A_2 - A_1)B_2 = 0. \tag{5.74}$$

For both cases, we obtain

$$A_1B_1 + A_2B_1 + A_2B_2 - A_1B_2 = \pm 2. \tag{5.75}$$

Let $p(a_1, a_2, b_1, b_2)$ denote the joint probability that before the measurements are performed the total system is in state $A_1 = a_1, A_2 = a_2, B_1 = b_1$, and $B_2 = b_2$.

Using the expectation value (2.1), we have

$$\mathbb{E}(A_1B_1 + A_2B_1 + A_2B_2 - A_1B_2) \tag{5.76}$$
$$= \sum_{a_1a_2b_1b_2} p(a_1, a_2, b_1, b_2)(a_1b_1 + a_2b_1 + a_2b_2 - a_1b_2) \tag{5.77}$$
$$\leq \sum_{a_1a_2b_1b_2} p(a_1, a_2, b_1, b_2)(2) \tag{5.78}$$
$$= 2. \tag{5.79}$$

We can also deduce that

$$\mathbb{E}(A_1B_1 + A_2B_1 + A_2B_2 - A_1B_2) \tag{5.80}$$
$$= \sum_{a_1a_2b_1b_2} p(a_1, a_2, b_1, b_2)a_1b_1 + \sum_{a_1a_2b_1b_2} p(a_1, a_2, b_1, b_2)a_2b_1 \tag{5.81}$$
$$+ \sum_{a_1a_2b_1b_2} p(a_1, a_2, b_1, b_2)a_2b_2 - \sum_{a_1a_2b_1b_2} p(a_1, a_2, b_1, b_2)a_1b_2 \tag{5.82}$$
$$= \mathbb{E}(A_1B_1) + \mathbb{E}(A_2B_1) + \mathbb{E}(A_2B_2) - \mathbb{E}(A_1B_2). \tag{5.83}$$

Using both (5.79) and (5.83), we obtain the CHSH inequality

$$\mathbb{E}(A_1B_1) + \mathbb{E}(A_2B_1) + \mathbb{E}(A_2B_2) - \mathbb{E}(A_1B_2) \leq 2. \tag{5.84}$$

This can be re-written using the quantum theoretic notation for expectation value

$$\boxed{\langle A_1B_1\rangle + \langle A_2B_1\rangle + \langle A_2B_2\rangle - \langle A_1B_2\rangle \leq 2.} \tag{5.85}$$

This is the equation we saw earlier (4.28) as a means to detect entanglement. More precisely, using the probabilistic aspects of quantum theory, we saw the CHSH inequality violated using Bell state (4.29), resulting in the equation

$$\langle A_1B_1\rangle + \langle A_2B_1\rangle + \langle A_2B_2\rangle - \langle A_1B_2\rangle = 2\sqrt{2}. \tag{5.86}$$

The value of $2\sqrt{2}$ is the maximum quantum value and is known as Tsirelson's bound. This violation has been experimentally verified [183, 184] in numerous quantum scenarios. Hence, these measurement correlations are stronger than

could ever exist in classical systems. It implies a profound consequence in that these quantum correlations overthrow the classical picture of the world; at least one of the three assumptions made to derive the CHSH inequality is wrong.

b) Implications: It is dominantly viewed that the assumption of locality is the one that is incorrect. Hence, the mathematical characterization for Bell *non-locality across space* is expressed as the violation of (5.85). Note that this Bell notion of locality (5.85) is distinct from the term locality used in other areas of quantum physics which describes the case that operators defined in spacelike separated regions commute. In this section, when we refer to non-locality we shall mean a Bell non-locality across space.

It also is common in the literature to interchange between the terms entanglement and non-locality. Such use may in principle be sufficient for a large number of cases, but falls short of the precision required for an adequate scientific taxonomy. We proceed to emphasize the differences between an entanglement and non-locality. The most obvious difference is that former is an algebraic property residing in the mathematics of quantum theory (4.2), whereas the latter is rooted from the measurement outcomes/correlations of experiments (5.85).

Nevertheless, to obtain non-local correlations from measurements on a quantum state, it is necessary that the state is entangled. This implies that the observations of non-local correlations means the state is entangled. Hence our use in (4.28). In a converse direction, it only true that all pure entangled states are non-local. This means for any entangled pure state one can obtain local measurements such that the measurement correlations violate the CHSH inequality. (The only pure states that do not violate it are product states.) However, there are entangled mixed states, such as (4.45), that do not violate the CHSH inequality. Therefore, not all entangled states are non-local.

In the language of quantum information, we can say that the interdependence of certain quantum information systems, that violate the CHSH inequality, would be impossible to replicate by classical information systems, which cannot violate the inequality. One of the utilities of this is that it allows one to detect entangled quantum information systems directly from measurement data without any reference to the physical experiment. This is known as device independence.

From the perspective of quantum foundations, non-locality across space suggests that for a subset of entangled cases, a quantum measurement on one system has the ability to *instantaneously* influence another system that can be arbitrarily spatially far. Hence an alternative perspective to gain is that it sheds a partial understanding on the non-trivial properties of certain quantum measurements. Furthermore, this instantaneous characteristic implies the *lack of a time interval* in this scenario. With a time interval involved, the non-local influence could be explained away by some hidden causal signal. Hence the *interdependence in this non-locality across space is shocking due to the absense of a time interval involved.*

The result also has a influence on philosophy, which can be highlighted by the subject being termed by some as 'experimental metaphysics' [185]. We provide a brief discussion. The decision to forgo the assumption of locality so to explain the experimental violation of the CHSH inequality is not based on any rigorous evidence. There is no mathematical or experimental proof to warrant such a decision. It may very well be the case that our concept of physical realism needs to be radically altered. From the perspective of this thesis, we find that there is more weight to the argument that one should drop the free will assumption. Entanglement in time already suggests the eternalists' view that the past and future are as real as the present. This provides an ideal scaffold to build a argument for the loss of free will, also known in this context as superdeterminism [186, 187].

c) Multipartite systems: The definition of Bell non-locality across space has been extended to more than two systems. Furthermore, it can be shown that all pure entangled n-partite states are are non-local [61].

Another important point to discuss within multipartite scenarios is what is referred to as the monogomy of entanglement [188, 189]. Let the left hand side of (5.85) be denoted

$$S^{AB}_{CHSH} \equiv \langle A_1 B_1 \rangle + \langle A_2 B_1 \rangle + \langle A_2 B_2 \rangle - \langle A_1 B_2 \rangle. \tag{5.87}$$

One property of this spatial non-locality is that a violation of the CHSH inequality precludes a simultaneous violation with another spatially separated system.

This is mathematically characterized as

$$S^{AB}_{CHSH} + S^{BC}_{CHSH} \leq 4, \tag{5.88}$$

for systems A, B, and C. A similar set of inequalities (5.88) hold for combinations (AB, AC) and (AC, BC).

d) Entropic version: In [190], an information-theoretic CHSH inequality was put forth. This provides a perspective in terms of systems storing information, as opposed to measurement correlations. They assumed the same scenario as in the original case involving the two spatially separated parties. Once again Alice has observables A_1 and A_2, whereas Bob has observables B_1 and B_2. These have respective values a_1, a_2, b_1 and b_2. The assumption of local realism (along with free will) is used to establish the existence of the joint probability $p(a_1, a_2, b_1, b_2)$. Using the Shannon conditional entropy (2.22), the information-theoretic CHSH inequality can be expressed as

$$\boxed{H(A_1|B_1) \leq H(A_1|B_2) + H(B_2|A_2) + H(A_2|B_1).} \tag{5.89}$$

To derive such a quantity, one makes use of the assumption that four objective quantities cannot carry less information than two of them,

$$H(A_1, B_1) \leq H(A_1, B_2, A_2, B_1). \tag{5.90}$$

Nevertheless, certain quantum entangled systems violate (5.89). An alternative entropic version can be found in [191].

5.2.2 CHSH game

a) Preliminaries: We have seen the use of guessing games in articulating the entropic uncertainty relations (3.128). More broadly the relationship between quantum theory and game theory is explored in [192, 193, 194]. Pertinent to this section is that Bell's theorem (CHSH inequality) have also been viewed through the lens of game. These are commonly referred to as nonlocal games, and the best

known example is the CHSH game which we will briefly describe below; in this scenario the participants can win the game at a higher probability with quantum resources, as opposed to having access to only classical resources. There has also been work on the relationship between Bell's theorem and Bayesian game theory [195, 196, 197]; in a subset of cases it was shown that quantum resources provide an advantage, and lead to quantum Nash equilibria. In [198], it was shown that quantum nonlocality can outperform classical strategies in games where participants have conflicting interests. However, in [199], a nonlocal game was constructed where quantum resources did not offer an advantage.

b) CHSH game: In this game, we consider spatially separated players Alice and Bob, as well as an outside party known as the referee that plays against Alice and Bob. Based on some probability distribution,

$$\pi : X \times Y \rightarrow [0, 1], \tag{5.91}$$

the referee chooses a question $x \in X$ for Alice and $y \in Y$ for Bob from some set of possible questions X and Y. With respect to the CHSH inequality, these questions can be thought of as labels for measurement settings. After receiving the questions, Alice and Bob respectively return answers $a \in R_A$ and $b \in R_B$ from some set of possible answers R_A and R_B. Relaying this to the CHSH inequality, one can view the answers as measurement outcomes. The referee is also tasked with deciding whether these answers are the winning answers for the questions that was posed according to the rules of the game. These rules are expressed through

$$V : R_A \times R_B \times X \times Y \rightarrow \{0, 1\}, \tag{5.92}$$

where $V(a, b, x, y) = 1$ if and only if Alice and Bob win against the referee by giving answers a and b for questions x and y. In this game, Alice and Bob have access to both the rules V and the probability distribution π. However, the constraint they face is that they cannot communicate once the game starts. This implies that each player is unaware of what question is given to the other player.

To see the direct relationship to the CHSH inequality (5.85), let $X = Y = \{0, 1\}$ and $R_A = R_B = \{0, 1\}$. The rules of the game are such that Alice and Bob win if

and only if

$$x \cdot y = a \oplus b, \tag{5.93}$$

where $\oplus$ represents modulo 2 addition. From this one can compute that the winning probability for a CHSH game is

$$p_{Win}^{CHSH} = \frac{1}{2}\left(1 + \frac{S}{4}\right), \tag{5.94}$$

where S is the CHSH expression (5.85). This provides us with an alternative view of the non-classical features of quantum resources. The probability that Alice and Bob win using only classical resources is at most probability 0.75, given $S \leq 2$. This is in contrast to utilizing quantum resources where Alice and Bob have the ability to win the game at a probability of almost 0.85 since $S = 2\sqrt{2}$.

5.2.3 PBR theorem

a) PBR theorem: From the perspective of this thesis, the PBR theorem demonstrates the discovery of a new quantum non-locality across space. However, the original intention of the theorem was to answer the foundational question: What is the nature of the quantum state (or quantum information)? The answer to this question can be aided by philosophical terminology. An ontic state refers to a state of reality meaning something that exists objectively in the world independent of an observer; it can be thought of as realism for the system in consideration. An epistemic state is a state of knowledge and refers to only what an observer currently knows about a physical system. The PBR theorem answers the question: Is the quantum state ontic or epistemic?

The mathematical characterization of these concepts is carried out through the framework of ontological models [200]. It can be thought of as a refinement of the hidden variable models found in the literature regarding Bell's theorem [201]. In the ontological model, when a system is prepared in some quantum state $|\Psi\rangle$, it is really in an ontic state λ, which describes a state of reality. The set of ontic states is denoted Λ. Due to our ignorance on what ontic state the system is in, the model assigns each quantum state $|\Psi\rangle$ an epistemic state μ_Ψ, which is a

probability distribution over Λ. These satisfy

$$\mu_\Psi(\lambda) \geq 0, \quad \text{and} \quad \int \mu_\Psi(\lambda) d\lambda = 1. \tag{5.95}$$

It also models a measurement and the outcome of that measurement in terms of the ontic state. For a measurement M we can denote the probability of obtaining outcome f in the state λ as $\xi_M(f|\lambda)$. These satisfy the conditions

$$\xi_M(f|\lambda) \geq 0, \quad \text{and} \quad \sum_f \xi_M(f|\lambda) = 1. \tag{5.96}$$

In order to reproduce the measurement predictions of quantum theory (3.75), we demand that

$$\int_\Lambda \xi_M(f|\lambda)\, \mu_\Psi(\lambda)\, d\lambda = |\langle f|\Psi\rangle|^2, \tag{5.97}$$

for all $|\Psi\rangle$ and f. It is important to emphasize that this ontological model includes standard quantum theory as a special case. Furthermore, note that the assumption of realism is implicit through the existence of an ontic state.

We now have the required tools to mathematically define what it means for a quantum state to be a state of reality or a state of knowledge. We say that an ontological model is Ψ-epistemic if there exists at least one pair of distinct quantum states $|\Psi_1\rangle$ and $|\Psi_2\rangle$, such that the corresponding epistemic states μ_{Ψ_1} and μ_{Ψ_2} have a non-zero overlap. If a model is not Ψ-epistemic, then it is Ψ-ontic.

When we say non-zero overlap we mean,

$$1 - \delta_C(\mu_{\Psi_1}, \mu_{\Psi_2}) > 0, \tag{5.98}$$

where the classical trace distance is defined as

$$\delta_C(p, q) \equiv \frac{1}{2} \int |p(x) - q(x)| dx, \tag{5.99}$$

for probability distributions $p(x)$ and $q(x)$. The underlying idea is that if there is no overlap in the epistemic states then distinct quantum states refer to distinct ontic states, thereby warranting the quantum state itself as a state of reality.

However if there is an overlap in the epistemic states, then a single ontic state can relate to two different quantum states through the two respective epistemic states. Hence, a unique quantum state cannot be associated with the ontic state. In this case, a quantum state signifies itself merely as a state of knowledge.

The aim of the PBR theorem is to show that models must be Ψ-ontic. The proof for the PBR theorem starts by assuming a Ψ-epistemic model and then arriving at a contradiction. More precisely suppose that for two quantum states $|\Psi_1\rangle$ and $|\Psi_2\rangle$, the corresponding epistemic states μ_{Ψ_1} and μ_{Ψ_2} overlap. This implies that there exists an ontic state $\lambda_* \in \Lambda$ where

$$\mu_{\Psi_1}(\lambda_*) > 0, \quad \text{and} \quad \mu_{\Psi_2}(\lambda_*) > 0. \tag{5.100}$$

In this case, even if one had access to the underlying ontic state λ_*, it would be impossible to tell which of the two quantum states was prepared. Alternatively, regardless of which of these quantum states were prepared, the ontic state λ_* will be occupied a non-zero fraction $P_* > 0$ of the time (where the value of P_* does not need to be specified). Next, let two copies of the system be prepared in one of the four quantum (separable or product) states,

$$\begin{aligned} |\Psi_{11}\rangle &= |\Psi_1\rangle \otimes |\Psi_1\rangle\,, \quad |\Psi_{12}\rangle = |\Psi_1\rangle \otimes |\Psi_2\rangle\,, \\ |\Psi_{21}\rangle &= |\Psi_2\rangle \otimes |\Psi_1\rangle\,, \quad |\Psi_{22}\rangle = |\Psi_2\rangle \otimes |\Psi_2\rangle\,. \end{aligned} \tag{5.101}$$

These two systems are prepared spatially far apart, and the choice to prepare either $|\Psi_1\rangle$ or $|\Psi_2\rangle$ is made independently at each spatial location. For this task, we make use of the assumption of *preparation independence*. This comprises of two components. The first is that each system obtains its own copy of Λ. The total state space of the two systems is the product of two copies of Λ, and therefore the ontic states are written as

$$\boxed{(\lambda_1 \times \lambda_2) \in \Lambda \times \Lambda.} \tag{5.102}$$

This implies that the quantum state $|\Psi_{jk}\rangle$ corresponds to epistemic state $\mu_{\Psi_{jk}}(\lambda_1, \lambda_2)$ and that joint measurements take the form $\xi_M(\Phi|\lambda_1, \lambda_2)$ for some vector $|\Phi\rangle$ in a

measurement basis. The second component is that the epistemic state $\mu_{\Psi_{jk}}(\lambda_1, \lambda_2)$ associated with quantum state $|\Psi_{jk}\rangle$ factorizes as

$$\boxed{\mu_{\Psi_{jk}}(\lambda_1, \lambda_2) = \mu_{\Psi_j}(\lambda_1)\mu_{\Psi_k}(\lambda_2)} \tag{5.103}$$

where $\mu_{\Psi_j}(\lambda_1)$ is the epistemic state for $|\Psi_j\rangle$ and $\mu_{\Psi_k}(\lambda_2)$ for $|\Psi_k\rangle$. Notice the resemblance to (2.9) and (4.2) through its factorization.

To illustrate the core points, let us first consider the simple case of qubits. Suppose that $|\Psi_1\rangle = |0\rangle$ and $|\Psi_2\rangle = |+\rangle \equiv (|0\rangle + |1\rangle)/\sqrt{2}$. If we prepare the two systems in one of the four states $|\Psi_{jk}\rangle$ and use preparation independence, then at least P_*^2 of the time we arrive at the situation that total system will be in ontic state $(\lambda_{*,1}, \lambda_{*,2})$. In this scenario, both systems are in ontic state λ_*; it will be impossible to decide whether $|\Psi_1\rangle$ or $|\Psi_2\rangle$ was prepared. Next we introduce the following two-qubit measurement using basis

$$\begin{aligned}
|\Phi_{11}\rangle &= \tfrac{1}{\sqrt{2}}(|0\rangle \otimes |1\rangle + |1\rangle \otimes |0\rangle), \\
|\Phi_{12}\rangle &= \tfrac{1}{\sqrt{2}}(|0\rangle \otimes |-\rangle + |1\rangle \otimes |+\rangle), \\
|\Phi_{21}\rangle &= \tfrac{1}{\sqrt{2}}(|+\rangle \otimes |1\rangle + |-\rangle \otimes |0\rangle), \\
|\Phi_{22}\rangle &= \tfrac{1}{\sqrt{2}}(|+\rangle \otimes |-\rangle + |-\rangle \otimes |+\rangle),
\end{aligned} \tag{5.104}$$

where $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$. The four states are antidistinguishable (3.106) using this measurement basis. In other words, we have that

$$|\langle\Phi_{jk}|\Psi_{jk}\rangle|^2 = 0 \tag{5.105}$$

for every choice of $j, k = 1, 2$. We see therefore that quantum theory predicts that the measurement outcome $|\Phi_{jk}\rangle$ never occurs when the quantum state $|\Psi_{jk}\rangle$ is prepared. Referring back to the ontological model we have the certainty that whichever quantum state is prepared, a fraction P_*^2 of the time the system is in ontic state $(\lambda_{*,1}, \lambda_{*,2})$. We also see that if the system is in this ontic state, we may get outcome $|\Phi_{jk}\rangle$ when we measure in basis (5.104). Moreoever, this ontic state occurs when the quantum state $|\Psi_{jk}\rangle$ is prepared a non-zero fraction of the time. However to reproduce the predictions of quantum theory (5.105), the

measurement outcome $|\Phi_{jk}\rangle$ should never occur for this ontic state. Therefore, a non-zero fraction of the time the measurement device contradicts the predictions of quantum theory. This provides us with the desired contradiction. This argument can be generalized using similar concepts (as described below). Therefore, the PBR theorem can be stated as: Any ontological model that reproduces quantum predictions and satisfies preparation independence is Ψ-ontic.

b) Implications: An equally weighted perspective is that at least one of the assumptions used to arrive at the contradiction must be false. This allows for the position (held by most physicists) that the quantum state is simply a state of knowledge (Ψ-epistemic); such a view is desirable in dissolving away many conundrums including the measurement collapse which is only a problem if the quantum state has a physical existence. To decipher which assumption must be incorrect, we relay our thoughts back to the CHSH inequality. It was of consensus in that scenario to maintain realism and adopt non-locality. In the PBR case, an analogous choice is to therefore adopt the *violation of preparation independence*. This resulting effect can be described as a new type of non-locality across space, and mathematically characterized as a violation of (5.102) and/or (5.103).

The violation of preparation independence is a far more perplexing spatial interdependence than the Bell non-locality across space. First, it applies to product states, and hence does not require entanglement (unlike Bell non-locality). The second point to note is that preparation independence is perhaps the most natural assumption to make in that two spatially separated systems should possess their own separate states of reality; such a notion of separability should be natural for product states. As an example, if one person prepares a system in one part of the universe and another person prepares a system in the other part of the universe, then there should be no correlations between these preparations; if this was not the case as insisted above, then an extrapolation on this effect is that one requires every system in the universe in order to determine all the parameters that are of relevance for a system prepared on Earth. A non-locality of such magnitude would lead to a radical destruction of basic assumptions.

Einstein wrote [202, 203] about the dangers of abandoning such assumptions (but within another context), and we quote "*Further, it appears to be essential for*

this arrangement of the things introduced in physics that, at a specific time, these things claim an existence independent of one another, insofar as these things 'lie in different parts of space.' Without such an assumption of the mutually independent existence of spatially distant things, an assumption which originates in everyday thought, physical thought in the sense familiar to us would not be possible." From the perspective of the theme of this thesis, notice that Einstein mentions the words *specific time* which signifies the *lack of a time interval*. Whether we accept the quantum state as a state of reality, or instead give up realism, or introduce the violation of preparation independence, the PBR theorem has most certainly emphasized the large gap in our fundamental understanding of quantum physics.

c) Multipartite systems: Using certain assumptions, we have shown that the epistemic states for $|0\rangle$ and $|+\rangle$ cannot overlap. Generalizing this to any pair of quantum states implies that a quantum state can uniquely correspond to an ontic state, thereby signifying itself as a physical property of the system. To prove this, we can let

$$|\Psi_0\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + \sin\left(\frac{\theta}{2}\right)|1\rangle\,, \tag{5.106}$$

$$|\Psi_1\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle - \sin\left(\frac{\theta}{2}\right)|1\rangle\,, \tag{5.107}$$

represent arbitrary non-orthogonal qubits, where $0 < \theta < \pi/2$. As in the previous case, suppose there is a non-zero probability of at least P_* that the ontic state of the system is compatible with either preparation. This means the corresponding epistemic states overlap. Suppose we prepare n of these systems independently. The total system can be described by one of the quantum states,

$$|\Psi(x_1, \ldots x_n)\rangle = |\Psi_{x_1}\rangle \otimes \cdots \otimes |\Psi_{x_{n-1}}\rangle \otimes |\Psi_{x_n}\rangle\,, \tag{5.108}$$

where $x_i \in \{0, 1\}$ for each i. Assuming preparation independence, we have the probability that at least P_*^n that the ontic state is compatible with any one of the quantum states. The contradiction to quantum theory is obtained if we can derive a measurement basis that makes these quantum states antidistinguishable. This

can be achieved if the number of systems n satisfies

$$2\arctan\left(2^{1/n}-1\right) \leq \theta. \tag{5.109}$$

Furthermore, the exact measurement circuit consists of a unitary evolution,

$$U_{\alpha,\beta} = H^{\otimes n} R_\alpha {Z_\beta}^{\otimes n}, \tag{5.110}$$

where

$$Z_\beta = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\beta} \end{pmatrix}. \tag{5.111}$$

and where H is Hadamard gate (3.49); the gate R_α operates as $R_\alpha |0\cdots 0\rangle = e^{i\alpha} |0\cdots 0\rangle$ and acts as an identity operator on the other computational basis states. To achieve the desired result the unitary evolution is chosen based on certain values of α and β. This is followed by a measurement of each qubit in the computational basis states (3.4). The result is that each outcome has zero probability given one of the 2^n possible preparations.

More precisely, the probability of obtaining the basis state $|x_1, \ldots x_n\rangle$ given that the state $|\Psi(x_1, \ldots x_n)\rangle$ is prepared is the squared absolute value of

$$\langle x_1 \ldots x_n| H^{\otimes n} R_\alpha Z_\beta^{\otimes n} |\Psi_{x_1}\rangle \otimes \cdots \otimes |\Psi_{x_n}\rangle \tag{5.112}$$

which can be shown to equate to

$$\frac{1}{\sqrt{2^n}} \left(\cos\frac{\theta}{2}\right)^n \left(e^{i\alpha} + \left(1 + e^{i\beta}\tan\frac{\theta}{2}\right)^n - 1\right). \tag{5.113}$$

Moreover, α and β can be derived so that

$$e^{i\alpha} + \left(1 + e^{i\beta}\tan\frac{\theta}{2}\right)^n - 1 = 0. \tag{5.114}$$

Hence, the required quantum probabilities are zero, and as a result we found a measurement that provides antidistinguishability for these quantum states.

d) Entropic version: From the perspective of quantum information science,

it also interesting to note that the PBR theorem been interpreted through the language of classical and quantum communication protocols [204, 205]. This program crucially involved the use of the Shannon entropy (2.20). In a related work antidistinguishability, which is a core concept in the PBR theorem, was used to provide an advantage in a two-player communication task [206].

Other notable developments on the PBR theorem and Ψ-epistemic models have been carried out in [207, 208, 209, 210, 211, 212] including on the issue of quantum indistinguishability [213, 214, 215].

5.2.4 PBR theorem as a Monty Hall game

a) Preliminaries: Analogous to the game formulation of CHSH inequality, a desirable construction is to view the PBR theorem through the lens of a game. One instantiation of this is in an exclusion game where the participant's goal is to produce a particular bit string [216, 217]; this has been shown to be related to the task of quantum bet hedging [218]. Furthermore, concepts involved in the PBR proof have been used for a particular guessing game [219].

In this subsection, we reformulate the PBR theorem into a Monty Hall game [71], which is *part of the original component of this thesis* (which as done in collaboration with my supervisor). This particular gamification of the theorem highlights that winning probabilities, for switching doors in the game, depend on whether it is a Ψ-ontic or Ψ-epistemic game; we also show that in certain Ψ-epistemic games switching doors provides no advantage. This may have consequences for an alternative experimental test of the PBR theorem

b) PBR elements: We extract certain parts of the PBR proof. Recall that two quantum systems are prepared independently, and each system is prepared in either state $|0\rangle$ or state $|+\rangle$. This means that the total system is in one of the four possible non-orthogonal quantum states (5.101) which we rewrite as:

$$\begin{aligned} |\Psi_1\rangle &= |0\rangle \otimes |0\rangle\,, \quad |\Psi_2\rangle = |0\rangle \otimes |+\rangle\,, \\ |\Psi_3\rangle &= |+\rangle \otimes |0\rangle\,, \quad |\Psi_4\rangle = |+\rangle \otimes |+\rangle\,. \end{aligned} \tag{5.115}$$

The total system is brought together and measured in the basis (5.104) which we re-label as:

$$\begin{aligned} |\Phi_1\rangle &= \tfrac{1}{\sqrt{2}}(|0\rangle \otimes |1\rangle + |1\rangle \otimes |0\rangle), \\ |\Phi_2\rangle &= \tfrac{1}{\sqrt{2}}(|0\rangle \otimes |-\rangle + |1\rangle \otimes |+\rangle), \\ |\Phi_3\rangle &= \tfrac{1}{\sqrt{2}}(|+\rangle \otimes |1\rangle + |-\rangle \otimes |0\rangle), \\ |\Phi_4\rangle &= \tfrac{1}{\sqrt{2}}(|+\rangle \otimes |-\rangle + |-\rangle \otimes |+\rangle), \end{aligned} \tag{5.116}$$

Invoking the Born probabilities, $|\langle\Phi_i|\Psi_h\rangle|^2$, where $i, h = 1, 2, 3, 4$, we found that for $i = h$, $|\langle\Phi_i|\Psi_i\rangle|^2 = 0$. This means that for any value i, the outcome $|\Phi_i\rangle$ never occurs when the system is prepared in quantum state $|\Psi_i\rangle$. The PBR proof showed that in Ψ-epistemic models there is a non-zero probability q (whose value does not need to be specified) that outcome $|\Phi_i\rangle$ occurs when state $|\Psi_i\rangle$ is prepared, thereby contradicting the predictions of quantum theory; hence one can infer that the quantum state corresponds to a Ψ-ontic model.

c) Ψ-ontic Monty Hall game: Antidistinguishability, where there is a measurement for which each outcome identifies that a specific member of a set of quantum states was definitely not prepared, is highlighted in the PBR proof by $|\langle\Phi_i|\Psi_i\rangle|^2 = 0$ for all i. We will exploit this to construct our game, which can be thought of as a quantum Ignorant Monty Hall game (2.16).

For state $|\Psi_1\rangle$ in (5.115), we have

$$\begin{aligned} |\langle\Phi_1|\Psi_1\rangle|^2 &= 0, \quad |\langle\Phi_2|\Psi_1\rangle|^2 = 1/4, \\ |\langle\Phi_3|\Psi_1\rangle|^2 &= 1/4, \quad |\langle\Phi_4|\Psi_1\rangle|^2 = 1/2. \end{aligned} \tag{5.117}$$

For the other states in (5.115), the same probability distribution (0, 1/4, 1/4, 1/2) occur but across the different outcomes (5.116); hence we will focus our game on $|\Psi_1\rangle$, but similar constructions hold for the other states.

The Monty Hall gamification is as follows: There are four doors labelled $\{1, 2, 3, 4\}$, and these correspond to the different measurement outcomes listed in (5.116). The prize door A_i, where i takes one of the door labels, is the outcome $|\Phi_i\rangle$ that the state $|\Psi_1\rangle$ collapses to upon measurement. For a Ψ-ontic game, through the

Born probabilities (5.117), we have $P(A_i) = |\langle\Phi_i|\Psi_1\rangle|^2$.

The contestant on the show does not know what state from (5.115) is used, and is only aware of the possible measurement outcomes (5.116). Based on this limited information, the contestant randomly picks one of the doors which we denote B_j where j is the corresponding door label; hence we have $P(B_j|A_i) = 1/4$, for all values i, j.

Monty's decision corresponds to the predictions of quantum theory. He is aware that state $|\Psi_1\rangle$ was used, and has access to the Born probabilities (5.117). The door opened by Monty is denoted C_k where k is one of the door labels. The main insight to construct this game is that when Monty opens a goat door, he is opening a door that has probability zero of having a prize in it. And for our game, a door that definitely does not have a prize in it corresponds to outcome $|\Phi_1\rangle$ as $P(A_1) = |\langle\Phi_1|\Psi_1\rangle|^2 = 0$. Hence in this game, Monty will open door C_1 unless the contestant has already chosen this door as their pick (as Monty cannot open the door chosen by the contestant); in that case Monty will open one of the other remaining doors with equal probability, and there is a chance he may open up the prize door as in the Ignorant Monty Hall game. From these factors, one can compute,

$$P(C_k \mid B_j \cap A_i) = \begin{cases} \frac{1}{3}, & \text{if } j = 1 \text{ and } k = 2, 3, 4, \\ 1, & \text{if } j \neq 1 \text{ and } k = 1, \\ 0, & \text{otherwise,} \end{cases} \tag{5.118}$$

where we have adopted the notation for joint probabilities as $P(A, B) \equiv P(A \cap B)$. The probability that Monty opens the prize door is

$$P(\text{opens prize door}) = \sum_{i=k\neq j} P(A_i \cap B_j \cap C_k) = \frac{1}{12}. \tag{5.119}$$

This implies that the probability that he opens a goat door is 11/12. Monty then offers the option to stick or switch. Suppose the contestant always sticks with the initial choice. Then the probability of winning if sticking and Monty opening

a goat door is

$$\sum_{i=j\neq k} P(A_i \cap B_j \cap C_k) = \frac{1}{4}. \tag{5.120}$$

With that, we can compute the conditional probability

$$P(\text{win if stick} \,|\, \text{opens goat door}) = \frac{1/4}{11/12} = \frac{3}{11}. \tag{5.121}$$

Suppose the contestant decides to always switch to one of the other two unopened doors with equal probability $1/2$. Let $|\Phi_l\rangle$ be the outcome switched to and let D_l be the corresponding door. With that, we can compute $P(A_i \cap B_j \cap C_k \cap D_j) = P(D_l \,|C_k \cap B_j \cap A_i) P(C_k \,|\, B_j \cap A_i) P(B_j|A_i) P(A_i)$. Hence, the probability of winning if switching and Monty opening a goat door is

$$\sum_{i=l\neq j\neq k} P(A_i \cap B_j \cap C_k \cap D_j) = \frac{1}{3}. \tag{5.122}$$

From that, one can calculate

$$P(\text{win if switch} \,|\, \text{opens goat door}) = \frac{1/3}{11/12} = \frac{4}{11}. \tag{5.123}$$

Hence, we see in a Ψ-ontic game, switching provides an advantage.

d) Ψ-epistemic Monty Hall game: In the PBR proof, for the Ψ-epistemic model, there is a non-zero probability q that outcome $|\Phi_1\rangle$ occurs when state $|\Psi_1\rangle$ is prepared. This implies that in a ψ-epistemic game, $P(A_1) = q \neq 0$. To allow for a comparison with the Ψ-ontic game, let $q = q_1 + q_2 + q_3$, and with that let the other prize door probabilities take values $P(A_2) = (1/4) - q_1$, $P(A_3) = (1/4) - q_2$ and $P(A_4) = (1/2) - q_3$.

As in the ψ-ontic game, $P(B_j|A_i) = 1/4$, for all values i, j. Monty as a character corresponds to the predictions of quantum theory (5.117); he will assume C_1 is definitely a goat door since $|\langle\Phi_1|\Psi_1\rangle|^2 = 0$. This means the probabilities in (5.118) apply in this game as well. Hence, the probability that Monty opens the prize

door

$$P(\text{opens prize door}) = \sum_{i=k\neq j} P(A_i \cap B_j \cap C_k) = \frac{1}{12} + \frac{2q}{3}. \tag{5.124}$$

This implies that the probability that Monty opens a goat door is $(11/12)-(2q/3)$. The probability of winning if always sticking and that Monty opens a goat door is

$$\sum_{i=j\neq k} P(A_i \cap B_j \cap C_k) = \frac{1}{4}. \tag{5.125}$$

From this we compute

$$P(\text{win if stick} \mid \text{opens goat door}) = \frac{3}{11-8q}. \tag{5.126}$$

If a switching strategy is adopted then:

$$\sum_{i=l\neq j\neq k} P(A_i \cap B_j \cap C_k \cap D_j) = \frac{1}{3} - \frac{q}{3}, \tag{5.127}$$

$$P(\text{win if switch} \mid \text{opens goat door}) = \frac{4-4q}{11-8q}. \tag{5.128}$$

Thus the probabilities depend on whether the game is a Ψ-ontic or Ψ-epistemic game. For value $q = 1/4$, we can calculate that $P(\text{win if switch} \mid \text{opens goat door}) = P(\text{win if stick} \mid \text{opens goat door})$; hence for certain Ψ-epistemic games, switching offers no advantage.

e) Experimental implications: Comparing a Ψ-ontic game to a Ψ-epistemic game, Monty opens the prize door less often. This corresponds to certain probabilities in the PBR proof being zero; some work on the experimental tests [180, 220, 221, 222, 223] of PBR discuss this exact zero probability as an experimental difficulty. Through our game, we provide another viewpoint; the difference in the probabilities of winning conditioned that a goat door is opened are simply different for the two physical scenarios. This may provide insights to alternative experimental designs to test PBR.

5.3 Non-locality across Time

Our aim is to explore how the effects of non-locality across space extend into the temporal regime. In particular, one can qualitatively describe a non-locality across time as a characteristic where an action on a subsystem can instantaneously influence the same subsystem at a later or earlier time! We have seen this shocking property portrayed in the section regarding entanglement in time. However in this chapter, our focus be on the case of a single system across multiple times. This non-locality across time will be expressed mathematically through the violation of a temporal version of the Bell-CHSH inequality. The most prominent of these are known as Leggett-Garg (LG) inequalities [224].

Some refer to the effects, that we shall describe, as an 'entanglement in time.' However in this thesis we thread carefully and refer to these effects exclusively as a non-locality across time. There are three reasons for this taxonomy and they stem from the extensive review of the spatial case. The first reason is that the LG and related temporal inequalities are about measurement outcomes, and not about an algebraic property within the mathematics of quantum theory; in the spatial case, measurement outcomes related directly to non-locality, whereas the algebraic property defined the concept of entanglement; setting the spatial case as precedence allows us to forgo using the words entanglement in time to describe these scenarios of (temporal) measurement correlations. The second and perhaps the more cautious reason is that not all spatially entangled states are spatially Bell non-local; hence assuming the nature of the relationship between an entanglement in time and non-locality across time prior to rigorous results is not very prudent. The third reason is that the necessary algebraic construction to help define an entanglement in time between a single system over various times is met with considerable technical problems [225]. It must therefore be emphasized that there is a large degree of unknown aspects to this area. However, for an extensive review on LG inequalities, refer to [226].

In the spatial case, we reviewed Bell-CHSH, PBR and its games. In this section we will articulate non-locality across time using the concepts in Bell-CHSH, PBR, and games. This serves to provide a systematic view into the subject.

5.3.1 Through Bell-CHSH concepts

a) LG inequalities: The LG inequalities [224] can be thought of as a temporal version of the Bell-CHSH inequalities (5.85). It was derived within the context of macroscopic coherence which can be thought of as property of an object, consisting of many quantum particles, existing in superpositions of macroscopically distinct states. (A fictional example is the Schrödinger's cat.) The result largely follows the same style of derivation that was used in the spatial Bell-CHSH case. We start with the following classically intuitive assumptions:

i) (A1) Macroscopic Realism (MR): a macroscopic system with two or more macroscopically distinct states available to it will at all times be in one or the other of these states.
ii) (A2) Non-Invasive Measurability (NIM) at the macroscopic level: it is possible, in principle, to determine the state of the system with arbitrarily small perturbation on its subsequent dynamics.
iii) (A3) Induction: the outcome of a measurement on the system cannot be affected by what will or will not be measured on it later.

NIM has also been described as that a measurement of an observable at any instant of time does not influence its subsequent evolution [227]. NIM has also been described as nondisturbance in that a measurement can be performed such that it does not influence the outcome of a measurement on the same system at a later time [228]. Another temporal Bell-CHSH inequality [130] was more direct to replace NIM with the assumption that the results of measurements performed at some time is independent of any other measurement at another time; they referred to this as a locality in time. The assumptions used in the LG inequalities are still of great debate. In this thesis, we view both (A2) and (A3) as the assumption of locality in time, with (A1) taking the role of realism. Notice the resemblance with the Bell-CHSH case where it was a locality in space paired with realism.

Using these assumptions, we can define a macroscopic dichotomic variable $Q = \pm 1$ for a system. We aim to measure its two-time correlation function

$$C_{ij} = \langle Q(t_i)\, Q(t_j) \rangle. \tag{5.129}$$

This quantity is computed from the joint probability $P_{ij}(Q_i, Q_j)$ of obtaining $Q_i = Q(t_i)$ and $Q_j = Q(t_j)$ from measurement times t_i, t_j as

$$C_{ij} = \sum_{Q_i, Q_j = \pm 1} Q_i \, Q_j \, P_{ij}(Q_i, Q_j). \tag{5.130}$$

The assumption (A1) implies the observable has a defined value at all times regardless of whether it is measured. Hence one can obtain a two-time probability as the marginal of a three-time distribution as follows

$$P_{ij}(Q_i, Q_j) = \sum_{Q_k; k \neq i,j} P_{ij}(Q_3, Q_2, Q_1). \tag{5.131}$$

Using (A2) and (A3), we find that the three probabilities $P_{21}(Q_3, Q_2, Q_1)$, $P_{32}(Q_3, Q_2, Q_1)$ and $P_{31}(Q_3, Q_2, Q_1)$ become the same. Hence we can write this simply as

$$P(Q_3, Q_2, Q_1) = P_{ij}(Q_3, Q_2, Q_1). \tag{5.132}$$

One can proceed to use this single probability to compute the following correlation functions:

$$\begin{aligned} C_{21} = {} & P(+1,+1,+1) - P(+1,+1,-1) - P(-1,-1,+1) + P(-1,-1,-1) \\ & - P(+1,-1,+1) + P(+1,-1,-1) + P(-1,+1,+1) - P(-1,+1,-1); \end{aligned} \tag{5.133}$$

$$\begin{aligned} C_{32} = {} & P(+1,+1,+1) + P(+1,+1,-1) + P(-1,-1,+1) + P(-1,-1,-1) \\ & - P(+1,-1,+1) - P(+1,-1,-1) - P(-1,+1,+1) - P(-1,+1,-1); \end{aligned} \tag{5.134}$$

$$\begin{aligned} C_{31} = {} & P(+1,+1,+1) - P(+1,+1,-1) - P(-1,-1,+1) + P(-1,-1,-1) \\ & + P(+1,-1,+1) - P(+1,-1,-1) - P(-1,+1,+1) + P(-1,+1,-1); \end{aligned} \tag{5.135}$$

Given that

$$\sum_{Q_3, Q_2, Q_1} P(Q_3, Q_2, Q_1) \equiv 1, \tag{5.136}$$

this implies

$$\begin{aligned} K_3 &\equiv C_{21} + C_{32} - C_{31} \\ &= 1 - 4(P(+1,-1,+1) + P(-,+,-)). \end{aligned} \tag{5.137}$$

If $P(+1,-1,+1) = P(-,+,-) = 0$, then $K_3 = 1$ which is the upper bound. On the other hand, the choice $P(+1,-1,+1) + P(-,+,-) = 1$ gives lower bound $K_3 \geq -3$. From this, we obtain the simplest LG inequality,

$$\boxed{-3 \leq K_3 \leq 1.} \tag{5.138}$$

This LG inequality has been violated through various quantum systems. It can be shown that the maximum violation by a two-level quantum system (qubit) is $K_3^{max} = 3/2$; at least one of the three assumptions made to derive the LG inequality is wrong.

b) Implications: If one takes the spatial Bell-CHSH case as an analogy but also as precedence, then we leave (A1) intact. This means that assumptions (A2) and (A3), which express locality in time, are incorrect. Hence the mathematical characterization of *non-locality across time* is expressed as the violation of (5.138).

c) Multi-measurements: The LG inequality has been extended to n-meaurements. Let us denote the variable,

$$K_n = C_{21} + C_{32} + C_{43} + \cdots + C_{n(n-1)} - C_{n1}. \tag{5.139}$$

Using the assumptions (A1-3), one can obtain the following LG inequalities

$$\boxed{\begin{aligned} -n \leq K_n \leq n-2 \qquad & n \geq 3, \text{ odd}; \qquad (5.140) \\ -(n-2) \leq K_n \leq n-2 \qquad & n \geq 4, \text{ even}, \qquad (5.141) \end{aligned}}$$

where the only requirement on the variable is to be bounded $|Q| \leq 1$. Using various symmetry properties one derive further inequalities. One in particular

is written as

$$\boxed{-2 \leq C_{21} + C_{32} + C_{43} - C_{41} \leq 2.} \tag{5.142}$$

Note that (5.142) and other LG inequalities describe a situation where there are a set of measurements on the same operator at $n \geq 3$ different times. There is in fact another temporal Bell-CHSH inequality [130] that considers a different physical scenario. In this case, there are two different times but different operator choices at each time. More precisely, in this scenario Alice measures at time t_1 while Bob measures at time $t_2 > t_1$. These measurements involve dichotomic variables. Each of them have two measurement choice ($i = 1, 2$), which can be denoted A_i and B_i for Alice and Bob respectively. Using assumptions (A1-3) where the locality of time assumption was explicity stated, one can derive the following temporal CHSH inequality,

$$\boxed{|\langle B_1 A_1\rangle + \langle B_1 A_2\rangle + \langle B_2 A_1\rangle - \langle B_2 A_2\rangle| \leq 2.} \tag{5.143}$$

Despite the physical differences, the equation (5.143) can be obtained directly from the LG inequality (5.142) by setting

$$Q(t_1) = B_2, \quad Q(t_2) = A_1, \quad Q(t_3) = B_1, \quad Q(t_4) = A_2. \tag{5.144}$$

Once again the violation of (5.143) provides a mathematical characterization of a *non-locality across time.* A qubit can be shown to violate (5.143) with a maximum value of $2\sqrt{2}$. Notice the resemblance of temporal CHSH case (5.143) to the spatial CHSH case (5.85).

d) Entropic version: We have seen an entropic version (5.89) of the spatial CHSH inequality. A natural question to consider is whether such a possibility exists for the temporal LG case. Such a curiosity has been answered in the affirmative in the works by [227, 131]. We provide a derivation of this entropic LG inequalities which utilizes the Shannon entropy (2.20).

In the LG scenario, we have a macroscopic system where $Q(t_i)$ represents an observable at time t_i. Let the outcome be denoted q_i with corresponding probability $P(q_i)$. Using assumption (A1), we have the existence of a joint probability

distribution $P(q_1, q_2, \ldots)$ due to the notion that the outcomes of observable at all instants of time exist whether the system has been measured or not. Using (A2) and (A3), we have the result that measurement at an earlier time t_i has no influence on the value at a subsequent time $t_j > t_i$; this implies that joint probabilities are written as convex combinations involving a hidden variable probability distribution $\rho(\lambda)$,

$$P(q_1, q_2, \ldots, q_n) = \sum_{\lambda} \rho(\lambda)\, P(q_1|\lambda) P(q_2|\lambda) \ldots P(q_n|\lambda) \tag{5.145}$$

where

$$0 \leq \rho(\lambda) \leq 1, \quad \sum_{\lambda} \rho_{\lambda} = 1, \tag{5.146}$$

and

$$0 \leq P(q_i|\lambda) \leq 1 \quad \sum_{q_i} P(q_i|\lambda) = 1. \tag{5.147}$$

One can harness the joint Shannon entropy (2.21) to an observable at two different times t_k and t_{k+l}, resulting in

$$H(Q_k, Q_{k+l}) = -\sum_{q_k, q_{k+l}} P(q_k, q_{k+l})\, \log_2\, P(q_k, q_{k+l}). \tag{5.148}$$

Using the conditional Shannon entropy (2.22), we can examine the information held by observable Q_{k+l} at time t_{k+l} given it had the values $Q_k = q_k$ at a previous time t_k. This can be shown to equate to

$$H(Q_{k+l}|Q_k = q_k) = -\sum_{q_{k+l}} P(q_{k+l}|q_k)\, \log_2\, P(q_{k+l}|q_k), \tag{5.149}$$

where the conditional probability is expressed as

$$P(q_{k+l}|q_k) = \frac{P(q_k, q_{k+l})}{P(q_k)}. \tag{5.150}$$

From this, one can easily derive the full conditional Shannon entropy,

$$H(Q_{k+l}|Q_k) = \sum_{q_k} P(q_k)\, H(Q_{k+l}|Q_k = q_k) \tag{5.151}$$

$$= H(Q_k, Q_{k+l}) - H(Q_k). \tag{5.152}$$

Re-arranging this, we obtain

$$H(Q_k, Q_{k+l}) = H(Q_{k+l}|Q_k) + H(Q_k) \tag{5.153}$$

One further set of inequalities that will be of use is given by the properties intrinsic to the Shannon entropy

$$H(Q_{k+l}|Q_k) \leq H(Q_{k+l}) \leq H(Q_k, Q_{k+l}), \tag{5.154}$$

where the right-hand inequality signifies that two variables can never hold less information than held by one of them. By combining (5.153) and (5.154) and extending it to three variables, we obtain

$$H(Q_k, Q_{k+m}) \leq \quad H(Q_k, Q_{k+l}, Q_{k+m}) = H(Q_{k+m}|Q_{k+l}, Q_k) + H(Q_{k+l}|Q_k) + H(Q_k). \tag{5.155}$$

This results in the entropic LG relation

$$\boxed{H(Q_{k+m}|Q_k) \leq H(Q_{k+m}|Q_{k+l}) + H(Q_{k+l}|Q_k),} \tag{5.156}$$

for times $t_k < t_{k+l} < t_{k+m}$. A similar line of argument allows one to obtain an n-measurement entropic LG inequality,

$$\boxed{H(Q_n|Q_1) \leq H(Q_n|Q_{n-1}) + H(Q_{n-1}|Q_{n-2}) + \ldots + H(Q_2|Q_1),} \tag{5.157}$$

for consecutive measurements $Q_1, Q_2, \ldots, Q_n$ for the various times $t_1 < t_2 < \ldots < t_n$.

Once again a violation of (5.156) or (5.157) is a mathematical characterization of

non-locality across time. Such a violation has been exhibited by quantum systems. Of great interest to this thesis is that *the interdependence of this non-locality across time is shocking due to the existence of a time interval.* To elaborate, let us consider the equation (5.156). It suggests the information content of the observable at three different times $t_k < t_{k+l} < t_{k+m}$ can never be smaller than the information content at two time instants. A quantum violation suggests the perplexing narrative that in fact having the knowledge of an observable at three different times corresponds to less information than knowing the observable at only two different times! From the view of this thesis, the added time interval to introduce the third time point makes this interdependence across time truly shocking.

5.3.2 Through PBR concepts

a) LG inequalities: One can provide an alternative derivation of the LG inequalities (5.138) using ontological models [200, 182, 226]. Recall the use of this framework in proving the PBR theorem. In this section, our aim is use these models to compute the correlation functions

$$C_{ij} = \sum_{Q_i, Q_j = \pm 1} Q_i \, Q_j \, P_{ij}(Q_i, Q_j), \tag{5.158}$$

and thereby re-derive the LG inequality. We start by describing the state of the system as outlined in (5.95). This is denoted by an epistemic state $\mu(\lambda)$ over the set of ontic states λ. Note that the ontic states capture assumption (A1). Next, a measurement (5.96) at time t_i is represented as

$$\xi_i(Q_i|\lambda), \tag{5.159}$$

which signifies the probability that of outcome Q_i given ontic state λ. We denote the probability of disturbance by the measurement on the ontic state $\lambda \to \lambda'$ as

$$\gamma_i(\lambda'|Q_i, \lambda). \tag{5.160}$$

In the ontological framework, the joint probability function of two measurements is then written as

$$P(Q_i, Q_j) = \int d\lambda' \, d\lambda \, \xi_j(Q_j|\lambda') \, \gamma_i(\lambda'|Q_i, \lambda) \, \xi_i(Q_i|\lambda) \, \mu(\lambda). \tag{5.161}$$

Using (A2) and (A3), we have the condition that the disturbance does not affect the ontic state. This can be expressed generally as

$$\gamma_M(\lambda'|Q, \lambda) = \delta(\lambda' - \lambda). \tag{5.162}$$

Hence (5.161) equates to

$$P(Q_i, Q_j) = \int d\lambda \, \xi_j(Q_j|\lambda) \, \xi_i(Q_i|\lambda) \, \mu(\lambda). \tag{5.163}$$

By substituting (5.163) into (5.158), we get

$$\langle Q_i \, Q_j \rangle = \int d\lambda \sum_{Q_i, Q_j = \pm 1} Q_i \, Q_j \, \xi_j(Q_j|\lambda) \, \xi_i(Q_i|\lambda) \, \mu(\lambda) \tag{5.164}$$

$$= \int d\lambda \langle Q_i \rangle_\lambda \, \langle Q_j \rangle_\lambda, \tag{5.165}$$

where $\langle \dots \rangle_\lambda$ denotes the expectation value for a given ontic state λ. We can then express (5.137),

$$K_3 \equiv C_{21} + C_{32} - C_{31}, \tag{5.166}$$

in the following way

$$K_3 = \int d\lambda \, \mu(\lambda) \Big(\langle Q_2 \rangle_\lambda \, \langle Q_1 \rangle_\lambda + \langle Q_3 \rangle_\lambda \, \langle Q_2 \rangle_\lambda \, - \langle Q_3 \rangle_\lambda \, \langle Q_1 \rangle_\lambda \Big). \tag{5.167}$$

Given that the expectation value of Q_i is bounded in magnitude by unity, the value K_3 is once again seen to satisfy the inequality

$$\boxed{-3 \leq K_3 \leq 1.} \tag{5.168}$$

b) Comments:

i) This expresses the notion that LG inequalities are valid for ontological models (A1) with locality in time, (A2) and (A3).

ii) Macroscopic realism can be considered as a specific form of the ontological framework through the formula

$$\mu(\lambda) = \sum_k p_k \nu_k(\lambda), \tag{5.169}$$

where $\nu_k(\lambda)$ is a distribution of states that all share macroscopic property k with respect to the relevant measurement M.

iii) The framework of ontological models has also found use in other temporal settings. In [229], these models were used in arguing that a time symmetric interpretation of quantum theory is not possible without retrocausality.

5.3.3 Through Games

a) Preliminaries: There are two considerable problems with the LG inequalities. The first is that the LG inequalities were designed for macroscopic systems, as opposed to a single evolving system. The second problem is that correlation functions that lead to violations of (5.143) can be classically simulated using a temporal version of the Toner-Bacon protocol [230, 231]. To counter these points, we will describe the use of games to develop a new formulation [230] of Bell's theorem for temporal correlations. We consider the case of a single quantum system measured at n points in time. The focus will be on a novel definition of nonclassicality for these temporal correlations, and provide the needed advantages over the LG inequalities.

b) modulo-(m,d) games: The particular n-player game which we will utilize are known as modulo-m, d games [232]. Each $n \geq 1$ players is given an integer $X_j \in [0, \ldots, d-1]$ for some fixed integer $d \geq 2$. The players are promised that d divides their sum

$$\sum_{j=1}^{n} X_j \equiv 0 \mod d. \tag{5.170}$$

The players are allowed to give answers in the form of integers $Y_j \in [0, \ldots, m-1]$ for some fixed integer $m \geq 2$. The condition for winning the game is if the

answers satisfy

$$\sum_{j=1}^{n} Y_j \equiv \frac{\sum_j X_j}{d} \mod d. \tag{5.171}$$

One can think of these games as a distributed computing task. It can be shown that these games cannot be solved with certainty using classical randomized algorithms. However, it possible to solve these games with certainty using a quantum GHZ state.

For a temporal scenario, a sequential version of the modulo-m, d game is desired. This can be achieved as follows. A sequential mudolo-m, d game is a communication task in which n separate players are given (log d)-bit inputs X_k with the condition that

$$\sum_{k=1}^{n} X_k \mod d = 0. \tag{5.172}$$

The requirement of the players is to provide values $Y_k \in [0, \ldots m-1]$ that satisfies

$$d \sum_{k=1}^{n} Y_k \equiv \sum_{k=1}^{n} X_k \bmod (md) \tag{5.173}$$

in a sequential protocol. In this sequential case, the kth stage allows the kth player to produce their local output Y_k and communicate a c_k-bit message M_k to the $(k+1)$st player.

c) Temporal correlations: Temporal correlations that have the same form as spatial correlations of an n-qu$m$it GHZ state

$$|GHZ\rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} |i\rangle^{\otimes n}, \tag{5.174}$$

are referred to as *temporal GHZ correlations*. It can be shown that the sequential mudolo-(m, d) game can be solved exactly using a sequence of POVM measurements on a single qu$m$it state which produces temporal GHZ correlations [230].

d) Nonclassicality of temporal correlations: We describe a new definition to capture the nonclassical properties of these temporal quantum correlations, and relate this later to the game. The impetus for this definition comes from the

notion that an m-level physical system has a classical information capacity of $\log_2 m$. The other motivation is that one wants to decide if a set of correlations is nonclassical purely based on the correlation function. We write the temporal correlation function as

$$E(Y_1 \ldots, Y_N | X_1 \ldots, X_N), \tag{5.175}$$

where a sequence of N consecutive measurements on a single quantum system with measurement settings provided by inputs X_k and measurement results given by numbers Y_k. We define a temporal correlation function (5.175) of the m-level physical system as *nonclassical* if all classical algorithms that simulate the function require more than $\log_2 m$ bits of classical communication at some step of the simulation. In other words, the correlation function is nonclassical if every classical simulation of it requires more communication that the classical communication capacity of the physical system in at least one stage of the simulation.

e) A temporal "Bell inequality": A key result is that every classical protocol that solves the sequential modulo-(m, d) games with certainty uses at least

$$\boxed{c_k = \log \frac{d}{m}} \tag{5.176}$$

bits of communication in all stages of the protocol except at most $md - 1$ (not necessarily consecutive) stages when d is an integer power of 2 and m is even.

This result can be thought of as a Bell inequality in that it limits what one can do with classical resources and also us to exhibit the nonclassicality of temporal quantum correlations. This latter piece can be described using the following result: The temporal GHZ correlations arising from the sequential measurements on a single qu*m*it, where m is even, are nonclassical for $n \geq 2m^3$. To prove this one simply uses result (5.176) and also shows the result that there exists a sequential modulo-(m, d) game for some d and n for which classical simulation uses in at least one stage of the protocol more than $\log m$ bits of communication.

5.3.4 Other works

In this subsection, we provide a brief overview on some other interesting works regarding non-locality across time.

a) Temporal Hardy's paradox: A temporal version of Hardy's non-locality paradox was proposed [233] and experimentally verified [228]. In this scenario, let Alice and Bob measure one after the other to signify the temporal property. Let $P(r, s|k, l)$ denote the probability that Alice obtains result r and Bob obtains result s given they chose detector settings a_k and b_l respectively. The temporal Hardy paradox is that under the LG assumptions (A1-3), the probabilities

$$P(+1, +1|1, 1) = 0, \tag{5.177}$$

$$P(-1, +1|1, 2) = 0, \tag{5.178}$$

$$P(+1, -1|2, 1) = 0, \tag{5.179}$$

$$P(+1, +1|2, 2) > 0, \tag{5.180}$$

are mutually inconsistent. Quantum theory on the other hand provides a way where these probabilities can be simultaneously be fulfilled.

b) Indefinite causal structures: There are many frameworks that employ the use of the Choi-Jamiolkowski isomorphism. One example of this is in the framework of quantum indefinite causal structures [234, 235]. It provides a framework that does not assume a pre-defined global causal structure but only that quantum theory holds locally. Central to the framework is the "process matrix" which can be thought of as a generalization of a density matrix. Of interest to the subject of this thesis is that this framework has been used to analyze temporal quantum correlations [236, 189]; in this work they experimentally observed multi-time quantum correlations that cannot be replicated by any spatial quantum state of equal dimension.

c) Pseudo-density matrix: Another generalization of a density matrix is known as the pseudo-density matrix [237]. This framework has been used to analyze various temporal quantum correlations including a weaker version than non-locality across time known as temporal steering [238]. In addition to that it has

been used in identifying the relationship between temporal correlations and aspects of quantum communications [239].

d) Quantum causal models: Classical causal models have found a wide range of use in areas of machine learning [240]. There have been various advances [241, 242] on quantum generalizations of classical causal models. This may lead to a deeper understanding of how quantum causality differs from classical causality. Moreover, from the perspective of this thesis, it may provide a platform for the development of temporal quantum machine learning algorithms.

e) Entangled histories framework: The entangled histories framework [243, 244] (and its consistent histories framework) are based on a analogous version of a unitary evolution operator known as the bridging operator. The concept of entanglement in time is introduced within this framework with a focus on studying the property of monogomy of entanglement (5.88).

6

Relativistic Quantum Information

> *"You know why we have come together: we must decide what to do about these new events. The universe is broken wide, and Lord Asriel has opened the way from this world to another."*
>
> – Philip Pullman, *His Dark Materials*

THE UNIVERSE contains both quantum physics as well as relativistic effects. However, a single theoretical description of these diverse phenomena remains elusive. More recently, there have been investigations on whether the conceptualization of quantum information could play a crucial role for this unification. On a coarse level, such research activities can be categorized in two directions. The first is known as relativistic quantum information (RQI), and it examines the effects of relativity on the concepts of quantum information science [245, 246, 247]. Besides fundamental reasons, this has important applications most notably to satellite based quantum communications [248]. The second direction explores how the concepts of quantum information science could be used to study relativistic structures [249, 250, 251]. A large motivation for this path stems from the holographic principle. Both directions use quantum field theory [252, 253, 254] which represents a partial unification; this is in contrast to the standard use of non-relativistic quantum mechanics to articulate quantum information. In this thesis we focus on the first direction of RQI, and explore how an entanglement in space and an entanglement in time manifest themselves in such a setting.

6.1 Review of Relativity

To understand the relativistic effects on quantum information science, we first provide a brief review on the subject of relativity. For a thorough introduction, we refer the reader to [255].

6.1.1 Special relativity

Perhaps the most shocking temporal effect in special relativity is time dilation. This can be mathematically described as

$$\Delta t = \gamma \Delta \tau, \tag{6.1}$$

where Δt and $\Delta \tau$ represents the time intervals under consideration, and

$$\gamma \equiv \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{6.2}$$

Ultimately, time dilation and other special relativistic effects are a consequence of the invariance of the spacetime interval between two events

$$(\Delta S)^2 = c^2(\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2, \tag{6.3}$$

where c represents the speed of light and where we have used coordinates (t, x, y, z). This implies that observers who are in motion to each other should always agree on the value of the spacetime interval (6.3) despite them disagreeing on the individual spatial intervals and time interval. The invariance of this particular combination of spatial and temporal intervals is what leads to the statement that space and time form one 'object' called spacetime.

A spacetime interval is called timelike if $(\Delta S)^2 > 0$. This means there is some frame of reference (coordinate system) where the events occur at the same spatial location, and there is no frame of reference where the events occur at the same instant of time. Moreover, an event that occurs first in one frame of reference, occurs first in all frames of reference. A spacetime interval is called spacelike if

$(\Delta S)^2 < 0$. This implies that there is some frame of reference where the events occur at the same instant of time, and there is no frame of reference where the events occur at the same spatial location. Lastly, a spacetime interval is called lightlike or null if $(\Delta S)^2 = 0$. This has the consequence that there is no frame of reference where the events occur at either the same instant of time or at the same spatial location. Furthermore, the event that occurs first in one frame of reference, occurs first in all frames of reference.

A spacetime diagram is one where the vertical axis corresponds to time (and where we set $c = 1$) and horizontal axis corresponds to one of the spatial coordinates, say z. The origin is an event, denoted by say E. Light rays move on lines $z = t$ and $z = -t$, which defines a light cone. Special relativity says that nothing can travel faster than the speed of light. This can be depicted in an alternative way in that spacelike intervals are the regions outside the light cone. These are the sets of events that are causally unrelated to event E.

The infinitesimal version of the spacetime interval (6.3) takes the form

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2. \tag{6.4}$$

6.1.2 General relativity

The theory of general relativity says that spacetime line element (6.4) is one of many possible spacetime line elements. Each describes a different spacetime. The particular case of (6.4) is known as flat spacetime, and the theory of general relativity includes curved spacetimes. To adequately describe such curvature, the subject utilizes the mathematics of differential geometry. The central tenet of differential geometry is that an intrinsic description of space could be accomplished by distance measurements made within that space.

For our brief review, we employ the standard use of the Einstein summation convention where one omits the summation symbol whenever a pair of contravariant and covariant indices appears in one term. We usually let the indices range over the four spacetime dimensions unless otherwise stated. Hence, (6.4) can be

written as

$$ds^2 = \eta_{ab}\, dx^a dx^b, \tag{6.5}$$

for coordinates dx^a ($dx^0 = cdt$) and the quantity η_{ab} is known as the Minkowski metric,

$$\eta_{ab} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{6.6}$$

More generally, we can represent an arbitrary line element as

$$ds^2 = g_{ab}\, dx^a dx^b, \tag{6.7}$$

where g_{ab} is known as the metric tensor or simply as metric. A Lorentzian metric is a metric with signature $(+---)$. This means that any given point in spacetime we can find coordinates such that

$$g_{ab} = \eta_{ab} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{6.8}$$

In the case of flat spacetime, the metric $g_{ab} = \eta_{ab}$ everywhere.

General relativity postulates that spacetime is a four-dimensional manifold equipped with a Lorentzian metric, g_{ab}. A manifold can be thought of as a collection of points which locally looks like $\mathbb{R}^4$. The quantity $ds^2 = g_{ab}\, dx^a\, dx^b$ is invariant. The metric tensor plays the crucial role of determining the geometry of the manifold and the important geometric quantities are built from this tensor and its derivatives. The connection (or Christoffel symbol) is given by

$$\Gamma^{\,a}_{bc} \equiv \frac{1}{2}\, g^{ad}\, (g_{db,c} + g_{dc,b} - g_{bc,d}). \tag{6.9}$$

The matrix inverse of metric g_{ab} is denoted g^{ab}. Furthermore, the commas denote partial derivatives: $X_{a,b} \equiv \partial_b X_a \equiv \partial X_a / \partial x^b$. We say a vector is timelike if

$g_{ab}X^aX^b > 0$, spacelike if $g_{ab}X^aX^b < 0$ and lightlike or null if $g_{ab}X^aX^b = 0$

This connection can be defined in terms of the covariant derivative of a tensor

$$\nabla_b V^a = \partial_b V^a + \Gamma^a_{\ cb} V^c. \tag{6.10}$$

This is a generalization of taking a derivative in curved spaces. Notice the deviation from flat space is represented by the connection. The Riemann curvature tensor is a quantity which measures the extent to which the covariant derivative fails to commute, and in that sense, the information about the curvature is located in the components of this tensor. The explicit formula for this tensor is given by

$$R^a_{\ bcd} \equiv \partial_c \Gamma^a_{\ bd} - \partial_d \Gamma^a_{\ bc} + \Gamma^a_{\ ec} \Gamma^e_{\ bd} - \Gamma^a_{\ ed} \Gamma^e_{\ bc}. \tag{6.11}$$

The Ricci tensor, Ricci scalar and Einstein tensor are respectively built out of the Riemann tensor as

$$R_{ab} \equiv R^c_{\ acb}, \tag{6.12}$$

$$R \equiv g^{ab} R_{ab}, \tag{6.13}$$

$$G_{ab} \equiv R_{ab} - \frac{1}{2} R\, g_{ab}. \tag{6.14}$$

From this, the theory of general relativity postulates the Einstein field equations

$$G_{ab} = \frac{8\pi G}{c^2} T_{ab}, \tag{6.15}$$

where G is Newton's constant of universal gravitation. The equations relate the curvature of spacetime (quantified by G_{ab}) to the distribution of matter and energy (as quantified by the stress-energy tensor T_{ab}).

The Einstein field equations allows one to obtain a spacetime from a given matter-energy distribution. Vacuum spacetimes are solutions where $T_{ab} = 0$ in (6.15). This can be shown to be equivalent to the statement that $R_{ab} = 0$ and is known as a Ricci-flat solution. The flat spacetime (6.4) one such solution. Another Ricci-flat solution is known as the Schwarzschild metric which in coordinates (t, r, θ, ϕ) is

written as

$$ds^2 = \left(1 - \frac{2\,m}{r}\right) dt^2 - \frac{1}{1 - \frac{2\,m}{r}} dr^2 - r^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2\right). \tag{6.16}$$

The parameter m measures the amount of mass inside the radius r, and in the region $r \leq 2m$ the metric describes a black hole region. The solution blows up at $r = 0$ and $r = 2m$; the former is known as a physical singularity whereas the latter is known as a coordinate singularity as it is simply an artefact of the use of this particular coordinate system.

In this thesis, we have seen the manifestation of the Schrödinger equation (3.63). Along with the Einstein field equations (6.15), these two pieces form the fundamental equations of modern theoretical physics. The aim to unify these descriptions is known as the problem of quantum gravity, and has so far remained unsolved despite considerable efforts. We make a few remarks on the similarities of these equations. Both require an energy input; the first through the energy-momentum tensor and the the second from the Hamiltonian. The first equation describes the dynamics of spacetime while the second equation describes the dynamics of a quantity whose direct relationship to the physical world is unknown. Nevertheless, both output solutions that describe point particles that behave in the most bizarre manner; general relativity says that point particles cannot exist but form singularities; quantum theory provides point particles with the most bizarre properties such as entanglement. Of particular relevance is that the authors Einstein and Rosen of the EPR paper [52], wrote another paper that same year titled "The particle problem in general relativity" [256]. In it they attempted (but failed) to build a model of a point particle without a singularity. The work was later termed the Einstein-Rosen bridge, and provided the pathway for the most shocking temporal structures in relativity, namely wormholes [4].

6.2 Quantum Fields

6.2.1 Quantum field theory

We showed in Chapter 3 how the quantum circuit model is based on the postulates of quantum theory. As a framework, quantum theory (ie those set of postulates) does not specify the state space, the state vector, or the Hamiltonian of a specific physical system under consideration. It merely provides the mathematical framework for the construction of various physical theories. The specification of such quantities allows the physics to arise, resulting in different physical theories. Non-relativistic quantum mechanics is only one such theory; the quantum circuit model corresponds to a non-relativistic two-level quantum system. Quantum field theory is another subset of quantum theory which describes (special) relativistic quantum particles. In this latter sense, quantum field theory can be viewed as a unification of quantum theory and special relativity.

In non-relativistic quantum mechanics, we have a position and a momentum operator; however the existence of these operators are not part of the postulates. In quantum field theory, these operators are not well-defined, and position is described as a label. We have an operator at each point in space and the collection of these position-dependent operators is known as a quantum field. Each quantum field has what is known as a conjugate momentum density which is also a function of the spatial label.

The framework of quantum theory can be expressed in the Schrödinger picture (where operators are time independent and states are time dependent), the Heisenberg picture (where operators are time dependent and states are time independent), or the Dirac picture (which is an intermediate of the two). In the Heisenberg picture, the quantum field then is also a function of time. So far in this thesis we have been using the Schrödinger picture which involves the equation (3.63). Portraying quantum field theory using the Schrödinger picture results the Hamiltonian expressed in terms of infinitely many degrees of freedom, and (3.63) taking the form of a functional differential equation. The quantum field state (or wave functional as its known in this case) is a function of time but also a functional of the classical field configuration. And the square of the wave func-

tional gives the probability density for measuring a certain field configuration. On a related note, all the foundational mysteries regarding quantum theory, such as the measurement problem, still remain.

Despite that the Heisenberg picture is rarely used to introduce the subject of non-relativistic quantum mechanics, it is precisely the Heisenberg picture that is often used to introduce quantum field theory. We aim to provide the most basic tools of this subject in order to progress towards to the entanglements in RQI.

6.2.2 Quantization

The procedure of quantization allows one to obtain a physical theory of a quantum system from an analogous classical system. As an example, it allows one to obtain a quantum Hamiltonian operator from a classical Hamiltonian function.

a) Harmonic oscillator: In non-relativistic quantum mechanics, one often quantizes a harmonic oscillator. A classical harmonic oscillator with external force $J(t)$ satisfies the equation of motion

$$\ddot{q} = -\omega^2 q + J(t), \tag{6.17}$$

where Hamiltonian is written as

$$H(p, q) = \frac{p^2}{2} + \frac{\omega^2 q^2}{2} - J(t)q, \tag{6.18}$$

where q is the spatial coordinate and p is the momentum. Quantization involves turning q and p into respective operators $\hat{q}(t)$ and $\hat{p}(t)$ that satisfy the following commutation relation

$$[\hat{q}, \hat{p}] = i, \tag{6.19}$$

where we have set $\hbar = 1$. From these quantities, one can define the annihilation and creation operators which can respectively be expressed as

$$\hat{a}(t) \equiv \sqrt{\frac{\omega}{2}} \left[\hat{q}(t) + \frac{i}{\omega}\hat{p}(t)\right], \quad \hat{a}^\dagger(t) \equiv \sqrt{\frac{\omega}{2}} \left[\hat{q}(t) - \frac{i}{\omega}\hat{p}(t)\right]. \tag{6.20}$$

These satisfy

$$[\hat{a}(t), \hat{a}^\dagger(t)] = 1, \tag{6.21}$$

at every moment of time. Through various computations and a final substitution into (6.18), one can obtain the quantum Hamiltonian operator

$$\hat{H} = \frac{\omega}{2}(\hat{a}^\dagger \hat{a} + \hat{a}\hat{a}^\dagger) - \frac{\hat{a}^\dagger + \hat{a}}{\sqrt{2\omega}} J(t). \tag{6.22}$$

One can proceed to construct a basis for the corresponding Hilbert space. This assumes the existence of normalized state $|0\rangle$ (note this is not the element from the computational basis states (3.4)) where

$$\hat{a}\,|0\rangle = 0. \tag{6.23}$$

This state is known as the vacuum state. One can create excited states

$$|n\rangle = \frac{1}{\sqrt{n!}}(\hat{a}^\dagger)^n\,|0\rangle\,, \tag{6.24}$$

for $n \geq 1$. All possible quantum states of the oscillator can be written as

$$|\psi\rangle = \sum_{n=0}^{\infty} \psi_n\ |n\rangle\,, \quad \sum_{n=0}^{\infty} |\psi_n|^2 = 1. \tag{6.25}$$

b) Field quantization: Classical relativistic fields can be described by equations such as the Klein-Gordon equation, the Dirac equation, and the Maxwell equations. Their role is analogous to harmonic oscillator in that they provide a classical Hamiltonian for quantization. The quantization of these classical field equations provides the quantum field theory (which in turn results in a description of relativistic quantum particles). More precisely, the Schrödinger equations corresponding to each of the classical field equations articulates the different quantum field theories. In this section, we will describe this quantization using the Heisenberg picture for the specific case of the Klein-Gordon equation. The scalar field satisfying this equation can be thought of as a set of infinitely many harmonic oscillators. Hence our quantization method will relate to the method

used to quantize a harmonic oscillator. To start this procedure, we have classical scalar field ϕ which satisfies the Klein-Gordon equation

$$\Box\phi = 0, \tag{6.26}$$

where the d'Alambertian operator $\Box$ is defined as

$$\Box\phi \equiv \frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}g^{\mu\nu}\partial_\nu\phi). \tag{6.27}$$

We have used notation $g = \det(g_{ab})$. In flat two-dimensional spacetime (6.5), the metric is given by $g_{\mu\nu} = \eta_{\mu\nu} = \{+-\}$. Hence we have

$$\Box\phi = (\partial_t^2 - \partial_z^2)\phi, \tag{6.28}$$

for coordinates (t, z). The solution to the Klein-Gordon equation are plane waves in Minkowski spacetime M,

$$u_{\omega,M}(t, z) = \frac{1}{\sqrt{4\pi\omega}}e^{-i\omega(t-\epsilon z)}, \tag{6.29}$$

where ϵ equates to $+1$ for positive momentum modes and to -1 for negative momentum modes. These plane wave solutions are known as global field modes. The modes where $\omega > 0$ are orthonormal with respect to a Lorentz invariant inner product

$$(\phi, \psi) = -i\int_\Sigma(\psi^*\partial_\mu\phi - (\partial_\mu\psi^*)\phi)d\Sigma^\mu. \tag{6.30}$$

Our next step is to quantize this scalar field. To do so we require a time-like Killing vector field. We say that K is a Killing vector field if

$$\mathcal{L}_K g_{\mu\nu} = 0 \tag{6.31}$$

where the Lie derivative of the metric tensor is defined as

$$\mathcal{L}_K g_{\mu\nu} \equiv K^\lambda\partial_\lambda g_{\mu\nu} + g_{\mu\lambda}\partial_\nu K^\lambda + g_{\nu\lambda}\partial_\mu K^\lambda. \tag{6.32}$$

If a spacetime has a Killing vector field, then one can find a basis for the plane

wave solutions of the Klein-Gordon equation such that

$$\mathcal{L}_K u_{k,M} = K^\mu \partial_\mu u_{k,M} = -i\omega u_{k,M}. \tag{6.33}$$

It can be shown that if K is a time-like Minkowski vector field, then the Lie derivative corresponds to ∂_t. Then (6.33) takes the form

$$\partial_t u_{k,M} = -i\omega u_{k,M} \tag{6.34}$$

$$\partial_t u^*_{k,M} = -i\omega u^*_{k,M} \tag{6.35}$$

where we identify $\omega > 0$ with a frequency. One can classify the plane wave solutions to the Klein-Gordon equation as

$$u_k \quad \longrightarrow \quad \text{positive frequency solutions} \tag{6.36}$$

$$u^*_k \quad \longrightarrow \quad \text{negative frequency solutions} \tag{6.37}$$

The quantized field can be obtained using these positive and negative frequency solutions. More precisely, the quantized field satisfies equation

$$\Box\hat{\phi} = 0, \tag{6.38}$$

and is given by the operator value function

$$\hat{\phi} = \int (u_{k,M} a_{k,M} + u^*_{k,M} a^\dagger_{k,M}) dk. \tag{6.39}$$

The operators $a^\dagger_{k,M}$ and $a_{k,M}$ are creation and annihilation operators which satisfy

$$[a_{k,M}, a^\dagger_{k',M}] = \delta_{k,k'}. \tag{6.40}$$

Observe that positive frequency solutions are associated with annihilation operators, whereas negative frequency solution correspond to creation operators. Furthermore, these creation and annihilation operators are analogous to case of the harmonic oscillator (6.20). We also have a vacuum state for the quantum field

which is defined by

$$a_{k,M}|0\rangle^{\mathcal{M}} = 0. \tag{6.41}$$

This is analogous (6.23). In fact, the vacuum state can be expressed as

$$|0\rangle^{\mathcal{M}} = \prod_k |0_k\rangle^{\mathcal{M}}, \tag{6.42}$$

where $|0_k\rangle^{\mathcal{M}}$ is the ground state of mode k. The vacuum can be physically thought of as 'empty space.' The action of the creation operators on the vacuum allows one to define particle states

$$|n_1, ..., n_k\rangle^{\mathcal{M}} = (n_1!, ..., n_k!)^{-1/2}\, a^{\dagger n_1}_{1,M} ... a^{\dagger n_k}_{k,M}\, |0\rangle^{\mathcal{M}}. \tag{6.43}$$

This procedure implies that only when there exists a time-like Killing vector field is the notion of a particle well-defined.

c) Bogoliubov transformation: For the scenario that a spacetime admits a time-like Killing vector field, the vector field is generally not unique. In our procedure we used ∂_t, and another such vector field could be denoted by $\partial_{\bar{t}}$. For each case, one can obtain a basis for the solutions which we respectively denote $\{u_k, u_k^*\}$ and $\{\bar{u}_k, \bar{u}_k^*\}$. Positive and negative frequency solutions can be identified for the other basis as well. Hence, the field can be equivalently quantized in both bases,

$$\hat{\phi} = \int (u_k a_k + u_k^* a_k^\dagger) dk = \int (\bar{u}_{k'} \bar{a}_{k'} + \bar{u}_{k'}^* \bar{a}_{k'}^\dagger) dk'. \tag{6.44}$$

By utilizing the inner product, it is possible to obtain a transformation, known as the Bogoliubov transformation, between the representations for the creation and annihilation operators

$$a_k = \sum_{k'} (\alpha^*_{kk'} \bar{a}_{k'} - \beta^*_{kk'} \bar{a}^\dagger_{k'}), \tag{6.45}$$

where $\alpha_{kk'} = (u_k, \bar{u}_{k'})$ and $\beta_{kk'} = -(u_k, \bar{u}^*_{k'})$ which are known as the Bogoliubov coefficients. Both vacuum states are defined as

$$a_k|0\rangle = \bar{a}_k|\bar{0}\rangle = 0 \tag{6.46}$$

and hence it is possible to derive a transformation between the two vacuum states.

d) For RQI: The field from the above procedure is just one type of quantum field theory. Collectively, quantum field theories provide a description of relativistic quantum systems. The field of RQI uses quantum field theory, as opposed to non-relativistic quantum mechanics, to express quantum information and its information tasks. In this sense, it allows one to investigate the effects of relativity on the concepts of quantum information science.

6.2.3 Locality in RQI

In Chapter 5, we defined locality as systems that obey the Bell-CHSH inequality (5.85). It is important to use the more precise terminology of Bell locality in relation to (5.85) given that there are other mathematical characterizations of the notion of locality [57]. In quantum field theory, locality is quantitatively expressed in a different manner to (5.85), and is also more commonly referred to as causality [257]. It captures the notion that a measurement at one spatial location say x cannot affect a measurement at another spatial location y, when x and y are not causally connected.

More rigorously, *causality* is the requirement that all operators commute for spacelike separation

$$\boxed{[O(x), O(y)] = 0 \quad \text{for} \quad (x-y)^2 < 0.} \tag{6.47}$$

One often says that the theory is causal if the commutators vanish outside the light cone. Nevertheless, (6.47) does not make a distinction between the forward light cone and backward light cone. In [258] it was shown that there is an implied arrow of causality (meaning what is the past and what is the future) which is connected to the sign of the imaginary number in the quantization procedure. Reversing the sign of the factors of i leads to a causal theory with the consequence of an arrow of causality running from large times to small times.

6.2.4 Entanglement in RQI

In Chapter 4, we defined entanglement as the nonseparability of a state (4.2). This nonseparability was also expressed through the density operator as (4.35). Despite RQI harnessing the framework of quantum field theory, it utilizes the same algebraic nonseparability definition of entanglement as non-relativistic quantum mechanics. An open question in RQI [246] is whether there exists a more general notion of quantum interdependence for relativistic quantum systems which maps to the standard notion of entanglement in the non-relativistic regime.

To articulate the current definition [133], consider a spacetime manifold M, with two subsets of it denoted R_1 and R_2. The states of the field restricted to each subset are described by respective Hilbert spaces, H_{R_1} and H_{R_2}. If the field operators commute between the two regions then one can say that H_{R_1} and H_{R_2} represent independent systems. Then the state of the quantum field ρ restricted to $R_1 \cup R_2$ is called *entangled* if it is not separable, meaning if it cannot be represented as

$$\boxed{\rho = \sum_i p_i \rho_1^i \otimes \rho_2^i,} \tag{6.48}$$

where ρ_1^i are density operators on H_{R_1} and ρ_2^i are density operators on H_{R_2}, with $p_i \geq 0$. Notice that this definition is analogous to (4.35).

Our aim in the next section is to show that the vacuum state (6.42) of a quantum field is an entangled state. We will do this through the state as opposed to a density operator framework. It is important to emphasize that whether a state is entangled or not depends on the tensor-product decomposition that is chosen for the total Hilbert space [247]. Hence the concept of entanglement in the field is to be understood within this context. To elaborate on matter, let H_{field} denote the Hilbert space associated to the free scalar field in Minkowski spacetime (6.39). If we decompose that Hilbert space into plane wave modes, then we have decomposition

$$H_{\text{field}} = \bigotimes_k L^2(R)_k \tag{6.49}$$

where $L^2(R)_k$ is the countably infinite harmonic oscillator state space with mode

k. Using this, the Minkowski vacuum is not entangled but can be decomposed into product state

$$|0\rangle^{\mathcal{M}} = \bigotimes_k |0_k\rangle^{\mathcal{M}}, \tag{6.50}$$

which is equivalent to (6.42). However, as we shall describe one can also decompose the field into a left and right half (known as Rindler wedges)

$$H_{\text{field}} = H_{\text{left}} \bigotimes H_{\text{right}}. \tag{6.51}$$

Then the Minkowski vacuum state is a tensor product of two-mode squeezed (TMS) states in pairs of (Rindler) modes indexed by ω

$$|0\rangle^{\mathcal{M}} = \bigotimes_\omega |TMS\rangle_{(\omega,I);(\omega,II)}\,. \tag{6.52}$$

TMS states are a central topic in the area of quantum optics [259]; physically in squeezed states the noise of the electric field at certain phases falls below that of the vacuum state; however our focus is solely on the mathematical description.

The above decomposition represents a bipartite entanglement across the left-right cut, and the explicit form of equation (6.52) will be articulated in the next section. We also would like to point out that the analysis in the next sections are all in 1 + 1 dimensions.

6.3 Spacelike Entanglement

6.3.1 Definition

Recall the RQI definition of entanglement as expressed through (6.48) using subsets R_1 and R_2 of spacetime M. A further distinction can be made [133]. If all the points in R_1 are spacelike separated with respect to all points in R_2, then we say that the quantum field in R_1 is *spacelike entangled* with respect to the quantum field in R_2. In other words, this nonseparability of the state of the quantum field can be regarded as an entanglement in space as described in Chapter 4. However in this relativistic setting, the spatial aspect is articulated far more precisely by

using the light cone structure i.e. spacelike intervals.

6.3.2 Left-Right entanglement

The two-dimensional Minkowski spacetime (t, z) can be broken up into regions using the light cone structure. The spacelike regions outside the light cone $|t| < -z$ and $|t| < z$ are respectively known as the left Rindler wedge and right Rindler wedge. We want to show that the Minkowski vacuum can be written as a spacelike entangled state between the left and right Rindler modes [260].

a) Independent systems: For the possibility of entanglement, we want that the fields within the left and right Rindler wedges are considered as independent systems. This is a requirement as we want to quantize them separately. Such a condition gets fulfilled if the commutators vanish

$$[\hat{\phi}(x_L), \hat{\phi}(x_R)] = 0. \tag{6.53}$$

For spacelike intervals, this vanishing holds for both massive and massless fields.

b) Minkowski plane waves: Our derivation rests on the following statement: Ordinary plane waves in Minkowski spacetime cover the spacetime. Our coordinates for Minkowski spacetime is (t, z) and hence the massless scalar field in two dimensions satisfies

$$\Big(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial z^2}\Big)\hat{\phi} = 0. \tag{6.54}$$

If we were to use light-cone coordinates

$$U = t - z, \qquad V = t + z, \tag{6.55}$$

then we can write the field in terms left and right moving sectors

$$\hat{\phi}(t, z) = \hat{\phi}_-(U) + \hat{\phi}_+(V). \tag{6.56}$$

Given that the left and right moving sectors do not interact, we can discuss the effect for the left-moving sector to simply the exposition. Through expansion,

one finds that

$$\hat{\phi}(V) = \int_0^\infty dk[\hat{b}_{+k}\, u_k(V) + \hat{b}^\dagger_{+k} u^*_k(V)], \tag{6.57}$$

where

$$u_k(V) = (4\pi k)^{-1/2} e^{-ikV}. \tag{6.58}$$

The Minkowski spacetime plane waves are given by (6.58) and its complex conjugate. The Minkowski vacuum which we denote $|0_M\rangle$ is defined as

$$\hat{b}_{+k}\,|0_M\rangle = 0, \tag{6.59}$$

for all k.

c) Rindler plane waves: For the right Rindler wedge ($0 < V$), we have the coordinate transformation

$$t = a^{-1}e^{a\epsilon}\sinh(a\tau), \qquad z = a^{-1}e^{a\epsilon}\cosh(a\tau). \tag{6.60}$$

Due to the conformal invariance of the massless wave equation in two dimension, the wave equation takes the same form as (6.54),

$$\left(\frac{\partial^2}{\partial\tau^2} - \frac{\partial^2}{\partial\epsilon^2}\right)_R \hat{\phi} = 0. \tag{6.61}$$

Obtaining analogous light-cone coordinates

$$\chi = \tau + \epsilon, \qquad \kappa = \tau - \epsilon, \tag{6.62}$$

we can express the left-moving sector as

$$\hat{\phi}_+(V) = \int_0^\infty d\omega[\hat{a}^R_{+\omega} g^R_\omega(\chi) + \hat{a}^{R\dagger}_{+\omega} g^{R*}_\omega(\chi)]. \tag{6.63}$$

The mode solutions or plane waves in Rindler coordinates are

$$g^R_\omega(\chi) = (4\pi\omega)^{-1/2} e^{-i\omega\chi}. \tag{6.64}$$

Without repeating all the details, similar calculations and results can be made

for analogous light cone coordinate, $\overline{\chi}$, in the left Rindler wedge ($V < 0 < U$). In particular the mode function takes the form

$$g^L_\omega(\overline{\chi}) = (4\pi\omega)^{-1/2} e^{-i\omega\overline{\chi}}. \tag{6.65}$$

The vacuum state for the left Rindler wedge and right Rindler wedge are the identical. It is known as the Rindler vacuum $|0_R\rangle$ and it is defined through

$$\hat{a}^R_{+\omega} |0_R\rangle = \hat{a}^L_{+\omega} |0_R\rangle = 0, \tag{6.66}$$

for all ω.

d) Bogoliubov transformation: The Minkowksi light cone coordinates are related to the analogous Rindler light cone coordinates through

$$V = a^{-1} e^{a\chi}, \qquad V = -a^{-1} e^{-a\overline{\chi}}. \tag{6.67}$$

Since the modes are complete in their region, we can expand the Rindler modes $g^R_\omega(\chi)$ and $g^L_\omega(\overline{\chi})$ in terms of Minkowski plane waves, $u_k(V)$ and $u^*_k(V)$, because the plane waves are defined over all spacetime. Hence, we can express one set of modes in terms of the other (i.e. using Bogoliubov transformations). The Heaviside function is used to make the expression valid in their respective quadrant,

$$\theta(V)\, g^R_\omega(\chi) = \int_0^\infty dk\, (\alpha^R_{\omega k} u_k(V) + \beta^R_{\omega k} u^*_k(V)), \tag{6.68}$$

$$\theta(-V)\, g^L_\omega(\overline{\chi}) = \int_0^\infty dk\, (\alpha^L_{\omega k} u_k(V) + \beta^L_{\omega k} u^*_k(V)). \tag{6.69}$$

These Rindler modes form a superposition of Minkowski plane waves and the coefficients, α and β, are the Bogoliubov coefficients. Solving equations (6.68) and (6.69) gives the relations

$$\beta^L_{\omega k} = -e^{-\pi\omega/a} \alpha^{R*}_{\omega k} \qquad \beta^R_{\omega k} = -e^{-\pi\omega/a} \alpha^{L*}_{\omega k} \tag{6.70}$$

e) L-R entanglement: By substituting (6.70) back into (6.68) and (6.69), we can

define a new set of modes known as the Unruh modes,

$$G_\omega(V) = \theta(V) g^R_\omega(\chi) + \theta(-V) e^{-\pi\omega/a} g^{L*}_\omega(\overline{\chi}), \tag{6.71}$$

$$\overline{G}_\omega(V) = \theta(-V) g^L_\omega(\overline{\chi}) + \theta(V) e^{-\pi\omega/a} g^{R*}_\omega(\chi). \tag{6.72}$$

These share the Minkowski vacuum

$$\hat{a}_{G\omega} |0_M\rangle = \hat{a}_{\overline{G}\omega} |0_M\rangle = 0. \tag{6.73}$$

The Unruh mode annihilation operators can be written in terms of the Rindler annihilation and creation operators as follows

$$\hat{a}_{G\omega} = (\hat{a}^R_\omega - e^{-\pi\omega/a} \hat{a}^{L\dagger}_\omega), \tag{6.74}$$

$$\hat{a}_{\overline{G}\omega} = (\hat{a}^L_\omega - e^{-\pi\omega/a} \hat{a}^{R\dagger}_\omega). \tag{6.75}$$

These expressions can be combined to produce the equation

$$(\hat{a}^{R\dagger}_\omega \hat{a}^R_\omega - \hat{a}^{L\dagger}_\omega \hat{a}^L_\omega) |0_M\rangle = 0. \tag{6.76}$$

We proceed to use the approximation that ω is discrete. From the previous equation we obtain,

$$\boxed{|0_M\rangle = \prod_i C_i \sum_{n_i=0}^{\infty} \frac{e^{-\pi n_i \omega_i/a}}{n_i!} (\hat{a}^{R\dagger}_{\omega_i} \hat{a}^{L\dagger}_{\omega_i})^{n_i} |0_R\rangle} \tag{6.77}$$

where $C_i = \sqrt{1 - e^{-2\pi\omega_i/a}}$. This is a *spacelike entanglement* of the Minkowski vacuum in terms of the left and right Rindler modes. One can re-express equation (6.77) as

$$|0_M\rangle = \prod_i C_i \sum_{n_i=0}^{\infty} e^{-\pi n_i \omega_i/a} |n^R_i\rangle \otimes |n^L_i\rangle \tag{6.78}$$

where $|n^R_i\rangle$ is the state of a Rindler mode restricted to the right wedge, containing n excitations of frequency ω_i. In an analogous manner, $|n^L_i\rangle$ is the Rindler mode restricted to the left wedge. This state is entangled as it is nonseparable between

the left and right wedges.

6.3.3 Implications

We discuss some observations of this spacelike entanglement. One can form a density operator using (6.77) and trace over either one of the regions. This results in a thermal state with temperature,

$$T = \frac{a\hbar}{2\pi kc}, \tag{6.79}$$

where $\hbar$ is the reduced Planck's constant and k is the Boltzmann's constant. This quantity is commonly referred to as the Unruh temperature. By utilizing the Rindler coordinates, one can intepret the temperature in the following way: While inertial observers describe the field to be the vacuum, observers in uniform acceleration observe a state thermalized with particles at the Unruh temperature. Hence the particle content of a field is observer dependent! This Unruh temperature is in fact analogous to the famous Hawking temperature of a black hole

$$T_H = \frac{\hbar c^3}{8\pi GMk}, \tag{6.80}$$

where M is the mass of the black hole (6.16). From an RQI point of view, one can say that the section of the quantum vacuum trapped behind the event horizon $r = 2m$ is spacelike entangled with that outside. From the temperature (6.80), one can derive the Bekenstein-Hawking entropy of a black hole

$$S_{BH} = \frac{c^3 Ak}{4G\hbar}, \tag{6.81}$$

where A is the surface area of the black hole (ie area of the event horizon). However it is not known what the microscopic nature the black hole entropy is, and many consider this formula as the crucial clue to quantum gravity.

In terms of quantum information protocols, in [261] it was shown that this Unruh effect reduces the fidelity of quantum teleportation. This alluded to the notion that entanglement is degraded in non-inertial frames. This was clearly shown

in [262] where degradation of spacelike entanglement was quantified when one of the observers moved in uniform acceleration. Such a case was also mapped into the scenario of an observer falling into a black hole resulting in a similar degradation. These results imply that entanglement is an observer dependent phenomenon! In Chapter 4, we mentioned the procedure of entanglement swapping. This notion carries over to RQI and is referred to as entanglement extraction or entanglement harvesting [247]. It is the process of extracting field entanglement by local quantum systems interacting with the field in a spacelike separated way. This harvesting procedure can be generalized into entanglement farming.

6.4 Timelike Entanglement

6.4.1 Definition

Recall the RQI definition of entanglement as expressed through (6.48) using subsets R_1 and R_2 of spacetime M. Analogous to the spacelike case, one can provide a timelike version [133]. If all the points in R_1 are timelike separated with respect to all points in R_2, then we say that the quantum field in R_1 is *timelike entangled* with respect to the quantum field in R_2. In other words, this nonseparability of the state of the quantum field can be regarded as an entanglement in time as described in Chapter 4. However in this relativistic setting, the temporal aspect is articulated far more precisely by using the light cone structure i.e. timelike intervals.

6.4.2 Future-Past entanglement

We want to focus on the regions inside the light cone in the two-dimensional Minkowski spacetime (t, z). The regions $t > |z|$ and $t < -|z|$ are known respectively as the Future and Past. We proceed to describe the work in [263] which showed the Minkowski vacuum can be written as a timelike entangled state between the future and past modes. It follows an analogous procedure to the spacelike case.

a) Independent systems: For massless fields, the commutator vanishes for timelike intervals

$$[\hat{\phi}(x_F), \hat{\phi}(x_P)] = 0. \tag{6.82}$$

Hence we can quantize the massless fields in F and P as independent systems. The concept of independent system also remains valid as an approximation when the commutator is small.

b) Minkowski plane waves: Exactly the same content as the spacelike case.

c) Future-Past plane waves: For the future quadrant F we have the coordinate transformation

$$t = a^{-1}e^{a\eta}\cosh(a\zeta), \qquad z = a^{-1}e^{a\eta}\sinh(a\zeta). \tag{6.83}$$

For the past quadrant P we have the coordinate transformation

$$t = -a^{-1}e^{a\overline{\eta}}\cosh(a\overline{\zeta}), \qquad z = -a^{-1}e^{a\overline{\eta}}\sinh(a\overline{\zeta}). \tag{6.84}$$

Due to conformal invariance of the massless wave equation, the wave equations take the same form as (6.54). With analogous light cone coordinates,

$$\nu = \eta + \zeta, \qquad \overline{\nu} = -\overline{\eta} - \overline{\zeta}, \tag{6.85}$$

we obtain mode functions in these coordinates (like in the spacelike case). These modes are called conformal modes and are written as

$$g^F_\omega(\nu) = (4\pi\omega)^{-1/2}e^{-i\omega\nu}, \tag{6.86}$$

$$g^P_\omega(\overline{\nu}) = (4\pi\omega)^{-1/2}e^{-i\omega\overline{\nu}}. \tag{6.87}$$

These conformal modes resemble the Rindler modes (6.64) and (6.65). In fact we will show that these are the Rindler modes. Hence their annihilation operators define the Rindler vacuum

$$\hat{a}^F_{+\omega}\,|0_R\rangle = \hat{a}^P_{+\omega}\,|0_R\rangle = 0, \tag{6.88}$$

for all ω.

d) Bogoliubov transformation: The main construct we require, regardless of spacelike or timelike case, are the Minkowski plane waves. For the spacelike case, we utilized the coordinate transformation

$$V = a^{-1}e^{a\chi}, \qquad V = -a^{-1}e^{-a\overline{\chi}}, \tag{6.89}$$

and for the timelike case we find that

$$V = a^{-1}e^{av}, \qquad V = -a^{-1}e^{-a\overline{v}}. \tag{6.90}$$

This shows that the light cone coordinate, V, has the same functional relationship to χ as to v. More precisely, $g^R_\omega(\chi)$ and $g^F_\omega(v)$ are identical functions of V since $\chi(V) = v(V)$. This implies that they are made of the same combination of Minkowski plane waves. Mathematically, these are the same quantities. A similar relationship holds between $g^L_\omega(\chi)$ and $g^P_\omega(v)$. In other words, these conformal modes are alternative expression for Rindler modes. Therefore when we expand these modes into Minkowski plane waves

$$\theta(V)\, g^F_\omega(v) = \int_0^\infty dk\, (\alpha^F_{\omega k} u_k(V) + \beta^F_{\omega k} u^*_k(V)), \tag{6.91}$$

$$\theta(-V)\, g^P_\omega(\overline{v}) = \int_0^\infty dk\, (\alpha^P_{\omega k} u_k(V) + \beta^P_{\omega k} u^*_k(V)), \tag{6.92}$$

and compare to (6.68) and (6.69) we obtain relationships

$$\alpha^F_{\omega k} = \alpha^R_{\omega k}, \quad \beta^F_{\omega k} = \beta^R_{\omega k}, \quad \alpha^P_{\omega k} = \alpha^L_{\omega k}, \quad \beta^P_{\omega k} = \beta^L_{\omega k}. \tag{6.93}$$

e) F-P entanglement: From here, we carry the same procedure as the spacelike case except to replace labels R to F, and L to P. This produces the final result,

$$\boxed{|0_M\rangle = \prod_i C_i \sum_{n_i=0}^{\infty} \frac{e^{-\pi n_i \omega_i/a}}{n_i!} (\hat{a}^{F\dagger}_{\omega_i} \hat{a}^{P\dagger}_{\omega_i})^{n_i} |0_R\rangle} \tag{6.94}$$

where $C_i = \sqrt{1 - e^{-2\pi\omega_i/a}}$. This shows a *timelike entanglement* between the past and the future in the Minkowski vacuum.

6.4.3 Implications

We make a few remarks on the *F*-*P* entanglement expressed in (6.94), in particular on its similarity with the temporal entanglement in Chapter 4. Although the quantum field is causally disconnected between F and P, measurements in F (for example, projections onto g_ω-particle number) can collapse the state of the field in P. Similarly measurements in P should collapse the state in F. This is analogous to the properties of entanglement in time discussed in Chapter 4. The similarities to temporal Bell states (4.174) become more striking when in a recent work [132] it was theoretically shown that this future-past entanglement (6.94) could be extracted to a pair of qubits that do not coexist at the same time.

Another extraction for this timelike entanglement was proposed in [133]. This involved the use of two detectors. One of the detectors interacted with the vacuum in the past while the other detector waits and interacts with the future vacuum. The two detectors end up becoming entangled. More precisely the timelike entanglement in the Minkowski vacuum is converted into bipartite entanglement between detectors at a constant time. However, the procedure requires a particular time correlation for the extraction to optimally occur. This was stated more shockingly through an example as, *"a detector that is switched on and off in the vicinity of a quarter to 12:00 can become entangled with a detector interacting with the field at the same spatial location in the future, but only if the later detector waits to be switched on and off at a quarter past 12:00."* It is therefore not surprising to see that the existence of a time interval (in this case thirty minutes) is what makes the interdependence of this timelike entanglement shocking.

7

Conclusion

> *"My soul, my soul, where are you? Do you hear me? I speak, I call you–are you there? I have returned, I am here again. I have shaken the dust of all the lands from my feet, and I have come to you, I am with you. After long years of long wandering, I have come to you again..."*
>
> – Carl Jung, *The Black Books*

IN THE REALM of quantum physics, we have witnessed a shocking interdependence across time in a system of multiple particles (in Chapter 4), a single particle (in Chapter 5), and zero particles (in Chapter 6). We found that the entanglement in time (in Chapter 4) and the timelike entanglement (in Chapter 6) are similar in nature. Both utilize the same algebraic definition of nonseparability. Both can be expressed in terms of an entanglement between qubits that do not coexist. However, the timelike entanglement articulates itself more clearly through the light cone terminology. Despite the common use of referring to non-locality in time (in Chapter 5) as an entanglement in time, we argue for the distinction to be clearly made. Non-locality in time is not an algebraic definition of nonseperability but rather the property of experimental measurement correlations. Moreover, further care needs to be taken in the RQI case where the term locality is characterized in a vastly different manner. In this chapter, we summarize the main achievements of this thesis and discuss future projects that relate to its topic. They range from conservative next steps to imaginatively speculative paths.

7.1 Summary

This thesis provides one of the first systematic expositions on the concept of entanglement in time of quantum systems. Furthermore, the thesis contrasts it with the more familiar concept of entanglement in space. The similarities and differences between these concepts are examined in the context of quantum information, quantum foundations and relativistic quantum information.

The thesis also achieved various original contributions:

a) Quantum blockchain: We designed a quantum information application of entanglement in time, namely a quantum blockchain [135]. Most other applications of quantum information harness an entanglement in space. Though the literature does refer to a few other applications of entanglement in time, they are fundamentally modifications of the spatially entangled case. Therefore, our work can be regarded as the first novel application of entanglement in time.

b) Monty hall teleportation: We designed a Monty Hall version of quantum teleportation [71]. The teleportation protocol is one of the most explored topics in quantum information and our work has added novel techniques into this area. Future work may involve porting these techniques to other quantum protocols.

c) Teleportation involving noise: We developed a variation of the teleportation protocol for the effect of noise on teleportation [71]. This work could be of great applicability to practical quantum communication networks.

d) PBR game: We provided one of the first (if not the first full) reformulation of the Pusey-Barrett-Rudolph theorem into a quantum game [71]. Given that game-theoretic versions of the CHSH inequalities have played a non-trivial role in quantum information and its foundation, time will only tell how impactful our gamification of this recent foundation result will be.

e) Density matrix in Gleason's theorem: We provided an explicit construction of the density matrix in Gleason's theorem [162]. Such a construction was missing in the vast literature concerning the foundations of quantum physics.

f) Geometric proof of KS theorem: We constructed a simplified geometrical proof [176] of the Kochen-Specker theorem in quantum foundations.

7.2 Quantum Time Machines

Our exploration moved from classical information (in Chapter 2), towards quantum information (in Chapter 3), to finally an examination of the entanglements within quantum information science (in Chapter 4). Much like classical information was the resource for the development of the 'Information Age', one can imagine that quantum information may transform the world to a 'Quantum Information Age.' More pragmatically, the applications of quantum information can be regarded as a technological frontier. Others have harnessed quantum information to build teleportation systems or create the most powerful computers. In this thesis, we have used quantum information to design a quantum blockchain, which can be viewed as a 'quantum time machine.'

Classically, a time machine is any system that permits one to travel into the past, and a rigorous definition can be found in [4]. Using quantum physics, we believe that a broader class of time machines may be possible; this includes functionalities that we have not yet imagined. To be more precise, we define a *quantum time machine as any quantum system that can perform information tasks across time in classically impossible ways.* A more rigorous definition could perhaps be formulated with the use of constructs known as steering inequalities [264, 265].

Our view is that quantum time machines (instead of quantum teleportation or quantum computers) will be the most shocking applications of quantum information, and the most exciting technologies for the world's transformation to the Quantum Information Age. We proceed to outline three possible projects regarding these temporal-based quantum information technologies.

7.2.1 Temporal cryptography

In Chapter 4, entanglement in space was described through the tools of qubits, density operators and entropy. The entropic analysis was particularly useful in the construction of entropic uncertainty relations with a spatially entangled memory (4.71). These relations were ultimately related to the security of quantum cryptographic protocols (4.116).

For the entanglement in time in Chapter 4, a description through entropy was not found in the literature. Rather than providing an analysis using the Shannon or von Neumann entropy, a novel next step is to make use of certain temporal entropic quantities. In classical information theory, two relatively recent quantities in the analysis of temporal data are the transfer entropy [266, 267, 268] and the past entropy [269, 270]. Modifications of these quantities for the quantum case may allow one to better analyze the entanglement in time. Furthermore, it may allow for the derivation of a unique set of entropic uncertainty relations with a temporally entangled memory. Following the spatial case, this may lead to the development of temporal quantum cryptographic protocols, which secure information across time in classically impossible ways.

7.2.2 Network consensus

In our quantum blockchain, the entanglement in time was harnessed for the data structure component. However, we believe that one of the best applications of entanglement in time could be in the network consensus component.

Prior to the inception of blockchain systems, network consensus was considered to be central topic within the subject of distributed algorithms [271]. An important aspect to this subject is the timing model which captures the timing of events in a distributed computer network. This can be synchronous (processors performing communication and computation in perfect lock-step synchrony), completely asynchronous (taking steps at arbitrary speeds and arbitrary order), or partially synchronous (where processors have partial information about the timing of events). Given this temporal environment, one research direction would be to harness an entanglement in time to develop quantum distributed consensus algorithms that can outperform the classical algorithms within each of scenarios of the timing model.

A far more interesting path would be when one considers the network consensus protocols that blockchain systems have recently introduced. Advanced blockchain systems such as proof-of-elapsed-time systems [272] and hedara hashgraph [273] have a significant temporal property to their design. But far more

important is that blockchain consensus overall have a probabilistic aspect that makes their system operational. Given that an entanglement in time is a temporal phenomenon with (quantum) probabilistic properties, it may be the case that developing probabilistic consensus protocols is their 'killer app.' One can imagine that these protocols would allow a network to achieve consensus across time in classically impossible ways.

7.2.3 Temporal logical machines

One can think of a digital computer as a machine that carries out boolean logic. Quantum computers or more precisely the quantum circuit model can be thought of as a quantum analogy of boolean operators. As an example, one refers to the Pauli operator σ_x as the quantum NOT operator. There has been recent work to reformulate the quantum circuit model into other frameworks; one such example utilizes category theory [274, 275] which provides various advantages.

Using a similar line of reasoning, we speculate that an entanglement in time may not be best captured through the quantum circuit model which is founded on boolean logic. There exists a well developed field known as temporal logic [276, 277, 278] which involves various temporal logical operators. An ambitious path would be to develop a model that can be considered a quantum analogue of temporal logic. This may better capture the the effect of entanglement in time than the boolean logic inspired quantum circuit model. We speculate that such a framework may lead to the derivation of radically new types of time machines i.e. temporal logical machines that may be as revolutionary as digital computing.

7.3 What is Quantum Information?

Entanglement in time is a most shocking temporal effect which is fundamentally mysterious. Quantum information (ie quantum state), such as the complex numbers in (3.40), allowed us to mathematically express this entanglement in time. We viewed similar temporal effects from a foundational perspective in Chapter 5 as well as in the relativistic regime in Chapter 6. If quantum information rep-

resents a physical quantity, then an entanglement in time magnifies the disruption of quantum physics onto the classical temporal world far greater than any other effect. Furthermore in Quantum Foundations, holding an epistemic view of quantum information has drastic consequences, and in RQI the unification of quantum information with the Einstein equations is still the most important open problem in theoretical physics. Therefore, from such a wide exploration we come to appreciate the most fundamental mystery: What is quantum information?

In this thesis, we take the view that there are two separate problems, a theoretically inclined problem and a physically inclined problem. The theoretical problem is what does quantum information or the quantum state physically represent? The physical problem is what is the physical state of the quantum system when it is not observed? (A refinement of the latter question is: Where is the mass of a quantum particle located or distributed when we do not observe it?) The connection between the quantum state and its corresponding unobserved system is not direct within the postulates of quantum theory (note that a quantum state is not a probability distribution; probability requires squaring it first). This lack of direct connection provides the ambiguity that results in the inception of the two problems described. We present three directions that relate to entanglement in time that could lead to an advancement towards answering these two problems.

7.3.1 Null tetrads

RQI is currently presented through the metric formulation of general relativity. There is an equivalent picture of general relativity known as the null tetrad formulation [279, 280]. It views the light cone structure as the fundamental entity, and elevates complex numbers as central quantities within relativity.

More precisely, one can define a set of light-like or null vectors as a basis. By tetrad, this implies a basis of four vectors. Hence, at each point on the spacetime manifold, there are four null vectors $l^a, n^a, m^a, \overline{m}^a$ with specific properties that we shall describe. The vectors l^a and n^a are real and satisfy $l^a\, n_a = 1$. The other two vectors, $m^a, \overline{m}^a$ are complex null vectors and have the property that they are complex conjugates of each other and satisfy the condition, $m^a\, \overline{m}_a = -1$. The

relationship of the null tetrad to the metric tensor (6.7) can be expressed as

$$g_{ab} = l_a\, n_b + n_a\, l_b - m_a\, \overline{m}_b - \overline{m}_a\, m_b, \tag{7.1}$$

$$g^{ab} = l^a\, n^b + n^a\, l^b - m^a\, \overline{m}^b - \overline{m}^a\, m^b. \tag{7.2}$$

The spacelike and timelike entanglements in RQI were based on regions separated by the light cone structure. Perhaps by using the null tetrad formulation, one may be able to express these entanglements more efficiently. This may provide some ideas on generalizing these entanglements in non-trivial curved spacetimes settings. This may lead to the discovery of a novel set of entanglements.

However, a far more ambitious reason is that since quantum information is complex valued, its unification to relativity may require relativistic structures to be complex valued [281]. This complex valued unification towards quantum gravity was already outlined in [31]. The novelty would be whether a null tetrad formulation of RQI could lead to a better insight into this program. This may help us answer the question of what is quantum information from the perspective of complex valued relativistic structures.

7.3.2 'Spacetime information theory'

Entanglement in time involves an interplay of both quantum physics and time at a fundamental level. In this thesis we have investigated quantum physics through an information-theoretic perspective. Could time itself also be studied using an information-theoretic method? If so, would that help understand entanglement in time (and ultimately quantum information) in a deeper way? Such a curiosity aligns to a program set by John Wheeler known as 'It from Bit' [282].

a) It from Bit: This research program puts forth the notion that the physical world emerges from information. Wheeler hypothesized that every physical quantity derives its ultimate significance from bits, and we quote, *"...every it – every particle, every field of force, even the spacetime continuum itself – derives its function, its meaning, its very existence entirely – even if in some contexts indirectly – from the apparatus-elicited answers to yes or no questions, binary choices, bits...all physical things are information-theoretic in origin"*. He emphasized the

goal to carry out such a program and we quote *"Tomorrow we will have learned to understand and express all of physics in the language of information"*. He proceeds to set an agenda by saying *"...capitalize on the findings and outlooks of information theory...search out every link each has with physics..."*.

We aim to contribute to Wheeler's program by outlining an *original* [283] speculative path on how time (and space) could be viewed from an information-theoretic perspective. Before embarking on these ideas, it wise to first identify the non-trivial concept that underpins all information theories.

b) Compression: Our view is that all information theories are fundamentally about compression and not about information. To support the previous statement, we examine several information theories and find that the concept of compression does indeed exist as the foundational result for each of them.

For the case of classical information theory, the fundamental result is the noiseless coding theorem (Theorem 2.3). It highlights that the Shannon entropy $H(X)$ can be operationally defined in terms of optimal compression. More precisely, for a sequence of n two-outcome random variables, uncompressed n bits can be optimally compressed to $H(X)n$ bits:

$$H(X)n_{uncompressed} = n_{compressed}. \tag{7.3}$$

Quantum information theory provides the quantum generalization of classical information theory. The fundamental result in this theory is the Schumacher's noiseless coding theorem (Theorem 3.4). This articulates that the von Neumann entropy $S(\rho)$ can be operationally defined in terms of optimal compression. More specifically, uncompressed n qubits can be optimally compressed to $S(\rho)n$ qubits:

$$S(\rho)n_{uncompressed} = n_{compressed}. \tag{7.4}$$

Therefore we see that compression is the key idea in the fundamental results of these theories, and it serves to provide the operational definition of the entropies.

The field of algorithmic information theory [284, 285], which we have not examined in this thesis, can also be seen to be fundamentally about compression. This

particular information theory is based on a notion of the information content of an individual object using concepts from computability theory. This is in contrast to classical information theory which is built on probability distributions. To elaborate, the information content of an object such as a finite binary string x can be captured by the algorithmic entropy $K(x)$ (also known as the Kolmogorov complexity) which is defined as the length of the shortest binary computer program that can produce x as the output. In other words, the algorithmic entropy of a string provides the length of its shortest possible compression. As an example, the string

$$0101010101010101010101010101010101010101010101010101010101010101 \quad (7.5)$$

has the compressed program description "31 repetitions of 01". This is opposed to the string

$$1100100001100001110111101110110011111010010000100101011110010 1 \quad (7.6)$$

which has no shorter description than writing down the string itself.

Using the algorithmic entropy, one can arrive at a definition of randomness of a single individual sequence which is not possible using probabilistic theories. The fundamental idea is that randomness equates to incompressibility. The inability to compress highlights that there is no structure of pattern to provide a concise description (as witnessed in string (7.6)). Stated more formally, a finite string x is random if and only if the algorithmic entropy of the string $K(x)$ is no shorter than the length of the string.

Therefore we see compression existing at a foundational level in all information theories. However the ambiguity arises when one probes what the information in each of the information theories is supposed to represent; classical information is fit for engineering purposes with no regard to meaning [11]; nobody knows what quantum information is [182]; a similar ambiguity exists on what algorithmic information is about [284]. Given each of the information theories has an associated entropy, perhaps that may add some clarification. Such a hope is quickly diminished when one reads [286] that Shannon called his term the

entropy because von Neumann suggested one reason as "*...nobody knows what entropy really is, so in a debate you will always have the advantage.*"

Our view to resolve the confusion is remove the focus on information in information theories. Rather it is compression in these theories that is fundamental. Building on these observations, we state that a structure that is compressible is what should constitute information, and the quantity involved in optimal compression is what should be termed the entropy.

c) Spacetime compression: If one takes 'It from Bit' as the underlying principle of the Universe, then there ideally must exist an information theory associated to time and space, just as there exists information theories for both classical and quantum systems. More precisely, this research direction demands the development of a 'spacetime information theory.' From an alternative direction, the notion that spacetime itself contains information is already alluded to by the Bekenstein-Hawking entropy (6.81). In fact the original inception of that result was based on observations regarding similarities between spacetime physics and aspects of information [287].

Furthermore if one takes compression as the primary information-theoretic technique, then such a spacetime information theory should be built on some mathematical notion of compression. To explore this idea, we provide a speculative path through a heuristic argument. We emphasize that these ideas are underdeveloped but proceed with the intention of conveying possibilities.

Our immediate aim is to simply identify whether a notion of compression can be found in the theory of relativity, which is our current framework for understanding spacetime. And predicated on that, develop various inferences that may lead to insights for developing a formal spacetime information theory on a firm basis.

We suggest the following idea: *A time interval itself should be treated as a form of information, and that time dilation as expressed in (6.1) can then be seen as a form of information-theoretic decompression.*

d) Mathematical details: To express this idea mathematically, we require a 'data compression' entropy, analogous to the other information theories. We identify this by suggesting that the reciprocal of the Lorentz factor (6.2) is an

entropy as it 'compresses' time intervals. Mathematically, this can be expressed as

$$\alpha = H(X), \tag{7.7}$$

where $\alpha \equiv 1/\gamma = \sqrt{1-(v^2/c^2)}$. (This is analogous to $S(\rho) = H(\lambda_x)$ in quantum information theory). For an uncompressed time interval $\Delta t_{uncompressed}$, optimal compression is achieved using the time dilation formula

$$\alpha \Delta t_{uncompressed} = \Delta t_{compressed}. \tag{7.8}$$

Note that this is analogous to the compression formulas in the classical (7.3) and quantum (7.4) cases. The bounds for this α-entropy are $0 \leq \alpha \leq 1$, with $\alpha = 0$ when $v = c$. The uncompressed interval can be thought of as the case when maximum entropy occurs.

A particular physical realization of this is when a given coordinate time Δt is contracted to various proper times $\Delta\tau$ depending on the velocity of different observers, ie $\Delta t = \gamma_1 \Delta\tau_1$, $\Delta t = \gamma_2 \Delta\tau_2$, etc. Different velocities compress a fixed time interval differently, and hence we can identify this entropy with velocity.

Using this assumption, we can state some similarities to the classical and quantum information theories. In classical information theory, the mutual information, $H(X{:}Y)$, of X and Y measure how much information X and Y have in common; the quantum mutual information for systems A and B is denoted by $S(A{:}B)$. This notion of commonality can be captured using the physical scenario of the relativistic velocity, $v_r = (v_1 - v_2)/(1-(v_1 v_2/c^2))$, of two observers, v_1 and v_2 with respect to a fixed coordinate time interval. Each observer can be identified with a compression entropy, $\alpha_1 = 1/\gamma_1$ and $\alpha_2 = 1/\gamma_2$. Their relative Lorentz factor, $\gamma_r = \gamma_1\gamma_2(1-(v_1 v_2/c^2))$, provides the inspiration to define the relativistic mutual information between α_1 and α_2:

$$\alpha_{1:2} \equiv \frac{1}{\gamma_r} = \frac{1}{\gamma_1\gamma_2(1-\frac{v_1 v_2}{c^2})} = \frac{\alpha_1\alpha_2}{(1-\frac{v_1 v_2}{c^2})}. \tag{7.9}$$

The classical joint entropy, $H(A,B)$, and the quantum joint entropy, $S(A,B)$, help

us define the relativistic joint entropy:

$$H(X, Y) = H(X) + H(Y) - H(X{:}Y), \tag{7.10}$$

$$S(A, B) = S(A) + S(B) - S(A{:}B), \tag{7.11}$$

$$\alpha_{1,2} \equiv \alpha_1 + \alpha_2 - \alpha_{1:2}. \tag{7.12}$$

Similarly, the relativistic conditional entropy can be developed using the classical and quantum analogue:

$$H(X|Y) \equiv H(X, Y) - H(Y), \tag{7.13}$$

$$S(A|B) \equiv S(A, B) - S(B), \tag{7.14}$$

$$\alpha_{1|2} \equiv \alpha_{1,2} - \alpha_2. \tag{7.15}$$

From here, we proceed to derive entropic properties concerning two systems and compare this with the classical and quantum case. For example, it can easily be shown that $\alpha_{1:2} = \alpha_{2:1}$ and $\alpha_{1,2} = \alpha_{2,1}$ which is like the classical case of $H(X{:}Y) = H(Y{:}X)$ and $H(X, Y) = H(Y, X)$. Furthermore, $H(Y|X) \geq 0$ and $H(X) \leq H(X, Y)$ with equality satisfied for each inequality, if and only if Y is a function of X; the last two properties fail for the quantum case, in particular for systems involving entanglement; for our relativistic case, $\alpha_{2|1} \geq 0$ and $\alpha_1 \leq \alpha_{1,2}$ is satisfied as long as $\gamma_r \geq \gamma_2$, ie physically the relative Lorentz factor has to be greater than or equal to the Lorentz factor of the second observer. The property of subadditivity holds for all three cases: $H(X, Y) \leq H(X) + H(Y)$, $S(A, B) \leq S(A) + S(B)$, and $\alpha_{1,2} \leq \alpha_1 + \alpha_2$ (which can be reduced to $\alpha_{1:2} \geq 0$). Classically, $H(Y|X) \leq H(Y)$ and it can be shown that $\alpha_{1|2} \leq \alpha_1$ (which can be reduced to $\alpha_{1:2} \geq 0$). In the last classical equation as well as second to last equation of classical subaddivity, equality is expressed if X and Y are independent variables. In the last two analogous relativistic equations, equality is expressed if $\alpha_{1:2} = 0$, which is when one of the observers has velocity c. Thus, the notion of 'independence' enters when one of the observers is moving at the speed of light.

To draw out physical implications, we assume this information-theoretic compression has the mathematical backbone of classical and quantum information theory. In classical information theory, a central result was the construction of ϵ-

typical sequences (2.41) with Theorem 2.2. A similar result held in the quantum information theory with ϵ-typical states with Theorem 3.3. In this 'spacetime information' case, one would ideally need to define an time interval Δt that is ϵ-typical. Hence in an analogous manner, time intervals come in typical and atypical forms. If this relativistic case is similar to the classical and quantum case, then one can easily see that typical time intervals would be those that can be compressed from Δt to $\alpha\Delta t$. Physically, this means typical time intervals obey the relativistic time dilation formula. Therefore, atypical time intervals are those that exhibit Lorentz violations. Furthermore, if a similar information theory result occurs in this case, then the probability that a time interval is ϵ-typical is

$$\Pr\{T_\epsilon^{(\Delta t)}\} > 1 - \epsilon \tag{7.16}$$

for sufficiently large duration Δt. Hence in this setting, Lorentz violations do occur for large time intervals but very rarely. In this sense, Lorentz violations is fundamentally not a matter of scale, but rather of probability; hence this violation may be experimentally detectable at large scales given a very large sample size.

It's important to note that this idea also applies to space intervals. For the x-direction, $\alpha\Delta x_{uncompressed} = \Delta x_{compressed}$, where the compressed space interval is due to length contraction. In the data compression subsection, it can be seen that optimal compression is achieved using $nH(X)$ bits (or $nS(\rho)$ qubits). For the case $nR > nH(X)$, compression without loss of information is achieved, but is not optimal. For $nR < nH(X)$, information is lost and compression is not reliable. These results seem to correspond to Lorentz transformations, which can be re-written as $\alpha\Delta t_{uncompressed} = \Delta t \pm \frac{v}{c^2}\Delta x$ where $\alpha\Delta t_{uncompressed} = \Delta t_{compressed}$ is the optimally compressed interval. The Lorentz equation for the the minus case, leads to $\Delta t > \Delta t_{compressed}$; this can be interpreted as compression is achieved but not optimally. For the plus case, $\Delta t < \Delta t_{compressed}$, which means some information is lost and gone to Δx.

e) Comments:

i) These set of ideas are merely speculations at this point and were inspired by Wheeler's program of 'It from Bit.' In our thesis it is precisely a time interval that was central to the shocking nature of the entanglements. Whether

a time interval could be treated as information in the way we stated is an exciting premise but requires much greater exploration.

ii) There are various problems with this heuristic argument. The first is regarding the problem on how this idea would work given the relative nature of Lorentz transformations between observers. A related problem is that information in standard compression is carried on less bits while boosting a particle contracts or expands its time completely relative to the reference frame the particle is observed in. With respect to these and other issues, it is important to emphasize that this is merely a heuristic argument for providing insights to develop a formal spacetime information theory. The undertaking of this latter subject would of course require far more sophisticated mathematical constructs along with appropriate postulates.

iii) However within such a formal spacetime information theory, we do believe compression will be found in its fundamental equations and help us view the Universe as Wheeler intended. Would the expansion of the Universe turn out to be an information-theoretic decompression? Would the Bekenstein-Hawking entropy correspond to an optimal compression? Through a unification with quantum information theory, would it help us achieve quantum gravity? And would that let us finally answer the question what is quantum information (and its related intrinsic randomness)?

7.3.3 Unordered time

Our current view of time is that it obeys a mathematical real number axis with an ordering from smaller values to larger values. The final speculation we state is that time may be physically unordered in the quantum realm compared to the classical realm. We believe that a development of this idea may provide a new interpretation of quantum physics, in particular to answering the physical and theoretical inclined problems that we stated earlier. Perhaps the shocking properties are nothing intrinsic with the quantum particles but rather a consequence of an unordering of time. When one observes (from the ordered world) a particle at various times, it gives the impression that the particle is behaving in a paradoxical manner. A measurement can be defined as the moment at which the temporal

ordered classical world meets the temporal unordered quantum world. From a theoretical view, the quantum state (ie quantum information) is a complex valued quantity and complex numbers are mathematically unordered; perhaps the quantum state through its unordered complex numbers is about spacetime and the Schrödinger equation is an evolution of spacetime in a unordered temporal manner; the dynamics of this unordering of time may stem from the energy associated with the quantum system in question. Providing a generalization of this bare idea with a mathematical framework is left for future work.

We do want to emphasize certain works and ideas that served as stimulus for such a speculation. First were the results on quantum causality [234, 235], and in particular a Bell's theorem for temporal order [288]. Relating to the unordered nature of complex numbers, the second influence comes from the use of imaginary time through the Wick rotation, as well as through the use of complex valued spacetime transformations in the Newman-Janis trick [281]; there are no deep reasons at present for why these tricks work. The third influence comes from the 'spacetime information theory' ideas we set out earlier; this suggests that time can be compressed in an information-theoretic manner; given that compression techniques can involve data reordering or a reduction of data this also leads to a notion of unordering in time. The final influence is that the metric fluctuations on the sub-Planckian scale is completely unknown [4]; a non-trivial metric fluctuation may provide the necessary basis for this unordered time.

In closing, we want to provide some clarity on the historical aspect of the subject. It is often emphasized that Einstein was critical of quantum theory. But of far greater importance is that it should be stated that he was one of its pivotal founders [289, 290]. In fact it was the discovery of entanglement (in space) in the EPR paper [52] that is the most cited of all his works! In align with the theme of this thesis, it should be noted that he also emphasized a time interval in his quest to truly understand quantum physics: *"All the fifty years of conscious brooding have brought me no closer to answer the question, 'What are light quanta?' Of course today every rascal thinks he knows the answer, but he is deluding himself."*

Bibliography

[1] D. Jennings and M. Leifer. No return to classical reality. *Contemporary Physics*, 57(1):60–82, 2016.

[2] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.

[3] E. Schrödinger. Discussion of probability relations between separated systems. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 31, pages 555–563. Cambridge University Press, 1935.

[4] M. Visser. *Lorentzian Wormholes: from Einstein to Hawking*. AIP-Press, 1996.

[5] T. C. Ralph and T. G. Downes. Relativistic quantum information and time machines. *Contemporary Physics*, 53(1):1–16, 2012.

[6] D. Applebaum. *Probability and Information: An Integrated Approach*. Cambridge University Press, 2008.

[7] A. Rodriguez and B. Mendes. *Probability, Decisions and Games: A Gentle Introduction Using R*. John Wiley & Sons, 2018.

[8] J. S. Rosenthal. Monty hall, monty fall, monty crawl. *Math Horizons*, 16(1):5–7, 2008.

[9] R. D. Gill. Monty hall problem. *International Encyclopaedia of Statistical Science*, pages 858–863, 2010.

[10] S. Lucas, J. Rosenhouse, and A. Schepler. The monty hall problem, reconsidered. *Mathematics Magazine*, 82(5):332–342, 2009.

[11] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[12] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.

[13] M. M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2017.

[14] A. M. Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42:230–265, 1936.

[15] R. Downey. *Turing's Legacy: Developments from Turing's Ideas in Logic*, volume 42. Cambridge University Press, 2014.

[16] C. Moore and S. Mertens. *The Nature of Computation*. OUP Oxford, 2011.

[17] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008.

[18] A. Narayanan and J. Clark. Bitcoin's academic pedigree. *Communications of the ACM*, 60(12):36–45, 2017.

[19] C. Cachin and M. Vukolić. Blockchain consensus protocols in the wild. *arXiv preprint arXiv:1707.01873*, 2017.

[20] W. Mao. *Modern Cryptography: Theory and Practice*. Prentice Hall PTR, 2004.

[21] S. Bano, A. Sonnino, M. Al-Bassam, S. Azouvi, P. McCorry, S. Meiklejohn, and G. Danezis. Consensus in the age of blockchains. *arXiv preprint arXiv:1711.03936*, 2017.

[22] L. Lamport, R. Shostak, and M. Pease. The byzantine generals problem. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 4(3):382–401, 1982.

[23] M. Castro and B. Liskov. Practical byzantine fault tolerance. In *OSDI*, volume 99, pages 173–186, 1999.

[24] D. J. Yaga, P. M. Mell, N. Roby, and K. Scarfone. Blockchain technology overview. *arXiv preprint arXiv:1906.11078*, 2019.

[25] S. D. Galbraith. *Mathematics of Public Key Cryptography*. Cambridge University Press, 2012.

[26] J. J. Sakurai. *Modern Quantum Mechanics, Revised Edition*. Addison Wesley, 1993.

[27] R. Shankar. *Principles of Quantum Mechanics*. Springer Science & Business Media, 2012.

[28] A. Bokulich and G. Jaeger. *Philosophy of Quantum Information and Entanglement*. Cambridge University Press, 2010.

[29] A. Einstein and L. Infeld. *The Evolution of Physics*. Cambridge University Press, 1938.

[30] F. Flamini, N. Spagnolo, and F. Sciarrino. Photonic quantum information processing: a review. *Reports on Progress in Physics*, 82(1):016001, 2018.

[31] R. Penrose. *The Road to Reality: A Complete Guide to the Laws of the Universe*. Vintage Books, 2007.

[32] D. Dieks. Communication by EPR devices. *Physics Letters A*, 92(6):271–272, 1982.

[33] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802, 1982.

[34] D. McMahon. *Quantum Computing Explained*. John Wiley & Sons, 2007.

[35] D. Ahn, C. R. Myers, T. C. Ralph, and R. B. Mann. Quantum-state cloning in the presence of a closed timelike curve. *Physical Review A*, 88(2):022332, 2013.

[36] T. A. Brun, M. M. Wilde, and A. Winter. Quantum state cloning using deutschian closed timelike curves. *Physical Review Letters*, 111(19):190401, 2013.

[37] A. K. Pati and S. L. Braunstein. Impossibility of deleting an unknown quantum state. *Nature*, 404(6774):164, 2000.

[38] T. A. Brun, J. Harrington, and M. M. Wilde. Localized closed timelike curves can perfectly distinguish quantum states. *Physical Review Letters*, 102(21):210402, 2009.

[39] J. L. Pienaar, T. C. Ralph, and C. R. Myers. Open timelike curves violate heisenberg's uncertainty principle. *Physical Review Letters*, 110(6):060501, 2013.

[40] C. W. Gardiner. *Quantum Noise*. Springer-Verlag, 1991.

[41] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher. Noncommuting mixed states cannot be broadcast. *Physical Review Letters*, 76(15):2818, 1996.

[42] C. M. Caves, C. A. Fuchs, and R. Schack. Conditions for compatibility of quantum-state assignments. *Physical Review A*, 66(6):062111, 2002.

[43] T. Heinosaari and O. Kerppo. Antidistinguishability of pure quantum states. *Journal of Physics A: Mathematical and Theoretical*, 51(36):365303, 2018.

[44] R. Han, G. Leuchs, and M. Grassl. Residual and destroyed accessible information after measurements. *Physical Review Letters*, 120(16):160501, 2018.

[45] D. Deutsch. Uncertainty in quantum measurements. *Physical Review Letters*, 50(9):631, 1983.

[46] K. Kraus. Complementary observables and uncertainty relations. *Physical Review D*, 35(10):3070, 1987.

[47] H. Maassen and J. B. Uffink. Generalized entropic uncertainty relations. *Physical Review Letters*, 60(12):1103, 1988.

[48] P. J. Coles, M. Berta, M. Tomamichel, and S. Wehner. Entropic uncertainty relations and their applications. *Reviews of Modern Physics*, 89(1):015002, 2017.

[49] P. J. Coles, V. Katariya, S. Lloyd, I. Marvian, and M. M. Wilde. Entropic energy-time uncertainty relation. *Physical Review Letters*, 122(10):100401, 2019.

[50] E. Witten. A mini-introduction to information theory. *arXiv preprint arXiv:1805.11965*, 2018.

[51] B. Schumacher. Quantum coding. *Physical Review A*, 51(4):2738, 1995.

[52] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47(10):777, 1935.

[53] J. Yin et al. Satellite-based entanglement distribution over 1200 kilometers. *Science*, 356(6343):1140–1144, 2017.

[54] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Reviews of Modern Physics*, 81(2):865, 2009.

[55] O. Gühne and G. Tóth. Entanglement detection. *Physics Reports*, 474(1-6):1–75, 2009.

[56] V. Vedral. *Introduction to Quantum Information Science*. Oxford University Press on Demand, 2006.

[57] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner. Bell nonlocality. *Reviews of Modern Physics*, 86(2):419, 2014.

[58] A. Einstein and M. Born. *The Born-Einstein Letters 1916–1955: Friendship, Politics and Physics in Uncertain Times*. Macmillan, 2004.

[59] S. R. Moulick and P. K. Panigrahi. Timelike curves can increase entanglement with LOCC. *Scientific Reports*, 6:37958, 2016.

[60] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880, 1969.

[61] S. Popescu and D. Rohrlich. Generic quantum nonlocality. *Physics Letters A*, 166(5-6):293–297, 1992.

[62] C. Simon. Towards a global quantum network. *Nature Photonics*, 11(11):678, 2017.

[63] S. Wehner, D. Elkouss, and R. Hanson. Quantum internet: A vision for the road ahead. *Science*, 362(6412):eaam9288, 2018.

[64] C. H. Bennett and S. J. Wiesner. Communication via one-and two-particle operators on einstein-podolsky-rosen states. *Physical Review Letters*, 69(20):2881, 1992.

[65] B. P. Williams, R. J. Sadlier, and T. S. Humble. Superdense coding over optical fiber links with complete bell-state measurements. *Physical Review Letters*, 118(5):050501, 2017.

[66] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels. *Physical Review Letters*, 70(13):1895, 1993.

[67] J. Ren et al. Ground-to-satellite quantum teleportation. *Nature*, 549(7670):70, 2017.

[68] W. Li, C. Li, and G. Guo. Probabilistic teleportation and entanglement matching. *Physical Review A*, 61(3):034301, 2000.

[69] H. Lu and G. Guo. Teleportation of a two-particle entangled state via entanglement swapping. *Physics Letters A*, 276(5-6):209–212, 2000.

[70] P. Agrawal and A. K. Pati. Probabilistic quantum teleportation. *Physics Letters A*, 305(1-2):12–17, 2002.

[71] D. Rajan and M. Visser. Quantum PBR theorem as a monty hall game. *Quantum Reports*, 2(1):39–48, 2020.

[72] R. Fortes and G. Rigolin. Fighting noise with noise in realistic quantum teleportation. *Physical Review A*, 92(1):012338, 2015.

[73] R. Fortes and G. Rigolin. Probabilistic quantum teleportation in the presence of noise. *Physical Review A*, 93(6):062330, 2016.

[74] L. T. Knoll, C. T. Schmiegelow, and M. A. Larotonda. Noisy quantum teleportation: An experimental study on the influence of local environments. *Physical Review A*, 90(4):042332, 2014.

[75] G. G. Carlo, G. Benenti, and G. Casati. Teleportation in a noisy environment: a quantum trajectories approach. *Physical Review Letters*, 91(25):257903, 2003.

[76] D. Kumar and P. N. Pandey. Effect of noise on quantum teleportation. *Physical Review A*, 68(1):012317, 2003.

[77] L. Chen et al. *Report on post-quantum cryptography*. US Department of Commerce, National Institute of Standards and Technology, 2016.

[78] G. Alagic et al. *Status report on the first round of the NIST post-quantum cryptography standardization process*. US Department of Commerce, National Institute of Standards and Technology, 2019.

[79] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden. Quantum cryptography. *Reviews of Modern Physics*, 74(1):145, 2002.

[80] C. H. Bennett and G. Brassard. Quantum cryptography: public key distribution and coin tossing. In *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, pages 175–179.

[81] A. K. Ekert. Quantum cryptography based on bell's theorem. *Physical Review Letters*, 67(6):661, 1991.

[82] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner. The uncertainty principle in the presence of quantum memory. *Nature Physics*, 6(9):659, 2010.

[83] M. Hillery, V. Bužek, and A. Berthiaume. Quantum secret sharing. *Physical Review A*, 59(3):1829, 1999.

[84] P. Benioff. The computer as a physical system: a microscopic quantum mechanical hamiltonian model of computers as represented by turing machines. *Journal of Statistical Physics*, 22(5):563–591, 1980.

[85] Y. Manin. Computable and uncomputable. *Sovetskoye Radio, Moscow*, 1980.

[86] R. P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488, 1982.

[87] R. Jozsa and N. Linden. On the role of entanglement in quantum-computational speed-up. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 459(2036):2011–2032, 2003.

[88] G. Vidal. Efficient classical simulation of slightly entangled quantum computations. *Physical Review Letters*, 91(14):147902, 2003.

[89] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser. Quantum computation by adiabatic evolution. *arXiv preprint quant-ph/0001106*, 2000.

[90] M. W. Johnson et al. Quantum annealing with manufactured spins. *Nature*, 473(7346):194–198, 2011.

[91] J. K. Pachos. *Introduction to Topological Quantum Computation*. Cambridge University Press, 2012.

[92] R. Raussendorf, D. E. Browne, and H. J. Briegel. Measurement-based quantum computation on cluster states. *Physical Review A*, 68(2):022312, 2003.

[93] F. Arute et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

[94] A. Montanaro. Quantum algorithms: an overview. *NPJ Quantum Information*, 2:15023, 2016.

[95] A. M. Childs and W. van Dam. Quantum algorithms for algebraic problems. *Reviews of Modern Physics*, 82(1):1, 2010.

[96] L. K. Grover. Quantum mechanics helps in searching for a needle in a haystack. *Physical Review Letters*, 79(2):325, 1997.

[97] P. W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pages 124–134. Ieee, 1994.

[98] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM review*, 41(2):303–332, 1999.

[99] J. Katz and Y. Lindell. *Introduction to Modern Cryptography*. Chapman and Hall/CRC, 2014.

[100] D. Deutsch. *The Fabric of Reality*. Penguin UK, 1998.

[101] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT press, 2016.

[102] D. Silver et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484, 2016.

[103] D. Silver et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354, 2017.

[104] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.

[105] M. Schuld and N. Killoran. Quantum machine learning in feature hilbert spaces. *Physical Review Letters*, 122(4):040504, 2019.

[106] V. Havlíček et al. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209, 2019.

[107] P. Rebentrost, M. Mohseni, and S. Lloyd. Quantum support vector machine for big data classification. *Physical Review Letters*, 113(13):130503, 2014.

[108] S. Hameroff and R. Penrose. Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. *Mathematics and Computers in Simulation*, 40(3-4):453–480, 1996.

[109] S. Hameroff. Quantum computation in brain microtubules? the penrose–hameroff 'orch or 'model of consciousness. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 356(1743):1869–1896, 1998.

[110] R. Penrose. *The Large, the Small and the Human Mind.* Cambridge University Press, 2000.

[111] E. Megidish, A. Halevy, T. Shacham, T. Dvir, L. Dovrat, and H. S. Eisenberg. Entanglement swapping between photons that have never coexisted. *Physical Review Letters*, 110(21):210403, 2013.

[112] E. Megidish, T. Shacham, A. Halevy, L. Dovrat, and H. S. Eisenberg. Resource efficient source of multiphoton polarization entanglement. *Physical Review Letters*, 109(8):080504, 2012.

[113] E. Megidish, A. Halevy, Y. Pilnyak, A. Slapa, and H. S. Eisenberg. Quantum tomography of inductively-created large multiphoton states. *arXiv preprint arXiv:1712.03633*, 2017.

[114] M. Victora, F. Kaneda, F. Bergmann, J. J. Wong, A. Graf, and P. Kwiat. Time-multiplexed methods for optical quantum information processing. In *Quantum Photonics: Pioneering Advances and Emerging Applications*, pages 179–206. Springer, 2019.

[115] M. Zukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert. "event-ready-detectors"bell experiment via entanglement swapping. *Physical Review Letters*, 71:4287–4290, 1993.

[116] A. Zeilinger. Light for the quantum. entangled photons and their applications: a very personal perspective. *Physica Scripta*, 92(7):072501, 2017.

[117] A. M. Goebel et al. Multistage entanglement swapping. *Physical Review Letters*, 101(8):080403, 2008.

[118] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller. Quantum repeaters: the role of imperfect local operations in quantum communication. *Physical Review Letters*, 81(26):5932, 1998.

[119] A. Peres. Delayed choice for entanglement swapping. *Journal of Modern Optics*, 47(2-3):139–143, 2000.

[120] X. Ma et al. Experimental delayed-choice entanglement swapping. *Nature Physics*, 8(6):479, 2012.

[121] X. Ma, J. Kofler, and A. Zeilinger. Delayed-choice gedanken experiments and their realizations. *Reviews of Modern Physics*, 88(1):015005, 2016.

[122] S. Yokoyama et al. Ultra-large-scale continuous-variable cluster states multiplexed in the time domain. *Nature Photonics*, 7(12):982–986, 2013.

[123] J. G. Cramer. The transactional interpretation of quantum mechanics. *Reviews of Modern Physics*, 58(3):647, 1986.

[124] R. P. Feynman and L. M. Brown. *Feynman's Thesis: A New Approach to Quantum Theory*. World Scientific, 2005.

[125] Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz. Time symmetry in the quantum process of measurement. *Physical Review*, 134:B1410, 1964.

[126] Y. Aharonov and L. Vaidman. Properties of a quantum system during the time interval between two measurements. *Physical Review A*, 41:11, 1990.

[127] M. Nowakowski, E. Cohen, and P. Horodecki. Entangled histories versus the two-state-vector formalism: Towards a better understanding of quantum temporal correlations. *Physical Review A*, 98(3):032312, 2018.

[128] C. Callender. *The Oxford Handbook of Philosophy of Time*. OUP Oxford, 2011.

[129] D. Zimmerman. *Presentism and the space-time manifold*. Citeseer, 2011.

[130] S. Taylor, S. Cheung, Č. Brukner, and V. Vedral. Entanglement in time and temporal communication complexity. In *AIP Conference Proceedings*, volume 734, pages 281–284. AIP, 2004.

[131] F. Morikoshi. Information-theoretic temporal bell inequality and quantum computation. *Physical Review A*, 73(5):052308, 2006.

[132] C. Sabín, B. Peropadre, M. del Rey, and E. Martín-Martínez. Extracting past-future vacuum correlations using circuit QED. *Physical Review Letters*, 109(3):033602, 2012.

[133] S. J. Olson and T. C. Ralph. Extraction of timelike entanglement from the quantum vacuum. *Physical Review A*, 85(1):012306, 2012.

[134] T. C. Ralph and N. Walk. Quantum key distribution without sending a quantum signal. *New Journal of Physics*, 17(6):063008, 2015.

[135] D. Rajan and M. Visser. Quantum blockchain using entanglement in time. *Quantum Reports*, 1(1):3–11, 2019.

[136] W. McCutcheon et al. Experimental verification of multipartite entanglement in quantum networks. *Nature communications*, 7:13251, 2016.

[137] D. Aggarwal, G. K. Brennen, T. Lee, M. Santha, and M. Tomamichel. Quantum attacks on bitcoin, and how to protect against them. *arXiv preprint arXiv:1710.10377*, 2017.

[138] Y. Gao, X. Chen, Y. Chen, Y. Sun, X. Niu, and Y. Yang. A secure cryptocurrency scheme based on post-quantum blockchain. *IEEE Access*, 6:27205–27213, 2018.

[139] W. A. A. Torres et al. Post-quantum one-time linkable ring signature and application to ring confidential transactions in blockchain. In *Australasian Conference on Information Security and Privacy*, pages 558–576. Springer, 2018.

[140] C. Li, X. Chen, Y. Chen, Y. Hou, and J. Li. A new lattice-based signature scheme in post-quantum blockchain network. *IEEE Access*, 7:2026–2033, 2018.

[141] E. O. Kiktenko et al. Quantum-secured blockchain. *Quantum Science and Technology*, 3(3):035004, 2018.

[142] J. Jogenfors. Quantum bitcoin: An anonymous and distributed currency secured by the no-cloning theorem of quantum mechanics. *arXiv preprint arXiv:1604.01383*, 2016.

[143] D. Sapaev, D. Bulychkov, F. Ablayev, A. Vasiliev, and M. Ziatdinov. Quantum-assisted blockchain. *arXiv preprint arXiv:1802.06763*, 2018.

[144] A. Behera and G. Paul. Quantum to classical one-way function and its applications in quantum money authentication. *Quantum Information Processing*, 17(8):200, 2018.

[145] L. Tessler and T. Byrnes. Bitcoin and quantum computing. *arXiv preprint arXiv:1711.04235*, 2017.

[146] K. Ikeda. qbitcoin: a peer-to-peer quantum cash system. In *Science and Information Conference*, pages 763–771. Springer, 2018.

[147] K. P. Kalinin and N. G. Berloff. Blockchain platform with proof-of-work based on analog hamiltonian optimisers. *arXiv preprint arXiv:1802.10091*, 2018.

[148] D. L. Hemmick and A. M. Shakur. *Bell's Theorem and Quantum Realism: Reassessment in Light of the Schrödinger paradox*. Springer Science & Business Media, 2011.

[149] A. M. Gleason. Measures on the closed subspaces of a hilbert space. *Journal of Mathematics and Mechanics*, pages 885–893, 1957.

[150] R. Cooke, M. Keane, and W. Moran. An elementary proof of gleason's theorem. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 98, pages 117–128. Cambridge University Press, 1985.

[151] G. Hellman. Gleason's theorem is not constructively provable. *Journal of Philosophical Logic*, 22(2):193–203, 1993.

[152] H. Billinge. A constructive formulation of gleason's theorem. *Journal of Philosophical Logic*, 26(6):661–670, 1997.

[153] I. Pitowsky. Infinite and finite gleason's theorems and the logic of indeterminacy. *Journal of Mathematical Physics*, 39(1):218–228, 1998.

[154] F. Richman and D. Bridges. A constructive proof of gleason's theorem. *Journal of Functional Analysis*, 162(2):287–312, 1999.

[155] F. Richman. Gleason's theorem has a constructive proof. *Journal of Philosophical Logic*, 29(4):425–431, 2000.

[156] P. Busch. Quantum states and generalized observables: a simple proof of gleason's theorem. *Physical Review Letters*, 91(12):120403, 2003.

[157] C. M. Caves, C. A. Fuchs, K. K. Manne, and J. M. Renes. Gleason-type derivations of the quantum probability rule for generalized measurements. *Foundations of Physics*, 34(2):193–209, 2004.

[158] D. Buhagiar, E. Chetcuti, and A. Dvurečenskij. On gleason's theorem without gleason. *Foundations of Physics*, 39(6):550–558, 2009.

[159] V. J. Wright and S. Weigert. A gleason-type theorem for qubits based on mixtures of projective measurements. *Journal of Physics A: Mathematical and Theoretical*, 2018.

[160] J. Hamhalter. *Quantum Measure Theory*, volume 134. Springer Science & Business Media, 2013.

[161] D. W. Cohen. *An Introduction to Hilbert Space and Quantum Logic*. Springer Science & Business Media, 2012.

[162] D. Rajan and M. Visser. Explicit construction of the density matrix in gleason's theorem. *arXiv preprint arXiv:1904.00533*, 2019.

[163] A. Alonso-Serrano and M. Visser. Coarse graining shannon and von neumann entropies. *Entropy*, 19(5):207, 2017.

[164] A. Zeilinger and R. Bertlmann. *Quantum [un] speakables II: half a century of Bell's theorem*. Springer, 2017.

[165] S. Kochen and E. P. Specker. The problem of hidden variables in quantum mechanics. In *The Logico-Algebraic Approach to Quantum Mechanics*, pages 293–328. Springer, 1975.

[166] J. S. Bell. On the problem of hidden variables in quantum mechanics. *Reviews of Modern Physics*, 38(3):447, 1966.

[167] A. Peres. Two simple proofs of the kochen-specker theorem. *Journal of Physics A: Mathematical and General*, 24(4):L175, 1991.

[168] M. Kernaghan. Bell-kochen-specker theorem for 20 vectors. *Journal of Physics A: Mathematical and General*, 27(21):L829, 1994.

[169] A. Cabello. A proof with 18 vectors of the bell-kochen-specker theorem. In *New developments on fundamental problems in quantum physics*, pages 59–62. Springer, 1997.

[170] A. Cabello, J. Estebaranz, and G. García-Alcaine. Bell-kochen-specker theorem: A proof with 18 vectors. *Physics Letters A*, 212(4):183–187, 1996.

[171] M. Kernaghan and A. Peres. Kochen-specker theorem for eight-dimensional space. *Physics Letters A*, 198(1):1–5, 1995.

[172] N. D. Mermin. What's wrong with these elements of reality? *Physics Today*, 43(6):9, 1990.

[173] N. D. Mermin. Simple unified form for the major no-hidden-variables theorems. *Physical Review Letters*, 65(27):3373, 1990.

[174] R. D. Gill and M. S. Keane. A geometric proof of the kochen-specker no-go theorem. *Journal of Physics A: Mathematical and General*, 29(12):L289, 1996.

[175] C. S. Calude, P. H. Hertling, and K. Svozil. Kochen-specker theorem: two geometric proofs. *arXiv preprint arXiv:1402.5195*, 2014.

[176] D. Rajan and M. Visser. Kochen-specker theorem revisited. *arXiv preprint arXiv:1708.01380*, 2017.

[177] M. Howard, J. Wallman, V. Veitch, and J. Emerson. Contextuality supplies the 'magic'for quantum computation. *Nature*, 510(7505):351, 2014.

[178] J. S. Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.

[179] A. Whitaker. John bell and the most profound discovery of science. *Physics world*, 11(12):29, 1998.

[180] M. F. Pusey, J. Barrett, and T. Rudolph. On the reality of the quantum state. *Nature Physics*, 8(6):475, 2012.

[181] E. S. Reich. Quantum theorem shakes foundations. *Nature*, 201(1), 2011.

[182] M. S. Leifer. Is the quantum state real? an extended review of ψ-ontology theorems. *arXiv preprint arXiv:1409.1570*, 2014.

[183] S. J. Freedman and J. F. Clauser. Experimental test of local hidden-variable theories. *Physical Review Letters*, 28(14):938, 1972.

[184] A. Aspect, P. Grangier, and G. Roger. Experimental realization of einstein-podolsky-rosen-bohm gedankenexperiment: a new violation of bell's inequalities. *Physical Review Letters*, 49(2):91, 1982.

[185] A. Shimony. Contextual hidden variables theories and bell's inequalities. *The British Journal for the Philosophy of Science*, 35(1):25–45, 1984.

[186] P. C. W. Davies and J. R. Brown. *The Ghost in the Atom: A Discussion of the Mysteries of Quantum Physics*. Cambridge University Press, 1993.

[187] L. Vervoort. Bell's theorem: Two neglected solutions. *Foundations of Physics*, 43(6):769–791, 2013.

[188] B. Toner. Monogamy of non-local quantum correlations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 465(2101):59–69, 2008.

[189] M. Ringbauer, F. Costa, M. E. Goggin, A. G. White, and A. Fedrizzi. Multi-time quantum correlations with no spatial analog. *NPJ Quantum Information*, 4(1):37, 2018.

[190] S. L. Braunstein and C. M. Caves. Information-theoretic bell inequalities. *Physical Review Letters*, 61(6):662, 1988.

[191] N. J. Cerf and C. Adami. Entropic bell inequalities. *Physical Review A*, 55(5):3371, 1997.

[192] J. Eisert, M. Wilkens, and M. Lewenstein. Quantum games and quantum strategies. *Physical Review Letters*, 83(15):3077, 1999.

[193] S. C. Benjamin and P. M. Hayden. Multiplayer quantum games. *Physical Review A*, 64(3):030301, 2001.

[194] F. S. Khan, N. Solmeyer, R. Balu, and T. S. Humble. Quantum games: a review of the history, current state, and interpretation. *Quantum Information Processing*, 17(11):309, 2018.

[195] N. Brunner and N. Linden. Connection between bell nonlocality and bayesian game theory. *Nature Communications*, 4:2057, 2013.

[196] A. Roy, A. Mukherjee, T. Guha, S. Ghosh, S. S. Bhattacharya, and M. Banik. Nonlocal correlations: Fair and unfair strategies in bayesian games. *Physical Review A*, 94(3):032120, 2016.

[197] M. Banik et al. Two-qubit pure entanglement as optimal social welfare resource in bayesian game. *Quantum*, 3:185, 2019.

[198] A. Pappa et al. Nonlocality and conflicting interest games. *Physical Review Letters*, 114(2):020401, 2015.

[199] M. L. Almeida, J. Bancal, N. Brunner, A. Acín, N. Gisin, and S. Pironio. Guess your neighbor's input: A multipartite nonlocal game with no quantum advantage. *Physical Review Letters*, 104(23):230404, 2010.

[200] N. Harrigan and R. W. Spekkens. Einstein, incompleteness, and the epistemic view of quantum states. *Foundations of Physics*, 40(2):125–157, 2010.

[201] J. S. Bell. *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*. Cambridge university press, 2004.

[202] A. Einstein. Quanten-mechanik und wirklichkeit. *Dialectica*, 2(3-4):320–324, 1948.

[203] D. Howard. Einstein on locality and separability. *Studies in History and Philosophy of Science Part A*, 16(3):171–201, 1985.

[204] A. Montina. Epistemic view of quantum states and communication complexity of quantum channels. *Physical Review Letters*, 109(11):110501, 2012.

[205] A. Montina. Communication complexity and the reality of the wave function. *Modern Physics Letters A*, 30(01):1530001, 2015.

[206] V. Havlíček and J. Barrett. Simple communication complexity separation from quantum state antidistinguishability. *arXiv preprint arXiv:1911.01927*, 2019.

[207] P. G. Lewis, D. Jennings, J. Barrett, and T. Rudolph. Distinct quantum states can be compatible with a single state of reality. *Physical Review Letters*, 109(15):150404, 2012.

[208] M. Schlosshauer and A. Fine. Implications of the pusey-barrett-rudolph quantum no-go theorem. *Physical Review Letters*, 108(26):260404, 2012.

[209] S. Aaronson, A. Bouland, L. Chua, and G. Lowther. ψ-epistemic theories: The role of symmetry. *Physical Review A*, 88(3):032111, 2013.

[210] M. K. Patra, S. Pironio, and S. Massar. No-go theorems for ψ-epistemic models based on a continuity assumption. *Physical Review Letters*, 111(9):090402, 2013.

[211] M. Schlosshauer and A. Fine. No-go theorem for the composition of quantum systems. *Physical Review Letters*, 112(7):070407, 2014.

[212] S. Mansfield. Reality of the quantum state: Towards a stronger ψ-ontology theorem. *Physical Review A*, 94(4):042124, 2016.

[213] M. S. Leifer. ψ-epistemic models are exponentially bad at explaining the distinguishability of quantum states. *Physical Review Letters*, 112(16):160404, 2014.

[214] J. Barrett, E. G. Cavalcanti, R. Lal, and O. J. E. Maroney. No ψ-epistemic model can fully explain the indistinguishability of quantum states. *Physical Review Letters*, 112(25):250403, 2014.

[215] C. Branciard. How ψ-epistemic models fail at explaining the indistinguishability of quantum states. *Physical Review Letters*, 113(2):020409, 2014.

[216] C. Perry, R. Jain, and J. Oppenheim. Communication tasks with infinite quantum-classical separation. *Physical Review Letters*, 115(3):030504, 2015.

[217] S. Bandyopadhyay, R. Jain, J. Oppenheim, and C. Perry. Conclusive exclusion of quantum states. *Physical Review A*, 89(2):022336, 2014.

[218] S. Arunachalam, A. Molina, and V. Russo. Quantum hedging in two-round prover-verifier interactions. *arXiv preprint arXiv:1310.7954*, 2013.

[219] W. C. Myrvold. ψ-ontology result without the cartesian product assumption. *Physical Review A*, 97(5):052109, 2018.

[220] D. Nigg et al. Can different quantum state vectors correspond to the same physical state? an experimental test. *New Journal of Physics*, 18(1):013007, 2015.

[221] D. J. Miller. Alternative experimental protocol to demonstrate the pusey-barrett-rudolph theorem. *Physical Review A*, 87(1):014103, 2013.

[222] M. Ringbauer, B. Duffus, C. Branciard, E. G. Cavalcanti, A. G. White, and A. Fedrizzi. Measurements on the reality of the wavefunction. *Nature Physics*, 11(3):249, 2015.

[223] K. Liao, X. Zhang, G. Guo, B. Ai, H. Yan, and S. Zhu. Experimental test of the no-go theorem for continuous ψ-epistemic models. *Scientific Reports*, 6:26519, 2016.

[224] A. J. Leggett and A. Garg. Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? *Physical Review Letters*, 54(9):857, 1985.

[225] D. Horsman, C. Heunen, M. F. Pusey, J. Barrett, and R. W. Spekkens. Can a quantum state over time resemble a quantum state at a single time? *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2205):20170395, 2017.

[226] C. Emary, N. Lambert, and F. Nori. Leggett–garg inequalities. *Reports on Progress in Physics*, 77(1):016001, 2013.

[227] A. R. U. Devi, H. S. Karthik, Sudha, and A. K. Rajagopal. Macrorealism from entropic leggett-garg inequalities. *Physical Review A*, 87(5):052103, 2013.

[228] A. Fedrizzi, M. P. Almeida, M. A. Broome, A. G. White, and M. Barbieri. Hardy's paradox and violation of a state-independent bell inequality in time. *Physical Review Letters*, 106(20):200402, 2011.

[229] M. S. Leifer and M. F. Pusey. Is a time symmetric interpretation of quantum theory possible without retrocausality? *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2202):20160607, 2017.

[230] S. Brierley, A. Kosowski, M. Markiewicz, T. Paterek, and A. Przysiezna. Nonclassicality of temporal correlations. *Physical Review Letters*, 115(12):120404, 2015.

[231] B. F. Toner and D. Bacon. Communication cost of simulating bell correlations. *Physical Review Letters*, 91(18):187904, 2003.

[232] M. Boyer. Extended GHZ n-player games with classical probability of winning tending to 0. *arXiv preprint quant-ph/0408090*, 2004.

[233] T. Fritz. Quantum correlations in the temporal clauser–horne–shimony–holt (chsh) scenario. *New Journal of Physics*, 12(8):083055, 2010.

[234] Č. Brukner. Quantum causality. *Nature Physics*, 10(4):259–263, 2014.

[235] O. Oreshkov, F. Costa, and Č. Brukner. Quantum correlations with no causal order. *Nature Communications*, 3:1092, 2012.

[236] F. Costa, M. Ringbauer, M. E. Goggin, A. G. White, and A. Fedrizzi. Unifying framework for spatial and temporal quantum correlations. *Physical Review A*, 98(1):012328, 2018.

[237] J. F. Fitzsimons, J. A. Jones, and V. Vedral. Quantum correlations which imply causation. *Scientific Reports*, 5:18281, 2015.

[238] H. Ku, S. Chen, N. Lambert, Y. Chen, and F. Nori. Hierarchy in temporal quantum correlations. *Physical Review A*, 98(2):022104, 2018.

[239] R. Pisarczyk, Z. Zhao, Y. Ouyang, V. Vedral, and J. F. Fitzsimons. Causal limit on quantum communication. *Physical Review Letters*, 123(15):150502, 2019.

[240] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT press, 2009.

[241] J. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens. Quantum common causes and quantum causal models. *Physical Review X*, 7(3):031021, 2017.

[242] J. Barrett, R. Lorenz, and O. Oreshkov. Quantum causal models. *arXiv preprint arXiv:1906.10726*, 2019.

[243] M. Nowakowski. Quantum entanglement in time. In *AIP Conference Proceedings*, volume 1841, page 020007. AIP Publishing, 2017.

[244] M. Nowakowski. Monogamy of quantum entanglement in time. *arXiv preprint arXiv:1604.03976*, 2016.

[245] I. Fuentes. Lecture series on relativistic quantum information. In *Diversities in Quantum Computation and Quantum Information*, pages 107–147. World Scientific, 2013.

[246] P. M. Alsing and I. Fuentes. Observer-dependent entanglement. *Classical and Quantum Gravity*, 29(22):224001, 2012.

[247] E. Martin-Martinez and N. C. Menicucci. Entanglement in curved spacetimes and cosmology. *Classical and Quantum Gravity*, 31(21):214001, 2014.

[248] D. E. Bruschi, T. C. Ralph, I. Fuentes, T. Jennewein, and M. Razavi. Spacetime effects on satellite-based quantum communications. *Physical Review D*, 90(4):045041, 2014.

[249] D. Harlow. Jerusalem lectures on black holes and quantum information. *Reviews of Modern Physics*, 88(1):015002, 2016.

[250] T. Nishioka. Entanglement entropy: holography and renormalization group. *Reviews of Modern Physics*, 90(3):035007, 2018.

[251] E. Witten. APS medal for exceptional achievement in research: Invited article on entanglement properties of quantum field theory. *Reviews of Modern Physics*, 90(4):045003, 2018.

[252] M. Srednicki. *Quantum Field Theory*. Cambridge University Press, 2007.

[253] V. Mukhanov and S. Winitzki. *Introduction to Quantum Effects in Gravity*. Cambridge University Press, 2007.

[254] N. D. Birrell and P. C. W. Davies. *Quantum Fields in Curved Space*. Number 7. Cambridge university press, 1984.

[255] S. M. Carroll. *Spacetime and Geometry*. Addison Wesley, 2004.

[256] A. Einstein and N. Rosen. The particle problem in the general theory of relativity. *Physical Review*, 48(1):73, 1935.

[257] D. Tong. Quantum field theory, 2007. *Lecture notes for University of Cambridge, Part III of the Mathematical Tripos*.

[258] J. F. Donoghue and G. Menezes. Arrow of causality and quantum gravity. *Physical Review Letters*, 123(17):171601, 2019.

[259] A. I. Lvovsky. Squeezed light. *Photonics: Scientific Foundations, Technology and Applications*, 1:121–163, 2015.

[260] L. C. B. Crispino, A. Higuchi, and G. E. A. Matsas. The unruh effect and its applications. *Reviews of Modern Physics*, 80(3):787, 2008.

[261] P. M. Alsing and G. J. Milburn. Teleportation with a uniformly accelerated partner. *Physical Review Letters*, 91(18):180404, 2003.

[262] I. Fuentes-Schuller and R. B. Mann. Alice falls into a black hole: entanglement in noninertial frames. *Physical Review Letters*, 95(12):120404, 2005.

[263] S. J. Olson and T. C. Ralph. Entanglement between the future and the past in the quantum vacuum. *Physical Review Letters*, 106(11):110404, 2011.

[264] H. M. Wiseman, S. J. Jones, and A. C. Doherty. Steering, entanglement, nonlocality, and the einstein-podolsky-rosen paradox. *Physical Review Letters*, 98(14):140402, 2007.

[265] Y. Chen et al. Temporal steering inequality. *Physical Review A*, 89(3):032112, 2014.

[266] T. Schreiber. Measuring information transfer. *Physical Review Letters*, 85(2):461, 2000.

[267] M. Staniek and K. Lehnertz. Symbolic transfer entropy. *Physical Review Letters*, 100(15):158101, 2008.

[268] L. Barnett and T. Bossomaier. Transfer entropy as a log-likelihood ratio. *Physical Review Letters*, 109(13):138105, 2012.

[269] A. D. Crescenzo and M. Longobardi. Entropy-based measure of uncertainty in past lifetime distributions. *Journal of Applied Probability*, 39(2):434–440, 2002.

[270] A. K. Nanda and P. Paul. Some properties of past entropy and their applications. *Metrika*, 64(1):47–61, 2006.

[271] N. A. Lynch. *Distributed Algorithms*. Elsevier, 1996.

[272] L. Chen, L. Xu, N. Shah, Z. Gao, Y. Lu, and W. Shi. On security analysis of proof-of-elapsed-time (poet). In *International Symposium on Stabilization, Safety, and Security of Distributed Systems*, pages 282–297. Springer, 2017.

[273] L. Baird. The swirlds hashgraph consensus algorithm: Fair, fast, byzantine fault tolerance. *Swirlds Tech Reports SWIRLDS-TR-2016-01, Tech. Rep.*, 2016.

[274] S. Abramsky and B. Coecke. A categorical semantics of quantum protocols. In *Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science, 2004.*, pages 415–425. IEEE, 2004.

[275] C. Heunen, M. Sadrzadeh, and E. Grefenstette. *Quantum Physics and Linguistics: A Compositional, Diagrammatic Discourse*. Oxford University Press, 2013.

[276] F. Krger and S. Merz. *Temporal Logic and State Systems*. Springer Publishing Company, 2008.

[277] A. Galton. *Temporal Logics and Their Applications*. Academic Press, 1990.

[278] D. M. Gabbay, I. Hodkinson, and M. Reynolds. *Temporal Logic Mathematical Foundations and Computational Aspects*. Clarendon Press, 1994.

[279] R. Penrose and W. Rindler. *Spinors and Space-time: Volume 1, Two-Spinor Calculus and Relativistic Fields*, volume 1. Cambridge University Press, 1984.

[280] P. J. O'Donnell. *Introduction to 2-spinors in General Relativity*. World Scientific, 2003.

[281] D. Rajan. Complex spacetimes and the newman-janis trick. *arXiv preprint arXiv:1601.03862*, 2016.

[282] J. A. Wheeler. Information, physics, quantum: The search for links. *Complexity, Entropy, and The Physics of Information*, 8, 1990.

[283] D. Rajan. Does god play dice with time itself? https://fqxi.org/community/forum/topic/3526. Accessed: 2020-06-21.

[284] P. D. Grünwald and P. M. B. Vitányi. Algorithmic information theory. *Handbook of the Philosophy of Information*, pages 281–320, 2008.

[285] R. G. Downey and D. R. Hirschfeldt. *Algorithmic Randomness and Complexity*. Springer Science & Business Media, 2010.

[286] J. S. Avery. *Information Theory and Evolution*. World Scientific, 2012.

[287] J. D. Bekenstein. Black holes and the second law. *Lett. Nuovo Cim*, 4(737):113, 1972.

[288] M. Zych, F. Costa, I. Pikovski, and Č. Brukner. Bell's theorem for temporal order. *Nature communications*, 10(1):1–10, 2019.

[289] W. Isaacson. *Einstein: His Life and Universe*. Simon & Schuster Audio, New York, USA, 2011.

[290] A. D. Stone. *Einstein and the Quantum: The Quest of the Valiant Swabian*. Princeton University Press, 2015.