Nonlocality & Communication Complexity

Wim van Dam
Wolfson College

Centre for Quantum Computation
Department of Physics, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy in Physics at the Faculty of Physical Sciences, University of Oxford.

Nonlocality & Communication Complexity

Wim van Dam
Wolfson College, University of Oxford

D.Phil. thesis, Department of Physics
Michaelmas Term 1999

Abstract

This thesis discusses the connection between the nonlocal behavior of quantum mechanics and the communication complexity of distributed computations. The first three chapters provide an introduction to quantum information theory with an emphasis on the description of entangled systems. The next chapter looks at how to measure the complexity of distributed computations. This is expressed by the 'communication complexity', defined as the minimum amount of communication required for the evaluation of a function $f(x, y)$—a communication necessary because the input strings x and y are distributed over separated parties. In the theory of quantum communication, we try to use the nonlocal effects of entangled quantum bits to reduce communication complexity. In chapters 5, 6 and 7, such an improvement over classical communication is indeed established for various functions. However, it is also shown that entanglement does not lead to a more efficient calculation of the inner product function. We thus reach the conclusion that nonlocality sometimes—but not always—allows a reduction in communication complexity. This subtle relationship between nonlocality and communication vanishes when we consider 'superstrong' correlations. We demonstrate that if a violation of the Clauser-Horne-Shimony-Holt inequality with the maximum factor of 4 is assumed, all decision problems have the same trivial complexity of a single bit. The thesis concludes with an overview of the current status of quantum communication theory, and a discussion of the experimental feasability of the suggested protocols.

Acknowledgements

I would like to thank Artur Ekert for his guidance during my D.Phil. period at the Centre for Quantum Computation. (I know that he does not like to be called a ‘supervisor’ of anything or anybody, so I will refrain from doing so.) He created the freedom for me to do what I thought was best, while giving me ample opportunities to make the most out of my Oxford years. His view—that doing science is a joint effort that should be fun—was refreshing during my first encounters with the academic profession. And fun it was.

Many people helped me during my scientific explorations. I feel privileged to thank the following researchers for both their insightful comments and the enjoyable discussions I had with them: Andris Ambainis, P.K. Aravind, Guido Bacciagaluppi, Simon Benjamin, André Berthiaume, Dik Bouwmeester, Dagmar Bruß, Harvey Brown, Harry Buhrman, Richard Cleve, Mauro D’Ariano, David Deutsch, Mark Ettinger, Lance Fortnow, Michael Freedman, Chris Fuchs, Daniel Gottesman, Frederic Green, Peter Grünwald, Rasmus Hvas Hansen, Lucien Hardy, Patrick Hayden, Leah Henderson, Alexander Holevo, Peter Høyer, Neil Johnson, Jonathan Jones, Richard Jozsa, Adrian Kent, Sophie Laplante, Helger Lipmaa, Hoi-Kwong Lo, Chiara Macchiavello, Frédéric Magniez, David Meyer, Mike Mosca, Ashwin Nayak, Michael Nielsen, Asher Peres, Martin Plenio, Sandu Popescu, John Preskill, John Rogers, Louis Salvail, Miklos Santha, Ben Schumacher, Jason Semitecolos, Sjoerd Stallinga, Andrew Steane, Alain Tapp, Barbara Terhal, John Tromp, Umesh Vazirani, Vlatko Vedral, Paul Vitányi, John Watrous, Harald Weinfurter, Reinhard Werner, Ronald de Wolf, and Anton Zeilinger.

This research would not have been possible without the financial support of the Institute for Logic, Language and Computation in Amsterdam, and the European TMR Research Network (ERP-4061PL95-1412). I also enjoyed the contributions and hospitality of a great many other groups and institutions. They enabled me to do an almost obscene amount of travelling while I was supposed to be working on this thesis, and I want to thank them for this. They are: Berkeley, BRICS Aarhus, CIDEC Estonia, CWI Amsterdam, Elsag-Bailey, Hewlett Packard Bristol, Imperial College, ISI Torino, LRI Orsay, Microsoft Redmond, the Novartis Foundation, NWO, UNISYS, the TMR-Network of the European Commission, Wolfson College, and the universities of Bielefeld, Delft, Helsinki, Milan, Pavia and Utrecht.

Finally, my parents and friends are thanked for their love, understanding and friendship. All of the above should realize however, that they are only ‘second place’, as my greatest gratitude goes to my dearest Heather. Besides her encouragement, trust and rigorous proofreading, I thank her for our love.

Preface

This is the D.Phil. thesis of Wim van Dam. It contains the work that I did as a graduate student under the supervision of Artur Ekert at the Centre for Quantum Computation, University of Oxford.

The main part of this thesis deals with the investigation of quantum communication protocols that have a smaller complexity than any possible classical protocol: that is, quantum communication complexity. This advantage of quantum over classical is made possible by the nonlocal correlations, which can be established with entangled quantum bits.

The first four chapters of this thesis are of an introductory nature. In them, I give a brief overview of, respectively, quantum information, quantum communication, nonlocality, and communication complexity theory. Chapter 5 gives an example of two quantum communication protocols that have a reduced complexity when compared to classical procedures. The results of this chapter are described in

- "Quantum entanglement and communication complexity", by Harry Buhrman, Richard Cleve, and Wim van Dam, Technical Report RS-97-40 in the BRICS Research Series, University of Aarhus; quant-ph archive, report no. 9705033,

where the phrasing of the quantum protocol is due to the author of this thesis.

The 6th chapter generalizes the above protocol to the multiparty setting. It was published earlier as a part of

- "Multiparty quantum communication complexity", by Harry Buhrman, Wim van Dam, Peter Høyer, and Alain Tapp, *Physical Review A*, Volume 60, No. 4, pp. 2737–2741 (1999); quant-ph archive, report no. 9710054.

The proof method of the classical lower bound is my main contribution to this article.

Together with Lucien Hardy, I published the paper that is described in Chapter 7,

- "Quantum communication using a nonlocal Zeno effect", Lucien Hardy and Wim van Dam, *Physical Review A*, Volume 59, No. 4, pages 2635–2640 (1999); quant-ph archive, report no. 9805037.

It shows how the quantum Zeno effect of an entangled pair of qubits can be used to reduce the error in a one-bit communication protocol. The derivation of the minimal classical error rate is by my hand.

Chapter 8 shows that there are distributed functions that do not allow a reduction in complexity by the use of entanglement. The analysis of the two-bit case is my contribution to this part, with the corresponding publication

- “Quantum entanglement and the communication complexity of the inner product function”, by Richard Cleve, Wim van Dam, Michael Nielsen, and Alain Tapp, *Proceedings of the 1st NASA International Conference on Quantum Computing and Quantum Communications*, in Lecture Notes in Computer Science, No. 1509, (editor: Colin P. Williams), Springer-Verlag, pages 61–74 (1999); quant-ph archive, report no. 9708019.

The last chapter before the Conclusion discusses the consequences of superstrong nonlocality for communication complexity. This work will be published in the near future as a single author article.

Contents

Chapter 1

Introducing Quantum Information and Communication

In this thesis we investigate the theory of quantum information and communication. The current interest in this field is fueled by the discovery that the use of quantum mechanical processes provides us with an advantage over the traditional, classical ways of manipulating information. In this chapter I will introduce the notion of 'quantum information', and the standard notation as it will be used throughout the rest of the thesis.

Section 1.1 Modeling Information

The term 'bit' stands for 'binary digit', which reflects the fact that it can be described and implemented by a two-level system. Conventionally, these two levels are indicated by the labels "zero" and "one", or "0" and "1". If we want to capture more than two possibilities, more bits are needed: with k bits we have 2^k different labels.

The abstraction from k two-level systems to the set $\{0,1\}^k$ of size 2^k takes us away from the physical details of the implementation of a piece of memory in a computer, and instead focuses on a more mathematical description of information. This 'physics independent' approach to standard information theory has been extremely successful in the past decades: it enables a general understanding of computational and communicational processes that is applicable to all the different ways of implementing these processes. It is for this reason that the Turing machine model of computation gives an accurate description of both the mechanical computer suggested by Charles Babbage and the latest Silicon based Pentium III processors, despite their obvious physical differences. This does not mean that Turing's model ignores the physical reality of building a computer, on the contrary. The observation that it would be unphysical to assume an infinite or unbounded precision in the components of a computer is expressed by

Turing's rule that per time-step only a fixed, finite amount of computational work can be done.[63] The proper analysis of algorithms in the theory of computational complexity relies critically on the exclusion of computational models that are not realistic. Such models often give the wrong impression that certain complicated tasks are easy. (A good example of this is the result that the factorization of integers can be done in polynomial time if we assume that addition, multiplication and division of arbitrary big numbers can be done in constant time. (See Chapter 4.5.4, Exercise 40 in [40] and [60].) There is, however, also a danger with this axiomatization of the physical assumptions in information theory: believing that the assumptions are true. This is what happened with the traditional view on information, forgotten were the implicit classical assumptions that ignore the possibilities of quantum mechanics. The realization that quantum physics describes a world where information behaves differently than in classical theory led to the blossoming of several fields—quantum information, quantum computing, quantum communication, et cetera. In this thesis I will focus on the differences in communication complexity between a classical and a quantum model of communication. Before doing so, it is necessary to define what we mean by quantum information and communication.

Section 1.2 Quantum Information

At the heart of quantum mechanical information theory lies the *superposition principle.* Where a classical bit is either in the state "zero" or "one", a quantum bit is allowed to be in a superposition of the two states. A qubit with the label q is therefore described in Dirac's bra-ket notation by the linear combination:

$$|q\rangle \quad = \quad \alpha|\text{“zero”}\rangle + \beta|\text{“one”}\rangle,$$

where for the complex valued amplitudes $\alpha, \beta \in \mathbb{C}$, the normalization restriction $|\alpha|^2 + |\beta|^2 = 1$ applies. In this formalism, the state space of a single qubit is built up by the unit vectors in the two-dimensional Hilbert space $\mathcal{H}_2$. For k qubits, there are 2^k basis states and hence the corresponding superposition is a linear combination of all 2^k possible strings of k bits:

$$|q_1 \cdots q_k\rangle \quad = \quad \sum_{i \in \{0,1\}^k} \alpha_i |i\rangle.$$

Again it is required that the amplitudes α_i obey the normalization condition $\sum_i |\alpha_i|^2 = 1$. (In Section 1.4 we will see the reason behind this stipulation.) The state space of k qubits is the k-fold tensor product of the state space of a single qubit. This space is identical with a single 2^k-dimensional Hilbert space:

$$|q_1 \cdots q_k\rangle \quad \in \quad \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_2 = \mathcal{H}_{2^k}.$$

For our purposes we will only use finite sets of quantum bits, so there is no need to look at infinite-dimensional Hilbert spaces.

Section 1.3 Time Evolution of Quantum Bits

Quantum mechanics only allows transformations of states that are linear and respect the normalization restriction. When acting on an n-dimensional Hilbert space, these are the $n \times n$ complex valued rotation matrices that are norm preserving: the unitary matrices of $\mathrm{U}(n)$. It is easy to show that this corresponds exactly to the requirement that the inverse of U is the complex conjugate $U^\dagger$ of the matrix. (The complex conjugate is defined by $U^\dagger[r,c] = (U[c,r])^*$, where M['row', 'column'] is used to denote the different matrix entries.)

The effect of a unitary transformation U on a state x is exactly described by the corresponding rotation of the vector $|x\rangle$ in the appropriate Hilbert space. For this reason, "U" stands both for the quantum mechanical transformation as well as for the unitary rotation:

$$U|x\rangle \;=\; U\left(\sum_i \alpha_i|i\rangle\right) \;=\; \sum_i \alpha_i U|i\rangle \;=\; \sum_i \alpha_i \sum_j U[j,i]|j\rangle.$$

It follows from the associativity of matrix multiplication that the effect of two consecutive transformation U_1 and U_2 is the same as the single transformation $(U_2 \cdot U_1)$. Just as matrix multiplication does not commute, so does the order of a sequence of unitary transformations matter: in general $U_2U_1 \neq U_1U_2$. We can restate this in a more intuitive way by saying that it makes a difference if we first do U_1 and then U_2, or the other way around. (A convincing example is that of the two actions "add five" and "multiply by two".)

Section 1.4 Measurements

When measuring the state $|x\rangle = \sum_i \alpha_i|i\rangle$, the probability of observing the outcome "i" equals $|\alpha_i|^2$. This explains the normalization restriction on the amplitudes: the different probabilities have to add up to one. But what exactly is a 'measurement' and an 'observation', and how do we describe this mathematically? These are thorny issues that this thesis will leave untouched. Here we will only give a formal description of the measurement process and a short explanation of why this is such a problematic part of quantum mechanics.

The possible outcomes "i" of x correspond to a set of orthogonal vectors $\{|m_i\rangle\}_i$ of the measuring device. This device can be our own eye or some kind of machine, but the crucial point is that 'measuring x' implies 'interacting with x'. The *effect on* x of such a measurement is that the state *collapses* according to the outcome "i" of our observation. This is described by the transformation:

$$\sum_i \alpha_i|i\rangle \quad \rightsquigarrow_{\text{outcome } i} \quad |i\rangle. \tag{1.1}$$

The collapse as described above is a non-unitary transformation. This is typical when we try to describe the behavior of x as it interacts with a system that lies outside of the state. (We say that x is an 'open system'.) When we view x and the measurement device *together* during the observation, the evolution becomes unitary again. Our current example is then described by the transformation:

$$\sum_i \alpha_i|i\rangle \otimes |\text{measurement device}\rangle \quad\longrightarrow\quad \sum_i \alpha_i|i\rangle|\text{outcome } m_i\rangle.$$

The problem with this last description is that it no longer specifies the specific outcome "i" that we seem to observe. It is here where the debate on the *measurement problem* starts and our discussion ends.

For the purposes of this thesis it is more convenient to use the terminology of the collapsing quantum state. We will therefore describe the effect of a measurement as in Equation 1.1 for practical reasons. (This does not imply that I really think that there is such a collapse, but this issue are not the topic of this text. In this thesis we are mainly interested in the differences between the classical and the quantum mechanical theory of information. These differences, expressed in probabilities et cetera, are independent of the viewpoint that one has on the measurement problem.)

We just described the traditional 'Von Neumann measurement' where we observe the state x in the canonical basis spanned by the basis vectors i. Other, more subtle, measurement procedures are also possible by choosing an in- or over-complete basis. We will postpone the description of these two options the point when we discuss the density matrix formalism, which is more suitable for the general theory of interacting quantum mechanical systems.

Section 1.5 Limitations of Dirac's Ket Notation

The braket notation that we discussed above is tailor-made for the description of closed quantum mechanical systems. By this we mean the evolution of states that do not interact with an exterior environment. When we also want to consider the behavior of open systems, the ket-notation becomes less suitable. This was already obvious in the discussion of the measurement procedure where we had to expand the set of unitary operations with a probabilistic procedure that 'collapses' the quantum state to one of the basis states. One cannot help but feel uncomfortable about this sudden change of rules: is it not possible to deal with open and closed quantum systems in the same way? Luckily, we find in the formalism of density matrices a positive answer to this question.

Section 1.6 Density Matrices

An n-dimensional pure state x can be expressed as a normalized vector $|x\rangle$ in the Hilbert space $\mathcal{H}_n$. The complex conjugate $|x\rangle^\dagger$ of this vector is the bra $\langle x|$, which is an element of the adjoint space $\mathcal{H}_n^\dagger$. By taking the direct product between the ket $|x\rangle$ and the bra $\langle x|$, we thus obtain an $n \times n$ complex valued, Hermitian matrix: the *density matrix* of x.

As an example, for the state $|x\rangle = \sum_i \alpha_i|i\rangle$, the density matrix is:

$$|x\rangle\langle x| \;=\; \left(\sum_i \alpha_i|i\rangle\right)\left(\sum_j \alpha_j^*\langle j|\right) \;=\; \sum_{i,j} \alpha_i\alpha_j^*|i\rangle\langle j|.$$

In the case of a single qubit with the ket description $|q\rangle = \alpha|0\rangle + \beta|1\rangle$, this leads to the 2×2 matrix in the standard basis

$$|q\rangle\langle q| \;=\; \begin{bmatrix} |\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2 \end{bmatrix}.$$

From now on, the density matrix of the state x will be denoted by the same symbol x, and the fact that a matrix is a density matrix will be indicated by its square brackets.

The great advantage of this formalism is that it also allows the description of an *ensemble* of pure quantum states. If we have such a state ρ, which is a probabilistic mixture of the pure states $|x_t\rangle$ with probabilities p_t, then the matrix ρ is the weighted linear combination of the corresponding pure states matrices,

$$\rho \;=\; \sum_t p_t \cdot |x_t\rangle\langle x_t|,$$

with $p_t \geq 0$ and $\sum_t p_t = 1$.

Every density matrix that can be written as such a convex combination of pure states is a legal, or ‘allowed’, state, where allowed means, “allowed by the laws of quantum physics”. It follows from linear algebra that this restriction coincides with the requirement that the matrix is a Hermitian, positive, semidefinite matrix with unit trace.

The *spectral decomposition* of a proper density matrix ρ is done in terms of its eigenvalues λ_t and eigenvectors $|\omega_t\rangle$, by the equality

$$\rho \;=\; \sum_t \lambda_t|\omega_t\rangle\langle\omega_t|. \tag{1.2}$$

This shows that we can interpret ρ as the mixture $\{(\lambda_t, |\omega_t\rangle)\}_t$, where the states ω_t are pure and mutually orthogonal.

The above decomposition gives a convenient way of assigning a mixture to a given density matrix. It is important to realize, however, that a density matrix corresponds to a whole family of possible mixtures. Take the two ensembles $\{(\frac{1}{2}, |0\rangle), (\frac{1}{2}, |1\rangle)\}$ and $\{(\frac{1}{2}, \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)), (\frac{1}{2}, \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle))\}$, which have the same density matrix:

$$\begin{aligned} \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \;&=\; \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \\ &=\; \frac{1}{2}\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} + \frac{1}{2}\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}. \end{aligned}$$

We shall see that this implies that these two mixtures are indistinguishable from each other; it is therefore more accurate and less confusing to consider them as equivalent mixtures.

The density matrix of a *qubit* ρ in the standard basis is always of the form

$$\rho \;=\; \begin{bmatrix} p & \alpha^* \\ \alpha & 1-p \end{bmatrix},$$

with the probability p between 0 and 1 and the 'off-diagonal term' $|\alpha|^2 \le p(1-p)$. If $|\alpha|^2 = p(1-p)$ then ρ is a pure state with $|\rho\rangle = \sqrt{p}|0\rangle + \frac{\alpha}{\sqrt{p}}|1\rangle$ (or $|\rho\rangle = |1\rangle$ if $p=0$); otherwise the qubit ρ corresponds to a mixture.

Section 1.7

Separated Systems

We need the technique of density matrices to be able to describe the evolution of an open system. By 'open' we mean that there is a possible interaction between the quantum mechanical state and its environment, where the information in the latter is ignored (traced out). We already saw how a (pure) qubit changed into a probabilistic mixture after it interacted with a measurement device outside the qubit system.

This thesis analyses the possibilities of communication between remote parties. An individual party in this setting is therefore an open system as it interacts with the other participants. Here we will show how we describe local actions and observations in such an extended environment.

Let A and B denote two separated parties Alice and Bob, each with their personal qubits X^A and X^B. The joint 2-qubit space of A and B is the tensor product of the two subspaces $\mathcal{H}_A \otimes \mathcal{H}_B$, which is a 4-level system. The question now is: if there is a state X^{AB} that lives in this joint space $\mathcal{H}_A \otimes \mathcal{H}_B$, how does A's part of X^{AB} look like? Or more specifically, how do we calculate the local 2×2 density matrix X^A from the global 4×4 state X^{AB}? The answer is that this is done by 'tracing out' the environment (here B's part) of the state. The state space for X^{AB} is spanned by the 4-dimensional basis $V^{AB} = \{|0^A0^B\rangle, |0^A1^B\rangle, |1^A0^B\rangle, |1^A1^B\rangle\}$, which can be decomposed as the product of the bases of the two subsystems, $V^{AB} = V^A \times V^B = \{|0^A\rangle, |1^A\rangle\} \times \{|0^B\rangle, |1^B\rangle\}$. When we want to consider A's part of X^{AB}, we have to express this in the basis V^A while ignoring B's state space $\mathcal{H}_B$. This is calculated by

$$X^A \;=\; \mathrm{Trace}_B(X^{AB}) \;=\; \langle 0^B|X^{AB}|0^B\rangle + \langle 1^B|X^{AB}|1^B\rangle.$$

Conversely, if we want to know the state on B's side, we trace out A's part of the state space, $X^B = \mathrm{Trace}_A(X^{AB})$.

The above method is easily extended to the general case. For a joint state X^{AB} (where V^B by itself can represent a multipartite system), the density matrix on A's side is calculated by performing a partial trace over a complete basis for the state space of $\mathcal{H}_B$. If $\{|b_i\rangle\}_i$ is such a basis, then this is thus done by the calculation

$$X^A \;=\; \sum_i \langle b_i|X^{AB}|b_i\rangle.$$

The experienced reader must have noticed by now that we use a notation for mixed states in this thesis that is perhaps unconventional. If X^{AB} is a (pure) distributed state, *then X^A will refer to the (mixed) subsystem on A's side. This means that we allow the symbols X, ϕ and even Ψ to refer to a mixed state. I realize that this is not in accordance with most of the literature, but it gives us a more natural way of denoting the different parts of a distributed system.*

Section 1.8 Von Neumann Entropy of Mixed States

The eigenvalues λ_i of a density matrix are always non-negative and sum up to one. If we decompose a mixture into a linear combination of orthogonal pure states, then the λ's will correspond to the probabilities of the respective eigenvectors. (See Equation 1.2.) Although the eigenvectors of a density matrix are not always unique, its eigenvalues are. This allows us to unambiguously define the *Von Neumann entropy* $S(\rho)$ of a state with spectral decomposition $\rho = \sum_t \lambda_t|\omega_t\rangle\langle\omega_t|$ by

$$S(\rho) \quad = \quad -\sum_t \lambda_t \log_2 \lambda_t .$$

If we calculate the logarithm of a matrix with the Taylor expansion: $\log(\rho) = (\rho - I) - \frac{1}{2}(\rho - I)^2 + \frac{1}{3}(\rho - I)^3 - \cdots$, this can also be written as $S(\rho) = -\mathrm{Trace}(\rho \log_2 \rho)$. The Von Neumann entropy $S(\rho)$ reflects how 'mixed' or random ρ is, where pure states have zero entropy.

Section 1.9 Operations on Mixed States

A unitary transformation U maps the state $|x\rangle$ to the new pure state $U|x\rangle$. The latter reads as density matrix $U|x\rangle\langle x|U^\dagger$. In the language of density matrices, the corresponding transformation $\hat{U}$ is therefore calculated by 'sandwiching' the matrix x between U and its conjugate $U^\dagger$:

$$\hat{U}\left(|x\rangle\langle x|\right) \quad = \quad U|x\rangle\langle x|U^\dagger .$$

If we have a mixed state ρ, then $\hat{U}$ acts linearly on the eigenvectors of ρ. The following equation shows us that this calculation can be done without having to decompose ρ, and that our sandwich expression therefore also holds for mixed states:

$$\begin{aligned}
\hat{U}\left(\rho\right) &= \hat{U}\left(\sum_t \lambda_t|\omega_t\rangle\langle\omega_t|\right) \\
&= \sum_t \lambda_t \cdot \hat{U}\left(|\omega_t\rangle\langle\omega_t|\right) \\
&= \sum_t \lambda_t \cdot U|\omega_t\rangle\langle\omega_t|U^\dagger \\
&= U\left(\sum_t \lambda_t \cdot |\omega_t\rangle\langle\omega_t|\right)U^\dagger \\
&= U \cdot \rho \cdot U^\dagger .
\end{aligned}$$

It is clear that the positive eigenvalues λ_t of ρ remain unchanged, and that $\hat{U}$ only rotates the eigenvectors ω_t to the new eigenstates $\hat{U}(\omega_t)$.

Unitary operations are an example of completely-positive, trace preserving maps: every positive, semidefinite matrix is mapped to (another) positive, semidefinite matrix, and the trace of the matrix remains unaltered. Complete-positivity, in combination with the preservation of the trace, assures us that the result of a transformation will be a proper state if we started with a proper one.

Besides the unitary functions, there are other transformations that are possible in quantum mechanics. Just as mixed states are composed of pure states, so can a positive map be a linear combination of matrix multiplications similar to the ones we discussed above. An example of such a non-unitary mapping is the mapping $\hat{P}$, corresponding to a measurement of a qubit in the standard basis $\{0, 1\}$. This function consists of two 'projectors' $P_0 = |0\rangle\langle 0|$ and $P_1 = |1\rangle\langle 1|$ that transform a qubit ρ into a probabilistic mixture of the states 0 and 1. Explicitly:

$$\begin{aligned}
\hat{P}(\rho) &= \hat{P}\left(\begin{bmatrix} p & \alpha^* \\ \alpha & 1-p \end{bmatrix}\right) \\
&= \hat{P}_0\left(\begin{bmatrix} p & \alpha^* \\ \alpha & 1-p \end{bmatrix}\right) + \hat{P}_1\left(\begin{bmatrix} p & \alpha^* \\ \alpha & 1-p \end{bmatrix}\right) \\
&= \begin{bmatrix} p & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1-p \end{bmatrix} \\
&= \begin{bmatrix} p & 0 \\ 0 & 1-p \end{bmatrix}.
\end{aligned}$$

We see that the eigenvalues of the new density matrix are p and $1-p$ with the corresponding eigenvectors $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$. In general, the eigenvalues of ρ will change under this transformation and hence there it is no unitary operation that can establish the above mapping. In the next section we will give a formal description of all transformations, such as the above $\hat{P}$, that are allowed by quantum physics.

Section 1.10 Operator Sum Representation

The following requirements for an operator $\hat{E}$ are necessary and sufficient for $\hat{E}$ to be a proper quantum mechanical transformation:

1. The mapping $\hat{E}$ can be written as a set of matrices $\{E_k\}_k$ with which it maps a state ρ to the linear combination $\sum_k E_k \cdot \rho \cdot E_k^\dagger$.

2. The set of operators $\{E_i\}_i$ has to obey the identity restriction $\sum_k E_k^\dagger \cdot E_k = I$. (Note the change of order of E and $E^\dagger$ in the multiplication.)

These two requirements exactly describe the set of *completely-positive, trace preserving maps.* Complete-positivity means that we require both $\hat{E}$ as well its trivial extensions $\hat{E} \otimes \hat{I}$ to higher dimensions to be positive. This is a stronger condition than positivity. An example of a positive but not completely-positive map is the partial transpose $\hat{T}$, which is defined by $\hat{T}(\rho) = \rho^T$.

We have truly extended the set of unitary transformations and measurements by the above 'operator sum' formalism. An example of this is the mapping that erases a qubit and replaces it with the value zero. This non-unitary function is the combination of two operators

$$\hat{E} \quad = \quad \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \right\},$$

and has the same effect on every qubit ρ, namely

$$\begin{aligned} \hat{E}(\rho) &= \hat{E}\left(\begin{bmatrix} p & \alpha^* \\ \alpha & 1-p \end{bmatrix} \right) \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{bmatrix} p & \alpha^* \\ \alpha & 1-p \end{bmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \\ &\quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{bmatrix} p & \alpha^* \\ \alpha & 1-p \end{bmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \\ &= \begin{bmatrix} p & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 1-p & 0 \\ 0 & 0 \end{bmatrix} \\ &= |0\rangle\langle 0|. \end{aligned}$$

We previously argued that a measurement has a non-unitary effect on a state because we ignored its interaction with an outside system (the measurement device). This lesson holds for all allowed transformations:

> *Every completely-positive, trace preserving transformation $\hat{E}$ of a system $\mathcal{H}_A$ can be viewed as a part of unitary mapping $\hat{U}_E$ on a bigger system $\mathcal{H}_A \otimes \mathcal{H}_B$. That $\hat{E}$* by itself *appears to be non-unitary is due to the fact that we ignore the space $\mathcal{H}_B$.*

It can be shown that for the extension of the system it is sufficient to assume that the dimension of the appended space $\mathcal{H}_B$ is twice as large as that of $\mathcal{H}_A$, and that its initial state is $|0\cdots 0\rangle$. Hence, for every allowed quantum mechanical transformation $\hat{E}$ that acts on an n-dimensional system, there exists a unitary matrix $U_E \in \mathrm{U}(3n)$ such that

$$\hat{E}(x) \quad = \quad \mathrm{Trace}_B \left[U_E(x \otimes |0^B \cdots 0^B\rangle\langle 0^B \cdots 0^B|) U_E^\dagger \right]$$

for all x. This is, in more general terms, the difference that we encountered between the Equations 1.1 and 1.2. The non-unitary 'collapse' associated with an observation, or any other kind of interaction, is again a unitary transformation when we incorporate the measurement device into the description of the event.

The converse of the earlier statement also holds: every mapping that can be written as a traced-out, unitary transformation on a larger Hilbert space is a completely-positive, trace preserving mapping. The reader is referred to the standard book by Asher Peres[53] and the article by Benjamin Schumacher[59] for a more extended and rigorous treatment of this 'operator sum representation'.

Section 1.11 A Few Elementary Operations

In quantum computing and communication we look at the possibilities of transforming information as is allowed by the laws of quantum mechanics. We usually decompose such quantum algorithms in a series of small elementary steps that consist of one and two qubit operations. The following unitary gates are so commonly used that we will define them here in the introduction; we can therefore then use them throughout the rest of the thesis without having to specify them.

The Not gate: This is the same gate that we know of in classical computation with the additional characteristic that it respects the superposition of a qubit:

$$\textsc{Not}(\alpha|0\rangle + \beta|1\rangle) \quad = \quad \beta|0\rangle + \alpha|1\rangle.$$

Phase Flip: The $\textsc{Flip}$ gate changes the phase of a qubit conditional on its value:

$$\textsc{Flip}(\alpha|0\rangle + \beta|1\rangle) \quad = \quad \alpha|0\rangle - \beta|1\rangle.$$

Phase Rotation: A more general phase rotation is provided by the $\textsc{Phase}$ operation which has a free variable ϕ that determines the angle of the phase change:

$$\textsc{Phase}(\phi)(\alpha|0\rangle + \beta|1\rangle) \quad = \quad \alpha|0\rangle + \mathrm{e}^{\mathrm{i}\phi}\beta|1\rangle.$$

(Note: $\textsc{Flip} = \textsc{Phase}(\pi)$.)

Hadamard transform: This transformation H maps the zero and one state to a superposition of the two basis states:

$$H|0\rangle \quad = \quad \tfrac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \quad \text{and} \quad H|1\rangle \quad = \quad \tfrac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

The Hadamard is its own inverse ($H^2 = I$) and is often used in parallel on a k qubit register. Such a k-fold application of H translates the information of a classical string into the phases of a full superposition and back again:

$$|x_1 \cdots x_k\rangle \quad \longleftarrow H^{\otimes k} \longrightarrow \quad \tfrac{1}{\sqrt{2^k}} \sum_{y\in\{0,1\}^k} (-1)^{(x,y)}|y\rangle,$$

where (x,y) is the inner product modulo 2 of the k bit vectors x and y.

Rotations: With the rotation R with an angle ϕ, we mean the unitary one qubit transformation:

$$R(\phi) \quad = \quad \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}.$$

Controlled-Flip: The controlled-flip is a two-qubit operation that applies the $\textsc{Flip}$ gate to the target bit, if the control bit equals “1”; otherwise it leaves the target unchanged:

$$\text{C}\textsc{Flip}|x,y\rangle \quad = \quad (-1)^{xy}|x,y\rangle,$$

for all $x, y \in \{0,1\}$.

Section 1.12 No Influence-at-a-Distance

We conclude this chapter by a brief look at the typical example of a two qubit entangled state. Let Φ^{AB} be the pure state $|\Phi^{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. As a density matrix, this reads in the standard basis:

$$\Phi^{AB} \;=\; |\Phi^{AB}\rangle\langle\Phi^{AB}| \;=\; \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}.$$

When 'viewed' from either side as a subsystem, this pure state equals the maximally mixed qubit:

$$\Phi^{B} \;=\; \Phi^{A} \;=\; \langle 0^B|\Phi^{AB}|0^B\rangle + \langle 1^B|\Phi^{AB}|1^B\rangle \;=\; \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}.$$

The complete, entangled state is therefore fundamentally different from the tensor products of its subsystems: $\Phi^{AB} \neq \Phi^{A} \otimes \Phi^{B}$. In the next chapters we will see how different these entangled states are from states that can be written as tensor products. But before doing that, we will finish our introduction with an explanation why entanglement *cannot* be used for instantaneous information transfer.

Entanglement between Alice and Bob does not allow Bob to indirectly influence the state of Alice's system. Let X^{AB} be the joint state (and hence X^{A} the system Alice's side). Everything that Bob can do with his part X^{B}, can be described with the operator-sum representation. That this does not effect Alice's system can be expressed by the following equations. In the most general setting, the system X^{AB} is a mixture of pure states $|X^{AB}_t\rangle$, where each state can be written as a bipartite superposition $|X^{AB}_t\rangle = \sum_i \alpha_{ti}|X^{A}_{ti}X^{B}_{ti}\rangle$. This gives the following summation with probabilities p_t, and amplitudes α_{ti}:

$$\begin{aligned} X^{AB} &= \sum_t p_t \cdot |X^{AB}_t\rangle\langle X^{AB}_t| \\ &= \sum_t p_t \sum_{ij} \alpha_{ti}\alpha^*_{tj} \cdot |X^{A}_{ti}X^{B}_{ti}\rangle\langle X^{A}_{tj}X^{B}_{tj}| \\ &= \sum_{tij} p_t\alpha_{ti}\alpha^*_{tj} \cdot |X^{A}_{ti}\rangle\langle X^{A}_{tj}| \otimes |X^{B}_{ti}\rangle\langle X^{B}_{tj}|. \end{aligned}$$

From this expression we can now calculate the density matrix of Alice's subsystem X^{A} with the partial trace over Bob's part. This shows us that X^{A} is independent of the local transformations that Bob may have applied to his part X^{B}. For, if we assume that this action has the operator sum representation $\hat{E}_B(\rho) = \sum_k E_k \cdot \rho \cdot E^{\dagger}_k$, then the ‘new’ state $\tilde{X}^A$ on Alice’s side equals

$$\begin{aligned}
\tilde{X}^A &= \mathrm{Trace}_B[\hat{I}_A \otimes \hat{E}_B(X^{AB})] \\
&= \sum_{tij} p_t \alpha_{ti} \alpha^*_{tj} \cdot |X^A_{ti}\rangle\langle X^A_{tj}| \otimes \mathrm{Trace}\left[\hat{E}_B\left(|X^B_{ti}\rangle\langle X^B_{tj}|\right)\right] \\
&= \sum_{tij} p_t \alpha_{ti} \alpha^*_{tj} \cdot |X^A_{ti}\rangle\langle X^A_{tj}| \otimes \mathrm{Trace}\left[\sum_k E_k |X^B_{ti}\rangle\langle X^B_{tj}| E^\dagger_k\right] \\
&= \sum_{tij} p_t \alpha_{ti} \alpha^*_{tj} \cdot |X^A_{ti}\rangle\langle X^A_{tj}| \otimes \mathrm{Trace}\left[|X^B_{ti}\rangle\langle X^B_{tj}|\right].
\end{aligned}$$

(The last step in the above derivation uses the fact that $\mathrm{Trace}(A \cdot B) = \mathrm{Trace}(B \cdot A)$ in combination with the restriction that $\sum_k E^\dagger_k \cdot E_k = I_B$.) Clearly, the final outcome $\tilde{X}^A$ does not depend on the remote operation $\hat{E}_B$ of Bob, and hence equals the original state X^A.

Chapter 2

Quantum Communication

The theory of quantum communication looks at the consequences for information transfer if we allow the settings where we can send qubits and use entanglement. In this chapter some of the possibilities and impossibilities of quantum communication are explored. We pay special attention to the procedure of teleporting quantum states with classical signals. Also Holevo's bound on the amount of information that can be transfered with quantum signals is discussed.

Section 2.1 Entanglement

At the end of the previous chapter, we encountered a combination of two entangled qubits distributed over two parties A and B: $|\Phi_{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle+|11\rangle)$. It is impossible to write this pure state as a tensor product $a \otimes b$. Even if we allow a mixture of tensor products, such a decomposition remains impossible. We say that the $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is an *entangled state.* More explicitly, the definition of entanglement is as follows:

> *A bipartite system Ψ_{AB} is separable if and only if it can be expressed as a mixture of tensor products:*
>
> $$\Psi_{AB} = \sum_i p_i \cdot a_i \otimes b_i,$$
>
> *where p is a probability distribution, and a_i and b_i are quantum states on A and B's side respectively.*
>
> *A state that is* not separable *is* entangled.

The condition of entanglement is stronger than that of traditional correlations. Two classical bits that are either 00 or 11 (with a 50%-50% distribution) can be written as the unentangled mixture $\frac{1}{2}|00\rangle\langle 00|+\frac{1}{2}|11\rangle\langle 11|$. A system over A and B is uncorrelated if it can be written as a single tensor product $a \otimes b$ with again a and b (mixed) states on A and B's side. It is the difference between the '$\frac{1}{\sqrt{2}}$' amplitudes of the entangled state and the '50%' probabilities of the classical correlation that plays a crucial role here.

The question of how to decide, with an efficient procedure, whether a mixed state is entangled or not is still unresolved.[36, 37, 54] Although we have a clear definition of what it means for a state to be separable, it is still not clear how to search the continuous space of possible decompositions $\sum_i p_i \cdot a_i \otimes b_i$ with an algorithm that always gives a reliable answer. This is not only due to the finite precision of the algorithm, but, more important, also to the fact that we do not know when we can stop our search for an unentangled decomposition.

Here we will limit ourselves to an example for the entanglement of two qubits in the presence of noise: the Werner states. After that, we continue with a description of the protocols for *superdense coding* and *teleportation*, which highlight the usefulness of entanglement for purposes of communication.

Section 2.2 An Example: Werner States

We will use the family of Werner states[66] to clarify the difference between classical and quantum correlations. The state $|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is entangled, whereas the two random bits are $\frac{1}{4}I$ not even correlated. Hence, if we consider the one parameter family $\Phi^\lambda = \lambda \cdot \Phi + \frac{1-\lambda}{4} \cdot I$ for $0 \leq \lambda \leq 1$, we cover the whole spectrum from uncorrelated bits ($\lambda = 0$), to maximally entangled qubits ($\lambda = 1$). The critical point for Φ^λ to be an entangled state is $\lambda = \frac{1}{3}$. We will prove this in two parts: the separability of Φ^λ if $\lambda \leq \frac{1}{3}$, and the entanglement of Φ^λ for every $\lambda > \frac{1}{3}$.

Define the following six qubit states: $|\zeta_z^+\rangle = |0\rangle$, $|\zeta_z^-\rangle = |1\rangle$, $|\zeta_x^\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$, and $|\zeta_y^\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm \mathrm{i}|1\rangle)$. The reader is invited to check for him or herself that we can decompose $\Phi^{(1/3)}$ into a sum of zeta tensor-products:

$$\begin{aligned}
\Phi^{(\frac{1}{3})} &= \begin{pmatrix} \frac{1}{3} & 0 & 0 & \frac{1}{6} \\ 0 & \frac{1}{6} & 0 & 0 \\ 0 & 0 & \frac{1}{6} & 0 \\ \frac{1}{6} & 0 & 0 & \frac{1}{3} \end{pmatrix} \\
&= \tfrac{1}{6}\left(\zeta_x^+ \otimes \zeta_x^+ + \zeta_x^- \otimes \zeta_x^- + \zeta_y^+ \otimes \zeta_y^- + \zeta_y^- \otimes \zeta_y^+ + \zeta_z^+ \otimes \zeta_z^+ + \zeta_z^- \otimes \zeta_z^-\right),
\end{aligned}$$

hence $\Phi^{(1/3)}$ is separable. From this it follows that for any $\lambda \leq \frac{1}{3}$, Φ^λ is a mixture of two disentangled states: $\Phi^\lambda = 3\lambda \cdot \Phi^{(1/3)} + \frac{1-3\lambda}{4} \cdot I$.

The case when λ is bigger than $\frac{1}{3}$ is analyzed with inseparability criterion of Asher Peres[54]. This sufficient condition tells us that a bipartite state ρ^{AB} is not separable if the 'partially transposed' matrix $(\hat{I}_A \otimes \hat{T}_B)\rho^{AB}$ has negative eigenvalues. The reason for this is the following. If the matrix ρ^B represents a valid state, then so does its transpose $\hat{T}(\rho^B) = (\rho^B)^T$. Hence, for every disentangled state $\rho^{AB} = \sum_i p_i \cdot \rho_i^A \otimes \rho_i^B$, the partially transposed matrix $\sum_i p_i \cdot \rho_i^A \otimes (\rho_i^B)^T$ will also correspond to a proper state. If this is not the case—if the transposed matrix has negative eigenvalues—then we can conclude that the original matrix cannot be written as a sum of tensor products, and hence that ρ^{AB} is entangled. It is straightforward to verify that Φ^λ has negative eigenvalues under the transformation $\hat{I}_B \otimes \hat{T}_B$ if $\lambda > \frac{1}{3}$. This concludes our proof that the mixture $\Phi^\lambda = \lambda \cdot \Phi + \frac{1-\lambda}{4} \cdot I$ is entangled if and only if $\frac{1}{3} < \lambda \leq 1$. (Another proof of the entanglement of Φ^λ can be given in terms of 'distillable entanglement'. This is done in [10], where it is shown how one can create near perfect entangled states from an unlimited supply of Φ^λ pairs, under the assumption that $\lambda > \frac{1}{3}$.)

Section 2.3 Superdense Coding

The procedure of superdense coding shows us how we can transmit two bits of information with only one qubit. This result was published by Charles Bennett and Stephen Wiesner in 1992[8], and was one of the first examples of 'entanglement enhanced communication'.

Take two parties Alice and Bob (A and B) that want to communicate with each other. More specifically, Alice wants to send two classical bits of information to Bob with a minimum amount of effort. The setting is such that the two parties initially share one entangled pair of qubits $\Phi_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, and that Alice is allowed to use qubits for her signal, rather than classical bits. The following single qubit protocol establishes the 2 bit transmission. (The bits that Alice wants to send are labeled (x, y). The NOT and FLIP operations are unitary, one qubit transformations, defined by $\text{NOT}(\alpha|0\rangle + \beta|1\rangle) = \beta|0\rangle + \alpha|1\rangle$ and $\text{FLIP}(\alpha|0\rangle + \beta|1\rangle) = \alpha|0\rangle - \beta|1\rangle$.)

1. Depending on the x and y values, Alice performs the unitary operation $\text{NOT}^x \cdot \text{FLIP}^y$ to her qubit Φ_A of the entangled pair Φ_{AB}. (The qubit that is the result of this transformation is indicated by Φ_A^{xy}.)

2. Next, Alice send her qubit to Bob.

3. At this stage, Bob, who now possesses both the qubit Φ_A^{xy} and Φ_B, measures this entangled pair Φ_{AB}^{xy} in the four-dimensional basis

$$\begin{cases} |b_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), & |b_{01}\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \\ |b_{10}\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), & |b_{11}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \end{cases}$$

   It turns out that the label of the outcome $b_{??}$ of this measurement tells Bob exactly which bits x and y Alice wanted to convey, as he will always measure b_{xy}.

The correctness of this protocol is best proven by a case-by-case analyses of all four possibilities $xy \in \{00, 01, 10, 11\}$.

1. If $xy = 00$, then Alice did not change her qubit and hence the pair that Bob possessed before the measurement was indeed $|\Phi_{AB}^{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = |b_{00}\rangle$.

2. If $xy = 01$, then the FLIP action by Alice corresponded to the joint transformation $\text{FLIP} \otimes I_2$ on the pair Φ_{AB}, yielding the pure state $\frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) = |b_{01}\rangle$.

3. In a similar way, does the NOT on Alice's side (the case $xy = 10$) and the identity on Bob's side give the entangled pair $|\Phi_{AB}^{10}\rangle = |b_{10}\rangle$.

4. The combination of FLIP and NOT on Φ_A results in the state $|\Phi^{11}_{AB}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$, which can be detected with 100% reliability, corresponding to the right answer $xy = 11$.

As the four b states are mutually orthogonal, no confusion over the outcome is necessary if we assume that Alice and Bob are capable of perfect manipulation, transmission and observation of their qubits.

Superdense coding shows us how one entangled pair and one qubit of communication can be used to transmit two classical bits of information. The obvious question is: Can we improve this result by either increasing the number of bits transmitted, or by reducing the resources needed for this protocol? The answer is: No, this is not possible. In the next section, we will collect some of the evidence for this answer by looking at the *teleportation* protocol.

Section 2.4

Teleportation

How can we get a qubit across? If we have a perfect quantum channel between two parties, then we can simply send the quantum information from A to B. But what if we do not have have this possibility and there is only a classical channel at our disposal? The surprising answer is that the reliable transmission of qubits is still possible if the two parties share some entanglement between them. In 1993 Charles Bennett, Gilles Brassard, Claude Crépau, Richard Jozsa, Asher Peres and Bill Wootters showed how one entangled pair of qubits and two bits of classical communication are sufficient to transmit an unknown qubit between two parties.[9] (Note that when the properties of the qubit *are* known, a classical description of its parameters can be broadcasted over the classical channel and no entanglement is required.) This protocol, which has been coined ‘teleportation’, is in a sense the complement of superdense coding, which uses one entangled pair and one qubit of communication to convey two classical bits of information. The procedure is as follows.

Let Alice have a qubit q that she wants to convey to Bob. Both parties share the standard entangled pair $|\Phi_{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. The parameters of the qubit are $|q\rangle = \alpha|0\rangle + \beta|1\rangle$, but are unknown to the parties. Hence a complete classical description of the qubit is impossible to obtain. Instead, Alice will let q interact with her part Φ_A of the entangled pair Φ_{AB} by means of a measurement on the two qubits. The basis b of this measurement is equivalent to the one that we used in the superdense coding protocol:

$$\left\{ \begin{array}{rclcrcl} |b_{00}\rangle & = & \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), & \qquad & |b_{01}\rangle & = & \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \\ |b_{10}\rangle & = & \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), & & |b_{11}\rangle & = & \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \end{array} \right.$$

After this, q and Φ_A are ‘collapsed’ according to the outcome b_{xy}, and Bob’s qubit is no longer entangled with the system of Alice. Instead, his new Φ^{xy}_B is correlated with the initial qubit q and the measurement outcome xy in the following way:

outcome xy	Alice's 2 qubits	Bob's qubit $|\Phi_B^{xy}\rangle$
00	b_{00}	$\alpha|0\rangle + \beta|1\rangle$
01	b_{01}	$\alpha|0\rangle - \beta|1\rangle$
10	b_{10}	$\beta|0\rangle + \alpha|1\rangle$
11	b_{11}	$\beta|0\rangle - \alpha|1\rangle$

.

It is straightforward to verify that for every combination $xy \in \{0,1\}^2$ it holds that

$$\textsc{Not}^x \cdot \textsc{Flip}^y\, |\Phi_B^{xy}\rangle \quad = \quad \alpha|0\rangle + \beta|1\rangle \quad = \quad |q\rangle. \tag{2.1}$$

After the measurement, Alice therefore broadcasts the two classical bits x and y to Bob who then corrects his qubit Φ_B^{xy} according to Equation 2.1. This completes the teleportation procedure as Bob has now obtained a qubit with the same parameters $\alpha|0\rangle + \beta|1\rangle$, while on Alice's side no trace of the original qubit q is left. It is an important aside that during the protocol no information about q is obtained: all four measurement outcomes xy are equally likely and independent of the amplitudes α and β. We also do not 'copy' the qubit q as Alice loses all her information about q (see the next section for an explanation of why this is important).

There seems to be a close connection between superdense coding and teleportation: both use the same measurement basis, transformations and ingredients. This similarity can be used to prove that the two procedures are optimal with respect to their resources. But before establishing this result, we need to convince ourselves of a some important upper bounds on the transfer of information with quantum mechanical means.

Section 2.5 Information versus Information Representation

Thinking about qubits as states with complex valued parameters is sometimes misleading. The uncountably many different *mathematical expressions* $\alpha|0\rangle + \beta|1\rangle \in \mathcal{H}$ for a qubit, suggest that a single qubit contains an infinite amount of information, which is not the case. If we have a single copy of a qubit q, then only a small amount of information about its amplitudes can be obtained via a measurement. After this, the quantum state q has changed according to the observed outcome and no more information about the original amplitudes α and β is accessible.

Furthermore, the *no-cloning theorem* tells us that it is also impossible to copy an unknown quantum state q in order to obtain the tensor product $q \otimes q$.[67] This prevents us from creating a large set of identical qs, which would enable us to estimate α and β with arbitrary accuracy.

It is often claimed that the above are typical features of quantum information, but this is a misconception. To see this, it is instructive to realize that the same theorems also hold for classical, probabilistic information. It is impossible to infer more than one bit of information from the mixture $\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$, although for every $0 \leq p \leq 1$ this probabilistic bit ρ is different. Nor is it possible to reliably clone the unknown state ρ to $\rho \otimes \rho$. The conclusion should therefore be that in both cases of probabilities and probability amplitudes, the real and complex values of the state *description* are highly redundant when compared to the amount of accessible information in the state itself.

It is the combination of superpositions with the phenomenon of *interference* that makes the crucial difference between classical and quantum information. The possibility of the superpositions $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ to evolve to the different pure states $|0\rangle$ and $|1\rangle$ (after a Hadamard transform), somehow suggests that a quantum mechanical superposition is more 'real' than the probabilistic combination of two bit values. It seems as if for a qubit $\alpha|0\rangle + \beta|1\rangle$, both states are really present, whereas in the probabilistic case, the mixture $p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$ 'in reality' has already decided which binary value it represents. But this does not allow us to confuse a quantum mechanical superposition with its deterministic description as a density matrix on a piece of paper. Such confusion leads too easily to an overestimation of the inherent complexity of a single quantum state.

Section 2.6 Holevo's Bound and an Appendix to It

A more accurate analysis on the limitations of qubits to carry classical information is provided by Alexander Holevo's theorem on quantum sources[35] and an addendum to this result by Michael Nielsen (see the original [24] and the appendix of thesis).

For the purpose of this thesis we will focus here on the latter, but the reader is encouraged to familiarize him or herself with Holevo's result as well as with a recent generalization of this theorem by Ashwin Nayak[48].

Nielsen's result reads as follows. If Alice wants to transmit k bits of information to Bob and A and B start as unentangled systems, then this can only be done with at least k (quantum) bits of communication between the two parties. This can be further specified as a lower bound on the amount of communication from Alice to Bob (being k_{AB} qubits), and on the total amount of communication, $k_{AB} + k_{BA}$. (where k_{BA} is the number of qubits that Bob sends to Alice during the protocol). The bounds are in accordance with what we already know to be possible with superdense coding:

- For the communication from Alice to Bob: $k_{AB} \geq \lceil \frac{k}{2} \rceil$.
- For the total amount of communication: $k_{AB} + k_{BA} \geq k$.

We can reach Nielsen's bounds if we let Bob distribute k_{BA} (with $k_{BA} \leq \lfloor \frac{k}{2} \rfloor$) entangled pairs by sending k_{BA} qubits to Alice, who then uses k_{BA} qubits for superdense coding and $k - 2k_{BA}$ qubits for traditional communication. For every allowed value of k_{BA}, this protocol indeed uses $k_{AB} = k - k_{BA} = \lceil \frac{k}{2} \rceil$ (quantum) bits from Alice to Bob, and $k_{AB} + k_{BA} = k$ qubits in total.

Section 2.7 Optimality of Superdense Coding and Teleportation

A direct consequence of Nielsen's bound is that when Alice sends k qubits to Bob, she can only convey $2k$ classical bits of information. This, in combination with the protocols for superdense coding and teleportation, gives the following useful limits:

1. If Alice and Bob share initial entanglement and Alice sends k *classical* bits, then only k bits of information can be transmitted from Alice to Bob.

2. Superdense coding cannot be used to transmit more than two classical bits per qubit.
3. It is impossible to teleport a qubit with less than two classical bits of communication from Alice to Bob.

These three results are easily proven by the strong similarity between superdense coding and teleportation. Respectively:

1. By running two such protocols in parallel, Alice would be able (using superdense coding) to replace her $2k$ classical bits with k qubits. Hence we would have a protocol with $k_{AB} = k$ that transmits more than $2k$ bits of information from Alice to Bob. This is impossible.
2. This is a specific instance of Nielsen's result.
3. Assume that strictly less than 2 bits are necessary. For big enough N it should then be possible to teleport $N + 1$ qubits with $2N$ classical bits. Hence, if we would use the $N + 1$ qubits as part of a superdense coding procedure, we would transmit more than $2N + 2$ bits with $2N$ bits of classical information. This is not possible by the first result.

The preceding sections seems to suggest that the difference between quantum and probabilistic bits is 'a factor of two' and that teleportation and superdense coding summarize everything there is to know about (errorless) quantum communication. The fact that there are many more pages to follow in this thesis indicates that this is not the case. In the next chapter we will touch on a much discussed feature of quantum mechanics: nonlocality. We will see that there is a fundamental difference between the classical and the quantum theory of information after all, and that is *by definition* that there is no classical explanation of the 'nonlocal' correlations that are possible with entangled qubits.

Chapter 3

Nonlocality

In this chapter the issue of nonlocality is discussed. We look at how local hidden variable theories put a limit on the correlations that they can describe. The predictions of quantum mechanics violate these bounds, which tells us that the theory of quantum physics does not have a local, probabilistic model. Special attention is paid to the so-called 'loopholes' of experiments that try to verify the nonlocality of Nature.

Section 3.1 Bell's Inequality

It was in 1964 that John Bell gave a new impulse to the discussion on the foundations of quantum mechanics with his celebrated inequality of locality.[6, 7] Ever since then, other such inequalities have been derived, corresponding experiments have been performed, and heated debates are still being held about the exact implications of it all. It is the opinion of this author that the most important thing to understand about Bell's inequality is that does not try to say anything about the theory of quantum physics. Instead, it puts a general bound on all possible classical, local models for Nature. After the derivation of this bound there are two kinds of (possible) violations that draw our attention. The first one is the *mathematical fact* that conventional quantum mechanics gives predictions that are not possible to describe with a classical model. The *experimental verification* of the violation of the inequalities is the second and most important aspect of Bell's result. It is because of this dichotomy between theory and experiment that the nonlocality of Nature can be verified *independently* of the validity of our current theory of quantum mechanics.

Section 3.2 Classical or Hidden Variables Models

Crucial to a proper understanding of quantum nonlocality is the definition of what is meant with *classical locality*. In this thesis we adopt the (arguably conventional) interpretation of the terms 'local, realistic theory' and 'hidden variable model', which both refer to the same set of classical assumptions about a system. To avoid any unnecessary confusion, we will define these terms below.

When measuring a physical system $\bar{X}$, we observe certain outcomes with certain probabilities. Without loss of generality we assume here that we always have binary outcomes "yes" (1) or "no" (0). The probability of obtaining the answer "yes" when performing the measurement M on system $\bar{X}$ is denoted by $\mathrm{Prob}(M|\bar{X})$. A range of different measurements $M, M', \ldots$ on the same system leads to a corresponding range of probabilities $\mathrm{Prob}(M|\bar{X})$, $\mathrm{Prob}(M'|\bar{X}), \ldots$ We speak of a *deterministic system* X if for each measurement M_x, the outcome is completely predetermined. In this case, $\mathrm{Prob}(M_x|X)$ is always an element of $\{0,1\}$; and hence, with m different measurement settings ($m = |\{M_x\}_x|$), there are 2^m different deterministic systems.

A *probabilistic system* $\bar{X}$ is a mixture of deterministic systems X_i (indexed by i), with the probability distribution p: "$\bar{X} = \{(p_i, X_i)\}_i$". A measurement M on such a mixture $\bar{X}$ will therefore give the answer "yes" with probability $\mathrm{Prob}(M|\bar{X}) = \sum_i p_i \cdot \mathrm{Prob}(M|X_i)$. (Note that for the distribution p, it holds that $\sum_i p_i = 1$ and $p_i \geq 0$.) Just as the outcomes $\mathrm{Prob}(M|X) \in \{0,1\}$ for deterministic systems are predetermined, so are the probabilities $\mathrm{Prob}(M|\bar{X})$ completely specified in advance by the distribution p. This is the 'realistic' part of traditional theories: every characteristic that one can measure about a system is already described ('is real') in that system before the actual measurement.

Consider now a deterministic bipartite system X^{AB} that is distributed over Alice (whose subsystem is labeled X^A) and Bob (with his X^B). A model for X^A's behavior is considered 'local' if nothing outside the measurement setting M^A and the state X^A can influence the outcome of this specific experiment. This means that even though X^A was once part of a larger system X^{AB}, X^A *by itself* contains all the information about the way it will 'react' to the measurement M^A. For two different measurements M_1^A and M_2^A there are 4 deterministic subsystems X_i^A. The same applies for experiments done by Bob on his part X^B. From this it follows that we have 16 possible states $X^{AB} = (X_i^A, X_j^B)$ if X^{AB} has to give a local and deterministic description for the combinations of separated experiments (M_1^A, M_1^B), (M_1^A, M_2^B), (M_2^A, M_1^B) and (M_2^A, M_2^B). Note that when we drop the locality requirement, each experiment has four possible outcomes, leading to much more, $4^4 = 64$, different deterministic models.

A probabilistic bipartite system can again be described as a mixture p of deterministic states: $\bar{X}^{AB} = \{(p_{ij}, (X_i^A, X_j^B))\}_{ij}$. In such a scenario the probabilities for a measurement M^A are calculated by

$$\mathrm{Prob}(M^A|\bar{X}^{AB}) \quad = \quad \sum_{i,j} p_{ij} \cdot \mathrm{Prob}(M^A|X_i^A),$$

and similarly for Bob's side by

$$\mathrm{Prob}(M^B|\bar{X}^{AB}) \quad = \quad \sum_{i,j} p_{ij} \cdot \mathrm{Prob}(M^B|X_j^B).$$

The locality restriction does of course not forbid the existence of correlations between the two parts of $\bar{X}^{AB}$. It is very well possible to construct a distribution p such that

$$\mathrm{Prob}(M^A \cdot M^B|\bar{X}^{AB}) \quad \neq \quad \mathrm{Prob}(M^A|\bar{X}^{AB}) \cdot \mathrm{Prob}(M^B|\bar{X}^{AB}).$$

If there is a local, realistic theory for a system, then the behavior of this $\bar{X}$ is completely specified by its underlying distribution. Such a theory is therefore also called a 'hidden variable model', where the variables are understood to be defining function p. Bell's inequality gives us a limit to what is possible with systems that admit such a classical description.

Section 3.3 Two-Party Nonlocality

I will present here the variant of Bell's inequality as it was phrased by John Clauser, Michael Horne, Abner Shimony and Richard Holt in 1969: the CHSH inequality.[22] The traditional labeling with spin directions is replaced with an equivalent description in bit values as this is how we will use the result later in the thesis.

Consider two separated parties A and B who both receive a subsystem X^A and X^B. Each side chooses to perform one out of two experiments: M_0^A or M_1^A on Alice's side, and M_0^B or M_1^B for Bob's part. This procedure is repeated many times such that all four possible measurement settings can be examined. We are interested in the correlated ($M_\cdot^A = M_\cdot^B$) and anti-correlated ($M_\cdot^A \neq M_\cdot^B$) outcomes for those four possibilities. By using binary values in combination with modulo two arithmetic (with $1 \oplus 1 = 0$), we can rewrite these (anti)-correlations as

$$M^A \oplus M^B = \begin{cases} 0 & \text{if the outcomes } M^A \text{ and } M^B \text{ are correlated,} \\ 1 & \text{if the outcomes } M^A \text{ and } M^B \text{ are anti-correlated.} \end{cases}$$

After a sufficient number of experimental runs, Alice and Bob should be able to estimate the overall likelihood that the outcomes $M_x^A \oplus M_y^B$ equals $x \cdot y$ for $x, y \in \{0,1\}$. If the experimental settings xy are chosen at random on both sides, this correlation equals

$$\begin{aligned} \mathrm{Corr}_{\text{Bell}} &= \tfrac{1}{4}\sum_{x,y} \mathrm{Prob}(M_x^A \oplus M_y^B = xy) \\ &= \tfrac{1}{4}\mathrm{Prob}(M_0^A = M_0^B) + \tfrac{1}{4}\mathrm{Prob}(M_0^A = M_1^B) + \\ &\quad \tfrac{1}{4}\mathrm{Prob}(M_1^A = M_0^B) + \tfrac{1}{4}\mathrm{Prob}(M_1^A \neq M_1^B). \end{aligned}$$

Assume now that the state X^{AB} is a deterministic one, and hence that all occurring probabilities are 0% or 100%. Inspection of the 16 possible systems X^{AB} shows that the value $\mathrm{Corr}_{\text{Bell}}$ will always be bounded by $\mathrm{Corr}_{\text{Bell}}^{\text{det}} \leq \frac{3}{4}$. (Take, for example, $\mathrm{Prob}(M_0^A) = \mathrm{Prob}(M_0^B) = \mathrm{Prob}(M_1^B) = 1$ and $\mathrm{Prob}(M_1^A) = 0$, then $\mathrm{Corr} = \frac{1}{4}(1+1+0+1) = \frac{3}{4}$.) Allowing the system to be probabilistic (with $\bar{X}^{AB} = \{(p_i, X^{AB})\}_i$) does not change these bounds on $\mathrm{Corr}_{\text{Bell}}$ as the expected value is a weighted sum of the deterministic cases:

$$\mathrm{Corr}_{\text{Bell}}^{\text{prob}}\left(\bar{X}^{AB}\right) = \sum_{i=1}^{16} p_i \cdot \mathrm{Corr}_{\text{Bell}}^{\text{det}}(X_i^{AB}),$$

with $\sum_i p_i = 1$. The conclusion is therefore that for every system $\bar{X}^{AB}$ that can be described by a hidden variable model p, the restriction "$\mathrm{Corr}_{\text{Bell}}^{\text{clas}} \leq \frac{3}{4}$" holds.

The theory of quantum mechanics surpasses the above bound. Take instead of X^{AB} an entangled pair of qubits $|\Psi^{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$. Let the measurements M_0 and M_1 be the projection on the respective vector (for both sides):

$$|M_0\rangle = \sin\left(\tfrac{\pi}{16}\right)|0\rangle + \cos\left(\tfrac{\pi}{16}\right)|1\rangle \quad \text{and} \quad |M_1\rangle = \sin\left(\tfrac{3\pi}{16}\right)|0\rangle - \cos\left(\tfrac{3\pi}{16}\right)|1\rangle.$$

It does not involve much mathematics to verify that for this setting, the expected correlations have become:

$$\left.\begin{array}{ll} \mathrm{Prob}(M_0^A = M_0^B), & \mathrm{Prob}(M_0^A = M_1^B) \\ \mathrm{Prob}(M_1^A = M_0^B), & \mathrm{Prob}(M_1^A \neq M_1^B) \end{array}\right\} \;=\; \tfrac{1}{2} + \tfrac{\sqrt{2}}{4},$$

leading to the combined sum $\mathrm{Corr}_{\mathrm{Bell}}^{\mathrm{qm}} = \frac{1}{2} + \frac{\sqrt{2}}{4} \approx 0.853$. This shows that the theory of quantum mechanics cannot be captured by a classical model that uses local hidden variables. A more detailed analysis of what the crucial ingredients of the above argument are will be done after the following inequality for more than two parties is discussed.

Section 3.4 Three-Party Nonlocality

The following nonlocality proof involves three parties and is generally considered more ‘convincing’ than the results of the previous section. It was introduced by David Mermin[45, 46] as a rephrasing of the original four-party example by Daniel Greenberger, Michael Horne and Anton Zeilinger[30].

We will label the parties A, B and C, and the systems they receive X^A, X^B and X^C respectively. As in the previous example, we allow the participants to use one out of two measurement settings (M_0 and M_1). This time we are interested in the correlation term

$$\begin{aligned} \mathrm{Corr}_{\mathrm{GHZ}} \;=\; & \tfrac{1}{4}\mathrm{Prob}(M_0^A \oplus M_0^B \oplus M_0^C = 0) + \tfrac{1}{4}\mathrm{Prob}(M_0^A \oplus M_1^B \oplus M_1^C = 1) + \\ & \tfrac{1}{4}\mathrm{Prob}(M_1^A \oplus M_0^B \oplus M_1^C = 1) + \tfrac{1}{4}\mathrm{Prob}(M_1^A \oplus M_1^B \oplus M_0^C = 1), \end{aligned}$$

which is again estimated with the outcomes of many different experimental runs.

The scenario where X^{ABC} is a deterministic system bounds the possible value of Corr from above by $\frac{3}{4}$, as can easily be shown. Assume a local, deterministic system X that obtains a correlation ratio strictly bigger than $\frac{3}{4}$. For this to be possible, X has to fulfill the first three clauses of the $\mathrm{Corr}_{\mathrm{GHZ}}$ expression, and hence has to obey:

$$\begin{aligned} \mathrm{Prob}(M_0^A|X) \oplus \mathrm{Prob}(M_0^B|X) \oplus \mathrm{Prob}(M_0^C|X) &= 0 \\ \mathrm{Prob}(M_0^A|X) \oplus \mathrm{Prob}(M_1^B|X) \oplus \mathrm{Prob}(M_1^C|X) &= 1 \\ \mathrm{Prob}(M_1^A|X) \oplus \mathrm{Prob}(M_0^B|X) \oplus \mathrm{Prob}(M_1^C|X) &= 1. \end{aligned}$$

By adding these three equalities we can now infer that

$$\mathrm{Prob}(M_1^A|X) \oplus \mathrm{Prob}(M_1^B|X) \oplus \mathrm{Prob}(M_0^C|X) = 0.$$

(We used here the fact that all probabilities are zero or one and thus $\mathrm{Prob}(M|X) \oplus \mathrm{Prob}(M|X) = 0$ for any M.) This conclusion contradicts the fourth clause of the GHZ-term, proving that for this system $\mathrm{Corr}^{X}_{\mathrm{GHZ}} = \frac{3}{4}$.

This bound immediately implies that all probabilistic, hidden variable models for $\bar{X}^{ABC}$'s behavior have to obey the same bound:

$$\mathrm{Corr}^{\mathrm{clas}}_{\mathrm{GHZ}} \quad \leq \quad \tfrac{3}{4}. \tag{3.1}$$

By using a three qubit entangled state we can go beyond this limit and, in fact, reach the maximum possible value

$$\mathrm{Corr}^{\mathrm{qm}}_{\mathrm{GHZ}} \quad = \quad 1.$$

Below we will see how the theory of quantum mechanics establishes this correlation factor.

Distribute the three entangled qubits $|\Psi^{ABC}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$ over the parties A, B and C. All three positions use the same projectors for their two possible experiments:

$$|M_0\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \quad \text{and} \quad |M_1\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle - i|1\rangle).$$

With this set-up, the four correlation values are indeed all equal to one:

$$\left.\begin{array}{ll} \mathrm{Prob}(M_0^A \oplus M_0^B \oplus M_0^C = 0), & \mathrm{Prob}(M_0^A \oplus M_1^B \oplus M_1^C = 1) \\ \mathrm{Prob}(M_1^A \oplus M_0^B \oplus M_1^C = 1), & \mathrm{Prob}(M_1^A \oplus M_1^B \oplus M_0^C = 1) \end{array}\right\} \quad = \quad 1.$$

This adds up to the total value $\mathrm{Corr}_{\mathrm{GHZ}} = 1$, which violates the classical bound of Equation 3.1.

What do nonlocality proofs tell us about the difference between the classical and the quantum theory of information? We now know that it is sometimes impossible to mimic the joint behavior of entangled but spatially separated qubits with a classical system in the same setting. This impossibility disappears if we let go of some of the assumptions in the description of the experiment. In the next section we will try to get a better understanding of such 'locality loopholes' as this will give us a clearer insight into the subtleties of the above results.

Section 3.5 Locality Loopholes

When is a classical system $\bar{X}^{AB}$ able to simulate the predictions of quantum mechanics? A partial answer is that this simulation is possible when A's system X^A has knowledge about the setting $M_?^B$ on Bob's side, or vice versa. This knowledge can be obtained in different ways, each leading to a potential loophole for an experimental verification of Nature's nonlocality:

No-signaling requirement: It should be impossible for $\bar{X}^A$ to broadcast any information to Bob's side about the measurement setting $M_?^A$ that it has encountered.

The no-signaling requirement is fulfilled if both measurements M^A and M^B are space-like separated events in space-time. Special relativity then tells us then that no information can travel between the two acts of measurement. Note that this space-like separation is only a method to establish the no-signaling condition. It would be equally valid if we were able to prohibit the transfer of information between A and B by other means.

Unpredictable measurement settings: The transfer of information between the two parties is unnecessary if the measurement settings are known to the systems $\bar{X}^A$ and $\bar{X}^B$ from the start. It is straightforward to reproduce the statistics of quantum mechanics if the four different experimental settings $M_?^A M_?^B$ occur in a regular pattern that can be predicted by the system $\bar{X}^{AB}$ before it separates into two subsystems. The choice on both sides should therefore be made at random and independently of each other. (The independence can again be established by making the two decisions at space-like separated events.)

Besides the aforementioned two restrictions, there is a third, more practical, way for a model to mask its classical foundations: the detector efficiency loophole. In practice it will almost never be the case that every signal can be detected by the measurement apparatus. As an example, with current technology, the detection of both the polarizations of entangled photons succeeds with a success probability of less than one percent. In such situations it is possible to come up with a classical model where the photons only ‘reveal’ themselves at M^A and M^B if the setting of the devices is in accordance with a scheme that was agreed upon before $\bar{X}^A$ and $\bar{X}^B$ parted. When one of the photons encounters an undesired setting, this particle then ‘hides’ itself from the detector, resulting in just one of the many unsuccessful polarization measurements. Such (admittedly contrived) ‘conspiracy theories’ are able to give a local explanation for all the performed nonlocality experiments to date.[52]

The reader might wonder what the practical merits are of these academic objections to the acceptance of nonlocality as a feature of Nature. After all, if our quantum communication protocol works as desired, why contemplate the ins and outs of the model that describes it? The surprising counter argument to this critique is that the above conditions translate directly into the requirements for a quantum protocol that truly outperforms the classical ways of processing information. This is the exciting idea behind quantum communication as I will discuss it in this thesis: to use Nature’s nonlocality to save on the amount of communication that is necessary in certain settings.

In the next chapters, we will see how the above arguments about the foundations of quantum mechanics can be transformed into procedures that reduce the complexity of distributed calculations. But before we are able to do this, it will be necessary to introduce a notion from computer science: communication complexity.

Chapter 4

Communication Complexity

In this chapter we introduce the notion of communication complexity. It is first defined in the traditional, classical sense after which we expand it to the quantum case. Also the generalization to multiple parties is made. Special attention is paid to the notion of probabilistic protocols and how they can be viewed as mixtures of deterministic communication procedures.

Section 4.1 Introduction

Consider two remote parties Alice and Bob each in possession of data that is unknown to the other person. If Alice has a natural number x and Bob has y, how many bits does Bob have to send to Alice such that she will be able to determine if $x + y$ is even or odd? Clearly this can be done with a single bit of information as Alice (who knows x) is only interested in whether or not Bob's y is even or odd. But what if Alice wants to decide if $x+y$ is prime? Intuitively one expects that in order to determine this decision problem, Alice and Bob will have to exchange more information than the previous one bit, and that this amount of communication will depend on the sizes $|x|$ and $|y|$ of the input strings. But how will it depend on the input size? What is the most efficient protocol? And given this optimal solution, how do we prove that there does not exist a better procedure? The theory of *communication complexity* tries to answer questions like these.

Section 4.2 Two-Party Communication Complexity

The setting for communication problems where there are two cooperative parties who want to compute a joint decision problem is as follows.

Alice and Bob are given two strings x and y respectively, both of length n. They want to compute a Boolean function f on these two input strings, hence for a given n, the function f will be of the form $f_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. The communication complexity of this function is the minimal amount of communication between the two parties that is necessary for Alice to calculate the binary value $f_n(x,y)$. More precisely, the complexity of the distributed task f is expressed by the relation between the input size n and the amount of communication necessary for the evaluation of $f_n(x,y)$ for the *worst case input strings* x and y. The following observations should clarify this definition.

Section 4.3 Some Observations about Communication Complexity

The trivial example of the "$x+y$ even or odd?" problem in the beginning of this chapter is one of the simple cases where the communication complexity is constant and hence independent of the input size. The version where Alice tries to determine the primality of $x+y$ has the obvious upper bound of n (which holds for *any* f_n), because Bob can always send all his n bits to Alice who then finishes the computation "$x+y$ prime?" on her side. This underlines the fact that the *computational difficulty* of determining the function PRIME (or any other function f) does not play a role here.

Because of the worst case assumption, the following line of reasoning is incorrect: "The sum $x+y$ will be even (and hence composite) 50% of the time. This can be checked with a single bit of communication; therefore, the average communication complexity of the PRIME$_n$ problem will be less or equal to $\frac{n}{2}$." Instead, we should conclude that the complexity of PRIME is going to be determined by the values of x and y for which their sum is not divisible by two.

The fact that Bob does not have to know the answer after the protocol does not have any significant consequences: it will only require one additional bit of communication for Alice to tell the final answer $f(x,y)$ to Bob.

Section 4.4 Formal Definition of Deterministic Communication

A deterministic protocol D fully determines for every possible input (x,y) which party is going to communicate which bit at what stage of the protocol. At the start of the procedure, the parties are unaware of each others inputs; therefore, who is going to communicate the first bit has to be 'input independent' (and hence pre-determined). If we assume that this is Alice, then she has to act according to two *decision sets* A_0 and A_1 in that she sends a "zero" to Bob if and only if her input $x \in A_0$, and a "one" if and only if $x \in A_1$. Because we require the protocol to be unambiguous and well-defined for every x, it follows that $A_0 \cap A_1$ is empty *and* that $A_0 \cup A_1$ covers the whole set of possible inputs for Alice. For the second bit, the situation becomes more complicated as we now have the two situations where the first communicated bit was zero or one. We will make this distinction by putting the relevant history of communication in the upper indices of the decision sets. Hence, we could have the description in the form of the two couples (A_0^0, A_1^0) and (B_0^1, B_1^1), which would tell us that depending on the first bit, Alice or Bob announces the second bit. More specifically, if the first value was zero, then Alice continues according to the sets A_0^0 and A_1^0. Otherwise, Bob uses his decision sets B_0^1 and B_1^1. Again, each couple of sets obeys the above-mentioned restrictions for a deterministic protocol. In general, we will completely specify what happens after Alice and Bob have communicated the string $c \in \{0,1\}^*$ by either a pair (A_0^c, A_1^c) if Alice has to communicate the next bit, or (B_0^c, B_1^c) if it is Bob's turn. Notice that it is not possible to have a combination of A^c and B^c as this makes it ambiguous who is going to communicate. At the end of the protocol it is Alice who has to determine the function value $f(x, y)$. Similarly, we can represent this with two decision sets for A, as long as we understand that this time the lower index denotes the final decision and is no longer part of the communication.

We can visualize a deterministic protocol as a decision tree where the nodes are labeled by the strings c that express the 'communication so far' and the branches by the respective decision sets A_0^c, A_1^c, et cetera. Figure 4.1 shows such a tree for a simple three bit communication protocol, which we shall use for the following example.

Imagine a two-party protocol D where both Alice and Bob receive two bits (x_1x_2, $y_1y_2 \in \{0,1\}^2$) that is described by:

1. If Alice has $x_1 = 0$, then she sends a "zero" to Bob, who then knows that the protocol has ended.

2. Otherwise, Bob will receive the value "1" from Alice, telling him that he has to communicate back both his input values y_1 and y_2.

3. After the communication has ended, Alice calculates the outcome of the protocol: $D(x, y)$ is the bit value $x_1 \cdot (y_1 \oplus x_2y_2)$. She is able to do this either because $x_1 = 0$, or on the basis of her knowledge of Bob's two input bits y_1 and y_2 (in combination with her own input x_1x_2).

A description such as the one above easily becomes unclear for larger protocols. It is for this reason that we use the formalism of decision sets. The three steps of the above example are thus summarized by the pairs of sets:

1. Alice's sets $A_0 = \{00, 01\}$ and $A_1 = \{10, 11\}$.

2. Bob's first bit with $B_0^1 = \{00, 01\}$ and $B_1^1 = \{10, 11\}$; his second bit: $B_0^{10} = B_0^{11} = \{00, 10\}$ and $B_1^{10} = B_1^{11} = \{01, 11\}$.

3. And the final conclusion of Alice by the table:

$$\begin{array}{rcl@{\qquad}c@{\qquad}rcl}
A_0^0 & = & \{00, 01, 10, 11\} & \text{and} & A_1^0 & = & \{\} \\
A_0^{100} & = & \{00, 01, 10, 11\} & & A_1^{100} & = & \{\} \\
A_0^{101} & = & \{00, 01, 10\} & & A_1^{101} & = & \{11\} \\
A_0^{110} & = & \{00, 01\} & & A_1^{110} & = & \{10, 11\} \\
A_0^{111} & = & \{00, 01, 11\} & & A_1^{111} & = & \{10\}.
\end{array}$$

Notice that the 'completeness requirement' for the union $B_0^c \cup B_1^c = \{00, 01, 10, 11\}$ sometimes leads to a redundancy in the sets as they cover input states that do not apply to them. (For example, the fact that $y_1y_2 = 00$ is an element of B_0^{11}, which only is used when $y_1 = 1$.) By the tree construction we also see that the strings $\{c_i\}_i$ of the final conclusion sets for Alice form a complete, self-delimiting code such that no c_i is the prefix of another string c_j, and any sufficiently long bit string starts with one of the words c_i.

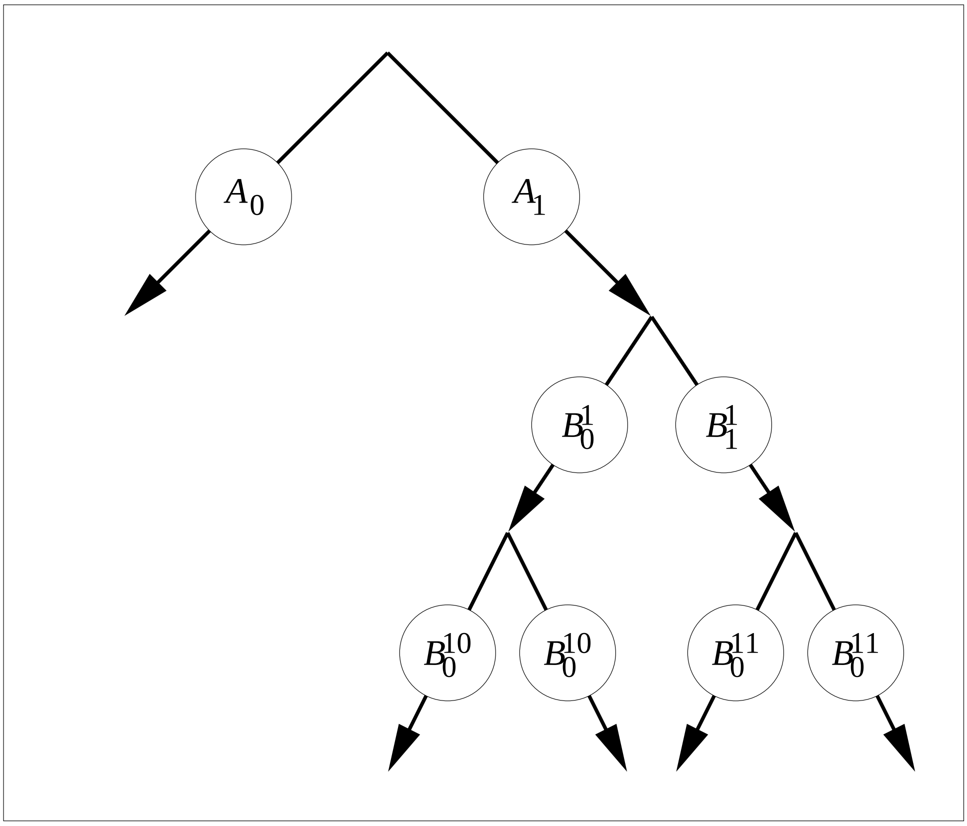

Figure 4.1: *The decision tree of a simple deterministic communication protocol for two parties. Alice starts by sending one bit to Bob. If this bit has value zero, then the procedure has ended and Bob does not communicate anything to Alice. Otherwise (the right part of the tree), Bob has to send back two bits of information before Alice is able to determine the outcome of the procedure. Notice that the sets that Alice uses for this conclusion are not shown here. (See the main text for a fully worked out example of this tree.)*

The two important characteristics of decision trees and their sets that we mentioned earlier will be repeated here formally. *For a deterministic protocol that calculates a function $f : X \times X \to \{0,1\}$, the decision sets have to obey the following:*

1. Every occurring node $c \in \{0,1\}^*$ in the tree either contains a pair (A_0^c, A_1^c) or (B_0^c, B_1^c).
2. For every pair (X_0^c, X_1^c) it holds that $X_0^c \cap X_1^c = \{\}$ and $X_0^c \cup X_1^c = X$.
3. All the 'leaves' (or end nodes) of the tree are A pairs as these contain the conclusion of Alice after she has completed the communication with Bob.

The amount of communication before Alice's conclusion corresponds exactly with the length $|c|$ of the string c that labels the leaves A^c. The worst case assumption tells us that the communication complexity of a protocol is the longest possible c that appears in a leaf. This is identical with the depth of the decision tree, minus one. Our tree-example for the calculation of the function $x_1 \cdot (y_1 \oplus x_2 y_2)$ therefore has complexity 3, despite the fact that for the case $x_1 = 0$, it only requires one bit of communication.

Section 4.5 Probabilistic Protocols

We speak of a probabilistic solution if the parties use a protocol that gives Alice the correct answer $f(x,y)$ with *high probability* for all combinations (x,y). The minimum correctness ratio $1 - \varepsilon$ (a protocol with probability of error ε) will in general influence the amount of communication that is necessary to obtain the confidence level. With $\varepsilon = 0$ we obviously return to the deterministic case.

An important aspect of the definition lies in the phrase "for all combinations (x,y)" which I will clarify here. Imagine a deterministic protocol D that is successful for all possible input combinations (x,y) except one. At first glance this may seem a reliable solution of the problem. But the worst-case assumption tells us, in fact, that with this protocol D, we should expect the values x and y for which the procedure fails. This teaches us that if there is an input on which a deterministic procedure makes an error, then this protocol has to be considered useless (the error ε equals 1). For a successful probabilistic approach, we need to add randomness as one of our ingredients.

The reason that our deterministic protocol failed was because its errors were also deterministic and hence predictable by the worst-case distribution μ that specifically 'aims' at the weak spots of proposed solutions. We counter this by randomizing the errors ourselves in the sense that we try to 'spread out' the values (x,y) for which we are likely to make a mistake. Such an approach requires Alice and Bob to share some random bits on the basis of which they execute their otherwise deterministic protocol. The following example should be instructive.

Assume that Alice and Bob try to calculate the distributed function $f : N \times N \to \{0,1\}$ (where N stands for the set of input values $\{1,\ldots,n\}$). Imagine that for each $(i,j) \in N \times N$ there exists an efficient deterministic protocol D_{ij} that works correctly except for the one combination $(x = i, y = j)$. That is, every protocol D can make a mistake, but the protocols differ in *where* they err. This allows A and B to adopt the following strategy:

1. Alice receives x, Bob receives his $y \in N$.
2. Both parties agree *at random* on a pair (i,j) that determines which one of the n^2 protocols D they are going to execute.
3. Alice and Bob perform the deterministic protocol D_{ij}.

(Note that the random bits ($2\log n$ in total) are used *before* the actual communication procedure.) Unlike the earlier non-randomized approach, this protocol is highly successful, as we can easily see. Given any combination (x,y), Alice and Bob have probability $\frac{1}{n^2}$ that they ended up executing the flawed protocol D_{xy} for this specific input. We have the remaining probability of $1-\frac{1}{n^2}$ that the two parties performed one of the n^2-1 procedures D that leads to the right answer $f(x,y)$. For reasonably large n, this will occur with high probability. As it is impossible for the distribution of the input values to 'anticipate' the protocol D_{ij} that A and B are going to perform, it is also impossible to force an error rate higher than the $\frac{1}{n^2}$ that holds for this probabilistic solution.

A few more words about the random bits that the parties use before we give our final characterization of probabilistic communication protocols. The randomization that we saw in the last example was *shared randomness:* both parties could agree on the random numbers (i,j) without having to communicate this specification to each other. This is also called the 'public coin model' of communication and can be viewed as the situation where Alice and Bob share an unlimited amount of classically correlated states

$$\rho_{\text{public coin}} \quad = \quad \tfrac{1}{2}|00\rangle\langle 00| + \tfrac{1}{2}|11\rangle\langle 11|.$$

A more restricted model of randomized communication is the one where the parties only have 'private coins'. In this model, shared randomness can only be achieved after one party has communicated some of his or her coin flips to the other participant. Hence, in this 'private coin' model we have to take into account the amount of shared randomness that the parties have to send to each other for the successful execution of their randomized procedure. In this thesis, we will always assume the 'public coin' model for reasons that will be explained in Section 4.10.

Another issue is that of the moment of randomization in the protocol. If we know beforehand the outcomes of all the random coin flips that will occur during a protocol, then we can again view the procedure as a deterministic one. And because there is nothing during the protocol that can influence the outcome of a coin flip, we might as well observe all of them before we start with the procedure P. As long as we establish our randomization *after* we received the x and y values, the input distribution cannot 'anticipate' any weaknesses specific for the outcomes of the coin flips.

It is for the above reasons that we can assume, without loss of generality, that the probabilistic protocol P is executed according to the three steps previously shown:

1. The parties receive their respective inputs.
2. With the help of public coin flips, Alice and Bob agree on a random number i (according to some fixed distribution π).

3. The deterministic protocol D_i is executed.

This can be summarized by the statement that the probabilistic procedure P is a mixture of deterministic protocols, $P = \{(\pi_i, D_i)\}_i$, and that on the input strings (x,y), it outputs the corresponding probabilistic bit $P(x,y) = \sum_i \pi_i \cdot D_i(x,y)$.

The quest for the optimal probabilistic protocol for a function f can be approached in two ways:

Given a desired success rate $1-\varepsilon$*, how many bits of communication are required?*

Or alternatively:

What is the minimum error rate ε *that can be obtained with* m *bits of communication?*

Although both kinds of questions will be asked in the coming chapters, we will here investigate problems of the second kind. The next section will show us how we can employ some standard techniques from game theory for our analysis of communication complexity.

Section 4.6 An Application of Von Neumann's Min-Max Theorem

Consider again the setting that we described before we ended the previous section. Alice and Bob want to calculate the distributed function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ with only m bits of communication between them. Our task is to optimize their randomized protocol in that we want the highest possible success ratio $1-\varepsilon$ for this limited amount of communication m. What do we do if every deterministic protocol (with $\varepsilon = 0$) requires more than m bits of communication?

As we now know, every probabilistic protocol P can be expressed as a mixture $\{(\pi_i, D_i)\}_i$, where π is the defining probability distribution over the set of *all* the deterministic protocols D_i. (Typically, we will have $\pi_i = 0$ for many unreliable protocols D_i.) Our second relevant distribution is the function μ that defines the probability $\mu(x,y)$ that Alice and Bob receive the input pair (x,y). Given the distributions π (defining the protocol P) and μ we calculate the error rate by

$$\begin{aligned}\varepsilon_{\pi|\mu} &= \sum_{x,y} \mu(x,y) \cdot \mathrm{Prob}(P(x,y) \neq f(x,y)) \\ &= \sum_{x,y}\sum_{i} \mu(x,y) \cdot \pi_i \cdot \delta(f(x,y), D_i(x,y)),\end{aligned}$$

where δ is the 'difference function' with $\delta(i,j) = 0$ if $i = j$ and $\delta(i,j) = 1$ otherwise. The worst case assumption for the input distribution can now be expressed as the maximization of the error over all μ's:

$$\varepsilon_\pi = \max_\mu \varepsilon_{\pi|\mu}.$$

But *we,* in the meantime, are looking for the protocol P that *minimizes* the error, and hence we are trying to reach the minimum of this maximum error, which is

$$\min_\pi \varepsilon_\pi = \min_\pi \max_\mu \varepsilon_{\pi|\mu} = \min_\pi \max_\mu \sum_{x,y,i} \mu(x,y) \cdot \pi_i \cdot \delta(f(x,y), D_i(x,y)),$$

where f is the problem in question, and all D_i are deterministic protocols. This is the situation from the viewpoint of Alice and Bob.

Nature, with her worst case behavior, on the other hand, aims for the highest possible error for each protocol P. This corresponds to an expression that is almost identical to the previous one,

$$\max_{\mu} \varepsilon_{|\mu} \quad = \quad \max_{\mu} \min_{\pi} \varepsilon_{\pi|\mu},$$

except for the changed order of the two optimizations.

This puts us in the situation where we have to analyze two conflicting strategies: the minimization of ε by Alice and Bob, and the maximization of the same term by Nature. It is therefore legitimate to ask if the whole setting is properly defined; for it is not inconceivable that the chain of arguments "Alice and Bob use protocol P", "Nature reacts by using the worst possible distribution μ", "knowing this specific μ, the two parties change to a better P'", "Nature counters with a new μ'", et cetera, has no well-defined end. Luckily, Andrew Yao's usage of Von Neumann's Min-Max theorem assures us that this is not the case, as there is a fixed point (π, μ) for this problem. (See the article [68] and the references [47, 49, 50, 51] for an introduction to game theory.)

It turns out that the above setting is an example of a 'zero-sum, two player game'. Let the first player be the duo Alice & Bob with their protocol P, and the opposing one Nature with her strategy μ. We can rephrase the conflicting goals for the participants by stating that Nature tries to maximize the error ε, whereas Alice and Bob try to maximize the negated term "$-\varepsilon$". This indeed makes it a zero sum game (for two players), and hence Von Neumann's celebrated Min-Max theorem applies. This result states that in the just-described setting there is always a fixed point (π, μ) (a *solution* of the game), such that the corresponding ε (the unique *value* of the game) solves the equation

$$\max_{\mu} \min_{\pi} \varepsilon_{\pi|\mu} \quad = \quad \min_{\pi} \max_{\mu} \varepsilon_{\pi|\mu}.$$

As this defines a *saddle point*, both parties know that any variation in their strategy (π for A and B, and μ for Nature) is not going to do them any good because, for the solution of the game, it holds that

> *For any alternative protocol π', we obtain an error that cannot be smaller.*
>
> *All other input distributions μ' give an error that cannot be bigger.*
>
> *In short, $\varepsilon_{\pi'|\mu} \geq \varepsilon_{\pi|\mu}$ and $\varepsilon_{\pi|\mu'} \leq \varepsilon_{\pi|\mu}$, for all π' and μ'.*

(Note that it is very well possible that several solutions exist for a game, but all of them will share the same value ε.) The above two characteristics in combination with the decomposition of a probabilistic protocol as a mixture of deterministic ones, also gives us a technique to easily 'recognize' the fixed-point solution of a setting. This will be explained in the next section.

Section 4.7 Proving Probabilistic Bounds

We now know that for every probabilistic communication setting, there is a solution in the form of an ideal protocol P, a worst case input distribution μ and the resulting error rate ε. But how do we determine such a solution? As stated earlier in this thesis, we are mainly interested in the setting where we try to minimize the error under a restriction on the amount of communication. In this setting, a relevant probabilistic protocol P will have to be a mixture $\{(\pi_i, D_i)\}_i$ of deterministic procedures D_i that all obey this limitation on the information transfer between Alice and Bob. Hence, if we can exhibit a specific input distribution μ and an error rate E for which every allowed *deterministic* protocol D_i obeys

$$\varepsilon_{D_i|\mu} \quad \geq \quad E,$$

then we can immediately conclude that any *probabilistic* protocol P will obey this bound E as well. And because ε is the maximum of $\varepsilon_{|\mu}$ over *all* distributions μ, we now also know that E is a lower bound on the value ε. In mathematical terms, this reasoning can be summarized by

$$\varepsilon \quad \geq \quad \varepsilon_{|\mu} \quad = \quad \min_{\pi} \varepsilon_{\pi|\mu} \quad = \quad \min_{\pi} \sum_i \pi_i \cdot \varepsilon_{D_i|\mu} \quad \geq \quad E.$$

Conversely, if we can prove for the same E the existence of a protocol P that obeys $\varepsilon_{\pi|\mu} \leq E$ for every input distribution μ and hence $\varepsilon_\pi \leq E$, then we can conclude that $\varepsilon = \min_\pi \varepsilon_\pi \leq E$. (The proof of this "$\varepsilon \leq E$" is usually done by showing that for every pair (x, y), the error probability of the protocol P is limited by $\mathrm{Prob}(P(x, y) \neq f(x, y)) \leq E$.)

The combination of the two bounds shows that indeed $\varepsilon = E$, and the solution for this value is obtained by the distributions π and μ that we used in the proof. In practice, it will almost always be the case that we simply suggest a distribution μ for which it is easy to verify that every *deterministic* protocol has an error rate of at least ε. After that lower bound, we then continue by describing a probabilistic protocol with $\mathrm{Prob}(P(x, y) \neq f(x, y)) \leq \varepsilon$ for every input pair (x, y). This is sufficient to prove that (π, μ) is a solution of the communication game with value ε. (It is typical for the worst-case distribution μ that it will be zero for any couple (x, y) for which the protocol performs above average.)

Some readers might find it unsatisfying that we just 'state' π and μ without giving a method for deriving such solutions. Such a derivation is possible because the Min-Max expression for ε is a linear equation, which can be solved in a straightforward way. But this is only possible if we are willing to deal with excessive amounts of data and variables. In this thesis, we try to avoid such an approach as it only gives a solution, but not much insight and understanding. Instead, the reader is invited to honestly try out all the possible deterministic protocols when this is suggested, and to see for him or herself that the error rate is indeed the minimum ε as stated. This dirty 'work by hand' is likely to give some insight in both the kind of instances (x, y) that are 'problematic' for Alice and Bob, and the set of optimal deterministic protocols that are used in the probabilistic mixture P. (This, at least, was my personal experience when I obtained the results.)

Section 4.8 Relations and Problems with a Promise

We can extend the setting of distributed decision problems of the form $f : X \times Y \to \{0,1\}$ to the broader notion of *relations.* Problems of this kind are described by a subset $R \subseteq X \times Y \times Z$, where Alice's input is an element $x \in X$ and Bob a $y \in Y$. The task for the two parties is to determine a value $z \in Z$ such that $(x,y,z) \in R$ with the minimum amount of communication required. A function is a relation for which every combination x and y has a uniquely defined $z = f(x,y)$ such that $(x,y,f(x,y)) \in R$, and a decision problem is a function with $Z = \{0,1\}$. (Note that for a general relation it is possible that there exist values of x and y with no corresponding z such that $(x,y,z) \in R$. In this case, the input combination (x,y) is *illegal* as there is no correct answer z to the problem.)

In the next chapters, we will use relations to describe the so-called *promise problems.* These are distributed functions $f : X \times Y \to \{0,1\}$ for which we are only interested in the protocol's behavior on a *subset* S of the possible inputs $X \times Y$. The promise is therefore that Alice and Bob only receive x and y such that $(x,y) \in S$. The standard way of describing such a promise problem is to express it as a relation $R_{f|S}$ which is a conventional decision problem f on the proper inputs, but a trivial relation on the inputs that lie outside S. Hence,

$$(x,y,z) \in R_{f|S} \quad \text{if and only if} \quad \begin{cases} (x,y) \in S \text{ and } z = f(x,y), \text{ or} \\ (x,y) \notin S, \end{cases}$$

which shows that, provided that the communication protocol has a well-defined outcome, every $P(x,y)$ for the improper inputs $(x,y) \notin S$ will be a correct outcome with $(x,y,P(x,y)) \in R_{f|S}$. The protocol therefore has a 100% success rate on those inputs, and hence it follows from the worst-case assumption that Alice and Bob do not have to expect such trivial cases. This is equivalent to the original setting where the distribution $\mu(x,y)$ is only non-zero for proper values of $(x,y) \in S$, with the difference, however, that we still require the protocol P to behave properly on *all* possible input values.

Section 4.9 Quantum Communication with Entanglement

It may come as a small surprise that we will not devote a separate chapter to the definition of the model for quantum communication, but after the preceding section and chapters this turns out to be unnecessary. In Section 4.5 we saw that the crucial ingredient for probabilistic communication protocols is the randomness that the parties can share. This randomness can be described as a 'public coin-flip' with the density matrix:

$$\rho_{\text{coin}} \quad = \quad \tfrac{1}{2}|00\rangle\langle 00| + \tfrac{1}{2}|11\rangle\langle 11|.$$

Besides the supply of these shared random bits, everything else is identical to the setting of deterministic protocols.

In short, we could say that the same holds for the difference between quantum and classical communication, with the exception that we replace the classical correlations of the state ρ_{coin} with the nonlocal correlations of entangled qubits of the form

$$\rho_{\text{ent}} \quad = \quad \tfrac{1}{2}|00\rangle\langle 00| + \tfrac{1}{2}|00\rangle\langle 11| + \tfrac{1}{2}|11\rangle\langle 00| + \tfrac{1}{2}|11\rangle\langle 11|.$$

In order to enable Alice and Bob to process their qubits, we also have to expand their set of local transformations with the unitary operations; our main interest however—the communication—is still done with classical bits. The four different stages of a quantum communication protocol are thus:

1. Alice and Bob share a sufficient amount of public coins ρ_{coin} and entangled qubits ρ_{ent}. They also agree on the protocol Q that they are going to use.
2. The parties receive their input values x and y with probability $\mu(x, y)$, where μ is the worst case distribution for the protocol Q.
3. The protocol Q is executed by A and B according to their inputs x and y.
4. Alice announces the (probabilistic) outcome $Q(x, y)$.

Obviously, we could replace the coin flips in step 2 by more entangled qubits as they behave exactly like ρ_{coin} when measured in the standard basis. Nevertheless, we will refrain from this, as it would advocate the usage of 'quantum resources' where it is not necessary.

The amount of classical bits of communication during the third phase of the protocol Q does not only depend on the pair (x, y), but also on the randomized measurement outcomes on the probabilistic states (classical or quantum) during the protocol. It is customary to define the overall communication complexity of a protocol Q as the highest possible amount of communicated bits between Alice and Bob, where the possibilities are over the input instance (x, y) *and* the randomized variables during step 3. (See Chapter 3 of [42] and references therein for a discussion on this 'worst-case versus average-case' complexity of randomized protocols.) The *communication complexity* of a distributed function f is the minimum complexity over all protocols Q that solve f.

Typical for the approach to communication complexity in this thesis is that we do not take into account the amount of correlated states that the parties need to use to perform the protocol (the number of distributed ρ states required in the first step). In the next section we will briefly look at some alternative models and discuss their relation to our 'entanglement model of quantum communication'. With this discussion I will also explain why the results in this thesis are not phrased in such 'qubit models'.

Section 4.10 Other Quantum Communication Models

Quantum communication is often understood as the transfer of information with qubits instead of classical bits. This approach has indeed been taken by the very first researchers of the field (Andrew Yao[69] and Ilan Kremer[41]). They considered the 'qubit model' where Alice and Bob are not allowed to share entanglement during step 2, but where the communication is done by quantum bits. A third possibility is the natural combination of the two models that deals with protocols where both initial entanglement and communication with qubits are allowed.

With the teleportation procedure that transfers one qubit at the cost of two classical bits and an entangled pair, the following reduction should be evident.

> *Any protocol that uses k qubits of communication and m entangled pairs can be perfectly simulated by a qubit protocol that uses $k + m$ qubits, or an entanglement protocol with $k+m$ entangled pairs and $2k$ classical bits of communication.*

From this it follows that the entanglement model will differ with at most a factor of two from the other models. What is not clear, however, is how to simulate the entanglement model with the qubit model within a constant factor. This is because we do not know any bound on the amount of entanglement that might be required for a k qubit communication protocol. If we could prove a theorem that states something along the lines of "any protocol that uses k qubits of communication, can be implemented with an a-priori entanglement of γk 2-qubit pairs", then it would be clear that with $(\gamma + 1)k$ qubits of communication in the qubit model, the same procedure could be executed. Currently, such a theorem, or a counterexample to it, is still lacking. We are thus faced with the distinct choice of analyzing quantum communication with or without prior entanglement. And in the entanglement case we have to make the additional decision if we allow communication with quantum bits or or if we restrict ourself to classical information transfer. In this thesis the latter option is chosen, and here I will briefly explain why.

If we want to compare the complexity of classical and quantum protocols, then we first have to agree on the measure that we use for our comparison. It is only with classical bits that we can express the complexity for both models in the same units. Otherwise, it is very tempting to 'explain' all of the differences between quantum and classical communication with a reference to the uncountable continuum of different quantum bits compared to the two possibilities for a classical bit. The entanglement model makes it immediately clear that something more subtle and interesting is going on. Another advantage of this entanglement based approach is that it allows us to study the relationship between nonlocality and communication complexity *without an explicit reference to the theory of quantum physics.* This will be done in Chapter 9 on 'Superstrong Correlations'. There is also the additional complication with qubit communication if more than two parties are involved: the no-cloning theorem[67] makes it impossible to send an unknown qubit to more than one party at the same time.

Section 4.11 Multiparty Communication Complexity

It is natural to generalize the two-party model to the setting where three or more participants are involved. As we will use this 'multiparty scenario' for several of our results, we discuss here the few choices that have to be made for such an extension.

For k parties $A_1, \ldots, A_k$ with their respective inputs $x_1, \ldots, x_k$, we look at the protocols that try to evaluate decision problems $f : X_1 \times \cdots \times X_k \to \{0, 1\}$ on Alice's side (who is labeled A_1). Almost all characteristics of the analysis for the two-party scenario translate directly to this setting. It is about the initial ‘information sharing’ and the communication by the participants that we have to make some additional choices. From now on we will always assume that

1. Initially, every party A_i only knows its value x_i.
2. A bit that is sent by one party becomes known to all the others at the cost of one bit of communication.

Because of the second characteristic, this setting is also known as the ‘broadcast model’. (A well-known alternative for this is the ‘number on the forehead model’ in which a participant A_i initially knows every value x_j *except* his own input x_i.)

Section 4.12 Assumptions throughout this Thesis

We have now come to the end of the introductory chapters of this thesis. But before we continue with the actual analysis of quantum communication complexity, I will summarize below the assumptions that are understood to hold for the following chapters.

- The communication complexity is measured in classical bits.
- The amount of initial entanglement or classical correlations is ignored in the complexity analysis.
- The final stage of a proper protocol is reached when the first party, Alice, knows the function value f (with high probability).
- The complexity is determined by the worst case behavior of both the input distribution μ and the randomized measurement outcomes during the execution of the protocol.
- The protocols for ‘promise problems’ have to be well defined for illegal inputs.
- In the multiparty setting, the ‘broadcast model’ is assumed, where initially each party A_i only knows his or her input value x_i and where every communicated bit becomes known to all parties.
- Measurements are always done in the standard basis $\{|0\rangle, |1\rangle\}$.
- The names of the first three parties are Alice, Bob and Carol.

Despite this listing, we will reiterate our assumptions when it is especially relevant for a better understanding of our results. Before investigating the differences between quantum and classical communication complexity, we conclude with a very brief historical overview.

Section 4.13 History and References

Two-party communication complexity as described here was introduced by Andrew Yao in the influential 1979 paper “Some complexity questions related to distributive computing”[68]. The recent book “Communication Complexity” by Eyal Kushilevitz and Noam Nisan[42] gives a thorough and up-to-date overview of this field of research.

Quantum communication complexity was first mentioned by, again, Yao in the article “Quantum Circuit Complexity”[69], where he used communication complexity as a method to derive lower bounds for quantum computation. Already in 1995, Ilan Kremer, a student of Noam Nisan, wrote an entire Master’s thesis about quantum communication complexity. Both Kremer and Yao use the qubit communication model where the parties communicate with qubits rather than with classical bits. Neither authors showed an improvement over the classical scenario. Richard Cleve was the first to establish such a separation (of one bit) in a three-party setting in 1997. This result, in the entanglement communication setting, was published together with Harry Buhrman in the article “Substituting Quantum Entanglement for Communication”.[23]

Shortly after this unexpected observation, several other results appeared. Most notable are the generalization of the original protocol of Cleve to the k-party setting, thereby obtaining an unbounded difference of $k \log k$ versus k bits between classical and quantum communication[19]; the proof that the quantum communication complexity of the INNER PRODUCT function cannot substantially be improved with the use of entanglement[24]; and—most recently—the square-root and even exponential separations in the two-party scenario[1, 17, 57].

Chapter 5

Two Simple Quantum Communication Protocols

This chapter describes two quantum protocols that have a communication complexity that is lower than is possible with classical means. The results were published in the article "Quantum Entanglement and Communication Complexity" by Harry Buhrman, Richard Cleve and myself.[16] Although both protocols are straightforward applications of the nonlocality arguments of Chapter 3, they do make an important step from the correlated outcomes of a set of entangled qubits to a protocol that can be used to perform reliable distributed computation.

Section 5.1 Reducing Errors with Nonlocality

Consider the two-party function with input size 2:

$$f(x_1, x_2, y_1, y_2) \quad = \quad x_1 \oplus y_1 \oplus (x_2 \cdot y_2).$$

The table of this function f looks like

$f(x,y)$	00	01	10	11
00	0	0	1	1
01	0	1	1	0
10	1	1	0	0
11	1	0	0	1

,

where the columns are indexed by x_1x_2 and the rows by y_1y_2.

Alice wants to estimate the value $f(x,y)$ with the highest possible correctness probability, under the restriction that only one bit of communication is allowed. We will show here that the classical bound on the success rate of 0.75 can be improved to approximately 0.85 if we allow Alice and Bob to use an entangled pair of qubits. The quantum protocol that establishes this probability is closely related to Bell's inequality and is implemented by the following procedure.

1. Before Alice and Bob receive their inputs x and y, they share the entangled qubit pair $\Phi_{AB} = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$.
2. Bob performs the rotation $R(\frac{-\pi}{16})$ on Φ_B if he has $y_2 = 0$; otherwise, he applies $R(\frac{3\pi}{16})$ to Φ_B.
3. After this rotation, Bob measures his entangled qubit in the standard basis, yielding an outcome $b \in \{0, 1\}$.
4. The same procedure holds for Alice. First, if $x_2 = 0$, apply $R(\frac{-\pi}{16})$; if $x_2 = 1$ then apply $R(\frac{3\pi}{16})$ to Φ_B. Then, measure the rotated qubit in the standard basis. The bit that is the result of Alice's observation is labeled a.
5. The one bit of communication from Bob to Alice is the parity bit $(y_1 \oplus b)$.
6. Having received the bit from Bob, Alice now finishes the protocol by estimating the value $f(x, y)$ with the binary value $x_1 \oplus a \oplus (y_1 \oplus b)$.

From the analysis of Bell's inequality in Section 3.3, we know that the probability for $a \oplus b$ to equal $x_2 \cdot y_2$ is approximately 85.3%, and hence that the correctness ratio of the above protocol will be the same:

$$\text{Prob}(x_1 \oplus y_1 \oplus a \oplus b = f(x_1, x_2, y_1, y_2)) \quad = \quad 0.853\ldots$$

(Note that this success rate is independent of the distributed x and y values.)

Such a correctness ratio cannot be obtained by a classical protocol as we will see now.

Section 5.2 Limitations of Classical Protocols

Section 4.5 taught us that a probabilistic protocol can be viewed as a mixture of deterministic protocols. Hence, it is sufficient to prove a lower bound on the error for all deterministic procedures in order to prove the same bound for randomized protocols. We will do this here for the uniform input distribution $\mu(x, y) = \frac{1}{16}$.

For a deterministic protocol, we can define the set B_0 that contains the values y_1y_2 for which Bob broadcasts the bit value "0" to Alice, and similarly the set B_1 for the communicated value "1". (With $B_0 \cup B_1 = \{0, 1\}^2$ and $B_0 \cap B_1 = \{\}$.) One of these two sets will at least contain two strings and without loss of generality we assume this set to be B_0. By inspecting the function table, we can conclude immediately that if Alice receives a zero from Bob, she will make a mistake at least 25% of the time. (If, for example, $B_0 = \{00, 01\}$, then Alice can only guess f's value if she has $x_2 = 1$.) Furthermore, if B_0 contains three strings, then at least one out of three announcements by Alice is wrong if she receives "0" from Bob. (As an example for $B_0 = \{00, 01, 10\}$: A's most successful guess will be $(x_1 \oplus x_2)$, which fails $\frac{1}{3}$ of the time. The degenerate case where $|B_0| = 4$ does not convey any information to Alice who therefore will have to make a blind guess with 50% probability of success.)

This shows that for the three possible partitions by B_0 and B_1, the error probability will be at least $\frac{1}{4}$:

- $(|B_0|,|B_1|) = (2,2)$: In both cases B_0 and B_1, Alice will make an error with 0.25 probability.
- $(|B_0|,|B_1|) = (3,1)$: The case B_0 occurs $\frac{3}{4}$ of the time with error rate $\frac{1}{3}$, hence the overall expected probability of error is $\frac{3}{4}\cdot\frac{1}{3} = \frac{1}{4}$.
- $(|B_0|,|B_1|) = (4,0)$: Alice makes a mistake 50% of the time.

We have thus established the result that any deterministic, one bit protocol will have a success ratio of at most $\frac{3}{4}$ (for the uniform distribution over x and y). From our previous result about probabilistic procedures (see Section 4.7), this implies that any classical protocol is bounded by this value. Hence we can indeed conclude that Alice and Bob have an advantage of 85% versus 75% if they use a pair of entangled qubits.

Section 5.3 An Exact Three-Party Quantum Protocol

The previous quantum protocol decreased the error probability by the use of entanglement but still left us with a ‘flawed’ procedure. Here we will define a three-party problem that has an errorless quantum protocol. This will be in sharp contrast to the classical setting where Alice, Bob and Carol have an error probability of at least 25%. The example of this section was the first published quantum communication protocol (see Cleve *et al.* in [23] and [16]), and was inspired by Mermin’s clarification[46] of the nonlocality proof by Greenberger, Horne and Zeilinger[30].

The three parties A, B and C receive their input x, y and $z \in \{0,1,2,3\}$, with the promise that the sum $x+y+z$ is an even number. The task for Alice is to decide after two bits of communication whether $x+y+z \bmod 4 = 0$ or $x+y+z \bmod 4 = 2$. As with the nonlocality proof of Section 3.4, A, B and C initially share the three qubit state $|\Psi_{ABC}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$. Next, depending on their input, each party performs a phase rotation on his or her entangled qubit. Alice rotates Ψ_A with $\text{P}\textsc{hase}(\frac{x\pi}{2})$, Bob applies $\text{P}\textsc{hase}(\frac{y\pi}{2})$, and Carol acts with $\text{P}\textsc{hase}(\frac{z\pi}{2})$ on her Ψ_C. After these three independent actions, the joint state has become

$$\begin{aligned}|\text{Phased }\Psi_{ABC}(x,y,z)\rangle &= \text{P}\textsc{hase}\left(\tfrac{x\pi}{2}\right)\otimes\text{P}\textsc{hase}\left(\tfrac{y\pi}{2}\right)\otimes\text{P}\textsc{hase}\left(\tfrac{z\pi}{2}\right)|\Psi_{ABC}\rangle\\ &= \tfrac{1}{\sqrt{2}}\left(|000\rangle + \mathrm{e}^{\frac{1}{2}(x+y+z)\pi\mathrm{i}}|111\rangle\right),\end{aligned}$$

and hence by the promise on the input values (ignoring normalization from now on):

$$|\text{Phased }\Psi_{ABC}(x,y,z)\rangle = \begin{cases}|000\rangle + |111\rangle & \text{if } x+y+z = 0 \bmod 4\\ |000\rangle - |111\rangle & \text{if } x+y+z = 2 \bmod 4\end{cases}.$$

Before measuring the bit values of Ψ, all parties rotate their individual qubits with a Hadamard transform (the global transformation $H\otimes H\otimes H$), resulting in the final state

$$|\text{Final}(x,y,z)\rangle = \begin{cases}|000\rangle + |011\rangle + |101\rangle + |110\rangle & \text{if } x+y+z = 0 \bmod 4,\\ |001\rangle + |010\rangle + |100\rangle + |111\rangle & \text{if } x+y+z = 2 \bmod 4\end{cases}$$

for the two cases. This shows that after the standard measurement (yielding the outcomes a, b and c), the decision problem is answered by the parity bit $(a\oplus b\oplus c)$ as this will be “0” if $x+y+z=0 \bmod 4$ and “1” otherwise. It is therefore sufficient for Bob and Carol to broadcast their outcomes b and c to Alice who then is able to announce the final answer with 100% reliability.

Section 5.4 Impossibility of a Two Bit Classical Protocol

Again we will use the approach where we analyze the error rate of all deterministic, two bit protocols to obtain a bound for all possible probabilistic procedures. (We assume a uniform distribution over all 32 input cases.) Although there is a promise on the values x, y and z, the inputs for two parties (ignoring the third) can be any of the 16 combinations of $\{0,1,2,3\} \times \{0,1,2,3\}$. From this observation it follows that Alice has to receive information from both Bob and Carol (and hence one bit from each) and that the protocol has to specify beforehand if B or C broadcasts the first bit. Without loss of generality we assume that this will be Bob.

The set B_0 (B_1) corresponds to the values of y for which Bob broadcasts a “zero” (“one”) to Carol and Alice. Carol can make her procedure dependent on the bit b that she receives from Bob, and hence she will use the sets C_0^0 and C_1^0 if $b=0$, and the sets C_0^1 and C_1^1 otherwise. After Carol announces her bit c to the other parties, Alice can infer that the two input values of her partners obey $(y,z) \in B_b \times C_c^b$.

The cases where $|B_0| = 3$ or 4 ($|B_0| = 0$ or 1) imply immediately that there are two values y_1 and y_2 together in B_0 (B_1) such that $y_1 + 2 = y_2$. Hence, for one of those values the deterministic protocol will always make a mistake in its calculation of $x+y+z \bmod 4$. Under the uniform distribution this error will happen 25% of the time. The protocol where B_0 or $B_1 = \{0,2\}$ suffers from the same deficit as the above scenarios, and hence has an equivalent error rate. This leaves us with the case where the sets of Bob are of the form $B_b = \{\beta_b, \beta_b + 1\}$ (with $\beta_b \in \{0,1,2,3\}$ and addition is modulo 4). The line of reasoning that we applied to the sets B also holds for the sets of Carol; hence, we can also assume that every C set is of this form $C_c^b = \{\gamma_c^b, \gamma_c^b + 1\}$ with $\gamma_c^b \in \{0,1,2,3\}$. Therefore, for every combination of b and c, the input values of Bob and Carol obey $(y,z) \in \{\beta_b, \beta_b+1\} \times \{\gamma_c^b, \gamma_c^b+1\}$. Hence there is a critical ambiguity for Alice between the cases $y+z = \beta_b + \gamma_c^b \bmod 4$ and $y+z = \beta_b + \gamma_c^b + 2 \bmod 4$. As we are dealing with deterministic protocols, one of these cases will always be resolved incorrectly, thus again giving a 25% error rate.

The upper bound of 75% correctness was derived for all possible deterministic protocols, and hence any probabilistic protocol will have to obey the same bound. This concludes the proof that the “mod four” problem cannot be solved classically with an error rate of less than 0.25. The quantum protocol obtained a significant improvement over this with its errorless protocol.

We can obtain a classical procedure without error if we allow Bob to communicate one bit more, for he can then broadcast the exact value y to Alice and Carol. This, in combination with the most significant bit of z (that is: $C_0 = \{0,1\}$ and $C_1 = \{2,3\}$), is sufficient for Alice to determine the outcome $x+y+z \bmod 4$ (using the promise on the input values). The result of this section can therefore also be restated as “The three-party entanglement allowed Alice, Bob and Carol to save one bit of communication”.

The next chapter will generalize this three-party problem to a distributed function for k parties. This will give us an unbounded difference in terms of communication complexity between the quantum and the classical settings.

Chapter 6

Multiparty Quantum Communication Complexity

The three-party problem of the previous chapter can be generalized to a k party distributed computation. In this chapter we will show how this leads to a quantum protocol that gives a reliable answer after only $k-1$ bits of communication, whereas in the classical setting the parties would require of the order of $k \log k$ bits. This result was published in the October 1999 issue of Physical Review A under the title "Multiparty Quantum Communication Complexity", with co-authors Harry Buhrman, Peter Høyer and Alain Tapp.[19]

Section 6.1 Multiparty Problem and Its Quantum Complexity

In the k party scenario, every participant A_i receives a real number $x_i \in [0,2)$, with the joint promise that the total sum $\sum_i x_i$ is a natural number. The task for Alice (A_1) is to determine whether this sum is an even or an odd number: the calculation of the function value

$$\text{Odd?}(x_1,\dots,x_k) \quad = \quad \sum_{i=1}^{k} x_i \bmod 2.$$

Our first result will be that if the parties share a k qubit entangled state, this problem can be solved with a protocol where each party communicates only one bit to Alice. This is shown by the following procedure which therefore has a total communication complexity of $k-1$ bits.

Section 6.2 The Quantum Protocol

Let the parties $A_1,\dots,A_k$ initially share the entangled state

$$|\Psi_{1:k}\rangle \quad = \quad \tfrac{1}{\sqrt{2}}\left(|0\cdots 0\rangle + |1\cdots 1\rangle\right),$$

where each party A_i possesses the qubit Ψ_i. Depending on their input value x_i, a phase rotation $\textsc{Phase}(\pi x_i)$ is applied to this qubit Ψ_i. Just as for the three-party scenario, this leads to essentially two different possibilities (ignoring normalization)

$$\begin{aligned}|\text{Phased }\Psi_{1:k}\rangle &= \textsc{Phase}(x_1\pi)\otimes\cdots\otimes\textsc{Phase}(x_k\pi)|\Psi_{1:k}\rangle\\ &= |0\cdots0\rangle + \mathrm{e}^{\mathrm{i}\pi\sum_i x_i}|1\cdots1\rangle\\ &= \begin{cases}|0\cdots0\rangle+|1\cdots1\rangle & \text{if } \sum_i x_i \text{ is even,}\\ |0\cdots0\rangle-|1\cdots1\rangle & \text{if } \sum_i x_i \text{ is odd.}\end{cases}\end{aligned}$$

Hence, locally applying the Hadamard transform on all k sites results in a superposition of bit strings with an even (odd) Hamming weight, if the summation $\sum_i x_i$ equals an even (odd) number. The continuation of the protocol should now be obvious: every party measures his or her qubit Ψ_i in the standard basis and (except Alice) broadcasts the outcome a_i to A_1. She concludes by calculating the total parity $\sum_i a_i \bmod 2$ of the measurement outcomes, which directly gives the correct answer to the original question: "Is $\sum_i x_i$ even or odd?" A protocol with less than $k-1$ bits of communication is impossible as this implies the exclusion of one of the parties, which cannot give an errorless procedure.

The core of this chapter is the proof that in the classical case, $\Theta(k\log k)$ bits of communication are required. This is far from trivial, as we will see in the next sections: a case-by-case analysis (as we did in the three-party case) is no longer possible.

Section 6.3 Bounds for Promise Functions

In the introductory chapter, "Communication Complexity", we discussed how a deterministic communication protocol corresponds to a tree where the leaves label the final decisions, and the depth of the tree equals the communication complexity of the protocol. What follows is a general method for obtaining lower bounds on deterministic protocols that we will use in the next section for our "even versus odd" problem.

First we remind ourselves that a protocol has to be well defined for all inputs $x_1,\ldots,x_k$, even if they do not obey the promise of the function $f :\subseteq X^n \to \{0,1\}$. This means that for every instance $x_1,\ldots,x_k$, the protocol has to follow one uniquely determined path from the root to a 'decision leaf'. Because our tree is binary, this implies that with depth d, the number of leaves is at most 2^d. Hence, if we have a minimum of t leaves, then this corresponds to a communication complexity of at least "depth -1" $= \log(t)-1$ bits. This gives us a good reason to look at the number of leaves that is required for a deterministic and exact protocol.

For different input combinations, a protocol can end up in the same decision leaf. Assume now that we know of two such combinations $(x_1,\cdots,x_k)$ and $(\tilde{x}_1,\cdots,\tilde{x}_k)$ that both lead to the same 'path down the decision tree' to a leaf L. For a given protocol, the direction that the protocol takes at a specific branch is solely determined by the input of the then speaking party A_i. Note now that by our assumption, these directions are the same for both x_i and $\tilde{x}_i$. This means that we can combine the x and $\tilde{x}$ values and still take the same path down to L. In other words, every element of the Cartesian product set $\{x_1, \tilde{x}_1\} \times \cdots \times \{x_k, \tilde{x}_k\}$ gives the same path down the tree. We will therefore call the subset R of inputs that lead to the same leaf a *rectangle*, as they will always be of the form $R = X_1 \times \cdots \times X_k$, with $X_i \subseteq X$ for every i.

If we want our deterministic protocol to be exact, then the elements of a specific rectangle R that are allowed by the promise must all have the same function value f. We can assume that every leaf contains at least one allowed input for it would otherwise be possible to 'trim down' the decision tree—by removing this vacuous leaf—without changing the effectiveness of the protocol. These two characteristics are summarized by saying that the rectangles have to be *monochromatic*.

We thus have collected some facts about the leaves of a decision tree that we can use in the following standard method for proving bounds on deterministic communication protocols. (See also Section 1.2 in [42].)

Assume a (promise) k-party decision problem $f :\subseteq X^k \rightarrow \{0,1\}$ where there is a total of $|X|^k$ possible input states (both proper and improper in the case of promise problems).

1. *Show that for the function f the maximum volume for a monochromatic rectangle is bounded by some limit v.*
2. *Realize that this bound tells us that there have to be at least $\frac{|X^k|}{v}$ 'rectangles-as-leaves' in the decision tree.*
3. *Conclude that a deterministic protocol will therefore need at least $k \log |X| - \log v - 1$ bits of communication if it wants to be errorless.*

This is the approach that we will now apply to our "Odd?" function.

Section 6.4 The Lower Bound for the Classical Protocol

Before we start our lower bound proof, we will rephrase and restrict the Odd?-problem to the 'finite and integer input' version that we already encountered in Section 5.3. This is the situation where we are only interested in values of x up to n bits and where we (implicitly) multiply every number with 2^{n-1} such that every x_i is viewed as an element of $\mathbb{Z}_{2^n}$. We therefore have the new promise that $\sum_i x_i = 0 \bmod 2^{n-1}$, for which the parties try to determine whether

$$\sum_{i=1}^{k} x_i \quad = \quad \begin{cases} 0 \bmod 2^n & \text{("even"), or} \\ 2^{n-1} \bmod 2^n & \text{("odd")?} \end{cases}$$

It should be clear that the original Odd?-function is *at least* as difficult to solve as this problem: as n gets bigger the problem becomes harder and only the limiting case $n \rightarrow \infty$ corresponds to the continuous version with $x \in [0,2)$ rather than $x \in \{\frac{q}{2^{n-1}} | q \in \mathbb{Z}_{2^n}\}$. It turns out however, that for our current purposes the lower bound $n \geq \log k$ suffices (in combination with $k \geq 2$). We will now prove the $\Theta(k \log k)$ lower bound for this restricted problem.

Let R be a k-dimensional rectangle that corresponds to a leaf in the decision tree, hence $R = R_1 \times \cdots \times R_k \subseteq \mathbb{Z}^k_{2^n}$. With modulo 2^n arithmetic, we define the sum of two sets $A, B \subseteq \mathbb{Z}_{2^n}$ as $A + B = \{a + b | a \in A, b \in B\}$. We use this set addition and the 'sides' $R_1, \ldots, R_k$ to define a sequence of $k + 1$ sets S_i by: $S_0 = \{0\}$ and $S_i = S_{i-1} + R_i$ for $i = 1, \ldots, k$. Obviously, for every combination $x_1 \cdots x_k$ in the rectangle R, the corresponding value $\sum_i x_i$ is an element of $S_k = R_1 + \cdots + R_k$. The requirement that R is monochromatic tells us that *either* $0 \in S_k$ *or* $2^{n-1} \in S_k$, but not both. This puts a bound on the size of S_k and hence on the volume of the rectangle R as we will see with the help of some group theory.

Assume without loss of generality that R is a zero-rectangle, and hence that $0 \in S_k$ and $2^{n-1} \notin S_k$. one: $H = \{0\}$. Now we are at the place where we have to use an application of *Kneser's addition theorem* for groups[39, 43], which states:

> *For every pair of subsets $A, B \subseteq \mathbb{Z}_{2^n}$ there exists a subgroup G of $\mathbb{Z}_{2^n}$ such that $A + B + G = A + B$ and $|A + B| \geq |A + G| + |B + G| - |G|$.*

We apply this theorem to the cases where $A = S_{i-1}$ and $B = R_i$ (and hence $S_i = A + B$ with $i = 1, \ldots, k$), which gives us k groups G_i that obey

$$\begin{aligned} S_k &= S_{i-1} + R_i + R_{i+1} + \cdots + R_k \\ &= S_{i-1} + R_i + G_i + R_{i+1} + \cdots + R_k \\ &= S_k + G_i. \end{aligned}$$

This readily shows that for every i the only possible G_i is the trivial subgroup $G_i = \{0\}$, for if G_i is a bigger subgroup, then $\{0, 2^{n-1}\} \subseteq G_i \subseteq S_k$ and S_k would not correspond to a monochromatic rectangle (remember that $0 \in S_k$). By the second part of Kneser's theorem we therefore know that $|S_i| \geq |S_{i-1}| + |R_i| - 1$ for all i, which sums up to

$$|S_k| \quad \geq \quad 1 - k + \sum_{i=1}^{k} |R_i|.$$

The exclusion of the element 2^{n-1} for the set S_k gives us $|S_k| \leq 2^n - 1$, leading to

$$\sum_{i=1}^{k} |R_i| \quad \leq \quad 2^n - 2 + k$$

for the sides of the rectangle R, and hence for its maximum volume v

$$|R| \; = \; \prod_{i=1}^{k} |R_i| \; \leq \; \left(\frac{2^n - 2 + k}{k} \right)^k .$$

The last equation shows us an upper bound v on a monochromatic rectangle for an exact k-party solution to the Odd?-problem (with input size n). The decision tree of this protocol has to cover a total of 2^{nk} inputs and hence needs at least $\frac{2^{nk}}{v}$ leaves. This gives the tree a minimum depth of $k \log(2^n) - \log v > k \log k - k$, by the assumptions $n \geq \log k$ and $k \geq 2$.

The method of the previous section now concludes our proof that every classical and exact k party protocol for the Odd?-function requires at least $k \log k - k - 1 \in \Omega(k \log k)$ bits of communication.

Section 6.5 Conclusion

We just saw a difference between classical and quantum communication of the order of $k \log k$ for k parties. This assures us that the quantum protocol is more efficient *even if we want to take into account the distribution of the entangled qubits* during the initial stage of the protocol. If the cost of transmitting 1 qubit equals c classical bits, then the quantum complexity still remains linear in k. For large enough k (*i.e.* $\log k > c + 3$) the resources that the classical procedure requires are still more than the costs $(c + 1)k$ in the quantum scenario.

Chapter 7

Quantum Whispers

Probabilities in quantum mechanics depend quadratically on the amplitudes of a state. This phenomenon lies at the heart of the so called 'quantum Zeno effect'. In this chapter, we will use this effect to establish a difference between classical and quantum communication in the probabilistic model. We will show how the error probability of the one-bit quantum protocol will always be lower than that of the classical approach. This result is extended to the setting where each party, in a sequence of many parties, is only allowed to communicate one bit to its neighbor (akin to the game of 'Chinese Whispering'). It is argued that for this multiparty problem, there is no reliable classical solution, whereas with entanglement it can be solved easily. The results of this chapter were published in the article "Quantum communication using a nonlocal Zeno effect" by Lucien Hardy and myself.[34]

Section 7.1 Introduction

In this chapter, we will employ the 'quantum Zeno effect' to construct a new quantum communication procedure. This effect relies on the quadratic, rather than linear, relation between the amplitudes and the probabilities of a quantum state. Several authors have used this phenomenon to obtain proofs of nonlocality that are different from the ones that we encountered in the previous chapters of this thesis.[15, 33, 62]

First, we will define a two-party problem Jump?(x, y) for which prior entanglement enables us to significantly reduce the error over the classical scenario. After that, a multiparty version, "Quantum Whispers", of the same problem is introduced. For this task we exhibit a reliable quantum protocol and argue that without entanglement no such solution exists.

Section 7.2 The Two-Party Problem

The "jump or no jump?" question that we will discuss in this chapter is a typical example of a promise problem. In the two-party setting Alice and Bob receive their

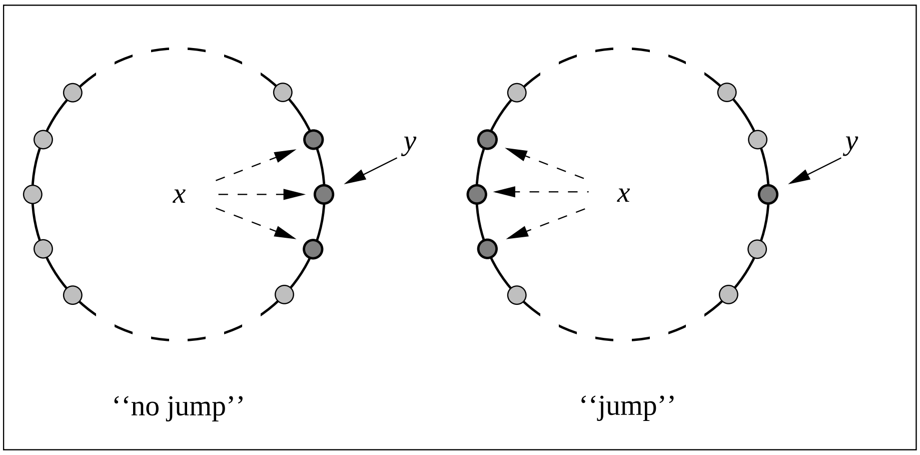

Figure 7.1: *Explanation of the communication problem. "No jump" means that x equals y or is adjacent to it; "jump" refers to the situation where x is opposite y or adjacent to that opposite position.*

input x and $y \in \{0, 1, \ldots, 2N-1\}$ under the promise that $x - y \in \{-1, 0, +1\} \bmod N$. The natural number N is a free parameter of the problem but has to be thought of as 'big'. Alice is allowed to communicate one bit to Bob, who then has to try to answer if the pair x, y makes a jump of size N or not:

$$\text{Jump?}(x,y) \quad = \quad \begin{cases} \text{"no jump"} & \text{if } x-y \in \{-1,0,+1\}, \\ \text{"jump"} & \text{if } x-y \in \{N-1, N, N+1\}. \end{cases}$$

(The subtraction is performed modulo $2N$.) See Figure 7.1 for a clarification of this setting.

For both the classical and the quantum solution of this problem, it is inevitable that Bob sometimes makes a mistake. It will be the difference in the minimum error rate that separates quantum from classical communication.

Section 7.3 The Quantum Protocol

Alice and Bob start in the setting where they share an entangled qubit pair $|\Phi_{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Depending on their input values they will perform a rotation R to their local qubits. Alice applies $R(\frac{x\pi}{2N})$ to her Φ_A, and Bob rotates his Φ_B with $R(\frac{y\pi}{2N})$. As a result of these two actions, the joint state is now of the form

$$\begin{aligned}R\left(\tfrac{x\pi}{2N}\right)\otimes R\left(\tfrac{y\pi}{2N}\right)|\Phi_{AB}\rangle &= \tfrac{1}{\sqrt{2}}\cdot\cos\left(\tfrac{(x-y)\pi}{2N}\right)|00\rangle +\\ &\quad -\tfrac{1}{\sqrt{2}}\cdot\sin\left(\tfrac{(x-y)\pi}{2N}\right)|01\rangle +\\ &\quad \tfrac{1}{\sqrt{2}}\sin\left(\tfrac{(x-y)\pi}{2N}\right)|10\rangle +\\ &\quad \tfrac{1}{\sqrt{2}}\cos\left(\tfrac{(x-y)\pi}{2N}\right)|11\rangle.\end{aligned}$$

After the rotations, when both parties measure their qubit in the standard basis, the parity of the two outcomes a and b is strongly correlated with the difference between x and y:

$$\begin{aligned}\mathrm{Prob}(a\oplus b=0|xy) &= \cos^2\left(\tfrac{(x-y)\pi}{2N}\right),\\ \mathrm{Prob}(a\oplus b=1|xy) &= \sin^2\left(\tfrac{(x-y)\pi}{2N}\right).\end{aligned}$$

By the periodicity of the trigonometric functions, the parity $(a\oplus b)$ will give a reliable indication of the "jump versus no jump" question. In the case where x and y do not make a jump, we have

$$\cos^2\left(\tfrac{(x-y)\pi}{2N}\right) \in \left\{1,\tfrac{1}{2}+\tfrac{1}{2}\cos\left(\tfrac{\pi}{N}\right)\right\},$$

and hence

$$\begin{aligned}\mathrm{Prob}(a\oplus b=0|\text{no jump between } x \text{ and } y) &\geq 1-\left(\tfrac{\pi}{2N}\right)^2,\\ \mathrm{Prob}(a\oplus b=1|\text{no jump between } x \text{ and } y) &< \left(\tfrac{\pi}{2N}\right)^2.\end{aligned}$$

Conversely, in the case where there is a jump between x and y we have

$$\sin^2\left(\tfrac{(x-y)\pi}{2N}\right) \in \left\{1,\tfrac{1}{2}+\tfrac{1}{2}\cos\left(\tfrac{\pi}{N}\right)\right\},$$

and consequently

$$\begin{aligned}\mathrm{Prob}(a\oplus b=1|\text{jump between } x \text{ and } y) &\geq 1-\left(\tfrac{\pi}{2N}\right)^2,\\ \mathrm{Prob}(a\oplus b=0|\text{jump between } x \text{ and } y) &< \left(\tfrac{\pi}{2N}\right)^2.\end{aligned}$$

The above analysis suggests the protocol where Alice sends her measurement outcome a to Bob, who then calculates the parity $(a\oplus b)$. By doing so, the error rate of Bob's estimation will be bounded by the quadratic expression

$$\varepsilon_{\text{quantum}} < \tfrac{\pi^2}{4N^2}.$$

Next, we will show that any *classical* protocol will have a error probability of the order $\Omega(\frac{1}{N})$.

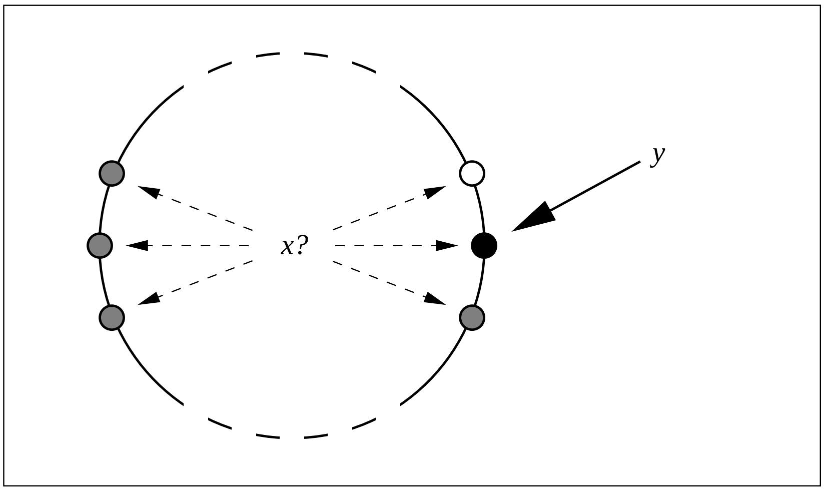

Figure 7.2: *Given that Bob knows his y, there are six possibilities for Alice's value x. Here we have chosen y such that for some allowed values x, Alice will send "one" (black) and for some other values "zero" (white).*

Section 7.4 Best Possible Classical Protocol

As in all of the previous proofs, we start by assuming that Alice and Bob use a deterministic protocol. Furthermore, we use the uniform distribution over the allowed input values. This is where each combination $(x,y) \in \{(x,x-1),(x,x),(x,x+1),(x,x+N-1),(x,x+N),(x,x+N+1)\}$ has equal probability $\frac{1}{12N}$. (Alice's input x ranges from 0 to $2N-1$.)

Let S_0 (S_1) be the set of x values for which Alice communicates a "zero" ("one") to Bob. If one of the sets S is empty, Alice does not convey any information to Bob, who then has to make a blind guess on the "jump versus no jump" question, leading to an error rate of 50%. Hence we will assume that both S_0 and S_1 are not empty. This implies that sometimes Bob will have a value y such that $(y-1) \in S_0$ and $y \in S_1$. (The probability that this happens will be calculated below.) Figure 7.2 shows us how Bob sometimes receives a "zero" from Alice (when $x = y-1$) in this scenario, and sometimes a "one" ($x = y$). We know that $(y-1) \in S_0$ and $y \in S_1$, but for the other four possible x values it is still unspecified to which set S they belong (hence we have a total of $2^4 = 16$ different cases). We continue with our assumptions by letting the three "jump" values $(y+N-1)$, $(y+N)$ and $(y+N+1)$ be members of S_0, and for the the remaining "no jump" input: $(y+1) \in S_1$. What should Bob conclude if he receives a "zero" from Alice? Figure 7.3 shows us that in 3 out of 4 cases, Alice's x value corresponds to a "jump", whereas only the $x = y-1$ case is a "no jump". Hence, Bob optimizes his answer by saying that there was indeed a jump between x and y. By the assumption of a uniform distribution over the values x, this means that Bob will be incorrect 25% of the time if he receives a "zero" from Alice in combination with this

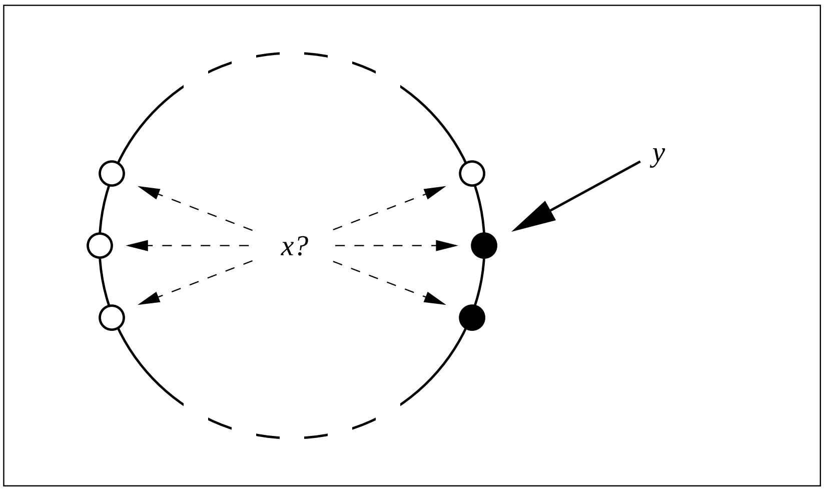

Figure 7.3: *A specific coloring of the four grey dots in Figure 7.2. For this example, the analysis in the text shows that Alice and Bob have an error probability of $\frac{1}{6}$ when trying to decide if x and y make a jump or not.*

y. If Bob receives a “one” from Alice, then he knows that either $x = y$ or $x = y + 1$, and hence he can state that there was no jump without the risk of making a mistake. The above shows that the total probability of error for the setting of Figure 7.3 is $\frac{1}{6}$ (the case $(x, y) = (y-1, y)$). It is straightforward to go through the other 15 possible cases of Figure 7.2, and to conclude that the minimum error probability is indeed $\frac{1}{6}$.

For any nontrivial protocol with $|S_0|$, $|S_1| \geq 1$ the scenario of Figure 7.2 will occur for at least 2 of the possible $2N$ values of y. Accordingly, the minimum error-rate for a deterministic protocol is bounded by

$$\varepsilon_{\mathrm{deterministic}} \quad \geq \quad \tfrac{2}{2N} \cdot \tfrac{1}{6}.$$

We can thus conclude that any classical, probabilistic procedure for deciding the jump/ no jump question will be incorrect at least $\frac{1}{6N}$ of the times, which is obviously linear in $\frac{1}{N}$.

The error rate of the quantum protocol is limited by

$$\varepsilon_{\mathrm{quantum}} \quad \leq \quad \tfrac{\pi^2}{4N^2},$$

whereas for classical protocol it always holds that

$$\varepsilon_{\mathrm{classical}} \quad \geq \quad \tfrac{1}{6N}.$$

We have thus proven that (for big enough N) the quantum protocol will always be more reliable than any possible classical procedure. In the next section, we will try to amplify the difference between the quantum and classical settings by looking at a multiparty scenario.

Section 7.5 Multiparty Communication Problem

Imagine the above two-party problem as the first step of a multiparty problem. In this scenario we consider a concatenation of jump/no jump questions for a line-up of M parties that are only allowed to communicate one bit to their next neighbor. This setting is reminiscent of the game "Chinese Whispers", where a message passes through a row of whispering people. Hence the name "Quantum Whispers".

Formally the problem is defined as follows. Take M parties $A_1, \ldots A_M$, each have two input values (y_i, x_i) except for the first party who only receives an x_1, and the last person A_M with his y_M. The promise on the input is the same that we used for the jump versus no jump problem:

$$y_{i+1} - x_i \quad \in \quad \{-1, 0, 1, N-1, N, N+1\} \bmod 2N,$$

for all $1 \leq i < M$ (note that x_i and y_i are uncorrelated). After having received their respective inputs, each party A_i is allowed to communicate one bit a_i to his or her neighbor A_{i+1}. After these $M-1$ bits of communication, the rightmost participant A_M has to decide whether there was an even or an odd number of jumps among the $M-1$ pairs (x_i, y_{i+1}).

We are interested in the error rate in the case where M is of the same order as N (that is, $M = \gamma N$ with $\gamma \in O(1)$). The quantum solution to the quantum whispers problem will be a straightforward application of the two-party procedure for each couple (A_i, A_{i+1}). For big enough N, this will give us a reliable quantum protocol. We then continue by arguing that in the classical setting such a reliable protocol does not exist. But it should be stressed that a formal proof of this claim is still lacking.

Section 7.6 The Quantum Whispers Protocol

Before receiving their input values, each pair (A_i, A_{i+1}) shares an entangled state $|\Phi^i_{i+1}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. (This notation as to distinguish the qubit Φ_i that is entangled with the system of A_{i-1} from Φ^i that is correlated with the qubit Φ_{i+1} of A_{i+1}.)

The first two parties start by performing the same rotations, on their respective qubits Φ^1 and Φ_2, as was described in the section on the two-party setting.

The standard measurement on both sides yields the classical bits a_1 and b_2 (of which a_1 will be sent to A_2 by the first party). From earlier investigations, we know that the parity $(a_1 \oplus b_2)$ indicates with $1 - \frac{\pi^2}{4N^2}$ certainty whether or not there is a jump between x_1 and y_2.

These rotations and measurements are done by all $M-1$ couples (A_i, A_{i+1}) on their qubits Φ^i and Φ_{i+1} (the rotations depending on the respective input values x_i and y_{i+1}). The estimations of the jump/no jump answers are the classical parity bits $(a_i \oplus b_{i+1})$, and hence the question "is there an even or odd number of jumps?" can be estimated by the total parity $\sum_{i=1}^{M-1}(a_i \oplus b_{i+1})$. Therefore, after A_1 has sent her a_1 to her neighbor, A_2 continues by communicating the bit $(a_1 \oplus b_2 \oplus a_2)$ to the third person in line. This procedure of sending the parity of the received bit and the two outcomes

a_i and b_i to A_{i+1} is repeated until A_M is reached. The bit that this rightmost party receives from A_{M-1} corresponds to the previously mentioned sum $\sum_i (a_i \oplus b_{i+1})$, minus his own outcome b_M. Hence, by taking the parity of the received bit and b_M, he will effectively estimate the odd/even jumps questions by $M-1$ repetitions of the two-party protocol. This allows an easy analysis of the expected error, as we will see below.

Section 7.7 Error of the Quantum Whispers Protocol

In the description of the quantum whispers approach to solving the odd/even jumps problem, it was mentioned how it is equivalent with $M-1$ applications of the two-party solution. We know that the error rate of this procedure is bounded from above by $\varepsilon_{\mathrm{q}} < \frac{\pi^2}{4N^2}$. Hence, the probability of success $1-\varepsilon_{\mathrm{q}}^M$ of the multiparty protocol is the sum of the probabilities corresponding to an even number of mistakes among the $M-1$ guesses. The lower bound on this probability is calculated by

$$1-\varepsilon_{\mathrm{q}}^M \quad > \quad \sum_{j=0}^{2j<M} \begin{pmatrix} M-1 \\ 2j \end{pmatrix} \left(\frac{\pi^2}{4N^2}\right)^{2j} \left(1-\frac{\pi^2}{4N^2}\right)^{M-1-2j}.$$

For some constant $\gamma = \frac{M}{N}$ and a large enough N, this can be approximated by only considering the $j=0$ case; it hence has a success probability of at least

$$1-\varepsilon_{\mathrm{q}}^M \;>\; \left(1-\frac{\pi^2}{4N^2}\right)^{M-1} \;\approx\; 1-\frac{\pi^2\gamma}{4N}.$$

This shows that the $M = \gamma N$ party version of the quantum whispers protocol has a maximum error rate of the order $O(\frac{1}{N})$. In the case of large N, this yields a reliable quantum protocol.

Section 7.8 Possible Classical Whispers

Here we will argue that there does not exist a reliable classical protocol for the even/odd jumps problem (again for large N and $\gamma = \frac{M}{N} \approx 10$).

We start by noting that the "jump" versus "no jump" problem is independent for each (x_i, y_{i+1}) pair because there is no correlation between x_i and y_i. The parity of the number of jumps relies critically on all the $M-1$ answers to the two-party problem which are independent of each other. This strongly suggests that it is necessary to solve the jump/no jump problem for all pairs (x_i, y_{i+1}). Because we only allow a single bit of communication between the parties A_i and A_{i+1}, this will always induce an error-rate of at least $\frac{1}{6N}$ *per pair.* (See the result of Section 7.4.) With the number of pairs $M-1 = \gamma N - 1$, the probability of an odd number of such errors (leading to an incorrect estimation of the parity) is bounded from below by

$$\begin{aligned}
\varepsilon_{\mathrm{c}}^{M} &\geq \sum_{j=0}^{2j+1<M} \left(\begin{array}{c} M-1 \\ 2j+1 \end{array} \right) \left(\frac{1}{6N}\right)^{2j+1} \left(1-\frac{1}{6N}\right)^{M-2j-2} \\
&= \frac{1}{2}-\frac{1}{2}\left(1-\frac{1}{3N}\right)^{M-1} \\
&\approx \frac{1}{2}-\frac{\mathrm{e}^{-\frac{\gamma}{3}}}{2} \quad \text{(assuming big } N\text{).}
\end{aligned}$$

Under the assumptions of $\gamma \approx 10$, the error probability $\varepsilon_{\mathrm{c}}^{M}$ is effectively 50%. This tells us that the answer of A_M is as successful as the outcome of a random coin toss.

The above argument for the unreliability of classical protocols is not water-tight, as we have not formally excluded every possible protocol. Such a proof turned out to be more difficult than expected (as is often the case for lower bound proofs[29]). This difficulty is due to the great number of potential M-bit protocols that the M parties can use. Consequently, although we have proven that there exists a reliable quantum protocol, we cannot be 100% certain that there is no classical procedure that is reliable, however unlikely this may be.

Chapter 8

Lower Bounds for Quantum Communication

In the previous chapters, we have only seen examples where entanglement reduced the complexity of communication protocols. Here we give our first lower bound for quantum communication. We will show that for the inner product problem, entanglement does not significantly improve the performance over the classical case. The quantum and classical scenario are nevertheless not equivalent. This difference is indicated by a small reduction of error in the $n = 2$ case of the inner product problem with only one bit of communication between the parties. See [24] for the original article by Richard Cleve, Alain Tapp, Michael Nielsen and myself.

Section 8.1 Introduction

How can we prove lower bounds for quantum communication? Obviously, some of the methods we use for classical communication fail; otherwise there would be no difference between the two. Hence, we have to look for 'impossibilities in quantum mechanics' and try to apply them in the setting of distributed computation. The problem is that, at the moment, we do not have many examples of tasks that are impossible in a nonlocal world. Instead, our path to understanding entanglement is paved with unexpected phenomena that do not exist in classical information theory. One of the few 'no-go's' in quantum communication is Holevo's bound, which tells us that quantum bits are not any better than classical ones for transferring information between parties. This bound, in combination with a result by Michael Nielsen (see Appendix A), will indeed be one of the cornerstones for this chapter's lower bound.

The crucial step in the proof that will follow is to go from the problem that only focuses on the single bit of the function $f(x, y)$ to the transfer of the whole bit string y. This is done by considering an f that is familiar both in communication complexity and in quantum computation: the inner product function.

Section 8.2 The Inner Product Problem

For two parties A and B with input strings $(x_1,\ldots,x_n)$ and $(y_1,\ldots,y_n)$, the inner product function $\mathrm{IP}:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}$ is defined by

$$\mathrm{IP}(x,y) \quad = \quad \sum_{i=1}^{n} x_i\cdot y_i \bmod 2.$$

This function is well studied as a classical communication problem. (See [20, 64] and Section 3.5 in [42].) It is easy to see that for a deterministic protocol, n bits of communication are required. This intractability of IP does not disappear in the probabilistic scenario. If Alice and Bob want to know $\mathrm{IP}(x,y)$ with a maximum error of ε, then $n-O(\log(\frac{1}{1-2\varepsilon}))$ bits of communication are still necessary. Hence for a fixed error rate, the communication complexity is $n-O(1)$. Below we will see that both for the deterministic and probabilistic cases, the situation does not much improve if the two parties are allowed to use prior entanglement. That is, the inner product function is also hard in the setting of quantum communication.

In quantum computation, the IP function is inextricably linked with the Hadamard transform. This was already mentioned in the section on quantum information, but we will briefly repeat it here. The one qubit Hadamard transform H is defined by

$$H|0\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle+|1\rangle) \quad \text{and} \quad H|1\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle-|1\rangle),$$

which also shows that H is its own inverse: $H^2=I$.

When we apply a Hadamard to each individual bit of the string $y=y_1\cdots y_n$, we obtain a superposition of all possible n bit strings x where the information about y is stored in the phases as the inner product between y and the different x's:

$$|y_1,\ldots,y_n\rangle \quad \longleftarrow H^{\otimes n} \longrightarrow \quad \tfrac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}(-1)^{\mathrm{IP}(x,y)}|x_1,\ldots,x_n\rangle. \tag{8.1}$$

This suggests the following protocol to extract n bits of information $y_1\cdots y_n$ with the help of the inner product function $\mathrm{IP}(\cdot,y)$. We start with the uniform superposition of all strings $x\in\{0,1\}^n$ and an empty output register:

$$\tfrac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}}|x\rangle|0\rangle.$$

Calculating the function values $\mathrm{IP}(x,y)$ in superposition then yields

$$\tfrac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}}|x\rangle|\mathrm{IP}(x,y)\rangle.$$

We now apply a conditional phase flip (CFLIP) on the output register,

$$\tfrac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}}(-1)^{\mathrm{IP}(x,y)}|x\rangle|\mathrm{IP}(x,y)\rangle,$$

after which we inverse the first computation of the function value, thereby giving the final state

$$\frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}} (-1)^{\mathrm{IP}(x,y)}|x\rangle|0\rangle.$$

Equation 8.1 shows us that this last state gives us all the n bits of the string y.

We thus see how the quantum calculation of the inner product for a superposition of x strings results in a protocol where we obtain n bits of information. We will use this phenomenon in combination with the known quantum limitations on information transfer to prove our lower bounds on the quantum communication complexity of the IP function.

Section 8.3 Informal Sketch of Proof

Before we write down the formal proof of the quantum lower bounds, we will first give an overview of the ideas that lie behind it. Assume that Alice and Bob can compute the inner product value $\mathrm{IP}(x,y)$ of their n bit strings with k classical bits of communication. Instead of classical inputs, let Alice use the superposition of all possible strings x while Bob fixes his y. The two parties then continue by executing the protocol 'in quantum parallel' that now uses k quantum bits of communication. Having finished this procedure, Alice now has a superposition of strings x entangled with the 2^n function values $\mathrm{IP}(x,y)$. As we saw in the previous section, this information can easily be converted into a description of Bob's n bits $y_1\cdots y_n$. Hence Alice has—with k qubits of communication—obtained n bits of information. The bounds on quantum information transfer tell us that this is only possible if $k > \frac{n}{2}$, which is tight by the method of Superdense coding (see Section 2.3).

For the probabilistic bound, we use the same argument by saying: " $\ldots$ the 2^n function values *approximating* $\mathrm{IP}(x,y)$ $\ldots$ into a description *strongly correlated* with y $\ldots$ only possible if k is of the order of n."

Our main concern for the formal proof will be the conversion of the protocol for classical inputs into its quantum superposition variant. This reduction provides a new method of analysis that can also be used for other, future lower bounds in quantum communication. The formal justification of it will be given in the next section.

Section 8.4 Quantum Parallelizing Communication Protocols

Let Alice and Bob share some initial entanglement Ψ_{AB} before they receive their inputs x and y. During the execution of the communication protocol, the two parties will most likely need some additional 'working space' to perform the appropriate computations, after which Alice writes down the outcome $f(x,y)$. Without loss of generality, we assume these ancillas to be initially set to zero, and hence the whole system starts in the state

$$|\text{begin}(x,y)\rangle \quad = \quad |x\rangle|0\rangle|0\cdots0\rangle|\Psi_{AB}\rangle|0\cdots0\rangle|y\rangle.$$

After the protocol is finished and Alice knows the outcome $f(x,y)$, both working registers have probably changed depending on the initial values x and y (as has the entangled state Ψ); we therefore write:

$$|\text{semi-final}(x,y)\rangle \quad = \quad |x\rangle|f(x,y)\rangle|\text{garbage}(x,y)\rangle|y\rangle.$$

Alice can now flip the phase of her state depending on the outcome $f(x,y)$. After that, we can perform the whole protocol in 'reverse', thereby removing the garbage and the outcome as well as restoring the initial entanglement Ψ, but maintaining the function depended phase $(-1)^{f(x,y)}$. This gives us the clean, final state

$$|\text{final}(x,y)\rangle \quad = \quad (-1)^{f(x,y)}|x\rangle|0\rangle|0\cdots0\rangle|\Psi_{AB}\rangle|0\cdots0\rangle|y\rangle.$$

We thus see that any k qubit communication protocol can be transformed in a clean, no garbage-producing, unitary procedure that requires $2k$ qubits of communication. By doing so, we enable the application of the protocol in superposition for different values of x, by which we mean the following.

Assume that instead of one string x, Alice has a uniform superposition input strings:

$$\tfrac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}|\text{begin}(x,y)\rangle \quad = \quad \tfrac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}|x\rangle|0\rangle|0\cdots0\rangle|\Psi_{AB}\rangle|0\cdots0\rangle|y\rangle.$$

For this situation, Alice and Bob can perform the same clean protocol described above. The communication of this 'quantum parallelized procedure' is done with the same number of qubits as in the original schema. Hence, the end result is a $2k$ qubit communication protocol with the final state:

$$\tfrac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}|\text{final}(x,y)\rangle \quad = \quad \tfrac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}(-1)^{f(x,y)}|\text{begin}(x,y)\rangle,$$

where the information about the function values $f(x,y)$ is now stored in the phases on Alice's side. This also shows why we required the initial protocol to be clean: it enabled us to create a superposition of strings $(-1)^{f(x,y)}|x\rangle$ without suffering from any entanglement with Bob's part.

Section 8.5 Bounds on Exact Inner Product Protocols

In the previous section, we saw how we can clean up a communication procedure while only doubling its complexity. We use this result in combination with the assumption that we have an errorless procedure for the calculation of IP : $\{0,1\}^n\times\{0,1\}^n\to\{0,1\}$, which requires k qubits of communication between Alice and Bob. This gives us a $2k$ qubit protocol that establishes the evolution (ignoring the workspaces and Bob's part as they remain unchanged for clean procedures)

$$|x\rangle \quad \longrightarrow_y \quad (-1)^{\mathrm{IP}(x,y)}|x\rangle,$$

for all possible $x,y\in\{0,1\}^n$.

If we apply the same $2k$ qubit procedure to the initial state where the x register is the uniform superposition, then the phases of Alice's final state will contain all the information about y:

$$\tfrac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}|x\rangle \quad \longrightarrow_y \quad \tfrac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}(-1)^{\mathrm{IP}(x,y)}|x\rangle.$$

From this it is straightforward for Alice to recover the bits $y_1\cdots y_n$ by using the Hadamard transform of Equation 8.1 on her qubits. The end result is thus that Bob has communicated n bits of information to Alice. This puts a lower bound on the value of k, as we will see now.

Because of the inverse part of the above clean protocol, exactly k of the $2k$ communicated qubits go from Bob to Alice. Hence, by the result proven in the appendix, k has to be bigger than $\lceil\frac{n}{2}\rceil$. Translating this back to our original (dirty) procedure, we have proven that any errorless, quantum protocol needs at least $\lceil\frac{n}{2}\rceil$ qubits of communication for the evaluation of the inner product function $\mathrm{IP}:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}$. This result is also tight as Bob can use superdense coding to send his n bits of information to Alice.

We will postpone the analysis of protocols with classical bits of communication, and return to it after we have looked at the quantum bounds for probabilistic protocols.

Section 8.6 Bounds on Probabilistic Inner Product Protocols

Assume that Alice and Bob have a dirty k qubit communication protocol P that computes the IP function correctly with probability of at least $1-\varepsilon$. We will again look at what effect the clean protocol $P^\dagger\cdot\textsc{CFlip}\cdot P$ has on the superposition

$$|\text{begin}(x,y)\rangle \quad = \quad |x\rangle|0\rangle|0\cdots0\rangle|\Psi_{AB}\rangle|0\cdots0\rangle|y\rangle.$$

More specifically, we will put a lower bound on the fidelity between the probabilistic outcome and that of the ideal protocol of the previous section; this will enable us to calculate the probability that Alice obtains the bit string y after the $2k$ bits of communication.

We can always assume that for every x and y, the effect of P can be written as

$$\begin{aligned} P|\text{begin}(x,y)\rangle \quad = \quad & \sqrt{1-\beta^2}|x\rangle|\mathrm{IP}(x,y)\rangle|\text{garbage}(x,y)\rangle|y\rangle \\ & +\ \beta|x\rangle|\bar{\mathrm{IP}}(x,y)\rangle|\text{garbage}'(x,y)\rangle|y\rangle, \end{aligned}$$

with the real valued β limited by $\beta^2\le\varepsilon$. By setting $P|\text{begin}(x,y)\rangle=|x\rangle|G(x,y)\rangle|y\rangle$, we can rewrite the situation after the CFLIP as

$$(-1)^{\mathrm{IP}(x,y)}|x\rangle|G(x,y)\rangle|y\rangle+2\beta|x\rangle|G'(x,y)\rangle|y\rangle$$

with the 'erroneous part' $|G'(x,y)\rangle = -(-1)^{\mathrm{IP}(x,y)}|\bar{\mathrm{IP}}(x,y)\rangle|\text{garbage}'(x,y)\rangle$. Because

$$\langle x|\langle G'(x,y)|\langle y|P|\text{begin}(x,y)\rangle \quad = \quad -(-1)^{\mathrm{IP}(x,y)}\beta,$$

it follows for the inverse $P^\dagger$ that

$$P^\dagger|x\rangle|G'(x,y)\rangle|y\rangle \quad = \quad -(-1)^{\mathrm{IP}(x,y)}\beta^*|\text{begin}(x,y)\rangle + \sqrt{1-\beta^2}|\text{begin}^\perp(x,y)\rangle,$$

where $|\text{begin}^\perp(x,y)\rangle$ is orthogonal to $|\text{begin}(x,y)\rangle$. Hence, the final state of the protocol $P^\dagger \cdot \mathrm{C}\textsc{Flip} \cdot P$ equals

$$(1-2\beta^2)(-1)^{\mathrm{IP}(x,y)}|\text{begin}(x,y)\rangle + 2\beta\sqrt{1-\beta^2}|\text{begin}^\perp(x,y)\rangle. \tag{8.2}$$

This gives us a lower bound on the inner product between the ideal outcome $|\text{final}(x,y)\rangle$ and the result, $P^\dagger \cdot \mathrm{C}\textsc{Flip} \cdot P|\text{begin}(x,y)\rangle$, of our probabilistic procedure of

$$\langle\text{final}(x,y)|P^\dagger \cdot \mathrm{C}\textsc{Flip} \cdot P|\text{begin}(x,y)\rangle \quad = \quad 1-2\beta^2 \quad > \quad 1-2\varepsilon.$$

For different values $x' \neq x$, the inner product between $|\text{final}(x',y)\rangle$ and the state of Equation 8.2 is always zero, and hence we can also apply the above lower bound to the superposition of states:

$$\tfrac{1}{2^n}\sum_{x\in\{0,1\}^n}\langle\text{final}(x,y)|P^\dagger \cdot \mathrm{C}\textsc{Flip} \cdot P|\text{begin}(x,y)\rangle \quad > \quad 1-2\varepsilon.$$

This concludes our proof that Alice's probability of observing the string y after applying the n-fold Hadamard at the end of the protocol will be at least $(1-2\varepsilon)^2$. By Fano's inequality[25], this means that the mutual information between Bob's y and Alice's measurement outcome is at least $(1-2\varepsilon)^2n-1$ bits. By the same reasoning as was used in the errorless protocol, this is only possible if the amount of communication k is equal to or bigger than $\frac{1}{2}(1-2\varepsilon)^2n-\frac{1}{2}$. For a fixed error rate, this translates as $k \in \Omega(n)$.

Section 8.7 Communication with Classical Bits

The above analysis dealt with the scenario where A and B were allowed to communicate with quantum bits. This is a different setting from that of the previous chapters, where only classical communication was allowed. Here we will translate the bounds of the preceding section to that of our standard model. Essential for this reduction are the protocols of teleportation and superdense coding, which give a tight relation between quantum and classical information transfer.

Assume that A and B have an errorless protocol for the computation of $\mathrm{IP} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ that uses k bits of classical communication. Imagine now that each party has two n-bit strings (x and x' for Alice, y and y' for Bob) and that they perform their protocol on both (x,y) and (x',y') *in parallel*. This yields a procedure where the communication is always done in pairs of bits and after which Alice knows both $\mathrm{IP}(x,y)$ and $\mathrm{IP}(x',y')$, and hence that $\mathrm{IP}(xx',yy') = \mathrm{IP}(x,y) \oplus \mathrm{IP}(x',y')$. In other words, we have a $2k$-bit communication protocol for the computation of $\mathrm{IP} : \{0,1\}^{2n} \times \{0,1\}^{2n} \to \{0,1\}$ that allows superdense coding on the pairs of bits that are exchanged. Applying this coding method gives a procedure with k quantum bits of communication that calculates the inner product for input size $2n$. By the results of the previous section, we can now conclude that $k \geq n$, which is obviously tight.

For the probabilistic setting, we assume a k-bit protocol with error-rate ε. By the same reduction of the above paragraph, this yields a k qubit protocol with a probability of error bounded from above by $2(\varepsilon - \varepsilon^2)$ for strings $xx', yy' \in \{0,1\}^{2n}$. Again by the earlier results in this chapter, this gives us a lower bound on the number of communicated bits according to $k \geq (1-2\varepsilon)^4 n - \frac{1}{2}$, which for the limit $\varepsilon \to 0$ implies $k \geq n - O(1)$.

Section 8.8 Inner Product Problem for Two Bits

The above results on the limitations of quantum communication seem to suggest that entanglement does not help *at all* for the distributed calculation of the inner product function. This pessimism is not entirely justified as we will see in the remainder of the chapter. We will consider the IP function for input strings of size two, where we allow only one bit of communication from Bob to Alice. The latter who then has to guess the value $\mathrm{IP}(x_1x_2, y_1y_2)$ as reliably as possible. With classical communication, this success rate is bounded from above by 75%, whereas in the quantum case the probability for A to give the right answer can be almost 80%.

Section 8.9 The Classical Case for Two Bits

The function $\mathrm{IP} : \{0,1\}^2 \times \{0,1\}^2 \to \{0,1\}$ that we consider here is defined by

$$\mathrm{IP}(x_1x_2, y_1y_2) \quad = \quad x_1 \cdot y_1 \oplus x_2 \cdot y_2,$$

which has the following table

IP	00	01	10	11
00	0	0	0	0
01	0	1	0	1
10	0	0	1	1
11	0	1	1	0

.

It is clear that if Alice has $x = 00$, she will never make a mistake when inferring that $\mathrm{IP}(x,y) = 0$. This tells us—as we assume a worst case distribution over the inputs—that we have to focus on the errors that can occur when $x \neq 00$. Hence, we start our analysis with the assumption that the values x and y occur according to the probability distribution μ with

$$\mu(x,y) \quad = \quad \begin{cases} 0 & \text{if } x = 00, \\ \frac{1}{12} & \text{otherwise.} \end{cases} \tag{8.3}$$

It is also obvious that for this distribution μ, Alice has to receive some information from Bob if she wants a probability of success greater than a half (equaling a random guess). We will therefore look at the deterministic protocols D, where Bob starts by sending the bit value $b = 0$ if his y is element of some set $B_0 \subseteq \{00, 01, 10, 11\}$, and otherwise the value $b = 1$ if $y \in B_1$. (With $B_0 \cap B_1 = \{\}$ and $B_0 \cup B_1 = \{00, 01, 10, 11\}$, see the introduction in Section 4.4.) After this, Alice has to guess the value of $\text{IP}(x, y)$ on the basis of her string x and the received bit b. By simply going through all the options for B_0 and B_1 (essentially 8), in combination with the rational choose for Alice (in the light of the assumed distribution μ), we can see that the following four protocols each have the lowest possible error probability of 3 out of $12 = 0.25$.

D_1: The two B sets are $B_0 = \{00\}$ and $B_1 = \{01, 10, 11\}$. Alice's answer will simply equal this bit b.

D_2: If $B_0 = \{01\}$ and $B_1 = \{00, 10, 11\}$ then A's guess is calculated by $\text{NOT}(x_2)$.

D_3: The final answer is $\text{NOT}(x_1)$ in combination with Bob's sets $B_0 = \{10\}$ and $B_1 = \{00, 01, 11\}$.

D_4: When Bob uses the sets $B_0 = \{11\}$ and $B_1 = \{00, 01, 10\}$, Alice answers "IP $= 1$" if her $x = 11$ and "IP $= 0$" otherwise.

Moreover, if we look at the values x and y for which the above protocols make a mistake (which happens only when $b = 1$), then we see that the error probability is equally distributed over the 12 pairs (x, y) with $\mu(x, y) \neq 0$:

Protocol D_i	the 3 combinations (x, y) that D_i answers incorrectly
D_1	$(01, 10)$, $(10, 01)$ and $(11, 11)$
D_2	$(01, 11)$, $(10, 00)$ and $(11, 10)$
D_3	$(01, 00)$, $(10, 11)$ and $(11, 01)$
D_4	$(01, 01)$, $(10, 10)$ and $(11, 00)$

.

This suggests a probabilistic protocol that uses two public coin flips to choose randomly between $D_1, \ldots, D_4$. The following procedure is indeed the best that Alice and Bob can do in the classical case after having received their inputs $x, y \in \{00, 01, 10, 11\}$:

1: Randomization Alice and Bob determine at random which one of the four deterministic protocols D_1, D_2, D_3 or D_4 they are going to use.

2: Bob's communication Depending on his input y and the chosen protocol D_i, Bob sends a "zero" or a "one" to Alice. (See the above list of protocol description for the specifications).

3: Alice's answer If Alice has $x = 00$ she concludes "$\text{IP}(x, y) = 0$"; otherwise, she acts in accordance with the protocol D_i that was chosen at the first stage.

This communication protocol makes no mistake if $x = 00$ and by its randomization errs with 25% in all the other cases. Therefore, a) there exists no input distribution μ' that causes a higher error rate than $\frac{1}{4}$, and b) the distribution μ as defined in Equation 8.3 reaches the 25% bound and is hence an example of a worst case situation. In other words, 0.75 is the highest possible correctness ratio in the classical setting. The next section shows that this can be improved if we allow Alice and Bob to use prior entanglement.

Section 8.10 The Quantum Improvement

Here we will show that there is a one bit quantum protocol that that has a correctness rate of 78.9%, which improves the classical bound by almost four percent.

The two parties have prior entanglement by the standard pair $|\Phi_{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Depending on his input y, Bob first applies an operation B_y to his Φ_B. After this, a measurement is performed on the rotated qubit, yielding a classical bit b that is sent to the other party. If the receiver, Alice, has $x = 00$, she knows with certainty that $\mathrm{IP}(x, y) = 0$ without having to interact with Bob. If this is not the case ($x \neq 00$), she performs a unitary transformation A_x to her Φ_A and measures the qubit in the standard basis with outcome a. Alice's 'guess' for $\mathrm{IP}(x, y)$ will now be the parity bit $(a \oplus b)$.

With the following rotations for A and B, this protocol will have a correctness probability of $\frac{1}{2} + \frac{\sqrt{3}}{6} \approx 78.9\%$ for the worst case $x \neq 00$. (If $x = 00$, the correctness rate is 100%.) Hence, the quantum protocol is more reliable than any classical procedure for this particular problem. The A and B rotations that establish this separation are (with $\zeta = \mathrm{e}^{\mathrm{i}\frac{\pi}{6}} = \frac{1}{2}\sqrt{3} + \frac{1}{2}\sqrt{-1}$):

$$
\begin{aligned}
A_{01} &= \tfrac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{3+\sqrt{3}} & \zeta^5\sqrt{3-\sqrt{3}} \\ \zeta\sqrt{3-\sqrt{3}} & \sqrt{3+\sqrt{3}} \end{pmatrix} \\
A_{10} &= \tfrac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{3+\sqrt{3}} & -\sqrt{\sqrt{3}-3} \\ -\sqrt{\sqrt{3}-3} & \sqrt{3+\sqrt{3}} \end{pmatrix} \\
A_{11} &= \tfrac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{3+\sqrt{3}} & \zeta\sqrt{3-\sqrt{3}} \\ \zeta^5\sqrt{3-\sqrt{3}} & \sqrt{3+\sqrt{3}} \end{pmatrix},
\end{aligned}
$$

and

$$
\begin{aligned}
B_{00} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & B_{01} &= \tfrac{1}{\sqrt{3}} \begin{pmatrix} 1 & \sqrt{-2} \\ \sqrt{-2} & 1 \end{pmatrix} \\
B_{10} &= \tfrac{1}{\sqrt{3}} \begin{pmatrix} 1 & \zeta^7\sqrt{2} \\ \zeta^{11}\sqrt{2} & 1 \end{pmatrix} & B_{11} &= \tfrac{1}{\sqrt{3}} \begin{pmatrix} 1 & \zeta^{11}\sqrt{2} \\ \zeta^7\sqrt{2} & 1 \end{pmatrix}.
\end{aligned}
$$

(In case the reader is wondering about the ratio behind this protocol, the entries of the above matrices were obtained with the help of a computer program that performed a numerical search for the optimal solution of the inner product problem. The corresponding analytical expressions were inferred and verified by the author of the computer program.)

Section 8.11 Communication Complexity versus Quantum Mechanics

We have reached the stage where we are convinced that quantum mechanics *sometimes* allows a significant reduction in communication complexity, but also sometimes does not. "The influence that quantum physics has on the theory of distributed computation is significant but subtle”, should be the tantalizing conclusion for computer scientists. But, to paraphrase Dr Johnson[12], “If physics kicks computer science, should physics not be kicked back?” That is, what can we learn about nonlocality and quantum mechanics from the theory of communication complexity?

The next chapter—which will be the last one before the Conclusion of this thesis—tries to initiate such a ‘back-action’. It will be shown that the limits of nonlocality coincide with the limits of distributed computing and that both of them can be viewed as a refinement of the ‘no-signaling theorem’.

Chapter 9

Superstrong Correlations versus Communication Complexity

How close is the relation between nonlocality and communication with prior entanglement? What are the implications of the no-signaling theorem for the nonlocal behavior of physics and communication complexity? In this chapter, we will touch on these and related questions. We do this by imagining a toy-theory where the CHSH *inequality for locality is maximally violated (stronger than is possible in quantum mechanics) while still respecting the axiom of no-signaling. It will be shown that in such a scenario, the communication complexity of a distributed decision problem becomes a vacuous concept: it will always be one bit. This approach is inspired by the work of Popescu and Rohrlich who, in a series of articles, asked the question: "Why is Nature not more nonlocal than she is?"*

Section 9.1 Nonlocality Revisited

The Clauser, Horne, Shimony and Holt (CHSH) inequality for classical theories gives a bound on the strength of correlations between two separated experiments.[22] We described this nonlocality argument earlier in this thesis, so we will here directly state it in the form as we will use it in the rest of this chapter.

Imagine two separated parties A and B, each of which can perform one out of two experiments on a particle that they receive from a common source. There are therefore four experimental set-ups that can apply to the combined system: (M_0^A, M_0^B), (M_0^A, M_1^B), (M_1^A, M_0^B) and (M_1^A, M_1^B). The two possible outcomes of the measurements on each side are labeled "0" and "1", and we will call the 2-particle system Φ_{AB}. We repeat the experiment many times such that we have an accurate estimation of all the possible correlations between the different measurements and their outcomes. As it is understood that for each trial we will always use the same state-preparation of Φ_{AB}, we drop the conditional part when expressing the probabilities. For example, the probability that both Alice and Bob measure a "one" when they use the measurement settings M_0^A and M_1^B is denoted by $\mathrm{Prob}(M_0^A \cdot M_1^B = 1)$.

The main result of Bell[6] and CHSH is that *for any local, hidden variable theory about Φ and the measurements M^A and M^B, the following inequality most hold:*

$$\begin{array}{l}\mathrm{Prob}(M_0^A \oplus M_0^B = 0)\ +\ \mathrm{Prob}(M_0^A \oplus M_1^B = 0)\ + \\ \quad \mathrm{Prob}(M_1^A \oplus M_0^B = 0)\ +\ \mathrm{Prob}(M_1^A \oplus M_1^B = 1)\quad \leq \quad 3. \end{array} \tag{9.1}$$

We know by now that quantum mechanics violates this bound with

$$\sum_{x,y\in\{0,1\}} \mathrm{Prob}(M_x^A \oplus M_y^B = x \cdot y)\ =\ 2+\sqrt{2}\ \approx\ 3.14,$$

for the entangled pair of qubits $|\Phi_{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, thereby proving that the theory of quantum mechanics cannot be phrased as a local theory. This, however, is only of limited interest. More important is that thus far all experiments have confirmed the violation of the CHSH inequality (as predicted by quantum physics).[3, 27, 65] This is the more relevant side of the matter as it is not inconceivable that in the future we will replace the theory of quantum mechanics by a more accurate or more general model of Nature. But no matter its exact formulation, this succeeding model will have to agree with the experimental results that we have already obtained. And as *the empirical data* by itself rules out a local explanation, any proper future candidate theory will have to be nonlocal as well. The study of ‘nonlocality-as-such’ should, for the above reasons, extend to all theories that violate the bound of Equation 9.1, rather than only investigating the version of nonlocality that we encounter in standard quantum mechanics. In this chapter, we will indeed study nonlocal correlations that are not possible with our current theory of quantum physics.

Section 9.2 The Question of Popescu and Rohrlich

In a series of articles, Sandu Popescu and Daniel Rohrlich ask the question why Nature seems to allow a violation of the CHSH inequality with a correlation term of $2+\sqrt{2}$, but not with more.[55, 56, 58] (See, for example, the article by Boris Cirel’son for a proof that $2+\sqrt{2}$ is indeed the limit.[21]) They rhetorically ask themselves: “Could the requirement of relativistic causality restrict the violation to $[2+\sqrt{2}]$ instead of 4?”[56] Such a result would be great step towards a better understanding of Nature for “If so, then nonlocality and causality would together determine the quantum violation of the CHSH inequality, and we would be closer to a proof that they determine all of quantum mechanics.” Unfortunately, this turns out not to be the case. The authors prove this by constructing a toy-theory where the nonlocality Inequality 9.1 is surpassed by a correlation value of 4. The non-zero probabilities of this super-nonlocal theory are simply

$$\begin{cases} \mathrm{Prob}(M_x^A M_y^B = 00) = \mathrm{Prob}(M_x^A M_y^B = 11) = \frac{1}{2} & \text{if } xy \in \{00, 01, 10\}, \\ \\ \mathrm{Prob}(M_1^A M_1^B = 01) = \mathrm{Prob}(M_1^A M_1^B = 10) = \frac{1}{2} & \text{otherwise.} \end{cases} \tag{9.2}$$

This leads indeed to the maximum violation

$$\sum_{x,y\in\{0,1\}} \mathrm{Prob}(M_x^A \oplus M_y^B = x \cdot y) \quad = \quad 4,$$

while the randomization of the outcomes still prevents Alice or Bob from transferring information to the other party without the use of conventional communication.

So, if causality is still respected with the correlations of Equation 9.2, why does Nature not allow it? Are there any (obvious) first principles that forbid a violation stronger than that of quantum mechanics? Or, to put it more dramatically and to the point, *what is so bad about stronger-than-quantum-mechanics nonlocality?* Here I will try to provide a partial answer to this question by pointing out the far-reaching consequences of the toy-model by Popescu and Rohrlich for distributed computing. It will be shown that the maximum violation of the CHSH inequality leads to a model of Nature where the notion of communication complexity is vacuous: all decision problems can be solved deterministically with only one bit of communication.

First, we will find a general way of expressing all possible distributed functions in a standard format that coincides with the inner product problem for two parties. Then we will see how, with superstrong correlations, the IP problem (and hence all problems) can be solved with the minimal amount of one bit of communication from Bob to Alice. The concluding section of this chapter is used for a discussion of both this result and the prospects for continuing this line of investigation.

Section 9.3 Distributed Decision Problems as Inner Products

Any function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ can be expressed as a multi-variable polynomial with modulo two arithmetic (where $1+1=2\equiv 0$). This is most easily seen by the fact that elementary Boolean operations like AND, OR, NOT or 'equivalence' can be calculated with addition and multiplication over $\mathbb{Z}_2$:

$$\begin{cases} x \text{ AND } y \;\equiv\; x\cdot y, & x \text{ OR } y \;\equiv\; x+y+x\cdot y, \\ \text{NOT}(x) \;\equiv\; 1+x, & (x \Leftrightarrow y) \;\equiv\; 1+x+y. \end{cases}$$

Just as any Boolean function $f : \{0,1\}^n \to \{0,1\}$ can be constructed from those primitives, so can f also be constructed from the elementary $\text{mod}2$ operations "$+$" and "$\cdot$". The 2-bit equivalence relation EQUIV, for example, thus becomes

$$\text{EQUIV}(x_1x_2, y_1y_2) \;=\; (x_1 \Leftrightarrow y_1)\text{AND}(x_2 \Leftrightarrow y_2) \;\equiv\; (1+x_1+y_1)\cdot(1+x_2+y_2).$$

Furthermore, as long as x and y are of finite length, we can rewrite such polynomials

$$f(x_1,\dots,x_n,y_1,\dots,y_n) \quad\in\quad \mathbb{Z}_2[x_1,\dots,x_n,y_1,\dots,y_n]$$

as a finite summation of products $\sum_i P_i(x)\cdot Q_i(y)$, where P and Q are polynomials in the input strings x and y respectively. Moreover, we can restrict the Q functions to the products of the form $Q(y) = \Pi_j y_j^{c_j}$, with c one of the 2^n characteristic vectors $c\in\{0,1\}^n$. In total, there are therefore 2^n different polynomials $Q_i(y)$ that we have to consider, and hence the index i in the summation is bounded by $1\leq i\leq 2^n$. This gives us a way of representing the function f as an inner product problem of input size 2^n:

$$f(x_1\cdots x_n, y_1\cdots y_n) \equiv \sum_{c\in\{0,1\}} P_c(x)\cdot\prod_{j=1}^{n} y_j^{c_j} \tag{9.3}$$

$$\equiv \sum_{i=1}^{2^n} P_i(x_1,\ldots,x_n)\cdot Q_i(y_1,\ldots,y_n). \tag{9.4}$$

For the 2-bit EQUIV function, for example, this is shown by

$$\begin{aligned}\text{EQUIV}(x_1x_2, y_1y_2) &\equiv (1+x_1+y_1)\cdot(1+x_2+y_2)\\ &\equiv (1+x_1+x_2+x_1x_2)+(1+x_2)\cdot y_1+(1+x_1)\cdot y_2+y_1y_2\\ &\equiv \sum_{i=1}^{4} P_i(x_1,x_2)\cdot Q_i(y_1,y_2),\end{aligned}$$

with the $2^n = 4$ polynomials on each side:

i	$P_i(x_1,x_2)$	$Q_i(y_1,y_2)$
1	$1+x_1+x_2+x_1x_2$	1
2	$1+x_2$	y_1
3	$1+x_1$	y_2
4	1	y_1y_2

.

We can view this as an inner product problem because all the bit values $P_i(x)$ will be known on Alice's side without any communication from the other party. The same holds for the bits described by the polynomials Q on Bob's side. Hence, if A and B are able to compute the IP function for input sizes of 2^n with the one bit of communication, then they also are able to calculate *any* decision problem $f:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}$ with a single bit of information exchange. We will see in the next section that this indeed possible with a maximum violation of the CHSH inequality.

Section 9.4 Inner Product and Nonlocality

Assume a model of Nature where the probabilities of Equation 9.2 are applicable, and hence where the correlation

$$\text{Prob}(M_x^A \oplus M_y^B = x\cdot y) = 100\%$$

holds for all $x,y\in\{0,1\}$. In such a world, Alice and Bob (with their bits x and y) can perform two separated measurements on their super-correlated states which yield the outcomes α and β that obey $\alpha+\beta\equiv xy$. From this, it follows that in the case of the inner product function IP$_n$, Alice and Bob can perform n measurements on an equal number of super-correlated particles in order to obtain—without any communication—a collection of bit values α_i and β_i, with again $\alpha_i+\beta_i\equiv x_iy_i$ for every i. By the commutativity of addition (modulo two), this allows the regrouping of the bits to the two separated sides of the communication protocol:

$$\mathrm{IP}(x_1\cdots x_n,y_1\cdots y_n) \;\equiv\; \sum_{i=1}^{n} x_i\cdot y_i \;\equiv\; \sum_{i=1}^{n}\alpha_i+\beta_i \;\equiv\; \underbrace{\sum_{i=1}^{n}\alpha_i}_{\text{Alice's side}} + \underbrace{\sum_{i=1}^{n}\beta_i}_{\text{Bob's side}}\;.$$

Because Bob can construct and add his β_i values without requiring any information from Alice, he can therefore compute the value $b = \sum_i \beta_i \bmod 2$ by himself and broadcast this single bit to Alice. She, on her part, creates the α_i values and finishes the protocol with the errorless conclusion IP$(x,y) = b+\sum_i \alpha_i \bmod 2$.

Section 9.5

Trivial Superstrong Communication Complexity

We just saw how the IP function has a communication complexity of one bit for every finite input size in the setting of superstrong correlations. Hence, we can apply the reduction of Section 9.3 to reach the result that any distributed decision problem $f(x,y)$ can be exactly computed with a single bit of communication. Equation 9.3 tells us that we can rewrite the function f to

$$f(x_1\cdots x_n,y_1\cdots y_n) \;\equiv\; \sum_{i=1}^{2^n} P_i(x_1,\dots,x_n)\cdot Q_i(y_1,\dots,y_n).$$

As Bob can compute all the Q_i values by himself, he and Alice can also remotely and independently create the α and β values such that $\alpha_i+\beta_i \equiv P_i(x)\cdot Q_i(y)$ for all $1\leq i\leq 2^n$. After the appropriate regrouping of the sum, the previous equation then becomes

$$f(x_1\cdots x_n,y_1\cdots y_n) \;\equiv\; \underbrace{\sum_{i=1}^{2^n}\alpha_i}_{\text{Alice's side}} \;+\; \underbrace{\sum_{i=1}^{2^n}\beta_i}_{\text{Bob's side}}\;.$$

It should now be clear that Bob can compute the bit $b = \sum_i \beta_i \bmod 2$ by himself and then communicate it to Alice who, just as for the IP function, concludes with $f(x,y) = b+\sum_i \alpha_i \bmod 2$.

This finishes the proof that with the help of the superstrong correlations of Equation 9.2, any distributed function can be decided on Alice's side without error after only one bit of communication from Bob. It is true that the amount of resources (the super-correlated states) grows exponentially with the input size n but this does not effect the conclusion that the communication complexity—after the inputs are distributed—is minimal. We will finish this chapter with a short discussion about the implications of the above result.

Section 9.6 Discussion

We can now rephrase our original question as, “What is so bad for Nature about super-efficient distributed computing?” It is not clear if there is a convincing answer to this question, as it does not seem to conflict with any physical intuition. But trivial communication complexity *does* disprove the existence of an intrinsic ‘complexity’ for distributed tasks. Even if we need an exponential amount of prior superstrong entanglement (as is indeed the case in the derivation of Section 9.5), the solution of all possible distributed functions with a single bit of communication surely does contradict our experiences in computer science. Much as in computability theory, there is a hierarchy of different ‘classes’ of communication problems.[4] Such hierarchies are at the core of theoretical computer science, and their ‘collapse’—as happened here by assuming superstrong correlations—goes against the intuition of most researchers in the field of complexity theory.

Future investigations along the lines of this chapter can be aimed at obtaining other implications of stronger-than-quantum-mechanics correlations. What happens, for example, if we assume a violation of the CHSH inequality with a factor less than the four we used here, say with $3\frac{1}{2}$? This is still stronger than allowed by the the nonlocality of quantum mechanics, but the consequences for communication complexity are less clear in this scenario. Also, the possible implications for *computational* classes deserve further research. One can thus investigate how the existence of superstrong correlations would enhance the power of, say, logarithmic depth circuits. Could it be the case that this class NC$_{ssc}$ contains all of P? And even beyond that, what can be computed in polynomial time under the assumptions of this chapter? The conjecture that superstrong nonlocality would imply NP $\subseteq$ P$_{ssc}$, is certainly a tempting thought.

Chapter 10

Conclusion

In this last, chapter I will describe some of the work on communication complexity done by other authors. This review is concluded with an outline for future research. Finally, the possibilities of experiments that implement one of the protocols will be discussed. We derive some threshold values for the realistic setting of noisy entanglement and faulty detectors. These criteria for a proper quantum communication protocol refer directly to some of the loopholes that exist for the experimental verification of Nature's nonlocality.

Section 10.1 Other Work on Quantum Communication

Soon after the first publication of a separation between quantum and classical communication complexity[23], other, more spectacular, results were obtained. Besides the ones described in this thesis, the following results should be mentioned.

In 1997, Harry Buhrman, Richard Cleve and Avi Wigderson obtained an almost quadratic separation for the well-known two-party disjointness problem in the randomized setting.[17] In the same article, they also showed an exponential difference between quantum and classical communication for a deterministic, promise problem that is based the on Deutsch-Jozsa algorithm.[26] The function for the quadratic separation tells if two sets $X, Y \subseteq \{1, \ldots, n\}$ are disjoint or not. Hence,

$$\textsc{Disjoint}(x_1 \cdots x_n, y_1 \cdots y_n) \quad = \quad \prod_{i=1}^{n} \textsc{Not}(x_i \cdot y_i),$$

where x and $y \in \{0,1\}^n$ are the characteristic vectors for the sets X and Y. The authors recognized that this function is a distributed case of a database search for an index i such that $(x_i \cdot y_i) = 1$. Hence, by applying Lov Grover's search algorithm[14, 31, 32] while sending the index register of size $\log n$ back and forth $O(\sqrt{n})$ times, Alice and Bob can solve this problem with $O(\sqrt{n} \cdot \log n)$ bits of communication. As it was already known that a probabilistic solution of the $\textsc{Disjoint}$ function requires $\Theta(n)$ bits of communication[38] in the classical setting, this established a near quadratic separation.

The exponential difference is possible for the deterministic, promise problem where the function DDJ ('distributed Deutsch-Jozsa') is defined by

$$\mathrm{DDJ}(x_1 \ldots x_n, y_1 \ldots y_n) \quad = \quad \begin{cases} 0 & \text{if the string } x \oplus y \text{ is `balanced',} \\ 1 & \text{if the string } x \oplus y \text{ is `constant',} \end{cases}$$

where $x \oplus y$ is the bitwise EXCLUSIVE OR of the n bits $(x_1 \oplus y_1, \ldots, x_n \oplus y_n)$, and the promise on x and y is that this string is either balanced or constant. For the quantum solution of this problem, Bob starts by preparing the state of $\log n$ qubits $\frac{1}{\sqrt{n}} \sum_i (-1)^{y_i} |i\rangle$, which is then teleported to the other party (requiring $2 \log n$ bits of communication). Alice, on her side, changes this received state to the final superposition $\frac{1}{\sqrt{n}} \sum_i (-1)^{x_i \oplus y_i} |i\rangle$, which enables her to decide without error if $x \oplus y$ is balanced or not. The proof that of the order of n bits of communication are necessary without quantum resources is rather involved and can be found in the original article. The exponential separation by Andris Ambainis, Leonard Schulman, Amnon Ta-Shma, Umesh Vazirani and Avi Wigderson improves the previous result in that it holds for the more realistic probabilistic setting.[1] It should be mentioned, however, that the distributed 'sampling' problem of this article lies outside the standard communication model: it is a not a decision problem, but a multi-valued function instead.

The strongest separation that we currently have was established by Ran Raz in 1999.[57] The problem that is analyzed in this article is defined follows. Alice receives a unit vector $\vec{x} \in \mathbb{R}^n$ and two mutually orthogonal subspaces M_0, $M_1 \subset \mathbb{R}^n$. The input of Bob consists of a rotation $T \in \mathrm{SO}(n)$. The question that Alice has to answer now is: "Is $T(\vec{x})$ an element (within some error margin) of M_0 or of M_1?" The input for both parties consists of n^2 real variables. In the approximating variant, each variable is described by $\Theta(\log n)$ bits, which leads to a total input size of $\Theta(n^2 \log n)$. The restriction on the inputs is the promise that $T(\vec{x})$ will lie in *either* M_0 *or* M_1, but not in both. Using teleportation, it is reasonably straightforward to design a quantum protocol with complexity $O(\log n)$ for this problem. The significance of Raz's work lies in the $\Omega(\sqrt{n})$ lower bound he obtains for the probabilistic, classical procedures.

Section 10.2 Open Problems and Future Research

An important open problem in the theory of quantum communication is the potential difference between the qubit and the entanglement model of communication (*cf.* Section 4.10). We know that every qubit of communication can be simulated with an entangled pair and two classical bits of communication, but what about the inverse of this simulation? Is it always the case that n bits of communication in combination with a potentially unbounded amount of entanglement can be converted into a protocol that uses $O(n)$ qubits of communication but no prior entanglement? If there exists such a conversion, then the qubit model and the entanglement model are effectively the same. But if this is not the case, then we have to conclude that the quantum complexity of a distributed function consists of two distinct components: 1) the amount of prior entanglement, and 2) the number of communicated bits, where the first can be much larger than the second.

We already mentioned the almost quadratic reduction in complexity for the DISJOINT function. This result is especially interesting as this function plays a role in communication complexity comparable to the SATISFIABILITY problem for computational tasks. In [4], it was shown by Babai *et al.* that it is a complete problem for the communication class co-NP^{cc}. Hence, the natural and open question of whether there exists an $O(\log n)$ quantum protocol for the DISJOINT problem is equivalent to asking "NP^{cc} $\subseteq$ BQP^{cc}?", where we already know that NP^{cc} $\nsubseteq$ BPP^{cc}.

This brings us to the most prominent open problem in in the field: *Does there exist an exponential separation between probabilistic classical and quantum communication for a function without a promise?* We know by the work of Beals *et al.* that in the black-box model of computation there can only be a polynomial difference between classical and quantum computers.[5] One wonders if this also holds for communication problems, and if so, if the 'lower bounds by polynomials' methods of this publication also translates to our setting. Recently, Harry Buhrman and Ronald de Wolf have made a significant first step in this direction[18]. Yet it is still unknown how far nonlocality will take us from traditional communication.

Section 10.3 Thresholds for Experimental Implementations

Our last discussion in this thesis will be about the experimental feasibility of simple quantum communication protocols. Various experimental results have already confirmed the nonlocal predictions of quantum mechanics.[3, 27, 65] More recent ideas like superdense coding or teleportation have also found their way into the laboratory. (See [11, 13, 28, 44] for some examples.) It is therefore natural to wonder if it is possible with current technology to implement the protocols of this thesis, and what the criteria are for a 'successful experiment'.

In communication complexity theory, we compare procedures on the basis of the amount of information that the parties have to exchange to solve a distributed task. This has the fortunate consequence that we are not concerned with the possibility of 'some kind of hidden signaling between A and B', and hence with the criterion of space-like-separation for the measurements M^A and M^B. It is perfectly in order for Alice and Bob, to have received the prior entanglement and the data a long time before the actual execution of the protocol. The only resource that counts in the context of this thesis is the number of (quantum) bits that the parties have to communicate. Hence, we can safely disregard the potential 'hidden' communication between the entangled quantum states. Not because it cannot occur, but because it does not count.

The detector efficiency, on the other hand, *does* play an important role. It increases the complexity of the quantum protocol if Alice sometimes has to inform Bob that her photon detection failed and that the experiment has to be repeated. Assume, for example, that the error-rate of the detector is so high, that there exists a classical model for the experiment that gives the same predictions (see [52] for an explanation of this possibility). Then, by the same token, there also exists a classical procedure that achieves the same correctness ratio as the quantum protocol. Hence, the existence of such a 'detector loophole' indicates that the quantum mechanical experiment does not give an improvement of over the classical lower bound. In the next sections, we derive some thresholds for the error-rate of the detectors and the noise of the entangled states, that truly separate quantum communication from its classical pendant.

Section 10.4 Thresholds for a One Bit Protocol

Consider the two-party protocol that uses one bit of communication and an entangled pair $|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. (See, for example, the protocols of Chapter 5 or Section 8.8.) Let P_q be the success rate of the quantum procedure under perfect conditions, which is higher than the classical bound P_c. We will answer two questions for this setting: "What is the amount of noise that we can allow for Φ?" and "What efficiency is required for the detection or measurement of the qubits?"

Section 10.5 Noisy Entangled States

Let $(1-\eta)$ be the probability that Alice and Bob have a random 2 qubit state rather than the desired entangled pair Φ. The density matrix Φ^η of this mixture reads in the computational basis as

$$\Phi^\eta_{AB} = (1-\eta)\cdot\begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{bmatrix} + \eta\cdot\begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}.$$

The expected probability of success, P^η_q, for the quantum procedure with this noisy state is $\eta P_q + \frac{1-\eta}{2}$ (under the worst-case assumption that the quantum protocol with the completely noisy state corresponds to a coin flip, and hence has an error rate of 50%). By the equation $P^\eta_q = P_c$, this gives the threshold on the quality of the state Ψ^η of

$$\eta > \frac{P_c - \frac{1}{2}}{P_q - \frac{1}{2}}.$$

As an aside, it is interesting to combine the η-threshold with the observation in Section 2.2 that for $\eta \leq \frac{1}{3}$ the state Φ^η is disentangled. If this is the case, then the quantum protocol can always be simulated by a classical procedure using the decomposition in tensor-products of the state Ψ_η. Hence we have also a bound on how much P_q and P_c can differ: $P_q - P_c \leq 2P_c - 1$.

Chapter 5 gave the example of a function for which $P_q = \frac{1}{2} + \frac{1}{4}\sqrt{2}$ and $P_c = \frac{3}{4}$. These values give us the criterion $\eta > \frac{1}{\sqrt{2}} \approx 0.71$ for the purity of the state Φ. This is definitely feasible for entangled photons, for which η rates close to 99% are already possible.

Section 10.6 Inexact Measurement Devices

An apparatus that tries to implement a measurement can have two kinds of errors: it gives a random answer, or it gives no answer at all. (The advantage of the second is that the party then at least knows that something went wrong.) The probability of an incorrect answer can be rephrased as a perfect measurement on a noisy state, the case which we analyzed in the previous section. A faulty measurement that produces a random outcome with probability $(1-\mu)$ is equivalent to a perfect measurement on the noisy state Φ^μ. If we combine such an imperfect measurement device with a noisy states Φ^λ, we get a total eta factor of $\eta = \mu \cdot \lambda$.

Now we will focus on the second scenario, where there is a probability $(1-\tau)$ that a party is not able to read out the (otherwise perfect) measurement outcome. This gives the probability τ^2 that the protocol is executed without any problems, $2\tau(1-\tau)$ that one of the parties does not measure an outcome, and the remaining $(1-\tau)^2$ that *both* parties do not obtain a measurement result. When a detection error occurs, Alice and Bob can adopt the strategy that they try to execute the best possible classical procedure as an alternative. This back-up plan will have a correctness probability of P_c in the case that both sides have such an error (the probability $(1-\tau)^2$). If only one party has a problem with his or her measurement, then the protocol will correspond again to a blind guess. The overall success of this approach is thus calculated by

$$P_q^\tau \quad = \quad \tau^2 \cdot P_q + (1-\tau)^2 \cdot P_c + \tau(1-\tau),$$

leading to the lower bound for the 'visibility' of the qubits

$$\tau \quad > \quad \frac{2P_c - 1}{P_q + P_c - 1}.$$

For the earlier $P_q = \frac{1}{2} + \frac{1}{4}\sqrt{2}$ and $P_c = \frac{3}{4}$, this gives the threshold of τ bigger than $2\sqrt{2} - 2 \approx 0.83$. For photon detections, the detector efficiency threshold is by far the most problematic as state-of-the-art experiments are still limited by a τ of the order of 10 to 20%.

Section 10.7 Conclusion

In this thesis we saw how one can translate the nonlocal phenomena of quantum physics into communication protocols that are more efficient than classical procedures. These results highlight the differences between classical and quantum information in way similar to that of quantum computation.

The implementation of a quantum protocol that is truly more efficient than any classical procedure is problematic because of the detector inefficiency of our current measurement devices. But, as for the detector loophole for nonlocality experiments, it is not inconceivable that in the near future a sufficiently reliable measurement device can be employed to overcome this barrier.

It is debatable if quantum communication will ever reach the status of a 'commercial application'. But even if it does not, its ideas will still remain a powerful tool to underline, explain and investigate the differences between a universe that is governed by classical laws and the one that we are living in.

Appendix A

Appendix to Holevo's Bound

'Holevo's bound' puts a limit on the amount of classical information that can be transmitted with quantum signals. One consequence of this celebrated result is the observation that a d-dimensional closed quantum system can carry no more information than a classical system of the same dimension: $\log d$ bits. More precisely, Holevo's theorem establishes an upper bound on the mutual information $I(A:B)$ between the source Bob and the receiver Alice according to $I(A:B) \leq \chi(B)$. (See [25, 61] for the notion of 'mutual information'.) This bound $\chi(B)$ is calculated as follows.

By $B = \{(p_i, \rho_i)\}_i$, we indicate a source that transmits its codewords ρ_i with probability p_i. The 'average' codeword for such a B is thus expressed by $\rho = \sum_i p_i \rho_i$. With this ρ, the *chi* quantity of a source is defined by

$$\chi(B) \quad = \quad S(\rho) - \sum_i p_i \cdot S(\rho_i),$$

where S denotes the Von Neumann entropy of a quantum mechanical mixture.

Before we extend this result, we will first take a closer look at the above theorem and try to understand *why* the information transfer is bounded by the difference between the two terms $S(\sum p_i \rho_i)$ and $\sum_i p_i S(\rho_i)$.

Section A.1 Information Transfer with Quantum States

Bob can send information if he is able to change the state of Alice in such a way that she on her side can detect this change. The bigger the state space of the change is, the more information can be transfered by it. This is captured by the positive term "$S(\sum_i p_i \cdot \rho_i) = S(\rho)$" in Holevo's bound, which expresses the randomness $S(\rho)$ that Bob can cause on Alice's side. Here we already see that the information transfer is fundamentally bounded by the dimension of the system (because $S(\rho) \leq \log(\mathrm{Dim}(\rho))$), independently of the number of messages ρ_i that Bob uses. This phenomenon is most clearly at work when we allow the whole continuum of one-qubit states $\rho(\alpha, \beta) = \alpha|0\rangle + \beta|1\rangle$. Such a source gives an uncountable infinite set of possible signals that Alice can receive, yet she will not be able to infer more than one bit of information per received signal from it. This is due to the small distinguishability between most of the messages: on paper it may seem that $|\phi\rangle = |0\rangle$ and $|\psi\rangle = \cos(\frac{1}{100})|0\rangle + \sin(\frac{1}{100})|1\rangle$ are very different, but the physical reality is that they are very similar. In every possible situation (allowed by quantum mechanics), ϕ will behave almost the same as ψ. It is therefore very hard for Alice to discover which one of the states she has received. All this is captured by the small entropy of the equal mixture $S(\frac{1}{2}|\phi\rangle\langle\phi| + \frac{1}{2}|\psi\rangle\langle\psi|) \approx 0.00042$.

The second, negative, term in Holevo's bound tells us that the randomness $S(\rho)$ has to be related to the probabilities p_i for the different messages, and not the individual entropies $S(\rho_i)$. It is not sufficient for Bob to cause a big random effect on the other party's side. For if he wants to convey a message i, there also has to be a strong, detectable correlation between the i and its carrier ρ_i. Such a correlation is indicated if the randomness of ρ disappears when we know the index i that Bob has used. But if the signals ρ_i *by themselves* are very random, *i.e.* their entropies $S(\rho_i)$ are high, then this will *decrease* the effectiveness of the source $B = \{(p_i, \rho_i)\}$. This justifies the subtraction of the sum $\sum_i p_i \cdot S(\rho_i)$.

Section A.2 Holevo's Bound versus Superdense Coding

Is superdense coding not a violation of Holevo's bound? Are we not using one qubit to transmit two bits of information? No, we are not. It is true that Bob only sends a single qubit to Alice but this signal is part of a bigger four-dimensional (2-qubit) system that *can* carry the two bits of information. This is again an example where we have to pay attention to the fact that entangled qubits should be viewed as a single system.

But is it then also possible to come up with a protocol where the parties initially share k entangled pairs, where after the transmission of one qubit more than two bits of information has been communicated? The following extension of Holevo's bound by Michael Nielsen tells us that this is not possible and that superdense coding is indeed the best we can do.

Section A.3 Holevo's Bound in the Presence of Entanglement

How much can the mutual information $I(A:B)$ of Alice increase if she receives one qubit from Bob? Let $B = \{(p_i, \rho_i)\}_i$ (with the average state $\rho = \sum_i p_i\rho_i$) be the situation before the communication of the qubit q, and $B' = \{(p_i, \rho'_i)\}$ (with $\rho' = \sum_i p_i\rho'_i$) the situation afterwards. This ρ' is the joint system of ρ and q, hence we can use both the subadditivity rule and the Araki-Lieb inequality[2], which tell us that $S(P) - S(Q) \leq S(PQ) \leq S(P) + S(Q)$ for quantum systems P and Q. In our case P is Alice's initial state ρ_i and Q is the single qubit q_i (with $0 \leq S(q_i) \leq 1$), implying the bound

$$\begin{aligned} \chi(B') &= S(\rho') - \sum_i p_i \cdot S(\rho'_i) && \text{(A.1)} \\ &\leq [S(\rho)+1] - \sum_i p_i \cdot [S(\rho)-1] \ = \ \chi(B)+2, && \text{(A.2)} \end{aligned}$$

where the square brackets indicate the application of the subadditivity/Araki-Lieb result.

Besides the above scenario, other activities during a quantum communication protocol are: 1) unitary operations on Alice or Bob's side and 2) the communication of quantum information from Alice to Bob. The entropy S of a quantum system is invariant under unitary transformations, and it also known that tracing out a qubit (which is what effectively happens on Alice's side when she sends a qubit to Bob) cannot increase the value of χ. We have thus reached the conclusion that the mutual information on Alice's side, $I(A:B)$, can only increase if she receives a qubit from the other party, and that this increase per qubit is bounded from above by 2 bits. This bound is obtained if A and B share initial entanglement and use the superdense coding protocol from Bob to Alice to send information.

Furthermore we can also analyze the scenario where the parties do not share initial entanglement. This can be rephrased as the situation where Alice starts with a fixed pure state ρ^0 and hence with $S(\rho^0) = 0$. This term $S(\rho)$ bounds the mutual information from above by $I(A:B) \leq S(\rho)$ and can only increase with one bit when Alice or Bob sends a qubit to the other party. Therefore, in this setting the total amount of communication has to be at least n bits if Alice wants to obtain n bits of mutual information from Bob. This result will be clarified for two standard protocols.

Classical communication from Bob to Alice: Alice starts with a zero register ρ^0 and every time Bob sends her a classical bit of information, the chi factor increases with one bit as $S(\rho') = S(\rho) + 1$ and $\sum_i p_i S(\rho_i')$ remains zero. After n bits of communication, this establishes $I(A:B) = n$.

Superdense coding from Bob to Alice: First, Alice distributes $\frac{n}{2}$ entangled pairs between her and Bob (we assume that n is even). This requires $\frac{n}{2}$ qubits of communication from A to B and yields the intermediate situation on her side with $S(\rho) = \sum_i p_i S(\rho_i) = \frac{n}{2}$ and hence $\chi = 0$. After that, Bob uses the entangled pairs for superdense coding, thereby reaching the bound of Equation A.2 and communicating n bits to Alice.

The main implication of this is that even if Alice is allowed to send out an unlimited amount of qubits (to create entanglement between her and other parties, for example), it will still be necessary for B to send $\lceil \frac{n}{2} \rceil$ qubits back to A to convey n bits of information to her. This immediately puts a limit on the usefulness of entanglement for information transmission: the factor 2 reduction of superdense coding is the highest possible.

This result is an expansion of Holevo's bound, as it encapsulates a more general setting where two-way communication is allowed between A and B instead of the one way communication of the earlier theorem.

Bibliography

[1] Andris Ambainis, Leonard J. Schulman, Amnon Ta-Shma, Umesh Vazirani, and Avi Wigderson. The quantum communication complexity of sampling. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS'98)*, pages 342–351, Los Alamitos, California, November 1998. IEEE Computer Society Press.

[2] Huzihiro Araki and Elliott H. Lieb. Entropy inequalities. *Communications in Mathematical Physics*, 18:160–170, 1970.

[3] Alain Aspect, Jean Dalibard, and Gérard Roger. Experimental test of Bell's inequalities using time-varying analyzers. *Physical Review Letters*, 49:1804–1807, 1982.

[4] Lászlo Babai, Phyllis G. Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *Proceedings of the 27th IEEE Symposium on Foundations of Computer Science*, pages 337–347, Los Angeles, Ca., USA, October 1986. IEEE Computer Society Press.

[5] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS'98)*, pages 352–361, Los Alamitos, California, November 1998. IEEE Computer Society Press. On the quant-ph archive, report no. 9802049.

[6] John S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.

[7] John S. Bell. *Speakable and unspeakable in quantum mechanics*. Cambridge University Press, 1987.

[8] Charles Bennett and Stephen Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69:2881–2884, 1992.

[9] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Josza, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70:1895–1899, 1993.

[10] Charles H. Bennett, Gilles Brassard, Sandu Popescu, Benjamin Schumacher, John A. Smolin, and William K. Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Physical Review Letters*, 76:722–725, 1996. On the quant-ph archive, report no 9511027.

[11] Danilo Boschi, Salvatore Branca, Francesco De Martini, Lucien Hardy, and Sandu Popescu. Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolski-Rosen channels. *Physical Review Letters*, 80:1121–1125, 1998.

[12] James Boswell. *The Life of Samuel Johnson*. Penguin Classics. Penguin, London, 1979. Edited by Christopher Hibbert.

[13] Dik Bouwmeester, Jian-Wei Pan, Klaus Mattle, Manfred Eibl, Harald Weinfurter, and Anton Zeilinger. Experimental quantum teleportation. *Nature*, 390:575, 1997.

[14] Michel Boyer, Gilles Brassard, Peter Høyer, and Alain Tapp. Tight bounds on quantum searching. *Fortschritte der Physik*, 46(4–5):493–505, 1998. On the quant-ph archive, report no. 9605034.

[15] Samuel L. Braunstein and Carlton M. Caves. Writing out better Bell inequalities. In Shun ichi Kobayashi *et al.*, editor, *Proceedings of the 3rd International Symposium on the Foundations of Quantum Mechanics in the Light of New Technology*, pages 161–170. Physical Society of Japan, 1990.

[16] Harry Buhrman, Richard Cleve, and Wim van Dam. Quantum entanglement and communication complexity. Technical Report in the BRICS Research Series RS-97-40, BRICS, University of Aarhus (Denmark), January 1998. On the quant-ph archive, report no. 9705033.

[17] Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC'98)*, pages 63–68. ACM Press, 1998. On the quant-ph archive, report no. 9705033.

[18] Harry Buhrman and Ronald de Wolf. Communication complexity lower bounds by polynomials. cs.CC report 9910010, Los Alamos archive, October 1999.

[19] Harry Buhrman, Wim van Dam, Peter Høyer, and Alain Tapp. Multiparty quantum communication complexity. *Physical Review A*, 60(4):2737–2741, October 1999. On the quant-ph archive, report no. 9710054.

[20] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, April 1988.

[21] Boris S. Cirel'son. Quantum generalizations of Bell's inequality. *Letter in Mathematical Physics*, 4:93–100, 1980.

[22] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23:880–884, 1969.

[23] Richard Cleve and Harry Buhrman. Substituting quantum entanglement for communication. *Physics Review A*, 56(2):1201–1204, August 1997. On the quant-ph archive, report no. 9704026.

[24] Richard Cleve, Wim van Dam, Michael Nielsen, and Alain Tapp. Quantum entanglement and the communication complexity of the inner product function. In Colin P. Williams, editor, *Proceedings of the 1st NASA International Conference on Quantum Computing and Quantum Communications (QCQC'98)*, number 1509 in Lecture Notes in Computer Science, pages 61–74. Springer-Verlag, 1999. On the quant-ph archive, report no. 9708019.

[25] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons, Inc., 1991.

[26] David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London A*, 439:553–558, 1992.

[27] Stuart J. Freedman and John F. Clauser. Experimental test of local hidden-variable theories. *Physical Review Letters*, 28:938–941, 1972.

[28] Akira Furusawa, Jens Lykke Sørensen, Samuel L. Braunstein, Christopher A. Fuchs, H. Jeff Kimble, and Eugene S. Polzik. Unconditional quantum teleportation. *Science*, 282(5389):706–709, 1998.

[29] Michael R. Garey and David S. Johnson. *Computers and Intractability (A guide to the theory of NP-completeness)*. W.H. Freeman and Company, New York, 1979.

[30] Daniel M. Greenberger, Michael Horne, and Anton Zeilinger. Going beyond Bell's theorem. In Menas Kafatos, editor, *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, pages 69–72. Kluwer Academic, 1989.

[31] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th Annual ACM Symposium on the Theory of Computing (STOC'96)*, pages 212–219, Philadelphia, Pennsylvania, May 1996. ACM. On the quant-ph archive, report no. 9605043.

[32] Lov K. Grover. Quantum mechanics helps in searching for a needle in a haystack. *Physical Review Letters*, 79(2):325–328, july 1997. On the quant-ph archive, report no. 9706033.

[33] Lucien Hardy. A new way to obtain Bell inequalities. *Physics Letters A*, 161:21–25, 1991.

[34] Lucien Hardy and Wim van Dam. Quantum communication using a nonlocal Zeno effect. *Physical Review A*, 59(4):2635–2640, April 1999. On the quant-ph archive, report no. 9805037.

[35] Alexander S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973. English Translation in *Problems in Information Transmission*, 9:177–183, 1973.

[36] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of mixed states: Necessary and sufficient conditions. *Physics Letters A*, 223(1–2):1–8, November 1996. On the quant-ph archive, report no. 9605038.

[37] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Mixed-state entanglement and distillation: Is there "bound" entanglement in nature? *Physical Review Letters*, 80(24):5239–5242, June 1998. On the quant-ph archive, report no. 9801069.

[38] Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM Journal on Discrete Mathematics*, 5(4):545–557, November 1992.

[39] Martin Kneser. Abschätzung der asymptotische Dichte von Summenmengen. *Mathematische Zeitschrift*, 58:459–484, 1953.

[40] Donald E. Knuth. *The Art of Computer Programming: Seminumerical Algorithms*, volume 2. Addison-Wesley, Reading, Massachusetts, third edition, 1998.

[41] Ilan Kremer. Quantum communication. Master's thesis, Computer Science Department, The Hebrew University of Jerusalem, March 1995.

[42] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambride University Press, Cambridge, United Kingdom, 1997.

[43] Henry B. Mann. *Addition theorems: The addition theorems of group theory and number theory*, volume 18 of *Interscience tracts in pure and applied mathematics*. Interscience Publishers, New York, 1965. See page 6 for a proof of Kneser's Theorem.

[44] Klaus Mattle, Harald Weinfurter, Paul G. Kwiat, and Anton Zeilinger. Dense coding in experimental quantum communication. *Physical Review Letters*, 76:4656–4659, 1996.

[45] N. David Mermin. *Boojums All the Way Through: Communicating Science in a Prosaic Age*. Cambridge University Press, Cambridge (England), 1990.

[46] N. David Mermin. What's wrong with these elements of reality? *Physics Today*, 43(6):9–11, June 1990.

[47] Roger B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge, Massachusetts, 1991.

[48] Ashwin Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proceedings of the 40th Annual Symposium on Foundations of Computer Science (FOCS'99)*, pages 369–377, New York, NY, October 1999. IEEE Computer Society Press.

[49] John von Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100:295–320, 1928. Translated in English in [50].

[50] John von Neumann. On the theory of games and strategy. In Harold W. Kuhn and Albert W. Tucker, editors, *Contributions to the Theory of Games, Volume IV*, volume 40 of *Annals of Mathematics Studies*, pages 13–42. Princeton University Press, Princeton, New Jersey, 1959. English translation of [49].

[51] John von Neumann and Oskar Morgenstern. *Theory of games and economic behavior*. Princeton University Press, Princeton, third edition, 1953.

[52] Philip M. Pearle. Hidden-variable example based upon data rejection. *Physical Review D*, 2(8):1418–1425, 1970.

[53] Asher Peres. *Quantum Theory: Concepts and Methods*, volume 72 of *Fundamental Theories of Physics*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.

[54] Asher Peres. Separability criterion for density matrices. *Physical Review Letters*, 77:1413–1415, 1996. On the quant-ph archive, report no. 9604005.

[55] Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994.

[56] Sandu Popescu and Daniel Rohrlich. The relativistic EPR argument. In Robert S. Cohen, Michael Horne, and John J. Stachel, editors, *Potentiality, entanglement and passion-at-a-distance: quantum mechanical studies for Abner Shimony, Volume Two*, volume 194 of *Boston studies in the philosophy of science*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997. On the quant-ph archive as "Action and Passion at a Distance: An Essay in Honor of Professer Abner Shimony", report no. 9605004.

[57] Ran Raz. Exponential separation of quantum and classical communication complexity. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing (STOC'99)*, pages 358–367, Atlanta, Georgia, May 1999. ACM Press.

[58] Daniel Rohrlich and Sandu Popescu. Nonlocality as an axiom for quantum theory. In Ady Mann and Micha Revzen, editors, *The dilemma of Einstein, Podolsky and Rosen, 60 years later: International symposium in honour of Nathan Rosen*, volume 12 of *Annals of the Israel Physical Society*, chapter 16. Israel Physical Society, Haifa, Israel, 1996. On the quant-ph archive, report no. 9508009.

[59] Benjamin Schumacher. Sending quantum entanglement through noisy channels. *Physical Review A*, 54(4):2614–2628, October 1996. On the quant-ph archive, report no. 9604023.

[60] Adi Shamir. Factoring numbers in $O(\log n)$ arithmetic steps. *Information Processing Letters*, 8(1):28–31, January 1979.

[61] Claude E. Shannon and Warren Weaver. *The mathematical theory of communication*. University of Illinois Press, 1949.

[62] Euan J. Squires, Lucien Hardy, and Harvey R. Brown. Nonlocality from an analogue of the quantum Zeno effect. *Studies in History and Philosophy of Science*, 25:425–435, 1994.

[63] Alan Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society, Series 2*, 42:230–265, 1937.

[64] Umesh V. Vazirani. Towards a strong communication complexity theory or generating quasi-random sequences from two communicating slightly-random sources. In *Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing*, pages 366–378, Providence, Rhode Island, May 1985. ACM Press. Extended abstract.

[65] Gregor Weihs, Thomas Jennewein, Christoph Simon, Harald Weinfurter, and Anton Zeilinger. Violation of Bell's inequality under strict Einstein locality conditions. *Physical Review Letters*, 81:5039–5043, 1998. On the quanth-ph archive, report no. 9810080.

[66] Reinhard F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Physical Review A*, 40(8):4277–4281, 1989.

[67] William K. Wootters and Wojceich H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, October 1982.

[68] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing (STOC'79)*, pages 209–213, New York, 1979. ACM Press.

[69] Andrew Chi-Chih Yao. Quantum circuit complexity. In *Proceedings of the 34th IEEE Symposium on Foundations of Computer Science*, pages 352–361, Los Alamitos, California, 1993. IEEE Computer Society Press.