HOW TO TEACH AN OLD DOG NEW TRICKS: QUANTUM INFORMATION, QUANTUM COMPUTATION, AND THE PHILOSOPHY OF PHYSICS

by

Armond Duwell

B.S. Physics, Georgia Institute of Technology, 1998

Submitted to the Graduate Faculty of
the Department of History and Philosophy of Science in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2004

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF HISTORY AND PHILOSOPHY OF SCIENCE

This dissertation was presented

by

Armond Duwell

It was defended on

October 11th 2004

and approved by

John Norton, Department of History and Philosophy of Science, University of Pittsburgh

Jeff Bub, Department of Philosophy, University of Maryland, College Park

John Earman, Department of History and Philosophy of Science, University of Pittsburgh

Laura Reutsche, Department of Philosophy, University of Pittsburgh

Dissertation Director: John Norton, Department of History and Philosophy of Science, University of Pittsburgh

HOW TO TEACH AN OLD DOG NEW TRICKS: QUANTUM INFORMATION, QUANTUM COMPUTATION, AND THE PHILOSOPHY OF PHYSICS

Armond Duwell, PhD

University of Pittsburgh, 2004

My dissertation consists of two independent parts. Part one of my dissertation examines concepts of quantum information. I clarify three very different concepts of information and assess their implications for understanding quantum mechanics. First I clarify the concept of information due to Shannon, and its relation to physical theories. Using the Shannon concept, I examine two purportedly new concepts of quantum information. I argue that a fundamental philosophical mistake is made regarding these concepts. Advocates of these new concepts do not properly distinguish between the properties of information due to the physical medium it is stored in from the properties of information per se. This distinction is crucial for developing a new concept to help us understand quantum mechanics and evaluating its merits.

Part two of my dissertation examines explanations of the efficiency that quantum computers enjoy over classical computers for some computational tasks, and the relationship between explanations of efficiency and interpretations of quantum mechanics. I examine the so-called quantum parallelism thesis, that quantum computers can perform many computations in a single step, a feat thought not to be possible on classical computers. The truth of this thesis is not obvious, and contested by some. I develop a set of general criteria for computation that any computing device must satisfy. I use these criteria to demonstrate that the quantum parallelism thesis is true. As an application of these general criteria for computation I articulate three distinct concepts of parallelism and demonstrate that classical computers can compute in parallel as well. This demonstrates that the truth of the quantum parallelism thesis alone does not provide a complete explanation of the efficiency of quantum computers. I supplement the quantum parallelism thesis to provide a complete explanation. Finally, I address the claim that only the many-worlds interpretation of quantum mechanics can underwrite the truth of the quantum parallelism thesis. The general criteria for computation provide support for the quantum parallelism thesis independent of any interpretation of quantum mechanics.

TABLE OF CONTENTS

LIST OF FIGURES

1.0 INTRODUCTION

Ever since modern quantum theory was born, physicists and philosophers have struggled to understand it. Substantial progress has been made over the years, perhaps not in the sense of producing a fully satisfactory interpretation, but by a clear understanding of the problems and restrictions on possible interpretations of the theory. In particular, I have in mind results such as Bell's theorem and the Kocher-Specker theorem. The two main problems of quantum theory, the measurement problem and failure to understand how EPR correlations are possible, still loom large.

For the physicist and philosopher, the options regarding quantum theory seem limited. One may hold out hope that a new theory will replace quantum theory and remove its problems by explaining or dissolving them. For those not content to wait for this promised and elusive theory, one can pick their favorite interpretation which is, no doubt, riddled with philosophical difficulties, or try to come up with a new interpretation. This latter option, constructing a new interpretation, may recently have become a more attractive option. The current situation in physics, with the development of quantum information theory (QIT) and quantum computation theory (QCT), may enable a new line of research in the foundations of quantum theory. Rather than honing interpretations[1] using Schrödinger's cat and EPR-B experiments, QIT and QCT have recently provided a host of new phenomena, teleportation, dense coding, quantum parallelism, etc., and hence conceptual difficulties, for the physicist and philosopher to ponder. Physicists have expressed great hope for these new sub-disciplines to help understand the foundations of quantum theory.

[1] By "interpretations of quantum theory", I mean, perhaps misleadingly, but following standard terminology, to include other quantum theories, like Bohm's theory and the GRW theory, in addition to interpretations of standard quantum theory.

> One of the most fascinating aspects of recent work in fundamental quantum theory is the emergence of a new notion, the concept of quantum information, which is quite distinct from its classical counterpart. It provides a new perspective for all foundational and interpretational issues and highlights new essential differences between classical and quantum theory (Jozsa 1998, 49).

> The computational differences between quantum and classical physics are if anything more striking and can be expected to offer new insights into the nature of quantum physics(Bernstein and Vazirani 1997, 1415).

Several other physicists, the Horodekis, Cerf and Adami, Deutsch and Hayden, and others, have expressed sentiments like the ones above. It remains to be seen if QIT or QCT can deliver the insights that philosophers and physicists desire.

Some philosophers such as, Bub, Clifton, Halvorson, MacCallum, Pitowski, Timpson, and myself have begun to explore the resources that QIT and QCT have to offer, but a great deal of work has yet to be done. This dissertation is devoted to foundational issues surrounding QIT and QCT.

QIT and QCT have already become too large to provide a comprehensive analysis, and I make no attempt at doing so. Part I of this dissertation deals with QIT. One of the fundamental foundational questions in QIT is what is quantum information? In this part of the dissertation, I examine two different concepts of quantum information, one due to Jozsa (1998), another due to Deutsch and Hayden (1999). I assess the warrant for the introduction of these new concepts and also assess any value they have for better understanding quantum mechanics. I also examine several new phenomena in order to critically evaluate concepts of quantum information.

Part II of this dissertation deals with QCT. David Deutsch has suggested that the many-worlds interpretation of quantum theory has special priority in explaining the efficiency of quantum computers. In particular, the many worlds interpretation allows one to claim that a quantum computer can compute many values of a function in a single computational step, presumably a feat that no conventional computer could perform. Andrew Steane has argued that a quantum computer does not compute many values of a function in a single computational step. As such, he severs the connection between the many worlds interpretation and explanations of the efficiency of quantum computers. In this part of the dissertation, I examine a crucial process that is part of efficient quantum mechanical algorithms to decide, first, whether quantum computers can compute many values of a function in a single step, and secondly, whether the many worlds interpretation has any special explanatory advantage over other interpretations of quantum mechanics.

2.0 INTRODUCTION PART I

The foundations of QIT are far from settled. The central theoretical entity, quantum information, remains poorly defined. Possible candidate definitions are as banal as a simple renaming of Shannon information stored in quantum systems, to suggestions along the lines that it is the primary stuff of the world. Until the concept of quantum information is clarified, it is difficult, perhaps impossible, to see what possible insight can be gained from QIT about the foundations of quantum theory.[1]

This part of the dissertation is devoted to analyzing different concepts of quantum information due to Jozsa (1998) and Deutsch and Hayden (1999). My method for analyzing these different concepts is to compare and contrast the properties claimed for quantum information with properties of Shannon information. To do so, I devote a chapter to the Shannon information theory. I pay particular attention to the relationship between the Shannon theory and classical physics, and also the locality properties of information. Understanding these features is crucial to understanding the need for a new concept of information for quantum systems.

In chapter four I critically access Jozsa's suggestion that a new concept of information is required for quantum systems. I will argue that all properties that Jozsa attributes to quantum information are simply just properties of the Shannon information when stored in quantum systems. So, Jozsa provides no convincing reasons to introduce a new concept of information, nor does he introduce a new concept.

In chapter five I evaluate Deutsch and Hayden's concept of quantum information. Deutsch and Hayden develop a formalism for quantum theory specifically designed to track the flow

[1] Of course, this does not limit the use of new phenomena as a conceptual tools to hone interpretations. Here I'm thinking of using the concept of quantum information as a potential explanatory tool.

of information in quantum systems. They claim that special properties of quantum information account for the appearance of non-locality and holism in quantum systems, all while demonstrating that quantum theory is explicitly local. I argue that quantum information cannot have the properties that Deutsch and Hayden attribute to it.

3.0 THE SHANNON THEORY

3.1 INTRODUCTION

In this chapter I seek to clarify the concept of information due to Shannon (1948). In particular, I seek answers to the following questions:

> How is the Shannon information content of a system quantified?
> How is Shannon information transferred?
> What is the physical status of Shannon information?
> What role, if any, can a Shannon information theoretic analysis of a physical phenomenon play in an explanation of a physical phenomenon?

Answers to these questions help us better understand what kind of concept Shannon information is, and what the concept can do. As we will see in later chapters, the Shannon concept of information is a useful reference when exploring new concepts of quantum information.

In section two I describe the Shannon theory. I discuss Shannon's measure of information, and its interpretation via the Shannon coding theorem. In section three I discuss features of the Shannon information especially regarding locality, its physical status, and its potential use in explanations of physical phenomena.

3.2 SHANNON INFORMATION THEORY

In 1948 Shannon published his landmark paper *The Mathematical Theory of Communication.* Shannon suggested that the fundamental problem of communication is reproducing a message selected at one spatial location at spatial location. He emphasizes *the meaning or* *interpretation* of a message is irrelevant to his theory. Rather, the "significant aspect is that the actual message was selected from a set of possible messages"(Shannon and Weaver 1948, 3). Here is an example. Suppose Alice chooses to send one of the following two messages to Bob with equal probability: (0) Go to the grocery store and buy one lb shrimp, two lbs plum tomatoes, olive oil, garlic, 6 oz Feta cheese, and 6 oz Parmesan cheese. or (1) Do nothing. Intuitively, one might be compelled to say that message (0) contains more information than message (1). This type of consideration is irrelevant to the Shannon theory. The Shannon theory is meant to quantify the resources required to communicate a message from a *prearranged choice of possibilities*, not to quantify resources needed to describe the meaning of a message. Assuming the above numbering system is prearranged, all that is required to communicate one of the above messages is to specify whether message (0) or (1) was chosen.

The Shannon theory is intended to apply to a *communication system* that consists of the following (Shannon and Weaver 1949, 5-6):

1. The *information source*, which produces a message or sequence of messages to be communicated.
2. The *transmitter*, which operates on the message in some way to produce a signal suitable for transmission over the channel.
3. The *channel*, which is merely the medium used to transmit the signal from transmitter to receiver.
4. The *receiver*, which ordinarily performs the inverse operation of that done by the transmitter, reconstructing the message from the signal.
5. The *destination*, which is the person (or thing) for whom the message is intended.

Shannon's theory determines the minimum resources required to remove uncertainty at a destination about the choice of a message at the information source. The uncertainty at a destination about the choice of a message or a sequence of messages from an information source is characterized by probability distributions on sets of states of the components of a communication system.[1] Intuitively, as the uncertainty of a message increases, so too will the resources required to remove the uncertainty. Shannon made this vague idea precise

[1] For simplicity discussion is restricted to probability distributions on sets of finite cardinality.

by creating a quantitative measure of uncertainty, and demonstrating that this measure quantifies the resources required to remove uncertainty, i.e. to transmit information.

3.2.1 Shannon's measure

Shannon's measure of uncertainty, H, is a function of a probability distribution on a set. A probability distribution on a set is denoted by $X := \{x, p_x\}$, where x is a variable that labels possible messages and p_x is the message's probability. We suppose that the cardinality of the ensemble is n. Shannon required a measure of information to satisfy three properties (Shannon and Weaver 1949, 19):

1. H should be continuous in the p_x.
2. If all the p_x are equal, $p_x = \frac{1}{n}$, then H should be a monotonic increasing function of n.
3. If a choice be broken down into two successive choices, the original H should be a weighted sum of the individual values of H.[2]

Property one is simply a natural requirement for a well behaved mathematical function. Property two is a seemingly necessary condition for any measure of uncertainty associated with ensembles to satisfy. If the probability distribution is flat, we require that our uncertainty associated with that distribution increase as the number of different messages increases. The motivation for property three is the most difficult to understand. When a message is chosen from an ensemble, there are different ways one could make a choice. One can choose a message outright, or alternatively, one could break the choice into a series of choices between disjoint subsets of messages in such a way as to leave the original probability distribution for the choice of a message invariant.

For example, consider the probability distribution $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{6}$. (See Figure 1.) The choice of an outcome (message) from this distribution can be viewed mathematically as choosing either 1 with probability $\frac{1}{2}$ or outcomes 2 *and* 3 with probability $\frac{1}{2}$, and then choosing again between 2 and 3 with the probabilities p_2 and p_3 renormalized, i.e. $p_2 = \frac{2}{3}$ and $p_3 = \frac{1}{3}$. Shannon required that his measure of uncertainty reflect this freedom by requiring that $H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2}H(\frac{2}{3}, \frac{1}{3})$. So, the uncertainty associated with sequences of

[2] This third postulate was shown to be equivalent to the following rule: For every $n \geq 2, H(p_1, p_2, \ldots, p_{n-1}, q_1, q_2) = H(p_1, p_2, \ldots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right)$, where $p_n = q_1 + q_2$ (Faddeev, 1957).

choices is the uncertainty associated with the original choice, plus the weighted uncertainty of the second choice. Surprisingly, the only function satisfying conditions 1-3 is of the form (Shannon and Weaver 1948, 19):

$$H = -K \sum_{x=1}^{n} p_x \log p_x. \tag{3.1}$$

It is convenient to choose K so that $H = -\sum_{x=1}^{n} p_x \log p_x$ (where logarithms have base 2).

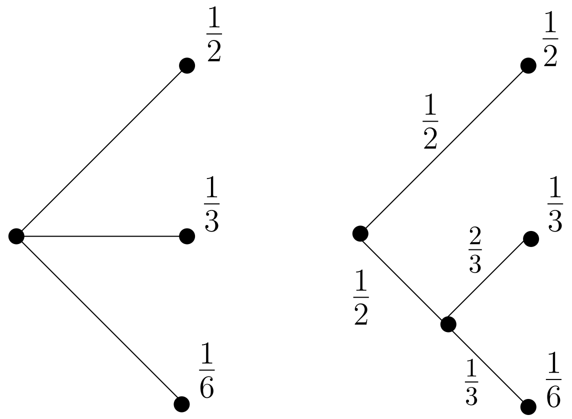

Figure 1: Suppose that we must make a choose a message 1, 2, and 3 with probability distribution $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, and $p_3 = \frac{1}{6}$. We can make one choice from the three, represented by the left hand side of the figure. Alternatively, we can make two choices. First, one can choose between message 1 or messages 2 and 3, and then, if 2 and 3 were chosen, choose again with renormalized probabilities for events 2 and 3. The probabilities for the messages remain invariant under these two different methods for choosing a message.

In addition to knowing that the Shannon measure satisfies properties 1-3, to understand the Shannon measure as a good measure of uncertainty, it is instructive to consider how changes in a probability distribution change the measure. The Shannon measure basically tells us how flat or peaked a probability distribution is, with the Shannon measure at a maximum for a flat distribution and a minimum when the probability is concentrated at a point (an event with probability one). If we think about the probability distribution the Shannon measure is applied to as describing a set of possible outcomes, the relationship of the Shannon measure to uncertainty is clear. When a probability distribution is flat, we are maximally uncertain about which outcome will occur; there are no outcomes to favor. When a probability distribution is sharply peaked, we can be fairly certain as to which result will occur. We note that the Shannon measure takes the form of an expectation value of the function $-\log p_x$ (Timpson 2003, 5). $\log p_x$ is called the "surprisal" associated with outcome x (Dretske 1981, 52). The term "surprisal" is appropriate because the surprisal function increases monotonically with decreased probabilities. This matches our intuition that high probability events are not surprising, but low probability events are. So, the Shannon measure quantifies the expected or average uncertainty (the expected "surprisal" associated with a probability distribution).

The Shannon measure is not a unique measure of uncertainty. Uffink, in his 1990, shows that the Shannon measure is one of many possible measures (Timpson 2003, 4). The Shannon measure gains its preferred status as a measure of information content due to its appearance in the Shannon coding theorem which quantifies the resources minimally necessary and sufficient to transmit information.

3.2.2 The Shannon coding theorem

Shannon's idea was to exploit the statistical properties of an information source to minimize the resources required to represent sequences of messages. For typical sequences of N messages, for all individual messages x, we can expect there to be roughly Np_x occurrences of message x. As N increases, it becomes unlikely that a sequence will not be typical in this sense. Shannon was able to prove that the number of different typical sequences of messages is approximately $2^{NH(X)}$, where $H(X) = -\sum_x p_x \log p_x$. To communicate N messages, one need only indicate which of the $2^{NH(X)}$ typical sequences were chosen. Hence, the information content of a sequence of messages can be quantified by counting the binary numbers needed to identify the message.

It is useful to take a closer look at the proof of Shannon's coding theorem. It will serve as a reference in the next chapter where comparisons are made between the Shannon coding theorem and the Schumacher coding theorem. The bulk of the proof of Shannon's coding theorem is in the proof the theorem of typical sequences which bounds the total number of typical sequences and their probability of occurrence. My presentation of the theorem follows Nielsen and Chuang (2000) closely.

We define an ϵ-typical sequence of messages of length N from an ensemble X, $x_1, \ldots, x_N$ to be a sequence whose probability $p_{x_1}p_{x_2}\cdots p_{x_N}$ satisfies the inequality

$$\left|\frac{1}{N}\log\frac{1}{p_{x_1}p_{x_2}\cdots p_{x_N}} - H(X)\right| \leq \epsilon.$$

Let $T(N,\epsilon)$ be the set of all ϵ-typical sequences of length N from X. It is possible to prove the following (Nielsen and Chuang (2000), 539):

1. Fix $\epsilon > 0$. Then for any $\delta > 0$, for sufficiently large N, the probability that a sequence is ϵ-typical is at least $1-\delta$.
2. For any fixed $\epsilon > 0$ and $\delta > 0$, and for sufficiently large N, the number of ϵ-typical sequences, $|T(N,\epsilon)|$, satisfies

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

The idea is to code each ϵ-typical sequence with a number. Since the number of ϵ-typical sequences is bounded above by $2^{N(H(X)+\epsilon)}$, $N(H(X)+\epsilon)$ binary numbers suffice to distinguish each ϵ-typical sequence. Any fewer and not all of the typical sequences can be distinguished. We can choose a codeword that indicates error when an atypical sequence is produced. From item one above, we know that the probability of an atypical sequence occurring is less than δ, which we can make as small as we like by making N sufficiently large. Since the number of typical sequences is bounded from above by $2^{N(H(X)+\epsilon)}$, reliable information transfer of a sequence of N messages can be done using approximately $NH(X)$ binary numbers in the limit of large N. So, the Shannon measure also indicates the minimal resources necessary and sufficient for reliable information transfer.

From here it is a short step to proving the Shannon coding theorem. Following Shannon and Weaver (1949) the Shannon coding theorem states:

> If the information source has an uncertainty of H (bits/message) and a channel is available with capacity C (bits/second), it is possible to code the output of the information source such that $C/H - \varepsilon$ (messages/second) are communicated, where ε is arbitrarily small and $\varepsilon > 0$. There is no coding scheme such that reliable communication can take place at a rate exceeding C/H.

The coding theorem demonstrates that the Shannon measure of an information source determines the best possible rate of communication for a given channel.

A communication system often falls short of the above maximum rate of information transfer. Uncertainty at the destination about the message increases due to noise in the communication system between the information source and the receiver. This has a deleterious effect on the rate of the information transfer. Not surprisingly, the maximum rate of information transfer will be a function of the individual and joint probability distributions for the information source and the destination. Similar to the ideal case, the maximum rate at which information can be transferred will be a function of the Shannon measure of information.

The Shannon measure of uncertainty can be used to define the uncertainty associated with a message from the information source given an outcome at the destination. Let $X := \{x, p_x\}$ denote an ensemble that represents an information source, $Y := \{y, p_y\}$ denote an ensemble that represents a destination, and let $\{p_{x,y}\}$ be joint probability distribution for the information source and destination. Given the individual and joint probability distributions, the conditional probability of message x given that y occurred at the destination is:

$$p_{x|y} = \frac{p_{x,y}}{\sum_x p_{x,y}}. \tag{3.2}$$

We define the conditional uncertainty of the information source given the destination as

$$H(X|Y) \equiv -\sum_{x,y} p_{x,y} \log p_{x|y}. \tag{3.3}$$

Since $H(X)$ measures the uncertainty of a message from the information source, and $H(X|Y)$ measures the uncertainty that remains at the destination about the information source's message, the total reduction in uncertainty will be the difference between these two quantities, and hence a measure of the quantity of information transferred. We define the information transferred in this case to be the mutual uncertainty or information, $H(X : Y)$, where $H(X : Y) \equiv H(X) - H(X|Y)$. In the case of perfect correlations between the information source and the destination, the mutual information reduces to the uncertainty of the source, $H(X : Y) = H(X)$. Shannon demonstrated that the maximum rate of information transfer is a function of $H(X : Y)$ (Shannon and Weaver 1949, 39-40).

Given the above overview of the Shannon theory and its important results, we can now proceed to discuss the more philosophical aspects of information.

3.3 FEATURES OF SHANNON INFORMATION

In this section we seek to clarify the status of information. In particular, we ask how is it that information is transferred? In doing so, we investigate the locality properties of information. We also inquire as to the physical status of information. Is information a physical primitive? Is it a property of a physical system? We will see that understanding how Shannon information is transferred underwrites answers to all of these questions.

3.3.1 The Shannon information and locality

One often speaks loosely about the location of information in a communication system as if information were a real physical quantity somehow located *in* the signal from the transmitter to the receiver or, similarly, that the signal itself is the information. These are not correct characterizations of information. If information is neither the signal itself nor located in the signal, then how does it get transferred? Information transfer generically refers to a situation where an attempt is made to remove uncertainty about the choice of a message by the information source at the destination by using a communication system.

The appropriate question to ask is not, “How does a signal store or contain information?”, but rather, “How can a signal be used to remove uncertainty about the information source’s message at the destination?” The answer to the latter question is straightforward. The physical state of the signal is correlated with the choice of a unique message at the source and at the destination. The relationship between the signal states and possible messages is arranged prior to communication. Correlations are the key to information transfer. The physical state of the signal is not information, and neither are the correlations. It is the correlations between the information source, the signal, and the destination that allow uncertainty to be removed at the destination about the information source’s message. So, rather than talking about the location of *information* in a communication system, we should really talk about correlations between physical states of parts of a communication system. In doing so, we can talk *as if* "information" were localized in a communication system. Tracking parts of a communication system whose states are progressively correlated with the information source is the means to talk precisely about the location and transfer of information.

The Shannon theory places no restrictions on *how* correlations can be established in a communication system. Thus, there is no requirement that there must be a local carrier of information that travels a continuous path from the information source to the destination, even though information transfer may typically proceed this way, e.g. a letter sent by mail. The significant aspect is that progressive correlations *can be* established in a communication system. The physical theory that applies to the components of a communication system will detail how correlations can be progressively established.

It should be clear that the Shannon theory places no restrictions on the type of physical systems that can form a communication system, so long as those systems are such that progressive correlations can be established throughout the system. Again, the Shannon theory is dependent on the physical theory that applies to the components of the communication system to determine the ensembles that characterize these components.

3.3.2 The physical status of Shannon information

It is conceivable that someone might suggest that the Shannon information of a physical system is a physical quantity. One could simply argue that if an ensemble objectively characterizes a physical system, in the sense that physical properties of a system determine the ensemble, and not properties of ourselves, i.e. our ignorance of the properties of a system, then the Shannon measure measures a genuine feature of the system, and hence defines a physical quantity.

To conclude on the basis of the above remarks that the Shannon information is a physical quantity would be a mistake. If an ensemble objectively characterizes a physical system, it is reasonable to suggest that it is a genuine physical feature of the system, with a physical status similar to something like energy perhaps. Clearly the Shannon measure of information can be applied to such a probability distribution, but that does not mean that the system contains information. As we have learned in the last section, for a system to contain information about an information source, it must be correlated with the information source. Containing Shannon information is thus a *relational property* of a system and an information source. It is only in virtue of the fact that a system can remove uncertainty about an information source that it can be said to contain information. When a probability distribution objectively characterizes an individual physical system, the question of whether the system contains Shannon information cannot arise without reference to an information source, except in error.

The Shannon measure indeed can be said to quantify an important physical feature of a system that is objectively characterized by an ensemble. That feature is not information. The concept of Shannon information arises only when we want to communicate events described by such an ensemble. Communication of course, presupposes both a source of information and a destination, and hence implicates relational properties of physical systems. These relational properties can, of course, be underwritten by individual non-relational properties. The point is that such properties considered in isolation have nothing to do with Shannon information.

3.3.3 Shannon information and transfer explanations

In physics, there is a natural type of explanation of a broad class of phenomena that one often encounters which explains phenomena in terms of transfer of a quantity, sometimes a conserved quantity. Many explanations refer to the transfer of heat, energy, or momentum from one physical system to another to explain the behavior of the systems involved. For example, we can think of a hot rock thrown into a pool of cool water. We can explain what happens, the temperature of the rock decreases and the temperature of the water increases until equilibrium is reached, in virtue of heat being transferred between the two bodies. We might explain the effects of a collision between two bodies in terms of a transfer of energy or momentum from one system to another. The general idea is clear. One explains the physical effect in virtue of the transfer of a physical quantity from one system to another.

Why is this explanatory? The properties of the transferred quantity are typically sufficient to determine the resulting behavior of the system, i.e. the physical effect or phenomena that was in need of explanation. Even if the properties of the transferred quantity do not completely determine the resulting behavior of the system, minimally the effect or phenomenon will co-vary with the quantity transferred. This qualifies the quantity as a partial cause of the phenomenon. Also the transfer of a quantity from one system to another is thought of by some, Dowe (2000), as the hallmark of causal interaction, which is almost by definition, explanatory. Note that transfer explanations are sometimes seen as more useful than reductive explanations of effects, e.g. appealing to the microphysics of the situation in the case of heat transfer, or a detailed account of the forces that bodies experience and their resulting trajectories in the case of collisions.

It is not difficult to imagine a situation where a physical phenomenon can be characterized as transferring Shannon information. For a phenomenon to be interpreted as transferring Shannon information some component of the physical system must be capable of being interpreted as an information source and also that correlations between the source and other systems are generated in the course of the phenomenon constituting information transfer. Now, we know that the physics of the underlying situation determines all of the information theoretic properties of the system (which are fixed once the individual and joint probabilities for the ensembles characterizing the system are fixed). An explanation in terms of the underlying physics might be intractable, or just uninformative. This seems very similar to situations in which “transfer” explanations are successful. A natural question to ask is whether an explanation that appeals solely to transfer of information can be implicated to explain the phenomenon just as in the cases above.

Unfortunately, information cannot be implicated in a transfer type explanation in the way heat or energy, or other quantities can. Most importantly, the Shannon information is not a physical quantity. Information is a relation between at least two physical systems. In order for information to be transferred, there need be no transfer of any quantity. To transfer information, one only needs to instantiate correlations between physical systems. Surely, one way to do this is by exchanging some quantity, perhaps a conserved quantity, but this is not required by Shannon's theory. So, the information theoretic properties alone do not ensure that a physical quantity is transferred when information is transferred. Thus, the seemingly necessary feature that makes transfer explanations possible, that some quantity is transferred, cannot be assumed given the information-theoretic properties alone.

Even if one mistakenly thought that the Shannon information was a quantity, it does not have the right properties to explain a physical effect. In particular, the Shannon information does not determine any properties of a system. When Shannon information is transferred to a physical system, we know *that* the state of the system is correlated to the information source. This tells us nothing about the process through which the correlations were instantiated. So, indicating Shannon information was transferred does not explain why the state of the system is what it is. Moreover, it is the state of the system in conjunction with the physical laws of that system that determines the behavior of that system, i.e. the effect or phenomenon to be explained. In transfer explanations, the quantity that was transferred is what determines the properties of the system that the quantity was transferred to. This is not the case with information transfer. The underlying physics of the system does the work here, not information.

For example, imagine that two communication systems are set up that are identical in every respect besides the following. Communication system one has an information source described by the probability distribution $X := \{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$, and communication system two has an information source described by $X' := \{\frac{1}{3} + \epsilon, \frac{1}{3} - \epsilon, \frac{1}{3}\}$ where ϵ is close to 0. Imagine that identical sequences are produced from the information sources in the two communication systems. The Shannon information content of the two sequences will be different, but physically, everything will be exactly the same. This undermines the use of quantity transfer explanations of a phenomenon using information. Information content is not a partial cause of the phenomenon.

Hence, transfer type explanations involving the information-theoretic properties alone are not appropriate for physical phenomena interpreted as communication systems. Explanations of physical effects based on the information theoretic properties alone are those that can be explained based solely on the individual and joint probability distributions of the components of physical system that is interpreted as a communication system.

3.4 CONCLUSION

At the beginning of this paper we sought to clarify the Shannon concept of information. In particular we wanted answers to the following questions:

> How is the information content of a system quantified?
> How is information transferred?
> What is the physical status of information?
> What role, if any, can an information theoretic analysis of a physical phenomenon play in an explanation of a physical phenomenon?

We have learned that the Shannon information content of a system is quantified using the Shannon measure of information. It is a good measure because it quantifies uncertainty and, following the coding theorem, quantifies the resources required to communicate a sequence of events described by an ensemble. Furthermore, the Shannon mutual information quantifies our uncertainty about a sequence of events given some probabilistic correlation to the sequence.

We have learned that information is transferred in a communication system in virtue of systems progressively becoming correlated with a choice at an information source. It is then clear that the Shannon theory is independent of any particular physical theory. Any set of systems, regardless of which physical theory applies to them, can be used in a communication system, so long as they enable the establishment of correlations in the components of a communication system.

We have learned that Shannon information is not a physical quantity or absolute property of a system. For a system to contain information, it must bear a certain kind of relationship to an information source. So, to say that a system contains information is a statement of the relational properties of that system to other systems. This fact indicates that we must not confuse information with the property of a physical system identified by the Shannon measure, which may have status similar to the internal energy of a system.

Finally, we have learned that the Shannon information alone is not sufficient to provide a transfer explanation of a physical phenomenon. Only when additional physical details are provided might a transfer explanation be possible. With this understanding of the Shannon information, we will proceed to use it to explicate notions of quantum information which may or may not be distinct from the Shannon information, and assess the warrant for their introduction.

4.0 QUANTUM INFORMATION DOES NOT EXIST

This chapter is adapted from Duwell, A. (2003) “Quantum information does not exist”. *Stud. Hist. Phil. Mod. Phys.* (September 2003) pp. 479-499

4.1 INTRODUCTION

The field of quantum information theory has exploded in the last ten years. Given all the interest, it seems reasonable to ask what quantum information is. There are two types of answers. The first is that “quantum information” simply refers to the behavior of classical or Shannon information when it is stored in quantum systems. This sentiment is expressed by Caves and Fuchs (1996), “Quantum information refers to the distinctive information-processing properties of quantum systems, which arise when information is stored in or retrieved from nonorthogonal quantum states”(Caves and Fuchs 1996, 1). In this case, “information” takes its meaning from the Shannon theory.[1]

The second way to answer the question is to suggest that quantum information is something radically different than Shannon information. Consider a quote from “Quantum Information and its Properties” (Jozsa 1998, 49):

> One of the most fascinating aspects of recent work in fundamental quantum theory is the emergence of a new notion, the concept of quantum information, which is quite distinct from its classical counterpart. It provides a new perspective for all foundational and interpretational issues and highlights new essential differences between classical and quantum theory.

[1] See Fuchs (2002), 132 or Fuchs (2001), 35 that the Shannon concept of information is appropriate for quantum information theory.

Clearly, Jozsa does not think that quantum and Shannon information concepts are the same. In this paper I will examine Jozsa's argument that a new concept of information is needed. I focus on Jozsa because, to my knowledge, his paper is the most complete statement of that position on quantum information.

Jozsa's argumentative strategy is to demonstrate that quantum information has properties that Shannon information does not have and also that the Shannon theory cannot account for peculiar information-theoretic properties of recently discovered quantum phenomena. Unfortunately, Jozsa never explicitly and completely formulates the properties of Shannon information nor the properties of quantum information he wishes to use to contrast the two concepts. Further, no serviceable definition of quantum information is offered.

In this paper, I will suggest that Jozsa's analysis fails to recognize three important features of the Shannon theory of information: (I) The Shannon theory does not utilize or depend on classical physics. (II) Transfer of Shannon information does not require a local carrier of information. (III) The Shannon information content of a physical system can only be quantified in the context of a communication system.

I will suggest that Jozsa's failure to recognize the features (I), (II), and (III) of Shannon information mislead him into suggesting that quantum information has properties that cannot be accounted for using the Shannon theory. I will argue that these properties Jozsa ascribes to quantum information are either not well motivated, or can be traced to properties of Shannon information when stored in quantum systems. I will also suggest failure to recognize features (I), (II), and (III) mislead Jozsa into suggesting that dense coding, teleportation, and Schumacher coding, are phenomena that cannot come under the purview of the Shannon theory. I will argue that these phenomena present no challenge to the Shannon theory.

In what follows, I will demonstrate that the Shannon theory has features (I), (II), and (III). Next, I will examine the special properties of quantum information that Jozsa thinks cannot be captured by the Shannon theory. I will argue that these properties depend on the type of physical system used to store information, not on new properties of information. I will examine the phenomena that Jozsa cites as manifestations of the special properties of quantum information, dense coding, teleportation, and Schumacher coding, and demonstrate that the Shannon theory can be applied to each phenomenon. I will argue that insofar as the properties Jozsa ascribes to quantum information can be made precise and well motivated, they coincide with properties of quantum states. Hence, no *new* concept of information is required.

4.2 THE SHANNON INFORMATION THEORY

(I) The Shannon theory is a theory about the statistical properties of a communication system. Once the statistical properties of a communication system are specified, all information theoretic properties of the communication system are fixed. The details of how the statistical properties are fixed are a matter of the physical theory that applies to the components of a communication system, not the Shannon theory proper. Hence, the Shannon theory can be applied to any communication system regardless if its parts are best described by classical mechanics, classical electrodynamics, quantum theory, or any other physical theory.

(II) Information is transferred in virtue of progressive correlations being established between the components of a communication system. The Shannon theory places no restrictions on *how* such correlations are established. Therefore, information transfer can take place without a localized carrier of information that travels a continuous spatial path from the information source to the destination. It is well know that quantum systems can be used to establish correlations between physical systems that cannot be explained by a local causal process (EPR experiments). Hence, it is possible that there is no carrier of information that describes a continuous path from the information source to the destination in a communication system that utilizes quantum systems to transfer information.

(III) In the Shannon theory, the information content of a system is not an absolute property of the system. If one examines a string of binary numbers, there is no telling what the information content of that string is until the statistical properties of the communication system that that string is a part of are specified.

4.3 QUANTUM STATES AND INFORMATION

Given the above features of Shannon information, we are now ready to analyze so called "quantum information" and its features or properties. In this section I articulate three properties that Jozsa attributes to quantum information. First, quantum information is largely inaccessible. Second, it cannot be copied. Third, qubits, two dimensional quantum systems, have a greater representational capacity than an equal number of cbits, two state systems.[2] I will argue that these properties can be construed as properties of Shannon information when stored in quantum systems.

4.3.1 Inaccessability

We turn to Jozsa to get an idea of what he means by "quantum information." In his article "Quantum Information and its Properties", Jozsa suggests that quantum information is embodied in a quantum state (Jozsa 1998, 49). Here is what Jozsa has in mind (Jozsa 1998, 50):

> The general state of a qubit may be labeled by two real parameters θ and ϕ: $|\psi\rangle = \cos\theta|0\rangle + e^{i\phi}\sin\theta|1\rangle$. Thus, we can apparently encode an arbitrarily large amount of classical information into the state of just one qubit (by coding the information into the sequence of digits of θ and ϕ). However, in contrast to classical systems, quantum measurement theory places severe restrictions on the amount of information we can obtain about the identity of a given quantum state by performing any conceivable measurement on it. Thus, most of the quantum information is "inaccessible" but it is still useful – for example it is necessary to predict any future evolution of the state and to carry out the processes of quantum computation.

The first claim we need to understand is that a quantum state can encode an arbitrarily large amount of information. We can imagine a finite ensemble, S, whose members are n equiprobable ordered pairs from the set $\{(\theta_1, \phi_1), \ldots, (\theta_n, \phi_n)\}$. In this case $H(S) = \log n$. We know that as n increases, $H(S)$ increases without bound. We can suppose that the choice of possible θ and ϕ pairs is correlated with the state of a qubit. Information is encoded in a quantum state in virtue of the correlation between the classical choice of a message from

[2] I use the terminology cbits to distinguish two state (physical) systems from bits, binary digits, which are units of information, or simply numbers.

S, say (θ_1, ϕ_1),and the state of a quantum system, ψ, where $|\psi\rangle = \cos\theta_1|0\rangle + e^{i\phi_1}\sin\theta_1|1\rangle$. This is the information-theoretic sense in which a quantum state can encode an arbitrarily large amount of information.[3]

Jozsa next points out that there are restrictions on the information obtainable by measurements on a quantum state. An arbitrarily large amount of information could be communicated by sending a qubit from a transmitter to a receiver if the receiver were able to precisely identify the quantum state that was sent, in which case, the information source's choice from an arbitrarily large ensemble would be revealed at the destination. In general this is not possible. As we will discuss later, Holevo's theorem shows that at most one bit of information can be communicated using a qubit in the above manner. Jozsa writes, "This phenomenon of inaccessibility has no classical analogue"(Jozsa 1998, 50).

Quantum measurement theory tells us that the state of a qubit cannot be determined completely by a measurement. This is in contrast to a system described by classical mechanics, whose state can, in theory, be determined precisely. Here Jozsa has put his finger on a difference between the information storage and retrieval capacities of classical and quantum systems. Since the Shannon theory does not require classical mechanics, this point does not require a new concept of information to describe.

4.3.2 Quantum information cannot be copied

The second property of quantum information Jozsa discusses is that quantum information cannot be copied. First, we must ask ourselves how Shannon information is copied. Copying Shannon information amounts to copying the signal state in a communication system whereby the correlations necessary for information transfer are preserved in the copy. If we choose as our signal in a communication system a quantum state, the no cloning theorem indicates that we will not be able to make a copy of that state. This fact points to limitations on how Shannon information can be manipulated and transferred when quantum systems are used to store information. Since Shannon's theory does not require classical systems, this fact does not suggest that a new concept of information is required.

[3] There are alternative means of encoding Shannon information in a qubit, see Caves and Fuchs (1996), 14.

4.3.3 Quantum systems embody more information than classical systems

Jozsa points out that there, "...is another intrinsically more quantum mechanical sense in which quantum states can embody vastly more "information" than classical states"(Jozsa 1998, 52). An arbitrary state of n qubits will be a superposition requiring 2^n *coefficients* to specify, one for each possible term in the superposition. A string of n cbits has only 2^n *states* available to it and can be specified using n binary numbers. Even if terms in the state description of a superposition of n qubits, only take the values -1, 0, 1, exponentially more binary numbers will be required to specify the quantum state. This is meant to emphasize that even when the continuous nature of qubits is not exploited, n qubits have exponentially more states available to them than n cbits. As a result, qubits can store more information than cbits. Again, Shannon's theory does not place restrictions on the types of physical systems that can be used in a communication system. This property of quantum systems does not require a new concept of information.

4.4 QUANTUM INFORMATION-THEORETIC PHENOMENA

I will now proceed to examine several phenomena, quantum dense coding, teleportation, and the compression of quantum information, each regarded by Jozsa as demonstrating a special property of quantum information. I will argue that each of these "properties" result from manipulations of qubits rather than cbits and provides no evidence that a new concept of information is required. First, we must consider the Holevo bound, which demonstrates mathematically our limited ability to access Shannon information stored in quantum states. The Holevo bound will be a useful conceptual tool for analyzing quantum dense coding, teleportation, and Schumacher coding.

4.4.1 The Holevo bound

Prior to discussing the Holevo bound, it is necessary to define the von Neumann entropy. The von Neumann entropy of a quantum system described by the density operator ρ is defined to be

$$S(\rho) = -tr(\rho \log \rho). \tag{4.1}$$

The formal relation of the von Neumann entropy to the Shannon entropy is straightforward. Density operators admit an orthogonal decomposition, $\rho = \sum_i \lambda_i |i\rangle\langle i|$, where the λ_i's are the eigenvalues of ρ. From (4), the von Neumann entropy is

$$S(\rho) = -\sum_i \lambda_i \log \lambda_i. \tag{4.2}$$

Since ρ is a density operator, the λ_i's will sum to one, are positive, and hence define a probability distribution. If Λ is an ensemble defined by the λ_i's, $H(\Lambda) = S(\rho)$. We may now proceed to discuss the Holevo bound.

The Holevo bound provides a bound on the total amount of Shannon information that can be transferred using a quantum state. The presentation of the theorem follows Nielsen and Chuang (2000). We suppose that the information source is characterized by finite ensemble $X := \{\rho_x, p_x\}$ where $x = 1, \ldots, n$. We assume that the transmitter sends a quantum state corresponding to the notation used for X. If x is the message chosen by the information source, the transmitter sends the quantum state ρ_x to the receiver. The receiver will perform a POVM measurement with elements $\{E_y\} = \{E_0, \ldots, E_{n'}\}$. We assume that element E_i occurs with probability p_i. This probability distribution will define the ensemble that characterizes the destination, $Y := \{y, p_y\}$ where $y = 1, \ldots, n'$. The Holevo bounds limits the mutual information to

$$H(X : Y) \leq S(\rho) - \sum_x p_x S(\rho_x),$$

where $\rho = \sum_x p_x \rho_x$.

The Holevo bound demonstrates how the choice of signal states and the particular measurements performed on a quantum system will determine the amount of Shannon information that can be communicated by a quantum system. The maximum of $H(X : Y)$ over all possible POVMs is referred to as the *accessible information*.

4.4.2 Quantum dense coding

The Holevo bound limits the accessible information of a qubit to one bit. Jozsa writes, "Remarkably however, if the qubit is entangled with another qubit then it may be used to communicate *two* bits of classical information from Alice to Bob"(Jozsa 1998, 53). This phenomenon is known as quantum dense coding. Jozsa claims that this is an instance of doubling the information capacity of a single qubit(Jozsa 1998, 54). The protocol is as follows.(See Figure 2.)

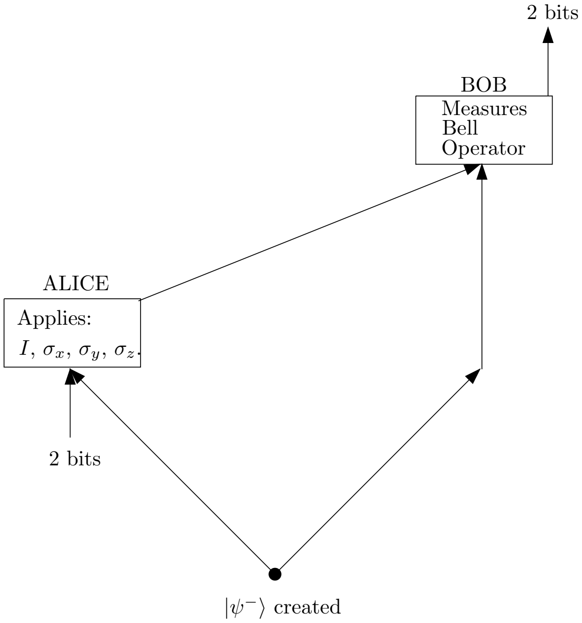

Figure 2: Dense coding. An EPR pair is created and Alice and Bob each receive one member of the pair. Alice encodes 2 bits of information in the pair by performing one of four operations on it. Alice sends her member of the EPR pair to Bob where he performs a joint measurement on the pair. The results of his measurement reveal the information Alice sought to communicate.

Alice and Bob each share one member of an EPR pair in a singlet state, $|\psi^-\rangle$, defined in (4.3) below. Alice performs one of four unitary operations specified by the operators I, the identity, or $\sigma_{x,y,z}$, the Pauli matrices, on her member of the EPR pair. This prepares one of the four states

$$|\psi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|0\rangle|1\rangle \pm |1\rangle|0\rangle) \;\; or \;\; |\phi^{\pm}\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle \pm |1\rangle|1\rangle).^{4} \tag{4.3}$$

These are the complete set of eigenstates of the Bell operator. If Alice sends her member of the EPR pair to Bob, he can measure the Bell operator on the entangled pair and perfectly distinguish all of the states in (4.3). Hence, it appears that despite the Holevo bound, Alice can transfer two bits of information by only sending one qubit to Bob. I claim that from a Shannon information-theoretic point of view, there is nothing out of the ordinary happening in quantum dense coding. I demonstrate this by applying the notion of a communication system to the dense coding protocol.

Alice will be the information source. We assume she can be characterized by an ensemble, A, with four equiprobable messages. The transmitter will utilize quantum states to communicate Alice's message. In fact, we denote the members of A to correspond to the quantum states the transmitter will use to code Alice's message as $\{|\psi^+\rangle, |\psi^-\rangle, |\phi^+\rangle, |\phi^-\rangle\}$. We let Bob, the destination, be represented by the same ensemble which we denote as B for clarity. We assume that the receiver performs the Bell operator measurement and determines Alice's message with certainty. The EPR *pair* is the channel that carries the signal, the quantum state, from the transmitter to the receiver. Shannon's framework for communication applies perfectly well to this situation.

Jozsa seems to be worried that Alice packs too much information in one qubit. Formally, there is no problem. This is verified by a quick look at the Holevo bound. Let $\rho = \sum_i p_i \rho_i$, where the ρ_i are density operators that describe the signal states in (4.3). We know that $H(A : B) \leq S(\rho) - \sum_i p_i S(\rho_i)$. Since the ρ_i are pure states, the $S(\rho_i)$'s vanish, so $H(A : B) \leq S(\rho)$. Since ρ is the maximally mixed state of two qubits, $S(\rho) = 2$. Therefore, the Holevo bound suggests that it is possible for Bob to extract two bits of information from the qubits. Is there a conceptual problem with dense coding?

Information transfer using the dense coding protocol is mysterious only if one fails to recognize features (II) and (III) of the Shannon theory. Namely, transfer of Shannon information does not require a local carrier of information and the Shannon information content

[4] I ignore global phase factor due to the application of σ_y.

of a physical system can only be quantified in the context of a communication system. If we fail to recognize these features, one is tempted to suppose that information must be located "in" Alice's member of the EPR pair because it is the only potential carrier of information traveling a spatially continuous route from Alice to Bob. Since we know that the total quantity of information transferred to Bob is two bits, our supposition will also lead us to mistakenly assign an information content of two bits to Alice's member of the EPR pair. This is in contradiction to the Holevo bound.

If we do recognize features (II) and (III), it is an easy matter to explain how information is transferred in the dense coding protocol. Information is transferred in virtue of the correlations established between Alice's choice and the state of the EPR pair. What is peculiar about the situation is that Alice has the ability to prepare the joint state of *two* qubits by acting locally on only *one* member of the pair in a non-classical way. Alice can change the global state of the system without changing the individual state of the system she acts upon. I.e. Alice changes the density operator for the EPR pair while leaving the density operator for the qubit she acts upon invariant. The fact that Alice can do this is not perplexing property of information, but a perplexing fact about quantum systems. Alice prepares a two qubit system to send to Bob in one of four distinct states capable of being perfectly distinguished. This allows her to encode two bits of information in the EPR pair. The Holevo bound is never challenged. Only when one fails to recognize features (II) and (III) will one be puzzled by the information-theoretic aspects of dense coding. Quantum dense coding simply is a phenomenon that highlights differences in our ability to transfer information depending on whether it is stored in classical or quantum systems.

4.4.3 Quantum teleportation

Quantum teleportation is a process whereby a quantum state is transferred from some spatial location to another spatial location without physically moving the quantum state on a spatially continuous path connecting the two locations. (See Figure 3.)

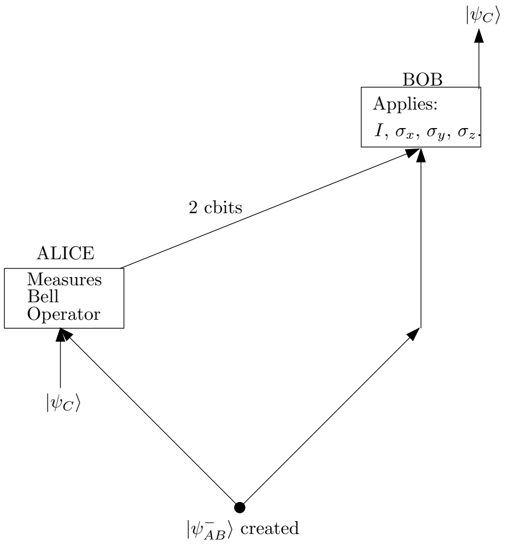

Figure 3: Teleportation. Alice and Bob are given one member of an EPR pair. Alice submits her member of the pair to a joint measurement with an unknown state. Alice sends the results of her measurements to Bob. Bob performs a rotation conditional on the results of Alice's measurement on his member of the EPR pair to ensure successful teleportation.

The teleportation protocol is as follows. Alice seeks to send Bob a quantum state, $|\psi\rangle_C = a|0\rangle_C + b|1\rangle_C$, itself perhaps unknown to Alice. Alice and Bob begin by each receiving a member of a fully entangled pair of qubits in the singlet state. The state vector describing the entangled pair will be labeled with a subscript. The subscript "A" indicates the member of the EPR pair sent to Alice, and the subscript "B" indicates the member of the EPR pair sent to Bob. In what follows we keep the definitions established in (4.3). The entire system is described by the state

$$|\Psi\rangle_{ABC} = \frac{1}{\sqrt{2}}(a|0\rangle_C + b|1\rangle_C)|\psi^-\rangle_{AB}. \tag{4.4}$$

This equation can be rewritten as

$$\begin{array}{rcl} rcl|\Psi\rangle_{ABC} = & \frac{1}{2}(|\psi^-\rangle_{CA}(-a|0\rangle_B - b|1\rangle_B) \\ + & |\psi^+\rangle_{CA}(-a|0\rangle_B + b|1\rangle_B) \\ + & |\phi^-\rangle_{CA}(\ \ b|0\rangle_B + a|1\rangle_B) \\ + & |\phi^+\rangle_{CA}(-b|0\rangle_B + a|1\rangle_B)). \end{array} \tag{4.5}$$

If Alice performs a Bell operator measurement on the quantum system whose state she seeks to teleport and her member of the EPR pair, the state of Bob's qubit will always be one of four fixed unitary transformations, I, or $\sigma_{x,y,z}$, away from $a|0\rangle + b|1\rangle$ according to (4.5).[5] In order to complete the teleportation, Alice need only send 2 cbits to Bob indicating which unitary transformation to perform for successful teleportation.

Jozsa identifies what he thinks is puzzling about quantum teleportation(Jozsa 1998, 58):

> The question is this: Alice succeeds in transferring the quantum state $|\psi\rangle_{[C]}$ to Bob by sending him just two bits of classical information. Clearly these two bits are vastly inadequate to specify the state $|\psi\rangle_{[C]}$ so how does the remaining information get across to Bob? What carries it? What route does it take?

Jozsa's answers to the three questions he poses above are as follows. First, the quantum information travels backwards in time. It proceeds from the Bell operator measurement, where the quantum system whose state is to be teleported interacts with Alice's member of the EPR pair, to the EPR source via Alice's member of the EPR pair. Next, the quantum information travels forward in time to Bob in his member of the EPR pair. All but two bits are transferred in this way. Alice provides the rest using cbits (Jozsa 1998, 58).

Jozsa offers an exotic explanation of information transfer in quantum teleportation. His particular explanation is motivated by the demand that a satisfactory explanation provide answers to the questions: "How does the remaining information get across to Bob?", "What carries it?", and "What route does it take?". I will argue that these demands are unreasonable to require of an explanation of information transfer in quantum teleportation.

To make things precise I will apply the Shannon theory to information transfer using the teleportation protocol. We suppose that Alice is the information source and is characterized by an ensemble, A', whose members are pairs of real numbers from the set $\{(a_1, b_1), \ldots, (a_n, b_n)\}$, and are equiprobable. We know that $H(A') = \log n$, and $H(A')$ increases without bound as n increases. We suppose that there is a transmitter that encodes Alice's choice from A' such that a quantum state has coefficients corresponding to that choice. Hence, choice (a_1, b_1) would be encoded as the quantum state $|\psi\rangle_C = a_1|0\rangle + b_1|1\rangle$. The signal will be transferred using a quantum channel. This channel is described by the teleportation protocol above. Hence, we assume that the transmitter will send two cbits

[5] Again, the global phase factor associated with σ_y is ignored.

to the receiver indicating the result of the Bell operator measurement. The receiver will perform the transformation on qubit requisite for successful teleportation. At this point, the receiver will have a quantum state perfectly correlated with Alice's choice from A'.

Shannon information is transferred in the teleportation protocol from Alice to Bob only in the sense that Bob possesses a quantum state perfectly correlated with Alice's choice from an ensemble. Note, this is not the conventional sense of information transfer, i.e. that uncertainty is removed at the destination about the information source's choice of message. Bob is no less uncertain about Alice's message when perfect teleportation occurs than when it does not because no measurements are performed on the teleported qubit. Even if the receiver did perform measurements, the Holevo bound would restrict the accessible information to one bit, an uncertainty far less than messages from Alice's ensemble. Nonetheless, there is a precise Shannon information-theoretic sense in which information is transferred, namely that a system whose state is perfectly correlated with an information source is in Bob's possession.

Regarding Jozsa, it is incorrect to suppose that the two cbits that indicate the appropriate unitary transformation for successful teleportation somehow provides two bits that go into identifying Alice's choice from her ensemble. Even if A' only had one member, corresponding to a perfectly certain choice, whereby no information would be transferred, the cbits would still be required in a communication system utilizing quantum teleportation. Even in the fully general case, where A' has arbitrarily many members, no story need be told regarding how the "remaining" bits of information are transferred.

The questions, "What carries it?", and "What route does it take?", clearly suppose the notion of a carrier of information which travels a continuous spatial path in a communication system. Feature (II) of the Shannon theory indicates that this is not required for an account of information transfer. Any method of instantiating correlations in parts of a communication system is allowable.

There is nothing puzzling about how information is transferred in a teleportation protocol. The teleportation protocol indicates exactly how correlations can be established between Alice and Bob's ends of the communication system. Information is always transferred in virtue of progressive correlations being established. Unless one asks questions that are ill motivated, the Shannon notion of information works perfectly well. Thus, quantum teleportation does not demand a new notion of information.

4.4.4 Schumacher coding

The Schumacher coding theorem indicates interesting new ways to transfer information. It appears, in some cases, that the Schumacher coding theorem allows us to transfer information at a rate that exceeds the bound set by Shannon's coding theorem. Contrary to appearances, the Schumacher coding theorem presents no challenge to the Shannon theory. In what follows, it will be illuminating to discuss both the Schumacher coding theorem and Schumacher coding itself.[6]

The information source in the Schumacher information transfer scheme is replaced with a quantum signal source, $M := \{\rho_x, p_x\}$ which emits qubits in the state ρ_x with probability p_x. Similar to Shannon coding, the Schumacher compression method exploits the statistical properties of the quantum signal source, and the quantum physics of the source too, to minimize the resources required to transfer the sequence emitted from the source with high fidelity.[7] Roughly, it works as follows. We suppose the sequence of messages emitted by a quantum signal source is length N. Encoding information in a sequence of N qubits amounts to specifying a vector in a Hilbert space of dimension 2^N. Similar to Shannon, Schumacher demonstrates that with high probability the vector in this 2^N dimensional Hilbert space that specifies the sequence has support in a limited region of the space. If ρ is the average density operator of M, the vector describing the sequence will have support in a region of approximate dimension $2^{NS(\rho)}$. To specify a vector in a $2^{NS(\rho)}$ dimensional space, only $N(S(\rho))$ qubits are required. So, one need only send $NS(\rho)$ qubits to recover the initial sequence with high fidelity.

Similar to the Shannon coding theorem the bulk of the proof of the Schumacher coding

[6] In this section, we limit discussion to the coding of pure states following Schumacher's (1995) closely. We refer the reader to Barnum, et al. (2000) for a fuller discussion of coding mixed states as well.

[7] To define the fidelity of the transfer, suppose that M emits a sequence of N qubits described by the density operator π_a with probability p_a, where $a = 1, \ldots, |N^*|$ where $|N^*|$ is the number of sequences of length N from M. Suppose that the final state after Schumacher coding of π_a is ω_a. The fidelity of the coding scheme is defined to be $F = \sum_a p_a Tr(\pi_a \omega_a)$. The fidelity measures how well the states ω_a represent the states π_a.

theorem is in the proof of the typical subspace theorem. We examine the theorem to better understand the relationship between the Shannon and Schumacher coding theorems.

This presentation closely follows Nielsen and Chuang (2000). Schumacher coding is actually much like Shannon coding formally. We suppose that a quantum signal source $M := \{\rho_x, p_x\}$ emits a sequence of qubits. We also suppose that the ρ_x form an orthogonal decomposition of the density operator describing M, $\rho = \sum_x p_x \rho_x$.[8] We can define an ϵ-typical sequence of qubits, $m_1, \ldots, m_N$ as a sequence whose probability, $p_{m_1} p_{m_2} \cdots p_{m_N}$ satisfies

$$\left| \frac{1}{N} \log \frac{1}{p_{m_1} p_{m_2} \cdots p_{m_N}} - S(\rho) \right| \leq \epsilon.$$

Define an ϵ-typical subspace, $\mathcal{T}(N, \epsilon)$, to be the space spanned by all ϵ-typical sequences of qubits, and the projector onto that subspace to be $P(N, \epsilon)$. We denote the N-fold tensor product of ρ as $\rho^{\otimes N}$. With the above, we can prove the following:

1. Fix $\epsilon > 0$. Then for any $\delta > 0$, and for sufficiently large N, the probability that a sequence of qubits is ϵ-typical, $tr(P(N, \epsilon)\rho^{\otimes N})$, is greater than or equal to $1 - \delta$.
2. For any fixed $\epsilon > 0$ and $\delta > 0$, and for sufficiently large N, the dimension, $|\mathcal{T}(N, \epsilon)|$, of the ϵ-typical subspace satisfies

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

The state of any N length sequence of qubits can be written as a superposition of vectors from $\mathcal{T}(N, \epsilon)$ and $\mathcal{T}(N, \epsilon)^{\perp}$, where $\mathcal{T}(N, \epsilon) \otimes \mathcal{T}(N, \epsilon)^{\perp}$ is the Hilbert space describing N qubits. Item one tells us that we can make the probability of projecting onto the typical subspace arbitrarily close to 1. Item two tells us that the dimension of the typical subspace always is less than or equal to $2^{N(S(\rho)+\epsilon)}$.

In order to implement Schumacher coding, first perform a projective measurement described by the complete set of operators $P(N, \epsilon)$ and $I - P(N, \epsilon)$. As item one suggests, this will almost always project the system to $\mathcal{T}(N, \epsilon)$. If the system projects to $\mathcal{T}(N, \epsilon)^{\perp}$, set

[8] This assumption is made for simplicity, but nothing depends on it. The Schumacher coding theorem holds for any quantum source that emits pure states. Trivially, the same compression that the Schumacher coding theorem achieves with pure states can be achieved when mixed states are sent. One simply sends pure states whose average density operator is equivalent those that describes a source that emits mixed states (Barnum, et al. 2000).

the state of the qubits to some standard state in $\mathcal{T}(N, \epsilon)$. Next, perform a unitary transformation, U. This unitary transformation essentially splits the N length sequence into two parts, one of length $N(S(\rho) + \epsilon)$, denoted by D, and another of length $N - N(S(\rho) + \epsilon)$, denoted by F. We assume that that the state of D is correlated to the state of the qubits prior to the application of U. We assume F is in some standard state, $|F\rangle$. We discard F and send only D to the destination. At the destination a system of qubits denoted F', with state $|F'\rangle$, is adjoined to D. We choose $|F'\rangle$ so $|F'\rangle = |F\rangle$. U^{-1} is applied to combined system $D + F'$ which results in a length N sequence of qubits in a state from $\mathcal{T}(N, \epsilon)$. Since item one indicates we will almost always project to a typical subspace, the above coding technique will transfer quantum states with high fidelity.

At this point, the coding theorem is mostly proved. Define the quantum channel capacity, Q, to be the rate at which qubits are available for transfer. Following Schumacher (1995), the coding theorem states:

> Let M be a quantum signal source described by the density operator ρ and let $\delta, \epsilon > 0$. For sufficiently large N, a sequence of length N from M can be transmitted at a rate $Q/(S(\rho) + \epsilon)$ (qubits/second) with fidelity $1 - \delta$. There is no high fidelity coding scheme for rates greater than $Q/(S(\rho) - \epsilon)$.

The Schumacher coding theorem determines the best rate at which a sequence of N qubits can be transferred with high fidelity.

Jozsa suggests that the Schumacher coding theory offers new possibilities for information transfer. The following example is from Jozsa's (1998). Suppose that we consider the ensemble, J, whose members are equiprobable and from the set $\{|\psi_0\rangle, |\psi_1\rangle\}$. We will compare the information transfer rates of two different communication systems. One will use cbits and Shannon coding to transfer information, the other will use qubits and Schumacher coding to transfer information. We will refer to the former as the standard communication system and the latter as the quantum communication system. We assume that the information source for both communication systems is characterized by J. We suppose that the transmitter in the quantum communication system encodes an information source's choices from J in quantum states according to the notion used for members of J. In addition, we assume that this transmitter performs Schumacher compression on those quantum states and the receiver performs decompression and possibly measurements on the decompressed qubits. Suppose the qubit states corresponding to members of J are

$$|\psi_0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ |\psi_1\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$

In this case, $S(\rho) = .601$ whereas $H(J) = 1$. An N length sequence of messages will require approximately $NH(J)$ cbits to represent, whereas only approximately $NS(\rho)$ qubits are required to represent the messages.

Each message from J has an uncertainty $H(J)$. The information content of a system perfectly correlated with N messages from J is $NH(J)$ bits. Let us make the idealizing assumption that the decompressed state of the N qubits is perfectly correlated with choices at the information source. In this case quantum communication system transfers $NH(J)$ bits using approximately $NS(\rho)$ qubits.[9] The standard communication system uses $NH(J)$ cbits to transfer $NH(J)$ bits. Assume that the transmitter in the standard communication system emits one cbit per second and transmitter in the quantum communication system emits one qubit per second. In this case, the information transfer rate for the quantum communication system will be $1/S(\rho + \epsilon)$ whereas the information transfer rate for the standard communication system is $1/H(J)$. Since $H(J) > S(\rho)$ in the above case, the quantum communication system has a better rate of information transfer than the standard communication system.

This result is no challenge to the Shannon coding theorem. The Shannon coding theorem demonstrates that the rate of information transfer is bounded above by C/H where C is the capacity of the channel and H is the entropy of the information source. In the above example, two types of channels were used, one utilizing cbits, the other utilizing qubits. We know from our discussion in sections 4.3.1 and 4.3.3 that qubits have a greater capacity store or encode information than cbits. So, the channels in the standard and quantum communication system described above have different capacities. The Shannon coding theorem will place

[9] As in the teleportation case, the sense of information transfer in Schumacher coding is restricted. No uncertainty about a message from J is removed at the destination. Nonetheless, we can regard the Schumacher coding as transferring information in the sense that the state at the destination is perfectly correlated to choices at the information source. The accessible information of N qubits from a quantum signal source described by density operator ρ is limited to only $NS(\rho)$ bits(Hausladen, et al. (1996), 1869). So, the quantum communication system described above can remove at most $NS(\rho)$ bits of uncertainty at the destination whereas the standard communication system can remove all $NH(J)$ bits of uncertainty.

different bounds on information transfer rates in the above communication systems based on the different channel capacities. Hence, the Shannon coding theorem could have been used to *predict* that different information transfer rates were possible in the above example.

Some might worry that the resolution of the problem above depends on the information storage capacities of qubits discussed in section 4.3.3, that it is arbitrarily large, but that the technique used for encoding information into a sequence of qubits prior to Schumacher compression does not seem to utilize these capacities. This worry is resolved by pointing out that we can have similar compression behavior in a standard communication system. It doesn't matter whether the sequence of messages produced by the information source is initially encoded in a string of cbits to maximize compression, or, as in the Schumacher compression scheme, the sequence of messages is encoded in a string of cbits in an uncompressed manner, and a compression operation is then applied to the string of cbits.

As we have seen, the Shannon theory can be applied to a communication system that utilizes Schumacher compression to transfer information. The Shannon coding theorem is not challenged by the Schumacher coding theorem. The Schumacher coding method indicates interesting new ways to transfer information when qubits instead of cbits are used in a communication system. Since Shannon's theory places no restrictions on the types of systems that can be used to transfer information, Schumacher coding cannot be regarded as demonstrating the need for a new concept of information.

4.5 QUANTUM INFORMATION = QUANTUM STATE?

The advocate of a new concept of information might agree that the Shannon concept of information can be applied to quantum systems perfectly well. Nonetheless, they might suggest there remains a collection of properties that are well motivated that warrant a new concept of information. They might suggest that the quantum information is an absolute property of a quantum system identified by the quantum state, it is inaccessible(we cannot determine the state of a quantum system perfectly), it cannot be copied (no cloning theorem), it takes a large amount of information to specify(quantum states require continuous parameters to specify), it can be transferred nonlocally(teleportation), and it can be compressed(Schumacher coding). It is obvious that there is already a concept that covers all of these properties: the quantum state. The term “quantum information” is then just a synonym for an old concept.

This use of the term “quantum information” as a synonym for “quantum information” is highly misleading. Typically, for purposes of clarity, when a common term is used technically, there is a strong connection between the technical use of the term and its common usage. The suggested usage of “quantum information” fails in this respect. When “quantum information” is used as a synonym for “quantum state”, there is no connection between our everyday usage of the term “information” and the meaning of “quantum information”. For example, central to our common everyday concept of information is the notion that when information is transferred, uncertainty is removed at a destination. When the term “quantum information” is used as a synonym for “quantum state”, transfer of quantum information amounts to transfer of a quantum state. It is difficult interpret this kind of information transfer as removal of uncertainty at a destination. The connection between the common usage of the term “information” and its use in the term “quantum information” is lost. So, there seems to be little motivation for using “quantum information” as a synonym for “quantum state”. I would urge that the term “quantum information” be reserved to refer to Shannon information stored in quantum systems.

4.6 CONCLUSION

I have argued that the considerations that Jozsa offers in “Quantum Information and its Properties” do not force us to introduce a new notion of information. All properties meant to contrast quantum information and Shannon information were shown to be particular to the system that transfers or stores information, not properties of information. Several quantum phenomena, dense coding, teleportation, and Schumacher coding, have been examined. All puzzling information-theoretic features of these phenomena can be traced to puzzling features of quantum systems rather than puzzling features of information. All of the phenomena considered can be characterized information-theoretically using the Shannon theory. So, these phenomena do not necessitate the introduction of a new concept of information. Finally, I have argued that insofar as the properties of quantum information can be made precise and well motivated, it coincides with the properties of a quantum state. So, based on Jozsa's arguments, no *new* concept seems to need to be introduced. This paper began by asking what quantum information is. The arguments in the paper point to one answer. Quantum information is just Shannon information stored in quantum systems.

5.0 TRADING OBSERVABILITY FOR LOCALITY: QUANTUM INFORMATION AND THE DEUTSCH AND HAYDEN APPROACH

5.1 INTRODUCTION

In their (1999) Deutsch and Hayden purport to have developed a version of quantum theory that is local and non-holistic. They claim that "...a complete description of a composite system can always be deduced from complete descriptions of its subsystems, where under those descriptions, 'the real factual situation of the system S_2 is independent of what is done with the system S_1, which is spatially separated from the former"' (Deutsch and Hayden 1999, 2). Deutsch and Hayden couch their analysis in terms of quantum information. Deutsch and Hayden claim that "All phenomena that have been thought to demonstrate nonlocality in quantum physics are actually due to the phenomenon of locally inaccessible information"(*ibid.* 19).

Deutsch and Hayden claim that quantum theory is explicitly local and non-holistic, and furthermore, that seemingly non-local and holistic behavior of quantum systems is due to the strange properties of an explicitly local quantity: locally inaccessible quantum information. Properties of quantum information obviously play a crucial role in substantiating Deutsch and Hayden's claims. As we will see, the properties that can be attributed to quantum information are critically dependent on the interpretation one gives to the Deutsch and Hayden formalism. I will argue that the concept of quantum information that Deutsch and Hayden introduce does not have the locality properties that Deutsch and Hayden need to support their contention that quantum theory is local and non-holistic. Furthermore, their concept of quantum information cannot be used to support explanations of EPR correlations and quantum teleportation.

In this paper, two interpretations of the Deutsch and Hayden formalism due to Timpson (2003) are adopted: the conservative interpretation and the ontological interpretation. It will be argued that on the conservative interpretation of the Deutsch and Hayden formalism, quantum information is neither an absolute property of a system, nor is it a local quantity. Hence, on this interpretation, quantum information will not yield the local non-holistic quantum theory that Deutsch and Hayden claimed to have developed. In the ontological interpretation, quantum information is again, not an absolute property of a system, but a relational one. In contrast to the conservative interpretation, quantum information is explicitly local on this interpretation. The reason quantum information is explicitly local is due to the introduction of new absolute properties of quantum systems. These properties are the carriers of information in quantum systems. These new properties support Deutsch and Hayden's claims to have developed a quantum theory free from non-locality and holism. However metaphysically soothing a local non-holistic quantum theory is, it comes at a price: our ontology is inflated by strange new properties of quantum systems are unobservable and underdetermined.

In addition to banishing non-locality and holism from quantum physics, Deutsch and Hayden also attempt to account for EPR correlations and quantum teleportation by tracking locally inaccessible information as it is transferred in quantum systems. According to the formalism they develop, information flows locally in both instances. Deutsch and Hayden are mysteriously silent as to how this is supposed to account for or explain EPR correlations or teleportation. In any case, the mere transfer of information is doing the explanatory work if anything is. So, Deutsch and Hayden can be read as attempting to employ transfer explanations of quantum phenomena. Transfer explanations explain a phenomenon by appealing to the fact that a physical quantity is transferred in an interaction between physical systems which at least partially determines their resulting behavior.

Explanations of EPR correlations and quantum teleportation depend on properties of quantum information. Since the properties of quantum information are interpretation dependent, so too will be the kinds of explanations that can be provided that implicate quantum information. As remarked above, quantum information is not a local quantity on the conservative interpretation. No local transfer explanation is possible using this quantity. On the conservative interpretation, quantum information is local, but it is not an absolute property of a system. Instead, a quantum system carries the quantum information it does in virtue of the strange underlying properties that Deutsch and Hayden introduce being correlated with an information source. Information is not a quantity, but a relational property of a system on this account. A quantity transfer explanation is again not appropriate. Instead, as I will argue, what Deutsch and Hayden can do, but do not, is utilize the properties introduced in the ontological interpretation to provide a local causal account of both EPR correlations and quantum teleportation.

In section 5.2 I will describe the formalism Deutsch and Hayden use to support their contentions that quantum theory is local and non-holistic. In section 5.3 I will outline possible interpretations of the Deutsch and Hayden formalism due to Timpson (2003). In section 5.4 I articulate Deutsch and Hayden's concept of information. In section 5.5 I examine the Deutsch and Hayden analysis of EPR experiments and quantum teleportation. I argue that quantum information plays no explanatory role for these phenomena. Finally, I supplement Deutsch and Hayden's analysis by pointing out that a local causal explanations of those phenomena can be provided.

5.2 THE DEUTSCH AND HAYDEN FORMALISM

Deutsch and Hayden adopt a Heisenberg picture of quantum mechanics.[1] In order to track the evolution of a system, one needs to keep track of how all observables of a system evolve. The Deutsch and Hayden formalism is designed to efficiently represent a complete description of a system of n qubits, i.e. the evolution of all quantum mechanical observables for the system.[2] The Deutsch and Hayden formalism allows one to describe the evolution of all system observables by tracking the evolution of a small set of local observables of each quantum system. Deutsch and Hayden describe each qubit by a triple whereby qubit k is

[1] See Appendix A for a description of differences between the Heisenberg and Schrödinger pictures of quantum mechanics.

[2] Efficient representations of quantum systems is of great importance for simulating quantum systems with classical computers.

given by

$$\hat{\mathbf{q}}_k(t) = \left(\hat{q}_{kx}(t), \hat{q}_{ky}(t), \hat{q}_{kz}(t)\right). \tag{5.1}$$

The components in the above equation satisfy

$$\begin{aligned}
[\hat{\mathbf{q}}_{\mathbf{k}}(t), \hat{\mathbf{q}}_{\mathbf{k'}}(t)] &= 0\; for\; (k \neq k') \\
\hat{q}_{kx}(t)\hat{q}_{ky}(t) &= i\hat{q}_{kz} \\
\hat{q}_{kx}(t)^2 &= I
\end{aligned} \tag{5.2}$$

with cyclic permutations over x, y, and z. The $\hat{\mathbf{q}}_k(t)$ are a representation of the Pauli operators, denoted by X, Y, and Z. Deutsch and Hayden choose the initial representation of $\hat{\mathbf{q}}_k(t)$ as

$$\hat{\mathbf{q}}_k(0) = \left(I^{k-1} \otimes X \otimes I^{n-k}, I^{k-1} \otimes Y \otimes I^{n-k}, I^{k-1} \otimes Z \otimes I^{n-k}\right), \tag{5.3}$$

where I^m is the m-fold tensor product of I's.

We refer to $\hat{\mathbf{q}}_k$ as the *descriptor* for qubit k. The descriptor evolves through Heisenberg evolution of each component matrix of the descriptor. It is a mathematical fact that suitable fixed mathematical operations on the components of the descriptors of a system are sufficient to determine the evolution of any observable of the system, global or local. This is established in appendix B. In conjunction with a fixed initial state, the descriptors can be used to recover all empirical predictions of quantum theory.[3] This formalism has the property that local transformations on individual qubits, or sets of qubits, only change the descriptors of the qubits involved in the process. Following Timpson (2003) we will call this property *contiguity.*

The descriptor of a system at time t will depend only on the local interactions it has been subject to and the history of interactions of any systems that it has had direct interactions with. For example, consider a system of two qubits. The initial descriptors will be

$$\begin{aligned}
\hat{\mathbf{q}}_1(0) &= (X \otimes I, Y \otimes I, Z \otimes I) \\
\hat{\mathbf{q}}_2(0) &= (I \otimes X, I \otimes Y, I \otimes Z).
\end{aligned} \tag{5.4}$$

[3] Interestingly, the evolution of the descriptors encodes the dynamics for *all* initial quantum states. This is in contrast to the Schrödinger picture where evolution is particular to a single initial state.

For all dynamical interactions associated with system 1 alone, spatial rotations, phase flips, etc., the evolution operator for the system, U, will factorize to $U_1 \otimes I$ (and similar for 2). The descriptors for systems 1 and 2 evolve as

$$\begin{aligned} \hat{\mathbf{q}}_1(1) &= U^\dagger \hat{\mathbf{q}}_1(0) U \\ \hat{\mathbf{q}}_2(1) &= U^\dagger \hat{\mathbf{q}}_2(0) U, \end{aligned} \tag{5.5}$$

where the evolution operators are distributed across the triple. More explicitly, in the case we are considering, we have

$$\begin{aligned} \hat{\mathbf{q}}_1(1) &= U_1^\dagger \otimes I (X \otimes I, Y \otimes I, Z \otimes I) U_1 \otimes I \\ \hat{\mathbf{q}}_2(1) &= U_1^\dagger \otimes I (I \otimes X, I \otimes Y, I \otimes Z) U_1 \otimes I. \end{aligned} \tag{5.6}$$

Clearly, $\hat{\mathbf{q}}_2(1) = \hat{\mathbf{q}}_2(0)$; U acts as the identity on system 2 in this case. For nonseparable evolution operators, those U such that $U \neq U_1 \otimes U_2$, the descriptors for systems 1 and 2 will display mutual dependence after the interaction. It is of interest to note that evolutions that might initially have vanishing effect on a descriptor at some initial time, might have a nontrivial effect on a descriptor at a later time. Suppose $U(1)$ and $U(2)$ represent a sequence of interactions at $t = 1, 2$. Recall that in the Heisenberg picture the descriptors evolve from $t = 0$ to $t = 2$ as $U(1)^\dagger U(2)^\dagger \hat{\mathbf{q}}_k(0) U(2) U(1)$. Notice the order of operations. In the case where U(1) is a local interaction on system 1, and U(2) is a nonseparable interaction on systems 1 and 2, even when the effect of U(1) is trivial for the descriptor for system 2 at $t = 1$, that same operation will generally change the descriptor of system 2 at $t = 2$. It will pick up a dependence on system 1 due to the way descriptors evolve with sequences of interactions.

Given the above formal machinery, and locality properties, we proceed to interpret that machinery in order to assess the potential benefits of adopting such a formalism for quantum theory.

5.3 INTERPRETING THE FORMALISM

Timpson (2003) has argued convincingly that there are two different interpretations that can be given to the Deutsch and Hayden formalism, which he dubs the conservative interpretation and the ontological interpretation.[4] Distinguishing between these two interpretations is crucial because they have very different stories to tell about information transfer and locality, and derivatively, how far Deutsch and Hayden go towards satisfying their goal of demonstrating that quantum theory can be interpreted as a local non-holistic theory.

5.3.1 The conservative interpretation

The conservative interpretation of the Deutsch and Hayden formalism is not much of an interpretation at all. The conservative interpretation takes the Deutsch-Hayden formalism simply as a different way of mathematically expressing standard quantum mechanics. Deutsch and Hayden view the Schrödinger picture of quantum theory as a non-local version of a demonstrably local theory, with standard quantum mechanics interpreted as demonstrably local in the Heisenberg picture using the Deutsch-Hayden formalism (Deutsch and Hayden (1999), 22-3). That the interpretive problems of quantum theory can be swept under the rug with a simple mathematical trick surely gives one reason to view Deutsch and Hayden's claims skeptically.

We should point out that the backdrop of the conservative interpretation is no-collapse quantum mechanics, with all of the advantages and drawbacks that come with it. No hidden variables are added and there are no nonlocal collapses of the state vector. Thus, any special claims to locality beyond the assumption of no-collapse quantum mechanics must be interpreted as pointing out that quantum theory is not holistic. There are two reasons for viewing the Schrödinger picture of quantum mechanics as holistic. First, the reduced density operators that describe individual quantum systems are not sufficient to reproduce the empirical results of the theory. The example to keep in mind here is an EPR pair where the reduced density operators of two entangled systems will be maximally mixed, whereas the

[4] It is clear that these two distinct interpretations are conflated in Deutsch and Hayden (1999). As Timpson (2003) points out, there is textual support for both readings.

quantum state that describes this pair is pure. The second reason to view the Schrödinger picture as holistic is that local operations on entangled subsystems have a global effect on the quantum state in fundamentally non-classical ways. The example to keep in mind here is dense coding, whereby a unitary operation on one member of an EPR pair can steer the global state of the pair into four different orthogonal states, but the reduced density matrixes of the pair submitted to the operation are unchanged.[5] Deutsch and Hayden view their formalism as demonstrating that these holistic properties can be vanquished. Recall the two claims to locality that Deutsch and Hayden championed: once the state-vector is fixed all empirical predictions of the theory can be deduced by tracking the evolution of descriptors of individual systems, and contiguity, whereby local unitary operations on particular systems only change the descriptors of those systems directly involved in the operations.

Regarding the first notion of holism, it must be pointed out that in the Deutsch and Hayden picture interpreted conservatively, properties of quantum systems are jointly determined by the descriptors and the initial quantum state. It should be noted that even though the conservative interpretation places emphasis on the role that descriptors play in determining properties of a system, these properties are no different than those determined in the Shrödinger picture, where the focus is on the role of the state in determining properties. The state vector of the system, the *global* state vector, still plays a crucial role in determining properties of a system, and the empirical predictions of the theory. This crucial role is especially emphasized when the initial state is non-factorizable.[6]

Regarding the second notion of holism, it is clear that contiguity does not vanquish the kind of holism it was meant to. The worry in this case is that changes to subsystems change the global state without changing the local state. There is a clear analogue in the Deutsch and Hayden formalism. It is true that descriptors do have the contiguity property, all local unitary interactions leave their mark on the descriptor for the local system, but this is not the whole story. Local interactions will change descriptors for the systems they act upon,

[5] Timpson (2003) is careful to point out that this not properly termed a local or nonlocal effect. Nothing hinges on whether we call this non-local or not, the key feature is that it is a fundamentally non-classical way to change a global state.

[6] It is true that when the global state is not entangled that local changes to subsystems will change the state of only the subsystem submitted to a local interaction, but this is a shared feature of the Schrödinger picture and the Deutsch and Hayden picture.

but in the cases we are concerned with, local interactions will not change any observables for these systems. Instead, the effects of such interactions will change *nonseparable* observables. More specifically, consider two systems S_1 and S_2. In dense coding, S_1 is subjected to a unitary interaction that leaves all observables of the form $O_1 \otimes I_2$ and $I_1 \otimes O_2$ invariant. This interaction does change observables of the form O_{12}, where $O_{12} \neq O_1 \otimes O_2$. When O_{12} is the Bell-operator observable, the local interaction on S_1 can steer the system into one of four eigenvalues of the operator. This is a sense of holism that is a brute feature of the theory, and no change in the way we describe quantum systems will change this fact. If this type of holism is mysterious in the Schrödinger picture, it is equally mysterious in the Deutsch-Hayden formalism interpreted conservatively.

In the Deutsch and Hayden formalism interpreted conservatively, not surprisingly, we are a far cry from a non-holistic theory. In particular, the properties of entangled systems are not seen as being determined by local properties of subsystems. Also, changes in the properties of entangled systems are not seen as arising from changes in local properties of subsystems. A more radical approach is required to remove these holistic features of the theory.

5.3.2 The ontological interpretation

The ontological interpretation is considerably more ambitious than the conservative interpretation. The ontological interpretation views the quantum state of a system as nothing more than a rule for making empirical predictions given the appropriate descriptors for a quantum system. It is assumed that the quantum state of a system is always in a standard (unentangled) state $|00\ldots0\rangle$. The descriptors are interpreted as denoting some sort of real, occurrent, absolute properties of individual quantum systems, of which nothing more is said.[7]

The reader is probably wondering what we are to make of the ontological interpretation when the initial state is not the standard state; $|\psi(0)\rangle \neq |00\ldots0\rangle$. In this case we determine which unitary transformation is required to adjust the state such that $U|\psi(0)\rangle = |00\ldots0\rangle$. Using this unitary operation, the initial descriptors are adjusted to reflect the non-standard

[7] It should be mentioned that these properties do not imply definite values for observables, as some standard hidden variables assignment might. There is no conflict with Bell's theorem.

initial state. In the conservative interpretation, this change represents nothing more than our mathematical freedom to represent quantum systems by changes in states or changes in observables. In the ontological interpretation this adjustment reflects the real occurrent properties of individual systems that underwrite the properties identified by the state in standard no-collapse quantum mechanics. *All* properties of quantum systems, both global and local properties, are reducible to local properties of individual quantum systems, even entanglement! "Thus on the ontological interpretation…the global properties of the joint system are reducible to local, intrinsic properties of subsystems [and] *changes* in the global properties are reducible to changes in the currently possessed properties of subsystems" (Timpson, 2003, 13).

Timpson (2003) notes several drawbacks to the ontological interpretation. First, the descriptors of a quantum system are underdetermined. There is a continuum of descriptors compatible with the empirical predictions of the theory. Nonetheless, we are to interpret descriptors as being correlated one to one with the underlying properties of the system. So, it is an article of faith that there exists a true descriptor of every quantum system. Second, in the ontological interpretation the descriptors represent the intrinsic properties of a single quantum system. This is in contrast to the conservative interpretation whereby the descriptors track the evolution of continuously many possible initial states. The worry is that the mathematical machinery for the task is overkill. As Timpson acknowledges, neither of these are overwhelming objections to the ontological interpretation. We prefer to reserve judgment on the Deutsch and Hayden approach until evaluating any concept of quantum information it might contribute and its usefulness in providing explanations of quantum phenomena.

5.4 DEUTSCH AND HAYDEN'S CONCEPT OF QUANTUM INFORMATION

As remarked in the introduction, Deutsch and Hayden prefer to couch much of their analysis in terms of information. This is particularly apparent in their explanations of EPR experiments and quantum teleportation. In what follows we examine the Deutsch and Hayden notion of information to determine its status as a coherent notion and its potential role in explanations of quantum effects.

Deutsch and Hayden do not offer a precise definition of their concept of information. As such, it is difficult to distinguish whether Deutsch and Hayden intend to introduce a new concept of quantum information, different from the Shannon concept, or that they want to demonstrate that Shannon information stored in quantum systems has certain locality properties. In either case, Deutsch and Hayden intend to demonstrate that the (quantum) information content of a system is deducible from the information contained in its subsystems, and changes in the distribution of information can be understood in terms of subsystems carrying information from one subsystem to another in a continuous spatiotemporal path (Deutsch and Hayden (1999), 1). Since Deutsch and Hayden's discussion of quantum information is limited to these features, our understanding of their concept of quantum information is derivative on the locality properties of information.

Deutsch and Hayden do not offer a precise definition for a system to contain information. One has to extract a reasonable definition knowing the properties that Deutsch and Hayden desire for information, and that these properties must be underwritten by the local and non-holistic properties of their formalism. Timpson has argued that the most natural definition for a system to contain information is (Timpson (2003), 18):

> S contains information about $\theta \leftrightarrow$ its descriptor depends on θ and measurements on the global system $S \cup S^{\perp}$ depend probabilistically on θ.

Deutsch and Hayden distinguish a special type of information, the so-called *locally inaccessible information*. Locally inaccessible information about a parameter θ is information that is contained in a system but no measurements on that system alone depend probabilistically on θ (Deutsch and Hayden (1999), 11-12). The location of locally inaccessible information about θ is determined by the dependence of a descriptor on θ. As an example, suppose that a member, S_1, of a fully entangled pair of qubits S_1 and S_2 is subject to a rotation. We know that there are measurements whose results depend probabilistically on the parameter θ. So, the system S_{12} contains information about θ. We know that no measurements on either qubits S_1 or S_2 alone depend on θ. Since the descriptors satisfy contiguity, we know that the descriptor of S_1 will indeed exhibit dependence on θ do to the dynamic interaction it experienced. According to Deutsch and Hayden, locally inaccessible information is located in S_1.

Crucial to understanding the potential usefulness of explanations of quantum phenomena in terms of information transfer is an understanding of what properties are responsible for the local nature of information on the Deutsch and Hayden approach. Those properties might be understood as properties of quantum information, or perhaps in terms of the information-bearing properties of systems that contain information. Put a different way, quantum information might be an absolute property of a quantum system, or a relational one. To determine which, we must examine the properties of information with respect to the interpretations of the Deutsch and Hayden formalism.

On the conservative interpretation we maintain our standard understanding of properties in no-collapse quantum mechanics. The Deutsch and Hayden approach is nothing more that a new way to express the standard theory. It is useful to focus on locally inaccessible information. When a system contains locally inaccessible information about θ, θ appears in the descriptor for that system. On the conservative interpretation though, the local properties of a system are not in one to one correspondence to descriptor of a system in the presence of entanglement. In that case, the local properties of a system are not correlated to θ. So, in the conservative interpretation there are no local properties to substantiate the claim that information is contained locally in a system. As such, locally inaccessible information cannot be an absolute property of an individual system. The seeming locality of the inaccessible information is a feature of the *description* of the system that contains it, an artifact of the way we choose to describe a situation, rather than being or corresponding to a property of the local system. Information in entangled quantum systems is stored in virtue of the relational properties of two systems that, on the conservative interpretation, are not reducible to local absolute properties of the systems. Contrary to what Deutsch and Hayden desire, information is *distributed* in this interpretation. Locally inaccessible information denotes a fictitious quantity on the conservative interpretation.

Focusing our attention on information in general, there are seemingly no strong reasons to view it as an absolute or relational quantity on the conservative interpretation. We seem to be free to identify information as an absolute property of a system, which would be fixed by the state vector in the Schrödinger picture, or as a relational property whereby a system contains information in virtue of its properties being correlated with an information source. In any case, as we will see, deciding this matter is not crucial to understanding explanations of EPR correlations and teleportation in the conservative interpretation because they rely on the transfer of inaccessible information, which is, based on the above arguments, unacceptable.

Locally inaccessible information has a very different status on the ontological interpretation. On the ontological interpretation descriptors are in one to one correspondence to the local properties of a system. Hence, locally inaccessible information can be stored in a system in virtue of being correlated with these properties, making it a relational property of a system, or be identified with these properties outright as an absolute property of a system. On this interpretation, information does have the locality properties desired by Deutsch and Hayden.

We note in passing that when information is viewed as a relational property in either the conservative or ontological interpretation, it coincides with the Shannon qualitative concept of information.

5.5 INFORMATION FLOW IN ENTANGLED SUBSYSTEMS

We are now in a position to examine Deutsch and Hayden's analysis of EPR experiments and quantum teleportation. Deutsch and Hayden seek an analysis of information flow for these quantum phenomena that is purely local. In what follows we remind the reader that these phenomena are examined from the perspective of no-collapse quantum mechanics.

5.5.1 EPR

The crucial phenomena of interest in EPR experiments are the correlations that occur between measurements at opposite wings of the experiment that do not have a local hidden variables interpretation (Bell, 1964). Countless pages have been written in attempt to make sense of this fact. We examine this phenomenon from the perspective of Deutsch and Hayden's formalism, in order to make clear any beneficial effects of the formalism.

In order to model measurements in each wing of the experiment quantum mechanically, we require a total of four qubits for the experiment. The actual EPR pair will be qubits 2 and 3. Each member of the pair will be subject to an arbitrary rotation about the x-axis. Qubit 1 will measure qubit 2 and qubit 4 will measure 3. Qubits 1 and 4 will be brought together to verify the appropriate correlations. (See Figure 4.)

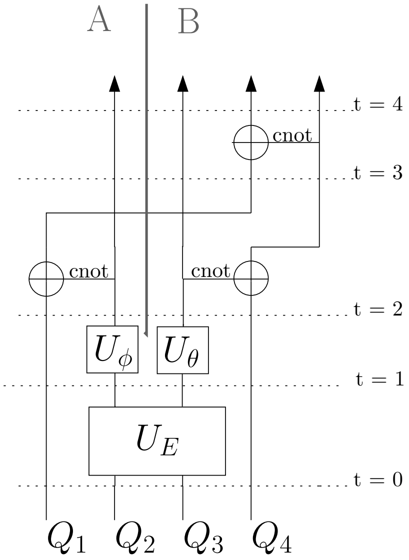

Figure 4: EPR experiment. Initially, qubits 2 and 3 are entangled via U_E. Qubit 2 is then rotated by θ and qubit 3 by ϕ. Qubits 1 and 2 are submitted to a measurement which stores the locally inaccessible quantum information about θ in qubit 1. Similarly, qubits 3 and 4 are submitted to a measurement which stores the locally inaccessible quantum information about ϕ in qubit 4. Finally, qubit 1 is transferred to the B wing of the experiment where qubits 1 and 4 interact to verify the standard correlations.

For clarity and brevity, I will describe the evolution of the system in the Schrödinger picture and simply indicate the relevant changes in the descriptors in the Deutsch-Hayden formalism.

We assume that the initial state of the system is

$$|\psi(0)\rangle = |0\rangle_1|0\rangle_2|0\rangle_3|0\rangle_4, \tag{5.7}$$

where the subscripts denote the state of a particular qubit. The descriptors are all in the initial state described in (5.3). At $t = 1$, qubits 2 and 3 are entangled by unitary operation U_E. The state of the system becomes

$$|\psi(1)\rangle = |0\rangle_1 \frac{i}{\sqrt{2}} \left(|0\rangle_2|0\rangle_3 - |1\rangle_2|1\rangle_3\right) |0\rangle_4. \tag{5.8}$$

The descriptors change as

$$\begin{aligned} \hat{\mathbf{q}}_1(1) &= \hat{\mathbf{q}}_1(0) \\ \hat{\mathbf{q}}_2(1) &= U_E^\dagger \hat{\mathbf{q}}_2(0) U_E \\ \hat{\mathbf{q}}_3(1) &= U_E^\dagger \hat{\mathbf{q}}_3(0) U_E \\ \hat{\mathbf{q}}_4(1) &= \hat{\mathbf{q}}_4(0). \end{aligned} \tag{5.9}$$

At $t = 2$ qubit 2 is submitted to a rotation θ about the x-axis, described by U_θ; qubit 3 is submitted to a rotation ϕ about the x-axis, described by U_ϕ. For convenience, we let

$$a = \frac{i}{\sqrt{2}} \left(\cos(\phi/2)\cos(\theta/2) + \sin(\phi/2)\sin(\theta/2)\right)$$

$$b = \frac{1}{\sqrt{2}} \left(\sin(\phi/2)\cos(\theta/2) - \cos(\phi/2)\sin(\theta/2)\right).$$

The state of the system is

$$|\psi(2)\rangle = |0\rangle_1 \left[a\left(|0\rangle_2|0\rangle_3 - |1\rangle_2|1\rangle_3\right) + b\left(|0\rangle_2|1\rangle_3 - |1\rangle_2|0\rangle_3\right)\right] |0\rangle_4 \tag{5.10}$$

$$\begin{aligned} \hat{\mathbf{q}}_1(2) &= \hat{\mathbf{q}}_1(0) \\ \hat{\mathbf{q}}_2(2) &= U_E^\dagger U_\theta^\dagger \hat{\mathbf{q}}_2(0) U_\theta U_E \\ \hat{\mathbf{q}}_3(2) &= U_E^\dagger U_\phi^\dagger \hat{\mathbf{q}}_3(0) U_\phi U_E \\ \hat{\mathbf{q}}_4(2) &= \hat{\mathbf{q}}_4(0). \end{aligned} \tag{5.11}$$

Deutsch and Hayden observe that there is locally inaccessible information about θ in qubit 2, and about ϕ in qubit 3. This is reflected in the above equations by the dependence of the descriptor for qubit 2 on U_θ and the dependence of the descriptor for qubit 3 on U_ϕ. No information about θ is in qubit 3, nor is there information about ϕ in qubit 2. This is reflected in the above equations by the independence of qubit 3 on U_θ and the independence of the descriptor for qubit 2 on U_ϕ.

At the next stage in the experiment, qubits 2 and 3 are measured. This measurement is described completely quantum mechanically. It is performed by a quantum gate, known as the CNOT gate, that performs the unitary evolution

$$|c\rangle|t\rangle \xrightarrow{U_{CNOT}} |c\rangle|c \oplus t\rangle,$$

where $\oplus$ is addition mod 2 and $c, t \in \{0, 1\}$. The evolution above takes the state of a target qubit, described by state $|t\rangle$, and flips it if the control qubit, described by $|c\rangle$, is in the state $|1\rangle$, and leaves unchanged if in the state $|0\rangle$. This is a useful way to model measurement because after the interaction there will be a perfect correlation between the state of the control bit and the state of the target bit. We will use the notation U_{ct} to denote the CNOT operation where c is a variable that ranges over control qubits, and t ranges over target qubits. So, U_{21} denotes a CNOT gate with qubit 2 as the control and 1 as the target. The state of the system after measurements are made on qubits 2 and 3 using qubits 1 and 4 respectively is

$$|\psi(3)\rangle = a(|0000\rangle - |0111\rangle) + b(|1011\rangle - |1101\rangle), \tag{5.12}$$

where we adopt the notation $|0000\rangle = |0\rangle_1|0\rangle_2|0\rangle_3|0\rangle_4$. The descriptors at $t = 3$ are

$$\begin{aligned}
\hat{\mathbf{q}}_1(3) &= U_E^\dagger U_\theta^\dagger U_{21}^\dagger \hat{\mathbf{q}}_1(0) U_{21} U_\theta U_E \\
\hat{\mathbf{q}}_2(3) &= U_E^\dagger U_\theta^\dagger U_{21}^\dagger \hat{\mathbf{q}}_2(0) U_{21} U_\theta U_E \\
\hat{\mathbf{q}}_3(3) &= U_E^\dagger U_\phi^\dagger U_{34}^\dagger \hat{\mathbf{q}}_3(0) U_{34} U_\phi U_E \\
\hat{\mathbf{q}}_4(3) &= U_E^\dagger U_\phi^\dagger U_{34}^\dagger \hat{\mathbf{q}}_4(0) U_{34} U_\phi U_E.
\end{aligned} \tag{5.13}$$

From the above descriptors, we can see that locally inaccessible quantum information about θ has spread via a the local interaction between qubits 1 and 2. Similarly, locally inaccessible information about ϕ has spread from qubit 3 to 4.

In the final stage of the experiment, qubit 4 acts as a control bit for qubit 1. The state becomes

$$|\psi(4)\rangle = a(|0000\rangle - |1111\rangle) + b(|0011\rangle - |0101\rangle). \tag{5.14}$$

The descriptors become

$$\begin{aligned}
\hat{\mathbf{q}}_1(4) &= U_E^\dagger U_\theta^\dagger U_\phi^\dagger U_{21}^\dagger U_{41}^\dagger \hat{\mathbf{q}}_1(0) U_{41} U_{21} U_\phi U_\theta U_E \\
\hat{\mathbf{q}}_2(4) &= U_E^\dagger U_\theta^\dagger U_{21}^\dagger \hat{\mathbf{q}}_2(0) U_{21} U_\theta U_E \\
\hat{\mathbf{q}}_3(4) &= U_E^\dagger U_\phi^\dagger U_{34}^\dagger \hat{\mathbf{q}}_3(0) U_{34} U_\phi U_E \\
\hat{\mathbf{q}}_4(4) &= U_E^\dagger U_\phi^\dagger U_{34}^\dagger U_{41}^\dagger \hat{\mathbf{q}}_4(0) U_{41} U_{34} U_\phi U_E.
\end{aligned} \tag{5.15}$$

Deutsch and Hayden emphasize that the information about what happens in the B wing of the experiment, the rotation of qubit 3 by ϕ, travels to the A wing in virtue of local interactions of qubit 4 with qubit 1, although it is locally inaccessible while in transit.

As I pointed out in my (1999), Deutsch and Hayden are not explicit as to whether their analysis of the EPR experiments (and teleportation too) are meant to simply to *describe* information transfer in quantum systems or whether they want to *explain* quantum phenomenon in terms of the transfer of information. As was discussed in chapter 1, the suitability of transfer explanations is limited with respect to information transfer. We will discuss both EPR experiments and teleportation with respect to the conservative and ontological interpretations of the Deutsch and Hayden formalism.

No local account of information transfer can be given in the conservative interpretation. The locality properties claimed for information are dependent on the concept of locally inaccessible information which cannot be supported by the interpretation of the formalism. In the above experiment, information will be distributed across the EPR pair, or with systems that interact with the pair. Additionally, since there is no local quantity that is transferred from one wing of the EPR experiment to the other, there is no way for Deutsch and Hayden to successfully employ a transfer explanation of the correlations.

The ontological interpretation fares far better than the conservative interpretation. In the ontological interpretation, locally inaccessible information is either correlated with or is identified with the properties identified by the individual descriptors for quantum systems, Deutsch and Hayden do not specify which. Locally inaccessible information, whose ontological status is underwritten by the local absolute properties, is carried by qubits that are transferred form one wing to the other. Deutsch and Hayden do get a purely local account of information transfer from one wing of the EPR experiment to the other. Note that whether information is an absolute or relational property of the system, the locality properties of information transfer that Deutsch and Hayden desire are still available.

The analysis of EPR correlations above allows us to say more about whether quantum information is an absolute or relational property of a system. As the above analysis demonstrates, locally inaccessible information spreads like a nasty virus. The slightest bit of entanglement is enough to pass it to any system, detectable through the appearance of θ in the descriptor of a system. This fact seems to eliminate the possibility of regarding quantum information as an absolute property of a system. For example, qubits 1 and 2 in the EPR experiment both contain locally inaccessible information about θ, but the descriptors for these qubits are different. If local properties of qubits are in one to one correspondence to descriptors, and the descriptors identify unique quantum information as an absolute property of the system, then the qubits cannot share the same quantum information. Of course, claiming that the qubits share the same quantum information is exactly what Deutsch and Hayden would like to do. This seems possible only if we regard the qubits as containing locally inaccessible information about θ in virtue their local properties being correlated with θ, i.e. we take quantum information to be a relational property. So, we can no longer entertain the identification of quantum information about a parameter θ directly with the properties identified by the descriptors.

If quantum information is relational, then quantum information is not a quantity that is transferred in the course of the EPR experiments. So, transfer explanations that depend on quantum information will not work. Instead, the EPR experiment information is transferred in virtue of progressive correlations being established in the absolute individual properties of systems. What Deutsch and Hayden seem to have available to them, but do not seem to realize, is that instead of a transfer explanation of EPR correlations, a local causal explanation of the correlations in terms of the properties identified by the descriptors of quantum systems is possible. The explanation of the correlations on the ontological interpretation would be as follows. The rotations on qubits 2 and 3 change the local properties of qubits 2 and 3. The measurement interactions correlate the properties of qubit 1 to qubit 2, and qubit 4 to qubit 3. The correlations that we get when qubit 1 and 4 interact are not mysterious at all, because systems with occurrent non-relational properties individually correlated to θ and ϕ are brought together and subjected to an interaction sensitive to these underlying properties. That fact that the correlations occur is no more mysterious than those correlations that occur when putting two halves of a piece of newspaper together.

What remains unexplained and fundamentally weird are these properties of qubits on the ontological interpretation that have heretofore gone unnoticed. Nonetheless, there are several things that can be said about these properties. First, these properties are not values of observables, but they do give rise to values of observables. We remind the reader that "values of observables" takes its meaning in the context of no-collapse quantum mechanics, whereby every possible value of an observable is occurrent if the system is not in an eigenstate of the observable. Following that, we note that these properties give rise to the discrete spectrum of values for certain observables. These special properties of a system give rise to all observables for that system but, these properties cannot always be detected locally. Sometimes, i.e. in the presence of entanglement, in virtue of some past interaction with another system, certain local system properties become locally unobservable. These local properties of a system have observable consequences only when they interact with other systems. This feature is the essence of Deutsch and Hayden's claim that, "All phenomena that have been thought to demonstrate nonlocality in quantum physics are actually due to the phenomenon of locally inaccessible information (*ibid.* 19)." The question that naturally arises when confronted with these properties is, is a local story about correlations so desirable that we are willing to introduce such strange non-classical properties?

5.5.2 Quantum teleportation

In this section we apply Deutsch and Hayden's analysis to quantum teleportation. As with the EPR experiments, it is not clear whether Deutsch and Hayden intend only to give an analysis of information flow in teleportation of if they want to explain teleportation in terms of information flow. We will treat both cases. Not surprisingly, much of what was said about EPR experiments will apply directly to teleportation.

Teleportation has been described by some as a example of nonlocal information transfer. A successful teleportation experiment transfers an unknown quantum state from one qubit to another without a direct interaction between these qubits. A teleportation experiment begins with an unknown quantum state prepared. This state interacts via a Bell operator measurement with one member of a fully entangled pair. The results of the measurement reveal which of four possible unitary transformations can be applied to the other member of the EPR pair, the one that did not interact with the unknown state, to ensure successful teleportation. When these unitary transformations are applied to the target system conditional on the results of the Bell operator measurement, the unknown state is transferred. (See Figure 5.)

At this point, we assume that the reader has a feel for the contiguity property of descriptors. So, rather than slog through the formal details of the experiment it suffices for our purposes to qualitatively describe the whereabouts of information according the Deutsch and Hayden formalism. As in the EPR case, the teleportation experiment begins with qubits in the standard state, $|00000\rangle$. At t=0, preparation of an unknown state is simulated by rotating qubit 1 by an angle θ about the x-axis. Qubits 4 and 5 are entangled. At $t = 1$, information about θ resides in qubit 1. Qubits 1 and 4 are submitted to measurement of the Bell operator. Qubits 1 and 4 are measured correlated to qubits 2 and 3, effectively storing the results of the measurement in qubits 2 and 3. At $t = 3$, the descriptors for qubits 2 and 3 contain information about θ, but that information is *locally inaccessible*. These qubits are brought to the other wing of the experiment. At $t = 4$ a unitary transformation, T, dependent on the measurement results stored in qubits 2 and 3 is performed on qubit 5 that correlates the descriptor for qubit 5 with the initial unknown state of qubit 1. Qubit 5 finally stores information about θ.

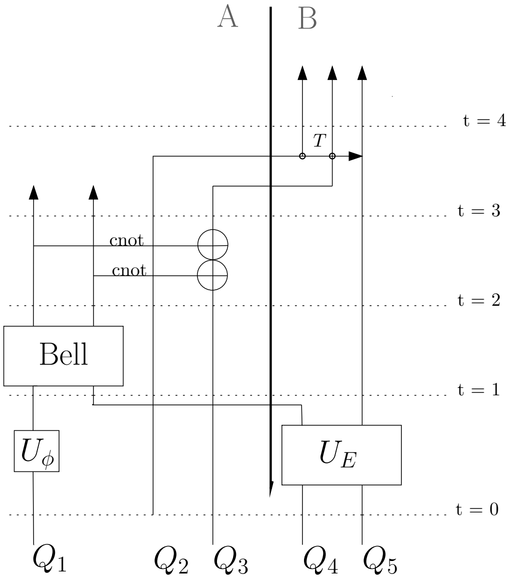

Figure 5: Teleportation. Initially, qubit 1 is submitted to a rotation θ to simulate the preparation of an unknown state. Also, qubits 4 and 5 are entangled via U_E. Qubit 4 is then transferred to the A wing of the experiment where it, along with qubit 1 is submitted to a measurement of the Bell operator. The results of the measurement are then stored in qubits 2 and 3 in virtue of CNOT interactions between 1 and 2 and 4 and 3 respectively. Finally, qubits 2 and 3 which store the locally inaccessible information about θ are transferred to the B wing of the experiment interact with qubit 5 to transfer the information about θ.

Since locally inaccessible information is not a concept suited to the conservative interpretation, the teleportation experiment cannot be interpreted as local transfer of quantum information on that interpretation. After the Bell operator measurement, because of entanglement, information about the unknown state will be contained in the relational properties of the system, not local properties. It is only the *descriptors* for qubits 2 and 3 that are correlated with θ, and not the systems themselves. Following Braunstein (1996), who examines a no-collapse version of teleportation, only qubits 2, 3, and 5 can be said to contain information about θ. So, on the conservative interpretation, Deutsch and Hayden's analysis is misguided and gets the locality properties of information wrong. A transfer explanation is not possible given that only the problematic locally inaccessible information is transferred from one wing of the experiment to the other.

The ontological interpretation of the Deutsch and Hayden approach is again far more promising than the conservative interpretation. Information is stored in a system in virtue of a correlation of the properties of a system to a parameter. Due to the contiguity properties of the descriptors, descriptors can only pick up dependence on other systems through local interactions. So, information transfer in teleportation will be local.

Just like in the EPR case, the ontological interpretation seems to be incapable of a transfer explanation of teleportation. A transfer explanation would seemingly have to depend on the transfer of locally inaccessible information. Locally inaccessible information is not a conserved quantity, but is a relational property of a system. Locally inaccessible information can be identified with the Shannon information, and that alone cannot support a transfer explanation.

What the ontological interpretation can do is give a local causal explanation of teleportation. That a system can become correlated with another via progressive interactions that correlate properties of systems to one another is no surprise. Here an analogy seems apt. On the ontological interpretation, teleportation is much like a phone call. A speaker's sound waves are like the unknown state. The telephone interacts with the sound waves to produce a signal suitable for transfer over electric wires. Similarly, an entangled qubit interacts with the unknown state to produce a signal suitable for transfer. A telephone transmits a message in virtue of the properties of the electric signal being correlated with the speaker's sound waves. The two qubits that make the trip from the A wing of the experiment to the B wing transfer the message in virtue of their actual properties being correlated with the measurement of the unknown state. Finally, with a phone, the electrical signal interacts with a system capable of transforming the signal into sound waves. With teleportation, the qubits whose properties are correlated to the unknown qubit must interact with another system suitable to allow the information about the unknown state to become accessible. What could be less mysterious? Well, the crucial disanalogy is that the electric signal can be used in conjunction with myriad systems to extract the message. The qubits that are transferred in teleportation can only interact with a very particular system for the information their properties are correlated with ever to become accessible. This strange feature of the underlying properties of quantum systems, Deutsch and Hayden would have to say, is exactly what makes them quantum mechanical, and not explicable by a conventional hidden variable theories.

5.6 CONCLUSION

Deutsch and Hayden wrote a very ambitious paper meant to banish nonlocality from quantum physics. We must be careful to distinguish between the different interpretations that the Deutsch-Hayden formalism can be given. We must further distinguish between claims of the locality broadly construed, and claims to the local nature of information. The benefits that the Deutsch-Hayden formalism enjoys are crucially dependent on these distinctions.

Regarding nonlocality, we have distinguished nonlocality in the sense of state vector collapse, and nonlocality in the sense of quantum holism. By adopting no-collapse quantum mechanics, Deutsch and Hayden trivially banish the first sense of nonlocality. We note that this is no special advantage due to the formalism they develop, but a shared one with other no-collapse interpretations, Everett, statistical, etc. On the other hand, quantum holism, the other sense of nonlocality, is not so easily dealt with. To deal with this, we must choose an interpretation for the formalism.

Earlier in this paper we distinguished two types of quantum holism: that the state of a quantum system cannot be determined by the state of its subsystems, or that changes in a local system change the global state in non-classical ways. Both of these types of holism can be subsumed into one notion because they are due to the fact that on the typical understanding of quantum mechanics, the individual properties of subsystems do not determine the properties of the total system. Do Deutsch and Hayden develop a quantum theory free of this feature? Yes and no. With respect to the Deutsch and Hayden formalism, it is clear that local interactions of systems change only the local descriptors implicated in the interaction. However, interpreted conservatively, the properties of a quantum system are determined not only by the descriptors of a system, but also the state. On this interpretation, quantum theory still has holism. The properties of the total system cannot generally be reduced to properties of subsystems. Moreover, locally inaccessible information is not an applicable concept because there are no properties that it corresponds to. As we have seen, Deutsch and Hayden's analysis of information transfer is simply wrong, and there is no hope of explaining quantum phenomena based information transfer.

The situation is quite different on the ontological interpretation of the Deutsch-Hayden formalism. Local properties of quantum systems do determine the global properties of the system. Holism is thereby vanquished. In its place, we have local properties of systems that are unobservable locally. These properties allow for a purely local account of information transfer in quantum systems. It was argued that inaccessible information must be understood as a relational property of quantum systems as opposed to an absolute property. Locally inaccessible information has properties that coincide exactly with the properties of Shannon information in quantum systems. Thus, transfer explanations that invoke this type of information are not possible. Instead of transfer explanations, it is possible to offer a local causal explanation of quantum phenomena based on the local properties of individual quantum systems, but note that these explanations are not based on a concept of quantum information.

One way to evaluate the status of a new interpretation of quantum mechanics, or perhaps more appropriately, a new quantum theory, in the case of the Deutsch-Hayden approach interpreted ontologically, is to determine how well it explains physical effects. Deutsch and Hayden seem to have found the holy grail of quantum theory that philosophers of physics have been seeking for so long: an account of quantum mechanics that allows for local causal explanations that do not have the slightest hint of conflict with the special theory of relativity, as other interpretations or quantum theories do. However, this kind of explanation comes at a price. We must inflate our ontology with a set of properties which are unlike any properties we are familiar with, are unobservable in principle, and underdetermined to boot. Given the sad history of tinkering with metaphysics to satisfy *our* predilections about the way the world works, we should be skeptical of Deutsch and Hayden's approach.

6.0 INTRODUCTION PART II

In 1985 David Deutsch provided an example of a computation that a quantum computer could perform faster than a classical computer.[1] Since Deutsch published his result, several new algorithms using quantum mechanical effects have been discovered that offer exponential speedup over existing classical algorithms. Perhaps the most significant is Shor's (1994) algorithm for factoring numbers in polynomial time. The best classical algorithm for factoring uses exponential time. The obvious question is, why are quantum computers faster than classical computers?

There are different levels of answers to such a question. One could attempt to identify types of quantum mechanical effects that are utilized in quantum algorithms that are not available for classical algorithms. The importance of such a task is obvious. If one can develop a list of quantum processes that perform tasks faster than classical computation processes, presumably new quantum algorithms that maximize the use of these processes will be more efficient than classical ones. This question can be answered on a more philosophical level as well. We can attempt to specify which features of the real world, the quantum world, allow quantum speedups to take place. Presumably, such a task is what Jozsa is referring to when he writes the following:

> ...the existence of interesting quantum algorithms, merely at the level of theoretical constructs, is of great value in itself as it points to new and essential differences between the fundamental structure of classical physics compared to quantum physics(Jozsa 2000, 105).

In this part of the dissertation, I examine a process called *quantum parallelism.* Deutsch and others (Jozsa, for instance) have singled this process out as a key feature of quantum

[1] This claim should be understood as made relative to a particular and perhaps restrictive model of computing.

algorithms that seemingly cannot be simulated efficiently by a classical computer. It has been suggested that the reason why the quantum parallelism process is of great importance in quantum computing is that it allows one to compute all values of a function in a single computational step. Let us call this the *quantum parallelism thesis*. The quantum mechanics of the process certainly seems compatible with the quantum parallelism thesis.

Let us consider a system of $n + m$ qubits. Let n qubits be the input register for our quantum computer and m qubits be the output register. Each qubit can represent a single binary number, 0 or 1. Our n qubit input register can represent any integer from 0 to $2^n - 1$. Further, our input register can be put into a superposition of all quantum states representing all of the integers from 0 to $2^n - 1$. For convenience, let us adopt the following notation: $|x\rangle$ will stand for the n qubit state representing the number x in binary. Suppose that $n = 2$. The number three would then be represented by the state $|3\rangle = |1\rangle|1\rangle$. With this notation, an n qubit system in state ψ that is a superposition of all numbers from 0 to $2^n - 1$, ignoring a normalization factor, is given by $|\psi\rangle = \sum_{x=0}^{2^n-1} |x\rangle$. Suppose we have a unitary gate, U_f, that performs the following evolution:

$$|x\rangle|y\rangle \xrightarrow{U_f} |x\rangle|f(x) \oplus y\rangle \tag{6.1}$$

where $f : \{0, \ldots, 2^n - 1\} \to \{0, 1\}^m$ and $\oplus$ is addition mod 2. Using the linearity of unitary evolution and the fact that quantum systems can be put into superpositions, it appears that all values of a function can be computed in one pass of a unitary gate if we put the input register into a superposition of all points in the domain of the function:

$$\sum_{x=0}^{2^n-1} |x\rangle|0\rangle \xrightarrow{U_f} \sum_{x=0}^{2^n-1} |x\rangle|f(x)\rangle. \tag{6.2}$$

Deutsch has suggested that the only interpretation of quantum mechanics that can support the quantum parallelism thesis is the many worlds interpretation. Each term in the initial superposition corresponds to a different world. The state vector describing the quantum mechanics of each world corresponds to a single value of the function to be computed. In each world, the quantum system representing the point of the function to be computed is then submitted to a quantum gate that computes each value of the function. After the quantum system passes through the gate, in each world, the value of the function corresponding to the point that identifies the world will have been computed. All values of the function are computed in virtue of all individual values being computed in different worlds.

Contrary to Deutsch, in “A quantum computer only needs one universe,” Andrew Steane argues that our understanding of quantum computers is not wedded to a many-worlds interpretation of quantum theory. Rather than attack the many-worlds interpretation directly, Steane's method is to argue that the quantum parallelism thesis is false. If the thesis is false, the many-worlds interpretation's special status with respect to quantum computing is undermined. Put another way, if it is *not* the case that the quantum parallelism process is a computation of many values of a function simultaneously, then the many-worlds interpretation enjoys no special status by underwriting the view that quantum computers *do* compute many values of a function simultaneously.

One of the key arguments Steane uses to undermine the quantum parallelism thesis is demonstrating that mathematical notation can be misleading. Steane gives an example of a trivial computational device whose mathematical description gives the impression that it computes all values of a function efficiently, but the device has the following two features:

1. The efficiency of the computation seemed to be due to the interpretation of the states of the computational device, rather than due to the intrinsic efficiency of the computation.
2. The computational device computes all values of the function, but the structure of the device precludes interpreting the evolution of the device as a computation of all values of the function simultaneously.

These two features raise questions. First, when is a machine performing a genuine computation? Second, given that a genuine computational process occurs and the machine outputs all values of a function in a single step, must this be as a result of doing all of the individual computations? In this part of the dissertation I address both questions.

The key issue in deciding the quantum parallelism thesis is deciding when multiple values of a function are computed. There is already a well known criterion for when the value of a function is computed, the Turing machine criterion. If a Turing machine halts on input x with $f(x)$ as its output, the value of the function at x has been computed. Of course, the problem with applying the Turing machine criterion to decide the quantum parallelism thesis is that we are not dealing with Turing machines. In this part of the dissertation, I develop a set of general criteria that any computational device must satisfy in order to have evaluated a function at multiple points. These set of general criteria are developed by abstracting from the features of the Turing machine criterion that make it a good criterion for a function to be evaluated in the first place.

If one adopts an objective view of the quantum state, it turns out that quantum parallelism satisfies the general criteria for multiple values of a function to be computed. This obviates appeals to interpretations of quantum mechanics to underwrite the quantum parallelism thesis. The many-worlds interpretation thus loses its favored status regarding quantum computation. Moreover, I will argue that at least one version of the many-worlds interpretation makes it more difficult to understand how the quantum parallelism thesis can be true.

The answer to the second question, “given that a genuine computational process occurs and all values of a function are computed simultaneously, must this be as a result of doing all of the individual computations?”, can be arrived at by a straight forward application of the general criteria for computation. In fact, the general criteria support at least three distinct concepts of parallel computation, where *parallel computation* here refers to multiple values of a function computed simultaneously. I offer three very different examples of computers that satisfy the general criteria for a function to be computed, but compute in fundamentally different ways. As we will see, one can compute in parallel without computing each value of the function individually. As such, our basic understanding of parallelism will be fundamentally revised.

As we will see, there are examples of classical computers that compute all values of a function in a single step, much like quantum computers that make use of quantum parallelism. Nonetheless, these computers appear to be hopelessly inefficient when we compare the spatial and energetic resources required for computation to the resources required by quantum computers. What this indicates is that the truth of the quantum parallelism thesis alone does not provide a complete explanation of quantum computational efficiency. I argue that quantum computers are efficient in virtue of them satisfying the following requirement:

> The physical dependencies of the computer mirror the logical dependencies of the compu-

> tational task.

I argue that a computer will be as efficient as a quantum computer at computing multiple values of a function if this requirement is met. In this sense, I offer an explanation of the special power of quantum computing.

The goal of this part of the dissertation is to argue that the quantum parallelism thesis is true, that the many-worlds interpretation has no special advantage with respect to quantum parallelism, and to provide an explanation of why quantum computers are faster than classical computers. To that effect, in chapter 7 I develop a set of general criteria for a computation of multiple values of a function. Using those criteria I argue that the quantum parallelism thesis is true. I will argue that the general criteria allow for at least three distinct notions of parallelism, and that quantum parallelism represents one of these distinct notions. In chapter 8 I provide an explanation of the special efficiencies of quantum computers that utilize quantum computation. In chapter 9 I examine Steane's arguments that the quantum parallelism thesis is false. I will show that none are conclusive. Also, I will demonstrate that the general criteria rule out possibly problematic computational devices Steane uses to undermine the quantum parallelism thesis. Finally, in chapter 10 I point out difficulties that one of the many-worlds interpretation has underwriting the quantum parallelism process.

7.0 GENERAL CRITERIA FOR COMPUTATION AND PARALLELISM

7.1 INTRODUCTION

When one examines the quantum parallelism process, there is at least an intuitive feeling that all values of the function are calculated. The quantum state, were it known, would indicate all values of the function. Intuition is no way to decide the matter, and in this chapter we turn to an objective criterion for a function to be evaluated: the Turing machine criterion. Roughly, the criterion is: If a Turing machine halts on input x with $f(x)$ as its output, the value of the function at x has been evaluated.[1]

The use of such a criterion to decide the quantum parallelism thesis faces obvious problems. First, the Turing machine criterion applies only to Turing machines, and not other models of computation. Second, the Turing machine criterion only addresses the case of the evaluation of a function at individual points of its domain. Both problems beg for an extension of the criterion to encompass all potential computing devices, and also to multiple values. In this chapter, I will abstract from the Turing machine criterion to develop a set of general criteria that any computational device must satisfy in order to compute multiple values of the function. For the purposes of this paper, I assume that the these general features will form a set of criteria that can be applied to any model of computation or any physical system to decide if points from a function's domain were computed. It will turn out that on the general criteria, the quantum parallelism thesis is true.

It is straightforward to demonstrate, using the general criteria, that there are several different concepts of parallelism that fit the description of many values computed in a single temporal step. In this chapter I will distinguish three concepts of parallelism: tube paral-

[1] In what follows, I will use the words "evaluated", "computed", and "calculated" interchangeably.

lelism, classical parallelism, and quantum parallelism. I will argue that these concepts of parallelism have important ramifications for explanations of quantum speedup. In particular, to explain the computation of multiple values of a function, no story need be told about how each individual value of the function was computed in the process.

In section 7.2 I discuss Turing machines and the criteria for computation that goes alone with them. In section 7.3 I develop a set of general criteria for multiple values of a function to be computed. In section 7.4 I introduce a model of computation that computes multiple values of a function in a single computational step according to the general criteria, but does the computation in a radically different way than computers that use classical or quantum parallelism. I use this model to help distinguish three concepts of parallelism in section 7.5. Finally, I discuss the ramifications of these three concepts of parallelism for explanations of quantum speedup.

7.2 TURING MACHINES AND COMPUTATION

We briefly discuss what a Turing machine is in order to understand why the Turing machine criterion is a good criterion for a function to be computed.

The following characterization of a Turing machine parallels Nielsen and Chuang (2000), pp.122-124. A Turing machine consists of a program, a finite processor control, a two way infinite tape, and a read/write tape head. The finite processor control consists of a countable number of machine states. The processor control has a starting state, which is the initial state of the machine before a computation, and a halting state, which is the final state after the computation. The tape consists of cells each indexed by an integer. The tape head can read the contents of the tape at a particular cell, move left, right, or not at all, and write a symbol on a tape. We assume that possible tape configurations always contain a finite number of nonblank cells. A program for a Turing machine is a set of instructions that determine the progress of a Turing machine. The instructions are of the form $\langle q, x, q', x', d\rangle$, where q and q' are processor configurations, x and x' are the symbols from the machine alphabet, and $d \in \{-1, 0, 1\}$ indicates the direction the tape head will move.

If the machine is in the state q with the symbol x under the read/write head, it will search out instructions with q and the symbol x in the leftmost two slots of the program instructions. The machine updates its state, writes a symbol and moves according to the right slots of the instructions. For example, if the machine is in the state q and the symbol x is under the read/write head and $\langle q, x, q', x', d \rangle$ is part of the program, the machine will write symbol x', update the processor configuration to q', and move the tape head along the tape according to d. A Turing machine operates by beginning operation in the initial configuration, scanning the tape, updating the processor configuration, and writing on the tape according to the program instructions until it is sent to its final or halting configuration, upon which the machine halts. The output of the Turing machine is the contents of its tape.

Let Σ denote the set of symbols available to a Turing machine, the alphabet for the machine. Let Σ^* denote the set of tape configurations (the set of tapes with a countable number of symbols from Σ written on the tape.) A Turing machine then always computes a function $g : \Sigma^* \rightarrow \Sigma^*$. So, Σ^* is the domain of g and some subset of Σ^* will be the range of g. Let us refer to g as the *machine function.* If $x \in \Sigma^*$ is the initial tape configuration, we refer to x as the input string to the Turing machine. Similarly, we denote the output string of the Turing machine by $g(x)$. Suppose that Σ is $\{0, 1, b\}$ where b is the blank symbol. We might be interested in computing functions $f : \{0, 1\}^n \rightarrow \{0, 1\}^m$. In this case, we want to choose the machine function g so that on input strings with sequences of symbols from $\{0, 1\}^n$ with the rest of the tape cells blank, the machine outputs a sequence of symbols from $\{0, 1\}^m$ with the rest of the tape cells blank. In other words, if we examine only the first n cells of the input, and only the first m cells of the output, $g(x) = f(x)$, i.e. the function f is a partial function of the machine function. In this case, we say that the Turing machine halts on input string x with string $f(x)$ as output, the function has been computed at x.

We can summarize the Turing machine criteria for a point of a function to be computed:

> For a Turing machine $\mathcal{M}$ with machine function g, $\mathcal{M}$ has computed the value of function f at x if and only if
>
> 1. f is a partial function of g such that $f(x) = g(x)$.[2]
> 2. On input string x, $\mathcal{M}$ has halted with output string $f(x)$.

[2] We suppress the qualifications about which cells to look at, etc. , detailed above.

We note several features of the Turing machine criteria that make it a good set of criteria for the computation of a function at a point. The functional behavior of the machine is such that it can be interpreted as beginning a computation with the tape correlated to the desired value of the function to be computed. The machine is deterministic and computes the appropriate value for any input for the domain of the function computed. We take these features to be constitutive of deterministic computation of a function. So, they must be preserved by any deterministic computational device.

7.3 GENERALIZED COMPUTATION

In this section a set of general criteria for multiple points of a function to be computed are developed that apply to any computation device. The criteria developed in this section preserve the essential features of the Turing machine criteria.

To compute the value of a function, two concepts are essential: the input and the output. For a Turing machine that computes the value of a function at a single point, the input is a variable over the domain of the function to be computed, and the output is a variable over the range of the function to be computed. For a computational device that evaluates multiple values of a function, the situation is slightly more sophisticated. For such devices, the input is a variable over the power set of the domain of the function to be computed (minus the empty set), $\mathcal{P}(dom(f)) - \emptyset$. The input indicates what value or values of the function are to be evaluated. The output is a variable over all possible partial graphs given the domain and range of the function to be computed. A graph of a function, $G(f)$ is the set of ordered pairs indicating points in the domain of the function and the value of those points for the function. A partial graph of a function, $G(f|_X)$ is a graph restricted to a subset, X of the domain of the function. More formally,

$G(f)$: $\{< x, y >: y = f(x)\}$, the graph of f.

$G(f|_X)$: $\{< x, y >: y = f(x) \,\&\, x \in X \,\&\, X \subseteq dom(f)\}$, the partial graph of f on X.

Note that the input and output of a computer are abstract notions that do not correspond to anything physical.

In order to actually perform a computation, there must be some way of representing the values of the input and output in physical systems. Let us call the systems that accomplish such a feat the *input system* and *output system*. One can successfully encode a value of the input variable into the input system by preparing the state of the input system such that it can be interpreted as representing the input points, a subset of the domain of the function to be computed. In order to represent all possible inputs there must be a distinct input state for every possible input. For example, if the domain of a function is $\{0, 1\}$, a successful encoding procedure would prepare a input system in one of three distinct states that are perfectly correlated with choice of an input values $\{0\}$, $\{1\}$, or $\{0, 1\}$. Similarly, a successful decoding procedure allows distinct physical states of the output system to be interpreted as representing distinct outputs. For example, consider a function $f : \{0, 1\} \rightarrow \{0, 1\}$. The output state must be capable of representing all possible combinations of input output pairs. In this case the output state must be able to represent $\{< 0, 0 >\}$, $\{< 0, 1 >\}$, $\{< 1, 0 >\}$, $\{< 1, 1 >\}$, $\{< 0, 0 >, < 1, 0 >\}$, $\{< 0, 0 >, < 1, 1 >\}$, $\{< 0, 1 >, < 1, 0 >\}$, and finally, $\{< 0, 1 >, < 1, 1 >\}$. A distinct state of the output system is required for each of these eight outputs.

We refer to an assignment of inputs to input states and output states to outputs as an *interpretation* of input states and output states. An interpretation is a pair of mappings (α, β). The mapping α is a one-one mapping from inputs into input states. We require the mapping to be one-one because distinct inputs must be represented by distinct input states. The mapping β is mapping from output states onto outputs. The mapping must be onto because all possible output states need to be interpreted as an evaluation of the function for some set of inputs.[3] All of this is straight-forward on a Turing machine. Inputs are encoded on distinct tape states, and similar for outputs. Typically, for a Turing machine, there is such an obvious interpretation of the input and output tape states that it is overlooked that we are interpreting those tape states as indicating inputs and outputs.

However obvious an interpretation of input states and output states may be, an interpretation is a crucial to the evaluation of a function. It is an interpretation which ultimately

[3] One might object that there needs to be an output state that indicates error, and the restriction that the mapping β be onto precludes this possibility. For the purposes of this paper, we consider only error free computation.

determines which function a machine computes. Consider a machine computes a function $f : \{0, 1\} \to \{0, 1\}$. As discussed above, such a machine needs three input states and eight output states. Suppose the input states are A, B, and C. Fix an interpretation of the input states so 0 is encoded by state A, 1 by B, and 0 and 1 by C. Suppose the output states are AX, BX, AY, BY, CXX, CYY, CXY, CYX. Suppose that we interpret the output states such that AX gets mapped to $\{0, X\}$, and similar for AY, BX, and BY. Suppose we interpret the output states such that CXY gets mapped to $\{< 0, X >, < 1, Y >\}$ and similar for CXX, CYY, and CYX. The interpretation of output states is completely fixed by assigning the values 0 and 1 to X and Y. We have the following two interpretations:

Interpretation 1. X stands for 0 and Y for 1.
Interpretation 2. X stands for 1 and Y for 0.

On interpretation 1., the output state AX would indicate that $f(0) = 0$, the output state CXY would indicate $f(0) = 0$ and $f(1) = 1$, etc.

Suppose that the computational device discussed above evolves from input state A to AX, from B to BY, and from C to CXY. On interpretation 1. this device computes the binary function $f(x) = x$. On interpretation 2. this machine computes the binary function $f(x) = x \oplus 1$ where $\oplus$ is addition mod 2. This simple example indicates that an interpretation of input and output states will determine what particular function a machine computes. So, an interpretation of input and output states is crucial for the evaluation of a function.

Successful computation occurs when input states are perfectly correlated with output states by the computational process such that the interpretation given to those output states indicates the correct values of the function at the points represented by the input state. For example, supposing that the input state represents the points x_1 and x_2, the output system must be in a state that represents the values of the function and the points at which they were evaluated at, $x_1; f(x_1)$ and $x_2; f(x_2)$. We can make this more precise.

Let $X \subseteq dom(f)$. Suppose that U describes the evolution of the computational device. In particular, let U be the mapping from the set of input states to the set of output states. For successful computation, we require that

$$\beta \circ U \circ \alpha(X) = G(f|_X).$$

Finally we are in a position to specify necessary and sufficient criteria for a function to be evaluated *on a computing device that computes multiple values of a function.*

The General Criteria:

A function f was evaluated at points $X \subseteq dom(f)$ by computing device $\mathcal{M}$ if and only if

1. An interpretation, (α, β), for a function f and the computing device $\mathcal{M}$ exists.
2. The computational device $\mathcal{M}$ instantiates correlations necessary to indicate the correct values of f according to the interpretation (α, β), i.e. if $\forall X \subseteq dom(f)$, $\beta \circ U \circ \alpha(X) = G(f|_X)$.
3. An input state was entered into device $\mathcal{M}$ associated with the input points $X \subseteq dom(f)$ according to α and the computational process of device $\mathcal{M}$ resulted in an output state that is associated with the output $G(f|_X)$ according to β.

The existence of an interpretation automatically implies that there are sufficient input and output states to represent all points of the domain of the function f and corresponding values of f at those points. The mapping α is required to be one-one and this can only be satisfied if the cardinality of the set of input states is greater than or equal to the set of inputs. The mapping β must be onto the set of outputs. So, there will exist a unique state of the output system for every possible output. All possible outcomes of function evaluations can be represented. One might ask why the computational device must be capable of representing all possible partial graphs compatible with a given domain and range, since the computer will only produce a set of outputs with the same cardinality as the set of inputs. If an interpretation is fixed that is limited to the actual partial graphs of the function, then knowing the interpretation renders computation redundant.

With regard to item two, were the condition not satisfied, the device $\mathcal{M}$ would indicate the incorrect values of the function f for some values of the function. In this case, the device would at best be computing a partial function of f, determined by which values the device could compute directly.

Item three is necessary to avoid the following type of counterexample. We can imagine a computing device freshly off the assembly line that accidentally had input and output states that corresponded to the correct evaluation of function points. In this case, there was no process that generated the output states of the machine.

An obvious objection to the general criteria is that no mention is made whether the output of a computational device is required to be accessible. The question can be raised, if that requirement is not made, in what sense has a computational device evaluated a function? If a computational device satisfies the general criteria, the input and output states of the device a correlated in just the right way so the *functional behavior* of the device perfectly mirrors the logical behavior of a function as a mapping between the domain and range of the function. Ontologically, the computational device will have done its job. If it turns out that there are epistemic limitations to determining the states of the physical device that implements a computation, this might be inconvenient, but that fact has no bearing on the functional behavior of the computational device. As a thought experiment we might imagine a Turing machine computing a computable function while it crosses the event horizon of a black hole. In this case, there is no doubt that a function will be evaluated, but in principle, there is no way that the results of the computation can be discovered. Ontologically, the Turing machine performed the calculation, but epistemically, it did not do its job. So long as one subscribes to a functional view of computational devices, the accessibility of results of computational processes has no bearing on whether a function has been computed.

7.3.1 Quantum parallelism and the general criteria

In this section, it is demonstrated that the process of quantum parallelism meets the general criteria and hence counts as an evaluation of a function at multiple points. We bear in mind that the accessibility of results, has no bearing on whether a computational device really computes the values of functions.

Recall that quantum parallelism is used to compute a function $f : \{0,1\}^n \longrightarrow \{0,1\}^m$. The computation involves $n+m$ qubits. The qubits are described in the computational basis which is denoted by $|0\rangle$ and $|1\rangle$ for each qubit. The input system consists of n qubits. The mapping from inputs to input states is straight forward. For every member of the domain of f, set the n-length sequence of qubits state according to the binary number indicating the choice from f. For example, the input point given by the n-length sequence $000\ldots11$ would be represented by the state $|0\rangle|0\rangle\ldots|1\rangle|1\rangle$. For brevity, we specify the state corresponding to the base ten number corresponding to the binary number it represents. So, $|3\rangle = |0\rangle|0\rangle\ldots|1\rangle|1\rangle$. Inputs corresponding to multiple points from the domain of f are mapped to the quantum mechanical superposition of the input states corresponding to the members of the domain of f the input represents. The input (0 and 1) gets mapped to the input state $|0\rangle + |1\rangle$. This mapping is obviously one-one.

The output system consists of $n+m$ qubits. For clarity, the subscript i indicates the state of an n-length sequence of qubits according to the notation above. The output states are all possible states of the form $|x\rangle|y\rangle$, where $0 \leq x \leq 2^n - 1$ and $0 \leq y \leq 2^m - 1$ and all superpositions of these states that have equal coefficients for each term. Define β by the following procedure. Map the states given by $|x\rangle|y\rangle$ to $\{< x, y >\}$ where $0 \leq x \leq 2^n - 1$ and $y \in \{0, 1\}$. Similarly, map superpositions of states, $|x\rangle|y\rangle + |x'\rangle|y'\rangle + \ldots$ to the output $\{< x, y >, < x', y' >, \ldots\}$. There is a state for every possible output, so the mapping is onto. So, an interpretation for quantum parallelism exists and criterion one is satisfied.

We proceed to demonstrate that criterion two is satisfied. The dynamics of the system ensure that if the initial state of the qubits is $|x\rangle_i|0\rangle$ it will evolve to the state $|x\rangle_i|f(x)\rangle$. By linearity, input states that are superpositions will evolve appropriately. For example, if the input state is $|0\rangle + |1\rangle$, the entire system will evolve to $|0\rangle_i|f(0)\rangle + |1\rangle_i|f(1)\rangle$. By the interpretation, this state will be mapped to the input $0; f(0), 1; f(1)$, which satisfies criterion two.

Criterion three is trivially satisfied when the quantum parallelism process actually takes place.

7.4 TUBE COMPUTERS

The following model of computation was suggested by John Norton. A tube computer, see Figure 6, has a set of slots for every individual input points and also slots for all combinations of input points. If we are computing a function $f : \{0, 1\} \rightarrow \{0, 1\}$, the tube computer would have a slot corresponding to $\{0\}$, $\{1\}$, and $\{0, 1\}$. These slots are connected to a set of tubes.

These tubes proceed to a set of slots that form the output system of the tube computer. The output system has a slot for every possible partial graph of the function of interest. So, for the function we are considering, there will be a slot for $\{< 0, 0 >\}$, $\{< 0, 1 >\}$, $\{< 1, 0 >\}$, $\{< 1, 1 >\}$, $\{< 0, 0 >, < 1, 0 >\}$, $\{< 0, 0 >, < 1, 1 >\}$, $\{< 0, 1 >, < 1, 0 >\}$, and finally, $\{< 0, 1 >, < 1, 1 >\}$. The tubes connect the input system and output system such that when a ball is dropped in the slot representing the input or inputs for the function to be computed, that ball is guided via the tubes that connect the input system to the output system into the slot representing the appropriate partial graph. This model of computing has an interpretation for input and output states, and the tubes set up the appropriate dynamic connections between the input system and the output system. So, were a ball dropped in the slot representing the entire domain of the function, this model of computing will satisfy the general criteria, and, moreover, compute all values of a function in a single computational step.

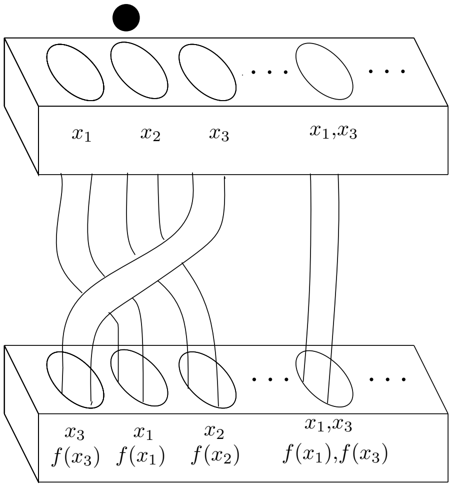

Figure 6: Tube Computer. There an input slot for every possible individual and multiple input for the function f. There is also an output slot for every possible partial graph given the domain and range of f. When a ball is dropped on the top of the tube computer, a tube directs the ball to the appropriate output slot.

Several objections can be raised to the above model of computing, which in virtue of its satisfaction of the general criteria, will be objections against them as well. First, the tube computer is not an interesting model of computation; one typically considers only programmable machines. The tube computer can't be thought of as implementing an algorithm on an input to produce the output. Second, the tube computer above is not practical because it requires a number of slots that is exponential with the size of the domain of the function. Regarding objections one and two, these are objections to the pragmatic value of the tube computer which are irrelevant to whether the machine really performs a computation. So, the tube computer, and also the general criteria, cannot fall prey to these objections.

7.5 THREE CONCEPTS OF PARALLELISM

Even if the tube computer really does compute all values in a single step, there is a sense that this cannot be parallelism. Typically, a computer that makes use of parallelism has multiple processors that perform computations simultaneously. There will be a set of computational devices(ala the general criteria) that constitute the parallel computer. Each of these computational devices will satisfy the general criteria for the computation of an individual value of a function. The computer as a whole, in virtue of all the mini-computational devices, will satisfy the general criteria for all values of a function to be computed. Furthermore, a parallel computer can perform the calculation of all values of a function in a single temporal step. We can refer to this notion of parallelism as *classical parallelism.* The tube computer computes all values in a single step, but certainly not in virtue of computing each individual value of the function.

One might suggest that the tube computer is really computing another function altogether. If the goal is to compute all values of the function f, the tube computer really computes the function, $\mathcal{P}(f)$, that maps the powerset (excluding the empty set) of the domain of f to the powerset (excluding the empty set) of all combinations of pairs from the domain and range of f. Put another way, the tube computer performs a computation of $\mathcal{P}(f) : \mathcal{P}(dom(f)) - \emptyset \rightarrow \mathcal{P}(\{< x, y > |x \in dom(f)\ y \in ran(f)\}) - \emptyset$, where the function $\mathcal{P}(f)$ is defined in the natural way from f. All computers that compute multiple values of a function f can be seen to be computing individual values of $\mathcal{P}(f)$ too, but there are seeming crucial differences. The difference between tube computers and computers that utilize classical parallelism is that the latter really do perform the computations on each point of the domain, whereas tube computers do not. Tube computers compute multiple points of a function exactly the same way as they compute the individual points. These methods of computing multiple values of a function present us with two very different concepts of parallelism.

In classical parallelism successful computation of multiple values of a function is physically related to the computation of the individual values. The computation of multiple points is physically determined by the computation of the individual values. Computation of multiple values in this case just is computation of each value individually. In tube parallelism, there is no physical relationship between computing individual values and the multiple values. The individual values of the function constrain the evolution of multiple values only in a logical sense, i.e. the input states that represent multiple values of the function must evolve to those that represent the same values as if each individual value was computed.

Where does quantum parallelism fit into the mix? Quantum parallelism seems to share characteristics of both classical parallelism and tube parallelism. Like classical parallelism, the computation of multiple values of a function using quantum parallelism is physically constrained by the computation of individual values of a function. As a matter of physics, once the evolution of the computer is fixed for individual input points, by linearity the evolution of the states that represent multiple points is determined. On the other hand, quantum parallelism is like tube parallelism in the sense that one cannot claim that the computation of multiple points of a function was performed in virtue of each individual value being computed. Computing using a superposition is not the same as computing every individual value. It is something less in virtue of the fact that the initial superposition in the quantum parallelism process precludes the possibility of assigning values (values of observables of the computational basis) corresponding to each and every point of the function computed. Thus, the three concepts of parallelism are each related to the next in some way. Claims to the effect that one concept is real parallelism and the others are impostors is a mere matter of terminology. More important than jargon, we must recognize that there are three different ways to compute multiple values of a function in a single temporal step.

7.6 CONCLUSION

In this chapter a set of criteria were developed for a function to be evaluated at several points in its domain that applies to any computational model and/or device. These criteria were developed by examining what made the TMC a good criterion for a function to be evaluated on a Turing machine. It turns out that on the general criteria, the quantum parallelism thesis is true, i.e., quantum computers are capable of computing the values of a function on multiple points in a function's domain in a single computational step. This result is significant because makes a potential explanation of quantum speedup, in the form of the quantum parallelism thesis, available to any interpretation of quantum theory that places ontological significance in the quantum state of a system.

Three concepts of parallelism have been presented: classical, quantum, and tube parallelism. All share the feature that multiple values of a function are computed in a single temporal step. That said, these do represent three distinct concepts because of the distinct ways in which computers can compute multiple values. Which concept is the correct concept of parallelism? I don't know, and it doesn't matter. The important question to ask is what are the implications for explaining the special efficiency of quantum computers? First, it is simply not true that in order to compute multiple values of a function in a single step one must compute each value of the function individually and simultaneously as in classical parallelism. So, to explain why quantum parallelism is the computation of many values of a function in a single temporal step, one is not required to tell a story about how the individual values are computed. Second, the tube computer indicates that to compute multiple values not only in a single temporal step, but a single computational step as well, the dropping of a single ball, one does not require a quantum mechanical system.

Given that the quantum parallelism process meets the general criteria, explanations of quantum speedup need not tell elaborate stories about how the criteria were satisfied, the quantum mechanics of the process takes care of this on its own. In the next chapter, we turn to what is left to be explained.

8.0 COMPUTATIONAL WELL-ADAPTEDNESS

8.1 INTRODUCTION

The quantum parallelism thesis states that quantum computers are capable of calculating the values of many points in a function's domain in a single computational step. If the thesis is true, it seems like it would be a reasonable explanation of why quantum computers are more efficient than classical computers at certain computational tasks. The sufficiency of this type of explanation as it stands is critically dependent on an assumption that no classical computing device could achieve the same behavior. It is clear that this assumption is not true. A classical computer that utilizes several processors will be able to calculate many values from a function's domain in a single temporal step. The tube computer gets the job done in a single roll of a ball. So, the truth of the quantum parallelism thesis alone is not enough to explain the efficiency of quantum computers. Clearly, more needs to be said.

Complexity theory evaluates the efficiency of computers based primarily on the *temporal* and *spatial* resources required to solve computational problems (Nielsen and Chuang 2000, 138). Sometimes energy considerations play a role as well. The truth of the quantum parallelism thesis would at best explain the advantages in *temporal* resources that quantum computers enjoy over other computers. It turns out that the quantum circuit model of computing utilizes fewer spatial resources than typical classical models of computation when evaluating all values of a function. A quantum computer can send a superposition representing all values from a function's domain to a gate that computes the function. A computer that employs classical logic circuits to calculate all values of a function simultaneously needs a gate for every member of the domain of the function to be computed. So, the spatial resources required for the computational task at hand for a classical computer are greater than those required for a quantum computer. Also, the tube computer utilizes spatial resources far greater than that required by a quantum computer. If we take complexity theory seriously, we realize that no explanation of the efficiency of quantum computers is complete unless it addresses the advantages in spatial resources that quantum computers seem to enjoy over other computers.

In this chapter, my task is to offer an explanation of the efficiency of quantum computers with respect to both temporal and spatial resources. To explain why quantum computers require minimal temporal resources for computing multiple values from a function's domain, I employ the conclusions of the last chapter, namely that the quantum parallelism thesis is true in virtue of the quantum parallelism process satisfying the general criteria. Since all values of a function can be computed in a single step of a quantum computer, the explanation is sufficient to explain the temporal resources of quantum computers tasked to compute multiple values of a function. I offer an explanation of the spatial efficiency of quantum computers for the computational task considered in section 8.2. It will be shown that the quantum parallelism process obeys a principle of efficient design: a computer will be well-adapted to computing a function if the physical dependencies of the machine mirror the logical dependencies of the computation.

8.2 AN EXPLANATION OF EFFICIENCY OF QUANTUM COMPUTERS

A satisfactory explanation of the efficiency of quantum computers must indicate why they consume fewer resources than other computers which perform the same tasks. A computer that can compute all values in a function's domain in a single step will be efficient with respect to temporal resources, but the tube computer indicates that this alone is not sufficient. In order to compute multiple values in a single step, models such as the tube computer use resources that are exponential with respect to the size of the computational problem. If the function to be computed has a domain $\{0,1\}^n$, the tube computer requires $2^{2^n} - 1$ slots to represent the inputs alone! In this section it is detailed why quantum computers are so well-adapted to computing multiple values of a function. A bit of background in complexity theory is required for that task to indicate the types of explanations available.

8.2.1 Complexity theory and explanations of efficiency

Complexity theory is usually concerned with asymptotic behavior. The notation below has been developed to quickly display asymptotic behavior. We suppose that $f, g : \mathbf{N} \to \mathbf{N}$.

$f \in O(g)$ if there are numbers $c, d \in \mathbf{N}$ such that $f(n) \leq c \cdot g(n)$ for all $n \geq d$.

$f \in \Omega(g)$ if $g \in O(f)$.

$f \in \Theta(g)$ if $g \in O(f)$ and $f \in O(g)$(Vollmer 1999, 234).

Often, the efficiency of an algorithm or computer is measured as a function of the size of the input of the function to be computed. For example, suppose we are computing a function $f : \{0, 1\}^n \to \{0, 1\}^m$. The "size" of the input is n. It is standard practice to assume that f can be computed efficiently with respect to time by some computer if any point of the function can be evaluated in a time that is $O(g(n))$, where $g(n)$ is a polynomial function of n.

The goal of algorithmic complexity theory is to prove bounds on the temporal, spatial, and energetic resources required to solve certain types of computational problems relative to particular models of computation and to construct efficient algorithms. The set of problems that can be solved with a particular model of computation and fixed resources is called a complexity class. Once complexity classes are established, one can demonstrate how certain classes are related to one another. For instance, if one model of computation can be used to simulate another model of computation, one can compare the resources necessary to solve certain types of problems between different models of computation. For example, consider the languages that a deterministic Turing machine can accept or reject in time polynomially related to the size of the problem, the class P, versus the problems that a non-deterministic Turing machine can solve with similar resources, the class NP. Obviously P is a subset of NP because a non-deterministic Turing machines can be programmed to be deterministic machines. This is one way complexity theory relates the efficiency of various models of computing to one another.

For any model of computing, there will be a technique for keeping track of the temporal or spatial resources consumed by any algorithm implemented for that model. In this way, the resources required for different algorithms run on different computational models can be compared.[1]

For the case of parallel computation of a function $f : \{0,1\}^n \to \{0,1\}^m$, we again choose to measure the size of the input by n. There are 2^n different possible individual inputs. There are $2^{2^n} - 1$ different combinations of individual inputs that can be constructed from $\{0,1\}^n$. It is useful to compare the resource requirements for tube, classical, and quantum parallelism. We distinguish the *representational resources*, those resources required to represent inputs and outputs, from *dynamic resources*, those resources required to dynamically relate input states to output states for successful computation. The representational resources required for the inputs to the tube computer are $2^{2^n} - 1$ slots. The spatial resources required for the dynamics are exponential. We need a tube for every combination of individual points of the function, i.e. $2^{2^n} - 1$ tubes. A computer that instantiates classical parallelism requires an input for every value to be computed, 2^n of them, and each input requires n cbits to represent. Classical parallelism requires a gate that computes f for every input, and so requires 2^n gates. A quantum computer using quantum parallelism requires only n qubits to represent any individual or combination of individual points to be computed. A quantum computer will only require a single gate to compute in parallel. The above are summarized in the table below.

Spatial Resource Type	Tube	Classical	Quantum
Representational	Exponential	Exponential	Linear
Dynamic	Exponential	Exponential	Constant

Clearly, the efficiency the quantum parallelism process is impressive.

Given the above framework for comparisons of computational models and algorithms, we can proceed to examine possible explanations of computational efficiency. Clearly, one type of explanation of efficiency is to give an account how a computer satisfies the general criteria using only the spatial and temporal resources that it did. This kind of explanation

[1] The units of space or time defined by complexity measures of different models of computation will differ by a multiplicative constants. Such differences are considered negligible. This is automatically taken care of by the asymptotic notation described above.

is given by indicating the algorithm for the computational model, and counting resources as complexity theory would have one do it. The resources available, and how they might be constrained are a matter of the model alone and the ways that appropriate functional behavior can be delivered within the context of the model. The way energy resources scale will be a matter of the physical theory of computer the model is intended for. This is a full account of why the resource consumed for a particular computational task, for a particular computational model, for a particular algorithm, were as they were. We have just done this for tube, classical and quantum parallelism, but in order to explain why quantum computers are more efficient than classical computers, something more is desired.

Of those who have tried to explain why quantum computers are more efficient than classical computers, e.g. Deutsch and Steane, neither have made any criteria explicit for what constitutes a satisfactory explanation. Deutsch seems to think that what must be explained is why a quantum computer can compute all values simultaneously. His method is to specify how the world might be, in Deutsch's case ala the many-worlds interpretation, to be compatible with the quantum formalism that allows multiple values to be computed in a single temporal step. Given that the quantum parallelism process satisfies the general criteria for multiple values of a function to be computed, this explanatory story is certainly not necessary, and, as I will argue in chapter four, misguided. Steane, as we will discuss later, simply points to quantum entanglement as an efficient way to represent multiple values of a function. He seems to point to the feature of quantum computers that is lacking in other models of computation that is crucial for efficient use of resources.

The explanatory task that Steane sets for himself seems to be useful for several reasons. By pinpointing the special features of quantum systems that are responsible for their efficiency in performing computational tasks, those who develop new algorithms can attempt to exploit this feature as best as possible to make these algorithms as efficient as possible. Second, we can look to see if the functional behavior instantiated by such a feature can be simulated efficiently in other models of computation to improve efficiency. Finally, we might learn something about the nature of the quantum world by pinpointing the special features of quantum systems responsible for their computational efficiency.

Keeping with Steane's explanatory endeavor, I seek to characterize what makes quantum computers that exploit parallelism efficient. Below I develop a concept of computational well-adaptedness. Well-adapted computers are always efficient at computing multiple values in a single temporal step. I will argue that a quantum computer that employs quantum parallelism is efficient because it is well-adapted. We note that quantum computers are well-adapted because they utilize entangled states.

8.2.2 Computational well-adaptedness

We say that a computer is *well-adapted* to computing multiple values of a function in parallel if and only if

1. The general criteria are satisfied in a single temporal step.
2. The computer in question is efficient computing individual values.
3. The physical dependencies of the computer mirror the logical dependencies of the computational task.

Condition 1 has been discussed at length in the preceding chapter. We focus on condition 3.

Condition 3 applies to two different features of a computer: how it represents inputs and outputs, and how the computational dynamics are instantiated. We focus first on representation of inputs(outputs). Combinations of individual inputs, or multiple inputs, will bear a logical relationship to the individual inputs. When computing the function

$$\mathcal{P}(f) : \mathcal{P}(dom(f)) - \emptyset \rightarrow \mathcal{P}(\{< x, y > | x \in dom(f)\; y \in ran(f)\}) - \emptyset,$$

note that the combination inputs can be constructed via set theoretic operations on the individual inputs. The individual inputs logically determine what the combination states are. Condition 3 requires that this logical dependence be reflected in a physical dependence. In particular, condition 3 requires that a type of physical relation to exist between individual states and multiple states so that the logical relationship between individual inputs(outputs) and multiple inputs(outputs) is represented physically. Condition 3 requires that once the resources required to represent any individual input(output) is determined, as a matter of physics so too are the combination inputs(outputs). So, the same resources required to represent individual inputs(outputs) represent the multiple inputs(outputs) as well.

Condition 3 places demands on the dynamics. Note that the function $\mathcal{P}(f)$ is defined when f is. In a similar vane, condition 3 requires that when the resources required to instantiate the computation of individual values are in place, as a matter of physics, the resources required for the computation of multiple values are in place. No additional resources must be consumed to instantiate the computational dynamics of the multiple states.

In summary, we have the following: If the physical dependencies of the computer mirror the mathematical dependencies of the computation, then

Constraint on Representation There is a physical relationship between the input(output) states such that once the resources required to represent any individual input(output) is fixed, as a matter of physics, so too are the resources required to represent multiple inputs(outputs).

Constraint on Dynamics As soon as the resources required for the computational dynamics of any individual input state are fixed, as a matter of physics, so too are the computational dynamics of the multiple input states.

Well-adaptedness is a sufficient condition for efficient parallel computation. First, condition 1 ensures efficient use of temporal resources. Conditions 2 and 3 together ensure efficient use of spatial and energetic resources. A well-adapted computer that satisfies the constraint on representation will require no additional spatial or energetic resources beyond those required to compute individual values. Moreover the multiple states will not require more resources than the individual values to represent. From condition 3, we know these are already reasonable requirements. Finally, we know that there will be no resources required for the computational dynamics beyond those required for computation of individual values by the constraint on dynamics. Again, from condition 2, we know these will be reasonable.

Quantum computers are well-adapted. First, they satisfy condition 1, by the general criteria. The constraint on representation is met because the states that represent multiple inputs(outputs) are superpositions of individual inputs(outputs). Of course, given that fact, the number of qubits required to represent multiple inputs(outputs) are equivalent to those required to represent individual inputs(outputs). Finally, the dynamics of the quantum computer are linear. Given that multiple inputs(outputs) are represented as a linear superposition of the states representing individual inputs, the constraint on dynamics is met. Finally, we know quantum computers computing individual values are efficient, so condition 2 is met.

It is useful to see how the above result manifests itself in the models of computation we have examined. The tube computer fails every condition but 1. With the tube computer, the representation and dynamics of the multiple states are completely physically independent of the input states and the dynamics of those states. For each combination of individual values, an additional slot is required as well as a tube to direct the ball to the output states. Moreover, there is no physical relationship between the individual inputs(outputs) and multiple inputs(outputs). Also, the tubes that implement the dynamics for the individual inputs are independent of the tubes that implement the dynamics for the multiple states. Far more resources are required to compute multiple values of a function than those required to compute individual values.

For a computer implementing classical parallelism, in order to represent multiple inputs(outputs) one uses the resources required to display every single individual input, each requiring some fixed set of resources. Each individual input requires n cbits to represent, and there are 2^n individual values to represent. So, computers that implement classical parallelism use exponential spatial resources to represent multiple inputs. For the classical computer to meet the contraint on representation, it would have to be capable of representing any value using n cbits, which is clearly impossible. The computer instantiating classical parallelism also fails the constraint on the dynamics. The dynamics for any individual value are instantiated by a single gate. To compute multiple values, a gate is required for each value to be computed. Hence the computer will use spatial resources exponential in n for the dynamics.

So, a tube computer, and a classical computer implementing parallelism both fail to meet the constraint on representation and the constraint on dynamics. Clearly, quantum computers are efficient because they are well-adapted to computing multiple values in parallel.[2]

8.3 CONCLUSION

In this chapter I have argued that in order to explain the efficiency of any model of computation, one must account for the efficient use of both temporal and spatial resources. The quantum parallelism thesis only accounts for efficient use of temporal resources by quantum computers. The tube computer demonstrated that the calculation of many values of a function simultaneously or in a single computational step does not guarantee that a computer will be efficient with respect to spatial resources. Hence, a further principle was proposed to account for the fact that quantum computers made more efficient use of spatial resources than other computers, namely that it is well-adapted to compute multiple values of a function.

Well-adaptedness is a good explanation of quantum efficiency with regard to spatial resources for several reasons. Well-adaptedness seems to pinpoint what features of quantum systems are responsible for their efficiency. Also, it is clear that other models of computation that can compute in parallel are not well-adapted. Now, this would be especially powerful if one could show that well-adapteness was a sufficient *and* necessary condition for efficient computation. Of course, there is little hope of doing so, given not only the myriad physical systems that can be used to perform a computation, but also, crucially, the myriad ways of interpreting the states of the system that performs the computation as inputs(outputs).

Now, necessary and sufficient conditions for computational efficiency are easy to construct, we simply just require that the temporal and spatial resources required for computation match those of a quantum computer. However, this gives absolutely no insight into the features of a computer that are responsible for the efficiency. Furthermore, it gives no insight into how an efficient computer might be constructed. Well-adaptedness does not suffer these ailments.

[2] It is not obvious that the constraint on representation and the constraint on dynamics are independent. In appendix A I demonstrate that the constraint on representation is independent of the constraint on the dynamics. It is not known if constraint on dynamics is independent of the constraint on representation.

9.0 STEANE'S OBJECTIONS

9.1 INTRODUCTION

In "A quantum computer only needs one universe," Andrew Steane argues that our understanding of quantum computers is not wedded to a many-worlds interpretation of quantum theory. The many-worlds interpretation is naturally associated with quantum computing because of the quantum parallelism process, which is often implicated in an explanation of the efficiency of quantum computers. Regardless of the truth or falsity of the quantum parallelism thesis, the quantum parallelism process at least gives the impression that many values of a function are being computed at once. The many-worlds interpretation can make sense of such a claim by suggesting that each value of the function is computed in a parallel world.

Steane's strategy for detaching understanding of quantum computing from the many-worlds interpretation is two-fold. First, Steane argues that the evidence that is sometimes provided for the quantum parallelism thesis is dubious. The intended conclusion is that the quantum parallelism thesis is false, and minimally he wants to shift the burden of proof to those who think the thesis is true. Second, Steane undermines the attractiveness of the quantum parallelism thesis by arguing that it has limited explanatory scope. Steane points out that for a particular model of quantum computation, so-called cluster state computers, there is no obvious instantiation of a quantum parallelism process. Yet, the efficiency of cluster state computers is similar to those that do employ quantum parallelism. Furthermore, Steane advances an explanation of the efficiency of quantum computers that is meant to apply to all quantum computers that is independent of the quantum parallelism thesis. So, to demonstrate the independence of quantum computers and the many-worlds interpretation, Steane argues against the quantum parallelism thesis. If the quantum parallelism thesis is not true, the usefulness of the many-worlds interpretation to our understanding of quantum computing becomes least less obvious.

Though I am sympathetic to Steane's cause, to argue the independence of explanations of quantum computational efficiency and the many-worlds interpretation, I think his strategy sacrifices too much. As I argued in chapter seven, there are good reasons to regard the quantum parallelism thesis as true; it satisfies the general criteria for a function to be computed. What is called for is an argument that the many-worlds interpretation offers no better understanding of quantum computing than other interpretations. I save this task for chapter 10. In this chapter, I wish to defend my conclusions in chapters 7 and 8 against Steane's criticisms.

In section 9.2 I address Steane's concerns regarding the evidence for the quantum parallelism thesis. I will argue that none of the objections levied at the thesis are successful. In section 9.3 I examine Steane's objections to the explanatory applicability and sufficiency of the quantum parallelism thesis. Finally, I will discuss Steane's alternative explanation of the efficiency of quantum computers.

9.2 OBJECTIONS TO THE QUANTUM PARALLELISM THESIS

In his attempt to shift the burden of proof to those who want to argue that quantum parallelism is true, Steane points out that the results of computations that are supposedly performed during quantum parallelism are not accessible. We have dealt with this objection in chapter 7. The inaccessibility of results is due to epistemic limitations of our knowledge of an unknown quantum state. This does not change the ontological fact that the state of the quantum computer evolved such that it satisfies the general criteria for a function to be computed. Epistemic limitations might have implications regarding the usefulness of quantum parallelism, but they have no bearing on whether values were computed or not.

In the literature on quantum computing, it is often remarked that to simulate a quantum computer, a classical computer would require exponentially more resources than the quantum computer (Jozsa 2000, 109). Though no one has explicitly endorsed this argument, it is alluded to that the above fact is an indication that quantum computers perform *exponentially more calculations* than classical computers as a brute fact of physical evolution. Steane correctly points out that the amount of computation performed by some algorithm is not measured by the resources required to perform that computation a different way. Steane offers a useful example. To find a root of a monotonic function $f : \{0, \dots, N-1\} \rightarrow \{\dots, -1, 0, 1, \dots\}$, it is sufficient to check all elements in the domain of the function to determine when the function takes the value 0. A much more efficient way is to begin by checking $f(N/2)$; if it is greater than 0, then check $f(N/4)$, if it is less then 0, check $f(3N/4)$, etc. One might be tempted to suggest, if we discover that $f(N/2) < 0$, then the computations $f(0) \neq 0$,$f(1) \neq 0$,..., $f(N/2) \neq 0$ were performed too. Really, one just uses the fact that the function is monotonic to eliminate the need to perform an exhaustive check of the domain of the function. Steane seems to advocate an analogous view of quantum computers. We might best think of them as running interesting efficient algorithms that solve computational problems in interesting new ways, but not that they perform the same calculations that classical computers do.

I am sympathetic to the Steane's views on the quantification of computations. Certainly there is no warrant to suggest that two different algorithms that solve the same computational problem always perform the same number of computations. In fact, I would go take this point a step further than Steane. It is simply false that a machine must perform calculations in order to compute a function. The tube computer is a good computational device to focus on.[1] Arguably, there are no calculations going on in the tube computer, only appropriate functional behavior. Rather than arguing about whether calculations are really being done in some algorithmic sense or not, the functional behavior is all that is relevant to the efficiency of the computers, and relevant to whether a function is evaluated.

It is exactly at this stage, the functional behavior of quantum computers described by the formalism of quantum mechanics, that Steane has further reservations about the quantum parallelism thesis. Steane argues even when we do closely examine the typical evidence for the quantum parallelism thesis, the evolution of the states of a quantum computer as

[1] Alternatively, one might think about a differential analyzer.

$\sum_{x=0}^{2^n-1}|x\rangle|0\rangle \xrightarrow{U_f} \sum_{x=0}^{2^n-1}|x\rangle|f(x)\rangle$, is insufficient to establish the truth of the thesis because mathematical notation can sometimes be misleading.

As an example, Steane asks us to consider a model of computing that operates on n needles. We suppose that the individual needles can point in one of three directions: north, east and northeast. We interpret the needle pointing up as the number zero, and label the state of such a needle as [0]. Similarly, we label the state of a needle pointing east as [1], and northeast as $\frac{1}{\sqrt{2}}([0]+[1])$. Steane purports to give an example of computation that appears parallel. Suppose that the needles are initially in the state [0] and then they are rotated to $\frac{1}{\sqrt{2}}([0]+[1])$. We adopt the notation that $[0][0]\ldots[0] = [00\ldots0]$. We identify the binary strings with their associated base ten number. So, 00000 will be identified with 0, and so on. The evolution is described by

$$[0] \longrightarrow \frac{1}{2^{n/2}}\sum_{x=0}^{2^n-1}[x] \tag{9.1}$$

First, after the needles are rotated, their state will be $\frac{1}{2^{n/2}}([0]+[1])([0]+[1])\ldots([0]+[1])$. Steane adds additional mathematical structure to the vector space that describes the position of the needles by adding a product rule: $(a[0]+b[1])(a'[0]+b'[1]) = aa'[0][0]+ab'[0][1]+ba'[1][0]+bb'[1][1]$. So, $\frac{1}{2^{n/2}}([0]+[1])([0]+[1])\ldots([0]+[1]) = \frac{1}{2^{n/2}}\sum_{x=0}^{2^n-1}[x]$. The interpretation given to such a state is that each term in the sum represents a value of the function. According to Steane, (9.1) misleadingly represents the evaluation of the function $f(x) = x$ over the domain $\{0,\ldots,2^n-1\}$. The mathematical notation appears to indicate parallel computation because 2^n values were computed, but by only using n steps (rotations of the n needles).

Since mathematical notation can be misleading, it is crucial to have a way to vet bogus computations from real ones. That is exactly why the general criteria were developed in chapter 7. The purported computation described above fails the general criteria for a function to be computed at multiple points. For convenience the general criteria are:

The General Criteria:

A function f was evaluated at points $X \subseteq dom(f)$ by computing device $\mathcal{M}$ if and only if

1. An interpretation, (α, β), for a function f and the computing device $\mathcal{M}$ exists.

2. The computational device $\mathcal{M}$ instantiates correlations necessary to indicate the correct values of f according to the interpretation (α, β), i.e. if $\forall X \subseteq dom(f)$, $\beta \circ U \circ \alpha(X) = G(f|_X)$.
3. An input state was entered into device $\mathcal{M}$ associated with the input points $X \subseteq dom(f)$ according to α and the computational process of device $\mathcal{M}$ resulted in an output state that is associated with the output $G(f|_X)$ according to β.

Obviously, the needle computer fails item three, but this objection is a cheap as the function the needle computer purportedly computes. This is trivially corrected by assuming the state of the needles is initially $\frac{1}{2^{n/2}} \sum_{x=0}^{2^n-1} [x]$. The real problem comes in with item one. There is no interpretation available for the needle computer because there are not enough states of the computer to represent all possible inputs from $\mathcal{P}(dom(f))$. For example, let us assume we are only dealing with two needles, and want to evaluate a function $f : \{0,1\}^2 \longrightarrow \{0,1\}^2$. Suppose we wish to evaluate the function on 01 and 10. We assume that the vectors describing each needle are normalized, so for an arbitrary needle, $(a[0] + b[1])$, $a^2 + b^2 = 1$. According to the above rules, this input would be represented as $\frac{1}{2}([0][1] + [1][0])$. We know that the state of the needles is always separable, and can be described as $(a[0] + b[1])(a'[0] + b'[1])$. According to the product rule,$(a[0] + b[1])(a'[0] + b'[1]) = aa'[0][0]+ab'[0][1]+ba'[1][0]+bb'[1][1]$. Following this rule, for the input 01 and 10, $aa', bb' = 0$ and $ab' = ba' \neq 0$, and this entails a contradiction. Either $a = 0, a' = 0$, or $a, a' = 0$. But if any of these are true, it would imply ab' or ba' or both were equal to 0, but they are nonzero by assumption. So, plainly the needle computer would not and cannot satisfy the general criteria, and is not capable of computing multiple values of a function in parallel.

I agree with Steane that the above example does not count as multiple evaluations of a function. I also agree with Steane that mathematical notation can be misleading. This is precisely why it is important to use the general criteria to decide when a function was computed.

With the example of the needle computer, what Steane seems to be objecting to, but not stating directly, is the notion that quantum computers perform the same algorithmic procedures in less time and space than classical computers. It is precisely this that is irrelevant on a functional view of computing. The efficiency of computers is measured by the temporal and spatial resources required to get the appropriate functional behavior out of a computing device. Note that the quantum parallelism thesis, as an explanation of temporal efficiency, depends only on the functional behavior of the process. When the many-worlds interpretation is connected to the quantum parallelism process do we have a notion of many algorithmic procedures occurring simultaneously, but this is superfluous. Furthermore, one can object to that conception of computing without objecting the quantum parallelism thesis.

Steane offers a final objection to the notion that a quantum computer is doing exponentially many computations at once.

> An n-bit quantum computer is sensitive to decoherence to the level 1/Poly(n), not 1/exp(n), in the case that different qubits have independent decoherence. If the quantum computer were really "doing 2^n computations", and the result depended on getting a large proportion of them right, then we would expect it to be sensitive to errors at the level of $1/2^n$, which it is not(Steane 2003, 473-4).

It seems to me this remark is seriously misguided. It presupposes some relationship between the number of calculations performed and decoherence or error rates. Clearly the causal relationship that must underlie such a correlation, if there is one, is due to the physical systems that represent inputs and the manipulations performed on those systems for purposes of calculation. It would seem that what explains the low error rate is the ability of quantum systems to represent all points in a function's domain with the same resources required to represent an individual value. If physical systems are sensitive to decoherence at the level of 1/Poly(n) for individual values, we would expect the same level for multiple values represented by the same physical systems. Furthermore, the error rate for certain processes is dependent on the physics of the system used for the process. One should not expect to import intuitions from computers that utilize classical systems to computers that utilize quantum systems.

9.3 EXPLANATIONS OF QUANTUM SPEEDUP

In this section we examine Steane's objections to the scope of the quantum parallelism thesis as well as his explanation of quantum speedup. We begin by examining what Steane views as a model of computing that cannot be characterized as implementing parallelism, and hence constitutes a serious challenge to the view that the quantum parallelism thesis offers a reasonable explanation of the efficiency of quantum computers.

Raussendorf and Briegel (2000) have discovered an interesting new model for quantum computation, the cluster state computer. A cluster state computer manipulates quantum states using a unitary transformation that entangles qubits and single qubit measurements conditioned on classical information. The qubits that form the cluster computer begin the computation in some standard state. Qubits are submitted to an operation that entangles the qubits that form the cluster. Raussendorf and Briegel have shown that the state of an arbitrary qubit can be transferred to any other qubit using single qubit measurements conditioned on classical information. They have also shown that an arbitrary rotation in $SU(2)$ can be performed by making single qubit measurements on four qubits that are conditioned on classical information. Finally, they show that a CNOT gate can be simulated using single qubit measurements condition on classical information on four qubits. Since arbitrary rotations and CNOT gates are universal, meaning any arbitrary unitary transformation can be performed to any desired degree of accuracy using only these gates, the cluster state computer can perform the same transformations on qubits that a quantum Turing machine or a quantum circuit computer can. Since a constant number of qubits and measurements are required to simulate universal quantum gates, the spatial and temporal resources consumed by a cluster state computer for computational problems is the same order as the resources required by quantum circuit computers or quantum Turing machines.

Steane thinks the cluster state computer is interesting because,

> The evolution of the cluster-state computer is not readily or appropriately described as a set of exponentially many computations going on at once. It is readily described as a sequence of measurements whose outcomes exhibit correlations generated by entanglement (Steane 2003, 474).

Furthermore,

> ...the qubits play a passive role, in that they are prepared at the outset in a standard state, and thereafter simply measured one at a time. Rather than "performing computations in superposition," the role of the quantum information is to provide a resource, namely entanglement, which permits measurement outcomes to exhibit correlations of a different nature to those which would be possible with a set of bits (Steane 2003, 475).

Steane emphasizes the differences between cluster state computers and the quantum circuit computers, but given that both models can perform arbitrary unitary transformations on qubits, we should be careful what conclusions can be drawn from the differences between the computers. Steane claims that for cluster state computers, qubits play a passive role, in part because they are initially in a standard state, and that measurements are the active players in this model of computing. It should be pointed out, that the same is true in the circuit model of computing. It is typically assumed for the quantum circuit model that qubits enter the computer in a standard state in the computational basis. Various circuitry is required to transform that standard state into an input state. Similarly, for cluster computers, measurements on qubits transform the standard state to an appropriate input state. Also, just as gates in a quantum circuit model perform transformations according the input state to produce the output state, measurements performed on qubits in the cluster-state model perform transformations on qubits in the cluster state computer. Just as a gate is a device to perform a transformation, so too are measurement devices. What is really interesting about cluster state computers is that measurements on individual qubits conditioned on classical information are sufficient to perform any unitary transformation to arbitrary accuracy. Conceptually, the ways in which a cluster-state computer performs unitary transformations is a far cry from quantum circuit computers; however, the end result is the same. The functional relationship established between the input states and the output states is interpretable as correlating inputs and outputs according to some function, ala the general criteria.

It is also interesting to note, that cluster-state computers can implement quantum parallelism in a straight-forward way. In the quantum circuits model, the complexity cost of a quantum gate that instantiates an unknown function is assigned unit value. The true complexity of this gate, relative to some universal set of gates, is surely greater than one, except in limited cases. We know that the cluster state computer can simulate the action of a universal set of gates. So, we know that there are a set of measurements on qubits that perform the same transformation that an unknown quantum gate that instantiates some function does. Let us black-box the complexity cost of the resources required to instantiate the unknown transformation and assign it unit value, just as we do with the gate in the standard quantum parallelism process. We know that if a state representing a superposition of all individual input states is submitted to such a transformation, that the output state can be interpreted as the graph of the unknown function. Furthermore, we know that a cluster state computer can prepare such a state. So, within the cluster state model, we can have quantum parallelism.

The existence of the cluster-state computer has no bearing on whether the quantum parallelism thesis is true or false. What is in question is if the model challenges the explanatory adequacy of the quantum parallelism thesis. It cannot. First, as I've argued cluster state computers are capable of quantum parallelism. Second, no claim has been made that quantum parallelism will explain every instance of quantum computational efficiency. An explanation of quantum efficiency that involves quantum parallelism must be circumscribed. In particular, it must be circumscribed to those algorithms that indeed utilize parallelism. It may be the case that there are certain algorithms run on a circuit computer that cannot be explained by quantum parallelism. This is perfectly consistent with quantum parallelism being a good explanation of efficiency in other cases. So, Steane's criticism that quantum parallelism might not explain all instances of quantum computational efficiency is beside the point.

Steane offers his own explanation of why certain quantum algorithms are more efficient than others.

> A quantum computer can be more efficient than a classical one generating some specific computational results because quantum entanglement offers a way to generate and manipulate a physical representation of the correlations between logical entities without the need to completely represent the logical entities themselves (Steane 2003, 476).

By "logical entities" Steane simply means integers, numbers, values, etc. By "representation of the correlations" Steane means the quantum state $\sum_{x=0}^{2^n-1}|x\rangle|f(x)\rangle$. Using Steane's terminology, this state doesn't *completely* represent all values of the function because the values cannot be accessed non-probabilistically. What Steane seems to be getting at is that if we required that the values be completely represented, accessible non-probabilistically, we would have to commit more resources to represent the values. In order to be perfectly accessible, a sequence of qubits must be in an eigenstate that represents all values of the function. If it takes n qubits to represent an individual value in the computational basis, it would require $2^n * n$ bits to "completely represent" all values of the function. Put a different way, Steane is suggesting that quantum computers are more efficient than classical computers because they can exploit entanglement to efficiently represent multiple values or correlations, and classical computers, at least classical circuit computers, cannot efficiently imitate such behavior.

Of course, I am sympathetic to the above explanation insofar as it implicates entanglement. However, Steane's purported explanation of quantum efficiency is dangerously close to simply noting that of those processes that seemingly cannot be simulated efficiently by a classical circuit computer or its equivalent, that entanglement is present. Steane gives us no indication *why* entanglement can be exploited to efficiently represent multiple values of a function, nor why it allows for fast algorithms. The explanation of computational efficiency offered in the last chapter, addresses this explicitly with the requirement that the physical dependencies of the machine mirror the logical dependencies of the function to be computed. By ignoring *why* entanglement is a useful means of reducing resource consumption in computation, Steane neglects to entertain the possibility that well-adapted classical computers might exist.

9.4 CONCLUSION

In "A quantum computer only needs one universe," Steane offers several plausibility considerations that suggest that the quantum parallelism process really doesn't count as the evaluation of several values of a function in a single computational step. Each of these has been dealt with in turn: accessibility, misleading mathematical notation, and error rates. The general criteria give a good means of determining when a function was computed. Moreover, the criteria have been shown to rule out dubious examples of computations, such as the needle computer. Also, we see not only that cluster state computers really do compute on the general criteria, but they can implement parallelism too.

It has been suggested that Steane objects to the view that quantum computers are really performing several algorithmic procedures, as several processors in a classical computer would, when it implements parallelism. That is a view that suggests a many-worlds view of computation. As we have seen in the last chapter, we see a divergence between the idea of algorithmic calculation of values and appropriate functional behavior that can be interpreted as a computation of multiple values of a function. Once one adopts a functional view of computation, the favorable image that the many-worlds interpretation offers of quantum parallelism seems unnecessary. The results of this chapter and the last suggest that the status of the quantum parallelism thesis depends on its satisfaction of the general criteria alone, and is consequently independent of any interpretation of quantum theory.

We have seen that Steane's explanatory view implicates the ability of quantum computers to represent correlations (or values) efficiently, which on close examination, seems perfectly compatible with parallelism. So, it seems reasonable to utilize the quantum parallelism thesis to explain the temporal efficiency of certain quantum algorithms. Whether this will always be sufficient will have to be decided on an algorithm by algorithm basis, unless some strong lower bound results become available.

10.0 THE MANY-WORLDS INTERPRETATION AND QUANTUM COMPUTING

10.1 INTRODUCTION

Because the quantum parallelism process satisfies the general criteria for multiple values of a function to be computed, there is no need to appeal to interpretations of quantum mechanics to explain how a quantum computer can compute many values in a single step. Any interpretation that treats the quantum state of a system objectively can employ the truth of the quantum parallelism thesis to explain quantum computational efficiency. That said, there is something downright appealing about the many-worlds interpretation's story about how multiple values of a function are computed in the quantum parallelism process.

The seeming advantage the many-worlds interpretation (MWI) provides is a robust story that underwrites the truth of the quantum parallelism thesis. In particular, it allows a story to be told about quantum parallelism that is tantalizingly close to classical parallelism. Recall that according to the MWI, multiple values are computed in virtue of individual values each being computed, just as in a classical computer, but in multiple worlds simultaneously. Moreover, the MWI makes quick work of the accessibility objection to the quantum parallelism thesis; only one value is accessible because an observer is restricted to observing the value of the function in a single world.

Unfortunately, when we press the MWI beyond simply the quantum parallelism process, we run into difficulties. In this chapter, we examine a computation that utilizes the quantum parallelism process. The computational task is to determine whether a binary function is constant or balanced.[1] It turns out that a quantum computer performing this task is

[1] A function is balanced if exactly half of the values in the range of the function are 0 and the other half

more efficient than a classical circuit computer. Naturally, one would like to be able to say that one reason for this was that it can compute all values of the function in a single temporal step. Obviously the MWI excels at this. However, as we will see, when pressed to explain how to explain how the results of the quantum parallelism process are used to determine if the function is constant or balanced, the MWI is at a loss. As is the case with nearly every interpretation of quantum mechanics, when pressed beyond simple examples, the interpretation comes up lacking.

In section 10.2 I describe the MWI using the measurement problem to illustrate the supposed advantages of the interpretation. I also describe the quantum parallelism process in the context of the MWI. In section 10.3 I describe an algorithm for determining if a binary function is constant or balanced. Finally, I attempt to explain the success of the algorithm using fundamental explanatory machinery of the MWI, worlds. It will be argued that the success of the algorithm can only be explained by appealing to the global state that describes the quantum mechanics of every world.

10.2 THE MWI AND QUANTUM PARALLELISM

The MWI was developed to deal with the measurement problem. The measurement problem can be formulated as follows. Let's use a cat to measure x-spin of an electron. We will treat all systems quantum mechanically so they evolve deterministically according to the Schrödinger equation. We specify a unitary evolution, U, that evolves the cat and electron system as follows:

$$|\uparrow\rangle|live\ cat\rangle \xrightarrow{U} |\uparrow\rangle|live\ cat\rangle$$

$$|\downarrow\rangle|live\ cat\rangle \xrightarrow{U} |\downarrow\rangle|dead\ cat\rangle$$

Let us further suppose that the electron is initially in a superposition of x-spin eigenstates

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)$$

[1] A constant function takes a single value for every point in it's domain.

Let's now make a measurement of the electron using the cat according to the evolution U. The state prior to measurement is:

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)|live\ cat\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle|live\ cat\rangle + |\downarrow\rangle|live\ cat\rangle)$$

The state evolves, according to 1, to

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle|live\ cat\rangle + |\downarrow\rangle|dead\ cat\rangle).$$

According to the orthodox semantics for quantum mechanics, a system possesses a value for an observable if and only if it is in an eigenstate for that observable. Clearly the final state of the electron is not an eigenstate of x-spin. Similarly, the cat is not in an eigenstate of being alive or dead. So, according to the standard semantics and the Schrödinger evolution, there will be no determinate fact about the cat being alive or dead. Here's the rub; when I do such experiments at home, I really do end up with a dead cat or a live one! So, there is something wrong either with 1. the Hilbert space representation of quantum states that allows for superpositions, 2. the Schrödinger evolution, 3. the standard semantics, or 4. my perception of either a live or dead cat.

The MWI keeps 1. and 2., but changes 3. and has some things to say about 4. The MWI[2] suggests that there is a universal state vector for the universe that describes the evolution of many different worlds that evolves according to the Schrödinger equation. The worlds described by the state vector of the universe are in one-to-one correspondence with terms in the state vector. During a measurement interaction new worlds are created, a new world for each possible outcome of the measurement. In the measurement interaction described above, there will be a world created in which the cat is dead and another world created where the cat is alive. There is no interaction between worlds, so observers will never find themselves with superpositions of live and dead cats. The semantics of the MWI are slightly different than the standard semantics. The MWI replaces the eigenvalue-eigenvector link with something that might be called the eigenvalue-eigenvector-world link. An observable has a determinate value in a particular world if and only if it is in an eigenstate of that

[2] There is not one single MWI, but many. Here I present the Dewitt version of the MWI.

observable in that world. So, the eigenvalue-eigenvector link is simply relativized to worlds rather than applying to the universe as a whole.

Recall the quantum parallelism process:

$$\sum_{x=0}^{2^n-1} |x\rangle|0\rangle \xrightarrow{U_f} \sum_{x=0}^{2^n-1} |x\rangle|f(x)\rangle. \tag{10.1}$$

On the MWI, there is a world for each term in the superposition on the LHS of (10.1). The evolution described by U_f takes place in each world, yielding the RHS of (10.1), and each value of the function is computed in a different world. Again, the MWI proponent has a good answer to the objection to the claim that all values of the function are calculated, namely that only one value of the function is accessible. A measurement of the value of a function is subject to the laws of quantum mechanics. Once we specify the measurement evolution, the MWI makes the appropriate predictions about the accessibility of values of the function. Suppose that a measurement of the system has the following evolution:

$$|x\rangle|f(x)\rangle|`f(x) =?'\rangle \longrightarrow |x\rangle|f(x)\rangle|`f(x) = k'_x\rangle \tag{10.2}$$

where $|`f(x) =?'\rangle$ describes the ready state of a measurement device and $|`f(x) = k'_x\rangle$ is the state of the device that displays the value of the function; k_x is 0 or 1 according to the value of $f(x)$. So, a measurement of a value of the function after the parallel computation will evolve the system as

$$\sum_{x=0}^{2^n-1} |x\rangle|f(x)\rangle|`f(x) =?'\rangle \longrightarrow \sum_{x=0}^{2^n-1} |x\rangle|f(x)\rangle|`f(x) = k'_x\rangle. \tag{10.3}$$

The MWI stipulates that in *all* worlds the measurement will reveal a value of the function. The particular point at which the function is evaluated in a given world will depend on the state of the input register in that world. To an observer, a measurement will reveal the value of the function at a random point. So, it appears that the MWI can maintain that all values of the function are calculated and still explain why only one of those values will be accessible.

10.3 A PROBLEMATIC ALGORITHM

The many-worlds theorist seemingly will also have an easy time explaining why quantum computers can efficiently determine properties of a function that depend on the individual values of the function, e.g. if a function is constant or balanced. Since each individual value was computed in a single step, it is little surprise that global properties can be determined efficiently. Classically, there seems to be no way of determining if an unknown function is constant or balanced without computing more than half of its values. Also, if one is committed to the principle that multiple values of an unknown function must be computed to determine if it is constant or balanced, such behavior gives us evidence in favor of quantum parallelism. Since the MWI seems to explain in what sense each value of a function is computed, quantum parallelism, in particular its use in computing the global properties of a function, seems to support belief in the MWI.

Unfortunately for the MWI, a detailed look at how the global property of the function is computed in a concrete example serves to undermine the explanatory potential of the MWI over other interpretations. Below is an example, taken from Mermin (2000), of a quantum algorithm used to determine if a function is constant or balanced. Consider a function $f : \{0, 1\} \rightarrow \{0, 1\}$. Now, let the initial input and output states, systems i and o respectively, be $(|0\rangle - |1\rangle)/\sqrt{2}$. Allow these qubits to pass through quantum gate U_f. The state, ignoring normalization factors, becomes:

$$|0\rangle_i|f(0)\rangle_o - |1\rangle_i|f(1)\rangle_o - |0\rangle_i|\overline{f}(0)\rangle_o + |1\rangle_i|\overline{f}(1)\rangle_o \tag{10.4}$$

where $\overline{f} = 1 - f$. If the function is constant, $f(0) = f(1)$ and $\overline{f}(0) = \overline{f}(1)$, the state is

$$(|0\rangle_i - |1\rangle_i)(|f(0)\rangle_o - |\overline{f}(0)\rangle_o) \tag{10.5}$$

and if the function is balanced, $f(0) = \overline{f}(1)$ and $f(1) = \overline{f}(0)$, the state is

$$(|0\rangle_i + |1\rangle_i)(|f(0)\rangle_o - |\overline{f}(0)\rangle_o). \tag{10.6}$$

It is plain to see that a measurement on the input state will determine whether the function is constant or balanced. Furthermore, the two possible final states for i are orthogonal and can always be distinguished by a single measurement. The unitary operator that performs such a measurement will evolve a measurement qubit, M, according to the state of the input qubit as follows:

$$(|0\rangle_i - |1\rangle_i)|`f = ?'\rangle_M \rightarrow (|0\rangle_i - |1\rangle_i)|`f = \mathit{constant}'\rangle_M,$$

$$(|0\rangle_i + |1\rangle_i)|`f = ?'\rangle_M \rightarrow (|0\rangle_i + |1\rangle_i)|`f = \mathit{balanced}'\rangle_M \tag{10.7}$$

We can rewrite the state of $i + o + M$ after the computation, but before the measurement as

$$\begin{aligned}(|0\rangle_i - |1\rangle_i)(|f(0)\rangle_o - |\overline{f}(0)\rangle_o)|`f = ?'\rangle_M = \\ |0\rangle_i|f(0)\rangle_o|`f = ?'\rangle_M - |0\rangle_i|\overline{f}(0)\rangle_o|`f = ?'\rangle_M \\ - |1\rangle_i|f(0)\rangle_o|`f = ?'\rangle_M + |1\rangle_i|\overline{f}(0)\rangle_o|`f = ?'\rangle_M\end{aligned} \tag{10.8}$$

or as

$$\begin{aligned}(|0\rangle_i + |1\rangle_i)(|f(0)\rangle_o - |\overline{f}(0)\rangle_o)|`f = ?'\rangle_M = \\ |0\rangle_i|f(0)\rangle_o|`f = ?'\rangle_M - |0\rangle_i|\overline{f}(0)\rangle_o|`f = ?'\rangle_M \\ + |1\rangle_i|f(0)\rangle_o|'f = ?'\rangle_M - |1\rangle_i|\overline{f}(0)\rangle_o|'f = ?'\rangle_M\end{aligned} \tag{10.9}$$

depending on if the function was constant, (10.8) or balanced, (10.9).

If we perform the measurement on i, the measurement device will evolve in one of two ways. If the function is constant, the state evolves to

$$\begin{aligned}(|0\rangle_i - |1\rangle_i)(|f(0)\rangle_o - |\overline{f}(0)\rangle_o)|`f = \mathit{constant}'\rangle_M = \\ |0\rangle_i|f(0)\rangle_o|`f = \mathit{constant}'\rangle_M - |0\rangle_i|\overline{f}(0)\rangle_o|`f = \mathit{constant}'\rangle_M \\ - |1\rangle_i|f(0)\rangle_o|`f = \mathit{constant}'\rangle_M + |1\rangle_i|\overline{f}(0)\rangle_o|`f = \mathit{constant}'\rangle_M\end{aligned} \tag{10.10}$$

If the function is balanced, the state evolves to

$$(|0\rangle_i + |1\rangle_i)(|f(0)\rangle_o - |\overline{f}(0)\rangle_o)|`f = \mathit{balanced}'\rangle_M =$$

$$|0\rangle_i|f(0)\rangle_o|`f = \mathit{balanced}'\rangle_M - |0\rangle_i|\overline{f}(0)\rangle_o|`f = \mathit{balanced}'\rangle_M$$

$$+ |1\rangle_i|f(0)\rangle_o|`f = \mathit{balanced}'\rangle_M - |1\rangle_i|\overline{f}(0)\rangle_o|`f = \mathit{balanced}'\rangle_M \qquad (10.11)$$

The attraction the MWI has with respect to quantum parallelism is its ability to maintain that every value of a function is calculated and explain why it is that only one of those values is accessible to an observer. This ability stems from the fact that the quantum state of each world underwrites the claim that in processes such as (10.1), the function is evaluated at each point. Specifically, in each world, the input register is in an eigenstate that represents the point at which the function is evaluated and the output state is in an eigenstate representing the value of the function at that point. An explanation of the ability of a quantum computer to determine if a function is constant or balanced is not as straightforward.

Assuming that the computational basis is the z-basis, in each world described by the terms of (10.8) and (10.9) the input qubit is in a z-spin eigenstate. To determine if f is constant or balanced, we perform a measurement of spin in the x-direction on the input qubit. A measurement of spin in the x-direction is incompatible with z-spin. So, the x-spin of the input qubit should have no determinate value on the MWI according to the state vector assigned to the world. Thus, each world should dispose an x-spin measurement of the input qubit to be up or down with equal probability. Nonetheless, we see that there is always a sure-fire disposition for the x-spin measurement to indicate correctly that the function is constant or balanced. How might a proponent of the MWI try to explain the disposition in question?[3]

1. The disposition is grounded in a determinate value of some observable of $i + o + M$.
2. That disposition is grounded in some quantum state ascribed to $i + o + M$.
3. Both of the above.

Option 1. Suppose that the $f =?$ observable, x-spin of the input state, has a determinate value. Initially this seems plausible since after all, the function really is constant or balanced and the input system begins and ends the computation in an x-eigenstate. If we maintain this

[3] This argument against the MWI is adapted from Clifton (1996)

option then we are forced to abandon the idea that worlds are in one-to-one correspondence to terms in the state description, a fundamental tenant of the MWI.

We demonstrate this via appeal to the first terms in equations (10.8) and (10.9). Suppose that our function is constant and $f(0) = 0$, the state is then equal to

$$(|0\rangle|0\rangle - |1\rangle|0\rangle - |0\rangle|1\rangle + |1\rangle|1\rangle)|'f =?'\rangle. \tag{10.12}$$

We know from equation (10.5) that the system is in an x-spin eigenstate. So let's make the assumption that there is a definite value for x-spin. This will underwrite the sure-fire disposition of the measurement system to indicate that the function is constant. For consistency, the same type of story must be told about a system's dispositions when the function is balanced. Supposing that the function is balanced and $f(0) = 0$, the state is equal to

$$(|0\rangle|0\rangle - |1\rangle|1\rangle - |0\rangle|1\rangle + |1\rangle|0\rangle)|'f =?'\rangle. \tag{10.13}$$

We will assume that there is a definite value of x-spin, which happens to be in the up direction. Again, this underwrites the disposition of the measurement system to indicate that the function is balanced. But now we are forced into a corner. If the terms of the state of a system in the computational basis are to pick out the worlds, then the first terms in equations (10.12) and (10.13) are identical, but nonetheless they are supposed to describe two very different worlds!

Option 2. One way to fix this problem is to question the assignment of states to worlds in the first terms in (10.12) and (10.13). Let us only consider the $|0\rangle|0\rangle|f =?\rangle$ world in (10.12). For brevity let us refer to it as the (0,0,constant) world. The quantum state of a world is meant to predict the dispositions of all measurement devices in that world if it is to be a complete physical description of that world. The quantum state that will predict the sure-fire disposition of this world to be in a down eigenstate of x-spin is (10.12). But again, if the state vector assigned to the (0,0,constant) world is a complete description of it, it is a trivial matter to find another observable that always has a sure-fire disposition that cannot be accounted for by (10.12). Let us add another measurement device, *M'*, to our setup which measures z-spin of i. The measurement interaction is specified by the following:

$$|0\rangle|`z =?'\rangle \longrightarrow |0\rangle|`z = +'\rangle$$

$$|1\rangle|`z =?'\rangle \longrightarrow |0\rangle|`z = -'\rangle \tag{10.14}$$

From (10.12), it is plain to see that i is not in a eigenstate of z-spin. Nonetheless, there will be a surefire disposition for *M'* to take on the value 'z = +' in the (0,0,constant) world. If we assign a different state to the world, one in which z-spin is always up, we will be unable to explain the disposition of the world to indicate that x-spin is down. There is *no* state that we can assign to the (0,0,constant) world that will make the appropriate predictions about the dispositions of measurement devices.

Option 3. Finally, one might suggest that dispositions as in the last case can be explained by assigning (10.12) to the (0,0,constant) world and further specifying that the value of z-spin in that world has a preexisting value of 0. Surely this is an unacceptable situation for the many worlds theorist to be in. Advocates of MWI often scold others for not taking the Shrödinger evolution seriously. Now, we have the many worlds theorist assigning a state to each world which is not predictively complete without the addition of a hidden variable. That is to say, they are making an addition the Shrödinger evolution of the system. Surely this is unacceptable for them.

10.4 CONCLUSION

Suppose we are looking for an explanation of why we can get away with one quantum computation as opposed to the two needed for a classical computation in order to determine if the function is constant or balanced. The many worlds theorist will suggest that each world computes a different value of the function. As we have seen, if we take talk about computations in different worlds seriously there is no explanation in each individual world as to how it is decided that the function is constant or balanced. Only with recourse to the universal validity of the Schrödinger evolution of the entire quantum system can we explain our success with the quantum algorithm. Now, those who endorse MWI surely would accept this explanation. The universal validity of the Schrödinger evolution is a fundamental assumption for MWIs. However, surely quantum computations as the one above cannot be used as evidence for the MWI. Any interpretation that subscribes to the universal validity of the Schrödinger evolution and uses it to explain the mysterious abilities of quantum computers will be on equal footing.

Note that all difficulties are avoided by simply appealing to the general criteria for a function to be computed. All values are computed in virtue of the functional behavior of the quantum computer. The efficiency of the algorithm can be explained simply by pointing out that the quantum parallelism process resulted in a quantum state appropriately correlated with all values of the function computed. This state has measurable properties that vary depending on if the function is constant or balanced. All extra metaphysical baggage the MWI utilizes beyond the quantum formalism makes an explanation of the quantum computation more difficult than it would be otherwise. I take it that any such consequence is a strike against an interpretation.

APPENDIX A

THE SCHRÖDINGER AND HEISENBERG PICTURES

The empirical predictions of quantum theory come in two forms: expectation values of observables, and possible values of those observables (and their corresponding probabilities). Time evolution in quantum theory is unitary, and there is some mathematical freedom to choose how to determine empirical predictions. We will consider the Schrödinger picture and the Heisenberg picture of evolution in the theory.

Let us consider a quantum system in the state $|\psi(t=0)\rangle = |\psi\rangle$, subject to unitary evolution $U(t)$ that evolves the system to time t. In the Schrödinger picture the state vector evolves with time, $|\psi(t)\rangle = U(t)|\psi\rangle$. The operators that represent observables are constants, $A(t) = A$. To determine expectation values of an arbitrary observable A after the unitary evolution, one simply takes the inner product of that observable with the evolved state vector,

$$\langle A\rangle(t) = \langle\psi|U^\dagger(t)AU(t)|\psi\rangle \qquad (\text{ A.1})$$

Possible measurement results for A are given by the eigenvalues of A. Suppose that $A = \sum_a a|a\rangle\langle a|$, where the a's are eigenvalues of A and the $|a\rangle$'s are their corresponding eigenvectors. The probability to get eigenvalue a after unitary evolution of $|\psi\rangle$ is given by

$$Pr(a,t) = |\langle a|U(t)|\psi\rangle|^2 \,. \qquad (\text{ A.2})$$

In the Heisenberg picture, the quantum state is constant and the observables evolve with time, i. e., $A(t) = U^\dagger(t)AU(t)$ and $|\psi(t)\rangle = |\psi\rangle$. The expectation value of the observable A

at time t is given by

$$\langle A\rangle(t) = \langle\psi|A(t)|\psi\rangle \tag{ A.3}$$

If we substitute $U^\dagger(t)AU(t)$ for $A(t)$ in the above equation, we see equations (A.1) and (A.3) are equivalent. As before, the possible measurement results of an operator are given by the eigenvalues of the operator. The eigenvectors of $A(t)$, the $|a(t)\rangle$'s are given by $U^\dagger(t)|a\rangle$. The probability of getting eigenvalue a is given by

$$Pr(a,t) = |\langle a(t)|\psi\rangle|^2 . \tag{ A.4}$$

If we substitute $|a(t)\rangle = U^\dagger(t)|a\rangle$ in the above equation, we see equations (A.2) and (A.4) are equivalent. So, we conclude that the empirical predictions of the Schrödinger and Heisenberg pictures are equivalent.

In summary, we have, for an operator A and a state $|\psi\rangle$,

Property	Schrödinger Picture	Heisenberg Picture
State	$\|\psi(t)\rangle_S = U(t)\|\psi(0)\rangle$	$\|\psi(t)\rangle_H = \|\psi(0)\rangle$
Operators	$A(t)_S = A$	$A(t)_H = U^\dagger(t)AU(t)$
Eigenvalues	$\|a(t)\rangle_S = \|a\rangle$	$\|a(t)\rangle_H = U^\dagger(t)\|a\rangle$
Expectation values	$\langle A\rangle(t) = \langle\psi(t)\|_S A\|\psi(t)\rangle_S$	$\langle A\rangle(t) = \langle\psi\|_H A(t)\|\psi\rangle_H$
Probabilities	$Pr(a,t) = \|\langle a\|\psi(t)\rangle_S\|^2$	$Pr(a,t) = \|\langle a(t)\|_H\psi\rangle\|^2$

APPENDIX B

DESCRIPTORS DETERMINE EVOLUTION OF ALL QUANTUM MECHANICAL OBSERVABLES

Quantum mechanical observables for n qubits are described by $2^n \times 2^n$ matrices. In what follows we demonstrate how to exploit the mathematical properties of these matrices to efficiently track quantum mechanical evolution. The Pauli group, consisting of 4^n n qubit tensor products of the identity, I, and the Pauli matrices σ_x, σ_y, and σ_z with possible global phases of ± 1, or $\pm i$, spans the set of $2^n \times 2^n$ matrices (Gottesman (1998), 3).[1] Since unitary evolution is linear, tracking the evolution of the Pauli group suffices to determine the evolution of all possible observables on an n qubit system.

Group structure can be further exploited to reduce the number of matrices whose evolution must be determined to specify the evolution of all quantum mechanical observables of the n qubit system. Every group has a set of generators. A set of generators is a subset of a group that can be extended using group operations to "generate" the whole group. Consider the Pauli group for a single qubit. For convenience, let $X = \sigma_x$, and similar for the other Pauli matrices. The Pauli group is $\{\pm I, \pm iI, \pm X, \pm iX, \pm Y, \pm iY, \pm Z, \pm iZ\}$. This entire group can be generated by the elements X and Z using group operations alone.[2] Similarly the entire Pauli group for n qubits can be generated from a subset of the group. Using the notation X_2 for the n-fold tensor product $I \otimes X \otimes I \otimes \cdots \otimes I$, the Pauli group for n qubits

[1] Recall that the Pauli matrices have the properties $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = I$, and $\sigma_x \sigma_y = i\sigma_z$ plus cyclic permutations.

[2] The properties of the matrices under group operations listed in the last footnote make this obvious.

can be generated from the set $\{X_1, \ldots, X_n, Z_1, \ldots, Z_n\}$(Gottesman (1998), 4).

The mapping for time evolution of operators in the Heisenberg picture is a group homomorphism; for operators A and B, $U^\dagger ABU = U^\dagger AUU^\dagger BU$.[3] So, to track how the Pauli group evolves, we need only track the evolution of the generating set for the group. Since the evolution is a group homomorphism the group generated from the evolved generators will be the entire evolved group. Similarly, the mapping for time evolution of operators is linear. Thus, we can construct all evolved operators through linear operations from the evolved group. So, we need only keep track of evolution on a generating subset of matrices of the Pauli group to completely determine the evolution of any $2^n \times 2^n$ matrix. Hence, the evolution of Deutsch and Hayden's descriptors is sufficient to construct the evolution of any observable.

[3] Recall that a mapping $h : G \to G$ is a homomorphism of group G if for all $g, g' \in G$, $h(gg') = h(g)h(g')$.

APPENDIX C

AN ALMOST WELL-ADAPTED CLASSICAL COMPUTER

From section 8.2, the crucial feature of well-adaptedness seems to be condition 2, that the physical dependencies of the computer mirror the logical dependencies of the computational task. It is not obvious that a classical computer (one described by classical physics) can satisfy this condition. In this section an example is given that demonstrates a computer using classical physics can satisfy the constraint on the dynamics.

A variation of the tube computer will satisfy the constraint on the dynamics. (See Figure 7.) Let us call this model the multi-ball tube computer. The multi-tube computer is much like the tube computer, but it only has input slots for every point in the domain of the function to be computed. Similarly, it only has slots that represent partial graphs of the function restricted to individual points from the domain of the function to be computed. A ball in a slot will represent an individual input state. Balls in multiple slots are a state of the machine that represent the multiple inputs, and similar for outputs. If 2^n balls are dropped, one in every slot, at once, we get the simultaneous evaluation of several values of the function.

Clearly the constraint on the dynamics is met. Once the tubes that control the dynamics for each and every state are fixed, the dynamics for the multiple states are determined. The multi-ball tube computer does not satisfy the constraint on representation. One ball is required to compute any individual value, more than one ball is required for more than one value to be computed. Also, just like the tube computer, the slots that are part of the representation of inputs do not make efficient use of space. 2^n slots are required, exponential in n.

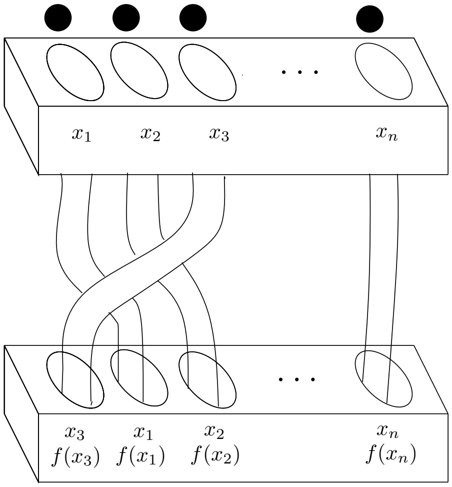

Figure 7: Multi-tube Computer. There an input slot for every individual input for the function f. There is also an output slot possible partial graph restricted to individual values given the domain and range of f. When balls are dropped, tubes direct the balls to the appropriate output slots for successful computation of all values of the function.

BIBLIOGRAPHY

[1] Barnum, H., Caves, C., Fuchs, C., Jozsa, R., Schumacher, B. (2000) "On quantum coding for ensembles of mixed states." preprint quant-ph/0008024

[2] Bennett, C.H.; Brassard, G.; Crepeau, C.; Jozsa, R; Peres, A; and Wooters, W.K. (1993), "Teleporting of unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels", *Physical Review Letters*, 70, 1895-1898.

[3] Caves, C. M., Fuchs, C. A. (1996). "Quantum information: how much information in a state vector?" preprint quant-ph/9601025.

[4] Bell, J. S. (1964) "On the Eintein-Podolsky-Rosen paradox." *Physics* 1: 195-200

[5] Clifton, R.K. (1996) "On What Being a World Takes Away" *PSA* 63 pp. S151-S158

[6] Deutsch, D. (1997) *The Fabric of Reality* Penguin

[7] Deutsch, D., Hayden, P. (1999). "Information flow in entangled quantum subsystems" preprint quant-ph/9906007.

[8] Dowe, P. (2000) *Physical Causation*, Cambridge: Cambridge University Press.

[9] Dretske, F. (1981). *Knowledge and the Flow of Information.* Oxford, Blackwells

[10] Duwell, A. (2000) "Explaining information transfer in quantum telepotation" *Philosophy of Science* **68** S288-S300

[11] Duwell, A. (2003) "Quantum information does not exist". *Stud. Hist. Phil. Mod. Phys.* (September 2003) pp. 479-499

[12] Faddeev, D. (1957). In Grell, H. (Ed.) *Arbeiten zum informationstheorie I* (pp. 88-91). Berlin: Deutscher Verlag er Wissenschaften.

[13] Fuchs, C.A. (2001). "Notes on a Paulian idea: foundational, historical, anecdotal, and forward-looking thoughts on the quantum" preprint quant-ph/0105039

[14] Fuchs, C.A. (2002). "Quantum States: What the Hell Are They?" http://netlib.bell-labs.com/who/cafuchs/PhaseTransition.pdf

[15] Gottesman, D. (1998) “The Heisenberg Representation of Quantum Computers” quant-ph/9807006

[16] Hausladen, P., Jozsa, R., Schumacher, B., Westmoreland, M., Wooters, W.K. (1996). “Classical information capacity of a quantum channel” *Physical Review A*, 54, No. 3, 1869-1876.

[17] Jozsa, R.(1998). “Quantum information and its properties”, In Lo, Hoi-Kwong, L., Popescu, S., Spiller, T. *Introduction to quantum computation and information*, Singapore: World Scientific.

[18] Jozsa, R. (2000) “Quantum Algorithms” in Boumeester, D., Ekert, A., Zeilinger, A. (Eds.) *The Physics of Quantum Information: Quantum Cryptography, Quantum Teleportation, Quantum Computation*, Germany: Springer

[19] Jozsa, R., Linden, N. (2002) “On the role of entanglement in quantum computational speed-up” quant-ph/0201143

[20] Mermin, N.D. (2000) ”The Contemplation of Quantum Computation” http://www.aip.org/pt/vol-53/iss-7/p11.html

[21] Nielsen, M.A., Chuang, I.L. (2000). *Quantum computation and quantum information*, Cambridge: Cambridge University Press.

[22] Shannon, C.E. and Weaver, W. (1949). *The mathematical theory of communication*, Urbana: The University of Illinois Press.

[23] Shor, P. (1994) “Algorithms for quantum computation: discrete log and factoring”, *Proceedings of the 35th Annual Symposium on the Foundations of Computer Science*, S. Goldwasser (editor), IEEE Computer Society Press, Los Alamitos, 1994, pp. 124-134.

[24] Schumacher, B. (1995). “Quantum coding” *Physical Review A* 51, 2738-2747

[25] Steane, A. (2003) “A quantum computer needs only one universe” *Stud. Hist. Phil. Mod. Phys.* 34, 469-478

[26] Timpson, C. (2003) “Nonlocality and information flow: The approach of Deutsch and Hayden” quant-ph/0312155

[27] Timpson, C. (2003) “On the supposed conceptual inadequacy of the Shannon Information in Quantum Mechanics” *Stud. Hist. Phil. Mod. Phys.*

[28] Uffink, J. (1990) Measures of Uncertainty and the Uncertainty Principle. Unpublished Ph.D. dissertation, University of Utrecht

[29] Vollmer, H. (1999) *Introduction to Circuit Complexity: A Uniform Approach* Italy: Springer-Verlag