Perspectives on Quantum Non-Locality

Daniel G. Collins

H. H. Wills Physics Laboratory
University of Bristol
&
Hewlett-Packard Laboratories
Bristol

A thesis submitted to the University of Bristol in
accordance with the requirements of the degree of
Doctor of Philosophy in the Faculty of Science
June 2002

Abstract

Quantum non-locality is a strange property of long distance quantum mechanics, which can only be described by classical physics using faster than light communication. Despite the fact that it cannot be used to build a superluminal phone line, it leads to many interesting effects, such as the teleportation of quantum mechanical states, the ability to send messages whose secrecy is guaranteed by the status of quantum mechanics as a fundamental theory, and the future possibility to perform certain computations in a fundamentally new and faster way. Furthermore, the very existence of quantum non-locality poses fundamental questions about the nature of reality. I explore this phenomenon from various perspectives.

I find close classical analogues for many features of quantum non-locality, in particular an analogue of the manipulation of bi-partite pure state entanglement under local operations and classical communication. I describe many new Bell inequalities which local classical theories must satisfy, but quantum mechanics does not. I give Bell inequalities for two quNits which are quite resistant to the presence of white noise. I show that non-local correlations exist even in states with very high fractions of noise, so long as they contain enough entanglement. I show that simulating n-body quantum mechanical correlations using superluminal classical communication requires communication which links all the bodies. I describe a new loophole in Bell's proof of quantum non-locality, based upon the possibility of a local model with memory. I then show how to close the loophole using a simple modification of Bell's inequality. Finally, I introduce a notion of the non-locality of quantum mechanical operations on multi-partite systems. This describes the non-local content in terms of the entanglement and classical communication required to implement such operations.

These perspectives give us qualitative and quantitative descriptions of many features of quantum non-locality, and so give us a greater understanding of the quantum mechanical world.

Thanks

This thesis will not fundamentally change our view of science, or even of physics, but the three years I have spent making it has changed me. The greatest thanks for this goes to my adviser, Sandu Popescu, whose clarity and simplicity of thought has done much to dispel the complexity and mechanicity which the maths course in Cambridge tried so hard to instill inside of me. Scientific thanks also go to Noah Linden and Richard Jozsa, for showing me there are many ways to do good science, to Adrian Kent, for exciting me about quantum information in the first place, to Daniel Oi, for talking physics with such enthusiasm, to Serge Massar, Benni Reznik, Nicolas Gisin, Valerio Scarani, Seth Lloyd and Vittorio Giovannetti for hosting me to work on science in such beautiful locations, and to Matthew Leifer, Jonathan Barrett, Simone Severini, Stuart Presnell, Andreas Winter and Leah Henderson for convincing me that none of us really understands quantum mechanics. More thanks go to the whole of the rest of my year, Ben Powell, who kept me sane when things seemed so hard at the start, and to the other physicists who persuaded me that I know almost no physics, Mark Dennis, Joe Brader, Andy Archer, Denzil Rodriguez, Danny Jervis, David Roberts, Jamie Walker and Andrew Mcdonald. Many friends and fellow PhD sufferers kept me from falling too far into physics, Chris Bradley, Matt Downton, Rory O'Neill, Gavin McStay, Krupa Pattni, Dan Simpson (whose little sister *is* called Lisa), Ailsa Powell and Paula Atkinson. Major thanks goes to my oldest and dearest friends, who show me how wide the world is and how many good things there are to do, Sejal Haria, James Montgomery, Tjun Tang, Mark Ferguson, Michael Bunter and Paul Heaton. Finally thanks go to my mum, dad and brother, who made me what I am, for which I am glad.

Declaration

I declare that the work in this thesis was carried out in accordance with the regulations of the University of Bristol, between October 1999 and June 2002. The work is original except where special reference is made to the work of others. It is the result of the author's investigations under the supervision of Professor Sandu Popescu, in collaboration with other scientists where indicated. No part of this work has been submitted previously for a degree at this or any other University either in the United Kingdom or overseas. Any views expressed in the thesis are those of the author and in no way represent those of the University of Bristol.

"*The eye sees only what the mind is prepared to comprehend.*"

Henri L. Bergson

Contents

List of Figures

List of Tables

Chapter 1

Overview

1.1 Background

In recent years, fundamental quantum mechanics has predicted several exciting new phenomena, which include quantum computation[1], teleportation[2] and quantum cryptography[3]. Furthermore, macroscopic objects have been shown to have microscopic quantum mechanical behavior. Buckyballs[4] and gasses of 10^{12} Cesium atoms[5] have been shown to exhibit wave-like interference. Such progress has made it more important than ever to have a good understanding of fundamental quantum mechanical behavior. One of the most striking aspects of large scale quantum mechanics is known as quantum non-locality. This is the fact that events in two distant regions of space can be linked with one another in ways which classically can only be explained using faster than light communication. This thesis is devoted to non-locality, which is one of the clearest ways in which the quantum world differs from the classical one.

Quantum non-locality occurs in entangled quantum mechanical systems, ie. those which are now spatially separated but which have interacted in the past. That such systems have strange properties was discussed in 1935 by Einstein, Podolsky and Rosen (EPR)[6]. They used such systems to suggest that quantum mechanics is not a fundamental theory, in the following way. In quantum mechanics, a particle can have a well defined position, or a well defined momentum, but not both. However, if

we entangle a second particle with the first, then depending upon what measurement we make on the second particle, we can learn either the position or the momentum of the first. Since we can do this when the second particle is spacially separated from the first, ie. without touching the first in any way, they reasoned that the first particle must already have a well defined position, and a well defined momentum. Quantum mechanics does not describe particles with well defined positions and momenta, and thus EPR claimed that it cannot be a fundamental theory. They said that there should be a deeper theory which describes particles with well defined position and momenta, and which, in order to coincide with quantum mechanics, describes how the world stops us knowing both simultaneously. Notice that locality is implicit in this argument, since we assume that measuring one particle does not affect the other.

Despite much debate, no-one was able to find such a theory, or a universally accepted critique of their argument. Lacking a solution, many physicists settled for the pragmatic point of view that whilst they did not understand the foundations of quantum theory, they could use it to make many successful predictions about various experiments. Furthermore, at that time the experiments which were most likely to test quantum non-locality were not feasible, and the problems in the foundations became to be considered a topic for philosophers.

In 1957, Aharonov and Bohm[7] used the data of a particle physics experiment to show that two particles in an entangled quantum mechanical state can be moved distant from one another and still remain entangled. They had demonstrated the existence of long distance quantum mechanical correlations for the first time. In a seminal work in 1964[8], Bell proved that the predictions of quantum mechanics for a certain gedanken experiment could not be reproduced in the way EPR desired, using a deeper local classical theory. Thus quantum mechanics must necessarily be non-local! This seemed to be a direct conflict between the two great theories of the early 20^{th} century, quantum mechanics and relativity. Since both theories were very well tested, and had appeared to be universally valid, this conflict was a great surprise and worry. However, quantum mechanics did not seem to explicitly break relativity, since it did not allow for faster than light communication. The two

theories seemed to sit uneasily side by side, without destroying one another. Because of this, and because quantum mechanics worked so well, it was some time before Bell's groundbreaking paper was widely appreciated. However inequalities better suited to experimentally testing for Bell's non-locality was given by Clauser, Horne, Shimony and Holt in 1969 [9], and by Clauser and Holt in 1974 [10]. Successively improved experiments have tested for non-locality [11]. Whilst each one gives close agreement with quantum mechanics, none have so far conclusively ruled out all possible local hidden variable theories, a fact which will be discussed in chapter 6.

In recent years, there has been a resurgence in interest in quantum non-locality. New, simpler proofs of the non-locality of quantum theory have been invented [12, 13]. Non-local correlations have been shown to exist in all entangled pure states, thus being generic [14, 15, 16]. Many new features of non-locality have been discovered. For example teleportation, superdense coding [17], remote state preparation [18], and the use of entanglement to reduce the number of bits needed for communication in certain distributed tasks, in the so called communication complexity scenario [19]. Furthermore, the recently discovered quantum cryptography and quantum computation seem related to non-locality. The study of entanglement, ie. quantum non-local states, has become fashionable. The introduction of information theory has led to methods for processing and transmitting such states, for example Schumacher compression [20], entanglement concentration and dilution [21] and entanglement purification [22]. These methods lead to the beautiful result that although there are many different pure, bi-partite entangled states, there is only one form of entanglement contained within them, which can be quantified by a single number. Entanglement in n-party systems has subsequently been shown to be more complicated, with new, inequivalent forms at every n[23, 24].

In parallel with these theoretical advances have come many experimental ones. It is now relatively easy to perform a version of Bell's gedanken experiment[25]. Multi-partite entangled states have been produced [26], as have prototype quantum cryptographic devices[3]. Teleportation[27, 28] has been performed. Prototype quantum computers have been demonstrated[29], and a huge effort is now devoted to increasing their size and power. In addition, the non-local collapse of the wave-

function has been tested[30].

Despite this progress, many basic issues remain unresolved. Perhaps the most fundamental question is "how does non-locality arise?" Whilst this remains unanswered, we cannot feel that we truly understand quantum mechanics. More approachable problems include the study of almost all of the more recent aspects of non-locality, the study of which is only partially developed. For example, mixed state and multi-partite entanglement are only partially understood. The precise link between Bell's non-locality and entanglement is not known: do all mixed entangled states contain non-local correlations? What is the best way to extract non-local correlations from a mixed state? Where does the power of the quantum computer come from? And what other tasks are possible? Can the loopholes (see chapter 6) in the current experimental tests of non-locality be removed in a clever way? Or do we need to wait for technology to improve to close them? These and other questions are the impetus for a huge effort to understand the foundations of quantum theory, and in particular quantum non-locality.

1.2 New Perspectives

In this thesis I focus upon quantum non-locality, trying to build a better intuition about its behaviour. I have several new perspectives, namely a) entanglement has a close classical analogue; b) Bell inequalities can be viewed as frustrated networks of correlations; c) non-local correlations can be detected even in states with an arbitrarily large fraction of noise; d) quantum mechanics contains true n-party non-locality; e) Bell's gedanken experiment has a new loophole, which we call the memory loophole; f) quantum operations have a notion of non-locality. I explain these six results in more detail below.

In my first perspective, in chapter 2, I show that entangled quantum mechanical states have a close classical analogue, namely secret classical correlations.

A quantum state of two particles ρ_{AB}, one held by Alice and one by Bob, is said to be entangled if and only if it cannot be written in the form $\sum_i p_i \rho_A^i \otimes \rho_B^i$, where the p_i's are probabilities. The idea behind the definition is that entangled

states are precisely those which cannot be made by Alice and Bob communicating classically, and using local actions (ie. measurements and local unitary operations just on Alice's particle, or just on Bob's particle). To make an entangled state requires coherent quantum interaction between the two particles. I describe an analogy between this and secret classical correlations. A secret classical correlation is a sample from a known probability distribution $P(X_A, X_B, X_E)$, where X_A goes to Alice, X_B to Bob, and X_E to Eve. This could arise by a secret coin toss which Alice and Bob see, but Eve does not.

That there exists a classical analogue of entanglement is very surprising, since entanglement is generally believed to be the most representative aspect of quantum mechanics, ie. that part which has no classical analogue. The fundamental analogy stems from the behavior of quantum entanglement under local operations and classical communication and the behavior of secret correlations under local operations and public communication. A large number of derived analogies follow. In particular teleportation is analogous to the one-time-pad[31], the concept of "pure state" exists in the classical domain, and entanglement concentration and dilution are essentially classical secrecy protocols. Further, for every single copy pure state secret classical correlation manipulation there is an analogous single copy pure state entanglement manipulation, and the majorization results[32, 33] which describe when such entanglement manipulations are possible are reproduced in the classical setting.

This analogy allows one to import questions from the quantum domain into the classical one, and vice-versa, helping to get a better understanding of both. Also, by identifying classical aspects of quantum entanglement it allows one to identify those aspects of entanglement which are uniquely quantum mechanical. I identify two such features: one is superdense coding, and the other is the Bell inequality, whose importance makes it the subject of four subsequent chapters. This work was published previously in Physical Review A[34] in collaboration with Sandu Popescu.

In chapter 3, I give a new interpretation of Bell inequalities, the basic tool for detecting non-locality, in terms of frustrated networks of correlations. I use this to derive new Bell inequalities for two systems with arbitrarily high numbers of levels.

The simplest idea of non-locality is that moving something in one region of space

causes something in a distant region of space to instantaneously move. Quantum mechanics does not possess such non-locality. This is good, since such non-locality would be in direct contradiction with special relativity, which assumes that nothing, not even forces, can move faster than light. Bell imagined a more complicated experiment in which two experimenters in different regions of space move things and watch the correlations between the instantaneous motion of the particles in their regions. The experiment is arranged so that there is not enough time for light to travel from one region to the other before the motion is observed. In the local classical picture the motion of the particles in each region can only depend upon the actions of the local experimenter, and upon any pre-arranged instructions. This puts constraints upon the type of correlated motion which such models can give. Bell formulated these constraints as inequalities (now called Bell inequalities) which any local classical theory must satisfy. He showed that quantum mechanics is not subject to the same constraints, and can in fact give correlations which violate such an inequality. Thus quantum mechanics predicts non-local correlations[1].

I introduce an interpretation of Bell Inequalities in terms of frustrated networks of correlations. This allows a clear understanding of the limitations upon the correlations which local classical theories can produce. It allows a better intuition for the kinds of non-local correlations quantum mechanics may produce. One can then easily construct many new Bell inequalities, a task which has previously been very difficult. For example, very few inequalities have previously been constructed even for two systems with more than two levels[38, 39, 40]. I construct a family of Bell inequalities for two n-level systems which are violated by the maximally entangled states. Since there are a huge number of possible Bell inequalities which could be created, and no complete classification is known, it is helpful to have a criterion to select some interesting ones. I use the idea[41] of adding a fraction of white noise to the maximally entangled (pure) state, and finding how much noise can be added whilst still retaining non-local correlations. The inequality which detects non-local

[1]By "classical" I am assuming that we don't believe in a "many worlds" scenario[35]. In such a scenario a local model can be made, for example see [36]. Unfortunately the many worlds interpretation has certain other undesirable features[37].

correlations in the presence of the largest amount of noise is the best one. My inequalities seem to be optimal, since they give an analytic description of previous numerical work, and generalise the previous work to arbitrarily large numbers of levels. These results were obtained in collaboration with Nicolas Gisin, Noah Linden, Serge Massar, and Sandu Popescu, and have previously appeared in Physical Review Letters[42].

In chapter 4, I re-examine the resistance to noise discussed in chapter 3. The numerical results presented in [41, 43] suggest that bi-partite entangled states with more than one third part noise do not violate any Bell inequality, even for systems with large numbers of levels. Thus it seems that non-local correlations are not robust against noise. This is surprising since the amount of entanglement in the maximally entangled state of two n-level systems is $logn$ e-bits, which increases with n. Adding only a fixed fraction of noise to such a state will, for large enough n, yield states which are still entangled. Thus *entanglement* is robust against noise. One might have thought that non-local correlations were also robust.

I argue that the class of gedanken-experiments previously considered was too restrictive to detect the non-locality in such states. The class contained only measurements which can be performed at a single time. By considering a more general class of gedanken-experiments involving *sequences* of measurements[44], I show that the non-local correlations are, similar to entanglement, robust under the addition of noise. Sequences of measurements are not interesting from a purely quantum mechanical point of view, since they behave very much like single time measurements. However they place further restrictions upon the allowed possible local classical theories. This allows us to demonstrate non-local correlations in states which behave locally under single time measurements. Note that the sequences of measurements considered here are different from the sequences of measurements considered in temporal Bell inequalities [45]. The latter do not assume locality, but instead derive a contradiction with quantum mechanics by assuming that one should be able to determine the state of a classical object without any influence being exerted upon the object, and hence without the object knowing it is being observed and hence without altering the future behaviour of the object. This work was performed under

the supervision of Sandu Popescu, and published in Journal of Physics A: Math and General[46].

In chapter 5, I consider the non-locality of more than two spatially separated quantum systems. Bell showed that entangled states of such systems produce correlations which cannot be reproduced by any local classical theory without superluminal communication. However, Svetlichny suggested that one could imagine reproducing such correlations using, in each trial of the experiment, superluminal communication between just two of the systems [47]. Any states which allow such a model contain only bi-partite non-locality: only those states whose correlations cannot be simulated in such a way contain true multi-partite non-locality. This concept of multi-partite non-locality is *a priori* different from that of entanglement: we do not know whether all multi-partite entangled states generate multi-partite non-local correlations. I use this idea to study the structure of correlations in n-partite systems. Generalising a result of Svetlichny[47] on three-partite non-locality, I give an inequality for n-partite correlations which demonstrates that n-partite quantum mechanics contains true n-partite non-locality. One family of states which contain such non-locality are the Schrödinger cat states. These are the n qubit state $\frac{1}{\sqrt{n}}(|0\rangle_1 |0\rangle_2 ... |0\rangle_n + |1\rangle_1 |1\rangle_2 .. |1\rangle_n)$, where each qubit is in a different location. This work was performed in collaboration with Nicolas Gisin, Sandu Popescu, David Roberts and Valerio Scarani, and previously appeared in Physical Review Letters[48].

In chapter 6, I re-examine Bell's gedanken experiment for detecting non-locality, and show that it has a previously unnoticed loophole, called the memory loophole. This is based upon the fact that any experiment has to be repeated many times to obtain useful statistics. In principle, a local classical model could remember what happens in the first $n-1$ trials, and use this to determine what it will do in the n^{th} trial. Thus, using memory, it may be able to bias the statistics and fool us into thinking that the world is non-local. This loophole is disturbing since it is based upon a fundamental flaw in Bell's gedanken experiment, unlike the other loopholes (such as the detection loophole[49]) which are based upon our practice of implementing modified versions of Bell's gedanken experiment due to current

technological limitations.

One way to avoid the memory loophole is to perform all n trials at the same time, each in a region of space distant from all the others. However, this is extremely difficult experimentally, and so I study the effect of memory on the usual sequential gedanken experiment. I prove that the memory loophole allows a systematic violation of Bell's inequality. However, the maximal violation is small, and is of order $\frac{1}{\sqrt{n}}$ as the number of trials, n, becomes large. In practice violations of close to the quantum mechanical limit of $2\sqrt{2}$ are seen (neglecting the effect of the other loopholes), and so the memory loophole does not change our conclusion that the world is (up to the other loopholes) non-local. However it does reduce slightly the number of standard deviations by which the inequality is violated. I present a linearised version of the Bell inequality which is unaffected by the presence of memory, and therefore is better suited for testing experimental data. Analysing previous experiments with this linearised version gives essentially the same number of standard deviations as were found by previous analysis using the standard Bell inequality. Thus the memory loophole does not make a difference to our final conclusions about the non-locality of the world. This work was carried out in collaboration with Jonathan Barrett, Lucien Hardy, Adrian Kent and Sandu Popescu, and has been accepted for publication in Physical Review A[50].

In chapter 7, I suggest a notion of the non-locality of quantum operations. This mirrors the non-locality of quantum states. For example, imagine that Alice and Bob are spatially separated, and each holds a qubit. The qubits begin in an unknown, possibly entangled, quantum state. Alice and Bob would like to perform a SWAP operation, which interchanges the state of Alice and Bob's qubits. It is not possible to perform this operation by local means alone. However, if they are allowed entanglement and classical communication, they can perform the operation. One method is for Alice to teleport her qubit to Bob, and Bob to teleport his qubit to Alice. I propose quantifying the non-locality of a general bi-partite operation by the amount of resources, both entanglement and classical communication, required to perform the operation. I discuss a number of operations for which I find the optimal amount of resources required. I also show that the creation of an operation

from resources is irreversible: once we have built a black box which performs the operation, we may be able to retrieve the entanglement *or* the classical communication, but certainly not both. I also show how entanglement can be used to catalyse classical communication from a quantum action. That is, entanglement can be used to perform otherwise impossible classical communication, without being destroyed in the process. This work was performed in collaboration with Noah Linden and Sandu Popescu, and has appeared in Physical Review A[51]. Two very closely related independent papers are by A. Chefles, C. R. Gilson and S. M. Barnett[52], and by J. Eisert, K. Jacobs, P. Papadopolous and M. B. Plenio[53].

My work suggests that quantum non-locality is not as strange or spooky as has often been suggested, and that a good understanding of its properties can be obtained. Different aspects require different points of view, some of which are presented here. Armed with these viewpoints, we can try to press forward into the quantum world and discover what it really means, and what possibilities it allows. In the next few chapters I elaborate on my perspectives, and show how they shed light upon quantum non-locality.

Chapter 2

Classical Analogue of Entanglement

2.1 Introduction

Quantum non-locality is considered to be one of, if not the most, representative aspects of quantum mechanics. It is one of the clearest ways in which quantum mechanics differs from classical mechanics. The most familiar manifestation of non-locality lies in entangled states. These are states of two (or more) spatially separated systems, one held by Alice and one by Bob, which cannot be made using local actions (by which we mean actions performed by either Alice or Bob on their system alone) and classical communication. Quite surprisingly I found, in collaboration with Sandu Popescu[34], that there exists a quite close classical analogue of quantum entanglement, namely *secret classical correlations*.

Our motivation in looking for a classical analogue of quantum entanglement was two-fold. Firstly, such an analogy allows us to identify aspects of quantum entanglement which were hitherto considered to be purely quantum but which are in fact not quantum at all. Indeed, all those aspects of entanglement which are common with the classical analogue, are not of a quantum nature. As a corollary we also get a better understanding of what are the true quantum features of quantum entanglement. Secondly, this analogy allows one to transfer questions from quantum entanglement

to the classical domain (classical information cryptography) and vice-versa and thus lead to a better understanding of both subjects. In fact, the inspiration for our work stems from the work of N. Gisin and S. Wolf [54] which asked if there is a classical analogue of bound entanglement.

The analogy we suggest is summarized in table 2.1.

Table 2.1: The Fundamental Analogy

quantum entanglement	—	secret classical correlations
quantum communication	—	secret classical communication
classical communication	—	public classical communication
local actions	—	local actions

Thus we suggest that a classical analogue of a pair of entangled particles is that of one sample of two secret, correlated, random variables (one at each remote party). Here by secret communication we mean communication through a channel to which an eavesdropper has no access. By public communication we understand communication through a channel to which an eavesdropper has full access (can hear everything), but cannot alter the messages sent, nor introduce new messages. Finally, in the quantum context by local actions we understand Alice or Bob subjecting their own system to unitary evolutions as well as to measurements and other non-unitary evolutions. The classical analogue of unitary transformations is that of replacing the value of the original random variable by some new value related to the old one by a one-to-one function, while the analogue of the case of quantum non-unitary evolutions is that of transformation by non bijective functions. [1]

[1]Note that when we replace the original value of the random variable by another via a non-bijective function, we consider that we actually erase the original information, so information is lost. This is completely analogous to what happens in the quantum case. Of course, one may argue that in neither case information is lost. For example, in the non-collapse interpretations of the quantum case all we have is an entanglement of the measured system with the measuring device; this entanglement however involves so many degrees of freedom that it cannot be reversed. Similarly, erasing say pencil markings from a paper still preserves the original information in some

The main idea of this analogy is that as with quantum entanglement, secret classical correlations act as a (fungible) *resource* and obey a "second law of thermodynamics" principle - the amount of secrecy doesn't increase under LOPC (local actions and public communication).

The modern paradigm is that of quantum non-locality as a *resource* as we describe below.

- Non-local correlations between two or more remote parties can be created by quantum communication, i.e. by sending quantum particles from a common source to the parties, or from one party to another.

- *Second law of thermodynamics*: The amount of non-locality between the remote parties cannot be increased by local actions and/or classical communication (LOCC).

  Indeed, one can view this statement as the very *definition* of what non-locality is.

- The remote parties can, by local actions and classical communication, transform non-locality from one form into another.

  For example, suppose two parties, Alice and Bob, have a large number of pairs of particles, each pair in some pure, non-maximally entangled state, $|\psi\rangle_{AB} = \sqrt{p}\,|0\rangle_A\,|0\rangle_B + \sqrt{1-p}\,|1\rangle_A\,|1\rangle_B$, where $0 < p < \frac{1}{2}$. By appropriate actions [21, 55] they can end up with a smaller number of pairs each in the maximally entangled state $|\phi\rangle_{AB} = \frac{1}{\sqrt{2}}(|0\rangle_A\,|0\rangle_B + |1\rangle_A\,|1\rangle_B)$. In effect, at least in the case of bi-partite pure states, non-locality is absolutely fungible - any form can be transformed into any other, and the transformation is reversible. Thus it doesn't really matter in which form the parties are supplied with non-locality, they can always convert it into the form which is required for implementing the specific task (for example teleportation) they want to do.

subtle arrangement of the graphite granules mixed with bits of paper and erasing gum, but this involves so many degrees of freedom that the original information cannot be recovered.

- Non-locality is consumed for producing useful tasks (teleportation, super-dense coding, remote implementation of joint unitary transformations [52, 51, 53], etc.).

As with quantum non-local correlations, secret correlations are also a resource.

- Secret correlations can be established between remote parties by secret communication.

- "Second law of thermodynamics": The amount of secret correlations cannot be increased by local actions and/or public communication (LOPC)[2].

  In fact, as with the case of non-locality, we can take this law to be the very definition of the amount of secret correlations, i.e. the amount of secret correlations between remote parties is that part of their correlations which cannot be increased by local actions and public classical communication.

- The remote parties can, by local actions and public communication, transform secret correlations from one form into another.

- Analogous to entanglement, secret correlations are a fungible resource - they can be stored, transformed from one form into another, and can be consumed to perform useful tasks, such as secret communication via the one time pad [31].

The possibility of transforming entanglement from one form to another allows us to obtain a *quantitative* definition of entanglement for pure, bi-partite states. We say that Alice and Bob have one $e-bit_{AB}$ for every copy of the state $\frac{1}{\sqrt{2}}(|0\rangle_A |0\rangle_B + |1\rangle_A |1\rangle_B)$ which they can reversibly produce using LOCC. We can quantify the classical communication between Alice and Bob by the number of bits (0's or 1's) they send. Similarly, we can quantify quantum communication by the number of qubits (two level quantum systems) that they send.

[2]In everyday practice, secret messages are exchanged by public communication by so called "public key distribution" protocols. We do not consider here this case since these are only pseudo secret messages - their secrecy is based on encoding which is difficult to decode due to computational complexity; in principle however an eavesdropper could decode the message.

We can quantify the amount of secrecy between Alice and Bob in an analogous way:

Table 2.2: Shared, Undirected Resource

$e - bit_{AB}$	—	shared secret bit_{AB}

Table 2.3: Directed Resources

$qubit_{A\to B}$	—	secret $bit_{A\to B}$
classical $bit_{A\to B}$	—	public classical $bit_{A\to B}$

This quantitative description of entanglement allows us to extend the above version of the second law for non-local correlations to allow for quantum communication, catalysis, etc.. For example [55] " By local actions, classical communication and exchange of n q-bits, the amount of non-locality between remote parties cannot be increased by more that n e-bits". Analogously we can extend the second law for secret classical correlations to allow for secret communication, catalysis, etc. For example " By local actions, public communication and exchange of n secret bits, the amount of secret correlations between remote parties cannot be increased by more that n secret correlation bits".

The situation of multi-partite secret correlations is more complicated, as is the situation of multi-partite entanglement. It is now clear that there are many different, irreducible, types of multi-partite entanglement [23], [24]; this is also the case for secret correlations.

At this point it is legitimate to ask what is the role of secrecy. That is, why do we consider *secret* classical correlations to be the analogue of entanglement and not simply *any* classical correlations. There are two main reasons. First of all, while such an analogy is certainly possible, it would be rather uninteresting. Indeed, one of the main aspects of manipulating entanglement is that there is a way in which the different parties may communicate (classical communication) which doesn't increase

the amount of entanglement. Similarly in the case of secret classical correlations, public communication doesn't increase the amount of secrecy. In the case of arbitrary classical correlations however there is no way in which the remote parties could communicate and not increase the correlations. So when trying to build an LOCC ("local operations and classical communications") analogue in the case of arbitrary classical correlations we have no choice but to completely eliminate the communication, which leads to a very uninteresting situation.

The second reason is far more profound. Consider for example two parties, Alice and Bob who share, say, a maximally entangled state $|\Psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle)$. Suppose now that Alice and Bob "degrade" the state by "erasing" the entanglement. They can do this *in a minimal way* by, say, Alice randomizing the phase of her basis state vectors $\{|0\rangle, |1\rangle\}$. Then Alice and Bob will be left with a mixture of $\frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle)$ and $\frac{1}{\sqrt{2}}(|0\rangle|0\rangle - |1\rangle|1\rangle)$ with equal probabilities. This mixture contains no entanglement (it is equivalent to an equal mixture of $|0\rangle|0\rangle$ and $|1\rangle|1\rangle$) but contains secret correlations between Alice and Bob. Thus secret correlations are in fact very closely related to entanglement.

The analogies described above are the "fundamental" analogies. From them follow an entire set of derived analogies. We would like to emphasize however that it is only the fundamental analogies (such as the behavior under LOCC/LOPC) which have truly deep significance and that one shouldn't expect the derived analogies to be very close (though many of them are). Derived analogies are summarized in table 2.4.

2.2 Quantum states and classical analogues

In the previous section we suggested that classical secret correlations are a good analogue for quantum entanglement. The basis of the analogy is the similar behavior of secret correlation and quantum entanglement under LOPC/LOCC. To make the analogy more detailed and to obtain the "derived" analogies mentioned above we need to define more precisely the analogy between quantum states and secret correlations.

Table 2.4: Derived Analogies

teleportation	—	one-time pad
entanglement concentration	—	secret correlation concentration
entanglement dilution	—	secret correlation dilution
entanglement purification	—	classical privacy amplification
single copy transformations	—	single copy transformations
catalytic transformations	—	catalytic transformations
bound entanglement	—	bound information ?

Consider two remote parties, Alice and Bob. A general quantum state is described by a density matrix ρ_{AB} or, equivalently, by a pure state Ψ_{ABE} in which A and B are entangled with a third party, the "environment". The classical equivalent of the general quantum state is a probability distribution $P(X_A, X_B, X_E)$ where X_A, X_B and X_E are random variables known to Alice, Bob and Eve (the eavesdropper) respectively. One copy of a quantum state Ψ_{ABE} corresponds to one sample of the probability distribution $P(X_A, X_B, X_E)$.

A quantum bi-partite pure state can always be written in the Schmidt basis [56] as

$$|\psi\rangle_{AB} = \sum_i \sqrt{p_i}\, |i\rangle_A\, |i\rangle_B\,. \tag{2.2.1}$$

If Alice and Bob measure their particles in the Schmidt basis then they get correlated random variables, X_A and X_B, which come according to the distribution $p(X_A = i, X_B = j) = \delta_{ij} p_i$. In other words, they both get the same sample from a random variable $X \sim \{p_i\}$. Furthermore, the values of X_A and X_B are secret - there is no third party E who knows them. We propose classical distributions of this form as the classical "pure" state. That is, a bi-partite classical pure state is a distribution

$$p(X_A = i, X_B = j, X_E = k) = \delta_{ij} p_i \tilde{P}(E_k) \tag{2.2.2}$$

where $\tilde{P}(E_k)$ is the distribution of eavesdropper's variable X_E and is completely

irrelevant, except for the fact that it is completely uncorrelated to the distribution of X_A and X_B[3]. Strictly speaking, we propose (2.2.2) as the classical analogue of the pure state Schmidt decomposition, and any classical state which is locally equivalent, ie. can be transformed into the above form by local, one-to-one mappings (the equivalent of local unitaries) we consider to be a pure state.

Another interesting case is that of distributions of the form $p(X_A = i, X_B = j, X_E = k) = P(X_A = i, X_B = j)\tilde{P}(E_k)$ in which E is completely uncorrelated with A and B, but A and B are not completely correlated with each other. Such a distribution is obtained when Alice and/or Bob measure a quantum pure state in some other basis than the Schmidt one. Such a distribution has some characteristics of a pure state and some characteristics of a mixed state. We will discuss in more detail this case in section 2.11.

For more than two parties the analogue of a density matrix $\rho_{ABC...}$ is a probability distribution $P(X_A, X_B, X_C, ...)$. It is not yet clear to us what the general analogue of a multi-partite pure state is. This is due, in part, to the fact that for multi-partite states the analogue of the Schmidt decomposition is far more complicated. We shall give some multipartite results in section 2.12.

2.3 Teleportation and the One Time Pad

The first "derived" analogy is probably the most striking of all. The fundamental quantum communication protocol that is teleportation[2] turns out to be analogous to the fundamental secrecy communication protocol[4], the one-time pad[31].

The scenario for teleportation is as follows. Alice would like to send a qubit to Bob, but is separated from him by a noisy environment, across which she cannot at present send any qubits. The environment even prevents her from carrying the state

[3]Note that quantum mechanically in order to say that the state of Alice and Bob is pure we don't need to specify that the state of Alice, Bob and the Environment is of the form $|\psi\rangle_{ABE} = |\psi\rangle_{AB}\left|\tilde{\psi}\right\rangle_E$, but it is enough to know the state ρ_{AB} of Alice and Bob alone. On the other hand, the classical correlations of Alice and Bob alone do not allow us to know if Eve is, or is not, correlated with Alice and Bob, therefore we must always describe the full state of Alice, Bob and Eve.

[4]This analogy has also been noticed by [57, 58, 3].

to him personally in some secure box. However, she does have a classical phone line, and some shared entanglement. We assume that she does not know that state of the qubit, and so she cannot just tell Bob the state using the phone line. Neither can she measure the state to determine it: quantum mechanics forbids it. However, using the entanglement and the classical phone line, she is able to transfer the state to Bob. This is teleportation.

The one-time pad works in a similar scenario. Alice would like to send a secret message to Bob, but can only communicate with him using a public phone line, which an eavesdropper can hear. However they do share some secret correlations. Using the secret correlations to encode the message, and transmitting it using the public phone line, they are able to communicate secretly. The basic protocol for this, which uses the secret correlations and the phone line in the most efficient way, and in which Eve cannot learn anything about the secret message (even if she has unlimited computational power), is the one-time pad. Note that we do not discuss public-key cryptography, which requires only a public channel, and no shared secret correlations, but does assume Eve has limited computational power.

Teleportation (the one-time pad) works in the following way[2, 31]. Alice begins with the qubit (secret bit) to be sent, which may be entangled (secretly-correlated) with any number of other particles (bits). She does a Bell measurement (addition modulo 2) on the qubit (secret bit) to be sent and the qubit (bit) of resource she holds. She then sends the outcome (result) of this operation as a classical bit (public bit) to Bob. He then does a conditional unitary (bit flip) upon his part of the e-bit (shared secret bit). Bob now holds the qubit (secret bit) Alice was sending him.

The necessary and sufficient resources are given by:

$$1e - bit_{AB} + 2classical\ bits_{A\to B} \Rightarrow 1qubit_{A\to B} \tag{2.3.1}$$

$$1shared\ secret\ bit_{AB} + 1public\ bit_{A\to B} \Rightarrow 1secret\ bit_{A\to B} \tag{2.3.2}$$

By necessary we mean that, if we were to try to do the teleportation with less than 1 e-bit - by using a less than maximally entangled state for example - the teleportation will not give a perfect output, and the classical information will give some information about the qubit we are sending. If we try to use a less than

completely correlated shared secret bit to send a secret bit then Eve gets some information about the secret bit. The resources are sufficient since we can achieve the operations using them.

Note that the resources are used up in the process: once we have used an e-bit (shared secret bit) to send a qubit (shared secret bit) we cannot reuse it. Quantum mechanically this is obvious, since the original maximally entangled state is destroyed by Alice's measurement. Classically however Alice and Bob do not lose their correlated bits - Alice and Bob need not erase or physically modify in any way their original correlated bits but just use them for some mathematical operations. What is lost however is the secrecy of these bits - they cannot be reused.

Furthermore, it is obvious to see that the one-time pad secret communication can be used to implement the analogue of teleportation of entangled states and of entanglement swapping.

Finally, let us note an important fact. Quantitatively the amount of resources in the classical and quantum cases are similar but not identical: but we need 2 classical $bits_{A\to B}$ to send 1 qubit, whereas only 1 public $bit_{A\to B}$ to send 1 secret bit.

2.4 Single Copy Entanglement and Secret Correlation Manipulations

The ability to manipulate entanglement, i.e. transforming entanglement from one form into another by local actions and classical communications is one of the most important aspects of entanglement. This leads to elevating entanglement to the status of a (fungible) resource: to a large extent it doesn't matter in which form entanglement is supplied, we can transform it into the specific form we need for different applications, very much as say, transforming the chemical energy stored in coal into electrical energy for use in electric engines. Similarly one can imagine that Alice and Bob are supplied with secret correlations in some given form, i.e. according to some specific probability distribution, and they want to obtain secret correlations obeying a different probability distribution. We find that the quantum

and classical scenarios are in very close analogy.

In this section we treat the case of bi-partite pure state single copy manipulations. In the quantum context this means that the two parties, Alice and Bob, share a single pair of particles in some pure state $|\Psi\rangle_{AB}$. In the classical context, Alice and Bob share a single sample of a classical pure-state (2.2.2).

In the case of a single copy, entanglement is not a completely interconvertible resource (as it is in the case of many copies (see section 2.8)), but many more restrictions apply.

For bipartite pure quantum states, it is possible to turn one state into another *with certainty* if and only if a certain set of conditions, collectively known as majorization, holds [32, 33]. We here show that for classical secret pure states, the transformation is possible if and only if an analogous condition holds.

Quantum mechanically the majorization condition is the following. Consider two quantum pure states $|\psi\rangle_{AB}$ and $|\phi\rangle_{AB}$, written in their Schmidt bases

$$|\psi\rangle_{AB} = \sum_i \sqrt{p_i}\,|i\rangle_A\,|i\rangle_B\,, \tag{2.4.1}$$

$$|\phi\rangle_{AB} = \sum_i \sqrt{q_i}\,|i\rangle_A\,|i\rangle_B\,, \tag{2.4.2}$$

with the squared Schmidt coefficients p_i and q_i arranged in decreasing order, $p_1 \geq p_2 \geq ...$ and $q_1 \geq q_2 \geq ....$ The vector $\vec{q} = \{q_i\}$ is said to majorize the vector $\vec{p} = \{p_i\}$ iff

$$\sum_{i=1}^{k} q_i \geq \sum_{i=1}^{k} p_i \quad \forall k. \tag{2.4.3}$$

$|\phi\rangle_{AB}$ is said to majorize $|\psi\rangle_{AB}$ iff $\vec{q}$ majorizes $\vec{p}$. The transformation $|\psi\rangle_{AB} \mapsto |\phi\rangle_{AB}$ is possible with certainty if and only if $|\phi\rangle_{AB}$ majorizes $|\psi\rangle_{AB}$ [33]. (Note that it is the final state which must majorize the starting one.)

For classical secret correlations, suppose Alice and Bob begin with an arbitrary classical bipartite pure state, which we may write as

$$p(X_A = i, X_B = j, X_E = k) = \delta_{ij} p_i \tilde{P}(E_k). \tag{2.4.4}$$

Their task it to produce some other state,

$$p(Y_A = i, Y_B = j, Y_E = k) = \delta_{ij} q_i \tilde{P}'(E_k). \tag{2.4.5}$$

We shall prove that they can do this iff $\vec{q}$ majorizes $\vec{p}$. However, to understand what is going on, let us first consider a simple example which has all the important features. The quantum version was first considered in [32].

Suppose Alice and Bob share one sample of the classical pure state X, where

$$p_1 = p_2 = p_3 = \frac{1}{3}, \tag{2.4.6}$$

and they would like to turn it into a sample of the pure state Y, where

$$q_1 = q_2 = \frac{1}{2}. \tag{2.4.7}$$

A probabilistic method (analogous to the procrustean method for the quantum case[21]) is for Alice to send message m_1 (which means "OK") if X is 1 or 2, and to send message m_2 (which means "not OK") if X is 3. If message m_1 is sent then Alice and Bob keep their sample, and they now have a shared secret random variable of the form Y. Indeed, in this case Eve only knows that the value of the secret variable is either 1 or 2 but she doesn't know which one - Alice and Bob's data is therefore still perfectly secret, and it is now either 1 or 2 with probability 1/2. If message m_2 is sent then the procedure failed and Alice and Bob have to throw away their sample. The reason is that Eve, who monitors the public communication, learns that Alice and Bob's variable is equal to 3, and there is no more Alice and Bob can do.

The above method works with probability $\frac{2}{3}$. Can Alice and Bob do better? The second distribution majorizes the first, since $\frac{1}{2} \geq \frac{1}{3}$, $\frac{1}{2} + \frac{1}{2} \geq \frac{1}{3} + \frac{1}{3}$ and $\frac{1}{2} + \frac{1}{2} + 0 \geq \frac{1}{3} + \frac{1}{3} + \frac{1}{3}$. Thus, according to the majorization theorem we shall shortly prove, there exists a method which works with certainty. The protocol for achieving this goes as follows. Alice reads the value of X. If it is 1, she flips an unbiased coin which tells her to send message m_1 or m_2 with equal probability. If $X = 2$ she flips an unbiased coin to send m_2 or m_3, and if $X = 3$ she flips an unbiased coin to send m_1 or m_3. She then publicly sends the message, so that everyone can read it. If m_1 is sent, Eve knows that X is 1 or 3 with equal probability. If m_2 is sent, Eve knows that X is either 1 or 2, with equal probability. And if m_3 is sent, Eve knows that X is 2 or 3 with equal probability. Now Alice and Bob just have to do a simple relabeling of

X to produce Y. If m_1 is sent, they both do $1 \mapsto 1, 3 \mapsto 2$. If m_2 was sent they do $1 \mapsto 1, 2 \mapsto 2$. If m_3 is sent they do $2 \mapsto 1, 3 \mapsto 2$. Whatever message was sent, Y is now a shared random variable which is (as far as Eve is concerned) a shared secret bit of the form (2.4.7).

Now we shall look at the general case. For which pure states X and Y is it possible to turn a single sample of X into a single sample of Y? Consider the most general possible protocol. We assume that Alice, Bob and Eve all know the protocol[5]. Alice and Bob start by having a single sample of the pure state X. They each have also access to some local source of secret randomness - they may each throw dice. Of course, Alice knows only the outcomes of her dice and Bob of his. During the protocol Alice and Bob may publicly communicate, perhaps in many rounds, with each message determined by X, the public messages already sent, and by the results of the local dice. At the end of the protocol there will be some total public message which consists of all the messages that were exchanged by Alice and Bob. All three parties, Alice, Bob and Eve know this total message. In addition, Alice and Bob know the value of X (which is common to both of them since the state is pure), and each of them knows the outcomes of his/her own dice. Based on all this knowledge Alice and Bob must decide on the values of Y_A and Y_B. Formally, we can write

$$Y_A = f_A(X_A, m, d_A) \tag{2.4.8}$$

$$Y_B = f_B(X_B, m, d_B) \tag{2.4.9}$$

where by m we denote the total message, and by d_A and d_B we denote the outcome of all Alice's and Bob's dice.

The above procedure can be simplified. Since we begin with a pure state, $X_A = X_B = X$. Furthermore, since we want to end with a pure state, we require $Y_A = Y_B$. This requirement implies that Y_A and Y_B cannot depend explicitly upon the outcome of the dice d_A and d_B (only implicitly through m). Also given the initial value X

[5]if Alice and Bob had a secret protocol, this would be like having an additional shared random variable, whose different outcomes told them which protocol to use. Thus they would have an additional resource. Here we insist they have only one shared resource, X.

and the message m, Alice and Bob must perform the same function f. Thus we get

$$Y_A = Y_B = f(X, m). \tag{2.4.10}$$

Furthermore, since Bob's actions may not depend on the outcomes of his dice but only on X and m, for every procedure which involves many rounds of communication between Alice and Bob, we can formulate an equivalent procedure in which the total message is entirely generated by Alice - she could simply throw all dice herself - and then communicate the message in a single transmission to Bob.

Let us now formalise this procedure for turning X into Y.

Alice looks at $X = x_i$, which occurs with probability p_i. She then throws a biased dice which tells her to send message m_j with some probability $p(m_j|x_i)$ which depends upon x_i. She then publicly announces m_j. Alice and Bob now follow the instructions in the message, which say to do $x_i \mapsto y_k(x_i, m_j)$. Forgetting what X is (ie. summing over x_i) this gives them some joint distribution for y_k and m_j, $p(y_k, m_j)$. Since Alice and Bob want y_k to be secret from Eve, who knows only the protocol and the message, this distribution must factorise: $p(y_k, m_j) = p(y_k)p(m_j)$. $p(y_k)$ is the final distribution, and so we want $p(y_k) = q_k$ (the distribution of Y).

This secrecy procedure can be thought of as a single party problem, which goes as follows. We begin with a sample from X, which occurs with probability p_i. We may look at the sample, and then roll some dice which gives outcome m_j with probability $p(m_j|x_i)$. We then perform the map $x_i \mapsto y_k(x_i, m_j)$. We then forget what X is, which gives some joint distribution for y_k and m_j, $p(y_k, m_j)$. We desire this distribution to factorise, $p(y_k, m_j) = p(y_k)p(m_j)$, and that $p(y_k) = q_k$. Note that this single party procedure is not a secrecy procedure, however it is possible iff the above secrecy transformation is.

To find for which p_i and q_k this single party problem is possible, and thus to find for which p_i and q_j the secrecy transformation is possible, we shall look at the time reversed problem. This goes as follows. We start with a sample from Y, which occurs with probability q_k. We then roll dice, which give outcome m_j with probability $p(m_j)$, independent of the outcome of Y. This gives a joint distribution $p(y_k, m_j) = q_k p(m_j)$. Now we must do the inverse of the map $x_i \mapsto y_k(x_i, m_j)$

to turn our Y into an X. If the map is one-to-one, and hence invertible, this will give us a distribution $p(x_i, m_j)$. Like any joint distribution, this can be written as $p(x_i, m_j) = p(x_i)p(m_j|x_i)$. If we now forget the value of Y and of m_k, we get a new distribution for X, $p(x_i)$. We desire $p(x_i) = p_i$. If the map is many-to-one, then we can give it a probabilistic inverse which is a "one-to-many" map where the probabilities of getting various x_i's given any particular y_k are given by the relative frequencies of the x_i's when y_k is produced in the forward time protocol. This probabilistic one-to-many map can be replaced by a probabilistic choice of several one-to-one maps, which will have the same effect upon the protocol since we forget which map we did at the end. Thus in the reversed time single party problem, we need only consider maps which are one-to-one. This also applies to the forward time single party problem, and to the forward time secrecy protocol: we only need consider maps which are one-to-one, ie. permutations.

As explained above, if we find the conditions for which the reversed time single party problem is possible, we will have the conditions for which the forward time secrecy transformation is possible. Physically, this time reversed single party problem goes as follows. We begin with a ball in some box according to the distribution q_i. We do not know which box the ball is in, and are not allowed to look to see where it is. We then apply some shuffle (one-to-one relabeling) to the boxes, choosing which shuffle to make according to a distribution, $p(m_j)$, which we may choose. We then forget which shuffle we did, and look at the new distribution of the balls, p_i. The question is for which q_i and p_i is this possible? Clearly p_i should be more random than q_i. This is a well-known problem, and is the context in which majorization appears in classical physics. The answer is that it is possible iff $\vec{p}$ majorizes $\vec{q}$. Intuitively this is easy to see, and the proof can be found, for example, in[59].

Above we have proved the majorization result in the classical context by using arguments referring solely to the classical context. We could have used however the known results for quantum entanglement manipulation to prove the classical ones. The reason is as follows. On one hand, it was found out that transforming pure quantum states (with certainty) from one into another involves only actions and measurements in the Schmidt decomposition basis. These actions do not in-

volve phases, but are simply classical actions upon the basis, which are performed coherently to make a quantum evolution. One could, however, imagine starting by measuring the quantum state in the Schmidt basis, and then performing the corresponding classical actions and measurements upon the state. This transforms one classical state into another, and will not give Eve any knowledge about the state since the quantum procedure did not entangle the quantum state with the environment. Thus, if we can transform with certainty a quantum pure state $|\Psi\rangle$ (2.4.1) into a quantum pure state $|\Phi\rangle$ (2.4.2), we can also transform with certainty X (2.4.4) the classical pure state equivalent of $|\Psi\rangle$, into Y (2.4.5), the classical pure state equivalent of $|\Phi\rangle$.

To prove the reverse, that is, that X can be transform with certainty into Y only if the quantum analogues can be transformed from one into the other, we note that we can turn any classical transformation of pure states into a quantum one, simply by applying the classical operations coherently, and performing the quantum actions in the Schmidt basis. Thus there cannot be any classical procedure which does better than the optimal quantum one. So the classical transformation is possible iff the quantum one is. [6]

2.5 Probabilistic Single Copy Manipulations

It may not be possible to transform a single copy of a resource from one form into another with certainty, but it may be possible to do it with some probability. What is the largest probability with which this can be done? For quantum states, the problem was considered in [32, 60], and the general answer is given in [60]. The maximum probability with which we can turn state

$$|\psi\rangle_{AB} = \sum_i \sqrt{p_i}\, |i\rangle_A\, |i\rangle_B\,, \tag{2.5.1}$$

[6] Note however that although we can use the quantum result to prove the classical one, we cannot use the classical result to prove the quantum result. The reason is that although we can turn any classical transformation into a quantum one, we cannot generate this way all possible quantum protocols - indeed, they may involve phases outside the Schmidt basis.

into state

$$|\phi\rangle_{AB} = \sum_i \sqrt{q_i}\,|i\rangle_A\,|i\rangle_B\,, \tag{2.5.2}$$

using LOCC is given by

$$\min_k \frac{1-\sum_{i=1}^k p_i}{1-\sum_{i=1}^k q_i}. \tag{2.5.3}$$

We shall now show that for classical secret states, the answer is the same.

As we did for the non-probabilistic transformations, we may simplify the most general protocol, which then goes as follows. Alice first looks at her sample which comes according to the distribution $p(x_i)$. She then chooses a message m_j according to $p(m_j|x_i)$. Most of the possible messages will be ones for which the transformation succeeds: these must say to do a one-to-one map[7] $X \mapsto Y$. The other messages say "fail": for these it does not matter what transformation we do, and it does not help to send more than one "fail" message. So we may assume we have only one "fail" message, m_{fail}, which says to do $x_i \mapsto y_1$. Alice and Bob then do $x_i \mapsto y_k(x_i, m_j)$ according to the message. This gives them a distribution $p(y_k, m_j)$. In the case they succeed, this distribution must factorise:

$$p(y_k, m_j) = \begin{cases} p(y_k)p(m_j) & \text{for j} \neq \text{"fail"} \\ \delta(y_k = 1)p(m_{fail}) & \text{for j} = \text{"fail"} \end{cases} \tag{2.5.4}$$

By defining $p(success) = \lambda$, so that $p(m_j) = \lambda p(m_j|success)$ for $j \neq$ "$fail$" and $p(m_{fail}) = 1 - \lambda$ and by requiring $p(y_k) = q_k$ (so that the protocol succeeds) we obtain:

$$p(y_k, m_j) = \begin{cases} \lambda q_k p(m_j|success) & \text{for j} \neq \text{"fail"} \\ (1-\lambda)\delta(y_k = 1) & \text{for j} = \text{"fail"} \end{cases} \tag{2.5.5}$$

The time reversed, single party version of this problem is to start by flipping a coin (H/T) with probabilities $(\lambda, 1-\lambda)$. We look at the result, and if it is T we start with $y_k = 1$, send a message m_{fail}, and are allowed to do anything (including probabilistic things) to transform $Y \mapsto X$. If the coin is H we get a sample y_k according to $p(y_k) = q_k$, but do not know which sample we get. We then pick some

[7] There is no loss in generality in forgetting about the many-to-one maps, for the same reasons as in the non-probabilistic manipulations.

message according to $p(m_j)$, and do the corresponding shuffle $y_k \mapsto x_i$. This gives some distribution $p(x_i, m_j)$. Finally, we forget whether the coin was H or T, and also which message was sent. This then gives us $p(x_i)$, which we would like to be p_i. Our aim is, for a given q_k and p_i, to find the maximal λ for which this is possible. This problem is closely related to the one where majorization first appeared in classical physics, and the maximal value of λ is as given at the start of this section. Once again the quantum and classical pure state manipulations are possible under the same conditions.

2.6 Catalysis of Single Copy Transformations

There is an interesting entanglement transformation called catalysis [61] which transfers easily to the classical case. This is where it is not possible to perform the transformation

$$|\psi\rangle_{AB} \overset{LOCC}{\longrightarrow} |\phi\rangle_{AB} , \tag{2.6.1}$$

but where we can perform the transformation

$$|\psi\rangle_{AB} |\chi\rangle_{AB} \overset{LOCC}{\longrightarrow} |\phi\rangle_{AB} |\chi\rangle_{AB} . \tag{2.6.2}$$

Thus the state $|\chi\rangle_{AB}$ acts as a catalyst. It enables the transformation of $|\psi\rangle_{AB}$ into $|\phi\rangle_{AB}$, but is not consumed in the process. To show this is possible, one has to find states such that $|\phi\rangle_{AB}$ does not majorize $|\psi\rangle_{AB}$, but the tensor product state $|\phi\rangle_{AB} |\chi\rangle_{AB}$ majorizes $|\psi\rangle_{AB} |\chi\rangle_{AB}$. One example[61] of such a catalysis is transforming the quantum state whose squared Schmidt coefficients are

$$p_1 = 0.4; p_2 = 0.4; p_3 = 0.1; p_4 = 0.1 \tag{2.6.3}$$

into the quantum state

$$q_1 = 0.5; q_2 = 0.25, q_3 = 0.25, \tag{2.6.4}$$

using the catalyst

$$r_1 = 0.6; r_2 = 0.4. \tag{2.6.5}$$

It is simple to check that the desired majorization conditions hold.

The classical analogue of this process follows immediately. That is, Alice and Bob may wish to turn the classical pure state defined by p_i into the classical pure state defined by q_j, using LOPC. This is only possible, as we showed in section 2.4, when q_j majorizes p_i. However there are cases when this is not possible, but if they also have a sample of the classical pure state r_k, then they can achieve the transformation

$$P \otimes R \stackrel{LOPC}{\longrightarrow} Q \otimes R \tag{2.6.6}$$

with certainty. The sample R is not revealed or altered by this process, and can be subsequently used independently elsewhere. As far as we know, this classical secret correlation catalysis has not been previously considered.

2.7 Shuffling with Catalysis

Another classical catalysis problem which has not (to our knowledge) been considered before is the single party, time reversed version[8] of the classical pure state catalysis discussed in the previous section. We call this "shuffling catalysis". We emphasize that this shuffling catalysis has, in itself, nothing to do with secrecy or secret correlations. However, it is possible to perform this shuffling catalysis iff the classical pure state catalysis is possible. Recalling (from section 2.4) that the majorization conditions are easier to prove in the shuffling scenario than in the classical secret correlation scenario, studying shuffling catalysis may help in finding exactly when classical secret correlation (and, by analogy, entanglement) catalysis is possible.

We state the problem of shuffling catalysis to make the idea clear. Suppose we have a sample from a distribution q_j and wish to turn it into a sample from a distribution p_i. We are not allowed to look at the sample to see what it is, we can only throw dice whose probabilities (which we choose) are independent of which sample we have. We then make some permutation (shuffle) upon the outcomes, which shuffle decided by the dice, and finally forget which one we did. As mentioned

[8] see section 2.4 for the meaning of the single party, time reversed version of the classical pure state transformation.

in section 2.4, this "shuffling" is possible iff q_j majorizes p_i. There are, however, distributions where q_j does not majorize p_i, and so cannot be turned into it directly, but where we can perform catalysis. This means that we can take a sample from a third distribution r_k, such that $q_j \otimes r_k$ majorizes $p_i \otimes r_k$, and then roll an independent dice and permute the possible outcomes of the tensor product distribution to turn $q_j \otimes r_k$ into $p_i \otimes r_k$. This catalysis is possible iff we can use r_k to turn the shared secret correlation pure state p_i into to the pure state q_j. Thus an example of this shuffling catalysis is the example given in section 2.6.

2.8 Pure State Concentration and Dilution

For many copies of a bipartite pure state, entanglement is a completely fungible resource. It can be converted from one form to another reversibly. Thus we can quantify the amount of entanglement by a single number, the entropy of entanglement. We shall show that the same is true for classical pure bipartite states. That is, for such states, secret correlations are a completely fungible resource. They can be converted from one form to another reversibly, and can be quantified by a single number, the entropy of secrecy.

We define the entropy of entanglement for a quantum pure state, $E(|\psi\rangle_{AB})$ as

$$E(|\psi\rangle_{AB}) = -\sum_i p_i \log p_i \tag{2.8.1}$$

where p_i are the squares of the Schmidt coefficients.

The physical meaning of the entropy of entanglement is the following. When Alice and Bob share a large number N of copies of some arbitrary pure state $|\psi\rangle_{AB}$, they can convert them, in a *reversible way*, using only local operations and classical communication into a number K of copies of the maximally entangled state

$$|\psi_s\rangle_{AB} = \frac{1}{\sqrt{2}}(|11\rangle_{AB} + |22\rangle_{AB}) \tag{2.8.2}$$

where

$$\frac{K}{N} \to E(|\psi\rangle_{AB}) \tag{2.8.3}$$

as $N \to \infty$. That is, the entropy of entanglement represents the yield of singlets per copy of the original state $|\psi\rangle_{AB}$. The operation of converting the states $|\psi\rangle_{AB}$ into maximally entangled states is called entanglement concentration[21] and the reverse operation is called entanglement dilution.

Since entanglement cannot increase under LOCC, the above procedures are optimal, in the sense that concentration and dilution cannot produce more copies: if they could, we would be able to produce entangled states from nothing[9]. We can thus quantify the amount of entanglement in a state by its entropy of entanglement. Any state is worth that many maximally entangled states, since it can be reversibly converted into that many states. We call one of these maximally entangled states an e-bit, and shall say that other states have an entanglement of E e-bits. Note that this quantity is additive. That is, if we have two states which individually have entanglement E_1 and E_2, together they have entanglement $E_1 + E_2$.

The quantum procedure of entanglement concentration can directly be mapped into an equivalent classical analogue. The reason for this is that all the quantum actions used for entanglement concentration take place in the Schmidt decomposition bases, i.e. the unitary actions are all permutations in the Schmidt basis while the measurements are of operators whose eigenstates are direct products in the Schmidt basis. Hence all these actions are essentially classical. Furthermore the quantum procedure does not require communication, so is completely secure.

The quantum dilution protocol also has a classical analogue. Indeed, the quantum dilution [21] involves only Schumacher compression of quantum information and teleportation. Both these protocols have classical analogues: Schumacher compression maps into Shannon data compression and teleportation is replaced by the one-time pad secret communication.

Since secret correlations cannot increase under LOPC, these procedures are optimal. They allow us to reversibly convert N copies of the classical pure state $X \sim p_i$ into K copies of the shared secret bit $Y \sim q_j$,

$$P(Y_A = 1, Y_B = 1) = P(Y_A = 2, Y_B = 2) = \frac{1}{2}, \tag{2.8.4}$$

[9] It would be like the Carnot cycle for a perpetual motion machine.

where

$$\frac{K}{N} = -\sum_i p_i log p_i. \tag{2.8.5}$$

We can thus quantify the amount of secret correlations by the entropy of secrecy, which is defined as the number of shared secret bits which can be produced per copy of the original state X. We note that this amount is equal to the mutual entropy between X_A and X_B, and is also equal to the local entropy of X_A, and to the local entropy of X_B.

2.9 Entanglement Purification and Privacy Amplification

An important procedure in quantum information is Entanglement Purification [22], which turns mixed states into pure states, at the many copy level. The number of pure states produced per input mixed state is the yield.

Analogous procedures for turning classical mixed states into classical pure states exist, though are usually subdivided into two stages. The first stage takes the mixed state $P(X_A, X_B, X_E)$ and turns it into a mixed state where Alice and Bob hold the same value, ie. of the form $P(i, j, k) = \delta_{ij} P(i, i, k)$. This stage is known as Information Reconciliation [62], because Alice and Bob are agreeing on a common value. The second stage takes the output of the first stage, and factors out Eve, to give a state of the form $\delta_{ij} p_i \tilde{P}(k)$. In other words it produces a pure state. This stage is known as Privacy Amplification [62], because Alice and Bob are increasing the secrecy of their key by reducing (to 0) Eve's knowledge of it.

In general it is not known what the optimal protocol is, and there may be different optimal protocols for different states. There are a few different schemes for the quantum and classical cases, but we do not wish to discuss the details here, just to draw the analogy. Firstly, any information reconciliation/privacy amplification protocol may be used as an entanglement purification protocol. Secondly any entanglement purification protocol may be used as an information reconciliation/privacy amplification protocol. We hope that a detailed study of the two problems together

will yield better understanding and new protocols in both the classical and the quantum case.

2.10 Bound Entanglement

One of our motivations for this work was a paper[54] by N. Gisin and S. Wolf suggesting a classical analogue of bound entanglement. A bound entangled state is a bi-partite mixed quantum state which cannot be created locally (without any prior entanglement), but from which no maximally entangled states can be distilled, even if there are many copies of the bound entangled state. It is as if the entanglement is "bound" inside the state, and cannot be released. They proposed the classical analogue to be a sample from a probability distribution on Alice, Bob and Eve, $P(X_A, X_B, X_C)$, in which Alice and Bob have strictly positive intrinsic information[10], but from which they cannot distill shared secret bits under LOPC, even if they have many samples from the distribution. Though it is not yet known if such a classical state exists, there is strong evidence that, by starting with a bound entangled state ρ_{AB}, taking a natural purification, $|\psi_{ABE}\rangle$, and measuring it in natural bases, we may produce a classical bound state. Here we simply note that bound information fits into our framework as a derived analogy, and is another consequence of the deeper analogy between entanglement and secret classical correlations.

2.11 Pure or Mixed?

We have mentioned in section 2.2 that it is not clear whether to classify classical states of the form $P(X_A, X_B)P(X_E)$ where X_A is *not* completely correlated with X_B as pure or as mixed. Such a distribution resembles a pure state because it is not correlated with Eve: this is like a pure state not being entangled with the environment. It also resembles a pure state because we can optimally distill shared

[10] a classical measure which, loosely speaking, is designed to test whether or not Alice and Bob share some information which Eve does not have and which they can use. The hope was that if positive, then they would have something useful, and if zero, then they would have nothing.

secret bits from many copies of such a state at a rate equal to the natural measure of shared correlations, the mutual information [63], [64]; this is the analogue of pure state entanglement concentration. However, it is not known whether such a distillation is reversible. That is, given the shared secret bits, can we produce the original states? If the answer is no, this would be typical behavior of a mixed state. Furthermore, a definite similarity to mixed states is that there is no Schmidt decomposition for such states: in other words there is no way, using local reversible transformations, to make Alice and Bob have the same values for their samples.

Another similarity to mixed states is that it is not possible, even probabilistically, to use LOPC to produce a pure state from one copy of such a distribution. For consider the bi-partite, 2-d case, where Alice and Bob both receive either a 0 or a 1, with probabilities $p_{00}, p_{10}, p_{01}, p_{11}$. We can assume that at least the first three probabilities are non-zero (otherwise they have a pure state). They wish to use LOPC to make a classical "entangled" pure state, ie. where $P(00) > 0$, $P(11) > 0$, $P(01) = P(10) = 0$. As discussed in section2.4, the most general thing they can do is to first communicate publicly, resulting in some total public message, m_i, where i may depend upon their local dice and upon their samples. They may then change their samples according to some map which is specified by the message. For example, the message could tell Alice to flip her bit, and Bob to leave his alone. Note that the message has to tell them what to do locally: it cannot tell them to look at the other person's bit to decide what they will do. Now, to make a pure state with any probability they need at least one map which is local in the sense described above and which produces both 00 and 11, and nothing else. We shall show that no such map exists.

Assume that such a map exists. Without loss of generality, we may assume the map does

$$00 \mapsto 00. \tag{2.11.1}$$

Since Bob has to act locally, this means that if he starts with a 0, he has to finish with a 0. Since they must finish with the same thing, this implies

$$10 \mapsto 00. \tag{2.11.2}$$

Since they are symmetric, similar reasoning gives

$$01 \mapsto 00. \tag{2.11.3}$$

Because they have to act locally, we now know that if Alice or Bob sees a 1, they have to finish with a 0. Thus

$$11 \mapsto 00. \tag{2.11.4}$$

And so the map takes everything to 00, which is no good. For classical states in higher dimensions, the same type of reasoning shows that we cannot produce a classical pure state from a single copy of such a state.

So, as we have shown, classical states of the form $P(X_A, X_B)P(X_E)$ have some characteristics in common with pure quantum states, and some in common with mixed quantum states.

2.12 Multi-Partite Results

It is well known that entanglement is much more complicated for multi-partite systems than for bi-partite systems[23, 24, 65]. In particular, already in the case of three parties, it is known that tri-partite entanglement is fundamentally different to bi-partite entanglement, even in the many copy scenario. Furthermore, there might even exist many different inequivalent forms of tri-partite entanglement. As more systems are added the problem becomes vastly more complicated, but we have a few results to guide us, such as the fact that there is genuine entanglement at every level (again, even in the many copy scenario). Here we show that many of these features have classical analogues.

First, we shall look at the tripartite case. We propose that the classical equivalent of the GHZ state,

$$|GHZ\rangle_{ABC} = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle), \tag{2.12.1}$$

is a probability distribution of the form

$$P(X_A, X_B, X_C, X_E) = P(X_A, X_B, X_C)\tilde{P}(X_E), \tag{2.12.2}$$

where $P(X_A, X_B, X_C)$ is given by

$$P(0,0,0) = P(1,1,1) = \frac{1}{2}. \tag{2.12.3}$$

We shall call this the C-GHZ (classical GHZ), and the classical singlet (ie. the bipartite shared secret bit) we shall call the C-EPR. Is is easy to see that out of 1 GHZ copy we may generate one C-EPR, ie.

$$C - GHZ \stackrel{LOPC}{\longrightarrow} C - EPR. \tag{2.12.4}$$

Clare simply forgets her bit. This may sound unsatisfactory since in the quantum case Alice and Bob end with an EPR which Clare has no control over, whereas here Clare could always later remember her bit, and so one may argue that we have not really performed the classical transformation. However, since Alice, Bob and Clare all begin with the same information and communicate only publicly, it is impossible for Alice and Bob to agree upon anything without Clare knowing it. Thus the "stronger" form of the transformation is impossible, and the best we can do is this weak form, with Clare forgetting her bit.

The above transformation is irreversible: ie. given one C-EPR, it is impossible to make a C-GHZ[23]. This is because the bi-partite entropy of secrecy can only decrease under LOPC, and viewing the system as (AB) vs. C a $C - EPR_{AB}$ will have 0 entropy, whereas the $C - GHZ_{ABC}$ has entropy of 1 (and is symmetric with respect to all the parties). It is possible, however, to do

$$C - EPR_{AB} + C - EPR_{BC} \stackrel{LOPC}{\longrightarrow} C - GHZ. \tag{2.12.5}$$

This is done as it would be in the quantum case: Bob makes a joint measurement on his bits (addition modulo 2), and publicly announces the result. Bob now forgets his second bit, and if the public message was 1, Clare flips her bit. They are then done. This procedure can be viewed as Bob using the $C - EPR_{BC}$ as a one-time pad to send Clare the value of the $C - EPR_{AB}$. It is again clear that we cannot do the reverse transformation: viewing the system as (AC) vs. B, the C-GHZ has an entropy of secrecy of 1, whereas the two C-EPR's together have an entropy of 2.

The entropy of secrecy can be used to show that there exists more than just bi-partite secrecy, even in the many-copy case. Specifically, the 4-party Cat state, which has distribution $P(X_A, X_B, X_C, X_D)$ given by

$$P(0,0,0,0) = P(1,1,1,1) = \frac{1}{2} \tag{2.12.6}$$

(where Eve factors out) cannot be converted reversibly into C-EPR pairs. The proof of this is exactly the proof used for the analogous quantum problem [23], and is done by partitioning the 4 parties into pairs in various ways, and looking at the entropy of entanglement, which must be asymptotically conserved under reversible transformations.

Suppose that we could reversibly convert asymptotically a single 4-party Cat state into C-EPR pairs: n_{AB} between A and B, n_{AC} between A and C, etc. Partitioning the system into (A) vs. (BCD) we get the equation

$$n_{AB} + n_{AC} + n_{AD} = 1. \tag{2.12.7}$$

Partitioning the system as (B) vs. (ACD), (C) vs. (ABD) and (D) vs. (ABC) gives

$$n_{AB} + n_{BC} + n_{BD} = 1, \tag{2.12.8}$$

$$n_{AC} + n_{BC} + n_{CD} = 1, \tag{2.12.9}$$

$$n_{AD} + n_{BD} + n_{CD} = 1. \tag{2.12.10}$$

On the other hand, partitioning the system as (AB) vs. (CD), (AC) vs (BD) and (AD) vs. (BC) gives

$$n_{AC} + n_{AD} + n_{BC} + n_{BD} = 1, \tag{2.12.11}$$

$$n_{AB} + n_{AD} + n_{BC} + n_{CD} = 1, \tag{2.12.12}$$

$$n_{AB} + n_{AC} + n_{BD} + n_{CD} = 1. \tag{2.12.13}$$

Summing the first 4 equations together gives

$$2 \sum_{allpairs} n_{ij} = 4, \tag{2.12.14}$$

whilst summing together the next 3 gives

$$2 \sum_{allpairs} n_{ij} = 3. \tag{2.12.15}$$

Thus the transformation is impossible, and the 4 party classical Cat state really is more than just bi-partite shared secret correlations.

We thus conclude that there are different types of multi-partite secret correlations.

2.13 Conclusion

We have described a fundamental analogy between entanglement and secret classical correlations. The analogy is quite simple to state. Both are resources, and the main objects involved in the study of such resources have a one-to-one correspondence, as given in the table on the first page. Due to this basic analogy, many derived analogies follow. In particular, we have shown that teleportation and the one-time-pad are deeply connected, that the concept of "pure state" exists in the classical domain, that entanglement concentration and dilution are essentially classical secrecy manipulations, and that the single copy entanglement manipulations have such a close classical analogue that the majorization results are reproduced in the classical setting. We have pointed out that entanglement purification is analogous to classical privacy amplification, and hope that the search for better protocols in the two areas can go hand in hand. We finally showed that, as with entanglement, one can look at multipartite shared secret correlations, and gave a flavor of how results in the quantum setting easily transfer into the classical world. Despite all these useful derived analogies, our main point is the fundamental one: entanglement and shared secret correlations are deeply related, and one should never be viewed without the other.

We want to emphasize that by no means do we claim that quantum entanglement is a fundamentally classical effect or that there exists a classical explanation of entanglement. The classical analogue of entanglement is nothing more nor less than a simple analogue, and has a value of its own. On the other hand, all the aspects

of quantum entanglement which are common with the classical analogue cannot be considered to be quantum. Thus many aspects which were hitherto considered to be genuinely quantum lose their status.

The main thrust of our work was to identify the common aspects of quantum entanglement and classical secret correlations. An even more interesting question to find those aspects which are *not* common. One such aspect appears to be the Bell inequality, which I shall discuss at length in the following chapters. In addition, we have not found any (and believe there is no) analogue of super-dense coding. Super-dense coding is the fact that by sending one qubit we may only send one classical bit of information, but by using an additional e-bit we may send two classical bits with just a single qubit. However, it is not the case that by having 1 secret correlation bit and by sending 1 secret bit we can send 2 public bits. The lack of super-dense coding manifests itself, implicitly, also by a difference in the quantitative descriptions of teleportation and one-time pad secret communication: in the case of teleportation we have to send 2 classical bits while in the one-time pad we have to send only 1 public bit. It is only such aspects which are not common to the two settings which are genuinely quantum. We hope that getting rid of those aspects which were believed to be quantum but are not, and identifying the genuine quantum ones will lead to a better understanding of quantum entanglement. And of secret communication.

Chapter 3

Bell Inequalities for Arbitrarily High Dimensional Systems

3.1 Introduction

In his celebrated paper[8], J. Bell showed that the correlated outcomes of a particular gedanken experiment could not be reproduced by any local classical theory (often called local realistic theory, and hereafter called local hidden variables, or LHV, theory). Following him, we call any such correlations non-local. The fact that quantum mechanics contains such correlations was a great surprise, and remains one of the most representative features of quantum mechanics, one for which no classical analog has been found. In this chapter I shall give a new interpretation of this irreproducibility in terms of frustrated networks of correlations. This gives us a better understanding of how quantum mechanics differs from classical theories, and a simple understanding of most previous Bell inequalities, which are the basic tool for detecting non-locality. It also allows us a simpler method for studying the non-local correlations in systems which were previously very difficult, such as bi-partite systems with more than 2 levels (dimensions) in each subsystem. I constructed Bell inequalities for such systems which are strongly resistant to noise, and which therefore may prove useful in experimental detection of non-local correlations. This work was performed in collaboration with Nicolas Gisin, Noah Linden, Serge Massar

and Sandu Popescu, and appeared in Physical Review Letters [42].

3.2 A New Interpretation of Bell Inequalities

The experiment analyzed by Bell is the following. A source prepares a pair of particles in some entangled state. One particle is sent to Alice and one to Bob, Alice and Bob being situated far from each other. When the particle arrives at Alice, she subjects it to a measurement A or $\tilde{A}$, deciding randomly which one to perform. Similarly, Bob subjects his particle to a measurement B or $\tilde{B}$. Each measurement may have d possible outcomes, $A, \tilde{A}, B, \tilde{B} = 0, \ldots, d-1$. The experiment is repeated many times. Everything is arranged such that each pair of measurements performed by Alice and Bob is space-like separated. After the experiment ends, Alice and Bob come together and compare their results. They are interested in the joint probability $P(A = j, B = l)$, which is the probability that $A = j$ and $B = l$ when A and B are measured, and the other joint probabilities $P(A = j, \tilde{B} = m)$, $P(\tilde{A} = k, B = l)$, and $P(\tilde{A} = k, \tilde{B} = m)$.

There are certain constraints upon the joint probabilities which may arise in LHV models. We use these constraints to construct (Bell) inequalities which all LHV models satisfy, but which quantum mechanics does not satisfy. The simplest kind of LHV theory is a deterministic one in which the outcomes $jklm$ of all possible measurements are fixed in advance. There are only a finite number (d^4) of such models, and we can make an inequality which they satisfy in the following way. Since $A = j$, $\tilde{A} = k$, $B = l$ and $\tilde{B} = m$ we have

$$\begin{aligned} r' &\equiv B - A = l - j \ , \\ s' &\equiv \tilde{A} - B = k - l \ , \\ t' &\equiv \tilde{B} - \tilde{A} = m - k \ , \\ u' &\equiv A - \tilde{B} = j - m \ . \end{aligned} \tag{3.2.1}$$

We see that the difference, r', between A and B can be freely chosen by choosing j and l. Similarly the difference, s', between B and $\tilde{A}$ and the difference, t', between

$\tilde{A}$ and $\tilde{B}$ can be freely chosen. But then the difference u' between $\tilde{B}$ and A is constrained since we necessarily have

$$r' + s' + t' + u' = 0 \ . \tag{3.2.2}$$

Thus in a LHV theory the relation between three pairs of operators can be freely chosen, but then the last relation is constrained.

This constraint plays a central role in our Bell inequalities. Indeed they are written in such a way that their maximum value can be attained only if this constraint is frustrated. The simplest such Bell expression is

$$I \equiv P(A = B) + P(B = \tilde{A} + 1) + P(\tilde{A} = \tilde{B}) + P(\tilde{B} = A), \tag{3.2.3}$$

where we have introduced the probability $P(A = B + k)$ that the measurements A and B have outcomes that differ, modulo d, by k:

$$P(A = B + k) \equiv \sum_{j=0}^{d-1} P(A = j + k, B = j \text{ mod d}) \ . \tag{3.2.4}$$

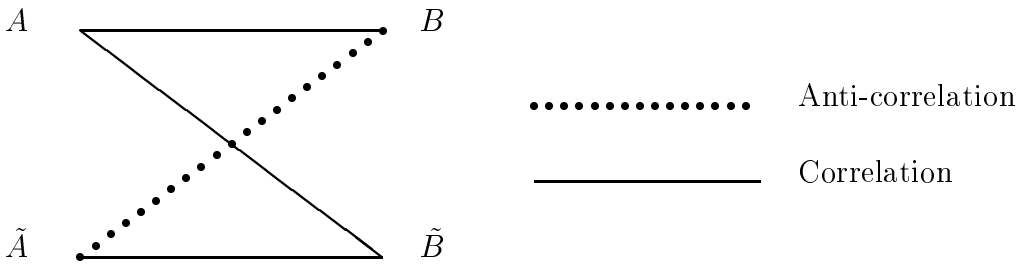
Because the difference between A and B is evaluated modulo d, all the outcomes of A and B are treated on an equal footing. As we see in eq. (3.2.3) this symmetrization is the key to reducing Bell inequalities to the logical constraint that is imposed by LHV theories. Indeed because of the constraint eq. (3.2.2) any choice of local variables $jklm$ can satisfy only three of the relations appearing in eq. (3.2.3), eg. $A = B$, $B = \tilde{A} + 1$, etc.... Hence for deterministic local classical theories, $I \leq 3$. On the other hand non-local correlations can attain $I = 4$ since they can satisfy all 4 relations.

A different way to look at the constraint is displayed in figure 3.1.

Here the LHV model would like to give correlations which satisfy all the links, ie. give $A = B$ where there is a solid line, and $A = B + 1$ where there is a dotted line. However the network is frustrated, and the most links a LHV model can satisfy is 3, ie. $I \leq 3$.

A more complicated class of models is described by probabilistically deciding in the past the outcomes $jklm$ of the different possible measurements. Such a model

Figure 3.1: Network for the CHSH inequality



can be described by d^4 probabilities c_{jklm} $(j,k,l,m = 0,\ldots,d-1)$. Since they are probabilities the c_{jklm} are positive $(c_{jklm} \geq 0)$ and sum to one $(\sum_{jklm} c_{jklm} = 1)$. The joint probabilities take the form $P(A=j, B=l) = \sum_{km} c_{jklm}$, and similarly for $P(A=j, \tilde{B}=m)$, $P(\tilde{A}=k, B=l)$ and $P(\tilde{A}=k, \tilde{B}=m)$. For such a strategy,

$$I = \sum_{ijkl} c_{ijkl}\, I_{ijkl} \leq \sum_{ijkl} c_{ijkl}\, 3 \leq 3, \tag{3.2.5}$$

where I_{ijkl} is the value of I for the deterministic strategy where the outcomes are $ijkl$. Thus such probabilistic models also satisfy the inequality $I \leq 3$.

In fact, provided we assume that the LHV model has no memory between one trial and the next (see chapter 6), the above class of models covers all possibilities. One might imagine a more complicated strategy where Alice and Bob's particles randomly decide what outcomes to give at the last minute, but this randomness can be absorbed into the probabilities c_{ijkl} thus giving an equivalent model with no such local randomness (see for instance [66].) Thus the inequality holds for all possible LHV models, and our method of proof is essentially to check all d^4 local classical deterministic models. Quantum mechanics does not satisfy the same constraints, and in fact violates the inequality. We shall later give explicit states and measurements which do this.

In the case of two dimensional systems the inequality $I(\text{LHV}) \leq 3$ is equivalent

to the CHSH inequality [9]. But the power of our reformulation is already apparent since this inequality generalizes the CHSH inequality to arbitrarily large dimensions. In fact the above formulation of the constraint imposed by LHV theories allows one to write in a unified way most previously known Bell inequalities![1] It can also serve to write completely new Bell inequalities and this is the subject of the remainder of this chapter. Specifically we will generalise in a non trivial way (see sections 3.3 and 3.6 below) the Bell expression (3.2.3) to d dimensional systems (for any $d \geq 2$). We feel that this gives a natural generalisation of the CHSH inequality. Based upon work in [67], we have also generalised the CH inequality [10] to arbitrarily high dimension (see section 3.8.)

3.3 Three Dimensional CHSH Inequality

One can imagine using our approach to write down many new Bell inequalities in d dimensional systems, each based upon constraints which LHV theories must satisfy. However there are many possible inequalities at each dimension, and a complete classification is at present unknown. Indeed, before our work it was very difficult to write down any inequalities at all. Now that we can find many we need some criteria for selecting interesting ones. One of the interests of our new Bell expressions is that they are highly resistant to noise. Indeed Bell inequalities are sensitive to the presence of noise and above a certain amount of noise the Bell inequalities will cease to be violated by a quantum system. However it has been shown by numerical optimization [41] that using higher dimensional systems can increase the resistance to noise. The measurements that are carried out on the quantum system in order to obtain an increased violation have been described analytically in [43]. And an analytical proof of the greater robustness of quantum systems of dimension 3 was given in [68]. When we apply our new Bell inequalities to the quantum state and measurement described in [43] for those dimensions ($d \leq 16$) for which a numerical optimisation was carried out in [43], we obtain the same resistance to noise as in

[1] We believe that all inequalities can be written in this way, however we have no formal proof, and whilst we checked many different Bell inequalities, we did not check all of them.

[43].

The first generalisation of the Bell expression eq. (3.2.3) is

$$\begin{aligned} I_3 = \quad &+ \quad \Big[P(A=B) + P(B=\tilde{A}+1) + P(\tilde{A}=\tilde{B}) + P(\tilde{B}=A)\Big] \\ &- \quad \Big[P(A=B-1) + P(B=\tilde{A}) + P(\tilde{A}=\tilde{B}-1) + P(\tilde{B}=A-1)\Big]. \end{aligned} \tag{3.3.1}$$

This is similar to I_2, but has 4 extra terms, each with a weight of -1. The idea of these extra terms is to severely reduce the value of I_3 compared with I_2 for any LHV model, whilst not reducing the quantum mechanical value by so much. For example, suppose we have a LHV model which gives $I_2 = 3$, by satisfying the first three terms, ie. $A = B$, $B = \tilde{A} + 1$ and $\tilde{A} = \tilde{B}$. This then fixes the relation $\tilde{B} = A - 1$. So we can penalize the LHV model by adding a term $P(\tilde{B} = A - 1)$ to I_2. This is the last term in equation (3.3.1). Since there are 4 ways in which an LHV model could give $I_2 = 3$, we have to subtract 4 different terms to give the expression I_3, which for LHV models is at most 2. Quantum mechanics gives its maximum value of I_2 in a different way, one which does not have such a strong correlation between $\tilde{B}$ and $A - 1$. We find that the penalty quantum mechanics gets from the extra 4 terms in I_3 is less than 1, and so I_3 is better at distinguishing LHV from QM than I_2.

Notice that there are two different fashions in which a deterministic LHV model can score 4. The first, as already discussed, is to satisfy three relations with weight $+1$, and one with weight -1. The second is to satisfy two of the relations with weight $+1$, and none of the others. One could imagine trying to modify the inequality to make an even bigger difference between QM and LHV models. We have only the following intuition, and certainly no proof, that this cannot be done. The inequality we have made has a lot of symmetry, and we wished to preserve this. Thus I_3 has two groups, each of 4 terms, and inside each group all terms have the same weight. We felt that we should not modify the relative weights of terms within the groups. One could try to put different weights between the groups, for instance by replacing -1 by -0.5, or -1.4. However the first modification would give a smaller penalty to LHV models than I_3 does, and so is not as good as I_3, whilst the latter has a large penalty, but a LHV model which satisfies only two relations with $+1$ and no

others will avoid the large penalty, and still give a value of 2, whereas QM would get penalized by more. Thus neither modification improves the inequality. It seems that the best inequality is when the weights are such that there are many ways for a LHV model to give the maximum value.

Another way to modify the inequality would be to add the four other possible terms $P(A = B + 1)$, $P(B = \tilde{A} - 1)$, etc, with some weighting. However, since $\sum_k P(A = B + k) = 1$, these terms are not independent of those already in I_3, and so this would not make a new inequality at all.

Note that for $d = 2$ the inequality $I_3(\text{LHV}) \leq 2$ is equivalent to the inequality $I \leq 3$ and therefore to the CHSH inequality. But for $d \geq 3$ the inequality based on I_3 is not equivalent to that based on I. For the quantum measurement described below (when $d \geq 3$) the inequality based on I_3 (and its generalisations I_d given below) is more robust than that based on I. We therefore feel that I_3 is a good three dimensional generalisation of the CHSH inequality.

3.4 Four Dimensional CHSH Inequality

For four dimensional systems one could use the inequality $I_3 \leq 2$, however we have found the following modification of I_3 to be more useful:

$$\begin{aligned} I_4 = \quad & + \quad \Big[P(A = B) + P(B = \tilde{A} + 1) + P(\tilde{A} = \tilde{B}) + P(\tilde{B} = A)\Big] \\ & - \quad \Big[P(A = B - 1) + P(B = \tilde{A}) + P(\tilde{A} = \tilde{B} - 1) + P(\tilde{B} = A - 1)\Big] \\ & +\tfrac{1}{3} \quad \Big[P(A = B + 1) + P(B = \tilde{A} + 2) + P(\tilde{A} = \tilde{B} + 1) + P(\tilde{B} = A + 1)\Big] \\ & -\tfrac{1}{3} \quad \Big[P(A = B - 2) + P(B = \tilde{A} - 1) + P(\tilde{A} = \tilde{B} - 2) + P(\tilde{B} = A - 2)\Big]. \end{aligned} \tag{3.4.1}$$

The maximum value for any (non-local) theory is 4. The maximum for any LHV theory is still 2. However the addition of the two groups each of 4 extra terms allows QM to obtain a larger value than it could for I_3. The intuition behind the weights in front of the new terms is to put as large weights as possible, without changing the

maximum possible LHV value. This should give QM the best chance to increase its score relative to the LHV score. The new weights were calculated as follows. First note that a LHV model could satisfy two terms from the first set, ie. $A = B$ and $B = \tilde{A}+1$, and two of the new terms, $\tilde{A} = \tilde{B}+1$ and $\tilde{B} = A-2$. Whatever weight we give the first of these new terms we should give the opposite (ie. multiplied by -1) weight to the other new term, in order that such an LHV model gives 2. Looking at the other ways of taking two terms from the first set, and two new terms, sorts the 8 new terms naturally into the 2 groups of 4 shown in I_4. Finally, to determine the weight of the third set of terms, note that one can satisfy one term from the first line, and three from the third line. Thus to give this LHV model a total value of 2, we put a weight $\frac{1}{3}$.

Note that the 16 terms are grouped into two sets of four with weights $+1$ and -1, and two sets of four with weights $+\frac{1}{2}$ and $-\frac{1}{2}$. In a sense they are one set of 8 terms with weight 1, and one set of 8 terms with weight $\frac{1}{2}$. This pattern continues to higher dimensions, with always several groups of 8, only with different weightings. To calculate what the weightings should be, one must look at the different ways an LHV model can try to get a large score, and arrange them so that it never gets more than 2. In fact, the weights are arranged so that they usually give a value of 2: this is in order to make the weightings as large as possible and so give QM the best chance of getting a big score. This intuition was not enough for us to go immediately to the general case: we first used the intuition to find a good inequality in five dimensions.

3.5 Five Dimensional CHSH Inequality

The five dimensional inequality is very similar to I_4, only that the weight of the last two sets is different. Previously this was determined by noting that one could, in four dimensions, satisfy one term in the first line and three terms in the third line. In five dimensions this is no longer the case, however one can satisfy all four terms from the second line, thus fixing the weight to $\frac{1}{2}$.

$$
\begin{aligned}
I_5 = \quad & + \quad \Big[P(A=B) + P(B=\tilde{A}+1) + P(\tilde{A}=\tilde{B}) + P(\tilde{B}=A)\Big] \\
& - \quad \Big[P(A=B-1) + P(B=\tilde{A}) + P(\tilde{A}=\tilde{B}-1) + P(\tilde{B}=A-1)\Big] \\
& +\tfrac{1}{2} \quad \Big[P(A=B+1) + P(B=\tilde{A}+2) + P(\tilde{A}=\tilde{B}+1) + P(\tilde{B}=A+1)\Big] \\
& -\tfrac{1}{2} \quad \Big[P(A=B-2) + P(B=\tilde{A}-1) + P(\tilde{A}=\tilde{B}-2) + P(\tilde{B}=A-2)\Big] .
\end{aligned}
\tag{3.5.1}
$$

3.6 Bell Inequalities for High Dimensional Systems

The Bell expression I_3 can be further generalised when the dimensionality is greater than 5 following similar reasoning. The extra terms in I_d do not change the maximum value attainable by LHV theories ($I_d^{max}(\text{LHV}) = 2$), nor do they change the maximum value attainable by completely non local theories ($I_d^{max} = 4$). However these extra terms allow a better exploitation of the correlations exhibited by quantum systems.

These new Bell expressions have the form:

$$
\begin{aligned}
I_d \;=\; \sum_{k=0}^{[d/2]-1} \left(1 - \frac{2k}{d-1}\right) \Big(\; & + \Big[P(A=B+k) + P(B=\tilde{A}+k+1) \\
& \quad + P(\tilde{A}=\tilde{B}+k) + P(\tilde{B}=A+k)\Big] \\
& - \Big[P(A=B-k-1) + P(B=\tilde{A}-k) \\
& \quad + P(\tilde{A}=\tilde{B}-k-1) + P(\tilde{B}=A-k-1)\Big]\Big).
\end{aligned}
\tag{3.6.1}
$$

As mentioned above the maximum value of I_d is 4. This follows immediately from the fact that the maximum weight of the terms in (3.6.1) is +1. And the maximum value of I_d for LHV theories is 2. We now prove this last result.

The proof consists of checking all the possible relations between A, B, $\tilde{A}$ and $\tilde{B}$ allowed by the constraints (3.2.2). This is most easily done by first changing

notation. We do not use the coefficients r', s', t', u' defined in (3.2.1), but use new coefficients r, s, t, u defined by the relation

$$A = B + r \ , \ B = \tilde{A} + s + 1 \ , \tilde{A} = \tilde{B} + t \ , \ \tilde{B} = A + u \ , \tag{3.6.2}$$

which obey the constraint

$$r + s + t + u + 1 = 0 \bmod d \ . \tag{3.6.3}$$

Furthermore we restrict (without loss of generality) r, s, t, u to lie in the interval

$$-[d/2] \leq r, s, t, u \leq [(d-1)/2] \tag{3.6.4}$$

With this notation the value of the Bell inequality for a given choice of r, s, t, u is

$$I_d(r, s, t, u) = f(r) + f(s) + f(t) + f(u) \tag{3.6.5}$$

where f is given by

$$f(x) = \begin{cases} -\frac{2x}{d-1} + 1 & , \quad x \geq 0 \\ -\frac{2x}{d-1} - \frac{d+1}{d-1} & , \quad x < 0 \end{cases} \tag{3.6.6}$$

We now consider different cases according to the signs of r, s, t, u.

1. r, s, t, u are all positive. Then (3.6.3) and (3.6.4) imply that $r+s+t+u = d-1$. Inserting into (3.6.5) and using (3.6.6) one finds $I_d = 2$.

2. Three of the numbers r, s, t, u are positive, one is strictly negative. Then (3.6.3) and (3.6.4) imply that either $r+s+t+u = d-1$ or $r+s+t+u = -1$. Inserting into (3.6.5) and using (3.6.6) one finds either $I_d = -2/(d-1)$ or $I_d = 2$.

3. Two of the numbers r, s, t, u are positive, two are strictly negative. Then (3.6.3) and (3.6.4) imply that $r + s + t + u = -1$. Inserting into (3.6.5) and using (3.6.6) one finds $I_d = -2/(d-1)$.

4. One of the numbers r, s, t, u is positive, three are strictly negative. Then (3.6.3) and (3.6.4) imply that either $r + s + t + u = -1$ or $r + s + t + u = -d - 1$. Inserting into (3.6.5) and using (3.6.6) one finds either $I_d = -2(d+1)/(d-1)$ or $I_d = -2/(d-1)$.

5. The numbers r, s, t, u are all strictly negative. Then (3.6.3) and (3.6.4) imply that $r + s + t + u = -d - 1$. Inserting into (3.6.5) and using (3.6.6) one finds $I_d = -2(d+1)/(d-1)$.

(Note that for small dimensions d not all the possibilities enumerated above can occur. For instance for $d = 2$, the only possible values are $I_d = \pm 2$.) Thus for all possible choices of r, s, t, u, $I_d(\text{local classical theories}) \leq 2$. This concludes the proof.

3.7 Quantum Violations of the Bell Inequalities

Let us now consider the maximum value that can be attained for the Bell expressions I_d for quantum measurements on the maximally entangled quantum state. We have carried out a numerical search for the optimal measurements. It turns out that the best measurements that we have found numerically give the same value as the measurements described in [43]. We do not have a proof that these measurements are optimal, but our numerical work and the numerical work that inspired [43] suggests that this is the case.

We therefore first recall the state and the measurement described in [43]. The quantum state is the maximally entangled state of two d-dimensional systems

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{d}} \sum_{j=0}^{d-1} |j\rangle_A \otimes |j\rangle_B \ . \tag{3.7.1}$$

The measurements is carried out in 3 steps. First Alice and Bob give each of the states $|j\rangle$ a phase, $e^{i\phi(j)}$ for Alice and $e^{i\varphi(j)}$ for Bob (or phases $e^{i\tilde{\phi}(j)}$ if Alice measures $\tilde{A}$, and $e^{i\tilde{\varphi}(j)}$ if Bob measures $\tilde{B}$. The state thus becomes

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{d}} \sum_{j=0}^{d-1} e^{i\phi(j)} e^{i\varphi(j)} |j\rangle_A \otimes |j\rangle_B \ . \tag{3.7.2}$$

where $\phi(j) = \frac{2\pi}{d}\alpha j$, $\tilde{\phi}(j) = \frac{2\pi}{d}\tilde{\alpha} j$, $\varphi(j) = \frac{2\pi}{d}\beta j$ and $\tilde{\varphi}(j) = \frac{2\pi}{d}\tilde{\beta} j$ with $\alpha = 0$, $\tilde{\alpha} = 1/2$, $\beta = 1/4$ and $\tilde{\beta} = -1/4$. The second step consists of each party carrying

out a discrete Fourier transform to bring the state to the form

$$|\psi\rangle_{AB} = \frac{1}{d^{3/2}} \sum_{j,k,l=0}^{d-1} \exp\left[i \left(\phi(j) + \varphi(j) + \frac{2\pi}{d} j(k-l) \right) \right] |k\rangle_A \otimes |l\rangle_B \ . \qquad (3.7.3)$$

The final step is for Alice to measure the k basis and Bob to measure the l basis. Thus the joint probabilities are

$$\begin{aligned} P_{QM}(A=k, B=l) &= \frac{1}{d^3} \left| \sum_{j=0}^{d-1} \exp\left[i \frac{2\pi j}{d} (k-l+\alpha+\beta) \right] \right|^2 \\ &= \frac{1}{d^3} \frac{\sin^2[\pi(k-l+\alpha+\beta)]}{\sin^2[\pi(k-l+\alpha+\beta)/d]} \\ &= \frac{1}{2d^3 \sin^2[\pi(k-l+\alpha+\beta)/d]} \end{aligned} \qquad (3.7.4)$$

where in the last line we have used the values of α and β given above. The probabilities for $\tilde{A}$ and $\tilde{B}$ have a similar form.

Equation (3.7.4) shows that these joint probabilities have several symmetries. First of all we have the relation

$$P_{QM}(A=k, B=l) = P_{QM}(A=k+c, B=l+c)$$

for all integers c. This symmetry property is related to the fact that in (3.2.4), we considered only the probabilities that A and B differ by a given constant integer c, thus

$$\begin{aligned} P_{QM}(A=B+c) &= \sum_{j=0}^{d-1} P_{QM}(A=j+c, B=j) \\ &= d\ P_{QM}(A=c, B=0) \ , \end{aligned} \qquad (3.7.5)$$

And similarly for $\tilde{A}$ and $\tilde{B}$. Furthermore we have the relation

$$P_{QM}(A=B+c) = P_{QM}(B=\tilde{A}+c+1) = P_{QM}(\tilde{A}=\tilde{B}+c) = P_{QM}(\tilde{B}=A+c). \qquad (3.7.6)$$

Using eqs. (3.7.4 to 3.7.6) we can rank these probabilities by decreasing order. Let us denote

$$q_c = P_{QM}(A=c, B=0) = 1/\left(2d^3 \sin^2[\pi(c+1/4)/d]\right) \ .$$

Then we have

$$q_0 > q_{-1} > q_1 > q_{-2} > q_2 > \ldots > q_{-[d/2]} \; (> q_{[d/2]})$$

where $[x]$ denotes the integer part of x and the last term between parenthesis occurs only for odd dimension d. This suggests that the quantum probabilities violate the constraints imposed by local classical theories. Indeed the probabilities in (3.7.6) are maximized by taking $c = 0$, but then the 4 relations that appear in (3.7.6) are incompatible with LHV theories. In fact replacing the above probabilities in the expression (3.2.3) yields a value $I_{QM} = 4dq_0 > 3$ for all dimensions d.

However a stronger violation is obtained if instead of using the Bell expression I, one uses the Bell expressions I_d. In fact for two d dimensional quantum systems, one can use all the Bell expressions I_k for $k \leq d$, but the strongest violation is obtained by using the Bell expression I_d. This value, denoted $I_d(QM)$, is given by

$$I_d(QM) = 4d \sum_{k=0}^{[d/2]-1} \left(1 - \frac{2k}{d-1}\right) (q_k - q_{-(k+1)}) \; . \tag{3.7.7}$$

For instance we find

$$\begin{aligned} I_3(QM) &= 4/\left(-9 + 6\sqrt{3}\right) \simeq 2.87293 \; , \\ I_4(QM) &= \frac{2}{3}\left(\sqrt{2} + \sqrt{10 - \sqrt{2}}\right) \simeq 2.89624 \; , \\ \lim_{d\to\infty} I_d(QM) &= \frac{2}{\pi^2} \sum_{k=0}^{\infty} \frac{1}{(k+1/4)^2} - \frac{1}{(k+3/4)^2} \\ &= 32 \; \text{Catalan}/\pi^2 \simeq 2.9696 \end{aligned}$$

where Catalan $\simeq 0.9159$ is Catalan's constant.

In the presence of uncolored noise the quantum state becomes

$$\rho = p|\psi\rangle\langle\psi| + (1-p)\frac{\mathbb{1}}{d^2}$$

where p is the probability that the state is unaffected by noise. The value of the Bell inequality for the state ρ is

$$I_d(\rho) = pI_d(QM)$$

Hence the Bell inequality I_d is certainly violated if

$$p > \frac{2}{I_d(QM)} = p_d^{min} \ . \tag{3.7.8}$$

(If there is a quantum measurement giving a value of I_d greater than that given by eq. (3.7.7), then of course the Bell inequality would be violated with even more noise. This remark applies to the various p^{min} below).

As a function of d one finds that p_d^{min} is a decreasing function of d. For instance:

$$\begin{aligned} p_3^{min} &= (6\sqrt{3}-9)/2 \simeq 0.69615 \\ p_4^{min} &= 3/(\sqrt{2}+\sqrt{10-\sqrt{2}}) \simeq 0.69055 \\ \lim_{d\to\infty} p_d^{min}(d) &= \pi^2/\,(16 \text{ Catalan}) \simeq 0.67344 \end{aligned}$$

For $d = 3$ this reproduces the analytical result of [68]. And combining eqs. (3.7.7) and (3.7.8) reproduces the numerical results of [43] for all dimensions ($2 \leq d \leq 16$) for which a numerical optimization was carried out.

3.8 Generalisation of the CH Inequality

After completing the work for the previous chapters, we learned of a Bell inequality for qutrits[67] that exhibits the same resistance to noise as that obtained in [41, 43, 68]. The inequality[67] is a 3 dimensional generalisation of the CH[10] inequality for qubits, in the same way that our inequalities are generalisations of the CHSH inequality for qubits. It is fruitful to look at the connections between the two three dimensional inequalities.

Experiments to test the CHSH inequality [11] are usually conducted with pairs of photons, and often use polarization to encode the two dimensions necessary to give the two outcomes of the measurements. However measurements of the polarization often fail to detect any photon at all, thus giving a third possible outcome: "no detection". It has been common to ignore such outcomes, and simply collect the data when both photons are detected. Unfortunately this procedure leads to the detection loophole, reviewed in chapter 6. One must collect all the data, including the "no detection" events, and show that this three outcome data violates a Bell inequality

directly. It seems that the CHSH inequality is useless for this task, as it only involves two outcomes. Motivated by this problem, Clauser and Horne derived the CH inequality, which deals with precisely this situation. In effect, the CH inequality has terms explicitly involving "no detection" outcomes, whereas the CHSH inequality does not. Despite this, the CHSH inequality and the CH inequality are closely related [10], since each can be derived from the other (see chapter 6.)

It is natural to look at our higher dimensional inequalities in this light. I have checked that I_3 is equivalent to using the "generalised CH" inequality for qutrits [67]. Thus, in this sense, the two three dimensional inequalities are equivalent. In a similar way our inequalities I_n can be used for systems of any dimensionality which have "no detection" outcomes, thus generalising the CH inequality to arbitrarily high numbers of dimensions. An alternative generalisation of the CH inequality to higher dimensions, also based upon I_n, has been given in [69].

3.9 Conclusion

We have given a new interpretation of Bell inequalities, in terms of frustrated networks of correlations. This has clarified the limitations upon the correlations which LHV theories can produce. This new understanding has allowed us to construct a large family of Bell inequalities for systems of large dimension. The numerical work of [41, 43] and a numerical search of our own suggest that these Bell inequalities are optimal in the sense that they are maximally resistant to white noise. For this reason we hope that the Bell inequalities presented here will have as much interest for physicists studying entanglement of systems of large dimensionality as the CHSH inequality has had for bi-dimensional systems.

Chapter 4

Violations of Local Realism by Two Entangled QuNits

4.1 Introduction

Two recent papers [41, 43] have studied the question of robustness of nonlocal correlations. I advanced this study in chapter 3 by giving Bell inequalities which gave an analytic description of their results for low dimensions, and generalised them to arbitrarily high dimensions. These results, however, seem to indicate a very surprising result. Namely, it appears that in a certain sense (which I shall define more precisely later), quantum nonlocal correlations are not very robust. Here I shall argue that nonlocal correlations are actually very robust. While I do not disagree with the specific results found in [41, 43], I show that the class of gedanken experiments they have considered (though very interesting in itself) is in fact quite limited and not sensitive enough. I present a different class of experiments which shows that nonlocal correlations are robust. This work was performed under the supervision of Sandu Popescu, and appeared in J. Phys. A: Math and Gen.[46].

The authors of [41, 43] have considered two quantum particles, each living in an d dimensional Hilbert space, which are in the maximally entangled state mixed with random noise. ie. states of the form

$$\rho_d(p_d) = (1 - p_d) \left|\Psi_d\right\rangle_{AB} \left\langle\Psi_d\right| + p_d \frac{1}{d^2} \mathbb{1}_{d\times d}, \tag{4.1.1}$$

where

$$|\Psi_d\rangle_{AB} = \frac{1}{\sqrt{d}} \sum_{m=1}^{d} |m\rangle_A |m\rangle_B \,, \tag{4.1.2}$$

p_d is a constant $0 \leq p_d \leq 1$ which describes the fraction of noise and $\mathbb{1}_{d\times d}$ is the identity matrix. They have asked, "what is the maximum fraction of noise, p_d, which can be added to the maximally entangled state so that the state still generates nonlocal correlations?"

It is useful here to make a clear distinction between two different issues which are relevant for our discussion. The first is the issue of *entanglement* or *non-separability*. A quantum state is *separable* if it can be written as

$$\rho_{AB} = \sum_i p_i \rho_A^i \otimes \rho_B^i, \tag{4.1.3}$$

and it is non-separable otherwise.

It has been shown [70, 71, 72] that if too much noise is added to the maximally entangled state, the state ceases to be entangled. Obviously, at this moment the quantum state ceases to have any nonlocal aspects whatsoever.

The other issue is whether or not the results of all possible measurements performed on the state can be explained by a local hidden variable model. If they cannot we say, following Bell, that the state generates nonlocal correlations (sometimes this is called a "violation of local realism").

It is clear that when there is so much noise that the state becomes separable, the state cannot generate any nonlocal correlations. It is however possible that the state ceases to generate nonlocal correlations at smaller levels of noise, i.e. while it is still entangled. Indeed, it is not known if every entangled (mixed) state generates nonlocal correlations or not - this is one of the most important issues in quantum nonlocality.

It appears from the results of [41] and [43] that the nonlocal correlations are not robust, meaning that for fractions of noise greater than $p_d \approx 0.33$ none of the states $\rho(p_d)$ produce nonlocal correlations. This is very surprising since the entanglement property of the maximally entangled states is robust - for any fraction of noise, when the dimensionality of the systems is large enough (how large depending on the

fraction of noise), the states of form (4.1.1) are entangled. Furthermore, these mixed entangled states exhibit most other aspects of nonlocality - for example they can be used for teleportation, super-dense coding, and can be purified to yield singlets. So it would be quite strange if they couldn't also generate nonlocal correlations.

We shall show that nonlocal correlations are, similar to entanglement, robust. More precisely we shall show that for any fraction of noise there are states (and experiments to perform upon those states) which exhibit nonlocal correlations. The reason that [41] and [43] did not find these experiments is because they only looked at experiments in which a single von-Neumann measurement is made on each particle; here we look at *sequences* of von-Neumann measurements.

The present discussion is, to some extent, a repeat of the history concerning Werner's density matrices. In 1989 Werner [73] presented some density matrices which are entangled but which are such that if single von-Neumann measurements are made on each particle, the results can be explained by a local hidden variables model. At that time it was tacitly assumed that performing single von-Neumann measurements on each particle essentially covers all possibilities. However it was subsequently shown [44] that the outcomes of *sequences* of von-Neumann measurements are nonlocal - they cannot be explained by any hidden variables model. This work was then extended in [74, 75, 76].

We shall next explain why performing sequences of measurements puts additional constraints on local hidden variable models, then use this to prove that there are states with arbitrarily high fractions of noise which exhibit nonlocal correlations.

4.2 Sequences of Measurements

Consider two observers, Alice and Bob, situated in two space-like separated regions. The standard assumption of LHV is that if Alice performs any arbitrary measurement A and Bob performs any arbitrary measurement B, and the measurements are timed so that they take place outside the light-cone of each other, then there exists a shared random variable λ, with distribution $\mu(\lambda)$, and local distributions $P_A(a;\lambda)$ and $P_B(b;\lambda)$ such that the joint probability that the measurement of A yields a and

the measurement of B yields b is given by

$$P_{AB}(a,b) = \int P_A(a;\lambda)P_B(b;\lambda)\mu(\lambda)d\lambda, \tag{4.2.1}$$

for all possible measurements A and B.

Consider now that Alice and Bob, instead of subjecting their particles to a single measurement, perform two measurements one after the other, say A^1 followed by A^2 and B^1 followed by B^2. Then a LHV model implies that

$$P_{A^1A^2B^1B^2}(a^1,a^2,b^1,b^2) = \int P_{A^1A^2}(a^1,a^2;\lambda)P_{B^1B^2}(b^1,b^2;\lambda)\mu(\lambda)d\lambda. \tag{4.2.2}$$

Quantum mechanically the two measurements on each side could be viewed as a single POVM. For LHV models however, doing one measurement after the other gives us the extra constraint that we must be able to write $P_{A^1A^2}(a^1,a^2;\lambda)$ in the form

$$P_{A^1A^2}(a^1,a^2;\lambda) = P_{A^1}(a^1;\lambda)P_{A^2}(a^2;A^1,a^1,\lambda). \tag{4.2.3}$$

Here $P_{A^1}(a^1;\lambda)$ is the probability that Alice's particle yields the answer a^1 when the first measurement to which is subjected is A^1 and given that the hidden variable has the value λ. $P_{A^2}(a^2;A^1,a^1,\lambda)$ is the probability that Alice's particle yields the outcome a^2 when the second measurement is A^2, given that the hidden variable has the value λ and given that it was first subjected to a measurement of A^1 to which it yielded the outcome a^1. The reason is that when Alice's particle has to give the outcome of measurement A^1, it does not yet know what exactly will be the measurement A^2 that will be subsequently performed, and so cannot use that information to decide which outcome a^1 to give. We must write Bob's probabilities in a similar way.

Now, let us look at the probabilities of outcomes of the second measurement, conditioned on some fixed result of the first.

$$P_{A^2B^2}(a^2,b^2;A^1,a^1,B^1,b^1) = \frac{P_{A^1A^2B^1B^2}(a^1,a^2,b^1,b^2)}{P_{A^1B^1}(a^1b^1)}. \tag{4.2.4}$$

Substituting (4.2.2) and (4.2.3) into (4.2.4), and defining

$$\tilde{\mu}(\lambda) = \frac{P_{A^1}(a^1;\lambda)P_{B^1}(b^1;\lambda)\mu(\lambda)}{\int P_{A^1}(a^1;\lambda)P_{B^1}(b^1;\lambda)\mu(\lambda)d\lambda}, \tag{4.2.5}$$

we have that

$$P_{A^2B^2}(a^2,b^2;A^1,a^1,B^1,b^1) = \int P_{A^2}(a^2;A^1,a^1,\lambda)P_{B^2}(b^2;B^1,b^1,\lambda)\tilde{\mu}(\lambda)d\lambda. \quad (4.2.6)$$

We shall now only consider experiments in which the first measurements are fixed and give some particular fixed outcomes, and thus can drop the indices A^1, a^1, B^1 and b^1, which leaves us with

$$P_{A^2B^2}(a^2b^2) = \int P_{A^2}(a^2;\lambda)P_{B^2}(b^2;\lambda)\tilde{\mu}(\lambda)d\lambda. \quad (4.2.7)$$

We further note that $\tilde{\mu}(\lambda)$ is positive and $\int \tilde{\mu}(\lambda)d\lambda = 1$, thus it can be viewed as a probability distribution analogously to $\mu(\lambda)$. Thus, if the whole experiment could be explained by a local hidden variables model, then the probabilities of outcomes for the second measurement conditioned upon any result of the first measurement have to be given by a LHV model themselves. This is a consequence of doing the measurements one after the other rather than together. In particular, we can look at Bell inequalities for these conditioned probabilities, and know that if they are violated, then the initial state is nonlocal. For example suppose that the second measurement which is performed by Alice is either A^2 or $\tilde{A}^2$ and that performed by Bob is either B^2 or $\tilde{B}^2$. Then using the CHSH inequality [9] and (4.2.7) it follows that

$$E(A^2B^2) + E(A^2\tilde{B}^2) + E(\tilde{A}^2B^2) - E(\tilde{A}^2\tilde{B}^2) \leq 2. \quad (4.2.8)$$

For the CHSH inequality we specify that A and B each have two outcomes, which we label as $+1$ and -1, and $E(A^2B^2) = Tr\tilde{\rho}A_2B_2$ is the expectation value of the product of the operators A^2 and B^2 in the state $\tilde{\rho}$ which is the state of the system after the first measurements (assuming that we indeed obtained the particular fixed outcomes we have chosen).

4.3 Non-Local Correlations Are Robust Against Noise

We shall now use (4.2.8) to show that for sufficiently large d, the states defined in equation (4.1.1) generate nonlocal correlations. We take the first measurement on

Alice's side, A^1, to be the projection onto the subspace $\{|1>_A, |2>_A\}$. The first measurement on Bob's side, B^1, is the projection onto the subspace $\{|1>_B, |2>_B\}$. We just look at the cases where the state is indeed in the first two subspaces, in which case the state becomes (after the first measurements):

$$\tilde{\rho} = \frac{(1-p_d)d}{d(1-p_d)+2p_d} |\Psi_2\rangle \langle\Psi_2| + \frac{2p_d}{d(1-p_d)+2p_d} \frac{\mathbb{1}_{2\times 2}}{2^2}. \tag{4.3.1}$$

We now take the second measurements (A^2, $\tilde{A}^2$, B^2, $\tilde{B}^2$) to be those which give the maximal violation of the CHSH inequality on the state $|\Psi_2\rangle_{AB}$ (ie. $2\sqrt{2}$), and we note that if the CHSH inequality is violated, the initial state is nonlocal. This occurs when

$$p_d < \frac{d}{d+c}, \tag{4.3.2}$$

where $c = \frac{2}{\sqrt{2}-1} \approx 4.83$. Therefore, for any fraction of noise we can, by taking d large enough, find states which give nonlocal correlations. Thus we have shown that the nonlocal correlations are robust to noise.

Finally, we note that we have not completely solved the problem of which states of the form (4.1.1) generate nonlocal correlations. Recalling that [70, 71, 72] states of this form are separable iff $p_d \geq \frac{d}{d+1}$, we can see that the states for which $\frac{d}{d+c} \leq p_d < \frac{d}{d+1}$ are entangled but do not violate the Bell inequality we have considered. It is an interesting and open question as to whether these states generate nonlocal correlations or not.

Chapter 5

Bell Inequalities to Detect True Multipartite Non-Locality

5.1 Introduction

In the last two chapters I have studied the non-local correlations between two entangled systems, each with arbitrarily high numbers of dimensions. In this chapter I shall look at the non-local correlations between more than two parties, in particular those between many qubits. I shall show that such systems exhibit true multipartite non-locality. By this I mean that any classical model which reproduces the quantum correlations must utilise superluminal communication which links all the parties. This is a different idea to that of multipartite entanglement, which is a question of whether the quantum mechanical state can be written in a separable way. My work generalises that of Svetlichny [47] on tri-partite systems. It was performed in collaboration with Nicolas Gisin, Sandu Popescu, David Roberts and Valerio Scarani, and published in Physical Review Letters [48].

The usual way to look at non-local properties of multipartite quantum states is in terms of entanglement, otherwise known as non-separability. This classification through entanglement presupposes that the system admits a quantum-mechanical description. Thus, any state of the system is described by a density matrix ρ. The two most common classifications are in terms of the entanglement of formation

[77], ie. the quantum interactions required to create the state, and in terms of the entanglement of distillation [21], ie. the useful entanglement one can create using just local operations and classical communication given an instance of the state. One can also look at these properties for many identical copies of a state, and perhaps under catalysis, or with a limited amount of quantum communication between the parties. These latter scenarios are often useful in simplifying the problem and allowing us to focus upon the most important aspects of non-locality of a state, eg. the entropy. Here we shall be concerned with classification in terms of the simplest and perhaps the most fundamental scenario, the entanglement of formation for a single copy, without any additional resources.

There are strong connections between the entanglement one puts into a state, and that which one can get out given an instance of the state. One cannot get out entanglement which one did not put in. Just how much of the entanglement can be recovered is an area of active research (eg. see [78] and references within). Here we shall just look at the type of entanglement needed to create the state, asking whether it is bi-partite or tri-partite or n-partite, and not ask how much of any type of entanglement is required. In this scenario the entanglement one can distill is very similar to that which one is required for formation of the state. We cannot distill n-partite entanglement from a state which is created with $n-1$-partite entanglement. On the other hand, a state created with n-partite entanglement will often allow us to distill n-partite entanglement (so long as there is not too much noise in the state).

To classify a given ρ in terms of the entanglement of formation, one must consider all possible decompositions of the state as a mixture of pure states $\rho = \sum_i p_i \left|\Psi^i\right\rangle \left\langle\Psi^i\right|$. Suppose we have three parties, A_1, A_2 and A_3. Then (i) If there exist a decomposition for which all $\left|\Psi^i\right\rangle_{A_1A_2A_3}$ are product states $\left|\psi^i\right\rangle_{A_1} \left|\psi^i\right\rangle_{A_2} \left|\psi^i\right\rangle_{A_3}$, then $\rho_{A_1A_2A_3}$ is not entangled at all, that is, it can be prepared by acting on each party separately. For this situation, we use the acronym *1/1/1QM*. (ii) If all $\left|\Psi^i\right\rangle_{A_1A_2A_3}$ can be written as either $\left|\psi^i\right\rangle_{A_1A_2} \left|\psi^i\right\rangle_{A_3}$ or $\left|\psi^i\right\rangle_{A_1A_3} \left|\psi^i\right\rangle_{A_2}$ or $\left|\psi^i\right\rangle_{A_2A_3} \left|\psi^i\right\rangle_{A_1}$, and at least one of the $\left|\psi^i\right\rangle_{A_jA_k}$ is not a product state, then $\rho_{A_1A_2A_3}$ is entangled, but there is no true three-particle entanglement. We shall say that $\rho_{A_1A_2A_3}$ exhibits two-particle entanglement, and use the acronym *2/1QM* to refer to it. (iii) Finally, if for any de-

composition there is at least one $|\Psi^i\rangle_{A_1A_2A_3}$ that shows three-particle entanglement, then to prepare $\rho_{A_1A_2A_3}$ one must act collectively on the three subsystems: $\rho_{A_1A_2A_3}$ exhibits true three-particle entanglement (acronym *3QM*).

It is difficult to establish to which class a given $\rho_{A_1A_2A_3}$ belongs, because in principle one should write down *all* the possible decompositions of $\rho_{A_1A_2A_3}$ into pure states. In fact, to date no general criterion is known. However, we know a sufficient criterion: there exists an operator $\mathcal{M}_3$ such that: (a) if $\mathrm{Tr}(\rho_{A_1A_2A_3}\mathcal{M}_3) > 1$, then certainly $\rho_{A_1A_2A_3}$ is entangled; (b) if $\mathrm{Tr}(\rho_{A_1A_2A_3}\mathcal{M}_3) > \sqrt{2}$, then certainly $\rho_{A_1A_2A_3}$ exhibits true three-particle entanglement. The operator $\mathcal{M}_3$ is the Bell operator that defines the so-called Mermin inequality [79]; we shall come back to it later.

In this chapter, we focus on an alternative classification, in terms of non-locality. Specifically, we look at how many parties must have communicated non-locally in order to reproduce the quantum correlations. This is closest in spirit to the entanglement of formation, asking how many parties must have shared quantum mechanical interactions in the past in order to create the state. One could consider looking at which non-local correlations which can be distilled, given an instance of the state. However, to my knowledge, no such process has ever been demonstrated, even theoretically. Thus from now on we shall compare the non-locality required to reproduce the correlations of the state with the entanglement required to produce the state.

The classification through non-locality does not presuppose that the system admits a quantum-mechanical description. Rather, we have the following cases:

(i) there exists a shared random variable λ, with distribution $\mu(\lambda)$, and local distributions $P_{A_1}(a_1;\lambda)$, $P_{A_2}(a_2;\lambda)$ and $P_{A_3}(a_3;\lambda)$ such that the joint probability that the measurement of A_1 yields a_1,the measurement of A_2 yields b_2, and the measurement of A_3 yields a_3 is given by

$$P_{A_1A_2A_3}(a_1a_2a_3) = \int P_{A_1}(a_1;\lambda)P_{A_2}(a_2;\lambda)P_{A_3}(a_3;\lambda)\mu(\lambda)d\lambda, \tag{5.1.1}$$

for all possible measurements A_1, A_2 and A_3.

(ii) The intermediate case, first considered by Svetlichny[47], is a *hybrid local - nonlocal model*: for each triple of particles, we allow arbitrary (i.e. nonlocal) correlation between two of the three particles, but only local correlations between

these two particles and the third one; which pair of particles is non-locally correlated may be different in each repetition of the experiment. If we define $p_{i,j}$ to be the probability that particles i and j are non-locally correlated, then in this model

$$P_{A_1A_2A_3}(a_1,a_2,a_3) = \sum_{k=1}^{3} p_{i,j} \int d\lambda \big[\mu_{i,j}(\lambda) P_{A_i,A_j}(a_i,a_j|\lambda) P_{A_k}(a_k|\lambda)\big], \qquad (5.1.2)$$

where $\{i,j,k\}$ is an even permutation of $\{1,2,3\}$. We refer to this situation by the acronym *2/1S*. Note that *2/1S* is more general than *2/1QM*, since we don't require that the two correlated particles are correlated according to QM.

(iii) The last situation (*3S*) is the one without constraints: we allow all the three particles to share an arbitrary correlation.

It is not evident *a priori* whether three-particle entanglement *3QM* is stronger, equivalent or weaker than *2/1S*. The proof that *3QM* is actually *stronger* than *2/1S* was given some years ago by Svetlichny [47], who found an inequality for three particles that holds for *2/1S* and is violated by *3QM*. In this chapter, we are going to exhibit a generalized Svetlichny inequality for an arbitrary number of particles n, that is, an inequality that allows to discriminate n-particle entanglement *nQM* from any hybrid model *k/(n-k)S*.

The plan of the chapter is as follows. First, we introduce the family of the Mermin-Klyshko (MK) inequalities [79, 80, 81, 82], that will be the main tool for this study. With this tool, we re-derive Svetlichny's inequality for three particles and compare it to Mermin's. We move then to the case of four particles, and show that the MK inequality plays the role of generalized Svetlichny inequality. Finally, we generalize our results for an arbitrary number of particles n.

5.2 Mermin-Klyshko Inequalities

We consider from now onwards an experimental situation in which *two dichotomic measurements* A_j and $\tilde{A}_j$ can be performed on each particle $j = 1, ..., n$. The outcomes of these measurements are written a_j and $\tilde{a}_j$, and can take the values ± 1.

Letting $M_1 = a_1$, we can define recursively the *MK polynomials* as

$$M_k = \frac{1}{2} M_{k-1} (a_k + \tilde{a}_k) + \frac{1}{2} \tilde{M}_{k-1} (a_k - \tilde{a}_k), \tag{5.2.1}$$

where $\tilde{M}_k$ is obtained from M_k by exchanging all the tilde and non tilde a's. The intuition behind this recursive definition comes from considering the value of M_k under a deterministic local hidden variable model. In such a model all of the a_j and $\tilde{a}_j$ take the value 1 or -1, with certainty. So either $\frac{1}{2}(a_j + \tilde{a}_j) = 1$ and $\frac{1}{2}(a_j - \tilde{a}_j) = 0$, or vice-versa. Thus M_k takes the same value as $M_k - 1$, for all k. Thus for such models M_k is 1 or -1, and a general local hidden variable model is bounded between -1 and 1. However as M_k contains many terms, quantum mechanics can produce a much bigger value (in fact $2^{\frac{k-1}{2}}$, see section 5.5). Writing out these polynomials explicitly, we have

$$M_2 = \frac{1}{2}\big(a_1 a_2 + \tilde{a}_1 a_2 + a_1 \tilde{a}_2 - \tilde{a}_1 \tilde{a}_2\big), \tag{5.2.2}$$

$$M_3 = \frac{1}{2}\big(a_1 a_2 \tilde{a}_3 + a_1 \tilde{a}_2 a_3 + \tilde{a}_1 a_2 a_3 - \tilde{a}_1 \tilde{a}_2 \tilde{a}_3\big). \tag{5.2.3}$$

The recursive relation (5.2.1) gives, for all $1 \leq k \leq n-1$:

$$M_n = \frac{1}{2} M_{n-k} (M_k + \tilde{M}_k) + \frac{1}{2} \tilde{M}_{n-k} (M_k - \tilde{M}_k). \tag{5.2.4}$$

We shall interpret these polynomials as sums of expectation values, eg. we shall interpret M_2 as

$$\frac{1}{2}\Big(E(A_1 A_2) + E(\tilde{A}_1 A_2) + E(A_1 \tilde{A}_2) - E(\tilde{A}_1 \tilde{A}_2)\Big), \tag{5.2.5}$$

where $E(A_1 A_2)$ is the expectation value of the product $A_1 A_2$ when A_1 and A_2 are measured (note that A_1 and $\tilde{A}_1$ cannot be measured at the same time). We call quantities such as $E(A_1 A_2 A_3)$ correlation coefficients. We shall look at the values of these polynomials under QM and hybrid local/non-local variable models, and show that they give generalised Bell inequalities.

We shall first look at hybrid local/non-local variable models. For technical simplicity, throughout this chapter we consider only *deterministic* versions of the hybrid variable models, which means that the script λ in eq. (5.1.1) and (5.1.2) *completely*

determines the outcome of the measurements (i.e the probabilities $P_{A_1}(a_1|\lambda)$ and similar are either zero or one). It is known that any non-deterministic local variable model can be made deterministic by adding additional variables [66]. In addition, we can also use the script λ to determine which particles are allowed to communicate non-locally: eg. for 3 parties the probabilities are now given simply by

$$P_{A_1A_2A_3}(a_1,a_2,a_3) = \int d\lambda \mu(\lambda) P_{A_1A_2A_3}(a_1,a_2,a_3|\lambda), \tag{5.2.6}$$

where for each λ the probabilities must factorize as some 2/1 grouping (though not necessarily the same for different λ). Now, for any λ, the outcomes of all products $A_1A_2A_3$ etc. are fixed, and so we can define the fixed quantity M_n^λ. The value of M_n is just the probabilistic average over λ of M_n^λ. Thus if we can put a bound upon all possible M_n^λ then we have a bound upon M_n. For example, it can be shown that for any LHV model, $M_n \leq 1$. This can be easily seen from (5.2.1) using a recursive argument, noting that for any script of local variables it holds that either $a_n = \tilde{a}_n$ or $a_n = -\tilde{a}_n$. In particular, $M_2 \leq 1$ for LHV is the Clauser-Horne-Shimony-Holt inequality for two particles [9]. On the opposite side, if we consider the model without constraints nS, then M_n can reach the *algebraic limit* M_n^{alg}, achieved by setting at +1 (resp. -1) all the correlations coefficients that appear in M_n with a positive (resp. negative) sign. So for example $M_2^{alg} = M_3^{alg} = 2$.

Turning to QM: here we consider only Von-Neumann measurements with two outcomes acting upon qubits. In this case, the observable that describes the measurement A_j can be written as $\vec{a_j} \cdot \vec{\sigma} \equiv \sigma_{a_j}$, with $\vec{a_j}$ a unit vector and $\vec{\sigma}$ the Pauli matrices. Thus we interpret M_n as the expectation value of the operator $\mathcal{M}_n$ obtained by replacing all a's by the corresponding σ_a. It is known that QM violates the inequality $\mathrm{Tr}(\rho\,\mathcal{M}_n) \leq 1$. More precisely, it is known [80, 81, 82] that: (I) The maximal value achievable by QM is $\mathrm{Tr}(\rho\,\mathcal{M}_n) = 2^{\frac{n-1}{2}}$, reached by the generalized Greenberger-Horne-Zeilinger (GHZ) states $\frac{1}{\sqrt{2}}\big(\,|0...0\rangle + |1...1\rangle\,\big)$; (II) If ρ exhibits m-particle entanglement, with $1 \leq m \leq n$, then $\mathrm{Tr}(\rho\,\mathcal{M}_n) \leq 2^{\frac{m-1}{2}}$. In other words, if we have a state of n qubits ρ such that $\mathrm{Tr}(\rho\,\mathcal{M}_n) > 2^{\frac{m-1}{2}}$, we know that this state exhibits at least $(m+1)$-particle entanglement. This means that the MK-polynomials allow us to detect multipartite entanglement (at least in some states). But do they

allow also us to detect true multipartite non-locality? The answer to this question is: yes for n even, no for n odd. As announced, we demonstrate this statement first for $n = 3$, then for $n = 4$, and finally for all n. For odd n we shall describe some new polynomials which are closely related to the Mermin polynomials, and which allow us to detect true multipartite non-locality, thus giving us a natural generalisation of Svetlichny's three party result.

5.3 Three Party Non-Locality

In order to detect tri-partite non-locality, one might hope to use the Mermin polynomial M_3 given in (5.2.3). We have already discussed the following bounds: $M_3^{LHV} = 1$, $M_3^{2/1QM} = \sqrt{2}$, $M_3^{3QM} = M_3^{alg} = 2$. We lack the bound for *2/1S*. This is easily calculated: consider a script in which particles 1 and 2 are correlated in the most general way, and particle 3 is uncorrelated with the others. Then we use (5.2.1), that reads $M_3 = \frac{1}{2} M_2 (a_3 + \tilde{a}_3) + \frac{1}{2} \tilde{M}_2 (a_3 - \tilde{a}_3)$. For any particular script, as we said above, a_3 can only be equal to $\pm\tilde{a}_3$. Without loss of generality, we choose $a_3 = \tilde{a}_3 = 1$, whence $M_3^{2/1S} = \max M_2$. Since particles 1 and 2 can have the highest correlation, $\max M_2 = M_2^{alg}$ here. In conclusion, $M_3^{2/1S} = 2$. Thus, for Mermin's polynomial

$$M_3^{2/1S} = M_3^{3QM} = M_3^{alg} = 2 : \tag{5.3.1}$$

the Mermin polynomial does not discriminate between the deterministic variable models *2/1S* and *3S*, and the quantum-mechanical correlation due to three-particle entanglement.

One of the problems with M_3 is the fact that M_3 has only four terms: the correlations $a_1a_2a_3$, $\tilde{a}_1\tilde{a}_2a_3$, $\tilde{a}_1a_2\tilde{a}_3$ and $a_1\tilde{a}_2\tilde{a}_3$ do not appear in M_3 (eq. (5.2.3)). But these correlations are those that appear in $\tilde{M}_3$; thus we are lead to check the properties of the polynomial

$$S_3 = \frac{1}{2}(M_3 + \tilde{M}_3) = \frac{1}{2}(M_2\tilde{a}_3 + \tilde{M}_2 a_3)\,. \tag{5.3.2}$$

For both LHV and 2/1S, the calculation goes as follows: we choose $a_3 = \tilde{a}_3 = 1$, and we are left with $S_3^{\cdot\cdot} = \frac{1}{2} \max(M_2 + \tilde{M}_2)$. But $M_2 + \tilde{M}_2 = a_1\tilde{a}_2 + \tilde{a}_1 a_2$, which can take

the value of 2 in both LHV and 2/1S. Therefore $S_3^{LHV} = S_3^{2/1S} = 1$, and this implies immediately $S_3^{2/1QM} = 1$ since *2/1QM* is more general than LHV and is a particular case of *2/1S*. The algebraic maximum is obviously $S_3^{alg} = 2$. We have to find S_3^{3QM}. As above, we define an operator $\mathcal{S}_3$ by replacing the a's in the polynomial S_3 with Pauli matrices. On the one hand, we have

$$\mathrm{Tr}(\rho\mathcal{S}_3) \;=\; \frac{1}{2}\big[\mathrm{Tr}(\rho\,\mathcal{M}_2\sigma_{\tilde{a}_3}) + \mathrm{Tr}(\rho\,\tilde{\mathcal{M}}_2\sigma_{a_3})\big] \;\leq\; \sqrt{2} \qquad (5.3.3)$$

since by Cirel'son theorem [83] each term of the sum is bounded by $\sqrt{2}$. On the other hand, we know [84] that the eigenvector associated to the maximal eigenvalue for such an operator is the GHZ state $\frac{1}{\sqrt{2}}\big(|000\rangle + |111\rangle\big)$. For some settings [1], we have $\langle GHZ|\mathcal{S}_3|GHZ\rangle = \sqrt{2}$: the bound can be reached, that is, $S_3^{3QM} = \sqrt{2}$. Thus the GHZ state generates genuine 3-party non-separability (non-locality). We note that, in fact, S_3 is one of Svetlichny's two inequalities [the second inequality is equivalent, and is associated to $\frac{1}{2}(M_3 - \tilde{M}_3)$].

The results for Mermin's and Svetlichny's inequalities for three particles are summarized in Table 5.1. We see that by combining Mermin's and Svetlichny's

Table 5.1: Maximal values of the Mermin and Svetlichny inequalities for various different models.

	LHV	*2/1QM*	*2/1S*	*3QM*	*3S* (alg.)
M_3	1	$\sqrt{2}$	2	2	2
S_3	1	1	1	$\sqrt{2}$	2
prod.	1	$\sqrt{2}$	2	$2\sqrt{2}$	4

Note: The last line is the product of the two previous values.

inequalities one can discriminate between the five models for correlations that we consider in this paper. This concludes our study of the case of three particles.

[1] To maximize $\frac{1}{2}\langle\mathcal{M}_n + \tilde{\mathcal{M}}_n\rangle_{GHZ}$ for n odd, the σ_{a_j} are taken of the form $\cos\alpha_j\sigma_x + \sin\alpha_j\sigma_y$. One possible choice for the settings is: $\alpha_k = \tilde{\alpha}_k + \frac{\pi}{2}$ for all k, $\tilde{\alpha}_1 = ... = \tilde{\alpha}_{n-1} = 0$, $\tilde{\alpha}_n = \frac{\pi}{4}$. For such settings, each correlation coefficient becomes equal to $\frac{1}{\sqrt{2}}$ in modulus, with the good sign. See [85].

5.4 Four Party Non-Locality

As above, we begin by considering the MK polynomial M_4. Like M_2, and unlike M_3, the polynomial M_4 is a linear combination of the correlation coefficients of *all* measurements. From the general properties of the MK inequalities [80, 81, 82], the following bounds are known: $M_4^{LHV} = 1$, $M_4^{1/1/2QM} = M_4^{2/2QM} = \sqrt{2}$, $M_4^{3/1QM} = 2$, $M_4^{4QM} = 2\sqrt{2}$. The algebraic limit is $M_4^{alg} = 4$ (sixteen terms in the sum, and a factor $\frac{1}{4}$ in front of all).

Now we have to provide the bounds for *1/1/2S*, *2/2S* and *3/1S*. This last one can be calculated in the same way as above: using (5.2.1), we have $M_4 = \frac{1}{2} M_3 (a_4 + \tilde{a}_4) + \frac{1}{2} \tilde{M}_3 (a_4 - \tilde{a}_4)$; we set $a_4 = \tilde{a}_4 = 1$, and since we allow the most general correlation between the first three particles we have $\max M_3 = M_3^{alg} = 2$. Therefore $M_4^{3/1S} = 2$.

One must be more careful in the calculation of *1/1/2S* and *2/2S*. This goes as follows: using (5.2.4), we have $M_4 = \frac{1}{2} M_{1,2} (M_{3,4} + \tilde{M}_{3,4}) + \frac{1}{2} \tilde{M}_{1,2} (M_{3,4} - \tilde{M}_{3,4})$, where to avoid confusion we wrote $M_{i,j}$ instead of M_2, with i and j the labels of the particles. Now, $M_{3,4} + \tilde{M}_{3,4} = a_3\tilde{a}_4 + \tilde{a}_3 a_4$, and $M_{3,4} - \tilde{M}_{3,4} = a_3 a_4 - \tilde{a}_3\tilde{a}_4$. So if we allow the most general correlation between particles 3 and 4, these two quantities are independent and can both reach their algebraic limit, which is 2. Consequently for both *1/1/2S* and *2/2S* we obtain $M_4^{\cdots} = \max(M_{1,2} + \tilde{M}_{1,2})$, which is again 2 in both cases. So finally

$$M_4^{1/1/2S} = M_4^{2/2S} = M_4^{3/1S} = 2 \quad < \quad M_4^{4QM} = 2\sqrt{2} \; : \tag{5.4.1}$$

for four particles, the MK polynomial M_4 detects both four-particle entanglement (this was known) and four-particle non-locality, and is therefore the natural generalization of Svetlichny's inequality.

5.5 Arbitrary Numbers of Parties

For a given number of particles n, we discuss only the maximal value allowed by QM, that is the case nQM, against any possible partition in *two* subsets of k and

$n-k$ particles respectively, with $1 \leq k \leq n-1$, that is the case *k/(n-k)S*. Partitions in a bigger number of smaller subsets are clearly special cases of these bilateral partitions. We are going to prove the following

Proposition: *Define the generalized Svetlichny polynomial* S_n *as*

$$S_n \;=\; \begin{cases} M_n & , \;\; n \;\; even \\ \frac{1}{2}(M_n + \tilde{M}_n) & , \;\; n \;\; odd \end{cases} . \tag{5.5.1}$$

Then all the correlations k/(n-k)S *give the same bound* S_n^k*, and the bound that can be reached by QM is higher by a factor* $\sqrt{2}$*:*

$$S_n^{nQM} \;=\; \sqrt{2}\, S_n^k \,. \tag{5.5.2}$$

The tools for the demonstration are the generalization to all the MK polynomials of the properties of M_2 and M_3 that we used above, namely: (I) the algebraic limit of M_k is: $M_k^{alg} = 2^{\frac{k}{2}} = M_k^{kQM}\sqrt{2}$ for k even, and $M_k^{alg} = 2^{\frac{k-1}{2}} = M_k^{kQM}$ for k odd. (IIa) For k even, M_k and $\tilde{M}_k$ are different combinations of all the correlation coefficients; $M_k + \tilde{M}_k$ and $M_k - \tilde{M}_k$ contain each one half of the correlation coefficients, and the algebraic limit for both is M_k^{alg}. (IIb) For k odd, M_k and $\tilde{M}_k$ contain each one half of the correlation coefficients. These properties are not usually given much stress, but can indeed be found in [80, 81, 82], or easily verified by direct inspection.

Let's first prove the Proposition for n even. In this case, the QM bound is known to be $S_n^{nQM} = 2^{\frac{n-1}{2}}$. As in the case of four particles, to calculate S_n^k we must distinguish two cases:

- For k and $n-k$ even: in (5.2.4), both $M_k + \tilde{M}_k$ and $M_k - \tilde{M}_k$ can be maximized independently because of property (IIa) above; therefore, we replace them by M_k^{alg}. We are left with $S_n^k = \frac{1}{2} M_k^{alg} \max(M_{n-k} + \tilde{M}_{n-k})$, and this maximum is again M_{n-k}^{alg}. So finally $S_n^k = \frac{1}{2} M_k^{alg} M_{n-k}^{alg} = 2^{\frac{n-2}{2}}$.

- For k and $n-k$ odd: in (5.2.4), M_{n-k} and $\tilde{M}_{n-k}$ can be optimized independently because of (IIb) above. We have then $S_n^k = M_{n-k}^{alg} \max M_k = M_{n-k}^{alg} M_k^{alg} = 2^{\frac{n-2}{2}}$.

Thus, we have proved the Proposition for n even.

To prove the Proposition for n odd, we must calculate both S_n^k and S_n^{nQM}. We begin with S_n^k. Inserting (5.2.4) in the definition of S_n for n odd, we find

$$S_n \;=\; \frac{1}{2}\, M_{n-k}\, \tilde{M}_k \;+\; \frac{1}{2}\, \tilde{M}_{n-k}\, M_k \,. \tag{5.5.3}$$

Without loss of generality, we can suppose k odd and $n-k$ even. Therefore, if we assume correlations *k/(n-k)S*, M_k and $\tilde{M}_k$ can both reach the algebraic limit due to property (IIb). So $S_n^k = \frac{1}{2} M_k^{alg} \max(M_{n-k} + \tilde{M}_{n-k})$; and due to property (IIa) this maximum is M_{n-k}^{alg}. Thus $S_n^k = 2^{\frac{n-3}{2}}$. Let's calculate S_n^{nQM}. From the polynomial S_n given by (5.5.3), we define the operator $\mathcal{S}_n$ in the usual way. Therefore for the particular case $k = 1$ we have

$$\mathrm{Tr}(\rho \mathcal{S}_n) \;=\; \frac{1}{2}\big[\mathrm{Tr}(\rho\, \mathcal{M}_{n-1}\sigma_{\tilde{a}_n}) + \mathrm{Tr}(\rho\, \tilde{\mathcal{M}}_{n-1}\sigma_{a_n})\big] \tag{5.5.4}$$

which is bounded by $2^{\frac{n-2}{2}}$ because each of terms in the sum is bounded by that quantity. This bound is reached by generalized GHZ states, for suitable settings (See footnote 1.) Therefore $S_n^{nQM} = 2^{\frac{n-2}{2}}$ for n odd, and we have proved the Proposition also for n odd.

5.6 Conclusion

In this chapter we have shown that quantum mechanics contains genuine n-party non-locality for all n. This leaves open the question of whether n-party non-locality is generic in quantum mechanics, or whether it only occurs for a handful of specific states. It is known that all n-party entangled pure n-party quantum states contain 2-party non-locality [14, 15, 16]. However, this question has not been previously addressed even for 3-party non-locality. We have a partial answer to this question: all tri-partite entangled pure tri-partite quantum states contain genuine 3-party non-locality. This is for the following reason. First note that from any such state one can use local operations to produce either the GHZ or W state[86], and give Alice and Bob local knowledge from which they can later deduce that they had produced the state. Secondly, we have already shown that the GHZ state violates S_3, and we found numerically that for suitable measurements the W state also

violates Svetlichny's inequality. If we therefore perform a sequence of measurements (see chapter 4) consisting of first the actions to produce either GHZ or W, and then the measurements required to violate S_3 on GHZ or W, we will demonstrate tri-partite correlations. Thus all pure tri-partite quantum states which are tri-partite entangled possess genuine 3-party non-locality.

One would also like to know whether multipartite non-locality occurs in nature. Although many experiments give evidence for bi-partite non-locality, current experiments do not appear to demonstrate genuine 3-party non-locality [85]. We believe that our inequalities will be useful in testing for the existence of such non-locality.

Chapter 6

Quantum Non-Locality and the Memory Loophole

6.1 Introduction

In the preceding chapters I have discussed Bell's non-locality, giving a new interpretation of Bell inequalities, and extending them to higher dimensions and the multipartite setting. In this chapter I return to Bell's fundamental gedanken experiment, and show that it has a previously unnoticed fundamental loophole, known as the memory loophole. This is based upon the fact that to measure the probabilities which appear in Bell's inequality we need to perform measurements on many pairs of particles. In practice we do this sequentially, one pair after another. It is normal to make the reasonable assumption that the pairs of particles in later rounds will not remember what measurements the previous pairs were subjected to, nor what outcomes they gave. However, if we wish to rule out all possible local classical models, we must not make this assumption. We must show that Bell inequalities allow us to distinguish between such memory LHV models and quantum mechanics. That is the purpose of this chapter, which is the result of a collaboration with Jonathan Barrett, Lucien Hardy, Adrian Kent and Sandu Popescu[50]. Similar work has been performed independently by Richard Gill [87]. The existence of the memory loophole was independently noticed by Accardi and Regoli [88], whose speculation that

it might allow local hidden variables to simulate quantum mechanics is refuted by Gill's (and our) analysis.

The memory loophole has a different status to other loopholes in the experimental tests of Bell's theorem, such as the detector efficiency loophole, the angular correlation loophole, or the locality loophole. The latter all refer to the possibility for LHV models to exploit additional assumptions which were made to compensate for imperfections in the experimental implementations of Bell's gedanken experiment, and were noted by CHSH [9]. These loopholes would not exist if the experiments were close to perfect [8]. However, the memory loophole is a loophole in Bell's original analysis of locality of the perfect gedanken experiment. As such it presented an unexpected and serious challenge to both the claim that quantum mechanics predicts non-local correlations, and to the experimental test of non-locality. In this chapter we shall show how to overcome this challenge in a simple fashion.

One might wonder what the point of considering all these loopholes is. Each seems to involve more conspiracy on Nature's part than the last, and none of them appears to lead to plausible physical models. Given the importance of the Bell-type experiments, however, and their consequences for our world view, we feel that it is important to analyze the experiments as rigorously as possible and in particular to distinguish between logical impossibility and physical implausibility of the models.

There is another more practical motivation[89, 90]. It is well known that quantum key distribution schemes which use entanglement have significant security advantages over other schemes; they can also be extended by the use of quantum repeaters to allow secure key distribution over arbitrary distances. The security of these schemes relies crucially on the fact that the states created and measured are genuinely entangled. The most obvious and seemingly reliable way to verify this is to use Bell-type tests as security checks within the protocols. However, any such tests need to be interpreted with care. If a quantum cryptosystem is acquired from a not necessarily reliable source, or possibly exposed to sabotage, then a cautious user must consider the possibility that devices have been installed which use classical communication to simulate, as far as possible, the behavior of quantum states, while allowing third parties to extract illicit information about the key. Such devices

effectively define a local hidden variable model, and the usual criterion of physical plausibility no longer applies. A saboteur could set up communication and computing devices that use any information available anywhere in the cryptosystem. In particular, saboteurs might well try to exploit memory loopholes, as well as other Bell experiment loopholes, if they could gain a significant advantage by so doing.

Before discussing the memory loophole, I shall give a review of the other loopholes, and the current status of the suggestive but inconclusive experimental tests for non-locality.

6.2 Testing for Non-Locality

Bell's ideal gedanken experiment requires two spacially separated parties, Alice and Bob. Each makes one of two possible measurements, (A or $\tilde{A}$, B or $\tilde{B}$) each with two possible outcomes. Each entire measurement, including the choice of which measurement to make and the recording of the measurement, must be performed in a region spacelike separated from the other. This procedure is then repeated many times, in order to measure probabilities such as $P(A = 1, B = 0)$. Bell [8] then derives a contradiction between any LHV theory and the quantum mechanical predictions of a particular experiment. The latter is the repeated production of the the singlet state of two spin-$\frac{1}{2}$ particles, and measurements of the local spin in one of two possible directions of each of the particles. Thus the usual theory of quantum mechanics, in which all states and measurements are physically realisable, cannot be described by a LHV model. Quantum mechanics is, in this sense, non-local. Except for the memory loophole, his derivation contains no other loopholes. He suggested that the inequality could be experimentally tested, to see whether the world really is non-local, or whether quantum mechanics would perhaps break down.

A first concern when performing an experiment is that the results will not be perfect. Bell [8] proved a stronger version of his contradiction by showing that the predictions of LHV theories and of QM were some finite distance apart. So LHV theories could not even approximately reproduce the QM correlations, so long as there was not too much noise in the latter. If quantum mechanics was correct, and

one had sufficiently good equipment, one could definitively rule out LHV models. Unfortunately the experiment was a long way ahead of current technology, and even today is not possible (though judging by the state of the art experiment [25] we are more than halfway there).

In order to make progress, additional, physically reasonable assumptions were made. This opened a great debate about what was "physically reasonable". Whilst Bell's theorem only assumed realism and locality, two widely held beliefs, there was less consensus on the validity of the additional assumptions. Each assumption led to a possible experimental test which behaved as quantum mechanics predicted, but each test only ruled out LHV models which satisfied the specific assumptions made. Though the assumptions are often very natural, time and again LHV models which do not satisfy the assumptions have been found which reproduce the correlations from the experiments.

One of the first assumptions, preceding even Bell's papers, was that the LHV model should really be just a quantum mechanical mixture [91]. That is, for some reason, when we try to spatially separate two particles in the spin singlet state, they decohere, and are described by a mixture of quantum mechanical product states. This clearly does not cover all possibilities, but I mention it as it is a particularly natural model. In fact, one point of view is that this is the only model worth considering, and that if this can be ruled out, then the case is closed, and one should accept quantum non-locality. The history of Bell's theorem and all the subsequent tests is based on the opinion that locality is so desirable that we should consider very carefully the possibility that there exists some, possibly less natural, local realistic model.

Furry pointed out that this assumption yields predictions which differ from quantum mechanics, and so which could be tested. He believed that standard quantum mechanics would be confirmed in any experiment. Aharonov and Bohm [7] analyzed existing experimental data, confirming his prediction. Thus entanglement was shown to be a true physical phenomenon. More recently long distance quantum mechanical correlations have been demonstrated over four kilometers in Malvern [92] and over tens of kilometers in Geneva [93]. Thus the natural model of decoher-

ence was ruled out. However, none of these experiments performed the ideal Bell experiment, and all left several loopholes for arguably less natural LHV models to exploit.

I now turn to the major loopholes in the current state of the art experiments.

6.3 The Detection Loophole

There have been many proposals designed to test Bell's inequalities [9, 10, 94, 95, 96, 97, 98, 99, 100]. And many experiments have been performed [101, 102, 103, 104, 105, 106, 107, 92, 108, 93, 25, 109]. Despite their excellent agreement with quantum mechanics, all but one of the experiments to date suffer from the detection loophole [10, 49]. Even the one experiment which does avoid this loophole [109] suffers from the locality loophole (see section 6.4.) The detection loophole arises because most of the experiments are performed with entangled pairs of photons, and when one tries to make a measurement on a photon, one frequently fails to detect it. The natural thing to do is to only consider the data in which both photons were detected, and throw away the rest. Unfortunately this can be exploited by a LHV model to reproduce quantum mechanical (or even stronger) correlations, in the following way.

Recall the CHSH inequality in the form introduced in chapter 3:

$$P_{CHSH} = P(A = B) + P(A = \tilde{B}) + P(\tilde{A} = B) + P(\tilde{A} = \tilde{B} + 1). \qquad (6.3.1)$$

$P_{CHSH} \leq 3$ for local hidden variable theories, where all measurements have two possible outcomes, 0 and 1, $P(A = B)$ is the probability that A and B have the same outcome (ie. are correlated), and $P(A = B + 1)$ is the probability that $A = B + 1$ modulo 2 (ie. are anti-correlated). If we choose our measurements appropriately QM can give its maximum value $P_{CHSH} = 2 + \sqrt{2}$. LHV models can give the value 3, for example by setting $A = B = \tilde{A} = \tilde{B}$, or $A = B = \tilde{A} = (\tilde{B} + 1)$. Suppose now that we only keep data in which both particles are detected. Then a LHV model could send instructions that $A = B$ and $\tilde{A} = \tilde{B} = nodetect$. This will give perfect correlation between A and B, and never any result if any other combination of measurements are made. By sometimes sending the previous instructions, and

sometimes sending instructions such as $A = B = \textit{nodetect}$ and $\tilde{A} = (\tilde{B} + 1)$, a LHV model can exactly reproduce the QM correlations, and even give $P_{CHSH} = 4$: stronger correlations than QM allows!

This example shows that we need to keep more of the data: at least all data in which at least one detector measures a photon. We cannot record the event in which no detectors measure a photon since we do not know when a pair of photons is created, and so cannot tell when it happens. Clauser and Horne made an inequality which applies to this circumstance, in which one needs only record the events when at least one detector fires. This is the CH inequality:

$$\begin{aligned} P_{CH} &= P(A = B = 0) + P(A = \tilde{B} = 0) + P(\tilde{A} = B = 0) \\ &\quad -P(\tilde{A} = \tilde{B} = 0) - P(A = 0) - P(B = 0). \end{aligned} \tag{6.3.2}$$

$P_{CH} \leq 0$ for LHV, and can be as large as $\sqrt{2}$ for QM.

Here $P(B = 0)$ is experimentally estimated as the number of times we find $B = 0$, $\#(B = 0)$, divided by the number of times a pair of particles was created and we measured B. We cannot in practice measure how many pairs of particles were created, which presents us with a seeming difficulty. However, since the right hand side of the inequality is 0, and all the terms on the left hand side are probabilities, which are counting rates divided by roughly the same denominator, one might expect that the following inequality on the actual counting rates holds:

$$\begin{aligned} &\#(A = B = 0) + \#(A = \tilde{B} = 0) + \#(\tilde{A} = B = 0) \\ &\quad -\#(\tilde{A} = \tilde{B} = 0) - \#(A = 0) - \#(B = 0) \quad \leq \quad 0. \end{aligned} \tag{6.3.3}$$

Here $\#(A = B = 0)$ is the number of times A and B are measured and found to be 0, and $\#(A = 0)$ is the number of times A and B are measured and A is found to be 0 (and B can be found to be anything, even not detected). It is important here that all the quantities are expected to be measured the same number of times. In other words, the choice between measuring A and $\tilde{A}$ must be equally likely. Also for $\#(A = 0)$ we should only count the results when, say, A is measured with B, and not those when A is measured with $\tilde{B}$ (we can interchange B and $\tilde{B}$ in this statement: it does not matter). Similarly for counting $\#(B = 0)$. This is a because

we assumed that all the probabilities had the same denominator when we moved from an inequality with probabilities to one with counting rates.

To prove that the inequality holds (at least on average), one can simply enumerate all the possible deterministic LHV models.

An alternative method is to show that, in a certain sense, the CH inequality is equivalent to the CHSH inequality [10]. Suppose we re-write P_{CHSH} using the identities

$$P(A = B) = 1 - P(A = 0) - P(B = 0) + 2P(A = 0, B = 0), \tag{6.3.4}$$

$$P(\tilde{A} = \tilde{B} + 1) = P(\tilde{A} = 0) + P(\tilde{B} = 0) - 2P(\tilde{A} = 0, \tilde{B} = 0), \tag{6.3.5}$$

and use equations similar to (6.3.4) to re-write $P(A = \tilde{B})$ and $P(\tilde{A} = B)$. Then we get an inequality which appears to be the CH inequality (6.3.2). At this moment, there is a subtle difference. In eqn (6.3.4),

$$P(A = 0) = P(A = 0, B = 0) + P(A = 0, B = 1), \tag{6.3.6}$$

whereas in eqn (6.3.2),

$$P(A = 0) = P(A = 0, B = 0) + P(A = 0, B = 1) + P(A = 0, B = nodetect). \tag{6.3.7}$$

However equation (6.3.4) was not designed to deal with no detection outcomes: only with outcomes 0 and 1. To deal with no detection we use a trick: any no-detection outcome is recorded as a 1. We can then use equation (6.3.4) freely, and we see that it is indeed equivalent to the CH inequality.

This relation between the CH and CHSH inequalities can also be extended to higher dimensions (see section 3.8).

The CH inequality allows us to apply a loophole free test for non-local correlations, but unfortunately QM predicts that we must have very efficient detectors in order to violate it. Eberhard has shown [49] that the lowest required efficiency in the presence of no other noise is to use a non-maximally entangled state, and requires an efficiency of 82.8%. Despite lower requirements for higher dimensional systems (see section 3.8 and [67, 110, 69]), no experiment with photons has come close to attaining this efficiency. The only experiment which closes the detection loophole

[109] uses entangled ions, which are so close together that the locality loophole remains open. In the most promising experiments, those with entangled photons, the detection loophole is the main problem in performing a loophole free test of quantum non-locality. However, it appears to be merely a technological problem, and one hopes that with better detectors it will be resolved.

6.4 The Locality Loophole

A very common assumption which has been made in almost every experiment to date is that the source emits pairs of particles independently of the settings of the measuring devices at Alice and Bob. Recall that in Bell's gedanken experiment, one should choose which measurement to perform at the last moment, at random. In most experiments, rather than choosing which measurement to make at the last moment, the measuring devices at Alice and Bob are fixed long in advance to measure observables A and B, or A and $\tilde{B}$, etc. This opens the locality loophole. It is possible that a (sub-luminal) signal goes from the measuring device to the source, tells it which observables are being measured, and then the source sends out a pair of particles correlated in the quantum mechanical way for that particular pair of measurements. This seems unlikely since no such signal has ever been directly detected, and such a signal would need to travel large distances [93]. However it is certainly a local possibility. As such we should perform a full version of Bell's gedanken experiment and rule it out.

The difficulty in removing this loophole is that light travels so fast. For example, if the two detectors are 300m apart, the choice, performance and registration of the measurements must happen within around 1μs. To choose to perform one of two measurements and to actually implement the measurement in such a short time is difficult. There is also the question of what it means to choose at random. The important issue is that the choice of measurement must be made in a way which cannot be influenced by, or predicted by, the source of the pair of particles. If the choice is influenced, then it is easy to make a LHV model which correlates together the choice of measurements and the outcomes and reproduces the quantum

mechanical predictions. To avoid this, we wish to choose the measurement using a true random number generator, and not any pseudo-random device, which could in principle be predicted. It seems impossible to prove that any random number generator is truly random, and thus to conclusively remove this issue. Since all our experience indicates that such devices exist, I shall assume that they do. Without this assumption, which we make in order to perform everyday physics, we simply cannot make progress.

We also would like a random number generator that chooses two outcomes with roughly equal probability. If they are not exactly equal, some of the previous inequalities are no longer valid under general LHV models. However, so long as the outcomes are chosen with close to equal probability, the possible violation of the inequalities will be small, and can be rigorously bounded, thus allowing a loophole free test. Further, those inequalities in terms of probabilities, such as $P(A = B = 0)$, which are experimentally estimated as $\#(A = B)$ divided by the number of times the pair (A, B) is measured, will be unaffected by biasing in the random choice of measurement. However inequalities which are purely in terms of counting rates, such as (6.3.3), are affected, and for these it is highly desirable to keep the biasing small, in order to make the task of finding experimental non-local correlations easier.

The first experiment to perform any random choice of measurement at all was the famous Aspect [104] experiment. However, this only has periodic switching of the choice of measurement, and so cannot be considered truly random. An improved experiment was recently performed in Innsbruck [25]. Each party, Alice and Bob, generates a local random number by firing separate photons onto a beam-splitter and measuring which direction they travel in. The local random numbers are then used to decide which of two sets of measuring apparatus to feed the entangled photon into, thus measuring A or $\tilde{A}$, and B or $\tilde{B}$. This experiment is the most ideal Bell experiment performed to date. If any source of numbers is truly random, this would seem to be it. A similar experiment has been performed in Geneva [93, 111], though is less appealing since the photon which generates the random number is in fact the entangled photon they perform their measurement on!

Even the Innsbruck experiment cannot be said to truly close the locality loophole,

however. There is a minor loophole concerning the time of arrival of the photons in the pair. Suppose for one moment that the experiment had been performed with perfect detectors, and that there were no other losses of photons. In practice one does not know when a photon pair will arrive, so the choice of which measurement is being performed is made very frequently. This ensures that whenever one photon of a pair does arrive, it will not have time to communicate the choice of measurement to the other photon. However, in practice the photons in a pair do not arrive at the two locations at precisely the same time: only within 6ns or so [25]. Now, suppose that the random selection of measuring device was performed every 1ns. Then, a photon in a LHV model could arrive at the detector, watch the measuring device switching at random for a couple of ns, and then let itself be measured when the measurement it preferred was being performed! Like this it can avoid giving correlations in contradiction with quantum mechanics, and so reproduce the quantum correlations (and even stronger ones, if desired). Any switching which is on a similar or much faster timescale than the time window between photon arrivals will open this loophole.

In addition, if the times of the random selection of which measurement to perform are pre-determined, then the LHV photon source could arrange to send the photons just before (ie. within the time window of) the random selection of the measurement. This increases the chance that the LHV photon gets the measurement it wants to $\frac{3}{4}$, even in the case that the random selection of measurement is made quite slowly compared to the time window between photon arrivals. In this case one can make a LHV model giving a large violation of the CHSH inequality: even larger than QM allows!

In the Innsbruck experiment, the choice of which measurement to perform does appear to be pre-determined, and the choice occurs at least every 75ns (the paper does not detail these facts more precisely). Thus there remains a locality loophole for both the reasons mentioned above, though the severity of the first decreases with the length of time between the choices of measurement. Fortunately it is simple to fix both problems. One should simply disregard all data obtained within one time window immediately preceding the choice of measurement. This procedure does not

introduce any new loophole: it is simply deciding (in advance of the experiment) during which time intervals we are going to record data. Of course, it is now important that the time between random choices of measurement longer than the time window (otherwise we will never record any data at all). This does not appear to present any particular difficulty to the experiment, indeed the current data may already have this feature. If this is the case, or if a slightly modified version of the experiment is performed, one would almost be able to say that the locality loophole is closed.

Unfortunately, there is one more issue concerning the locality loophole [112]. This is that the records of the outcome of the measurement (and of which measurement was made) must be made within the spacelike separated region. In other words, the records must be made very quickly. This begs the question of when can we say that the measurement has been recorded. In standard quantum mechanics, the measurement collapses the wavefunction, and at this stage, when we have a classical record, we are able to say that the result is recorded. Before the collapse occurs, the result is not recorded. In this picture, to close the locality loophole, the collapse must happen within the spacelike separated region. Unfortunately, we do not know when the collapse occurs, and have no direct way to test it. In fact, we have no direct way to test whether it exists at all! If we believe it is a genuine physical phenomenon, the only tests seem to be via Bell-type experiments [30, 112], and using these tests to vindicate our Bell experiments is rather circular reasoning.

One can postulate that the result of the measurement has been recorded when it is written in a macroscopic memory. Macroscopic is another hazy term in this analysis, one open to much debate. A conscious person is usually agreed to be macroscopic, but is probably too slow to record data to be of any use in the foreseeable future. Personally, I would be more than satisfied with a record of a thousand 0's to represent one possibility, and a thousand 1's to represent another possibility, perhaps written in the memory of a computer. However this is an issue of taste, and as such is open to debate. It is desirable to perform Bell experiments in which the data is recorded in the most macroscopic manner possible, whilst still occuring quickly enough to be in a space-like separated region. This point is perhaps the

most likely possibility for local realism to hold when all the other loopholes have been experimentally closed. These remarks about recording the results of the data macroscopically also apply to recording the choice of the measurement, and even, if we are feeling particularly conspiratorial, to making the choice of measurement itself. Removing then to the satisfaction of everyone may require human observers, which, given human reaction time of perhaps 0.1s, would require a spacelike separation of around $30000km$. Such experiments are a long way from the current situation.

6.5 The Angular Correlation Loophole

For historical value, I shall now discuss an older loophole, which applied to many of the first experiments, but does not apply to the modern ones. This is the angular correlation loophole.

In addition to the detection and locality loopholes, many of the early experiments, such as Freedman and Clauser's [101], and Aspect's [104], suffered from the angular correlation loophole [9, 10, 113, 114]. In both experiments, pairs of photons are emitted in spherical waves from a source. Thus the two photons may go in any two directions, not necessarily opposite to each other. Alice puts a detector which collects a certain range of directions on her side, and Bob does likewise. Sometimes photons are detected on one or both sides, but frequently they pass unnoticed, simply missing the detectors. This applies even if they have perfect detectors. This opens a loophole formally similar to the detection loophole, but of quite different experimental origin. One could of course make the detectors at Alice and Bob's side cover a large range of directions, almost a hemisphere on each side. However the polarization correlation between the photons conditioned upon such a measurement is quite weak, and decreases with the size of detector. Because of this, the experimental data do not violate the CH inequality, whatever size detector we put [10, 113, 114]. Aspect reported violating the CHSH inequality, but this was after ignoring all data where there was not a photon detected on both sides. This is equivalent to assuming that the LHV model would not try to exploit the no-detect outcomes, and so is an additional assuption.

In fact, based upon the angular correlation loophole, Clauser and Horne [10] have given an explicit LHV model which reproduces the quantum mechanical correlations of Freedman and Clauser's [101] experiment, and Santos [114] has done the same for Aspect's [104] experiment. Thus these experiments cannot demonstrate conclusively that the world is non-local, even with perfect detectors. Though these experiments are not currently in favour for testing non-locality, it was shown by Popescu [74] that despite the LHV models, these experiments did contain non-local correlations, which should be detectable if only we could perform the right measurements. He showed that if one performed a sequence of measurements (see section 4.2), one could detect the non-locality. The first measurement would be Alice and Bob each making a local non-demolition measurement to see whether Alice's photon was in a direction very close to one direction $\vec{n}$, and Bob's very close to being in the opposite direction, $-\vec{n}$. The second is an ordinary CHSH test on the photons, conditional on them being both detected in the first measurements. Popescu showed that with perfect (or very good) detectors, and with randomly selected measurements for the second measurements in the sequence, this would violate the CHSH (or CH) inequality, loophole free. Thus, at least in principle, the quantum mechanical description of these entangled particles is non-local.

The angular correlation loophole does not enter most modern experiments (eg. [93, 25]), which use parametric down conversion to produce pairs of photons in two narrow beams, thus removing the angular distribution almost completely.

Having reviewed the well known loopholes, I now turn to our contribution, the Memory loophole.

6.6 The Memory Loophole

In this section we present and formalise the memory loophole. In section 6.7, we emphasize the probabilistic nature of the Bell inequality, and introduce a linearised version of the CHSH inequality [9], which we will later show to be unaffected by the memory loophole. In section 6.8, we summarize our results. Sections 6.9 to 6.11 contain our main results. We analyze the inequality from the point of view of Bell's

original model and various different versions of the memory loophole. We show that the probability of violating a standard CHSH inequality is affected by the loophole, but that the effect is not significant for a large sample. Section 6.12 concludes.

Bell's gedanken experiment is performed by two observers, Alice and Bob, situated in two space separated regions. A source emits a pair of particles, one to Alice and one to Bob. The standard assumption of LHV is that if Alice performs any arbitrary measurement A and Bob performs any arbitrary measurement B, and the measurements are timed so that they take place outside the light-cone of each other, then there exists a shared random variable λ, with distribution $\mu(\lambda)$, and local distributions $P_A(a;\lambda)$ and $P_B(b;\lambda)$ such that the joint probability that the measurement of A yields a and the measurement of B yields b is given by

$$P_{AB}(a,b) = \int P_A(a;\lambda)P_B(b;\lambda)\mu(\lambda)d\lambda, \tag{6.6.1}$$

for all possible measurements A and B. One then makes Bell inequalities which the correlations in the outcomes generated by all such models must satisfy, and which quantum mechanics violates.

In order to determine correlations one has to perform measurements not on a single pair of particles but on many such pairs, and gather a large number of outcomes which will determine the statistics. The natural way to do this is sequentially, with the source emitting one pair after the other, and Alice and Bob making measurements upon each pair as they arrive. Now, according to the LHV model above (6.6.1), all the pairs in the ensemble are *uncorrelated*. This assumption appears natural from the perspective of quantum mechanics. In quantum theory, when we have a number of pairs, each pair being described by the same wave-function, the pairs are uncorrelated. However, we can imagine the following scenario. A first pair of particles is emitted by the source. One of the particles arrives at Alice and it is subjected to a measurement and gives an outcome according to the LHV model (6.6.1). However, it also leaves in the environment information indicating to what measurement it was subjected and what outcome it yielded. Now, when a particle in the second pair arrives at Alice, it will read this message and it will give an outcome which depends not only on the measurement, A^2, it is subjected to, but also on the

message left by the first particle, i.e. on which measurement, A^1, was performed upon the first particle, and what outcome, a^1, it gave. Particles on Bob's side behave in a similar way. The consequence is that the original LHV model (6.6.1) is now replaced by

$$P_{A^n B^n}(a^n, b^n) = \int P_{A^n}(a^n; M, \lambda)\, P_{B^n}(b^n; M, \lambda)\, \mu(\lambda) d\lambda\,, \tag{6.6.2}$$

where

$$P_{A^n}(a^n; M, \lambda) = P_{A^n}(a^n; A^1, ..., A^{n-1}, a^1, ..., a^{n-1}, \lambda) \tag{6.6.3}$$

and

$$P_{B^n}(b^n; M, \lambda) = P_{B^n}(b^n; B^1, ..., B^{n-1}, b^1, ..., b^{n-1}, \lambda)\,. \tag{6.6.4}$$

Here M stands for the local record, or *memory*, of the previous measurements. We call this a local hidden variable model with *1-sided memory*.

There is a further interesting variation of Bell's original model. Suppose that the source emits pairs of correlated particles one by one. Suppose too that on each pair Alice and Bob perform their measurements space-like separated from one another, so while Alice is performing her measurement no signal can arrive from Bob's measurement. However, the time between the measurements on the different pairs is long enough, so that by the time Alice measures her n-th particle, the particle could have received information about what has happened in Bob's measurements on all previous particles $(1, \ldots, n-1)$, and similarly for Bob. One could imagine local hidden variable models in which this information is indeed communicated and used, in which case the probability in (6.6.3) is replaced by

$$P_{A^n}(a^n; M, \lambda) = P_{A^n}(a^n; A^1, ..., A^{n-1}, a^1, ..., a^{n-1}, B^1, ..., B^{n-1}, b^1, ..., b^{n-1}, \lambda) \tag{6.6.5}$$

and similarly for the probability on Bob's side. This is a local hidden variable model with *2-sided memory*.

In principle, Bell's original argument can be extended to render both types of memory loophole irrelevant. We could require that separated apparatuses are used for each particle pair, and that *every* measurement is space-like separated from every other — but it seems unlikely that such an experiment will be done any time

soon with a large enough sample of particles to demonstrate statistically significant violations of Bell inequalities. Even the much weaker constraint that all of Alice's measurements are space-like separated from all of Bob's — which would exclude the 2-sided but not the 1-sided loophole — has not been satisfied in any experiment to date. (See, e.g., [11] and references therein).

Having established, therefore, that the original version of the local hidden variables model as proposed by Bell has to be modified, we now examine the consequences.

6.7 CHSH-Type Inequalities. General Considerations.

We first revisit the usual Bell inequalities experiment, and emphasize in more detail the statistical aspects of the measurements.

We shall use the version of the CHSH inequality in the form introduced in chapter 3:

$$P_{CHSH} = P(A = B) + P(A = \tilde{B}) + P(\tilde{A} = B) + P(\tilde{A} = \tilde{B} + 1) \leq 3 \qquad (6.7.1)$$

for local hidden variable theories, where all measurements have two possible outcomes, 0 and 1, $P(A = B)$ is the probability that A and B have the same outcome (ie. are correlated), and $P(A = B + 1)$ is the probability that $A = B + 1$ modulo 2 (ie. are anti-correlated). $A, \tilde{A}, B, \tilde{B}$ are chosen so that quantum mechanics predicts the maximal value, $P_{CHSH} = 2 + \sqrt{2}$. It is claimed that every ordinary (i.e. as originally constructed by Bell) local hidden variables model must obey the inequality.

Of course, even in an ideal experiment, an ordinary local hidden variables model can violate the CHSH bound. The quantities which figure in the CHSH expression are theoretical probabilities, which are abstract concepts. In reality each probability is determined by repeating a measurement a large number of times and estimating the probabilities as frequencies of events. These measured probabilities are subject to statistical fluctuations, which can yield violations of the CHSH bound. Our first

task is to examine the problem in detail, defining precisely the operational meaning of the different quantities, and get an accurate understanding of what exactly is the meaning of violation of Bell's inequalities. Only after all these are clarified will we be able to see the effect of the various memory loopholes. In particular, we will see that memory can allow particles to take advantage of statistical fluctuations and build them up into a systematic bias. We will also see, however, that, if the CHSH expressions are defined in the usual way, the biases that can thus be obtained tend to zero as the number of pairs tested increases. Moreover, we will see that a simpler linearised form of the CHSH expressions is "memory-proof", in the sense that the probability of a given level of violation is no greater for memory-dependent local hidden variable models than for optimally chosen memoryless models.

What we mean by (6.7.1) in an experimental context is the following. We suppose that Alice and Bob perform measurements on N pairs of particles. For each of their particles Alice and Bob choose at random what measurement to perform, A or $\tilde{A}$ for Alice and B or $\tilde{B}$ for Bob. We define $\#(A,B)$ to be the number of pairs on which operators A and B were measured, $\#(A=B)$ and $\#(A=B+1)$ to be the number of times the outcomes were correlated and anti-correlated in these measurements. Note that Alice and Bob should not pre-arrange the sequence of their measurements - this would introduce well-known loopholes; the entire experiments of Alice and Bob, including the decision of what to measure on each particle have to be space-like separated from each other. Consequently Alice and Bob do not have total control on how many times a specific pair of measurements, say A, B is performed, but this number, $\#(A,B)$ is a random variable.

We define

$$X_N = \frac{\#(A=B)}{\#(A,B)} + \frac{\#(A=\tilde{B})}{\#(A,\tilde{B})} + \frac{\#(\tilde{A}=B)}{\#(\tilde{A},B)} + \frac{\#(\tilde{A}=\tilde{B}+1)}{\#(\tilde{A},\tilde{B})}\,, \tag{6.7.2}$$

$$Y_N = \frac{4}{N}(\#(A=B) + \#(A=\tilde{B}) + \#(\tilde{A}=B) + \#(\tilde{A}=\tilde{B}+1))\,. \tag{6.7.3}$$

X_N is the experimental meaning of the CHSH inequality (6.7.1); the index N denotes that the experiment has been performed on N pairs. Indeed, the expression $\frac{\#(A=B)}{\#(A,B)}$ is the frequency of correlations between the outcomes of A and B, and it is

therefore the experimental definition of the correlation probability $P(A_1 = B_1)$ and so on.

Note that our definition of X_N assumes that $\#(A, B) > 0$ for all pairs of operators A, B. If not, X_N is undefined. Strictly speaking, our expressions for the expectation and other functions of X_N should thus all be conditioned on the event that X_N is defined. We will neglect this below, assuming that N is large enough that the probability of X_N being undefined is negligible. One could, alternatively, use an experimental protocol which ensures that X_N is defined. For instance, one could require that, if $\#(A, B) = 0$ for any A, B after N pairs have been tested, the experiment continues on further pairs until $\#(A, B) > 0$ for all A, B, and then terminates. Our analysis would need to be modified slightly to apply to such a protocol, but the results would be essentially the same.

Y_N is another experimental quantity closely related to X_N. The two quantities are equal if the four combinations of possible measurements (A, B), etc. are measured equal numbers of times. For large N the four combinations of possible measurement are very likely to be made almost equally often, and so X_N and Y_N are almost certain to be very close. Although it is traditional to use X_N in analyzing Bell experiments, Y_N is in fact much better behaved and easier to analyze, since it is a linear expression.

6.8 CHSH-type Inequalities. Expectation Values and Fluctuations.

X_N and Y_N represent quantities determined by making measurements on a batch of N pairs of particles. We do not assume the pairs behave independently: they may be influenced by memory, and we will analyze the different types of memories. We are interested in the maximum possible expectation value of X_N and Y_N, and the maximum probability of X_N or Y_N taking a value much larger than the expectation.

Obviously, the expectation and fluctuations of X_N and Y_N could be experimentally estimated only by repeating the whole series of N experiments a large number

of times, and then only under the assumption that different batches of N pairs behave independently. However, we do not wish to measure these quantities. All we wish to know is that the value of X_N which we measure in an experiment would be extremely unlikely to have occurred if the particles were governed by a local hidden variable model. For this, calculating the expectation and fluctuations of X_N and Y_N under the assumption of some LHV model are sufficient.

As we will see, it can be shown that the probability of obtaining experimental data consistent with quantum theory, given a local hidden variable theory using a memory loophole, for a large sample, is extremely small. Since the cumulative data in Bell experiments are indeed consistent with quantum theory, we conclude that they effectively refute the hypothesis of memory-dependent local hidden variables — so long, of course, as these hidden variables are assumed not also to exploit other well-known loopholes such as the detector efficiency loophole.

The results for which we have complete proofs can be summarized in the following table:

Table 6.1: Violations of Bell Inequalities by Memory LHV Models

LHV Model	$E(X_N)$	$P(\hat{X}_N > 5\delta)$	$E(Y_N)$	$P(\hat{Y}_N > \delta)$
Memoryless	≤ 3	$< 5f_N^\delta$	≤ 3	$< f_N^\delta$
1-sided Memory	$< 3 + o(N^{-1/2+\epsilon})$	$< 5f_N^\delta$	≤ 3	$< f_N^\delta$
2-sided Memory	$< 3 + o(N^{-1/2+\epsilon})$	$< 5f_N^\delta$	≤ 3	$< f_N^\delta$

Here $\hat{X}_N = X_N - 3$, $\hat{Y}_N = Y_N - 3$, and we have simplified the presentation by taking δ to be small enough that $(3+\delta) < (3+5\delta)(1-\delta)$. The expression $o(N^{-1/2+\epsilon})$ denotes a term that asymptotically tends to zero faster than $N^{-1/2+\epsilon}$ for any $\epsilon > 0$.

$$f_N^\delta = \frac{1}{\sqrt{2\pi}} \frac{\sqrt{3}}{\delta\sqrt{N}} \exp\left(-\frac{1}{6}\delta^2 N\right). \tag{6.8.1}$$

The proofs are given in the following sections.

The significance of these results is as follows. The memoryless case represents the results for standard local hidden variables which behave independently for each pair. The result $E(X_N) \leq 3$ is the standard expression of the CHSH inequality.

Although values of X_N larger than 3 can be experimentally obtained from a local hidden variables model, the probability of obtaining $3+5\delta$ decreases exponentially as $5f_N^\delta$. Hence, for a given δ and sufficiently large N, observing $3+5\delta$ when performing N experiments can be taken as a very good confirmation of the fact that it is not due to an LHV model. In the memoryless case, $E(Y_N) \leq 3$ and the fluctuations also decrease exponentially.

In the 2-sided memory case, the expectation value of Y_N again satisfies $E(Y_N) \leq 3$. Hence the existence of memory makes no difference here. Memory also makes no difference to the fluctuations: they still decrease exponentially. On the other hand, the expectation value of X_N can be larger than in the standard memoryless case. Hypothetically, if Bell experiments are analysed by using X_N and the effect of the memory loophole is neglected, a 2-sided memory LHV model could mistakenly be interpreted as exhibiting non-locality. Fortunately, we can put an upper bound of

$$E(X_N) \leq 3 + 5N^{-1/2+\epsilon} + 5\sqrt{3/2\pi}N^{-\epsilon}\exp(-N^{2\epsilon}/6)\,, \tag{6.8.2}$$

for any small $\epsilon > 0$. Thus, for large enough N, X_N is almost as good as Y_N at distinguishing quantum theory from local hidden variable models.

In the 1-sided memory case, we can use the 2-sided memory results to show that Y_N is unaffected by the presence of memory, and X_N is affected in a negligible way for sufficiently large N. Actually, we have not succeeded in finding a 1-sided memory model for which $E(X_N)$ or $P(\hat{X}_N > \delta)$ are larger than the maximal values attainable by memoryless models, for any N. We thus cannot exclude the possibility that 1-sided memory is of no use at all in helping LHV models come closer to reproducing quantum mechanics.

6.9 CHSH-Type Inequalities in Bell's No Memory Model

We first revisit the derivation of the CHSH inequality in Bell's model, using techniques which will be useful for analyzing the different memory models.

We first recall how these quantities are interpreted in standard analyses, when the Bell pairs are measured sequentially and the memory loophole is neglected. Let Z_N be a binomially distributed variable with N trials, each of which has the two possible outcomes 0 and 1, with probability $p \neq 0, 1$ of outcome 1 for each trial. The normal approximation to the binomial distribution gives us that

$$P(Z_N > pN + z\sqrt{Np(1-p)}) \to 1 - \mathcal{N}(z) \tag{6.9.1}$$

as $N \to \infty$, where

$$\mathcal{N}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{1}{2}y^2\right) dy \tag{6.9.2}$$

is the normal distribution function, which obeys

$$1 - \mathcal{N}(z) \approx \frac{1}{\sqrt{2\pi}} z^{-1} \exp\left(-\frac{1}{2}z^2\right) . \tag{6.9.3}$$

For large N, and for z large compared to 1 and small compared to $N^{1/2}$, the errors in these approximations are small and can be rigorously bounded[115]. Below we consider N and z in these ranges and neglect the error terms, which make no essential difference to the discussion.

Now

$$Y_N = \frac{4}{N} \sum_{n=1}^{N} Y_N^n \,, \tag{6.9.4}$$

where

$$Y_N^n = \delta^n(A = B) + \delta^n(A = \tilde{B}) + \delta^n(\tilde{A} = B) + \delta^n(\tilde{A} = \tilde{B} + 1) \,. \tag{6.9.5}$$

Here $\delta^n(A = B)$ is 1 if A and B are measured at the n^{th} round and found to be the same, and 0 otherwise, and $\delta^n(A = B + 1)$ is 1 if A and B are measured at the n^{th} round and found to be different, and 0 otherwise.

In a memoryless local hidden variable theory, the Y_N^n are independent random variables taking values 0 or 1. We have that

$$E(\delta^n(A = B)) = \frac{1}{4} p^n(A = B), \tag{6.9.6}$$

where $p^n(A = B)$ is the probability that $A = B$ if (A, B) is measured at the n^{th} trial, and similarly for the other three terms in (6.9.5). So, from (6.7.1) we have

that

$$y_n = E(Y_N^n) = \frac{P_{CHSH}}{4} \leq \frac{3}{4}\,. \tag{6.9.7}$$

Clearly, for any N and any $\delta > 0$, the probability $P(Y_N > 3+\delta)$ is maximised when the Y_N^n are identically distributed, with $y_n = 3/4$ for all n. For small δ we have that

$$\begin{aligned} P(Y_N > 3+\delta) &= P(NY_N/4 > 3N/4 + \delta N/4) \\ &\approx 1 - \mathcal{N}(\delta\sqrt{N}/\sqrt{3}) \\ &\approx \frac{1}{\sqrt{2\pi}}\frac{\sqrt{3}}{\delta\sqrt{N}} \exp\left(-\frac{1}{6}\delta^2 N\right)\,, \end{aligned} \tag{6.9.8}$$

for large N, which tends to zero fast as $N \to \infty$. A similar argument shows that quantum mechanics predicts that $P(Y_N < 2 + \sqrt{2} - \delta)$ tends to zero fast. A long run of experiments can thus distinguish quantum mechanics and memoryless local hidden variables with near certainty.

Although the analysis of Y_N is simpler and arguably more natural, Bell experiments are traditionally interpreted via the quantity X_N. Since

$$\begin{aligned} E\left(\frac{\#(A=B)}{\#(A,B)}\right) &= \sum_{n=1}^{N} p(\#(A,B)=n)\frac{E(\#(A=B)|\#(A,B)=n)}{n} \\ &= \sum_{n=1}^{N} p(\#(A,B)=n)\frac{nP(A=B)}{n} \\ &= P(A=B)\,, \end{aligned} \tag{6.9.9}$$

and similarly $E(\frac{\#(A=B+1)}{\#(A,B)}) = P(A=B+1)$, equations (6.7.1) and (6.7.2) imply that $E(X_N) \leq 3$. (Recall that we assume the $n=0$ terms in these sums have negligible probability.)

Moreover, since

$$P(\#(A,B) < N/4(1-\delta)) \approx \frac{\sqrt{3}}{\delta\sqrt{2\pi N}} \exp\left(-\frac{1}{6}\delta^2 N\right)\,, \tag{6.9.10}$$

we have that

$$P\left(X_N > \frac{1}{1-\delta}Y_N\right) \lesssim \frac{4\sqrt{3}}{\delta\sqrt{2\pi N}} \exp\left(-\frac{1}{6}\delta^2 N\right) \tag{6.9.11}$$

and

$$P\left(X_N > \frac{3+\delta}{1-\delta}\right) \lesssim \frac{5\sqrt{3}}{\delta\sqrt{2\pi N}} \exp\left(-\frac{1}{6}\delta^2 N\right)\,. \tag{6.9.12}$$

Similarly, quantum mechanics predicts that $P(X_N < 2 + \sqrt{2} - \delta)$ tends to zero fast. Thus, for large N, X_N distinguishes the predictions of quantum mechanics and memoryless local hidden variables almost as well as Y_N does.

6.10 The Two-Sided Memory Loophole

Now we consider the case where the LHV model for N trials is allowed to exploit the memory loophole, predicting results at each round of measurement which may depend upon the previous measurements and outcomes on both sides.

Since equations (6.9.6) and (6.9.7) still hold, we have that

$$E(Y_N) = \frac{4}{N}\sum_{n=1}^{N} E(Y_N^n) \leq \frac{4}{N}\sum_{n=1}^{N}\frac{3}{4} = 3\,. \tag{6.10.1}$$

Thus memory does not help increase $E(Y_N)$. We shall now show that it does not help the probability of a large fluctuation in Y_N. First, we note that Y_N is just (a constant times) the sum of Y_N^n, where Y_N^n is a random variable at the nth trial. Now, Y_N^n can only take values of 0 or 1. To maximize the probability of a large Y_N, we should try to maximize the probability of each Y_N^n being 1. This at first appears complicated, since with memory LHV models there will be correlations between $P(Y_N^n = 1)$ for different n. The key is to note that, regardless of what happens in later rounds, for all LHV memory models,

$$P(Y_N^n = 1 \,|\, \text{events in trials } 1 \ldots n-1) \leq 3/4\,. \tag{6.10.2}$$

This is because, for any fixed set of events in the earlier rounds, the model in round n is just an LHV model, whose probabilities have been chosen with no prior knowledge of the measurements which will be performed in round n, and must therefore satisfy the CHSH inequality.

It follows that, for any N and any $\delta > 0$, the probability $P(Y_N > 3 + \delta)$ is maximised when $P(Y_N^n = 1) = 3/4$ for all n. But an LHV model can maximize the probability that $Y_N^n = 1$, for any n, by a strategy independent of the outcomes of the previous measurements, for instance by predicting the outcome 1 for any measurement on either side. Since $Y_N^n = 0$ or 1, any such strategy maximizes the probability

$P(Y_N > 3 + \delta)$, and so equation (6.9.8) still holds even when the memory loophole is taken into account. The memory loophole does not alter the distinguishability of the predictions of quantum mechanics and local hidden variables, if Y_N is used as the correlation measure, since neither the maximal expectation nor the maximal variance of Y_N are increased by memory-dependent strategies.

Now let us turn to X_N. We know that if the particles are described by identical LHV models, then $E(X_N) \leq 3$. Also, even when the particles have memory, equations (6.9.10-6.9.12) hold. Suppose we take $\delta = N^{-1/2+\epsilon}$, for some small $\epsilon > 0$, and let N be large enough that $\frac{3+\delta}{1-\delta} < 3 + 5\delta$. Then from (6.9.12), since X_N is always bounded by 4, we have that

$$\begin{aligned} E(X_N) &\leq 4P(X_N > 3 + 5\delta) + (3 + 5\delta)(1 - P(X_N > 3 + 5\delta)) \\ &\lesssim 3 + 5N^{-1/2+\epsilon} + 5\sqrt{3/2\pi}N^{-\epsilon}\exp(-N^{2\epsilon}/6)\,, \end{aligned} \tag{6.10.3}$$

so that $(E(X_N) - 3)$ is bounded by a term that decays faster than $N^{-1/2+\epsilon}$, for any $\epsilon > 0$. This means that no LHV model can produce $E(X_N)$ much above 3 for large N; it also means that the X_N remain efficient discriminators of quantum mechanics and local hidden variable theories even when the memory loophole is taken into account.

So far we have shown that the memory loophole makes no essential difference to Bell inequalities, so long as we use a large number of particles. We shall now show that if we only use a small number of particles, the 2-sided memory loophole does indeed make a difference. We shall give a memory-dependent LHV model with $E(X_N) > 3$. To construct a simple example, we take a model which gives $X_N = 3$ with certainty, and modify it a little so that the expectation increases above 3. We set $N = 101$. We can get $X_{101} = 3$, with certainty, simply by outputing $+1$ regardless of the observables measured. Our new model is identical to this one except for the case when after 100 measurements we have measured (A, B), $(A, \tilde{B})$ and $(\tilde{A}, B)$ 33 times each, and $(\tilde{A}, \tilde{B})$ once. Our new model is allowed memory, so it can count how many times the various observables are measured, and thus tell when this is the case. In this (rather unlikely) case, the new model will output $+1$ on side A regardless of which measurement is performed, and output $B = +1$ if B

is measured, or $\tilde{B} = -1$ if $\tilde{B}$ is measured.

The two models will give identical values for X_{101} unless the above unusual state of affairs occurs after 100 rounds. Conditioned upon this event occurring, the old model still has an expectation of X_{101} equal to 3, whereas the new model has slightly more, almost 25/8. Since the expectation of the new model is the same as that of the old model in all other cases, this increases the unconditional expectation of the new model to very slightly greater than 3.

The intuition behind the modification is that if one term in X_N (e.g. $\frac{\#(A=B)}{\#(A,B)}$) has a small denominator compared to another term, then we will gain more by increasing the numerator in the term with the small denominator than in the term with the big denominator.

Now that we have this model with $E(X_N) > 3$, it is easy to see how to modify it to make a model which does better. The idea is to start trying to increase the numerator in the best places from the start. In each round, there are 4 possible pairs of observables which could be measured ((A, B), $(A, \tilde{B})$, etc.). We can send a list which is guaranteed to give the correct sort of correlation or anti-correlation to at most 3 of the possible pairs, where we can choose which ones. So at each stage our model must choose one pair which, if measured, will give the wrong sort of correlation. After all the measurements are finished, the model would like to give the "incorrect" correlation to the pair of observables which has been measured most (since this term has the biggest denominator). There is no way for it to be sure of doing this, since it does not know at the start which pair will be measured most. So, our new model simply guesses.

More precisely, the improved model is as follows. In the first round of measurements it gives outcome +1, whatever is measured. From the second round it looks to see which pair, eg. $(A, \tilde{B})$, has been measured most, and arranges that if that pair is measured in the next round, the correlations will be "incorrect", whereas if any other pair is measured in the next round the correlations will be "correct". It is easy to see this model produces $E(X_N) > 3$ for all N large enough that there is a negligible probability of one of the four observable pairs not being measured. Of course, our earlier bounds imply that $E(X_N) \to 3$ as $N \to \infty$. We conjecture

that the model produces the maximum value of $E(X_N)$ attainable by a local hidden variable theory with 2-sided memory.

6.11 The One-Sided Memory Loophole

We comment briefly on the case of the 1-sided memory loophole, represented by a model of the form (6.6.2). We do not know whether such models can increase the value of $E(X_N)$ above 3, or come any closer to simulating quantum theory than memoryless LHV models. Note, however, that 1-sided memory models are a restricted class of the 2-sided memory models, and thus all the upper bounds proven for 2-sided models still apply. In particular, $E(Y_N) \leq 3$, and equation (6.9.8) still holds, ie. $P(Y_N > 3 + \delta) \approx \frac{1}{\sqrt{2\pi}} \frac{\sqrt{3}}{\delta\sqrt{N}} \exp(-\frac{1}{6}\delta^2 N)$. These are in fact tight bounds, since they can be obtained without any memory.

The two sided bounds also apply for X_N. However, we do not know whether they are tight: it may be that 1-sided memory LHV models are no more powerful than memoryless LHV models.

6.12 Conclusion

We have seen that in the analysis of Bell-type experiments, one ought to allow for the possibility that the particles have memory, in the sense that outcomes of measurements on the nth pair of particles depend on both measurement choices and outcomes for the $1\text{st}, \ldots, (n-1)\text{th}$ pairs. The standard form for local hidden variable models, originally due to Bell and summarized in equation (6.6.1), does not allow for this possibility, so a new analysis is needed. We have distinguished 1-sided and 2-sided versions of this loophole and shown that in the 2-sided case, a systematic violation of a Bell-type inequality can be obtained. In the case of the CHSH inequality, however, we have derived an upper bound on the probability of large deviations and thereby shown that the expected violation tends to zero as the number of particle pairs tested becomes large. Thus the CHSH inequality is robust against the memory loophole and the corresponding experimental tests remain good

discriminators between quantum mechanics and local hidden variables — there is no need to design improved experiments in which more (or even all) measurements are space-like separated from one another.

We have also shown that if the analysis is performed in terms of the quantities Y_N rather than X_N then the memory models have no advantage over standard memoryless local hidden variable models. Thus these quantities are better suited to testing experimental data, and we advocate their use in the analysis of future experiments.

Chapter 7

The Non-Local Content of Quantum Operations

7.1 Introduction

In the past, and in this thesis, most of the research on quantum non-locality has been devoted to the issue of non-locality of *quantum states*. However I feel that an equally important issue is that of non-locality of *quantum evolutions*. That is, in parallel with the understanding of non-locality of quantum kinematics one should also develop an understanding of the non-locality of quantum dynamics.

Let us start with a simple example. Consider two qubits situated far from each other, one held by Alice and the other one by Bob. Suppose they would like to implement a two qubit quantum evolution described by the unitary operator U. (We wish to be able to apply U on *any* initial state of the two qubits). With exception of the case when U is a product of two local unitary operators, $U = U_A \otimes U_B$, no other quantum evolution can be accomplished by local means only. Thus almost all quantum evolutions are non-local. The main question I address in this chapter is how to describe, qualitatively and quantitatively, the non-locality of quantum evolutions. The framework I shall describe was proposed by Noah Linden and Sandu Popescu. I joined them to find the first results within this framework[51], results I give here. The framework and many of the results which we described have also been discovered

independently by A. Chefles, C. R. Gilson and S. M. Barnett [52], and by J. Eisert, K. Jacobs, P. Papadopolous and M. B. Plenio[53]. P. Zanardi and co-workers have studied the related issue of how much entanglement a unitary operation can create, averaged over all unentangled input states [116, 117, 118].

Since our research on this topic, considerable work was performed by various other parties. I shall survey this work at the end of this chapter.

In order to be able to describe the amount of non-locality contained by the unitary operator U the following approach was suggested. Consider that Alice and Bob, in addition of being able to perform any local operations, also have *additional resources*, namely they share entangled states, and they are able to communicate classically. The question then reduces to finding out how much of these resources are needed to implement U.

When thinking of non-locality, the role of quantum entanglement is clearly important, however in this scenario the role of the classical communication is equally important. Understanding the character of a quantum evolution requires knowing both the amount of entanglement and the amount of classical communication needed to perform the operation.

7.2 General Sufficiency Conditions

First of all, it is important to note that *any* unitary evolution can be implemented given enough shared entanglement and classical communication. Indeed, consider the case of two qubits, one held by Alice and one by Bob. Any unitary transformation U on these two qubits can be accomplished by having Alice teleport her qubit to Bob, Bob performing U locally and finally Bob teleporting Alice's qubit back to Alice. The resources needed for the two teleportation actions are: (1 e-bit plus two classical bits transmitted from Alice to Bob for the Alice to Bob teleportation) plus (1 e-bit plus two classical bits transmitted from Bob to Alice for the Bob to Alice teleportation). It is obvious now that any unitary operation involving any number of parties and any number of qubits can be accomplished by a similar procedure (teleporting all states to a single location, performing U locally and teleporting

back the qubits to their original locations).

The "double teleportation" procedure shown above is *sufficient* to implement any quantum evolution. The question is however whether so much resources are actually *needed*. We will discuss a couple of specific example below.

7.3 The SWAP Operation on Two Qubits

The SWAP operation defined by:

$$U_{\rm SWAP}|\psi\rangle \otimes |\phi\rangle = |\phi\rangle \otimes |\psi\rangle \tag{7.3.1}$$

is a particularly intriguing case, since although it takes product states to product states, it is, as we now show, the most non-local operation possible in the sense described above. That is, we will prove that in order to implement a SWAP on two qubits it is not only sufficient but also *necessary* to use 2 e-bits plus 2 bits of classical communication from Alice to Bob plus 2 bits of classical communication from Bob to Alice.

Proof: To prove that the SWAP operation *needs* as non-local resources 2 e-bits, we will show that if we have an apparatus able to implement the SWAP operation we can use it in order to create 2 e-bits. Thus, since entanglement cannot be created *ex nihilo*, the apparatus which implements the SWAP *must* use 2 e-bits as an internal non-local resource.

Let us show how to generate two singlets using the SWAP operation. Firstly Alice and Bob prepare singlets locally

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle_A |\uparrow\rangle_a + |\downarrow\rangle_A |\downarrow\rangle_a) \quad \text{and} \quad \frac{1}{\sqrt{2}}(|\uparrow\rangle_B |\uparrow\rangle_b + |\downarrow\rangle_B |\downarrow\rangle_b), \tag{7.3.2}$$

Alice's spins are labeled A and a and Bob's B and b. Now perform the SWAP operation on spins A and B:

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle_A |\uparrow\rangle_a + |\downarrow\rangle_A |\downarrow\rangle_a)\ \frac{1}{\sqrt{2}}(|\uparrow\rangle_B |\uparrow\rangle_b + |\downarrow\rangle_B |\downarrow\rangle_b) \mapsto$$

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle_B |\uparrow\rangle_a + |\downarrow\rangle_B |\downarrow\rangle_a)\ \frac{1}{\sqrt{2}}(|\uparrow\rangle_A |\uparrow\rangle_b + |\downarrow\rangle_A |\downarrow\rangle_b). \tag{7.3.3}$$

This state contains two singlets held between Alice and Bob.

To find the classical communication resources *needed* to implement the SWAP operation we will adapt an argument first given in [2]. We show that if we have an apparatus able to implement the SWAP operation we can use it in order to communicate 2 bits from Alice to Bob plus 2 bits from Bob to Alice. From this follows that it must be the case that the SWAP apparatus uses 2 bits of classical communication from Alice to Bob plus 2 bits of classical communication from Bob to Alice as an internal resource, otherwise Alice could receive information from Bob transmitted faster than light.

For suppose that the SWAP operation requires less than four bits of classical communication (two bits each way). Alice and Bob can produce an instantaneous SWAP operation which works correctly with probability greater than one sixteenth in the following way. Alice and Bob run the usual SWAP protocol, but instead of waiting for classical communication from each other, they simply guess the bits that they would have received. Since we have assumed that the SWAP operation requires less than 4 bits, the probability that Alice and Bob guess correctly is greater than one sixteenth and hence the SWAP operation also succeeds with probability greater than one sixteenth.

Thus using the protocol described previously can now use this imperfect, but instantaneous SWAP to communicate 4 bits instantaneously. The bits arrive correctly when the SWAP is implemented correctly. Hence the probability that 4 bits arrive correctly is larger than one sixteenth; 4 bits communicated correctly with probability greater than one sixteenth represents a non-zero amount of information. Thus Alice and Bob have managed to convey some information to each other instantaneously. We conclude therefore that the SWAP operation cannot be done with less that 4 bits of classical communication; otherwise it allows communication faster than the speed of light.

Earlier in this section we showed that the SWAP operation can be used to generate two singlets. We now show that the SWAP operation can be also be used to perform four bits of classical communication (two bits each way): the main idea is that of "super-dense coding" [17]. Suppose that initially Alice and Bob share two

singlets:

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle_A |\uparrow\rangle_B + |\downarrow\rangle_A |\downarrow\rangle_B) \quad \text{and} \quad \frac{1}{\sqrt{2}}(|\uparrow\rangle_a |\uparrow\rangle_b + |\downarrow\rangle_a |\downarrow\rangle_b .) \tag{7.3.4}$$

Now Alice chooses one of four local unitary operations $\mathbb{1}$ (identity), σ_x, σ_y, σ_z and performs it on her spin A. This causes the first singlet to be in one of the four Bell states. Bob also, independently chooses one of these four locally unitaries and performs it on his spin b, putting the second singlet into one of the Bell states. Then the SWAP operation is performed on spins A and b. Now both Bob and Alice have one of the Bell states locally; which one they have depends on which operation the other performed. By measurement, they can work out which of the four unitaries the other performed. Thus the SWAP operation has enabled two bits of classical communication to be performed each way.

7.4 The CNOT Operation on Two Qubits

Another important quantum operation is CNOT, defined as

$$|\uparrow\rangle |\uparrow\rangle \mapsto |\uparrow\rangle |\uparrow\rangle \tag{7.4.1}$$

$$|\uparrow\rangle |\downarrow\rangle \mapsto |\uparrow\rangle |\downarrow\rangle \tag{7.4.2}$$

$$|\downarrow\rangle |\uparrow\rangle \mapsto |\downarrow\rangle |\downarrow\rangle \tag{7.4.3}$$

$$|\downarrow\rangle |\downarrow\rangle \mapsto |\downarrow\rangle |\uparrow\rangle \tag{7.4.4}$$

As we prove below, the necessary and sufficient resources for CNOT are 1 e-bit plus 1 bit of classical communication from Alice to Bob plus 1 bit of classical communication from Bob to Alice.

Proof: Constructing a CNOT We now show how to construct the CNOT operation using one singlet and two bits of classical communication. We then show

how to generate one singlet or perform two bits of classical communication using the CNOT.

Firstly we will show how, using one singlet and one bit of classical communication each way, we can perform a CNOT on the state

$$(\alpha\,|\uparrow\rangle_A + \beta\,|\downarrow\rangle_A)\;(\gamma\,|\uparrow\rangle_B + \delta\,|\downarrow\rangle_B) \tag{7.4.5}$$

i.e. transform it to

$$\alpha\,|\uparrow\rangle_A\,(\gamma\,|\uparrow\rangle_B + \delta\,|\downarrow\rangle_B) + \beta\,|\downarrow\rangle_A\,(\gamma\,|\downarrow\rangle_B + \delta\,|\uparrow\rangle_B). \tag{7.4.6}$$

Here α, β, γ and δ are complex numbers such that $|\alpha|^2+|\beta|^2 = 1$, and $|\gamma|^2+|\delta|^2 = 1$. Since the operation behaves linearly, the protocol performs the CNOT on any input state (i.e. even if the qubits are entangled with each other or with other systems).

Step 1 The first step is to append a singlet held between Alice and Bob to the state (7.4.5):

$$(\alpha\,|\uparrow\rangle_A + \beta\,|\downarrow\rangle_A)\;(|\uparrow\rangle_a\,|\uparrow\rangle_b + |\downarrow\rangle_a\,|\downarrow\rangle_b)\;(\gamma\,|\uparrow\rangle_B + \delta\,|\downarrow\rangle_B), \tag{7.4.7}$$

then for Alice to measure the total spin of her spins A and a.

If the total spin is one, then the spins A and a are in the subspace spanned by $|\uparrow\rangle_A\,|\uparrow\rangle_a$ and $|\downarrow\rangle_A\,|\downarrow\rangle_a$. In this case the state becomes

$$(\alpha\,|\uparrow\rangle_A\,|\uparrow\rangle_a\,|\uparrow\rangle_b + \beta\,|\downarrow\rangle_A\,|\downarrow\rangle_a\,|\downarrow\rangle_b)\;(\gamma\,|\uparrow\rangle_B + \delta\,|\downarrow\rangle_B). \tag{7.4.8}$$

Now Alice disentangles the singlet spin by performing the following (local) operation:

$$|\uparrow\rangle_A\,|\uparrow\rangle_a \mapsto |\uparrow\rangle_A\,|\uparrow\rangle_a\,; \qquad |\downarrow\rangle_A\,|\downarrow\rangle_a \mapsto |\downarrow\rangle_A\,|\uparrow\rangle_a\,, \tag{7.4.9}$$

and the state becomes

$$(\alpha\,|\uparrow\rangle_A\,|\uparrow\rangle_b + \beta\,|\downarrow\rangle_A\,|\downarrow\rangle_b)\;(\gamma\,|\uparrow\rangle_B + \delta\,|\downarrow\rangle_B)\,|\uparrow\rangle_a\,. \tag{7.4.10}$$

If the total spin is zero, the the spins A and a are in the subspace spanned by $|\uparrow\rangle_A\,|\downarrow\rangle_a$ and $|\downarrow\rangle_A\,|\uparrow\rangle_a$. In this case, rather than (7.4.8) the state becomes

$$(\alpha\,|\uparrow\rangle_A\,|\downarrow\rangle_a\,|\downarrow\rangle_b + \beta\,|\downarrow\rangle_A\,|\uparrow\rangle_a\,|\uparrow\rangle_b)\;(\gamma\,|\uparrow\rangle_B + \delta\,|\downarrow\rangle_B). \tag{7.4.11}$$

In this case Alice can disentangle the a spin by

$$|\uparrow\rangle_A |\downarrow\rangle_a \mapsto |\uparrow\rangle_A |\uparrow\rangle_a ; \qquad |\downarrow\rangle_A |\uparrow\rangle_a \mapsto |\downarrow\rangle_A |\uparrow\rangle_a , \tag{7.4.12}$$

leading to

$$(\alpha |\uparrow\rangle_A |\downarrow\rangle_b + \beta |\downarrow\rangle_A |\uparrow\rangle_b) (\gamma |\uparrow\rangle_B + \delta |\downarrow\rangle_B) |\uparrow\rangle_a . \tag{7.4.13}$$

In order to get this state in the correct form, Bob needs to invert his b spin. Thus Alice must communicate one bit to Bob to tell him whether she found total spin one or zero, and thus whether he needs to invert his spin or not.

After these operations, the state is

$$(\alpha |\uparrow\rangle_A |\uparrow\rangle_b + \beta |\downarrow\rangle_A |\downarrow\rangle_b) (\gamma |\uparrow\rangle_B + \delta |\downarrow\rangle_B) |\uparrow\rangle_a . \tag{7.4.14}$$

Step 2 Now Bob performs a CNOT on the b and B spins, thus the total state is

$$[\alpha |\uparrow\rangle_A |\uparrow\rangle_b (\gamma |\uparrow\rangle_B + \delta |\downarrow\rangle_B) + \beta |\downarrow\rangle_A |\downarrow\rangle_b (\gamma |\downarrow\rangle_B + \delta |\uparrow\rangle_B)] |\uparrow\rangle_a . \tag{7.4.15}$$

Step 3 Bob now measures σ_x on his part of the singlet b. ie. He measures whether spin b is $|\uparrow_x\rangle_b = \frac{1}{\sqrt{2}}(|\uparrow\rangle_b + |\downarrow\rangle_b)$ or $|\downarrow_x\rangle_b = \frac{1}{\sqrt{2}}(|\uparrow\rangle_b - |\downarrow\rangle_b)$ Either the state becomes

$$[\alpha |\uparrow\rangle_A (\gamma |\uparrow\rangle_B + \delta |\downarrow\rangle_B) + \beta |\downarrow\rangle_A (\gamma |\downarrow\rangle_B + \delta |\uparrow\rangle_B)] \ |\uparrow\rangle_a \ (|\uparrow\rangle_b + |\downarrow\rangle_b), \tag{7.4.16}$$

or

$$[\alpha |\uparrow\rangle_A (\gamma |\uparrow\rangle_B + \delta |\downarrow\rangle_B) - \beta |\downarrow\rangle_A (\gamma |\downarrow\rangle_B + \delta |\uparrow\rangle_B)] \ |\uparrow\rangle_a \ (|\uparrow\rangle_b - |\downarrow\rangle_b). \tag{7.4.17}$$

In the former case (i.e. the x component of spin b was $+$) we have performed the protocol as desired. In the latter, Alice needs to perform a σ_z rotation by π. ie. she must perform the transformation $|\uparrow\rangle_A \mapsto |\uparrow\rangle_A$, $|\downarrow\rangle_A \mapsto -|\downarrow\rangle_A$. Thus Bob needs to communicate one bit to Alice to tell her whether or not to perform the rotation.

We have thus shown how to perform a CNOT using one singlet and one bit of classical communication each way.

Creating entanglement by CNOT We show now that a CNOT apparatus can be used to create 1 e-bit between Alice and Bob; thus (since entanglement cannot

be increased by local operations) 1 e-bit is a necessary resource for constructing a CNOT.

Creating 1 e-bit by a CNOT is straightforward:

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle_A + |\downarrow\rangle_A)\,|\uparrow\rangle_B \mapsto \frac{1}{\sqrt{2}}(|\uparrow\rangle_A\,|\uparrow\rangle_B + |\downarrow\rangle_A\,|\downarrow\rangle_B). \tag{7.4.18}$$

Classical communication by CNOT

Suppose that Alice and Bob have an apparatus which implements a CNOT and also share 1 e-bit. They can use these resources to communicate *at the same time* 1 classical bit from Alice to Bob and 1 classical bit from Bob to Alice. This proves (see preceding section) that communicating 1 classical bit each way is a necessary resource for constructing a CNOT.

Suppose the initial state is

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle_a\,|\uparrow\rangle_b + |\downarrow\rangle_a\,|\downarrow\rangle_b). \tag{7.4.19}$$

Alice can encode a "0" by not doing anything to the state and a "1" by flipping her qubit. Bob can encode a "0" by not doing anything to the state and a "1" by changing the phase as follows: $|\uparrow\rangle_b \to |\uparrow\rangle_b$ and $|\downarrow\rangle_b \to -|\downarrow\rangle_b$.

The four states corresponding to the different bit combinations are thus

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle_a\,|\uparrow\rangle_b + |\downarrow\rangle_a\,|\downarrow\rangle_b) \quad \textit{corresponds to} \quad 0_A0_B. \tag{7.4.20}$$

$$\frac{1}{\sqrt{2}}(|\downarrow\rangle_a\,|\uparrow\rangle_b + |\uparrow\rangle_a\,|\downarrow\rangle_b) \quad \textit{corresponds to} \quad 1_A0_B. \tag{7.4.21}$$

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle_a\,|\uparrow\rangle_b - |\downarrow\rangle_a\,|\downarrow\rangle_b) \quad \textit{corresponds to} \quad 0_A1_B. \tag{7.4.22}$$

$$\frac{1}{\sqrt{2}}(|\downarrow\rangle_a\,|\uparrow\rangle_b - |\uparrow\rangle_a\,|\downarrow\rangle_b) \quad \textit{corresponds to} \quad 1_A1_B. \tag{7.4.23}$$

After encoding their bits, Alice and Bob apply the CNOT operation. This results in the corresponding four states

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle_a\,|\uparrow\rangle_b + |\downarrow\rangle_a\,|\uparrow\rangle_b) = \frac{1}{\sqrt{2}}(|\uparrow\rangle_a + |\downarrow\rangle_a)\,|\uparrow\rangle_b \quad \textit{corresponds to} \quad 0_A0_B \tag{7.4.24}$$

$$\frac{1}{\sqrt{2}}(|\downarrow\rangle_a |\downarrow\rangle_b + |\uparrow\rangle_a |\downarrow\rangle_b) = \frac{1}{\sqrt{2}}(|\uparrow\rangle_a + |\downarrow\rangle_a) |\downarrow\rangle_b \quad \textit{corresponds to} \quad 1_A 0_B \tag{7.4.25}$$

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle_a |\uparrow\rangle_b - |\downarrow\rangle_a |\uparrow\rangle_b) = \frac{1}{\sqrt{2}}(|\uparrow\rangle_a - |\downarrow\rangle_a) |\uparrow\rangle_b \quad \textit{corresponds to} \quad 0_A 1_B \tag{7.4.26}$$

$$\frac{1}{\sqrt{2}}(|\downarrow\rangle_a |\downarrow\rangle_b - |\uparrow\rangle_a |\downarrow\rangle_b) = \frac{1}{\sqrt{2}}(|\downarrow\rangle_a - |\uparrow\rangle_a) |\downarrow\rangle_b \quad \textit{corresponds to} \quad 1_A 1_B. \tag{7.4.27}$$

Bob can now find out Alice's bit by measuring his qubit in the $\{|\uparrow\rangle_b, |\downarrow\rangle_b\}$ basis while Alice can find out Bob's bit by measuring her qubit in the $\{\frac{1}{\sqrt{2}}(|\uparrow\rangle_a + |\downarrow\rangle_a), \frac{1}{\sqrt{2}}(|\uparrow\rangle_a - |\downarrow\rangle_a)\}$ basis.

7.5 The Double CNOT Operation on Two Qubits

One might have thought that the SWAP operation was the *unique* maximally non-local operation, at least in the terms we use to classify such operations. We here demonstrate that there is another maximally non-local operator, which is the "Double CNOT", or "DCNOT" gate, formed by acting a CNOT from particle 1 onto particle 2, and then a second CNOT from particle 2 onto particle 1. It is defined by

$$|\uparrow\rangle |\uparrow\rangle \mapsto |\uparrow\rangle |\uparrow\rangle \tag{7.5.1}$$

$$|\uparrow\rangle |\downarrow\rangle \mapsto |\downarrow\rangle |\downarrow\rangle \tag{7.5.2}$$

$$|\downarrow\rangle |\uparrow\rangle \mapsto |\uparrow\rangle |\downarrow\rangle \tag{7.5.3}$$

$$|\downarrow\rangle |\downarrow\rangle \mapsto |\downarrow\rangle |\uparrow\rangle . \tag{7.5.4}$$

To show that DCNOT is maximally non-local, we shall first demonstrate that it can be used to create 2 e-bits. We shall then show that it can be used to communicate

2 bits of information from Alice to Bob, and simultaneously to send 2 bits from Bob to Alice. The argument used for the SWAP operation then proves that to build a DCNOT we need 2 e-bits plus 2 bits of classical communication from Alice to Bob plus 2 bits of classical communication from Bob to Alice. Since any transformation on two qubits can be performed using these resources via teleportation, we will then have shown that the DCNOT is maximally non-local, in terms of resources.

Creating 2 e-bits is easy. Alice and Bob prepare singlets locally, and then perform the DCNOT on spins A and B:

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle_A|\uparrow\rangle_a + |\downarrow\rangle_A|\downarrow\rangle_a)\ \frac{1}{\sqrt{2}}(|\uparrow\rangle_B|\uparrow\rangle_b + |\downarrow\rangle_B|\downarrow\rangle_b) \mapsto$$

$$\frac{1}{2}(|\uparrow\rangle_A|\uparrow\rangle_a|\uparrow\rangle_B|\uparrow\rangle_b + |\downarrow\rangle_A|\uparrow\rangle_a|\downarrow\rangle_B|\downarrow\rangle_b + |\uparrow\rangle_A|\downarrow\rangle_a|\downarrow\rangle_B|\uparrow\rangle_b + |\downarrow\rangle_A|\downarrow\rangle_a|\uparrow\rangle_B|\downarrow\rangle_b). \tag{7.5.5}$$

We now have a Schmidt decomposition of rank 4, ie. a 2 party state which is locally equivalent to 2 e-bits between Alice and Bob.

Transmitting 2 bits of information in both directions at the same time is a little more tricky. Alice and Bob need to have 2 e-bits in addition to the DCNOT operation. They first transform their e-bits (locally) into the state

$$\frac{1}{2}(|\uparrow\rangle_A|\uparrow\rangle_a|\uparrow\rangle_B|\uparrow\rangle_b + |\downarrow\rangle_A|\uparrow\rangle_a|\uparrow\rangle_B|\downarrow\rangle_b + |\downarrow\rangle_A|\downarrow\rangle_a|\downarrow\rangle_B|\uparrow\rangle_b + |\uparrow\rangle_A|\downarrow\rangle_a|\downarrow\rangle_B|\downarrow\rangle_b). \tag{7.5.6}$$

Alice now encodes 1 bit of information in the state by either applying, or not applying $\sigma_z \otimes \sigma_z$ to her 2 spins. She encodes a second bit of information by applying, or not applying σ_x to her first spin, A. Bob similarly encodes two bits of information, using the transformation σ_z on spin B to encode his first bit, and $\sigma_x \otimes \sigma_x$ to encode his second bit.

Having encoded the information, they make it locally accessible by applying the DCNOT to spins A and B. It is not obvious, but simple to check, that Alice and Bob now each have one of the 4 Bell states locally, and that Alice's particular state corresponds to Bob's encoded bits, and vice-versa.

7.6 The Double CNOT is Locally Inequivalent to the SWAP

Having shown that the DCNOT really is maximally non-local in our sense, we will check that it is locally inequivalent to the SWAP. Though it looks different, one might have thought that, since both gates are maximally non-local, they could be converted one into the other using local ancillas and some local unitaries. We shall show that this is not the case. We do not allow any classical communication or entanglement, since these are the resources we are counting. Local measurements are already included since without communication we can imagine a measurement to be merely a local unitary upon the system and an ancilla, perhaps with the ancilla thrown away at the end.

To show that the SWAP and the DCNOT are different under local unitaries and ancillas we first show that using just local unitaries on the qubits we cannot turn a SWAP into a DCNOT (or vice-versa). ie. we show that

$$DCNOT_{AB} \neq U_A \otimes U_B \; SWAP_{AB} \; V_A \otimes V_B, \tag{7.6.1}$$

for any choice of local unitaries U_A, U_B, V_A, V_B. This is because the SWAP takes product states, $|\psi\rangle_A |\phi\rangle_B$, to product states, and the addition of local unitaries does not change this fact, whereas the DCNOT entangles some product states, eg.

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle_A + i\,|\downarrow\rangle_A)\frac{1}{\sqrt{2}}(|\uparrow\rangle_B + |\downarrow\rangle_B) \mapsto |\uparrow\rangle_A \frac{1}{\sqrt{2}}(|\uparrow\rangle_B + i\,|\downarrow\rangle_B) + |\downarrow\rangle_A \frac{1}{\sqrt{2}}(i\,|\uparrow\rangle_B + |\downarrow\rangle_B). \tag{7.6.2}$$

Next we show that one cannot build a SWAP using local unitaries and ancillas. ie.

$$SWAP_{AB} \otimes \mathbb{1}_{ab} \neq U_{Aa} \otimes U_{Bb} \; DCNOT_{AB} \; V_{Aa} \otimes V_{Bb}. \tag{7.6.3}$$

We shall prove this by contradiction. First we shall show that, if the unitaries V_{Aa}, V_{Bb} entangle the ancillas a, b with the main qubits A, B, then the DCNOT will take some initial product states $|\psi\rangle_A |\phi\rangle_B$ to states which are entangled between Alice and Bob (ie. between the system (A, a) and the system (B, b).) Since we are

trying to perform a SWAP, the final state should have no entanglement between A and B, thus the final local unitaries U_{Aa}, U_{Bb} must transfer all of this entanglement onto the ancillas. However, this gives us a way to turn a DCNOT into a SWAP plus some additional entanglement. This is impossible, since this would allow us to use classical communication and 2 e-bits to make a DCNOT, then local operations to make a SWAP and some additional entanglement, and then use the SWAP to make 2 e-bits, thus finishing with more entanglement than we started with! Thus the unitaries V_{Aa}, V_{Bb} cannot entangle the ancillas a, b and the qubits A, B. But then our previous result without ancillas (eqn. (7.6.1)) applies and so the DCNOT and the SWAP are different.

We now just need to show that if the unitaries V_{Aa}, V_{Bb} entangle the ancillas a, b with the main qubits A, B, then the DCNOT will take some initial product states $|\psi\rangle_A |\phi\rangle_B$ to states which are entangled between Alice and Bob. Suppose we input the state $|\psi\rangle_A |\phi\rangle_B$, along with the ancillas in some reference state, $|\uparrow\rangle_a |\uparrow\rangle_b$. We then entangle our ancillas using V_{Aa}, V_{Bb}.

$$|\psi\rangle_A |\uparrow\rangle_a |\phi\rangle_B |\uparrow\rangle_b \mapsto (|\uparrow\rangle_A |\chi\rangle_a + |\downarrow\rangle_A |\eta\rangle_a) (|\uparrow\rangle_B |\mu\rangle_b + |\downarrow\rangle_B |\nu\rangle_b) \tag{7.6.4}$$

where $|\chi\rangle_a, |\eta\rangle_a, |\mu\rangle_b, |\nu\rangle_b$ are not necessarily normalised or orthogonal. In order for ancilla a to be entangled, we need that:

$$|\chi\rangle_a \neq 0; |\eta\rangle_a \neq 0; |\chi\rangle_a \neq z |\eta\rangle_a \, \forall complex z \tag{7.6.5}$$

and similarly for Bob's ancilla b.

Now we perform the DCNOT on our state (AB), mapping it to:

$$|\uparrow\rangle_A |\chi\rangle_a |\uparrow\rangle_B |\mu\rangle_b + |\downarrow\rangle_A |\chi\rangle_a |\downarrow\rangle_B |\nu\rangle_b + |\uparrow\rangle_A |\eta\rangle_a |\downarrow\rangle_B |\mu\rangle_b + |\downarrow\rangle_A |\eta\rangle_a |\uparrow\rangle_B |\nu\rangle_b . \tag{7.6.6}$$

Now, if the protocol is successful, there should be no entanglement between Alice's side and Bob's side, ie. it should be a product state, $|\epsilon\rangle_{Aa} |\theta\rangle_{Bb}$. We can make a (non-orthonormal) basis for Aa (and thus write the state ϵ_{Aa} in a unique way) using the linearly independent vectors

$$|\uparrow\rangle_A |\chi\rangle_a , |\downarrow\rangle_A |\chi\rangle_a , |\uparrow\rangle_A |\eta\rangle_a , |\downarrow\rangle_A |\eta\rangle_a , \tag{7.6.7}$$

and some further vectors orthogonal to those four, if necessary. If we write out Alice's part of the state $|\epsilon\rangle_{Aa}$ in this way, and leave Bob's part $|\theta\rangle_{Bb}$ as it is, then we see that all the coefficients in Bob's part of the state in equation 7.6.6 must be proportional to $|\theta\rangle_{Bb}$. ie.

$$|\uparrow\rangle_B |\mu\rangle_b = z' |\theta\rangle_{Bb} ; |\downarrow\rangle_B |\mu\rangle_b = z'' |\theta\rangle_{Bb} ; etc. \tag{7.6.8}$$

Clearly this is impossible, so ancillas have not helped us to turn a DCNOT into a SWAP, and thus the two operations really are different. We note that one could consider a more general scenario involving a catalyst of entanglement or classical communication, and ask whether the two operations are still inequivalent. We do not know if the two operations are equivalent in this scenario, though we suspect not.

7.7 The Time of Operations

So far we have looked at entanglement and classical communication as the important factors for performing a quantum action. However, another quantity which we believe to be important is the *time* it takes to implement the action. This is natural from the point of view of quantum computation, where we are mainly interested in implementing some operation (the program) in the shortest possible time. Taking a concrete example in our case, earlier we showed that one can perform any operation using a "double teleportation" method. However, this procedure uses two rounds of communication, the first from Alice to Bob, and the second from Bob to Alice. We say that it takes 2 units of time to perform the operation. One can ask whether it is possible to perform any operation in a single unit of time. A further interesting question may be whether there is some tradeoff between time and the amount of entanglement and classical communication required to perform some operation.

We have found three operations which can be performed in unit time. The first

is the SWAP. This can be viewed as simply Alice's qubit teleported to Bob, and Bob's qubit teleported to Alice. These operations can be carried out simultaneously, and so the SWAP is performed in unit time. The next such operation is the CNOT. The protocol we gave for performing this operation used one e-bit and one bit in each direction and two units of time. Alice first made a measurement on her system and the e-bit, and sent the result to Bob. He then performed a "correcting" unitary operation upon his qubit depending upon the result of Alice's measurement. Then he performed a joint unitary operation and a measurement upon his qubit and his half of the (now collapsed) e-bit, and sent the result to Alice. She finally made a "correcting" unitary operation depending upon the outcome of Bob's measurement. Although this method uses two units of time, this is only because Bob waits for the result of Alice's measurement before performing his own measurement. He could perform his measurement at the beginning, and only perform the correcting unitary when he receives Bob's result. In principle this could mess up the protocol, however one can check that it does not, and the CNOT operation is successfully implemented (up to an overall phase) in a single unit of time.

This implementation of the CNOT operation in one unit of time can be used to implement the DCNOT in one unit of time, by performing two CNOT's one after the other. A straightforward implementation would take two units of time, but if one uses the same trick as for the CNOT, and does not wait for the messages in the first CNOT to be received before beginning to perform the second CNOT, one finds that the procedure successfully implements the DCNOT.

Despite these gates being implementable in unit time, we believe that in general, this is not possible. Certainly the trick that we used above to implement the C-NOT in unit time does not work for general C-U gates (ie. a unitary upon Bob's qubit which is performed if and only if Alice's qubit is $|\downarrow\rangle$. Gates which can be performed in unit time appear to be special, and are in this sense easier to perform using local actions, classical communication and entanglement. In this sense they are less non-local than gates which require two units of time. We believe that the time required to perform an operation is a resource which will play an important role in the study of non-local operations.

7.8 Multi-partite Operations

In the previous sections we studied different bi-partite operations. What about multi-partite operations, such as the Toffoli or the Fredkin gates on three qubits? As we showed in section II, they can all be implemented by using the "double teleportation" method. On the other hand, finding the *necessary* resources is far more difficult than in the bi-partite case; indeed it is not possible at present. The reason is that there exist different inequivalent types of multi-partite entanglement [23, 20]. For example, it is known that singlets and GHZ states are inequivalent in the sense that they cannot be reversibly transformed into each other, not even in the asymptotic limit. Although GHZs (as all other entangled states) can be built out of singlets, such a procedure is wasteful. Hence, when investigating the minimal entanglement resources needed to implement multi-partite quantum operations, we have to use the different inequivalent types of entanglement. Unfortunately, at present multi-partite entanglement is far from being fully understood. Some interesting results on multi-partite operations, for example the Toffoli gate, were presented in [53] and [52].

7.9 "Conservation" Relations

In studying the non-locality of quantum states a most important issue is that of "manipulating" entanglement, i.e. of transforming some states into others [21]. Similarly we can ask: Given a unitary evolution, can we use it to implement some other unitary evolution?

In particular, for pure quantum states we have *conservation* relations [21, 119]. For example, when Alice and Bob share a large number n of pairs of particles, each pair in the same state Ψ, they could use these pairs to generate some other number, k, of pairs in some other state Φ. In the limit of large n, this transformation can be performed *reversibly*, meaning that the total amount of non-locality contained in the n copies of the state Ψ is the same as the total amount of non-locality contained in the k copies of the state Φ. Is something similar taking place for

unitary transformations?

For unitary transformations we did not study the case of the asymptotic limit, i.e. performing the same transformation U on many pairs of particles[1]. However, an interesting pattern emerges even at the level of a single copy.

Consider first the case of SWAP. We know what the minimal resources needed to implement a SWAP are. But suppose now that we are given a device which implements a SWAP. Could we could use it to get back the original resources needed to create the SWAP?

The balance of resources needed to implement a SWAP can be written as

$$2\text{e-bits} + 2\text{bits}_{A\to B} + 2\text{bits}_{B\to A} => \text{SWAP}. \tag{7.9.1}$$

The question is whether

$$\text{SWAP} => 2\text{e-bits} + 2\text{bits}_{A\to B} + 2\text{bits}_{B\to A}? \tag{7.9.2}$$

The answer is "no". That is, combining entanglement and classical communication resources to yield a SWAP is an *irreversible process* - we cannot use the SWAP to get the resources back. We shall prove that, if one uses the SWAP to make 2 e-bits, then one cannot use it to send any classical communication from Alice to Bob. The key element in our proof is that, if Alice sends a qubit to Bob and uses it to create an e-bit, she cannot use it to simultaneously send any classical bits. This we prove by contradiction: suppose we could do

$$1\text{qubit}_{A\to B} + x\text{e-bits} + y\text{bits}_{A\to B} \mapsto (x+1)\text{e-bits} + (y+z)\text{bits}_{A\to B}, \tag{7.9.3}$$

where the e-bits and bits on the left hand side are because we allow catalysis. Strictly speaking, we should add some bits from Bob to Alice to the left hand side of this equation. If we did this, they would just carry through all the equations and emerge at the end unchanged: they do not affect any of the arguments put here. Now, we

[1] since our work was published, some work has been performed upon the asymptotic limit [120, 121], though the problem is not as yet solved in general. See my review at the end of this chapter for more details.

could add $x+1$ qubits to both sides and use superdense coding to perform:

$$\begin{aligned}(1+(x+1))\text{qubits}_{A\to B} + x\text{e-bits} + y\text{bits}_{A\to B} \\ \mapsto \quad (x+1)\text{qubits}_{A\to B} + (x+1)\text{e-bits} + (y+z)\text{bits}_{A\to B} \\ \mapsto \quad (2(x+1)+(y+z))\text{bits}_{A\to B}.\end{aligned} \tag{7.9.4}$$

Clearly, we could have produced the original resources for this procedure by

$$\begin{aligned}((1+(x+1))+x+y)\text{qubits}_{A\to B} \\ \mapsto \quad (1+(x+1))\text{qubits}_{A\to B} + x\text{e-bits} + y\text{bits}_{A\to B}.\end{aligned} \tag{7.9.5}$$

Since we cannot use the transmission of one qubit to send more than one classical bit, we have that

$$2(x+1)+(y+z) \leq (1+(x+1))+x+y \tag{7.9.6}$$

which simplifies to

$$z \leq 0. \tag{7.9.7}$$

So, we have that if we use a qubit to create an e-bit, we cannot use it to send any classical bits.

Now, to show that we cannot use a SWAP to make 2 e-bits and some classical bits, note that we can perform a SWAP by sending a qubit in each direction. Thus if we could could use a SWAP to make two e-bits and send any classical communication, then we could use two qubits, one in each direction, to make two e-bits and send some classical communication, something we just proved was impossible. So we cannot recover all the resources required to implement a SWAP after it has been built.

Despite this irreversibility, looking back to the proof of the resources needed for SWAP, we see that we can write the following tight "implications":

$$2\text{e-bits} + 2\text{bits}_{A\to B} + 2\text{bits}_{B\to A} => 1\text{SWAP}. \tag{7.9.8}$$

$$2\text{e-bits} + 1\text{SWAP} => 2\text{bits}_{A\to B} + 2\text{bits}_{B\to A}. \tag{7.9.9}$$

$$1\text{SWAP} => 2\text{e-bits}. \tag{7.9.10}$$

The first of these three implications is to be read as "given 2e-bits and $2\text{bits}_{A\to B}$ and $2\text{bits}_{A\to B}$ we can produce the SWAP operation; also if we wish to produce the SWAP operation with e-bits, and bits communicated from Alice to Bob and vice-versa, we cannot do so with fewer than 2e-bits and $2\text{bits}_{A\to B}$ and $2\text{bits}_{A\to B}$."

The second and third implications have a slightly different meaning. For example we read the second implication as "given 1 SWAP and 2 e-bits, we can communicate 4 classical bits (two each way); also we cannot communicate more than 4 classical bits (two each way) ". On the other hand, it does not mean that "1 SWAP and 2 e-bits are necessary for communicating 4 classical bits (two each way) " - for example we can implement this classical communication with 2 SWAPs.

Exactly the same implications apply for the DCNOT.

$$2\text{e-bits} + 2\text{bits}_{A\to B} + 2\text{bits}_{B\to A} => 1\text{DCNOT}. \tag{7.9.11}$$

$$2\text{e-bits} + 1\text{DCNOT} => 2\text{bits}_{A\to B} + 2\text{bits}_{B\to A}. \tag{7.9.12}$$

$$1\text{DCNOT} => 2\text{e-bits}. \tag{7.9.13}$$

Furthermore, very similar implications can be written for the CNOT:

$$1\text{e-bit} + 1\text{bit}_{A\to B} + 1\text{bit}_{B\to A} => 1\text{CNOT}. \tag{7.9.14}$$

$$1\text{e-bit} + 1\text{CNOT} => 1\text{bit}_{A\to B} + 1\text{bit}_{B\to A}. \tag{7.9.15}$$

$$1\text{CNOT} => 1\text{e-bit}. \tag{7.9.16}$$

In fact these implications are very similar to the implications which describe teleportation and super-dense coding which appear, together with many other similar implications on Bennett's famous transparency presented at almost all early quantum information conferences:

$$\text{1e-bit} + \text{2bits}_{A\to B} \Rightarrow \text{1qubit} \tag{7.9.17}$$

$$\text{1e-bit} + \text{1qubit} \Rightarrow \text{2bits}_{A\to B} \tag{7.9.18}$$

$$\text{1qubit} \Rightarrow \text{1e-bit} \tag{7.9.19}$$

The above three implications (7.9.17,7.9.18,7.9.19) are generally thought to describe relations between classical information, quantum information and entanglement. However, we would like to argue that their true meaning may be more closely related to dynamics, and that a more illuminating form is probably

$$\text{1e-bit} + \text{2bits}_{A\to B} \Rightarrow \text{1teleportation}_{A\to B} \tag{7.9.20}$$

$$\text{1e-bit} + \text{1teleportation}_{A\to B} \Rightarrow \text{2bits}_{A\to B} \tag{7.9.21}$$

$$\text{1teleportation}_{A\to B} \Rightarrow \text{1e-bit} \tag{7.9.22}$$

We conjecture that similar relations hold between any quantum action and the resources needed to implement it, that is

$$Entanglement + ClassicalCommunication \Rightarrow Action \tag{7.9.23}$$

$$Entanglement + Action \Rightarrow ClassicalCommunication \tag{7.9.24}$$

$$Action \Rightarrow Entanglement \tag{7.9.25}$$

It may be that these relations hold, in general, only in the asymptotic limit of many copies of the quantum action.

7.10 Different Ways of Achieving the Same Task

It is interesting to note that although the transformation from resources to unitary actions is irreversible, sometimes the same end product can be achieved in two different ways. For example, there are two alternative ways to implement

$$2\text{CNOTs} => 1\text{bit}_{A\to B} + 1\text{bit}_{B\to A}. \tag{7.10.1}$$

The first way is to use one CNOT to transmit 1 classical bit from Alice to Bob and the other CNOT to transmit 1 classical bit from Bob to Alice, i.e.

$$1\text{CNOT} => 1\text{bit}_{A\to B} \tag{7.10.2}$$

and

$$1\text{CNOT} => 1\text{bit}_{B\to A}. \tag{7.10.3}$$

Another possibility is to use first one CNOT to create 1 e-bit (7.9.16) then the other CNOT plus the e-bit to transmit the 2 classical bits (7.9.15), i.e.

$$2\text{CNOTs} => 1\text{e-bit} + 1\text{CNOT} => 1\text{bit}_{A\to B} + 1\text{bit}_{B\to A}. \tag{7.10.4}$$

7.11 Catalysing Classical Communication

A very interesting phenomenon is that of "catalysing" classical communication. This phenomenon is similar in its spirit to that of "catalysing entanglement manipulation" [61, 23]. An example is the following.

On its own, the SWAP can only send one bit in each direction at the same time, and cannot be used for Alice to send 2 bits to Bob, even if Bob sends no information whatsoever. That is,

$$1\text{SWAP} \not=> 2\text{bits}_{A\to B}. \tag{7.11.1}$$

However, if Alice and Bob share 1 e-bit, Alice can send 2 bits to Bob *without destroying* the e-bit, i.e.

$$1\text{SWAP} + 1\text{e-bit} => 2\text{bits}_{A\to B} + 1\text{e-bit}. \tag{7.11.2}$$

This may be done as follows. Initially Alice and Bob share a non-local singlet; Bob also prepares a second singlet locally. Alice encodes the two bits she wishes to send to Bob by performing one of the four rotations 1, σ_x, σ_y, σ_z on her half of the non-local singlet. By performing the SWAP operation on Alice's particle from the non-local singlet and one particle of the singlet that Bob has prepared locally, Alice and Bob end up with a non-local singlet held between them; also Bob can find out the two bits by measurements on the local singlet he now holds. Specifically, we begin with the state:

$$(|\uparrow\rangle_A |\uparrow\rangle_{b1} + |\downarrow\rangle_A |\downarrow\rangle_{b1})(|\uparrow\rangle_B |\uparrow\rangle_{b2} + |\downarrow\rangle_B |\downarrow\rangle_{b2}), \tag{7.11.3}$$

where A is Alice's particle, and B, $b1$ and $b2$ are Bob's particles. Alice performs one of the rotations 1, σ_x, σ_y, σ_z on her particle. They then perform the SWAP on particles A and B, and get (if Alice performed 1):

$$(|\uparrow\rangle_B |\uparrow\rangle_{b1} + |\downarrow\rangle_B |\downarrow\rangle_{b1})(|\uparrow\rangle_A |\uparrow\rangle_{b2} + |\downarrow\rangle_A |\downarrow\rangle_{b2}) \tag{7.11.4}$$

If Alice performed one of the other rotations, Bob will get one of the other Bell states in system (B, $b1$). Bob now measures that system in the Bell basis to extract the information, and Alice and Bob are left with a singlet between systems A and $b2$.

In effect the SWAP acts as a double teleportation; one from Alice to Bob and one from Bob to Alice. Teleporting Alice's qubit, in conjunction with the e-bit, implements a transmission of two bits from Alice to Bob using super-dense coding; it destroys the e-bit in the process. Simultaneously, the Bob to Alice teleportation restores the e-bit.

7.12 Trading One Type of Action For Another

An interesting question is the following. There are cases in which two different actions require the same resources. For example the resources needed for 1 SWAP are the same as for 2 CNOTs, i.e., 2e-bits + $2\text{bits}_{A\to B}$ + $2\text{bits}_{B\to A}$. Now, suppose we had already used the resources to build 2 CNOTs, but we wanted to change our mind and we wanted to do 1 SWAP instead. Due to the irreversibility discussed above, we cannot simply get back the original resources and use them to construct the SWAP. Is it however possible to go *directly* from 2 CNOTs to 1 SWAP, without going back to the original resources? As far as we are aware, the answer is "No".

It turns out however that if we have many CNOTs it is nevertheless useful to build a SWAP from CNOTs directly rather than going back to the original resources. Indeed, to obtain the entanglement and classical communication resources needed for 1 SWAP, i.e. 2e-bits + $2\text{bits}_{A\to B}$ + $2\text{bits}_{B\to A}$ we need 4 CNOTs. However, it is well-known that one can construct 1 SWAP directly from 3 CNOTs. Indeed, we don't even need 3 CNOTs, but can realize a SWAP by

$$2\text{CNOTs} + 1\text{bit}_{A\to B} + 1\text{bit}_{B\to A} => 1\text{SWAP} \tag{7.12.1}$$

which uses less non-local resources than 3 CNOTs. To see this, it suffices to note that

$$1\text{CNOT} + 1\text{bit}_{A\to B} => 1\text{teleportation}_{A\to B} \tag{7.12.2}$$

and similarly

$$1\text{CNOT} + 1\text{bit}_{B\to A} => 1\text{teleportation}_{B\to A} \tag{7.12.3}$$

To implement (7.12.2) Alice starts with her qubit in the state $\Psi_A = \alpha \left|\uparrow\right\rangle_A + \beta \left|\downarrow\right\rangle_A$ which has to be teleported and Bob with his qubit in the state $\left|\uparrow\right\rangle_B$. After CNOT the state becomes:

$$\Psi_A \left|\uparrow\right\rangle_B = (\alpha \left|\uparrow\right\rangle_A + \beta \left|\downarrow\right\rangle_A) \left|\uparrow\right\rangle_B \mapsto \alpha \left|\uparrow\right\rangle_A \left|\uparrow\right\rangle_B + \beta \left|\downarrow\right\rangle_A \left|\downarrow\right\rangle_B \tag{7.12.4}$$

Alice then measures her qubit in the $|+>=\frac{1}{\sqrt{2}}(|\uparrow\rangle+|\downarrow\rangle)$ and $|->=\frac{1}{\sqrt{2}}(|\uparrow\rangle-|\downarrow\rangle)$ basis and communicates the result to Bob. If $(+)$ then Bob's qubit is already in the required state $\Psi_B = \alpha|\uparrow\rangle_B + \beta|\downarrow\rangle_B$; if $(-)$ then Bob's qubit is in the state $\Psi'_B = \alpha|\uparrow\rangle_B - \beta|\downarrow\rangle_B$ and Bob can obtain Ψ by changing the relative phase between $|\uparrow\rangle_B$ and $|\downarrow\rangle_B$ by π.

One can also ask how actions can be traded for one another at the many copy level. For example, we earlier showed that a single instance of the SWAP cannot be used to implement a DCNOT. However, it turns out that 2 copies of a SWAP gate and local operations can be used to perform 2 copies of any bi-partite gate upon qubits, provided that the 2 SWAP gates are allowed to be performed at different times. Thus one can use two copies of the SWAP to make a SWAP and a DCNOT. This works as follows. Firstly, 1 copy of the arbitrary operation can be performed using the "double teleportation" method. Alice uses the first SWAP to send her qubit to Bob, he then performs the required operation locally, and Bob then uses the second SWAP to return Alice's qubit. They perform the second copy of the arbitrary operation in a similar way: Bob uses the first copy of the SWAP to send his qubit to Alice, she performs the operation locally, and returns Bob's qubit using the second SWAP. Thus 2 copies of the SWAP can be used to make a SWAP and a DCNOT. If we view the SWAP as a catalyst, we have shown that a SWAP can be turned into a DCNOT! However, we do not know whether this is reversible. Despite both operations being maximally non-local in terms of the amount of entanglement and classical communication needed to implement them, we believe that this is not possible.

7.13 Subsequent Research

In the time between writing our paper [51], and coming to write this thesis, quantum operations have become an area of intense study, which is now a key part of quantum information. I shall use the remainder of this chapter to survey this work, and work which is closely related. As one might expect, opening the study of quantum non-local dynamics has led to many questions, and many different points of view. I shall

begin with those most closely related to our work.

7.14 Non-Integer Resources

All the operations considered in our work could be implemented in terms of integer resources. It was soon shown how to go beyond this, with a method of implementing a certain class of unitaries using less than one e-bit, and less than two classical bits in each direction [122]. The method takes a probabilistic number of steps, but always succeeds, and on average uses less than one e-bit, though in some cases uses more. It is not known whether the procedure is optimal.

The method is based upon a mathematical isomorphism between states and operations [123] which is of great use in importing results from the study of non-local states to that of operations, as well as directly in the study of quantum operations [124, 125, 126]. Since it is so useful, I shall describe it here, acting upon a single qubit (though it works for any dimension of hilbert space, and any number of (possibly spacially separated) qubits). Given any operation acting on the qubit A, make a state out of it by first maximally entangling the qubit with an ancillary qubit, a, and then applying the unitary on A. Thus

$$U_A \mapsto U_A \frac{1}{\sqrt{2}}(|00\rangle_{Aa} + |11\rangle_{Aa}). \tag{7.14.1}$$

Given a state, normally of two qubits, A and a, form an operation (on the state of a third qubit A', initially in state $|\psi\rangle_{A'}$) by measuring the Bell operator on the qubits A' and a, and postselecting the result where the outcome is $\frac{1}{\sqrt{2}}(|00\rangle_{A'a} + |11\rangle_{A'a})$. This sounds a strange idea, until we apply it to a state formed from an operation as in equation (7.14.1). Then we find qubit A is finally left in the state $U_A|\psi\rangle_A$, and so we have applied the initial operation to the state of qubit A'. This only works with probability $\frac{1}{4}$, but is still mathematically an isomorphism. Essentially we teleport the state of the qubit A' through the operation stored in the qubit A. Qubit a is a reference qubit for the storing of the operation in qubit A.

The isomorphism works for entangled states, and non-unitary operations, and for systems with many qubits. We just have to add a reference qubit for every

qubit in the system of interest. Using this isomorphism, it is simple to use a small amount of entanglement to implement weakly entangling bi-partite unitaries with a finite probability of success, and to know when we have succeeded. One would much rather implement a gate with certainty: this was also demonstrated [122]. This is an important step towards finding the necessary and sufficient resources required to implement a unitary.

7.15 Generating Entanglement From a Unitary

A great deal of subsequent work has been performed trying to find the maximum amount of entanglement one can produce from a non-local unitary operation. An important simplification for qubits came with the discovery that, using local unitaries before and after the non-local unitary, any non-local unitary is equivalent (up to an overall phase) to [127, 128]

$$U(\alpha_x, \alpha_y, \alpha_z) = exp\left(i \sum_{i \in \{x,y,z\}} \alpha_i \sigma_i^A \otimes \sigma_i^B\right). \tag{7.15.1}$$

Here α_x, α_y and α_z are the pauli matrices. This form has allowed great progress to be made in understanding two qubit unitaries. The lack of such a simple form is holding back progress for higher dimensional unitaries.

For single copies of qubit bi-partite unitaries, the maximal entanglement which one can produce has been found in some restricted cases. First, for infinitesimal unitaries, with ancillas and initial entanglement allowed [129]. Also, for finite unitaries with no initial entanglement and no ancillas [127]. Later, for finite unitaries with arbitrary initial entanglement but no ancillas [120]. Ancillas were discussed to some extent in [127, 120].

An important realisation is the fact that many copies of a unitary are no better at creating entanglement per copy asymptotically than a single copy, so long as we are allowed prior shared entanglement, and to perform joint local operations on the states, and ancillas [120, 121]. The reason for this is that two non-local operations performed at the same time could be performed one just after the other, and one

performed after the other is simply two independent unitaries, in the presence of shared entanglement and ancillas. Thus, unlike for states, allowing many copies of the unitary does not simplify our study.

7.16 Generating Classical Communication

In parallel with knowing how much entanglement a unitary operation can produce, we would also like to know how much classical communication it can generate. This has been studied in [121]. It has been suggested that for two qubit unitary operations, the entanglement and classical communication capacities may be the same [130, 131]. Whilst there is no proof, there are many promising lines of attack, and many partial results to support this claim. If it were true, it would be a great simplification, and one would hope to get a greater understanding of the relation between generating entanglement and classical communication for unitary operations.

The study of generating classical communication and entanglement from a bi-partite operation is related to that of generating the same from a quantum channel. Normally in a quantum channel, we assume that the system is being carried from Alice to Bob, and not in the other direction. However there is no reason not to consider a bi-directional channel, and such a channel would certainly be a bi-partite operation. One hopes that progress can be made by considering the two questions together.

7.17 Directly Interconverting Operations

A fundamental issue is the use of one, or many, non-local unitary operations to simulate one another. One of the main results in quantum information is the inter-convertibility under local operations and classical communication of pure, bi-partite, quantum states. I hope that some similar result may be found for quantum operations. Whilst it is interesting to do so, there is no fundamental reason to turn operations into entangled states and back again. One could try to convert opera-

tions one into the other directly. Indeed, this may well give a more fundamental view of the non-locality of an operation. It would be interesting to know the optimal ways to turn one operation into another directly.

In terms of non-infinitesimal non-local unitary operations, this has not yet been much studied. However, an interesting catalysis effect has already been discovered [132]. This is a scenario where a single unitary cannot be used to simulate another one using local operations and classical communication. However, if a maximally entangled catalyst is allowed, the simulation can be performed. This effect gives hope that, similar to bi-partite pure state entanglement, there may be only one form of bi-partite unitary non-local operation, whose content in each unitary operation we could quantify with a single number. An attempt to axiomatize this idea of measuring the interaction strength was made in [133].

Another work on interconverting operations found a method for converting several copies of any bi-partite unitary into a CNOT which is near optimal in the scenario without any ancillas or prior entanglement [134].

A key work looked at simulating one non-local hamiltonian with another, and found the optimal method for two qubit hamiltonians [135]. The results are somewhat similar to the majorization results found for interconverting non-local pure bi-partite states. The question of converting hamiltonians into unitary operations in the optimal way has been discussed [128, 136, 137]. Producing interactions of a desired form (perhaps for quantum computation) from a given form (perhaps some experimental implementation) is an area which has been much studied, (see [135] and references within).

7.18 Quantum Remote Control

Another related branch of quantum information is that of implementing operations at a distance [138]. Suppose that Alice wants to perform one of many possible unitary operations on Bob's qubit (with his help), but only Alice knows which one. Alice could just tell Bob which to perform, but this will take a very large amount of classical communication. If we use some entanglement, we can implement the

operation much more efficiently. The main question is, what is optimal? One could implement an arbitrary unitary by teleporting Bob's qubit to Alice, having Alice perform the operation, and then teleporting the qubit back to Bob. It was shown that, if the operation Alice wishes to perform could be any qubit unitary, this is optimal [138], in terms of the entanglement and classical communication resources required.

Subsequently it has been shown that if Alice and Bob know that the operation to be performed comes from a restricted class of operations, for example any rotation around the z-axis, then more efficient procedures exist [139]. In that particular case, only one e-bit and one bit in each direction are needed. Alternatively, suppose the operation is any unitary, but Alice knows only the angle of the rotation and Bob knows only the axis about which the operation is to be performed [140]. In this case one again only needs one e-bit and one bit in each direction. [140] also introduced a hybrid state-operator (stator) object which has proven useful in the study of non-local quantum operations.

Instead of performing a remote unitary operation, one may wish to perform a remote measurement, possibly a POVM. Alice knows which measurement is to be performed, and Bob holds the system to be measured. One possibility is to teleport Bob's system to Alice, using one e-bit. However this has been shown [141] to be inefficient: one can implement POVM's remotely with certainty using a non-maximally entangled state. This is one of very few non-local operations known which can be performed with certainty using a non-maximally entangled state (see also [122]).

7.19 Instantaneous Non-Local Measurements

A final somewhat related area is the question of the instantaneous measurement of non-local observables[142, 143, 144, 145, 146]. Suppose we wish to make a measurement of an observable with entangled eigenstates. Such a measurement cannot in general, with no additional resources, be made with just local operations and classical communication. If we were allowed plenty of time, we could teleport Al-

ice's particle to Bob and perform the measurement locally. Suppose, however, we have to perform all the quantum interactions instantaneously, and are subsequently only allowed to communicate classically. Which measurements can be performed? We allow unlimited prior shared entanglement, and allow the measurement to destroy the state, thus relaxing the usual condition of repeatability for Von-Neumann measurements.

The question has important consequences for our understanding of relativistic quantum mechanics. If there are variables which cannot be measured instantaneously, in what sense can we say that a system has a particular value of that variable at a particular instant in time? The interpretation would be somewhat problematic. Fortunately, it has recently been shown that all observables can be measured instantaneously [147, 148]. Though the questions are very old, the recent results were inspired by our work on non-local operations.

7.20 Conclusion

Quantum operations have a non-local content, similar to that of quantum states. We have laid out a basic framework for discussing this content, in terms of the necessary and sufficient amounts of entanglement and classical communication needed to implement an operation. In this sense the SWAP and the DCNOT gates are maximally non-local, whereas the CNOT has a lesser amount of non-locality. We have discussed these and several other features of this non-locality. With this framework, one can try to make analogies with many features of the non-locality of quantum states. For example one hopes to find an analog of the quantification of entanglement, arising via reversible asymptotic manipulations of quantum operations. It is now clear that the non-local content of quantum operations is of as fundamental interest to quantum information as the non-local content of quantum states. Our work is a new direction in quantum information, one towards a greater understanding of quantum dynamics.

Chapter 8

Conclusion

8.1 Summary

I have discussed various perspectives on quantum non-locality. First I looked at the well known non-local features of quantum mechanical states, and showed that many of these features had close classical analogs. Since the classical world is in general easier to understand than the quantum one, this gave us greater insight into those quantum features which had good classical analogs, and pinpointed the features which had no classical analog to be those which are most representative of quantum mechanics.

My second perspective looked at one of those features with no classical analog, the non-local correlations between measurements in different regions of space. I investigated various limitations on the correlations which classical models can produce, focusing on those limitations which quantum mechanics does not respect. I looked at measurements with many outcomes, systems with many parties, systems with noise, and systems with memory. This gave a better understanding of how quantum mechanics differs from classical mechanics.

Finally I introduced a new aspect of quantum non-locality, namely that of joint operations on several parties. Non-local operations are viewed as a resource, which can be created from entanglement and classical communication, converted from one form to another, and eventually used to perform a desired action. This led to the

discovery of several quantum mechanical processes involving quantum operations, and so a direct understanding of some of the possibilities which quantum non-locality allows.

8.2 Looking Forward

Since my thesis has described new perspectives on the study of quantum non-locality, there is much room for further development. It is hoped that the quantum-classical analogy which so far was applied only to existing quantum mechanical state manipulations will allow us to find new and improved procedures. The issue of which states and experiments exhibit non-local correlations is still unresolved, and the approach suggested here seems certain to yield further progress. The non-locality of quantum operations is as yet barely explored: one hope is that it will have a simple quantification, similar to entangled quantum states, allowing us to say that a bi-partite unitary operation has "x units of interaction".

Despite this progress, I have left untouched the deepest issues concerning quantum non-locality, in particular the implications it has for our world view. To date no-one has proposed a satisfactory physical mechanism for non-locality, nor an explanation for why it is that the world around us appears classical although the building blocks of our world, elementary particles, are governed by quantum mechanics. Fortunately quantum mechanics is a set of simple rules which tell us how to make accurate predictions about many experimental situations, in particular those which involve long distance quantum correlations. In this thesis I have developed a greater understanding of the consequences of these rules, and hence of the fundamentally non-classical phenomenon of quantum non-locality.

Bibliography

[1] *Introduction to Quantum Computation and Information*, edited by T. Spiller H.-K. Lo and S. Popescu (World Scientific, Singapore, 1998).

[2] C. H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).

[3] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002), a review.

[4] B. Brezger, L. Hackermüller, S. Uttenthaler, J. Petschinka, M. Arndt, and A. Zeilinger, Phys. Rev. Lett. **88**, 100404 (2002).

[5] B. Julsgaard, A. Kozhekin, and E. S. Polzik, Nature **413**, 400 (2001).

[6] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).

[7] D. Bohm and Y. Aharonov, Phys. Rev. **108**, 1070 (1957).

[8] J. S. Bell, Physics **1**, 195 (1964).

[9] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).

[10] J. F. Clauser and M. A. Horne, Phys. Rev. D **10**, 526 (1974).

[11] A. Aspect, Nature **398**, 189 (1999), a review.

[12] D. M. Greenberger, M. Horne, and A. Zeilinger, in *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, edited by M. Kafatos (Kluwer Academic, Dordrecht, 1989), pp. 69–72.

[13] L. Hardy, Phys. Rev. Lett. **68**, 2981 (1992).

[14] N. Gisin, Phys. Lett. A **154**, 201 (1991).

[15] N. Gisin and A. Peres, Phys. Lett. A **162**, 15 (1992).

[16] S. Popescu and D. Rohrlich, Phys. Lett. A **166**, 293 (1992).

[17] C. H. Bennett and S. J. Wiesner, Phys. Rev. Lett. **69**, 2881 (1992).

[18] A. Pati, Phys. Rev. A **63**, 014320 (2001).

[19] G. Brassard, arxiv.org e-print archive (2001), quant-ph/0101005.

[20] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).

[21] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Phys. Rev. A **53**, 2046 (1996).

[22] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, Phys. Rev. Lett. **76**, 722 (1996).

[23] C. H. Bennett, S. Popescu, D. Rohrlich, J. A. Smolin, and A. V. Thapliyal, Phys. Rev. A **63**, 012307 (2001).

[24] N. Linden, S. Popescu, B. Schumacher, and M. Westmoreland, (1999), quant-ph/9912039.

[25] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, Phys. Rev. Lett. **81**, 5039 (1998).

[26] D. Bouwmeester, J-W. Pan, M. Daniell, H. Weinfurter, and A. Zeilinger, Phys. Rev. Lett. **82**, 1345 (1999).

[27] D. Boschi, S. Branca, F. De Martini, L. Hardy, and S. Popescu, Phys. Rev. Lett. **80**, 1121 (1998).

[28] D. Bouwmeester, J-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger, Nature **390**, 575 (1997).

[29] *Experimental Quantum Computation and Information*, edited by F. De Martini and C. Monr (IOS Press, Nieuwe Hemweg 6B, 1013 BG Amsterdam, The Netherlands, 2002), a review.

[30] H. Zbinden, J. Brendel, N. Gisin, and W. Tittel, Phys. Rev. A **63**, 022111 (2001).

[31] G. S. Vernam, Journal of the American Institute for Electrical Engineers **55**, 109 (1926).

[32] H.-K. Lo and S. Popescu, Phys. Rev. A **63**, 022301 (2001).

[33] M. A. Nielsen, Phys. Rev. Lett. **83**, 436 (1999).

[34] D. Collins and S. Popescu, Phys. Rev. A **65**, 032321 (2002).

[35] H. Everett, Reviews of Modern Physics **29**, 454 (1957).

[36] D. Deutsch and P. Hayden, Proc. Royal. Soc. Lond. A **456**, 1759 (2000).

[37] J. Bub, in *Interpreting the Quantum World* (Cambridge University Press, Cambridge, 1997), Chap. 8.

[38] N. D. Mermin, Phys. Rev. D **22**, 356 (1980).

[39] A. Garg and N. D. Mermin, Phys. Rev. Lett. **49**, 901 (1982).

[40] A. Garg and N. D. Mermin, Phys. Rev. D **27**, 339 (1983).

[41] D. Kaszlikowski, P. Gnacinski, M. Zukowski, W. Miklaszewski, and A. Zeilinger, Phys. Rev. Lett. **85**, 4418 (2000).

[42] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, Phys. Rev. Lett. **88**, 040404 (2002).

[43] T. Durt, D. Kaszlikowski, and M. Zukowski, Phys. Rev. A **64**, 042101 (2001).

[44] S. Popescu, Phys. Rev. Lett. **74**, 2619 (1995).

[45] A. J. Leggett and A. Garg, Phys. Rev. Lett. **54**, 857 (1985).

[46] D. Collins and S. Popescu, J. Phys. A: Math. Gen. **34**, 6831 (2001).

[47] G. Svetlichny, Phys. Rev. D **35**, 3066 (1987).

[48] D. Collins, N. Gisin, S. Popescu, D. Roberts, and V. Scarani, Phys. Rev. Lett. **88**, 170405 (2002).

[49] P. H. Eberhard, Phys. Rev. A **47**, R747 (1993).

[50] J. Barrett, D. Collins, L. Hardy, A. Kent, and S. Popescu, arxiv.org e-print archive (2002), quant-ph/0205016.

[51] D. Collins, N. Linden, and S. Popescu, Phys. Rev. A **64**, 032302 (2001).

[52] A. Chefles, C. R. Gilson, and S. M. Barnett, arxiv.org e-print archive (2000), quant-ph/0003062.

[53] J. Eisert, K. Jacobs, P. Papadopolous, and M. B. Plenio, Phys. Rev. A **62**, 052317 (2000).

[54] N. Gisin and S. Wolf, Crypto 2000 482 .

[55] H.-K. Lo and S. Popescu, Phys. Rev. Lett. **83**, 1459 (1999).

[56] E. Schmidt, Math. Ann. **63**, 433 (1907).

[57] C. H. Bennett, private communication (unpublished).

[58] M. Koniorczyk, T. Kiss, and J. Janszky, J. Phys A (Math. Gen.) **34**, 6949 (2001).

[59] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).

[60] G. Vidal, Phys. Rev. Lett. **83**, 1046 (1999).

[61] D. Jonathan and M. B. Plenio, Phys. Rev. Lett. **83**, 3566 (1999).

[62] C.H. Bennett, C. Crepeau, and U. M. Maurer, IEEE Trans. Inform. Theory **41**, 1915 (1995), and references within.

[63] U. Maurer, IEEE Trans. Inform. Theory **39**, 733 (1993).

[64] R. Ahlswede and I. Csiszar, IEEE Trans. Inform. Theory **39**, 1121 (1993).

[65] E. F. Galv ao, M. B. Plenio, and S. Virmani, J. Phys. A **33**, 8809 (2000).

[66] I. Percival, Phys. Lett. A **244**, 495 (1998).

[67] D. Kaszlikowski, L. C. Kwek, J.-L. Chen, M. Zukowski, and C. H. Oh, Phys. Rev. A **65**, 032118 (2002).

[68] J.-L. Chen, D. Kaszlikowski, L. C. Kwek, C. H. Oh, and M. Zukowski, Phys. Rev. A **64**, 052109 (2001).

[69] S. Massar, S. Pironio, J. Roland, and B. Gisin, arxiv.org e-print archive (2002), quant-ph/0205130.

[70] K. Zyczkowski, P. Horodecki, A. Sanpera, and M. Lewenstein, Phys. Rev. A **58**, 883 (1998).

[71] G. Vidal and R. Tarrach, Phys. Rev. A **59**, 141 (1999).

[72] S.L. Braunstein, C.M. Caves, R. Jozsa, N. Linden, S. Popescu, and R. Schack, Phys. Rev. Lett. **83**, 1054 (1999).

[73] R. F. Werner, Phys. Rev. A **40**, 4277 (1989).

[74] S. Popescu, in *The Dilemma of Einstein, Podolsky and Rosen - 60 years after.*, edited by A. Mann and M. Revzen (IOP Publishing, Bristol, 1996).

[75] M. Zukowski, R. Horodecki, M. Horodecki, and P. Horodecki, Phys. Rev. A **58**, 1694 (1998).

[76] S. Teufel, K. Berndl, D. Dürr, S. Goldstein, and N. Zanghi, Phys. Rev. A **56**, 1217 (1997).

[77] C. H. Bennett, D. P. DiVincenzo, J. Smolin, and W. K. Wootters, **54**, 3824 (1996).

[78] G. Vidal and J. I. Cirac, Phys. Rev. Lett. **86**, 5803 (2001).

[79] N. D. Mermin, Phys. Rev. Lett. **65**, 1838 (1990).

[80] A.V. Belinskii and D.N. Klyshko, Phys. Ups. **36**, 653 (1993).

[81] N. Gisin and H. Bechmann-Pasquinucci, Phys. Lett. A **246**, 1 (1998).

[82] R. F. Werner and M. M. Wolf, Phys. Rev. A **61**, 062102 (2000).

[83] B. S. Cirel'son, Lett. Math. Phys. **4**, 83 (1980).

[84] V. Scarani and N. Gisin, J. Phys. A: Math. Gen. **34**, 6043 (2001).

[85] P. Mitchel, S. Popescu, and D. Roberts, arxiv.org e-print archive (2002), quant-ph/0202009.

[86] W. Dür, G. Vidal, and J. I. Cirac, Phys. Rev. A **62**, 062314 (2000).

[87] R. Gill, arxiv.org e-print archive (2001), quant-ph/0110137.

[88] L. Accardi and M. Regoli, arxiv.org e-print archive (2000), quant-ph/0007005, quant-ph/0007019, quant-ph/0110086.

[89] D. Mayers and A. Yao, in *Proceedings of the 39th Annual Symposium on the Foundations of Computer Science* (Computer Society Press, Los Alamitos, 1998), p. 503, also available as arxiv.org e-print archive quant-ph/9809039.

[90] N. Gisin, private communication (unpublished).

[91] W. H. Furry, Phys. Rev. **49**, 393, 476 (1936).

[92] P. R. Tapster, J. R. Rarity, and P. C. M. Owens, Phys. Rev. Lett. **73**, 1923 (1994).

[93] W. Tittel, J. Brendel, H. Zbinden, and N. Gisin, **81**, 3563 (1998).

[94] T. K. Lo and A. Shimony, Phys. Rev. A **23**, 3003 (1981).

[95] P. G. Kwiat, P. H. Eberhard, A. M. Steinberg, and R. Y. Chiao, Phys. Rev. A **49**, 3209 (1994).

[96] S. F. Huelga, M. Ferrero, and E. Santos, **51**, 5008 (1995).

[97] E. S. Fry, T. Walther, and S. Li, Phys. Rev. A **52**, 4381 (1995).

[98] M. Freyberger, P. K. Aravind, M. A. Horne, and A. Shimony, Phys. Rev. A **53**, 1232 (1996).

[99] C. Brif and A. Mann, Europhys. Lett. **49**, 1 (2000).

[100] A. Beige, W. J. Munro, and P. L. K. Knight, Phys. Rev. A **62**, 052102 (2000).

[101] S. J. Freedman and J. F. Clauser, Phys. Rev. Lett. **28**, 938 (1972).

[102] J. F. Clauser and A. Shimony, Rep. Prog. Phys. **41**, 1883 (1978).

[103] E. S. Fry and R. C. Thompson, Phys. Rev. Lett. **37**, 465 (1976).

[104] A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett **49**, 1804 (1982).

[105] Z. Y. Ou and L. Mandel, Phys. Rev. Lett. **61**, 50 (1988).

[106] Y. H. Shih and C. O. Alley, Phys. Rev. Lett. **61**, 2921 (1988).

[107] Z. Y. Ou, S. F. Peirera, H. J. Kimble, and K. C. Peng, Phys. Rev. Lett. **68**, 3663 (1992).

[108] P. G. Kwiat, K. Mattle, H. Weinfurter, A. Zeilinger, A. V. Sergienko, and Y. H. Shih, Phys. Rev. Lett. **75**, 4337 (1995).

[109] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, Nature **409**, 791 (2001).

[110] S. Massar, Phys. Rev. A **65**, 032121 (2002).

[111] N. Gisin and H. Zbinden, Phys. Lett. A **264**, 103 (1999).

[112] A. Kent, arxiv.org e-print archive (2002), quant-ph/0204104.

[113] E. Santos, Phys. Rev. Lett. **46**, 1388 (1991).

[114] E. Santos, Phys. Rev. A **46**, 3646 (1992).

[115] W. Feller, in *An Introduction to Probability Theory and its Applications*, 3^{rd} ed. (John Wiley, New York, 1970), Vol. 1, Chap. 7.

[116] P. Zanardi, C. Zalka, and L. Faoro, Phys. Rev. A **62**, 030301(R) (2000).

[117] P. Zanardi, Phys. Rev. A **63**, 040304 (2001).

[118] X. Wang and P. Zanardi, arxiv.org e-print archive (2002), quant-ph/0207007.

[119] S. Popescu and D. Rohrlich, Phys. Rev. A **56**, R3319 (1997).

[120] M. S. Leifer, L. Henderson, and N. Linden, arxiv.org e-print archive (2002), quant-ph/0205055.

[121] C. H. Bennett, A. Harrow, D. W. Leung, and J. A. Smolin, arxiv.org e-print archive (2002), quant-ph/0205057.

[122] J. I. Cirac, W. Dür, B. Kraus, and M. Lewenstein, Phys. Rev. Lett. **86**, 052310 (2001).

[123] A. Jamiolkowski, Rep. of Math. Phy. No. 4 **3**, (1972).

[124] W. Dür and J. I. Cirac, Phys. Rev. Lett. **64**, 012317 (2001).

[125] W. Dür, G. Vidal, and J. I. Cirac, Phys. Rev. Lett. **89**, 057901 (2002).

[126] W. Dür and J. I. Cirac, arxiv.org e-print archive (2002), quant-ph/021112.

[127] B. Kraus and J. I. Cirac, Phys. Rev. A **63**, 062309 (2001).

[128] N. Khaneja, R. Brockett, and S. J. Glaser, Phys. Rev. A **63**, 032308 (2001).

[129] W. Dür, G. Vidal, J. I. Cirac, N. Linden, and S. Popescu, Phys. Rev. Lett. **87**, 137901 (2001).

[130] D. W. Berry and B. C. Sanders, arxiv.org e-print archive (2002), quant-ph/0205181.

[131] D. W. Berry and B. C. Sanders, arxiv.org e-print archive (2002), quant-ph/0207065.

[132] G. Vidal and J. I. Cirac, Phys. Rev. Lett. **88**, 167903 (2002).

[133] M. A. Nielsen, C. M. Dawson, J. L. Dodd, A. Gilchrist, D. Mortimer, T. J. Osborne, M. J. Bremner, A. W. Harrow, and A. Hines, arxiv.org e-print archive (2002), quant-ph/0208077.

[134] M. J. Bremner, C. M. Dawson, J. L. Dodd, A. Gilchrist, A. W. Harrow, D. Mortimer, M. A. Nielsen, and T. J. Osborne, arxiv.org e-print archive (2002), quant-ph/0207072.

[135] C. H. Bennett, J. I. Cirac, M. S. Leifer, D. W. Leung, N. Linden, S. Popescu, and G. Vidal, Phys. Rev. A **66**, 012305 (2002).

[136] G. Vidal, K. Hammerer, and J. I. Cirac, Phys. Rev. Lett. **88**, 237902 (2002).

[137] K. Hammerer, G. Vidal, and J. I. Cirac, arxiv.org e-print archive (2002), quant-ph/0205100.

[138] S. F. Huelga, J. A. Vaccaro, A. Chefles, and M. B. Plenio, Phys. Rev. A **63**, 042303 (2001).

[139] S. F. Huelga, M. B. Plenio, and J. A. Vaccaro, Phys. Rev. A **65**, 042316 (2002).

[140] B. Reznik, Y. Aharonov, and B. Groisman, Phys. Rev. A **65**, 032312 (2002).

[141] B. Reznik, arxiv.org e-print archive (2002), quant-ph/0203055.

[142] L. Landau and R. Peierls, Z. Phys. **69**, 56 (1931).

[143] Y. Aharonov and D. Z. Albert, Phys. Rev. D **24**, 359 (1981).

[144] Y. Aharonov, D. Z. Albert, and L. Vaidman, Phys. Rev. D **34**, 1805 (1986).

[145] S. Popescu and L. Vaidman, Phys. Rev. A **49**, 4331 (1994).

[146] D. Beckman, D. Gottesman, M. A. Nielsen, and J. Preskill, Phys. Rev. A **64**, 052309 (2001).

[147] B. Groisman and B. Reznik, Phys. Rev. A **66**, 022110 (2002).

[148] L. Vaidman, arxiv.org e-print archive (2001), quant-ph/0111124.